Duração: 120 + 30 minutos

Justifique convenientemente todas as respostas

Pergunta 1 2 valores

Uma fábrica produz certo tipo de componentes para sondas. Resultados anteriores permitem concluir que 2% das componentes produzidas na fábrica têm defeitos. Nessa fábrica há um departamento de controlo de qualidade, onde as componentes são testadas para detetar a possível existência de defeitos. Mas o sistema de controlo de qualidade não é perfeito: 4% das componentes não são classificadas como tendo defeitos quando efectivamente têm defeitos; 5% das componentes são classificadas como tendo defeitos quando na verdade não têm defeitos.

Calcule a probabilidade de uma componente, escolhida ao acaso da produção da fábrica, ter efetivamente defeitos dado que foi classificada como não tendo defeitos.

- **Quadro de acontecimentos e probabilidades**

Evento	Probabilidade
D = componente com defeitos	$P(D) = 0.02$
CD = componente classificada como tendo defeitos	$P(CD) = ?$
	$P(\overline{CD} \mid D) = 0.04$
	$P(CD \mid \bar{D}) = 0.05$

- **Prob. pedida**

  Aplicando o teorema de Bayes, tem-se

$$
\begin{aligned}
P(D \mid \overline{CD}) &= \frac{P(\overline{CD} \mid D) \times P(D)}{P(\overline{CD})} \\
&= \frac{P(\overline{CD} \mid D) \times P(D)}{P(\overline{CD} \mid D) \times P(D) + P(\overline{CD} \mid \bar{D}) \times P(\bar{D})]} \\
&= \frac{P(\overline{CD} \mid D) \times P(D)}{P(\overline{CD} \mid D) \times P(D) + [1 - P(CD \mid \bar{D})] \times [1 - P(D)]} \\
&= \frac{0.04 \times 0.02}{0.04 \times 0.02 + (1 - 0.05) \times (1 - 0.02)} \\
&= \frac{0.04 \times 0.02}{0.9318} \\
&\simeq 0.000859.
\end{aligned}
$$

Pergunta 2 2 valores

Ao chegarem a um entroncamento, os condutores das viaturas devem — de forma independente de condutor para condutor — virar ou à esquerda ou à direita, sendo a probabilidade de virar à direita igual a 0.4 para qualquer condutor.

Determine a probabilidade de pelo menos 15 dos próximos 20 condutores virarem à direita, sabendo que pelo menos 2 destes 20 condutores efetuaram tal manobra.

- **V.a. de interesse**

  X = número de condutores que viram à direita, em 20 que chegam ao entroncamento

- **Distribuição de X**

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

  $n = 20$

  $p = P(\text{virar à direita}) = 0.4$

- **F.p. de X**

  $P(X = x) = \binom{20}{x} 0.4^x (1-0.4)^{20-x}, \quad x = 0, 1, 2, \ldots, 20$

- **Prob. pedida**

$$
\begin{aligned}
P(X \geq 15 \mid X \geq 2) &= \frac{P(X \geq 15, X \geq 2)}{P(X \geq 2)} \\
&= \frac{P(X \geq 15)}{1 - P(X < 2)} \\
&= \frac{1 - P(X \leq 14)}{1 - P(X \leq 1)} \\
&= \frac{1 - F_{binomial(20,0.4)}(14)}{1 - F_{binomial(20,0.4)}(1)} \\
&\overset{tabela}{=} \frac{1 - 0.9984}{1 - 0.0005} \\
&\simeq 0.001601.
\end{aligned}
$$

Pergunta 3	2 valores

O tempo (em horas) que decorre até à primeira avaria de um laser, usado na leitura do código de barras de artigos de supermercado, possui distribuição uniforme no intervalo $(600, b)$, com $b > 600$. Estudos prévios indicam que a probabilidade de o tempo até à primeira avaria deste laser exceder 720 horas é de 0.4.

