MAE116 – Noções de Estatística

Grupos B e D – II Semestre de 2020

Lista de exercícios 6 – Estimação I – CASA (gabarito)

Exercício 1.

Suponha que a renda mensal (em salários mínimos) por residência em uma determinada cidade paulista tenha distribuição normal. Selecionada um amostra de 20 residências dessa cidade, observou–se uma média de 8 salários mínimos e um desvio padrão de 2 salários mínimos.

Solução

(a) Determine um intervalo de confiança para a renda mensal dessa cidade, considerando um coeficiente de confiança igual a 95%. Qual é o erro amostral associado ao intervalo construído no item anterior?

Seja

X : Renda mensal (em salários mínimos) por residência em uma determinada cidade paulista.

Do enunciado, sabemos que $X \sim \mathrm{N}(\mu, \sigma^2)$, sendo a renda mensal média populacional μ dessa cidade e o desvio padrão populacional σ desconhecidos.

O intervalo de confiança para μ considerando σ desconhecido é dado por:

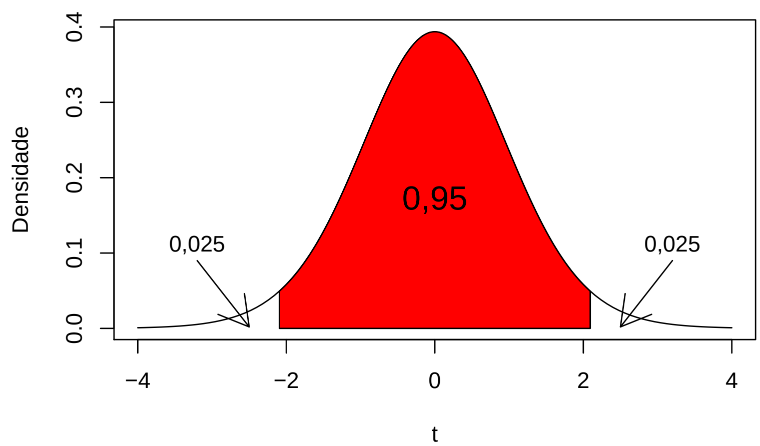
$$\left[\overline{x} - t^c_{n-1}\frac{s}{\sqrt{n}} \ ; \ \overline{x} - t^c_{n-1}\frac{s}{\sqrt{n}}\right],$$

em que, t^c_{n-1} é o quantil da distribuição t-Student com $n-1$ graus de liberdade tal que $\mathrm{P}(-t^c_{n-1} \leq T \leq t^c_{n-1}) = \gamma$, sendo $T \sim t_{n-1}$, $\overline{x}$ é o valor observado da média amostral e s é o valor observado do desvio padrão amostral, calculados a partir de uma amostra de n residências da cidade.

Temos que $\overline{x} = 8$ e $s = 2$ para uma amostra de $n = 20$ residências. Para calcular o intervalo de confiança, ainda precisamos obter o valor do quantil $t^c_{n-1} = t^c_{19}$. Temos que $\gamma = 0,95$ e $n = 20$, e portanto t^c_{19} é o quantil da distribuição t-Student com 19 graus de liberdade tal que

$$\mathrm{P}(-t^c_{19} \leq T \leq t^c_{19}) = 0,95.$$

Para utilizar a tabela da distribuição t-Student fornecida no e-disciplinas é necessário expressar t_{19}^c em termos de sua probabilidade acumulada à esquerda. Para ilustrar, o gráfico abaixo apresenta a área de 95% delimitada pelos quantis $-t_{19}^c$ e t_{19}^c na distribuição t-Student com 19 graus de liberdade. Note que o quantil t_{19}^c deixa uma probabilidade acumulada à sua esquerda igual a $0,975$. Com base na tabela obtemos que $t_{19}^c = 2,093$.



