Probabilidades e Estatística E
Ano lectivo 2014/15
1º teste - 29 de Abril de 2015

Aviso: Trata-se duma resolução sumária.

1. Considere a v.a. $X(t)$: "número de acessos em t minutos". $X(t) \sim P(4t)$.
   (a) A probabilidade de ocorrerem 3 acessos num minuto é $P(X(1) = 3) = e^{-4}\frac{4^3}{3!} = 0.195$.
   (b) A probabilidade de ocorrerem 6 acessos em dois minutos é $P(X(2) = 6) = e^{-8}\frac{8^6}{6!} = 0.122$.
   (c) $Y \sim B(5, p)$, com $p = 1 - P(X(1) = 0) = 0.9817$. Resulta que $E(Y) = 4.9084$, $V(Y) = 0.0899$ e $CV = 0.061$
   (d) $E(3Y - W) = 3E(Y) - E(W) = 13.3919$.
   $V(3Y - W) = V(3Y) + V(W) - 2Cov(3Y, W) = 9V(Y) + V(W) - 6Cov(Y, W) = \cancel{3.1424}$ 2.2535.

2. (a) $P(Y > X) = 16/45$.
   (b) $P(X = 2|X + Y = 3) = 1/4$.
   (c) $Cov(X, Y) = E(XY) - E(X)E(Y) = 1 - 1 = 0$.
   (d) Como $P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0)$, X e Y não são independentes.

3. (a) $P(X > 498) = 1 - P(X \leq 498) = 1 - P\left(\frac{X-469.9}{22} \leq \frac{498-469.9}{22}\right) = 1 - \Phi(1.28) = 0.1003$.
   (b) Seja $S = \sum_{i=1}^{5} X_i$, com $X_i \sim N(469.9, 484)$, a variável aleatória que representa o preço dos 5 computadores. Como os computadores foram comprados por 5 clientes diferentes, podemos assumir que as variáveis X_i, $i = 1, \ldots, 5$ são independentes. Então, $S \sim N(2349.5, 2420)$ (resulta das propriedades da distribuição normal). A probabilidade da loja ter uma receita inferior a 2380€, com a venda desses computadores e

$$P(S < 2380) = \Phi(0.62) = \cancel{0.1003}\ 0.7324$$

   (c) Como os computadores têm todos os mesmos componentes, podemos assumir que o preço dos 50 computadores é $50X$. A probabilidade da loja ter uma receita inferior a 23800€, com a venda desses computadores é $P(50X < 23800) = P(X < 476) = \Phi(0.28) = 0.6163$

4. (a) $\alpha = \frac{1}{625}$.
   (b)

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x^5/3125, & 0 \leq x \leq 5, \\ 1, & x > 5. \end{cases}$$

   (c) $P(X < 3) = F(3) = 0.07776$.
   Podemos concluir que 7.76% dos discos rígidos avariam dentro do período de garantia.
   (d) $x_1 = 3.7944 \quad x_2 = 4.5735$

5. Considere os acontecimentos:
G_i - "a equipa do clube A ganha o i-ésimo jogo contra a equipa do clube B";
E_i - "a equipa do clube A empata o i-ésimo jogo contra a equipa do clube B";
P_i - "a equipa do clube A perde o i-ésimo jogo contra a equipa do clube B";
D_i - "o treinador da equipa do clube A é despedido entre o $(i-1)$-ésimo jogo e o i-ésimo jogo entre as equipas rivais".

Nota: para a resolução de (a) e (b), os acontecimentos anteriores podem ser definidos sem o índice "i".

Então, $P(D_i) = 0.15$, $P(G_i|D_i) = 0.2$, $P(E_i|D_i) = 0.2$, $P(P_i|D_i) = 0.6$, $P(P_i|\overline{D}_i) = 0.4$, $P(E_i|\overline{D}_i) = 0.2$, e $P(P_i|\overline{D}_i) = 0.4$.