Obtenha o valor esperado, bem como o desvio padrão do tempo até à primeira avaria de tal laser.

- **V.a. de interesse**

  X = tempo até 1a. avaria do laser (em horas)

- **Distribuição de X**

  $X \sim \text{uniforme}(600, b)$.

- **F.d.p. de X**

$$
f_X(x) = \begin{cases} \frac{1}{b-600}, & 600 \leq x \leq b \\ 0, & \text{c.c.} \end{cases}
$$

- **Obtenção de b**

$$
b \quad : \quad \begin{cases} b > 600 \\ P(X > 720) = 0.4 \Leftrightarrow \int_{720}^{b} \frac{1}{b-600}\, dx = 0.4 \Leftrightarrow \left.\frac{x}{b-600}\right|_{720}^{b} = 0.4 \Leftrightarrow b - 720 = 0.4b - 240 \\ \qquad \Leftrightarrow b = \frac{720-240}{0.6} \Leftrightarrow b = 800 \end{cases}
$$

- **Valor esperado e desvio-padrão de X**

  Uma vez que $X \sim \text{uniforme}(a = 600, b = 800)$ segue-se:

$$
\begin{aligned}
E(X) &\overset{form}{=} \frac{a+b}{2} \\
&= \frac{600+800}{2} \\
&= 700; \\
DP(X) &= \sqrt{V(X)} \\
&\overset{form}{=} \sqrt{\frac{(b-a)^2}{12}} \\
&= \frac{800-600}{\sqrt{12}} \\
&= \frac{100}{\sqrt{3}} \\
&\simeq 57.7350269.
\end{aligned}
$$

Pergunta 4	2 valores

O *input* (X) de um canal é uma variável aleatória, medida em *volt*, com distribuição uniforme discreta em $\{\frac{1}{2}, 1\}$. O *output* (Y) desse canal é também medido em *volt* e é tal que: $Y \mid X = \frac{1}{2} \sim \text{exponencial}(2)$; $Y \mid X = 1 \sim \text{exponencial}(1)$.

Determine $P\left(Y \le 1 \mid X = \frac{1}{2}\right)$ e $P(Y \le 1)$. O que pode concluir sobre a independência das variáveis aleatórias X e Y?

- **Par aleatório**

  $X = \text{input (em volt)}$

  $Y = \text{output (em volt)}$

- **Distribuição de X**

  $X \sim \text{uniforme discreta}\left(\{\frac{1}{2}, 1\}\right)$

- **F.p. de X**

$$
P(X = x) = \begin{cases} \frac{1}{2}, & x = \frac{1}{2}, 1 \\ 0, & \text{c.c.} \end{cases}
$$

- **Distribuição de Y condicional a $\{X = x\}$**

  $Y \mid X = \frac{1}{2} \sim \text{exponencial}(2)$

  $Y \mid X = 1 \sim \text{exponencial}(1)$

- **F.d.p. de Y condicional a $\{X = x\}$**

$$
f_{Y|X=1/2}(y) = \begin{cases} 2\,e^{-2y}, & y \ge 0 \\ 0, & \text{c.c.} \end{cases}
$$

$$
f_{Y|X=1}(y) = \begin{cases} e^{-y}, & y \ge 0 \\ 0, & \text{c.c.} \end{cases}
$$

- **Probabilidades pedidas**

  Ao tiramos partido da distribuição condicional de Y, dado que $X = 1/2$, e ao aplicarmos a lei da probabilidade total, obtemos sucessivamente:

$$\begin{aligned}
P(Y \le 1 \mid X = 1/2) &= \int_{-\infty}^{1} f_{Y|X=1/2}(y)\, dy \\
&= \int_{0}^{1} 2\, e^{-2y}\, dy \\
&= -\, \mathrm{e}^{-2y}\Big|_0^1 \\
&= 1 - e^{-2} \\
&\simeq 0.864665; \\
P(Y \le 1) &= P(Y \le 1|X = 1/2) \times P(X = 1/2) + P(Y \le 1|X = 1) \times P(X = 1) \\
&= \int_0^1 2\, e^{-2y}\, dy \times \frac{1}{2} + \int_0^1 e^{-y}\, dy \times \frac{1}{2} \\
&= (1 - e^{-2}) \times \frac{1}{2} + (1 - e^{-1}) \times \frac{1}{2} \\
&\simeq 0.748393.
\end{aligned}$$