Para utilizar o pacote `Rcmdr` siga os passoa abaixo:

`Distribuições` $\rightarrow$ `Distribuições contínuas` $\rightarrow$ `Distribuição t` $\rightarrow$ `quantis t`. Forneça a probabilidade 0.975 (Lembre que no `R`, o "." é separador decimal); graus de liberdade igual a 19; e deixe selecionado a cauda esquerda. A saída é:

```
Rcmdr>  qt(c(0.975), df=19, lower.tail=TRUE)
[1] 2.093024
```

Logo, o intervalo de confiança é dado por

$$\left[8 - 2,093\frac{2}{\sqrt{20}} \; ; \; 8 + 2,093\frac{2}{\sqrt{20}}\right] = [7,0640 \; ; \; 8,9360] \, .$$

O erro amostral associado é dado por

$$\varepsilon = t_{n-1}^c \frac{s}{\sqrt{n}},$$

ou seja, o erro amostral associado ao intervalo construído supondo o desvio padrão populacional desconhecido é

$$\varepsilon = 2,093\frac{2}{\sqrt{20}} = 0,9360.$$

(b) Supondo que o desvio padrão populacional σ da renda mensal seja igual a 2 salários mínimos, que tamanho deve ter uma amostra para que o intervalo $10,0 \pm 1,0$ tenha 95% de confiança?

Supondo o desvio padrão populacional $\sigma = 2$ e utilizando um coeficiente de confiança $\gamma = 0,95$, o tamanho da amostra n é determinado como

$$n = \left(\frac{z}{\varepsilon}\right)^2 \sigma^2,$$

em que $\varepsilon = 1,0$ é o erro amostral, e z é o quantil da distribuição normal padrão tal que $\mathrm{P}(-z \leq Z \leq z) = 0,95$, sendo que $Z \sim N(0,1)$.

Para utilizar a tabela da distribuição normal padrão fornecida no e-disciplinas, também é necessário expressar o quantil z em termos de sua probabilidade acumulada à esquerda. Como

$$\mathrm{P}(-z \leq Z \leq z) = 0,95,$$

então z deixa uma probabilidade acumulada à esquerda igual a $0,975$. Verificando a tabela obtemos que $z = 1,96$. Logo,

$$n = \left(\frac{1,96}{1,0}\right)^2 2^2 = 15,3664 \approx 16,$$

ou seja, deve ser selecionada uma amostra de 16 residências.

Exercício 2

O tempo de reação a um estímulo luminoso pode ser considerado como tendo distribuição Normal com desvio padrão igual a 5 minutos (a média é desconhecida). O objetivo de uma pesquisa é estimar a média μ do tempo de reação ao estímulo luminoso.

Solução

(a) Qual deve ser o tamanho da amostra para que o erro cometido ao estimarmos o tempo médio μ, não seja superior a 20 segundos, com probabilidade 0,90?

Seja

$$X : \text{Tempo de reação a um estímulo luminoso, em minutos.}$$

Temos do enunciado que $X \sim N(\mu; 5^2)$.

O tamanho da amostra é calculado como

$$n = \left(\frac{z}{\varepsilon}\right)^2 \sigma^2,$$

em que ε é o erro amostral, e, sendo o coeficiente de confiança $\gamma = 0,90$, z é o quantil da distribuição normal padrão tal que P$(-z \leq Z \leq z) = 0,90$, com $Z \sim N(0,1)$.

Do enunciado, é desejado que o erro amostral cometido, não seja superior a 20 segundos, que equivale a $1/3$ de um minuto. Logo, $\varepsilon = 1/3$.

Como z é tal que P$(-z \leq Z \leq z) = 0,90$, então z deixa uma probabilidade acumulada à esquerda igual à $0,95$. Utilizando a tabela da distribuição normal padrão, obtemos que $z = 1,64$. Logo,

$$n = \left(\frac{1,64}{1/3}\right)^2 5^2 = 605,16 \approx 606$$

Portanto, o tamanho da amostra para que o erro cometido ao estimarmos o tempo médio μ, não seja superior a 20 segundos, com probabilidade 0,90 deve ser pelo menos de 606 pacientes.

Dez pacientes foram sorteados e tiveram seus tempos de reação anotados. Os dados foram os seguintes (em minutos): 3,5; 4,0; 4,6; 4,8; 5,2; 4,7; 5,9; 3,5; 6,0; 4,9.

(b) Obtenha uma estimativa pontual do tempo médio de reação μ.

Uma estimativa pontual para o tempo médio de reação μ é dada pela média amostral $\overline{x}$ das observações, que para esta amostra é dada por

$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{3,5 + \cdots + 4,9}{10} = 4,71.$$

(c) Encontre uma estimativa intervalar de 98% de confiança para μ. Qual é o erro amostral de sua estimativa pontual?

