Probabilidade e Estatística

LEAN/LEM/LEAmb/LEGM LEIC-A LEIC-T LERC-LEE LEAer LEBiol LEBiom LEEC LEMec LEQ

2º Semestre – 2021/2022
06/07/2022
18:00-20:00

Duração: **120** minutos **Exame Época Normal - (c)**

Justifique convenientemente todas as respostas

Pergunta 1	2 valores

É sabido que 3.4% das componentes electrónicas de um certo tipo apresentam defeitos e que as componentes electrónicas são produzidas pelos fabricantes 1 e 2 nas proporções de 70% e 30%, respetivamente.

Calcule a probabilidade de uma componente produzida pelo fabricante 2 apresentar defeitos sabendo que a probabilidade de uma componente produzida pelo fabricante 1 apresentar defeitos é o dobro da correspondente probabilidade para uma componente produzida pelo fabricante 2.

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

Acontecimento	**Probabilidade**
A_1 = "a componente foi produzida pelo fabricante 1"	$P(A_1) = 0.7$
A_2 = "a componente foi produzida pelo fabricante 2"	$P(A_2) = 0.3$
D = "a componente apresenta defeitos"	$P(D) = 0.034$
	$P(D \mid A_2) = \ ?$
	$P(D \mid A_1) = 2 \times P(D \mid A_2)$

- **Cálculo da probabilidade pedida**

$$P(D) = P(A_1) \times P(D \mid A_1) + P(A_2) \times P(D \mid A_2) \qquad \text{(lei da probabilidade total)}$$

$$\Leftrightarrow \quad P(D) = P(A_1) \times 2 \times P(D \mid A_2) + P(A_2) \times P(D \mid A_2)$$

$$\Leftrightarrow \quad P(D \mid A_2) = \frac{P(D)}{2 \times P(A_1) + P(A_2)}$$

$$\Leftrightarrow \quad P(D \mid A_2) = \frac{0.034}{2 \times 0.7 + 0.3} = 0.02.$$

Pergunta 2	2 valores

Admita que a variável aleatória X representa o número de doentes que chegam, por dia, às urgências de um hospital da região de Lisboa é uma variável aleatória com distribuição de Poisson. Indique o valor mínimo (use as tabelas) para a média das chegadas por dia quando se observam pelo menos 10 chegadas às urgências com uma probabilidade mínima de 0.8.

- **V.a. de interesse**
  X = número de doentes que chegam, por dia, às urgências do hospital.

- **Distribuição**

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

- **Cálculo do parâmetro λ**

$$P(X \geq 10) = 1 - F_X(9) \geq 0.8, \text{ isto é: } \min \lambda : F_X(9) \leq 0.2, \text{o que implica que } \lambda = 13.$$

Pergunta 3 — 2 valores

O nível de glicose no sangue, em pessoas adultas, pode ser considerado como uma variável aleatória com distribuição normal. Sabendo que o valor médio é de 80 mg/ml e que 30% das pessoas adultas têm níveis entre 80 mg/ml e 120 mg/ml, determine $E(3X^2 + \sqrt{2})$.

- **V.a. de interesse**
  X = Valor (mg/ml) da glicose no sangue.

- **Distribuição**
  $X \sim \text{normal}(\mu = 80, \sigma^2)$.
  Sabe-se que:

$$P(80 \leq X \leq 120) = 0.3.$$

- **Cálculo de σ**

  O valor de σ é determinado usando a relação:

$$\begin{aligned} P(80 \leq X \leq 120) = 0.3 &\iff P\left(\frac{80-80}{\sigma} \leq \frac{X-80}{\sigma} \leq \frac{120-80}{\sigma}\right) = 0.3 \\ &\iff P\left(0 \leq Z \leq \frac{40}{\sigma}\right) = 0.3 \\ &\iff \Phi\left(\frac{40}{\sigma}\right) - \Phi(0) = 0.3 \iff \sigma = \frac{40}{\Phi^{-1}(0.8)} \approxeq 47.52. \end{aligned}$$

- **Cálculo de $E(3X^2 + \sqrt{2})$**

$$\begin{aligned} E(3X^2 + \sqrt{2}) &= 3E(X^2) + \sqrt{2} \\ &= 3\left(V(X) + [E(X)]^2\right) + \sqrt{2} \\ &= 3(\sigma^2 + 80^2) + \sqrt{2} \\ &= 3 \times 47.52^2 + 3 \times 80^2 + \sqrt{2} \\ &\approxeq 25975.87. \end{aligned}$$

Pergunta 4	2 valores

Considere que o par aleatório (X,Y) tem função densidade de probabilidade:

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{3x}{2}, & 0<x<1, 0<y<2x, \\ 0, & \text{outros casos.} \end{cases}$$

Calcule o valor esperado de $Y|X=x$, $0<x<1$. Serão X e Y variáveis aleatórias independentes?