  Comentário

  X e Y dizem-se duas v.a. independentes sse $F_{X,Y}(x, y) = P(X \le x, Y \le y) = P(X \le x) \times P(Y \le y)$, $\forall (x, y) \in \mathbb{R}^2$. [Consequentemente, $(Y \mid X = 1/2) \sim (Y \mid X = 1) \sim Y$.

  Ora, não só $(Y \mid X = 1/2) \not\sim (Y \mid X = 1)$, como] $P(Y \le 1) \ne P(Y \le 1 \mid X = 1/2)$. Logo, X e Y são v.a. dependentes.

Pergunta 5	2 valores

O número de computadores de determinado modelo vendidos diariamente numa loja é representado pela variável aleatória X com distribução uniforme discreta em $\{0, 1, 2, 3, 4, 5, 6\}$.

Admita que a loja só pode fazer encomendas de 60 em 60 dias e que os números de computadores daquele modelo vendidos nessa loja, em dias diferentes, são variáveis aleatórias independentes e identicamente distribuídas a X.

Qual o número mínimo de unidades do referido modelo de computador que deve existir em *stock*, no início de um desses períodos de 60 dias, por forma a que o valor aproximado da probabilidade de haver rotura deste *stock* nesse período seja no máximo de 10% ?

- **V.a.**

  X_i = número de computadores de certo modelo vendidos no dia i, $\quad i = 1, \ldots, n$

  $n = 60$

- **Distribuição, f.p., valor esperado e variância comuns**

  $X_i \overset{i.i.d.}{\sim} X \sim$ uniforme discreta$(\{0, 1, 2, 3, 4, 5, 6\})$, $\quad i = 1, \ldots, n$

  $P(X = x) = \frac{1}{7}, \quad x = 0, 1, \ldots, 6$

  $E(X_i) = E(X) = \mu = \sum_{x=0}^{6} x \times P(X = x) = \frac{0+1+\cdots+6}{7} = \frac{6\times(6+1)}{2} = 3$

  $E(X_i^2) = E(X^2) = \sum_{x=0}^{6} x^2 \times P(X = x) - 3^2 = \frac{0^2+1^2+\cdots+6^2}{7} = \frac{6\times(6+1)\times(2\times6+1)/6}{7} = 13$

$V(X_i) = V(X) = \sigma^2 = E(X^2) - E^2(X) = 13 - 3^2 = 4$

- **V.a. de interesse**

  $S_n = \sum_{i=1}^{n} X_i$ = número de computadores de certo modelo vendidos 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), temos

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

- **Valor do stock mínimo**

$$\begin{aligned}
k \quad : \quad & P(S_n > k) \leq 0.10 \\
& 1 - P(S_n \leq k) \leq 0.10 \\
& P(S_n \leq k) \geq 0.90 \\
& \Phi\left(\frac{k - n\mu}{\sqrt{n\sigma^2}}\right) \geq 0.90 \\
& \frac{k - 60 \times 3}{\sqrt{60 \times 4}} \geq \Phi^{-1}(0.90) \\
& k \geq 180 + \sqrt{240} \times 1.2816 \\
& k \geq 199.854.
\end{aligned}$$

  O menor valor de k, por forma a que o valor aproximado de $P(S_n > k)$ não exceda 0.10 é $k_{min} = 200$.

