MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Exercícios de revisão

Exercício 1.
Um produtor de laranjas guardou as frutas colhidas de cada árvore em caixas separadas para estudar o número de laranjas por caixa (ou por árvore). Após um dia de colheita, 20 caixas foram contadas e os resultados brutos, após ordenação, são: 22, 27, 29, 33, 35, 37, 38, 43, 43, 44, 48, 48, 52, 53, 55, 57, 61, 62, 67 e 69.

(a) Calcule a mediana, Q_1 e Q_3.
Resposta:
mediana=(44+48)/2=46
Q_1*=(35+37)/2=36*
Q_3*=(55+57)/2=56*
□

(b) Calcule o intervalo interquartil e determine pontos atípicos.
Resposta:
O intervalo interquartil é $J = Q_3 - Q_1 = 20$ *e, portanto, os limites inferiores e superiores são* $Q_1 - 1,5J = 6$ *e* $Q_3 + 1,5J = 86$*. Logo, não existem pontos atípicos.* □

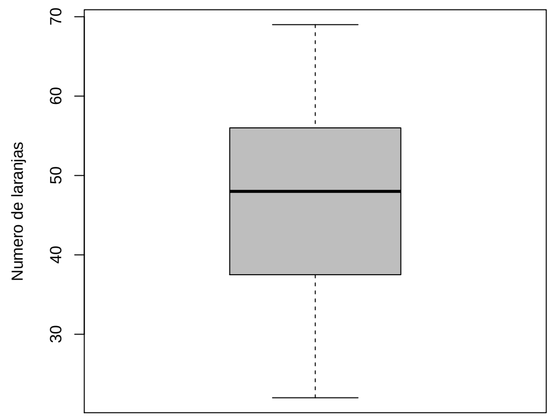
(c) O *boxplot* para a variável número de laranjas é apresentado na Figura 1. Comente.
Resposta:
Observa-se que não existem pontos atípicos. Note que a mediana está um pouco mais perto do terceiro quartil do que do primeiro quartil, indicando uma leve assimetria. □

Exercício 2.
O custo mensal de manutenção de determinado modelo de automóvel (excluindo-se combustível e trocas de óleo) está sendo analisado em função da idade do veículo. Nove automóveis fabricados em diferentes anos tiveram o custo averiguado. Os dados obtidos foram:

Idade do veículo (em anos)	1	2	3	4	5	6	7	8	9
Custo mensal (em reais)	8	13	18	15	24	26	25	32	33

(a) Observe o diagrama de dispersão apresentado na Figura 2. Interprete o relacionamento entre as variáveis idade do veículo e custo mensal.

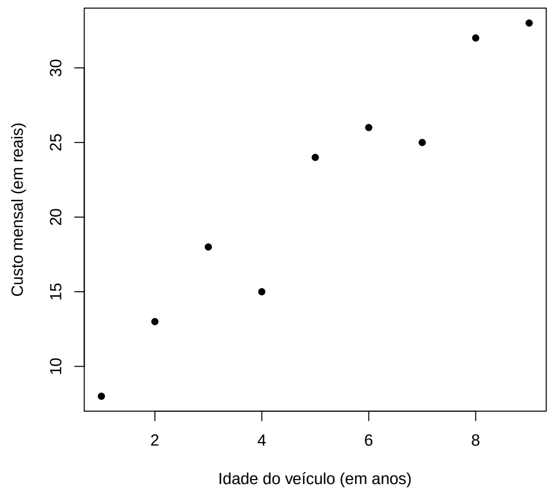
Figura 1: *Boxplot para a variável número de laranjas.*



Resposta:

Observa-se que conforme aumenta a idade do veículo, o custo mensal tende a aumentar. Nota-se também uma tendência linear.

Figura 2: *Diagrama de dispersão entre as variáveis idade do veículo (em anos) e custo mensal (em reais).*



□

(b) O coeficiente de correlação linear de Pearson entre idade do veículo e custo mensal é 0,97. Comente.

Resposta:

O coeficiente de correlação indica forte correlação linear positiva entre as variáveis idade do veículo e custo mensal. □

(c) A reta de regressão ajustada entre as as variáveis, Y: custo mensal X: idade do veículo é dada por

$$\widehat{Y}_i = 6,39 + 3,03X_i, \quad i = 1, \ldots, 9.$$

Interprete o valor do coeficiente angular "*b*" da reta ajustada.

Resposta:

Para um aumento de um ano na idade do veículo, o custo mensal aumenta, em média, 3,03 reais. □

(d) Considerando a reta ajustada em (c), estime o custo mensal esperado para um veículo com 4,5 anos.

Resposta:

O custo mensal esperado para um veículo com 4,5 anos é dado por:

$$\widehat{Y} = 6,39 + 3,03 \cdot 4,5 = 20,025 \quad \textit{reais.}$$

□

Exercício 3.
Uma determinada peça é manufaturada por 3 fábricas: A, B e C. Sabe-se que A produz o dobro de peças que B e que B e C produzem o mesmo número de peças. Sabe-se ainda que 2% das peças produzidas por A e por B são defeituosas, enquanto que 4% das produzidas por C são defeituosas. Todas as peças produzidas são misturadas e colocadas em um depósito. Do depósito é retirada uma peça ao acaso.