Como o desvio padrão é conhecido, o intervalo de confiança para o tempo médio de reação μ é dado por

$$\left[\overline{x} - z\frac{\sigma}{\sqrt{n}} \ ; \ \overline{x} + z\frac{\sigma}{\sqrt{n}}\right],$$

em que z é tal que, $\gamma = P(-z \leq Z \leq z)$.

Para $\gamma = 0,98$, z deixa uma probabilidade acumulada à esquerda igual a $0,99$. A partir da tabela da distribuição normal padrão obtemos que $z = 2,33$. Assim, o intervalo de confiança para μ com coeficiente de confiança 98% é dado por

$$\left[4,71 - 2,33\frac{5}{\sqrt{10}} \ ; \ 4,71 + 2,33\frac{5}{\sqrt{10}}\right] = [1,0259 \ ; \ 8,3941] \, .$$

O erro amostral associado ao intervalo construído acima é

$$\varepsilon = z\frac{\sigma}{\sqrt{n}} = 2,33\frac{5}{\sqrt{10}} = 3,6841.$$

(d) Suponha que no item (c) não fosse conhecido o desvio padrão populacional. Como você procederia para determinar o intervalo de confiança? Justifique. Construa o intervalo de confiança para μ com o mesmo coeficiente de confiança adotado no item (c).

Como o desvio padrão populacional não é conhecido, é necessário estimá-lo para que seja possível construir um intervalo de confiança para a média. Esta nova fonte de variação (a estimação pontual do desvio padrão) faz com que seja necessário corrigir o intervalo de confiança no item anterior (que utilizava quantis da distribuição normal padrão) pelos quantis da distribuição t-Student, a fim de que o coeficiente de confiança seja satisfeito corretamente. A fórmula do intervalo de confiança para a média com coeficiente de confiança γ quando o desvio padrão é desconhecido é da forma

$$\left[\overline{x} - t_{n-1}^{c}\frac{s}{\sqrt{n}}\ ;\ \overline{x} - t_{n-1}^{c}\frac{s}{\sqrt{n}}\right],$$

em que, t_{n-1}^{c} é o quantil da distribuição t-Student com $n-1$ graus de liberdade tal que $\mathrm{P}(-t_{n-1}^{c} \leq T \leq t_{n-1}^{c}) = \gamma$, sendo $T \sim t_{n-1}$ e s é o desvio padrão amostral.

Recorde que

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}.$$

Utilizando as observações da amostra, calculamos que

$$s = \sqrt{s^2} = \sqrt{\frac{(3,5-4,71)^2 + \cdots + (4,9-4,71)^2}{9}} = \sqrt{0,7566} = 0,8698.$$

Em resumo, temos que $\overline{x} = 4,71$ e $s = 0,8698$ para uma amostra de $n = 10$ pacientes. Para calcular o intervalo de confiança, ainda precisamos obter o valor do quantil $t_{n-1}^{c} = t_{9}^{c}$. Temos que $\gamma = 0,98$ e $n = 10$, e portanto t_9^c é o quantil da distribuição t-Student com 9 graus de liberdade tal que

$$\mathrm{P}(-t_9^c \leq T \leq t_9^c) = 0,98.$$

Para podermos utilizar a tabela fornecida no e-disciplinas é necessário expressar t_9^c em termos da probabilidade acumulada à sua esquerda. Note que se $\mathrm{P}(-t_9^c \leq T \leq t_9^c) = 0,98$, então t_9^c deixa uma área acumulada à esquerda igual a $0,99$. Com base na tabela, obtemos que $t_9^c = 2,821$.

Assim, o intervalo de confiança quando o desvio padrão σ não é conhecido fica dado por

$$\left[4,71 - 2,821\frac{0,8698}{\sqrt{10}}\ ;\ 4,71 + 2,821\frac{0,8698}{\sqrt{10}}\right] = [3,9341\ ;\ 5,4859].$$

Exercício 3

O intervalo $[26,164\ ;\ 29,947]$ é o intervalo de 98% de confiança, construído a partir de uma amostra de tamanho 72, para a média populacional μ da idade de indivíduos de certa população.

Solução:

(a) Supondo que a idade de indivíduos dessa população segue uma distribuição normal com desvio padrão σ conhecido, encontre σ e o erro amostral associado ao intervalo de confiança $[26,164\ ;\ 29,947]$.