Pergunta 6 2 valores

Seja X a variável aleatória que indica o número de pacotes de rede que chegam a um *router* num dado intervalo de tempo. Admita que X tem distribuição de Poisson com parâmetro desconhecido λ ($\lambda > 0$).

Considere uma amostra casual $(x_1, \ldots, x_n)$ de dimensão $n = 100$, proveniente da população X e tal que $\sum_{i=1}^{100} x_i = 650$.

Obtenha a estimativa de máxima verosimilhança da probabilidade de chegar pelo menos um pacote nesse intervalo de tempo.

- **V.a. de interesse**

  X = número de pacotes de rede que chegam a um router num dado intervalo de tempo

- **Distribuição de X**

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

- **Parâmetro desconhecido; espaço paramétrico**

  λ

  $\Theta = \mathbb{R}^+$

- **F.p.**

  $P(X=x) \stackrel{form.}{=} e^{-\lambda}\,\frac{\lambda^x}{x!}, \quad x \in \mathbb{N}_0$

- **Amostra**

$$\underline{x} = (x_1,\ldots,x_n) \quad : \quad n = 100$$
$$\sum_{i=1}^{n} x_i = 650$$

- **Obtenção da estimativa de MV de λ**

  Passo 1 — Função de verosimilhança

$$\begin{aligned}
L(\lambda \mid \underline{x}) &= P(\underline{X} = \underline{x}) \\
&\stackrel{X_i\ indep}{=} \prod_{i=1}^{n} P(X_i = x_i) \\
&\stackrel{X_i \sim X}{=} \prod_{i=1}^{n} P(X = x_i) \\
&= \prod_{i=1}^{n} \left(e^{-\lambda}\,\frac{\lambda^{x_i}}{x_i!} \right) \\
&= e^{-n\lambda}\,\lambda^{\sum_{i=1}^{n} x_i}\,\frac{1}{\prod_{i=1}^{n}(x_i!)}, \quad \lambda \in \Theta = \mathbb{R}^+
\end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(\lambda \mid \underline{x}) = n\,\lambda + \ln(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$$

  Passo 3 — Maximização

  A estimativa de MV de λ é aqui representada por $\hat{\lambda}$ e

$$\begin{aligned}
\hat{\lambda} \quad : \quad & \begin{cases} \left.\frac{d\ln L(\lambda|\underline{x})}{d\lambda}\right|_{\lambda=\hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\frac{d^2\ln L(\lambda|\underline{x})}{d\lambda^2}\right|_{\lambda=\hat{\lambda}} < 0 & \text{(ponto de máximo).} \end{cases} \\
& \begin{cases} -n + \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}} = 0 \\ -\frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}^2} < 0 \end{cases} \\
& \begin{cases} \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \\ -\frac{n}{\bar{x}} < 0 \quad \text{(prop. verdadeira já que } n \in \mathbb{N},\ \bar{x} \geq 0\text{).} \end{cases}
\end{aligned}$$

  Passo 4 — Concretização

  Para esta amostra tem-se

$$\hat{\lambda} = \frac{650}{100} = 6.5.$$

- **Outro parâmetro desconhecido**

$$h(\lambda) = P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - e^{-\lambda}.$$

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

  A propriedade de invariância dos estimadores de MV permite-nos concluir que a estimativa de MV de $h(\lambda)$ é

$$\begin{aligned}\widehat{h(\lambda)} &= h(\hat{\lambda}) \\ &= 1-e^{-\hat{\lambda}} \\ &= 1-e^{-6.5} \\ &\simeq 0.998497.\end{aligned}$$

Pergunta 7 2 valores

Um engenheiro pretende estimar a proporção de circuitos integrados defeituosos, p. Para o efeito efectua n séries de inspeções de circuitos integrados: na série i, regista o número de inspeções (x_i) que ele necessitou efetuar até encontrar o primeiro circuito integrado defeituoso.

