Dados da Companhia Milsa

Exemplo

Dados analisados: Informações sobre estado civil, grau de instrução, número de filhos, salário (fração do salário mínimo), idade (em anos e meses) e procedência de 36 funcionários da seção de orçamentos da Companhia Milsa.
Fonte: Morettin & Bussab, Estatística Básica.

N^o	Estado Civil	Instrução	Filhos	Salário	Idade	Região
1	solteiro	1^o grau	-	4,00	23 03	interior
2	casado	1^o grau	1	4,56	32 10	capital
⋮	⋮	⋮	⋮	⋮	⋮	⋮
36	casado	superior	3	23,30	42 02	interior

Obtendo Mediana e Quartis do número de filhos

A variável número de filhos foi observada na seguinte ordem:

-	1	2	-	-	0	-	-	1	-	2	-
-	3	0	-	1	2	-	-	1	-	-	0
2	2	-	0	5	2	-	1	3	-	2	3

Onde "-" corresponde a funcionários solteiros. Considerando apenas os funcionários casados, temos

1	2	0	1	2	3	0	1	2	1
0	2	2	0	5	2	1	3	2	3

Obtendo Mediana e Quartis do número de filhos

Ordenando as observações do maior para o menor, temos

$$\begin{array}{cccccccccc} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 5 \end{array}$$

Como $n = 20$ é par, a mediana é dada por

$$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

onde $x_{(k)}$ é a k-ésima observação ordenada.
Neste caso, a mediana do número de filhos dos funcionários casados é igual a

$$\frac{x_{(10)} + x_{(11)}}{2} = \frac{2+2}{2} = 2$$

Obtendo Mediana e Quartis do número de filhos

De modo semelhante, o primeiro quartil é dado por

$$Q_1 = \frac{x_{\left(\frac{n}{4}\right)} + x_{\left(\frac{n}{4}+1\right)}}{2} = \frac{1+1}{2}$$

E o terceiro quartil, por

$$Q_3 = \frac{x_{\left(\frac{3n}{4}\right)} + x_{\left(\frac{3n}{4}+1\right)}}{2} = \frac{2+2}{2}$$

E os valores são, respectivamente, $Q_1 = 1$ e $Q_3 = 2$.

Dureza de 30 peças de alumínio

Exemplo

Considere agora as seguintes observações, sobre a dureza de peças de alumínio.
Fonte: Hoaglin, Mosteller e Tukey, 1983, apud *Morettin & Bussab*, Estatística Básica.

53,0	70,2	84,3	69,5	77,8	87,5
53,4	82,5	67,3	54,1	70,5	71,4
95,4	51,1	74,4	55,7	63,5	85,8
53,5	64,3	82,7	78,5	55,7	69,1
72,3	59,5	55,3	73,0	52,4	50,7

Dureza de 30 peças de alumínio

As estatísticas descritivas da amostra, sobre localização e dispersão, são estas:

Média:	67,8	Mediana:	69,3
Desvio Padrão:	12,5	1^o Quartil:	55,3
Desvio Médio:	10,7	3^o Quartil:	77,0

Dureza de 30 peças de alumínio

Exemplo

Dados ordenados

1	2	3	4	5	6
50,70	51,10	52,40	53,00	53,40	53,50
54,10	55,30	55,70	55,70	59,50	63,50
64,30	67,30	69,10	69,50	70,20	70,50
71,40	72,30	73,00	74,40	77,80	78,50
82,50	82,70	84,30	85,80	87,50	95,40

Dureza de 30 peças de alumínio

- Média (Me) :
  $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{53,0 + 70,2 + ... + 52,4 + 50,7}{30} = 67,8.$
- Mediana (Md) : (média dos dois valores que ocupam a posição central no rol de dados ordenados, ou seja 15^a e 16^a posições), $Md = \frac{69,1+69,5}{2} = 69,3.$
- 1^o Quartil: (semelhante à mediana, mas consideramos a primeira metade do conjunto de valores), ou seja $Q1 = 55,3$ (elemento que ocupa a 8^a posição)
- 3^o Quartil: (semelhante à mediana, mas consideramos a primeira metade do conjunto de valores), ou seja $Q3 = 77,8$ (elemento que ocupa a 23^a posição)