- **Função densidade de probabilidade marginal de** X

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\, dy.$$

Se $0<x<1$,

$$\begin{aligned} f_X(x) &= \int_0^{2x} \frac{3x}{2}\, dy \\ &= \frac{3x}{2} \times 2x \\ &= 3x^2. \end{aligned}$$

Pode então escrever-se:

$$f_X(x) = \begin{cases} 3x^2, & 0<x<1, \\ 0, & \text{outros casos.} \end{cases}$$

- **Função densidade de probabilidade marginal de** $Y|X=x$

Seja x um valor tal que $0<x<1$, então

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}.$$

Se $0<y<2x$,

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{3x}{2\times 3x^2} \\ &= \frac{1}{2x}. \end{aligned}$$

Pode então escrever-se:

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

De onde se conclui que, para $0<x<1$, $Y|X=x \sim \text{uniforme}(0,2x)$.

- **Cálculo de** $\mathrm{E}(Y|X=x)$

$$\begin{aligned} \mathrm{E}(Y|X=x) &= (2x+0)/2 \\ &= x. \end{aligned}$$

Em alternativa,

$$\begin{aligned}
\mathrm{E}(Y|X=x) &= \int_{I\!R} y f_{Y|X=x}(y)\,dy \\
&= \int_0^{2x} \frac{y}{2x}\,dy \\
&= \frac{1}{2x}\left(\frac{y^2}{2}\right)\Bigg|_{y=0}^{y=2x} = x.
\end{aligned}$$

- **Possível independência entre X e Y**

$$f_Y(y) = \int_{I\!R} f_{(X,Y)}(x,y)\,dx.$$

Se $0<y<2$,

$$\begin{aligned}
f_Y(y) &= \int_{y/2}^{1} \frac{3x}{2}\,dx \\
&= \left[\frac{3x^2}{4}\right]_{y/2}^{1} \\
&= \frac{3}{4}\left(1-\frac{y^2}{4}\right).
\end{aligned}$$

Pode então escrever-se:

$$f_Y(y) = \begin{cases} \frac{3}{4}\left(1-\frac{y^2}{4}\right), & 0<y<2, \\ 0, & \text{outros casos,} \end{cases}$$

Assim sendo, X e Y são v.a. dependentes já que

$$\exists (x,y):\ f_{(X,Y)}(x,y) \neq f_X(x)\, f_Y(y).$$

Por exemplo, se $x=y=\frac{1}{2}$, então $\frac{1}{2}<2\times\frac{1}{2}$, o que nos garante que $0<y<2x$ e $0<x<1$. Assim sendo,

$$f_{(X,Y)}\left(\frac{1}{2},\frac{1}{2}\right) = \frac{3}{4} \neq \frac{3}{4}\times\frac{3}{4}\left(1-\frac{1}{16}\right),$$

uma vez que

$$f_X\left(\frac{1}{2}\right) = \frac{3}{4},\ \text{e}\ f_Y\left(\frac{1}{2}\right) = \frac{3}{4}\left(1-\frac{1}{16}\right).$$

Pergunta 5	2 valores

O tempo, em minutos, de reparação de um certo tipo de avaria numa linha de montagem é bem modelado pela distribuição exponencial de valor esperado 30 minutos. Calcule a probabilidade aproximada de o tempo total de reparação de 35 avarias do referido tipo, ocorridas de forma independente, exceder 18 horas.

- **V.a.** X_i

  Seja X_i a va tempo de reparação da i-ésima avaria, $i = 1, \ldots, n$, com $n = 35$.

- **Distribuição**

  $X_i \sim \exp(\lambda)$, $i = 1, \ldots, n$ são variáveis aleatórias iid.

- **Valor esperado de** X_i
  $\mathrm{E}(X_i) = 30 = 1/\lambda$, logo $\lambda = 1/30$.

- **Variância de** X_i
  $\mathrm{Var}(X_i) = 1/\lambda^2 = 30^2 = 900$.

- **V.a. de interesse**
  $S_n = \sum_{i=1}^n X_i$, tempo total de reparação de n avarias.

- **Valor esperado de** S_n

  $\mathrm{E}(S_n) = \mathrm{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{E}(X_i) = n/\lambda = 35 \times 30 = 1050$.

- **Variância de** S_n

  $\mathrm{Var}(S_n) = \mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) = n/\lambda^2 = 35 \times 900 = 31500$.