(a) Por exemplo, G_i e E_i, uma vez que $G_i \cap E_i$ é o acontecimento impossível.

(b) Como G_i e E_i são disjuntos, $P(G_i \cup E_i) = P(G_i) + P(E_i)$.
Usando o Teorema da probabilidade total,

$$\begin{aligned} P(G_i \cup E_i) &= P(G_i) + P(E_i) = \\ &= P(G_i|D_i)P(D_i) + P(G_i|\overline{D}_i)P(\overline{D}_i) + P(E_i|D_i)P(D_i) + P(E_i|\overline{D}_i)P(\overline{D}_i) = \\ &= 0.57 \end{aligned}$$

(c) A probabilidade pedida é dada por $P(D_{i+1}|P_i)$. Basta apresentar a expressão anterior para a resposta estar correcta. Se quisermos apresentar um valor numérico, temos de assumir uma condição adicional. Podemos assumir que os acontecimentos D e P ocorrem de modo invariante ao longo do tempo, ou seja, que $P(D_{i+1} \cap P_i) = P(D_i \cap P_i)$. Neste caso $P(D_{i+1}|P_i) = P(D_i|P_i) = \frac{P(P_i|D_i)P(D_i)}{P(P_i)} = \frac{P(P_i|D_i)P(D_i)}{1-P(G_i \cup E_i)} = 0.209$ (resulta da aplicação do teorema de Bayes).
No entanto, como também é possível que a condição que se assumiu não seja verdadeira, também se aceitou como resposta correcta $P(D_{i+1}) = P(D_i) = 0.15$, ou uma justificação da impossibilidade de apresentar um valor numérico.

Aviso: Trata-se duma resolução sumária.

1. (a) $\hat{\mu} = \bar{x} = 31\text{MB/s}$

   (b) O estimador de μ, $\bar{X}$, é um estimador centrado e consistente.

   (c) $\text{IC}_{95\%}(\mu) = ?$

   - Variável pivot: $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{a}{\sim} N(0,1)$

     A distribuição da população é desconhecida, a variância populacional é desconhecida e o tamanho de amostra $n = 51 > 30$ é suficientemente grande para permitir a aproximação da distribuição da estatística acima pela distribuição normal, através do TLC.

   -

$$
\begin{aligned}
&a: \; P(-a \leq Z \leq a) = 0.95 \;\Leftrightarrow\; P(Z \leq a) - P(Z < -a) = 0.95 \;\Leftrightarrow \\
&\Phi(a) - \Phi(-a) = 0.95 \;\Leftrightarrow\; 2\Phi(a) - 1 = 0.95 \;\Leftrightarrow\; \Phi(a) = 0.975 \\
&\;\Rightarrow\; a = z_{0.025} = 1.96 \quad \text{(tabela)}
\end{aligned}
$$

   -

$$
\begin{aligned}
&-1.96 \leq Z \leq 1.96 \;\Leftrightarrow\; -1.96 \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq 1.96 \;\Leftrightarrow \\
&-1.96\frac{S}{\sqrt{n}} - \bar{X} \leq -\mu \leq 1.96\frac{S}{\sqrt{n}} - \bar{X} \;\Leftrightarrow\; \bar{X} - 1.96\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{S}{\sqrt{n}}
\end{aligned}
$$

     Assim,

$$
\text{IC}_{95\%}(\mu) \;\equiv\; \left[\bar{X} - 1.96\frac{S}{\sqrt{n}} \;;\; \bar{X} + 1.96\frac{S}{\sqrt{n}}\right]
$$

     concretizando: $\text{IC}_{95\%}(\mu) = \left[31 - 1.96\frac{3}{\sqrt{51}} \;;\; 31 + 1.96\frac{3}{\sqrt{51}}\right] \;\equiv\; [30.18; 31.82]$

