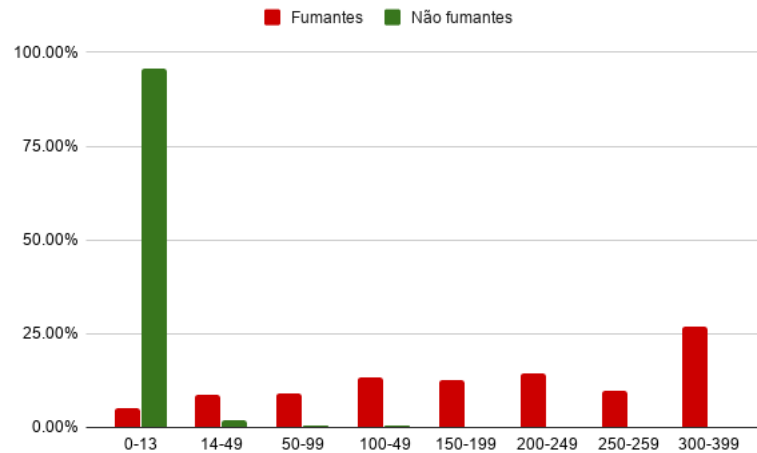


## Lista 1

1.
  - a. Vitamina (A, B1, B2, B6, B12); Variável qualitativa nominal.
  - b. Quantidade de calorias na batata frita; Variável quantitativa contínua.
  - c. Desfecho de uma doença (curado, não curado); Variável qualitativa nominal.
  - d. Classificação de uma lesão (lesão fatal; severa; moderada; pequena); Variável qualitativa ordinal.
  - e. Grupo sanguíneo (A, B, AB, O); Variável qualitativa nominal.
  - f. Paridade (primeira gestação, segunda gestação, terceira ...); Variável qualitativa ordinal.
  - g. Estado geral de um paciente (bom, regular, ruim); Variável qualitativa ordinal.
  - h. Número de nascidos vivos em certo hospital em junho/99; Variável quantitativa discreta.
  - i. Idade; Variável quantitativa discreta
  - j. Concentração de flúor na água; Variável quantitativa contínua.
  - k. Atividade esportiva preferida. Variável qualitativa ordinal.
2.
  - a. Não, pois todos os níveis de cotinina possuem diferentes frequências absolutas, isto é, por exemplo que o número de pessoas fumantes é maior ao número de não fumantes mas que em proporção ao total, poderiam ter o mesmo valor.
  - b. Resposta

Nível de cotinina	Fumantes( $p_i\%$ )	Não fumantes( $p_i\%$ )
0-13	5.07%	95.79%
14-49	8.64%	2.09%
50-99	9.23%	0.67%
100-49	13.39%	0.44%
150-199	12.80%	0.20%
200-249	14.29%	0.23%
250-259	9.81%	0.26%
300-399	26.77%	0.32%
Total	100.00%	100.00%

c. .



d. Da figura e a tabela, note que a grande maioria dos não fumantes tem baixo nível de cotinina.

3. a. Ordenando os dados temos:

21.3 22.1 22.8 23.5 24.6 65.4 67.2 71.7 76.3 84.5.

Como são 10 dados, a mediana seria a semissoma dos valores intermédios, posição 5ta e 6ta isto é:

$$\frac{24.6 + 65.4}{2} = 45.$$

É claro que a mediana não representa o conjunto de dados, pois veja que os dados estão divididos em dois grupos.

b. .

b1.

21.3 22.1 22.8 23.5 24 24.6 65.4 67.2 71.7 76.3 84.5.

Veja que mediana está na 6ta posição pois  $\frac{(11+1)}{2} = 6$ , então

$$Md = 24.6.$$

A mediana representa somente os dados do lado esquerdo.

b2.

21.3 22.1 22.8 23.5 24.6 65.4 67.2 71.7 75 76.3 84.5.

Veja que mediana está na 6ta posição pois  $\frac{(11+1)}{2} = 6$ , então

$$Md = 65.4.$$

A Mediana representa somente os dados do lado direito.

Do item b1. e b2. a Mediana seria uma medida não estável.