- **Distribuição aproximada de** S_n
  De acordo com o teorema do limite central, temos

$$\frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n/\lambda}{\sqrt{n/\lambda^2}} \overset{a}{\sim} \text{normal}(0,1).$$

- **Valor aproximado da probabilidade pedida**

$$\begin{aligned}
P\left(\sum_{i=1}^{35} X_i > 18 \times 60\right) &= 1 - P\left(\sum_{i=1}^{35} X_i \leq 1080\right) \\
&= 1 - P\left(\frac{S_n - n/\lambda}{\sqrt{n/\lambda^2}} \leq \frac{1080 - n/\lambda}{\sqrt{n/\lambda^2}}\right) \\
&= 1 - P\left(\frac{S_n - 1050}{\sqrt{31500}} \leq \frac{1080 - 1050}{\sqrt{31500}}\right) \\
&\overset{TLC}{\simeq} 1 - \Phi(0.17) \\
&= 1 - 0.5675 \\
&= 0.4325.
\end{aligned}$$

Pergunta 6	2 valores

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

$$f_X(x;\theta) = \begin{cases} \frac{x^3}{6\theta^4} e^{-\frac{x}{\theta}}, & x > 0, \\ 0, & \text{outros casos,} \end{cases}$$

com $\theta > 0$ desconhecido. A concretização de uma amostra aleatória de dimensão 4 proveniente de X conduziu a $x_1 = 3.6$, $x_2 = 2.3$, $x_3 = 4.2$ e $x_4 = 1.7$. Obtenha a estimativa de máxima verosimilhança do parâmetro θ.

- Seja $\underline{x} = (x_1, x_2, x_3, x_4)$ uma amostra aleatória de dimensão 4 proveniente da população X.

- **Obtenção do estimador da máxima verosimilhança de θ**
  Passo 1 - Função de verosimilhança

$$
\begin{aligned}
L(\theta|\underline{x}) &= f_{\underline{X}}(\underline{x}) \\
&\overset{X_i\ indep}{=} \prod_{i=1}^{4} f_{X_i}(x_i) \\
&\overset{X_i \sim X}{=} \prod_{i=1}^{4} f_X(x_i) \\
&= \prod_{i=1}^{4} \frac{x_i^3}{6\theta^4} e^{-\frac{x_i}{\theta}} \\
&= \frac{1}{(6\theta^4)^4} \left(\prod_{i=1}^{4} x_i^3\right) e^{-\frac{\sum_{i=1}^{4} x_i}{\theta}}.
\end{aligned}
$$

Passo 2 - Função de log-verosimilhança

$$
\ln L(\theta|\underline{x}) = -4\ln(6\theta^4) + \ln(\prod_{i=1}^{4} x_i^3) - \frac{\sum_{i=1}^{4} x_i}{\theta}.
$$

Passo 3 - Maximização

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

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

$$
\frac{d\ln L(\theta|\underline{x})}{d\theta} = 0 \Leftrightarrow -\frac{4}{6\theta^4}24\theta^3 + \frac{\sum_{i=1}^{4} x_i}{\theta^2} = 0 \Leftrightarrow \frac{\sum_{i=1}^{4} x_i}{\theta^2} = \frac{4}{6\theta^4}24\theta^3 \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{4} x_i}{16}
$$

$$
\left.\frac{d^2\ln L(\theta|\underline{x})}{d\theta^2}\right|_{\theta=\hat{\theta}} = -29.42 < 0, \ \text{(proposição verdadeira)}
$$

Passo 4 - Estimador e estimativa de MV de θ

$$
\text{Estimador } \hat{\theta}_{MV} = \frac{\sum_{i=1}^{4} X_i}{16}, \ \text{Estimativa } \hat{\theta} = \frac{11.8}{16} = 0.7375.
$$

Pergunta 7	2 valores

As medições de uma grandeza física obtidas com um dado instrumento distribuem-se de acordo com uma distribuição normal com desvio padrão igual a 1.4. Determine o tamanho mínimo da amostra necessário para construir um intervalo de confiança para o valor esperado das medições a um nível de confiança de 96%, e com uma amplitude não superior a 0.5.

- **V.a. de interesse**

  X = medição da grandeza física com um dado instrumento.

- **Situação** $X \sim \text{normal}(\mu, \sigma^2 = 1.4^2)$.