   (d)

$$
\begin{aligned}
&n: \;\; \Delta_{\text{IC}_{95\%}(\mu)} = \frac{1}{4}\Delta_{\text{IC}^{(c)}_{95\%}(\mu)} \;\Leftrightarrow \\
&\bar{X} + 1.96\frac{S}{\sqrt{n}} - (\bar{X} - 1.96\frac{S}{\sqrt{n}}) = \frac{1}{4}\big(\bar{X} + 1.96\frac{S}{\sqrt{51}} - (\bar{X} - 1.96\frac{S}{\sqrt{51}})\big) \;\Leftrightarrow \\
&2 \times 1.96\frac{S}{\sqrt{n}} = \frac{1}{4}\big(2 \times 1.96\frac{S}{\sqrt{51}}\big) \;\Leftrightarrow \\
&\frac{1}{\sqrt{n}} = \frac{1}{4\sqrt{51}} \;\Leftrightarrow\; \sqrt{n} = 4\sqrt{51} \;\Rightarrow\; n = 816
\end{aligned}
$$

   (e) - $H_0: \sigma \leq 2.8 \;\; vs. \;\; H_1: \sigma > 2.8$ (Teste unilateral direito)

- Estatística de teste: $X^2 = \frac{(n-1)S^2}{2.8^2} \overset{\text{sob } H_0}{\sim} \chi^2_{(n-1)} \equiv \chi^2_{(50)}$

  (Distribuição da população normal, com média desconhecida)
- Região de rejeição: $\alpha = 0.05$

$$R_{0.05} \equiv \left]\chi_{0.05}; +\infty\right[\equiv \left]67.5; +\infty\right[$$

- Decisão: $x^2_{obs} = \frac{(n-1)s^2}{2.8^2} = \frac{(50\times 3^2}{2.8^2} = 57.40 \notin R_{0.05}$

  Então não se rejeita H_0 ao nível de significância 5%, pelo que os dados indicam que o desvio padrão não é superior a 2.8.

2. (a) Após identificar as sequências ascendentes e descendentes e corrigir a dimensão da amostra, temos

41	30	28	41	28	26	28	41	30	34
·	−	−	+	−	−	+	+	−	+
40	36	30	20	43	36	36	20	42	43
+	−	−	−	+	−	=	−	+	+
42	40	32	26	28	41	34	24	42	40
−	−	−	−	+	+	−	−	+	−
28	41	34	46	48	21	24	22	44	41
−	+	−	+	+	−	+	−	+	−

$n = 40$ que, após correcção, $n = 39$, existindo um total de 23 sequências.

Como $n = 39 > 25$ usamos as propriedades assintóticas do teste, vindo para valor da estatística do teste:

$$z_{obs} = \frac{23 - (2\times 39 - 1)/3}{\sqrt{(16\times 39 - 29)/90}} = -1.03713 \simeq -1.04.$$

Como a estatística de teste é assintoticamente Normal reduzida quando a hipótese nula é verdadeira, e estamos perante um teste bilateral, temos, com $Z \sim N(0,1)$:

$$\mathrm{p}-value = 2\times min\{\mathbb{P}(Z < -1.04), \mathbb{P}(Z > -1.04)\} = 2\times\mathbb{P}(Z < -1.04) \simeq 2(1-0.8508) = 29.84\%.$$

Decisão: Não rejeitar H_0 para o nível de significância referido.

(b) Recorrendo às classes da sugestão teríamos:

Classe	Oi	pi	Ei	(Oi-Ei)^2/Ei
]-∞, 25]	6	0,10565	4,226	0,745
]25, 30]	10	0,16034	6,413	2,006
]30, 35]	4	0,23401	9,361	3,070
]35, 40]	6	0,23401	9,361	1,206
]40, 45]	12	0,16034	6,413	4,866
]45, ∞[	2	0,10565	4,226	1,173
	40	**1,0000**	**40**	**13,0656**

alpha=	10,00%
parâmetros =	
graus de liberdade =	
Região Crítica =	]9,23635
p-value =	2,28%

Mas observa-se que a primeira e a última classe não verificam as condições exigidas para a realização do teste não-paramétrico do Qui-quadrado. Assim, usando o agrupamento mais natural, com as classes vizinhas (ver tabela na página seguinte), temos $x^2_{obs} = 8.0387$.