4. a. Sejam  $m_i = \frac{L_i + U_i}{2}$  onde  $L_i$  é o limite inferior do intervalo e  $U_i$  é o limite superior do intervalo e  $n = \sum_{i=1}^6 n_i = 1000$

$$\bar{X} = \sum_{i=1}^6 \frac{n_i m_i}{n}$$

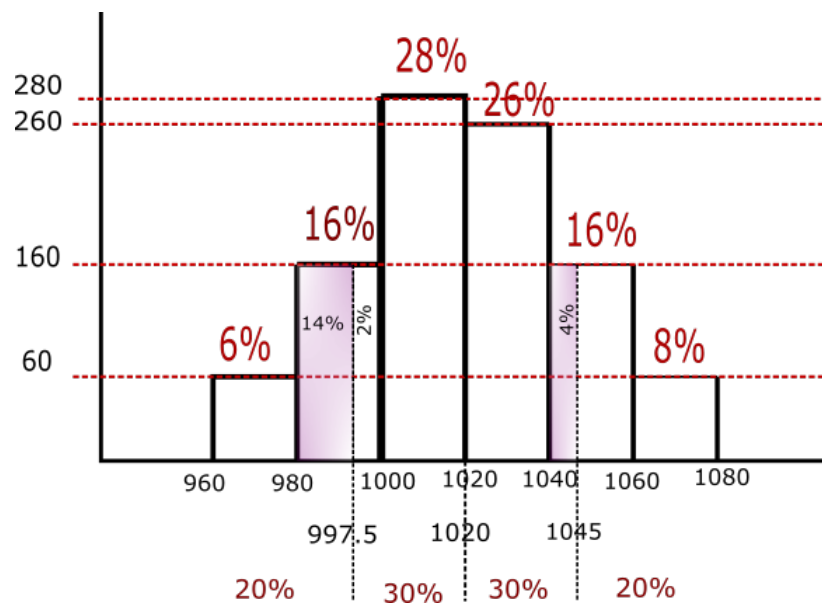
$$\bar{X} = 1020.8$$

b.

$$\text{Var}(X) = \sum_{i=1}^6 \frac{n_i (m_i - \bar{X})^2}{n - 1}$$

$$\text{Var}(X) = \frac{691360}{999} = 962.061$$

- c. Da seguinte figura podemos ver que o peso (em gramas) representativo está em torno de 1020 g, o que fornecido pela média calculada no item a..



- d. Para um melhor entendimento desta parte do exercício ver o histograma do item c..

- Para os primeiros 20%

$$\frac{1000 - 980}{16\%} = \frac{x - 980}{14\%} \quad (1)$$

$$x = 997.5 \quad (2)$$

então o intervalo seria  $(960, 997.5]$

- Os 30% seguintes:  $x = 1029$ , pois  $28\% + 2\% = 30\%$  então o intervalo seria dado por:  $(997.5, 1020]$
- Para os seguintes 30%

$$\frac{1060 - 1040}{16\%} = \frac{x - 1040}{4\%} \quad (3)$$

$$x = 1045 \quad (4)$$

então o intervalo seria  $]1020, 1045]$

- Veja que só falta  $12\% + 8\% = 20\%$ , assim o último intervalo é dado por  $(1045, 1085]$
5. a. Seja X: Diâmetros do coração dos adultos maiores normais.  
Para cálculo da média temos:

$$\bar{X} = \sum_{i=1}^{15} \frac{X_i}{15} = 131.0666667$$

Para calcular a mediana, ordenamos os dados em forma crescente:

$X : 103 \ 114 \ 114 \ 114 \ 121 \ 125 \ 125 \ 130 \ 130 \ 132 \ 135 \ 139 \ 146 \ 169 \ 169$

como são 15 dados é um número ímpar, então o cálculo é dado por: posição =  $\frac{15+1}{2} = 8$ , assim a mediana é dada pelo valor na oitava posição, isto é 130.

A moda é o valor mais frequente, neste caso é 114.

b.