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

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \text{normal}(0,1).$$

- **Obtenção dos quantis de probabilidade**

$$a_\alpha = \Phi^{-1}(0.02) = -2.0537;$$
$$b_\alpha = \Phi^{-1}(0.98) = 2.0537.$$

- **Inversão da desigualdade** $a_\alpha \le Z \le b_\alpha$

$$\begin{array}{rcl} P(a_\alpha \le Z \le b_\alpha) & = & 1-\alpha \\ & \Leftrightarrow & \\ P\left(-2.0537 \le \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \le 2.0537\right) & = & 0.98 \\ & \Leftrightarrow & \\ P\left(\bar{X} - 2.0537\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 2.0537\frac{\sigma}{\sqrt{n}}\right) & = & 0.98 \end{array}$$

- **Concretização**

$$IC_{96\%}(\mu) \quad = \quad \left[\bar{x} - 2.0537\frac{1.4}{\sqrt{n}}, \bar{x} + 2.0537\frac{1.4}{\sqrt{n}}\right].$$

  Amplitude $= 2 \times 2.0537\frac{1.4}{\sqrt{n}} \le 0.5 \iff n \ge 132.27$**, pelo que** $n_{\min} = 133$.

Pergunta 8	2 valores

Admita que a variável aleatória X representa o índice de poluição atmosférica (API). Os regulamentos fixam o índice máximo de poluição atmosférica em 30. O gestor de uma fábrica quer mostrar que está em conformidade com os regulamentos, afirmando que o índice de poluição atmosférica medido próximo da fábrica é igual a 30. Para tal recolheu uma amostra de X ao longo de 10 dias, tendo observado um valor médio amostral igual a 32.5, e um desvio padrão amostral igual a 5. Assumindo que os valores são independentes entre cada medição e que têm distribuição normal, teste se a pretensão do gestor é aceitável, tomando a decisão com base no valor-p.

- **V.a. de interesse**

  X= índice de poluição medido próximo de uma fábrica.

- **Situação**

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

- **Hipóteses**

  $$H_0 : \mu = \mu_0 = 30$$
  $$H_1 : \mu = \mu_0 > 30$$

- **Estatística de teste**

  $$T = \frac{\bar{X} - \mu_0}{S_{n-1}/\sqrt{n}} \sim t_{(n-1)}.$$

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

  O teste é unilateral, logo a região de rejeição de H_0 é do tipo $W = (c, \infty)$.

- **Decisão (com base no valor-p)** Atendendo a que o valor observado da estatística de teste e do valor-p são iguais a:

  $$t = \frac{32.5 - 30}{5/\sqrt{10}} = 1.58114$$
  $$\text{valor-p} = 1 - F^{-1}_{t_{(9)}}(1.58114) \in (1 - 0.95, 1 - 0.925) = (0.05, 0.075),$$

  pelo que H_0 é rejeitada para níveis de significância superiores ou iguais a 7.5%, não sendo rejeitada para níveis de significância inferiores ou iguais a 5%.

Pergunta 9 — 2 valores

O sistema de autenticação de uma plataforma informática bloqueia o acesso de qualquer utilizador que falhe 3 tentativas consecutivas de acesso. Um engenheiro informático sustenta que a distribuição do número de tentativas de acesso falhadas por utilizador se distribui de acordo com uma distribuição binomial com probabilidade de sucesso $p = 0.25$. Dos registos do sistema foram retirados ao acaso os seguintes dados referentes a 500 acessos:

Nº de tentativas falhadas	0	1	2	3
Frequência	262	186	46	6

Averigue, aplicando um teste apropriado, se a hipótese do engenheiro é contrariada pelos dados recolhidos.

- **V.a. de interesse**

  X = número de tentativas de acesso falhadas por utilizado.

- **Hipóteses**

  $H_0 : X \sim \text{binomial}(3, 0.25)$

  $H_1 : X \not\sim \text{binomial}(3, 0.25)$

- **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 = No. de classes = 4

  O_i = Frequência absoluta observável da classe i

  E_i = Frequência absoluta esperada, sob H_0, da classe i

- **Cálculo do valor observado da estatística de teste**
  As frequências absolutas esperadas sob H_0 são dadas por

$$p_i^0 = P(X = i \mid H_0) = F_{Bi(3,0.25)}(i) - F_{Bi(3,0.25)}(i-1),\ i = 0,1,2,3.$$

  e iguais a

i	O_i	p_i^0	$E_i = np_i^0$
0	262	0.4219	210.95
1	186	0.4219	210.95
2	46	0.1406	70.30
3	6	0.0156	7.80
	$n = 500$	1	500

  Assim temos

$$\begin{aligned} t_0 &= \frac{(262-210.95)^2}{210.95} + \frac{(186-210.95)^2}{210.95} + \frac{(46-70.30)^2}{70.30} + \frac{(6-7.80)^2}{7.80} \\ &= 24.12. \end{aligned}$$

- **Decisão (com base no valor-p)**

$$\begin{aligned} valor-p &= P(T > 24.12 \mid H_0) \\ &= = 1 - F_{\chi^2_{(3)}}(24.12) \\ &< 0.0005, \end{aligned}$$

  pelo que os dados mostram evidência para rejeitar H_0 a qualquer nível de significância usual (1%, 5% e 10%).