Seja

$$X:\ \text{idade de indivíduos de uma certa população.}$$

Temos do enunciado que

$$X \sim N(\mu, \sigma^2).$$

Supondo σ conhecido, um intervalo de confiança para a média populacional μ com coeficiente de confiança γ é dado por

$$\left[\overline{x} - z\frac{\sigma}{\sqrt{n}}\ ;\ \overline{x} + z\frac{\sigma}{\sqrt{n}}\right],$$

sendo z o quantil da distribuição normal padrão tal que $\mathrm{P}(-z \leq Z \leq z) = \gamma$, com $Z \sim \mathrm{N}(0,1)$.

Como $\gamma = 0,98$, então z acumula uma probabilidade à esquerda igual a $0,99$. Temos a partir da tabela da distribuição normal padrão que $z = 2,33$. Para $n = 72$, segue que

$$26,164 = \overline{x} - 2,33\frac{\sigma}{\sqrt{72}} \tag{1}$$

$$29,947 = \overline{x} + 2,33\frac{\sigma}{\sqrt{72}}. \tag{2}$$

Fazendo a diferença entre as equações (2) e (1), obtemos que

$$\begin{aligned} 29,947 - 26,164 &= 2 \times 2,33\frac{\sigma}{\sqrt{72}} \\ \Rightarrow \sigma &= \frac{\sqrt{72} \times 3,783}{4,66} \\ \Rightarrow \sigma &= 6,8884. \end{aligned}$$

O erro amostral é, portanto, dado por

$$\varepsilon = z\frac{\sigma}{\sqrt{72}} = 2,33\frac{6,8884}{\sqrt{72}} = 1,8915.$$

(b) Que tamanho deve ter a amostra para que o erro amostral calculado em (a) seja reduzido à metade?

O tamanho da amostra é calculado como

$$n = \left(\frac{z}{\varepsilon}\right)^2 \sigma^2,$$

em que, $z = 2,33$ e $\sigma = 6,8884$.

Como queremos a metade do erro amostral do item (a), devemos considerar agora $\varepsilon = 1,8915/2 = 0,9458$. Assim,

$$n = \left(\frac{2,33}{0,9458}\right)^2 6,8884^2 = 287,9718 \approx 288.$$

Portanto, para que o erro amostral associado ao intervalo de confiança apresentado no itam (a) seja reduzido à metade, o tamanho amostral deve ser de 288.

(c) Compare o tamanho da amostra obtido no item (b) com o tamanho da amostra dado no item (a). Que resultado você pode estabelecer?

Recorde que no item (a) o tamanho amostral utilizado para a construção do intervalo de confiança para a média μ foi $n = 72$. Com o objetivo de reduzir o erro amostral associado a este intervalo pela metade, calculamos no item (b) que devemos aumentar o tamanho da amostra para $n = 288$. Ou seja, para reduzir o erro amostral à metade neste caso, é necessário considerar um tamanho amostral quatro vezes maior.

Exercício 4

De um sítio geológico foram extraídas 40 amostras de rochas. Cada amostra foi analisada e a porcentagem da rocha tipo A foi determinada. Os resultados apurados foram:

69,3	63,4	73,1	72,0	68,6	64,1	67,9	66,5	65,8	59,0
63,7	67,4	64,5	79,0	57,7	76,9	67,6	66,7	51,7	66,0
63,3	65,5	62,0	64,4	63,8	58,9	62,3	68,6	71,5	76,4
53,1	68,1	67,7	66,4	62,0	71,7	61,8	60,9	64,5	64,5

Solução:

(a) Construa um histograma para mostrar que a porcentagem da rocha tipo A parece ter distribuição normal.

Neste caso, é mais fácil e rápido utilizar a própria linguagem R para ler os dados e construir o histograma, ao invés do pacote `Rcmdr`. Para isto, abra o R ou Rstudio, copie os seguintes comandos:

```
dados <- c(69.3,  63.4,  73.1,  72.0,  68.6,  64.1,  67.9,  66.5,
           65.8,  59.0,  63.7,  67.4,  64.5,  79.0,  57.7,  76.9,
           67.6,  66.7,  51.7,  66.0,  63.3,  65.5,  62.0,  64.4,
           63.8,  58.9,  62.3,  68.6,  71.5,  76.4,  53.1,  68.1,
           67.7,  66.4,  62.0,  71.7,  61.8,  60.9,  64.5,  64.5)

hist(dados, main = "", xlab = "Porcentagem da rocha tipo A",
 ylab = "Frequência")
```

cole no console, e aperte enter. O histograma gerado segue abaixo.

