Exemplo Foram observadas quatro amostras de duas colheitadeiras de milho, quanto a porcentagem de quebra de sementes, conforme a tabela a seguir:

	Colheitadeira	
Amostra	A	B
1	5	0
2	4	3
3	5	7
4	6	10
média	5	5

As medidas de dispersão

são estatísticas descritivas que visam fornecer o grau de variabilidade das observações em relação a um valor central (geralmente a média aritmética).

Principais medidas de dispersão:

- Amplitude
- Amplitude Interquartílica
- Desvio Médio
- **Variância**
- **Desvio Padrão**
- Coeficiente de Variação

Amplitude

$$A_x = \max(x) - \min(x)$$

	Colheitadeira	
Amostra	A	B
1	5	0
2	4	3
3	5	7
4	6	10
média	5	5

Colheitadeira A

$A = 6 - 4 \Rightarrow A = 2$

Colheitadeira B

4	6	10
média	5	5

Colheitadeira B

$A = 10 - 0 \Rightarrow A = 10$

Conferindo...

Exemplo Considerando-se o exemplo de porcentagem de quebra de sementes de milho, temos:

Colheitadeira	Amplitude
A	6-4 = 2
B	10-0 = 10

A amplitude interquartílica é dada por:

$$P_{100p} = \begin{cases} \frac{x_{[np]} + x_{[np+1]}}{2}, & \text{se } np \text{ for inteiro} \\ x_{[\text{int}(np)+1]}, & \text{se } np \text{ for não inteiro} \end{cases}$$

$$AIC = Q_3 - Q_1$$

$P_{25} = Q_1 \quad P_{75} = Q_3$

Exemplo Considerando-se o exemplo de porcentagem de quebra de sementes de milho, temos:

$n = 4$

nº inteiros

Rol:

Colheitadeira		x_1	x_2	x_3	x_4
	A	4	5	5	6
	B	0	3	7	10

$np = 4 \cdot 0{,}25 = 1$

$np = 4 \cdot 0{,}75 = 3$

Colheitadeira A

$Q_1 = \frac{x_1 + x_2}{2} = \frac{4+5}{2} = 4{,}5$

$Q_3 = \frac{x_3 + x_4}{2} = \frac{5+6}{2} = 5{,}5$

$AIQ = 5{,}5 - 4{,}5 = 1{,}0$

Colheitadeira B

$Q_1 = \frac{x_1 + x_2}{2} = \frac{0+3}{2} = 1{,}5$

$Q_3 = \frac{x_3 + x_4}{2} = \frac{7+10}{2} = 8{,}5$

$AIQ = 8{,}5 - 1{,}5 = 7{,}0$

Conferindo...

Colheitadeira	Q_1	Q_3	AIQ
A	4,5	5,5	**1,0**
B	1,5	8,5	**7,0**

- Desvio de uma observação em relação a uma constante:

$$d_i = x_i - k$$

- Desvio de uma observação em relação à média aritmética:

$$e_i = x_i - \bar{x}$$

$$\sum_{i=1}^{n} e_i = ?$$

	Colheitadeira	
Amostra	A	B
1	5	0
2	4	3
3	5	7
4	6	10
média	5	5

	e_A	e_B	$\|e_A\|$	$\|e_B\|$
	0	-5	0	5
	-1	-2	1	2
	0	2	0	2
	1	5	1	5
Σ	0	0	2	14

- **Desvio Médio**

$$Dm_x = \frac{1}{n}\sum_{i=1}^{n} |e_i| = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$

Exemplo: Considerando-se o exemplo de porcentagem de quebra de sementes de milho, temos:

Colheitadeira	Desvios				$\sum_{i=1}^{4} \|e_i\|/4$
A	0,0	-1,0	0,0	1,0	2/4 = 0,5
B	-5,0	-2,0	2,0	5,0	14/4 = 3,5

Conferindo...

Colheitadeira	Desvios				$\sum_{i=1}^{4} \|e_i\|/4$
A	0,0	-1,0	0,0	1,0	2,0/4 = **0,5**
B	-5,0	-2,0	2,0	5,0	14,0/4 = **3,5**

μ = média populacional
média [illegible]

Colheitadeira	Desvios				$\sum_{i=1} \lvert e_i \rvert/4$
A	0,0	-1,0	0,0	1,0	2,0/4 = **0,5**
B	-5,0	-2,0	2,0	5,0	14,0/4 = **3,5**

μ = popu...

$\bar{x}$ = média amostral

Variância

é a média dos quadrados dos desvios em relação à média aritmética.

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

σ^2 variância populacional

s^2 = variância amostral

Estimador da variância populacional:

$$S_X^2 = \frac{1}{\boxed{n-1}}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

	Colheitadeira	
Amostra	A	B
1	5	0
2	4	3
3	5	7
4	6	10
média	5	5

A $(x_i - \bar{x})^2$	B $(x_i - \bar{x})^2$
0	25
1	4
0	4
1	25
$\sum$ 2	$\sum$ 58

$n = 4$

$S_A^2 = \frac{2}{3}$

$S_A^2 = 0,6667\ \%^2$

$S_B^2 = \frac{58}{3}$

$S_B^2 = 19,3333\ \%^2$

Conferindo...