Dureza de 30 peças de alumínio

- Variância:

$$
\begin{aligned}
Var &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\
&= \frac{1}{30}(53,0 - 67,8)^2 + (70,2 - 67,8)^2 + \ldots \\
&\quad +(54,4 - 67,8)^2 + (50,7 - 67,8)^2 = 156,5
\end{aligned}
$$

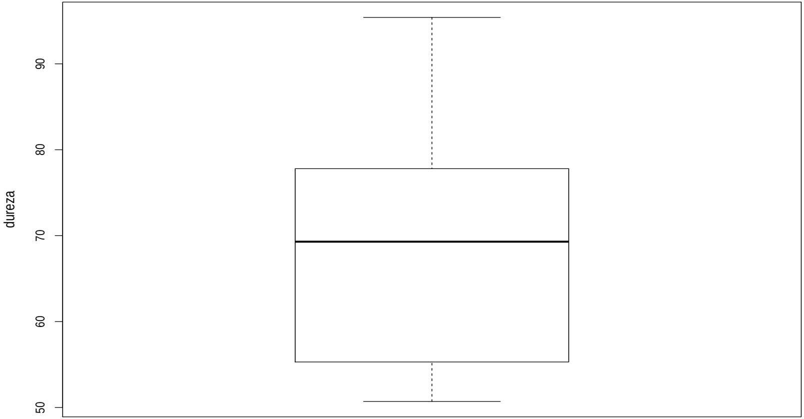
- Desvio-padrão: $DP = \sqrt{Var} = \sqrt{156,5} = 12,5$
- Construir uma distribuição de frequências para a variável dureza.

Distribuição de frequências

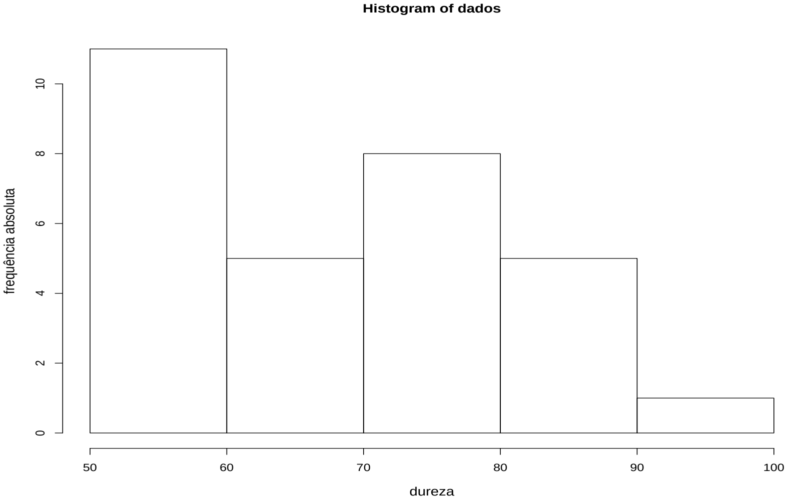
classe	fa	fr	fac	facr
$[50, 60)$	11,00	0,37	11,00	36,67
$[60, 70)$	5,00	0,17	16,00	53,33
$[70, 80)$	8,00	0,27	24,00	80,00
$[80, 90)$	5,00	0,17	29,00	96,67
$[90, 100)$	1,00	0,03	30,00	100,00

fa: frequência absoluta, fr: frequência relativa, fac: frequência acumulada crescente, facr: frequência acumulada crescente relativa. Além disso, $J_1 - (3/2)d_J = 21,55$ e $J_3 + (3/2)d_J = 111,55$ (limites para a identificação de outliers)

Box-plot da dureza



Histograma da dureza



Mediana e Moda

Mediana: Dado que está na posição central, ou média dos valores das posições centrais, dependendo do valor de n (o número de observações) ser par ou ímpar.