Dadas as caraterísticas da produção destes circuitos, é razoável assumir que as observações $(x_1,\ldots,x_n)$ resultaram de uma amostragem aleatória de uma população X que tem distribuição geométrica de parâmetro p.

Obtenha um intervalo aproximado de confiança a 95% para p, admitindo que dispõe de uma amostra com 100 observações e média amostral 5.3. Tire partido do facto de $\frac{\bar{X}-(1/p)}{\sqrt{\bar{X}(\bar{X}-1)/n}} \overset{a}{\sim} \text{normal}(0,1)$.

- **V.a. de interesse**

  X = número de inspeções a efetuar até encontrar o primeiro circuito integrado defeituoso

- **Situação**

  $X \sim \text{geométrica}(p)$

  p DESCONHECIDO

- **Obtenção do IC para p**

  Passo 1 — Seleção da v.a. fulcral

  [Dado que a $\frac{\bar{X}-(1/p)}{\sqrt{\bar{X}(\bar{X}-1)/n}}$ só depende de p e de um seu estimador e, todavia, possui distribuição aproximada completamente especificada, podemos considerar a seguinte v.a. fulcral para p:]

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

  Passo 2 — Obtenção dos quantis de probabilidade

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

$$\begin{cases} a_\alpha = \Phi^{-1}(\alpha/2) = -\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(1-0.05/2) = -\Phi^{-1}(0.975) \overset{tabelas,calc.}{=} -1.96 \\ b_\alpha = \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.96, \end{cases}$$

  que enquadram a v.a. fulcral Z com probabilidade aproximadamente igual a $1-\alpha$.

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

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

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

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

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

Passo 4 — Concretização

A expressão geral do IC aproximado para p é

$$IC_{(1-\alpha)\times 100\%}(p) \quad \simeq \quad \left[\frac{1}{\bar{X}-a_\alpha \times \sqrt{\frac{\bar{X}(\bar{X}-1)}{n}}}, \frac{1}{\bar{X}-b_\alpha \times \sqrt{\frac{\bar{X}(\bar{X}-1)}{n}}}\right].$$

Atendendo aos quantis acima e ao facto de $n = 100$ e $\bar{x} = 5.3$, o IC pretendido é

$$\begin{aligned} IC_{95\%}(p) \quad &\simeq \quad \left[\frac{1}{5.3+1.96\times\sqrt{\frac{5.3\times(5.3-1)}{100}}}, \frac{1}{5.3-1.96\times\sqrt{\frac{5.3\times(5.3-1)}{100}}}\right] \\ &= \quad [0.160367, 0.229131]. \end{aligned}$$

Pergunta 8	2 valores

Numa amostra casual de 10 baterias de lítio produzidas por determinada empresa, observou-se uma duração média de 24.125 milhares de horas. Admita que a duração (em milhares de horas) das baterias produzidas tem distribuição normal com valor esperado desconhecido e desvio padrão unitário.

Conjectura-se que a duração esperada das baterias produzidas é igual a 25 mil horas. Confronte esta conjectura com $H_1 : \mu < 25$ e decida com base no valor-p.

- **V.a. de interesse**

  X = duração (em milhares de horas) de bateria de lítio

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2 = 1)$

  μ DESCONHECIDO

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 25$

  $H_1 : \mu < \mu_0 = 25$

  Estatística de teste

  $T = \frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim_{H_0} \text{normal}(0,1)$

- **Região de rejeição de H_0**

  Trata-se de um teste unilateral inferior, logo a região de rejeição de H_0 é do tipo $W = (-\infty, c)$.

- **Decisão (com base no valor-p)**

  Atendendo a que

$$
\begin{aligned}
t &= \frac{\bar{x}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} \\
&= \frac{24.125-25}{\sqrt{\frac{1^2}{10}}} \\
&\simeq -2.77 \\
valor-p &= P(T < t \mid H_0) \\
&= \Phi(t) \\
&= \Phi(-2.77) \\
&= 1-\Phi(2.77) \\
&\overset{tabelas, calc.}{=} 1-0.9972 \\
&= 0.0028,
\end{aligned}
$$

  devemos:

  - não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.28\%$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 0.28\%$, designadamente a qualquer dos n.u.s. de 1%, 5% e 10%.