Decisão: Rejeitar H_0 para o nível de significância proposto.

Classe	Oi	pi	Ei	(Oi-Ei)^2/Ei
]-∞, 30]	16	0,26599	10,64	2,701
]30, 35]	4	0,23401	9,361	3,070
]35, 40]	6	0,23401	9,361	1,206
]40, ∞[	14	0,26599	10,64	1,061
	40	1,0000	40	8,0387

alpha=	10,00%
parâmetros =	
graus de liberdade =	
Região Crítica =	]7,77944(
p-value =	9,02%

3. (a) Dado que $\hat{\lambda}_A \sim E\left(\lambda, \frac{1}{n}\right)$ tem-se $E(\hat{\lambda}_A) = \lambda + \frac{1}{n} \neq \lambda$, logo $\hat{\lambda}_A$ não é estimador centrado de λ. Como $E(\hat{\lambda}_B) = E(\bar{X} - 1) = \cdots = E(X) - 1 = \lambda + 1 - 1 = \lambda$ então $\hat{\lambda}_B$ é estimador centrado de λ.

   (b) $EQM\left(\hat{\lambda}_A\right) = V\left(\hat{\lambda}_A\right) + \left(E\left(\hat{\lambda}_A\right) - \lambda\right)^2 = \frac{1}{n^2} + \left(\lambda + \frac{1}{n} - \lambda\right)^2 = \frac{2}{n^2}$.

   Como $\hat{\lambda}_B$ é estimador centrado de λ e as variáveis aleatórias $X_1, X_2, \ldots, X_n$ são i.i.d., então $EQM\left(\hat{\lambda}_B\right) = V\left(\hat{\lambda}_B\right) = V(\bar{X} - 1) = \cdots = \frac{V(X)}{n} = \frac{1}{n}$.

   (c) $P\left(-c \leq E \leq c\right) = 0.90 \Rightarrow P\left(E > c\right) = 0.05 \Rightarrow c \approx z_{0.05} = \Phi^{-1}(0.95) = 1.64$

   $$-1.64 \leq \sqrt{100}\left(\hat{\lambda}_B - \lambda\right) \leq 1.64 \Leftrightarrow \cdots \Leftrightarrow \hat{\lambda}_B - \frac{1.64}{10} \leq \lambda \leq \hat{\lambda}_B + \frac{1.64}{10}$$

   $$IC_{90\%}\left(\lambda\right) \equiv \left[\hat{\lambda}_B - \frac{1.64}{10}, \hat{\lambda}_B + \frac{1.64}{10}\right]$$

   Dado $\hat{\lambda}_B = \bar{x} - 1 = 1.86247$, tem-se $IC_{90\%}\left(\lambda\right) \equiv \left[1.86247 - \frac{1.64}{10}, 1.86247 + \frac{1.64}{10}\right] = [1.69847, 2.02647]$.

4. (a) Temos $S_{xx} = 60$, $S_{xy} = 23.79$ e $S_{yy} = 9.5102$. Então $\hat{\beta}_1 = 0.3965$, $\hat{\beta}_0 = -0.4369$ e $\hat{\sigma}^2 = 0.0110696$.

   (b) Pretende-se testar $H_0 : \beta_1 \leq 0 \;\; vs. \;\; H_1 : \beta_1 > 0$ (Teste unilateral direito)
   A estatística de teste é $T = \frac{\hat{\beta}_1 - 0}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \underset{Sob\, H_0 \text{ com } \beta_1 = 0}{\sim} t_{n-2} \equiv t_7$;
   O valor observado da estatística de teste é $t_{obs} = 29.191$;
   Região de rejeiçãop para o nível de significancia de 1%: $]t_{7:0.01}, \infty[=]3, \infty[$;
   Logo, rejeitamos H_0 ao nível de significância 1%.