Moda: Valor mais frequente. Se os dados estão tabelados, pode-se aproximar pelo x^* (o ponto médio) do intervalo com o maior f_j. Porém deve-se ter em conta que os dados assim apresentados podem não ter uma moda, e logo esta é uma aproximação imperfeita.

Dados Tabelados

Exemplo

Considere novamente os dados de *Morettin & Bussab*: Estatística Básica.

Nº	Estado Civil	Instrução	Filhos	Salário	Idade	Região
1	solteiro	1º grau	-	4,00	23 03	interior
2	casado	1º grau	1	4,56	32 10	capital
3	casado	1º grau	2	5,25	36 05	capital
4	solteiro	2º grau	-	5,73	20 10	outro
5	solteiro	1º grau	-	6,26	40 07	outro
⋮	⋮	⋮	⋮	⋮	⋮	⋮
36	casado	superior	3	23,30	42 02	interior

Dados Tabelados

Em forma de tabela, o salário pode ser apresentado como:

Intervalo	x^*	n	f	N	F
$[4 - 8)$	6	10	10	0,27	0,27
$[8 - 12)$	10	12	22	0,33	0,61
$[12 - 16)$	14	8	30	0,22	0,83
$[16 - 20)$	18	5	35	0,14	0,97
$[20 - 24]$	22	1	36	0,03	1

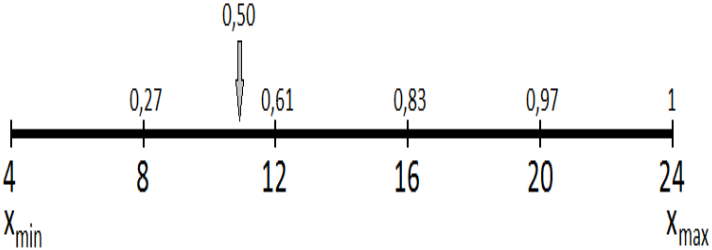
N: frequência relativa ; F: frequência relativa acumulada

Dados Tabelados - Medidas de Posição Central e Dispersão

Dados na Lista	Dados na Tabela
$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$	$\bar{x} = \sum_{j=1}^{k} x_j^* \frac{n_j}{n} = \sum_{j=1}^{k} x_j^* f_j$
$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$	$s^2 = \sum_{j=1}^{k} (x_j^* - \bar{x})^2 \frac{n_j}{n}$
$s = \sqrt{s^2}$	$s = \sqrt{s^2}$
$CV = \frac{s}{\bar{x}}$	$CV = \frac{s}{\bar{x}}$

onde k é o número de intervalos de classe, e x_j^* é o ponto médio do intervalo de classe $(a_j - b_j]$.

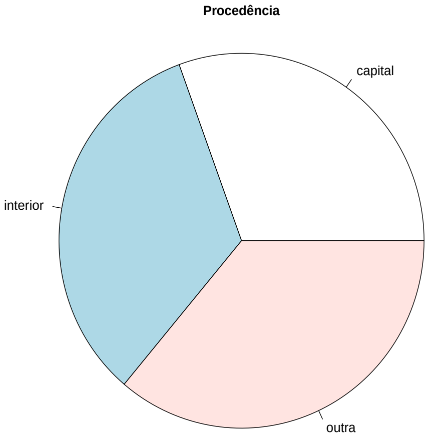
Dados Tabelados - Cálculo da Mediana



Podemos obter o valor da mediana, ν, com uma regra de três:

$$\frac{12-8}{0{,}61-0{,}27} = \frac{\nu-8}{0{,}5-0{,}27} \Leftrightarrow \nu = 10{,}7059$$

Representação gráfica de variáveis qualitativas



Representação gráfica de variáveis qualitativas