Pergunta 9 — 2 valores

Uma engenheira informática defende a hipótese H_0 de que o número de anomalias detetadas por hora, numa rede de computadores, segue uma distribuição de Poisson com variância igual a 3.

Os dados recolhidos casualmente foram sumarizados na tabela seguinte:

Número de anomalias detetadas por hora	0	1	2	≥ 3
Frequência absoluta observada	12	38	46	104
Frequência absoluta esperada sob H_0	E_1	29.86	44.82	E_4

Após ter calculado as frequências absolutas esperadas E_1 e E_4 (aproximando-as às centésimas), averigue se H_0 é consistente com este conjunto de dados. Decida com base no valor-p aproximado.

- **V.a. de interesse**

  X = número de anomalias detetadas por hora na rede de computadores

- **Hipóteses**

  $H_0 : X \sim \text{Poisson}(\lambda)$, onde $\lambda : V(X) = 3 \Leftrightarrow \lambda = 3$

  $H_1 : X \not\sim \text{Poisson}(\lambda)$

- **Estatística de teste**

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

onde:

- k = no. de classes = 4;
- O_i = freq. abs. observável da classe i;
- E_i = freq. abs. esperada sob H_0 da classe i;
- $\beta = 0$.

- **Frequência esperadas sob H_0 omissas**

$$\begin{aligned} E_1 &= n \times P(X=0 \mid H_0) \\ &= \left(\sum_{i=1}^{k} o_i\right) \times F_{Poisson(3)}(0) \\ &\overset{tabelas,calc.}{=} (12+38+46+104) \times 0.0498 \\ &\simeq 9.96 \\ E_5 &= n - \sum_{i=1}^{k} E_i \\ &= 400 - (9.96 + 29.86 + 44.82) \\ &= 115.36 \end{aligned}$$

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

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

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

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$\{0\}$	12	9.96	$\frac{(12-9.96)^2}{9.96} \simeq 0.42$
2	$\{1\}$	38	29.86	2.22
3	$\{2\}$	46	44.82	0.03
4	$\{3,4,\ldots\}$	104	115.36	1.12
		$\sum_{i=1}^{k} o_i = n$ $= 200$	$\sum_{i=1}^{k} e_i = n$ $= 200$	$t = \sum_{i=1}^{k} \frac{(o_i-e_i)^2}{e_i}$ $\simeq 3.79$

Dado que o valor observado da estatística de teste é $t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 3.79$ e $W = (c, +\infty)$, obtemos

$$\begin{aligned} valor - p &= P(T > t \mid H_0) \\ &\simeq 1 - F_{\chi^2_{(k-1)}}(t) \\ &= 1 - F_{\chi^2_{(3)}}(3.79) \\ &\simeq 0.285051 \end{aligned}$$

e devemos

- não rejeitar de H_0 a qualquer n.s. $\alpha_0 \le valor - p \simeq 28.5051\%$, designadamente aos n.u.s. $(1\%, 5\%, 10\%)$;
- rejeitar de H_0 a qualquer n.s. $\alpha_0 > valor - p \simeq 28.5051\%$.

Alternativamente e recorrendo às tabelas de quantis da distribuição do qui-quadrado, podemos obter um intervalo para o valor-p deste teste:

$$F^{-1}_{\chi^2_{(3)}}(0.70) = 3.665 \quad < \quad t = 3.79 < 4.642 = F^{-1}_{\chi^2_{(3)}}(0.80)$$
$$0.70 \quad < \quad F_{\chi^2_{(3)}}(3.79) < 0.80$$
$$0.20 = 1 - 0.80 \quad < \quad valor - p \simeq 1 - F_{\chi^2_{(3)}}(3.79) < 1 - 0.70 = 0.30.$$

Logo:

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 20\%$, nomeadamente aos n.u.s. $(1\%, 5\%, 10\%)$;
- devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 30\%$.