(a) Qual a probabilidade de que ela seja defeituosa?
Resposta:
Considerem-se os seguintes eventos:

D: A peça é defeituosa

A: A peça provém da fábrica A

B: A peça provém da fábrica B

C: A peça provém da fábrica C

Tem-se então que: $P(A) = 50\%$, $P(B) = P(C) = 25\%$, *uma vez que só existem as 3 fábricas e que A produz o dobro de B e esta por sua vez produz a mesma quantidade que C. Sabe-se também que* $P(D|A) = P(D|B) = 2\%$ *e que* $P(D|C) = 4\%$. *Então pode-se escrever que:* $P(D) = P(D \cap A) + P(D \cap B) + P(D \cap C) = P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) = 0,5 \times 0,02 + 0,25 \times 0,02 + 0,25 \times 0,04 = 0,025$ *isto, é 2,5%.* □

(b) Qual é a probabilidade de ter sido produzida pela máquina A, dado que a peça é defeituosa?
Resposta:

$$\begin{aligned} P(A|D) &= \frac{P(A \cap D)}{P(D)} \\ &= \frac{P(A)P(D|A)}{P(D)} \\ &= \frac{0,50 \times 0,02}{0,5 \times 0,02 + 0,25 \times 0,02 + 0,25 \times 0,04} \\ &= 0,40 \end{aligned}$$

Logo, 40% das peças defeituosas são produzidas pela fábrica A. □

Exercício 4.

De acordo com o último censo, 20% das famílias de uma região vivem abaixo da linha da pobreza. A seguir as probabilidades da distribuição binomial para $n = 12$ e $p = 0,2$.

```
> .Table
             Pr
0  6.871948e-02
1  2.061584e-01
2  2.834678e-01
3  2.362232e-01
4  1.328756e-01
5  5.315022e-02
6  1.550215e-02
7  3.321889e-03
8  5.190451e-04
9  5.767168e-05
10 4.325376e-06
11 1.966080e-07
12 4.096000e-09
```

(a) Uma amostra aleatória de 12 familias é selecionada

i). Calcule a probabilidade de que pelo menos 2 vivam abaixo da linha de pobreza

<u>*Resposta:*</u>

Seja X a variável aleatória que conta o número de famílias que vivem abaixo da linha de pobreza. Então entre as 12 selecionadas, $X \sim$ Bin(12; 0,2). Logo,

$$\begin{aligned} P(X \geq 2) &= 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - 0,069 - 0,206 \\ &= 0,725 \end{aligned}$$

$\square$

ii). Calcule a probabilidade de que entre 2 e 4 (inclusive) vivam abaixo da linha de pobreza

<u>*Resposta:*</u>

$$
\begin{aligned}
P(2 \leq X \leq 4) &= P(X=2)+P(X=3)+P(X=4) \\
&= 0,283+0,236+0,133 \\
&= 0,652
\end{aligned}
$$

□

(b) Uma amostra de 80 famílias é selecionada. Calcule a probabilidade de menos de 10 famílias viverem abaixo da linha da pobreza.

<u>*Resposta:*</u>

Seja, X : número de famílias que vivem abaixo da linha da pobreza, entre as 80 selecionadas, então, X ~Bin(80; 0,2). Dados os valores de n e p pode ser usada a aproximação normal para calcular as probabilidades.

Verificando: $np = 80 \cdot 0,2 = 16$. Assim, X segue aproximadamente uma distribuição normal com média $\mu = np = 16$ e desvio padrão $\sigma = \sqrt{np(1-p)} = \sqrt{80 \cdot 0,2 \cdot 0,8} = 3,578$.

Aproximando a distribuição binomial pela distribuição normal, a probabilidade de menos de 10 famílias amostradas viverem abaixo da linha da pobreza é:

$$
\begin{aligned}
P(X < 10) = P(X \leq 9) &\approx P\left(Z < \frac{9-16}{3,578}\right) \\
&= P(Z < -1,96) \\
&= 1 - P(Z < 1,96) \\
&= 1 - 0,975 \\
&= 0,025
\end{aligned}
$$

□

Exercício 5.

Pacientes, sofrendo de certa moléstia, são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma distribuição normal de média 15 e desvio padrão 2 (em dias).

(a) Qual é a proporção de pacientes que demoram mais de 17 dias para se recuperar?
<u>*Resposta:*</u>
Seja a variável X: tempo de cura do paciente. Logo, a proporção de pacientes que demoram mais de 17 dias em se recuperar é

$$P(X > 17) = P\left(\frac{X-15}{2} > \frac{17-15}{2}\right) = P(Z > 1) = 0,1587.$$

□

(b) Qual é a probabilidade de que um paciente, escolhido ao acaso, apresente tempo de cura inferior a 20 dias?
<u>*Resposta:*</u>

$$P(X < 20) = P\left(\frac{X-15}{2} < \frac{20-15}{2}\right) = P(Z < 2,5) = 0,9938.$$

□

(c) Qual é o tempo máximo necessário para a recuperação de 25% dos pacientes?
<u>*Resposta:*</u>
Temos que calcular o valor de t tal que $P(X < t) = 0,25$. Então,

$$P\left(\frac{X-15}{2} < \frac{t-15}{2}\right) = 0,25$$

e

$$\frac{t-15}{2} = -0,67 \Longrightarrow t = 13,66,$$

isto é, o 25% dos pacientes ficarão curados antes de, aproximadamente, 14 dias. □