Exemplo:

- **Colheitadeira A**:

$$S_{X_1}^2 = \frac{\sum_{i=1}^{4}(x_i - \bar{x})^2}{4-1}$$

$$= \frac{(5-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2}{3} = 0,67(\%)^2$$

- **Colheitadeira B**:

$$S_{X_2}^2 = \frac{\sum_{i=1}^{4}(x_i - \bar{x})^2}{4-1}$$

$$S^2_{X_2} = \frac{\sum_{i=1}^{4}(x_i - \bar{x})^2}{4-1}$$

$$\frac{(0-5)^2 + (3-5)^2 + (7-5)^2 + (10-5)^2}{3} = 19,33(\%)^2$$

Dados agrupados em tabelas de frequências

$$S^2_X = \frac{1}{n-1}\sum_{i=1}^{k} f_i e_i^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

em que k corresponde ao número de diferentes valores para a variável e $n = \sum_{i=1}^{k} f_i$

Ou ainda,

$$S^2_X = \frac{1}{n-1}\left[\sum_{i=1}^{k} f_i x_i^2 - \frac{\left(\sum_{i=1}^{k} f_i x_i\right)^2}{n}\right]$$

Exemplo: Foram observadas 76 goiabas quanto ao número de pintas pretas, cujos dados são apresentado na tabela a seguir:

$x_i \cdot f_i$	X_i	f_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
0	0	21	1,69	35,49
33	1	33	0,09	2,97
26	2	13	0,49	6,37
12	3	4	2,89	11,56
8	4	2	7,29	14,58
12	6	2	22,09	44,18
8	8	1	44,89	44,89
$\Sigma = 99$	Total	76		160,04

$- n = \sum f_i$

$n = 76$

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{99}{76}$$

$\bar{x} = 1,3$

$$S^2 = \frac{160,04}{75}$$

$S^2 = 2,13$ pintas^2

Conferindo...

- Média

$$\bar{x} = \frac{0 \times 21 + 1 \times 33 + \ldots + 8 * 1}{76}$$
$$= 1,30 \text{ pintas}$$

- Variância

$$S_X^2 = \frac{21(0-1,30)^2 + 33(1-1,30)^2 + \ldots + 1(8-1,30)^2}{76-1}$$

$$= 2,13 \text{ pintas}^2$$

Dados agrupados em tabelas de classes frequências

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i e_i^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i(m_i - \bar{x})^2$$

em que k corresponde ao número de diferentes valores para a variável e $n = \sum_{i=1}^{k} f_i$

Ou ainda,

$$S_X^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} f_i m_i^2 - \frac{\left(\sum_{i=1}^{k} f_i m_i\right)^2}{n}\right]$$

Exemplo: Foram avaliados 60 novilhos de dois anos quanto ao consumo de matéria seca (kg), cujos dados são apresentados na tabela a seguir.

Consumo (kg)	m_i	f_i	$m_i \cdot f_i$	$(m_i-\bar{x})^2 \cdot f_i$
10,0 ⊢ 10,4	10,2	10	102	2,5
10,4 ⊢ 10,8	10,6	30	318	0,3
10,8 ⊢ 11,2	11,0	20	220	1,8
Total		60	640	4,6

$n=60$

$$\bar{x} = \frac{\sum m_i \cdot f_i}{n} = \frac{640}{60} \qquad \bar{x} = 10,7 \text{ kg}$$

$$S^2 = \frac{\sum (m_i - \bar{x})^2 \cdot f_i}{n-1} = \frac{4,6}{59} \qquad S^2 = 0,078 \text{ kg}^2$$

- Média

$$\bar{x} = \frac{10,2 \times 10 + 10,6 \times 30 + 11,0 \times 20}{60} = 10,67 \text{ kg}$$

- Variância

$$S^2 = \frac{10 \times (10,2 - 10,67)^2 + 30 \times (10,6 - 10,67)^2 + 20 \times (11,0 - 10,67)^2}{60 - 1} = 4,53/59 = 0,0768 \text{ kg}^2$$

O desvio padrão

corresponde à raiz quadrada da variância,

$$S_X = \sqrt{S_X^2}$$

O desvio padrão tem a mesma unidade dos dados originais

Exemplo: Considerando os dados de consumo de matéria seca de novilhos de 2 anos:

- Variância

$$S^2 = \frac{10 \times (10,2 - 10,67)^2 + 30 \times (10,6 - 10,67)^2 + 20 \times (11,0 - 10,67)^2}{60 - 1} = 4,53/59 = 0,0768 \text{ kg}^2$$

$S^2 = 0,0768 \Rightarrow S = \sqrt{0,0768} \Rightarrow S = 0,2771 \text{ kg}$

Conferindo...