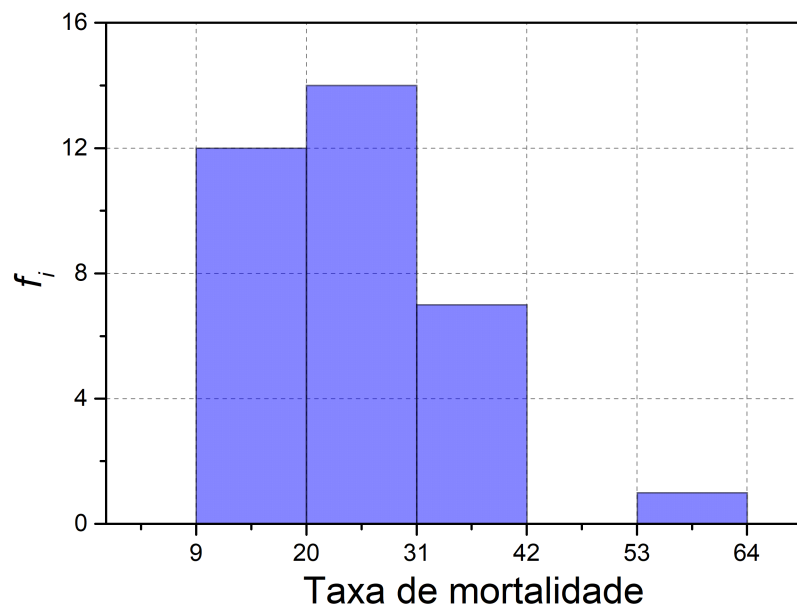
$$\text{Var}(X) = \frac{\sum_{i=1}^{15} (X_i - \bar{X})^2}{14} = 358.495 \quad (5)$$

$$Dp(X) = 18.933 \quad (6)$$

6. a. .

Nro	taxa de morta.	$f_i$	$N_i$	$p_i\%$	$P_i\%$
1	(9,20]	12	12	35.29%	35.29 %
2	(20,31]	14	26	41.18%	76.47 %
3	(31,42]	7	33	20.59%	97.06 %
4	(42,53]	0	33	0.00%	97.06%
5	(53,64]	1	34	2.94%	100 %
Total		34		100%	

b. .



c. Seja  $m_i = \frac{L_i + U_i}{2}$  onde  $L_i$  é o limite inferior do intervalo e  $U_i$  é o limite superior do intervalo.

$$n = \sum_{i=1}^5 f_i = 34$$

$$\bar{X} = \frac{\sum_{i=1}^5 f_i m_i}{n}$$

$$\bar{X} = 24.85$$

Para o cálculo da mediana utilizamos as proporções, para isso, observamos nas frequências acumuladas porcentuais tal que contenham 50 %, assim a  $Md \in (20, 31]$ .

$$\frac{Md - 20}{14.71\%} = \frac{31 - 20}{41.18\%}$$

$$Md = 23.92$$

Para o cálculo do 1º quartil, ou seja 25º percentil ( $q_1$ ), observamos nas frequências acumuladas percentuais tal que contenham 25 %, assim a  $q_1 \in (9, 20]$ .

$$\begin{aligned}\frac{q(0.25) - 9}{25\%} &= \frac{20 - 9}{35.29\%} \\ q(0.25) &= 16.79\end{aligned}$$

Para o cálculo do 3º quartil, ou seja 75º percentil ( $q_3$ ), observamos nas frequências acumuladas percentuais que contenham 75%, assim  $q_3 \in (20, 31]$ .

$$\begin{aligned}\frac{q(0.75) - 20}{(75 - 35.29)\%} &= \frac{31 - 20}{41.18\%} \\ q(0.75) &= 30.60.\end{aligned}$$

A variância e o desvio padrão é dado por:

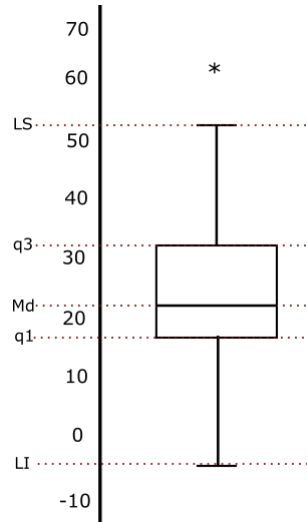
$$\begin{aligned}\text{Var}(X) &= \sum_{i=1}^5 \frac{f_i(m_i - \bar{X})^2}{n - 1} \\ \text{Var}(X) &= \frac{1888.61}{33} = 57.23 \\ Dp(X) &= 7.56.\end{aligned}$$

d. – Esquema de 5 números.