   (c) $\hat{y}|_{x=6.5} = 2.14$. O valor $\hat{y}|_{x=0}$ não pode ser calculado porque o modelo de regressão linear simples pode não ser válido quando $x < 1$ ou $x > 9$ (e neste caso não é válido).

Aviso: Trata-se duma resolução sumária.

1. (a) **i.** Dado X a procura mensal em toneladas, então $X \sim N(80, 10^2)$, logo $P(X > 90) = 1 - \Phi\left(\frac{90-80}{10}\right) = 1 - 0.8413 = 0.1587$

ii. $P(75 < X < 90) = \Phi\left(\frac{90-80}{10}\right) - \Phi\left(\frac{75-80}{10}\right) = \Phi(1) - (1 - \Phi(0.5)) = 0.8413 - (1 - 0.6915) = 0.5328$

iii. Seja h a produção que a empresa terá de garantir. Assim, $P(X > h) = 0.01 \Leftrightarrow P(X \leq h) = 0.99 \Leftrightarrow \Phi\left(\frac{h-80}{10}\right) = 0.99 \Leftrightarrow \frac{h-80}{10} = 2.33 \Leftrightarrow h = 103.3$

iv. Sejam as variáveis aleatórias X_i, procura em toneladas no mês i, $i = 1, ..., 36$. Assim $T = \sum_{i=1}^{36} X_i$ representa a procura total, em toneladas, durante os próximos 3 anos.
Dado que T é uma combinação linear de variáveis independentes e com distribuição normal, então

$$T \sim N\left(E(T), V(T)\right)$$

onde

$$E(T) = E\left(\sum_{i=1}^{36} X_i\right) = \sum_{i=1}^{36} E(X_i) = 36 \times 80 = 2880$$

e, como as variáveis $X_1, ..., X_{36}$ são independentes, tem-se

$$V(T) = V\left(\sum_{i=1}^{36} X_i\right) = \sum_{i=1}^{36} V(X_i) = 36 \times 100 = 3600$$

logo

$$P(T > 2950) = 1 - \Phi\left(\frac{2950-2880}{\sqrt{3600}}\right) = 1 - 0.8790 = 0.121$$

(b) Seja Y a variável aleatória que contabiliza o número de meses, em 12 meses, onde a procura foi inferior a 75 toneladas. Como a procura é independente de mês para mês, então Y contabiliza o número de sucessos em 12 provas de Bernoulli independentes com probabilidade de sucesso, $p = P(X < 75) = \Phi\left(\frac{75-80}{10}\right) = 0.3085$, constante. Assim

$$Y \sim B(12, 0.3085)$$

e

$$E(Y) = 12 \times 0.3085 = 3.702.$$

2. W - Número de computadores que chegam por dia para serem reparados

$W \sim$ Poisson(λ), $E[W] = \lambda = 1$, donde $W \sim$ Poisson(1) e

$$P(W = w) = \frac{e^{-1}(1)^w}{w!} = \frac{e^{-1}}{w!}, \quad w = 0, 1, 2, \ldots$$

(a) i. $P(W = 0) = \frac{e^{-1}}{0!} \simeq 0.37$.

ii. $P(W = 1) = \frac{e^{-1}}{1!} \simeq 0.37$.

iii. $P(W \geq 2) = 1 - P(W < 2) = 1 - \Big(P(W = 0) + P(W = 1)\Big) \simeq 1 - (0.37 + 0.37) = 0.26$.