Obs: Em caso de erro, verifique se as aspas foram copiadas corretamente.

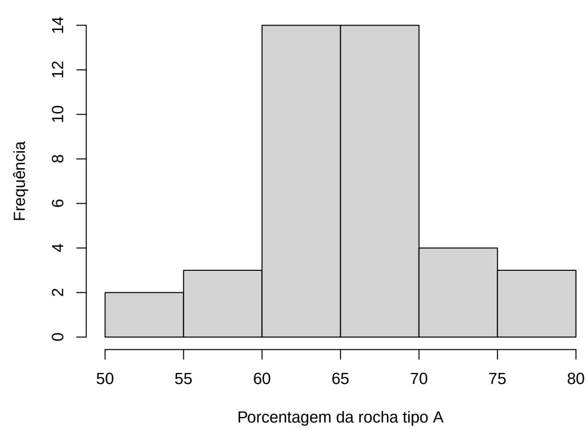


Figura 1: Histograma da porcentagem da rocha tipo A.

Caso opte pela utilização do pacote `Rcmdr`, siga os seguintes passos:

Passo 1 (Leitura dos dados): `Dados` → `Novo conjunto de dados` → Dê um nome para o conjunto (por exemplo, dados) → Digite os dados em uma mesma coluna, lembrando que o separador decimal no R é o ".". Após o término, clique em OK.

Passo 2 (Construção do histograma): `Gráficos` → `Histograma` → clique em OK.

Com base no histograma, é razoável supor que porcentagem da rocha tipo A segue uma distribuição normal.

(b) Calcule, com base nessa amostra, um intervalo de confiança para a porcentagem média populacional da rocha tipo A, com coeficiente de confiança igual a 95%.

Com base nas $n = 40$ amostras, primeiro devemos calcular as estimativas pontuais para a porcentagem média populacional μ e para o desvio padrão populacional σ, que são dadas por

$$\overline{x} = \frac{69,3 + \cdots + 64,5}{40} = 65,7075$$

e

$$s = \sqrt{s^2} = \sqrt{\frac{(69,3 - 65,7075)^2 + \cdots + (64,5 - 65,7075)^2}{39}} = \sqrt{32,0094} = 5,6577.$$

O intervalo de confiança para μ, sendo σ desconhecido, é dado por

$$\left[\overline{x} - t^c_{n-1}\frac{s}{\sqrt{n}} \ ; \ \overline{x} + t^c_{n-1}\frac{s}{\sqrt{n}}\right],$$

em que t^c_{n-1} é o quantil da distribuição t-Student com $n - 1$ graus de liberdade tal que $\mathrm{P}(-t^c_{n-1} \leq T \leq t^c_{n-1}) = \gamma$, sendo $T \sim t_{n-1}$.

Temos que $\gamma = 0,95$ e $n = 40$, e portanto $t^c_{n-1} = t^c_{39}$ é o quantil da distribuição t-Student com 39 graus de liberdade tal que

$$\mathrm{P}(-t^c_{39} \leq T \leq t^c_{39}) = 0,95.$$

A tabela fornecida no e-disciplinas não considera a distribuição t_{39}, e portanto temos que utilizar o pacote `Rcmdr`. Para isso, é necessário expressar t^c_{39} em termos da sua probabilidade acumulada à esquerda. Note que se $\mathrm{P}(-t^c_{39} \leq T \leq t^c_{39}) = 0,95$, então t^c_{39} deixa uma área acumulada à esquerda igual a $0,975$. Siga os segunites passos para obter o valor de t^c_{39} pelo `Rcmdr`:

`Distribuições` $\rightarrow$ `Distribuições contínuas` $\rightarrow$ `Distribuição t` $\rightarrow$ `quantis t`. Forneça a probabilidade 0.975 para o quantil expresso em termos da probabilidade acumulada à esquerda (Lembre que no `R`, o "." é separador decimal); graus de liberdade igual a 39. Deixe selecionado a cauda esquerda. A saída é:

```
Rcmdr>  qt(c(0.975), df=39, lower.tail=TRUE)
[1] 2.022691
```

Portanto, um intervalo de confiança para μ com coeficiente de confiança igual a 95% é dado por

$$\left[65,7075 - 2,0227\frac{5,6577}{\sqrt{40}} \ ; \ 65,7075 + 2,0227\frac{5,6577}{\sqrt{40}}\right] = [63,8981 \ ; \ 67,5169].$$