		34	
md		23.92	
$q$	16.79		30.60
md	9.9		62.2

– Para o Box Plots

$$\begin{aligned}d_q &= q_3 - q_1 = 30.69 - 16.79 = 13.9 \\ LI &= q_1 + 1.5(dq) = 16.79 - 1.5(13.9) = -4.06 \\ LS &= q_3 + 1.5(dq) = 30.60 + 1.5(13.9) = 51.45\end{aligned}$$



– Ramos de folhas

0	9
1	0 1 3 3
1*	5 7 8 8 8 9
2	0 0 1 2 2 2 3 3 3
2*	5 7 7 8 9 9
3	2 2 3 6 6 6 8 9
4	
5	
6	2

7. a.

$$\bar{Y} = \frac{\sum_{i=1}^{10} Y_i}{10} = 15 \quad (7)$$

$$\text{Var}(Y) = \frac{\sum_{i=1}^{10} (Y_i - \bar{Y})^2}{9} = 16.66 \quad (8)$$

$$Dp(Y) = 4.081 \quad (9)$$

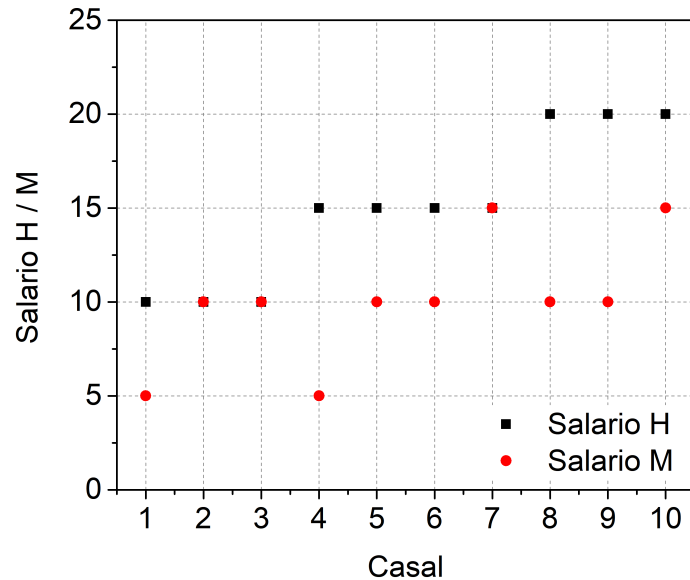
b.

$$\bar{X} = \frac{\sum_{i=1}^{10} X_i}{10} = 10$$

$$\text{Var}(X) = \frac{\sum_{i=1}^{10} (X_i - \bar{X})^2}{9} = 11.11$$

$$Dp(X) = 3.33$$

c. .



d.

$$\text{Corr}(X, Y) = \frac{1}{n} \sum_{i=1}^{10} \left( \frac{Y_i - \bar{Y}}{Dp(Y)} \right) \left( \frac{X_i - \bar{X}}{Dp(X)} \right) = 0.4088$$

Como  $0.480 > 0$ , então, existe uma correlação média entre as variáveis  $X$  e  $Y$ .

e.  $S$  = Salario total do casal.

Casal	1	2	3	4	5	6	7	8	9	10
$S=X+Y$	15	20	20	20	25	25	25	3	30	35
$Z=0.92Y+0.94X$	13.9	18.6	18.6	18.5	23.2	23.2	27.9	27.8	27.8	32.5

Desta tabela temos:

$$\bar{S} = \frac{\sum_{i=1}^{10} S_i}{10} = 25$$

ou também

$$\bar{S} = \bar{X} + \bar{Y} = 10 + 15 = 25.$$

Para a variância de  $S$ , da tabela tem-se:

$$\text{Var}(S) = \sum_{i=1}^{10} \frac{(S_i - \bar{S})^2}{9} = 39.89$$

f.  $Z$  = Salario com desconto



$$\begin{aligned}
Z &= (Y - 0.084Y) + (X - 0.06X) = 0.92Y + 0.94X \\
\Rightarrow \bar{Z} &= 0.92\bar{Y} + 0.94\bar{X} \\
\Rightarrow \bar{Z} &= 23.2
\end{aligned}$$

Para a variância de  $S$ , da tabela tem-se:

$$Var(Z) = \sum_{i=1}^{10} \left( \frac{Z_i - \bar{Z}}{9} \right)^2 = 30.18$$