(b) Representem, respetivamente, W_1, W_2 e W_3 o número de computadores que chegaram à loja ontem, há dois dias e há 3 dias. Sabe-se que W_1, W_2 e W_3 são independentes e que todas seguem distribuição Poisson(1). Pretendemos obter o valor de

$$P(W_1 = 1 | W_1 + W_2 + W_3 = 4) = \frac{P(W_1 = 1, W_1 + W_2 + W_3 = 4)}{P(W_1 + W_2 + W_3 = 4)}$$

Relativamente à probabilidade que se encontra no numerador da fracção anterior, note que

$$\begin{aligned} P(W_1 = 1, W_1 + W_2 + W_3 = 4) &= P(W_1 + W_2 + W_3 = 4 | W_1 = 1) P(W_1 = 1) = \\ &= P(W_2 + W_3 = 3) P(W_1 = 1). \end{aligned}$$

Alternativamente, também podemos obter o mesmo resultado fazendo

$$\begin{aligned} P(W_1 = 1, W_1 + W_2 + W_3 = 4) &= P(W_1 = 1, W_1 + W_2 + W_3 - W_1 = 4 - 1) = \\ &= P(W_1 = 1) P(W_2 + W_3 = 3). \end{aligned}$$

Então resulta que

$$\begin{aligned} P(W_1 = 1 | W_1 + W_2 + W_3 = 4) &= \frac{P(W_2 + W_3 = 3) P(W_1 = 1)}{P(W_1 + W_2 + W_3 = 4)} = (*) = \\ &= \frac{\frac{e^{-2}2^3}{3!} \cdot \frac{e^{-1}1^1}{1!}}{\frac{e^{-3}3^4}{4!}} \simeq 0.40 \end{aligned}$$

$(*)$ Pela propriedade aditiva da Poisson, $W_2 + W_3 \sim$ Poisson$(1 + 1) \equiv$ Poisson(2) e $W_1 + W_2 + W_3 \sim$ Poisson$(1 + 1 + 1) \equiv$ Poisson(3).

(c) $X-$ número de computadores que são recebidos, por dia, na loja; Pela capacidade da loja só pode tomar os valores 0, 1 ou 2. Temos que:

$P(X = 0) = P(W = 0) \underbrace{\simeq}_{(a)} 0.37$;

$P(X = 1) = P(W = 1) \underbrace{\simeq}_{(a)} 0.37$;

$P(X = 2) = P(W \geq 2) \underbrace{\simeq}_{(a)} 0.26$;

Então a função de probabilidade de X é dada por:

$$X \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ 0.37 & 0.37 & 0.26 \end{array} \right.$$

O seu valor esperado é dado por:

$$E[X]=\sum_{x=0}^{2} xP(X=x)=0\times 0.37+1\times 0.37+2\times 0.26=0.89$$

A sua variância é dada por:

$$V(X)=E[X^2]-E^2[X]=\sum_{x=0}^{2} x^2P(X=x)-0.89^2=$$
$$0^2\times 0.37+1^2\times 0.37+2^2\times 0.26-0.89^2=1.41-0.89^2\simeq 0.62$$

A moda de X é o seu valor mais provável, desde que esse valor seja único. Como neste caso são dois os valores mais prováveis, 0 e o 1, dizemos que a moda não existe.

(d) Y - número de computadores que são reparados com sucesso, por dia, na loja.

- Porque os computadores são reparados (com sucesso ou não) e entregues no próprio dia que são recebidos na loja, e porque só se recebe no máximo 2 computadores, Y só poderá tomar os valores 0, 1 e 2;
- Porque não se pode reparar (com sucesso ou não) mais computadores do que os recebidos, $P(Y=y|X=x)=0$ para $y>x$, o que implica que $P(X=x,Y=y)=0$ para $y>x$, ou seja $P(X=0,Y=1)=0$, $P(X=0,Y=2)=0$, $P(X=1,Y=2)=0$;
- Porque $P(X=0)=P(X=0,Y=0)+P(X=0,Y=1)+P(X=0,Y=2)\Leftrightarrow 0.37=P(X=0,Y=0)+0+0\Leftrightarrow P(X=0,Y=0)=0.37$;
- De enunciado P(X=2,Y=2)=0.1;
- De enunciado $P(X=2,Y=0)=P(X=2,Y=1)=p$. Porque $P(X=2)=P(X=2,Y=0)+P(X=2,Y=1)+P(X=2,Y=2)\Leftrightarrow 0.26=p+p+0.1\Leftrightarrow p=0.08$
- De enunciado, $E[XY]=0.56$. Então:

$$E[XY]=\sum_{x=0}^{2}\sum_{y=0}^{2} xyP(X=x,Y=y)=0.56\Leftrightarrow$$
$$1\times 1\times P(X=1,Y=1)+1\times 2\times 0+2\times 1\times 0.08+2\times 2\times 0.10=0.56\Leftrightarrow$$
$$P(X=1,Y=1)=0$$

- Finalmente, $P(X=1,Y=0)=P(X=1)-P(X=1,Y=1)-P(X=1,Y=2)\Leftrightarrow P(X=1,Y=0)=0.37$

Então a função de probabilidade conjunta de (X,Y) é dada por:

$X\backslash Y$	0	1	2	$P(X=x)$
0	0.37	0	0	0.37
1	0.37	0	0	0.37
2	0.08	0.08	0.10	0.26
$P(Y=y)$	0.82	0.08	0.10	1

3. **a)** Após identificar as sequências ascendentes e descendentes e corrigir a dimensão da amostra, temos

3	7	5	3	5	6	10	5	6	3	6	4	9	8	4
·	+	-	-	+	+	+	-	+	-	+	-	+	-	-
3	6	5	9	8	5	3	5	3	7	5	4	5	6	9
-	+	-	+	-	-	-	+	-	+	-	-	+	+	+

$n = 30$ e não necessita de correcção, existe um total de 19 sequências. Como $n = 30 > 25$ usamos as propriedades assintóticas do teste, vindo para valor da estatística do teste:

$$t_0 = \frac{19 - (2 \times 30 - 1)/3}{\sqrt{(16 \times 30 - 29)/90}} = -0.297812.$$

Como a estatística de teste é assintoticamente Normal reduzida quando a hipótese nula é verdadeira, e estamos perante um teste bilateral, temos, com $Z \sim N(0,1)$: Vindo para região crítica ou de rejeição:

$$R_{10\%}(H_0) =]-\infty, -1.645[\cup]1.645, +\infty[.$$

Decisão: Não rejeitar H_0 para o nível de significância referido.

$$\begin{aligned}\mathrm{p}-value &= 2 \times min\{\mathbb{P}(Z < -0.297812), \mathbb{P}(Z > -0.297812)\} = 2 \times \mathbb{P}(Z > |-0.297812|) \\ &\simeq 2(1 - 0.6179) = 2 \times 0.3821 = 76.42\%.\end{aligned}$$

Decisão: Não rejeitar H_0 para o nível de significância referido.

b)Podemos usar os estimadores centrados e consistentes, média e variância amostrais,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \quad S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{X}_n^2\right)$$

que, para esta amostra, nos permite obter:

$$\hat{\mu} = \bar{x}_{30} = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{167}{30} = 5.56667$$

$$\hat{\sigma}^2 = s_{30}^2 = \frac{1}{29}\left(\sum_{i=1}^{30} x_i^2 - 30\bar{x}_{30}^2\right) = 4.18506.$$

c) Neste caso pretendemos testar: $H_0 : \mu \leq 5$ vs $H_1 : \mu > 5$.

Como desconhecemos a distribuição, mas estamos perante uma amostra aleatória de dimensão igual a 30 de uma distribuição com valor esperado e variância, podemos apelar ao Teorema Limite Central e usar como estatística de teste

$$T(\vec{X}^n) = \frac{(\bar{X}_n - \mu_0)}{S_n}\sqrt{n} \overset{a}{\sim} N(0,1) \quad (sob\, H_0)$$

neste caso, temos, para valor da estatística de teste

$$t(\vec{x}^n) = \frac{(\bar{x}_n - 5)}{s_n}\sqrt{30} = \frac{(5.56667 - 5)}{\sqrt{4.18506}}\sqrt{30} = 1.51719$$

e para região crítica ou de rejeição:

$$R_{10\%}(H_0) =]1.28, +\infty[.$$

Decisão: Rejeitar H_0 para o nível de significância referido.