Pergunta 10 — 2 valores

Os registos de $n = 10$ voos de um certo tipo de avião comercial, numa mesma rota e a altitudes idênticas, indicam que o consumo de combustível (Y, em litros por hora) e a velocidade (x, em nós) foram tais que:

$$\sum_{i=1}^{10} x_i = 4\,603, \quad \sum_{i=1}^{10} x_i^2 = 2\,120\,005, \quad \sum_{i=1}^{10} y_i = 5\,985.1, \quad \sum_{i=1}^{10} y_i^2 = 3\,584\,489.89, \quad \sum_{i=1}^{10} x_i y_i = 2\,756\,640.3.$$

Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \ldots, n$.

Após ter enunciado as hipóteses de trabalho que entender convenientes, averigue se os dados permitem concluir a significância de tal modelo, ao nível de significância de 5%.

- **Modelo de RLS**

  Y_i = consumo de combustível por hora durante o voo i (v.a. resposta)

  x = velocidade durante voo i (variável explicativa)

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

- **Hipóteses de trabalho**

  $\varepsilon_i \stackrel{i.i.d.}{\sim} \text{normal}(0, \sigma^2), \quad i = 1, \ldots, n$

- **Estimativa de MV de β_1; estimativa de σ^2**

  Importa notar que

  - $n = 10$
  - $\sum_{i=1}^n x_i = 4\,603$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{4\,603}{10} = 460.3$

    $\sum_{i=1}^n x_i^2 = 2\,120\,005$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 2\,120\,005 - 10 \times 460.3^2 = 1\,244.1$
  - $\sum_{i=1}^n y_i = 5\,985.1$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{5\,985.1}{10} = 598.51$

    $\sum_{i=1}^n y_i^2 = 3\,584\,489.89$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 3\,584\,489.89 - 10 \times 598.51^2 = 2\,347.689$
  - $\sum_{i=1}^n x_i y_i = 2\,756\,640.3$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 2\,756\,640.3 - 10 \times 460.3 \times 598.51 = 1\,698.77.$

Logo,

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} \\
&= \frac{1698.77}{1244.1} \\
&\simeq 1.365461
\end{aligned}$$

$$\begin{aligned}
[\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\,\bar{x} \\
&= 598.51 - 1.365461 \times 460.3 \\
&\simeq -30.0117] \\
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2 \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&\simeq \frac{1}{10-2}\left(2347.689 - 1.365461^2 \times 1244.1\right) \\
&\simeq 3.510597
\end{aligned}$$

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$

  $H_1 : \beta_1 \neq 0$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \sim_{H_0} t_{(n-2)}$$

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

  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq 0$), pelo que a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo $W = (-\infty, -c) \cup (c, +\infty)$, onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, i.e.,

$$c = F^{-1}_{t_{(n-2)}}(1-\alpha_0/2) = F^{-1}_{t_{(10-2)}}(1-0.05/2) = F^{-1}_{t_{(8)}}(0.975) \stackrel{tabelas,calc.}{=} 2.306.$$

- **Decisão**

  O valor observado da estatística de teste é dado por

$$\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{1.365461 - 0}{\sqrt{\frac{3.510597}{1244.1}}} \\
&\simeq 25.704932.
\end{aligned}$$

  Dado que $t \simeq 25.704932 \in W = (-\infty, -2.306) \cup (2.306, +\infty)$, concluímos que devemos rejeitar a hipótese de o consumo de combustível por hora não ser influenciado pela velocidade ($H_0 : \beta_1 = 0$), quer ao n.s. de 5%, quer a qualquer outro n.s. maior que 5%.