Pergunta 10	2 valores

O interesse crescente na utilização da Internet para fins comerciais tem levado muitas companhias a vender os seus produtos através deste meio. Um estatístico levou a cabo um estudo para determinar até que ponto o grau de escolaridade e o uso da Internet estão ligados entre si. Para o efeito considerou uma amostra selecionada aleatoriamente de 20 adultos, para os quais registou o número de anos de escolaridade (x, com valores observados entre 8 e 14 anos) e o número de horas despendidas na Internet, Y, na semana anterior ao decorrer do questionário. Obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{20} x_i = 228, \quad \sum_{i=1}^{20} x_i^2 = 2696, \quad \sum_{i=1}^{20} y_i = 157, \quad \sum_{i=1}^{20} y_i^2 = 1671, \quad \sum_{i=1}^{20} x_i y_i = 1852.$$

Admita a validade do modelo de regressão linear simples, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, 20$. Para indivíduos com 14 anos de escolaridade, obtenha um intervalo de confiança a 90% para o número esperado de horas semanais despendidas na Internet.

- **Hipóteses de trabalho**

  No modelo de RLS, $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, consideraremos $\epsilon_i \overset{i.i.d.}{\sim} \text{normal}(0, \sigma^2)$, $i = 1, \ldots, n$.

  Pretende-se um intervalo de confiança a 90% para $E[Y|x = 14]$.
  Observar que $x = 14 \in [\min(x_i), \max(x_i)] = [8, 14]$

- **Variável aleatória fulcral**

$$T = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\right)\hat{\sigma}^2}} \sim t_{(18)},$$

  com $x_0 = 14$.

- **Intervalo aleatório** Como $1 - \alpha = 0.9$ então $\alpha = 0.10$ e $a = F^{-1}_{t(18)}\left(1 - \frac{\alpha}{2}\right) = F^{-1}_{t(18)}(1 - 0.05) = F^{-1}_{t(18)}(0.95) = 1.734$

  Como a distribuição da t-Student é simétrica vem:

$$P\left(-1.73 \leq \frac{(\hat{\beta}_0+\hat{\beta}_1 x_0)-(\beta_0+\beta_1 x_0)}{\sqrt{\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum\limits_{i=1}^{n} x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}} \leq 1.73\right) = 0.90$$

$$\equiv P\left((\hat{\beta}_0+\hat{\beta}_1 x_0)-1.73\sqrt{\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum\limits_{i=1}^{n} x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2} \leq (\beta_0+\beta_1 x_0) \leq (\hat{\beta}_0+\hat{\beta}_1 x_0)+1.73\sqrt{\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum\limits_{i=1}^{n} x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}\right) = 0.90.$$

Obtendo-se então o intervalo aleatório:

$$ICA_{90\%}(\beta_0+\beta_1 x_0) = \left((\hat{\beta}_0+\hat{\beta}_1 x_0) \pm 1.73\sqrt{\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{\sum\limits_{i=1}^{n} x_i^2-n\bar{x}^2}\right)\hat{\sigma}^2}\right).$$

- **Concretização:** Precisamos de calcular as estimativas de β_0 e β_1, que serão dadas por:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \\
&= \frac{1852-20\times 228/20\times 157/20}{2696-20\times(228/20)^2} \\
&\simeq 0.6426; \\
\hat{\beta}_0 &= \bar{y}-\hat{\beta}_1\times\bar{x} \\
&\simeq 157/20-0.6425\times 228/20 \\
&= 0.5244,
\end{aligned}$$

e também a estimativa de σ^2

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2-n\,\bar{y}^2\right)-(\hat{\beta}_1)^2\left(\sum_{i=1}^{n} x_i^2-n\,\bar{x}^2\right)\right] \\
&= \frac{1}{20-2}\left[438.55-0.6425^2\times 96.8\right] \\
&= 22.1432.
\end{aligned}$$

O intervalo de confiança, a 90% de confiança é dado por:

$$\begin{aligned}
IC_{90\%}(\beta_0+\beta_1\times 14) &= \left((0.5244+0.6426\times 14)\pm 1.73\sqrt{\left(\frac{1}{20}+\frac{(228/20-14)^2}{96.8}\right)22.1432}\right) \\
&= (6.6961, 12.3454).
\end{aligned}$$