d) Como temos um teste unilateral direito.

$$\mathrm{p}-value = \mathbb{P}(Z > 1.51719) = 1 - \mathbb{P}(Z \leq 1.51719) \simeq 1 - 0.9357 = 6.43\%.$$

e) Podemos definir, como variável aleatória lucro

$$L = L(X) = 200 \times 6000 - 200000 \times X = 1200000 - 200000X,$$

cujo valor médio é igual a $\mu_L = \mathbb{E}(L(X)) = 1200000 - 200000\mathbb{E}(X)$.

Como $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, é um estimador centrado e consistente do tempo médio de instalação de cada um dos sistemas com 200 processadores, podemos estimar μ_L recorrendo ao estimador:

$$\hat{\mu}_L = 1200000 - 200000\,\bar{X}.$$

Como $\mathbb{E}(\hat{\mu}_L) = 1200000 - 200000\mathbb{E}(\bar{X}_n) = 1200000 - 200000\mathbb{E}(X)$ podemos facilmente concluir que o estimador é centrado para μ_L. Como $\mathbb{V}(\hat{\mu}_L) = 200000^2\mathbb{V}(\bar{X}_n) = 200000^2\frac{1}{n}\mathbb{V}(X)$, resulta que o estimador é consistente.

Deste modo, a estimativa centrada e consistente do lucro esperado é:

$$\hat{\mu}_L = 1200000 - 200000 \times 5.56667 = 86666.$$

4. (a) $R^2 = \hat{\beta}_1^2 \frac{S_{xx}}{S_{yy}} = 0.9919$. Logo, como o coeficiente de determinação está bastante próximo do valor 1, podemos concluir que o modelo de regressão linear simples se ajusta bem aos dados.

   (b) O IC vai ser deduzido a partir da variável pivot,

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} = \sqrt{S_{xx}}\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}} \sim t_{n-2}$$

   Considerando o nível de confiança $1-\alpha$, temos que

$$P\left(-t_{n-2:\alpha/2} \leq \sqrt{S_{xx}}\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}} \leq t_{n-2:\alpha/2}\right) = \ldots =$$

$$= P\left(\hat{\beta}_1 - t_{n-2:\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{n-2:\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right) = 1-\alpha$$

   Logo, o intervalo 95% ($\alpha = 0.05$) de confiança para β_1 é,

$$IC_{95\%}(\beta_1) \equiv \left[\hat{\beta}_1 - t_{n-2:0.025}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}};\ \hat{\beta}_1 + t_{n-2:0.025}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right]$$

   Como $t_{7:0.025} = 2.36$, o intervalo de 95% de confiança para β_1 é

$$IC_{95\%}(\beta_1) = [0.3644445, 0.4285555]$$

   (c) $\mathrm{E}(\hat{\beta}_1) = \mathrm{E}\left(\sum_{i=1}^{n} \frac{(x_i-\bar{x})}{S_{xx}}Y_i\right) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})}{S_{xx}}\mathrm{E}(Y_i) = \ldots = \frac{\beta_1 S_{xx}}{S_{xx}} = \beta_1$, logo $\hat{\beta}_1$ é um estimador centrado para β_1.

5. $P(\overline{A}\cap\overline{B}) = 1 - P(A\cup B) = 1 - \big(P(A)+P(B)-P(A\cap B)\big)$. Logo $P(A)+P(B)-P(A\cap B) = 0.2$
   $P(A-B)+P(B-A) = P(A)+P(B)-2P(A\cap B) = 0.2 - P(A\cap B)$
   Como $P(A-B)+P(B-A) = 0.05$, resulta que $0.05 = 0.2 - P(A\cap B)$. Logo, $P(A\cap B) = 0.15$