$$S_X = \sqrt{0,0768} = 0,2772 \text{ kg}$$

desvio padrão

O coeficiente de variação é dado por

O **coeficiente de variação** é dado por

$$CV_X = 100\frac{S_X}{\bar{x}}$$

→ desv padrão

média

- comparação de dispersões quando as unidades de medida são diferentes

Orientação

$$CV \leq 10\% \Rightarrow \text{baixo}$$
$$10\% < CV \leq 20\% \Rightarrow \text{médio}$$
$$20\% < CV \leq 30\% \Rightarrow \text{alto}$$
$$CV > 30\% \Rightarrow \text{muito alto}$$

Exemplo: Considerando os dados de consumo de matéria seca de novilhos de 2 anos:

$$S_X = \sqrt{0,0768} = 0,2772 \text{ kg}$$

$$\bar{x} = 10,67 \text{ kg}$$

$$CV = \frac{100 \cdot 0,2772}{10,67} = 2,6\ \%$$

baixo coeficiente de variação

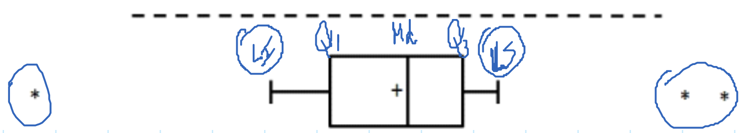
Conferindo...

$$CV_X = 100\frac{0,2772}{10,67} = 2,60\%$$

<u>Gráfico de Caixas (Box-Plot)</u>

Construção de um gráfico de caixas

1. Calcular o primeiro quartil (Q_1), a mediana (Md) e o terceiro quartil (Q_3);
2. Calcular a AIQ $= Q_3 - Q_1$;
3. Verificar a existência de observações atípicas, ou seja, valores menores do que $\boxed{Q_1 - 1,5\text{AIQ}}$ ou maiores do que $\boxed{Q_3 + 1,5\text{AIQ}}$ representados individualmente no gráfico de caixas por $*$; → discrepantes
4. Calcular os limites inferior e superior dos dados sem considerar as observações atípicas;
5. Construir o gráfico seguindo o esquema a seguir:



Exemplo: Para os valores observados de produção média diária de leite, tem-se:

$n = 10$

X_1	X_2	X_3	X_4	X_5
8,78	9,34	9,80	9,90	9,90
9,95	10,00	10,34	11,75	15,00
X_6	X_7	X_8	X_9	X_{10}

parte inteira

$P_{25} \to np = 10 \cdot 0,25 = 2,5$

$X_{(2+1)} = X_3$

$P_{50} \to 10 \cdot 0,5 = 5$

$P_{50} = \dfrac{X_5 + X_6}{2}$

$P_{75} = 10 \cdot 0,75 = 7,5$

$X_{(7+1)} = X_8$

$Q_1 = X_3 = 9,8$

$Md = \dfrac{X_5 + X_6}{2} = \dfrac{9,90 + 9,95}{2} = 9,925$

$Q_3 = X_8 = 10,34$

$AIQ = 10,34 - 9,8 = 0,54$

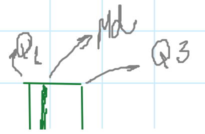
discrepantes

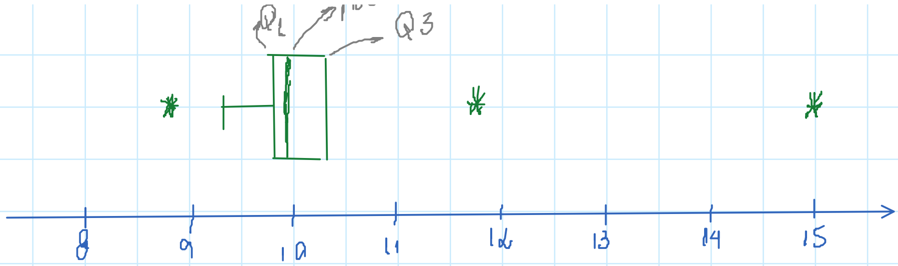
$Q_1 - 1,5 \cdot AIQ = 9,8 - 1,5 \cdot 0,54 = 8,99$

8,78 é discrepante $\quad LI = 9,34$

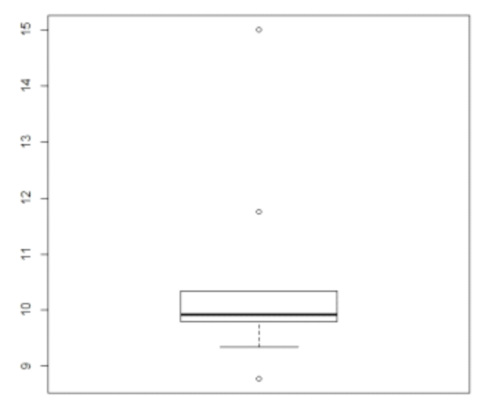
$Q_3 + 1,5 \cdot AIQ = 10,34 + 1,5 \cdot 0,54 = 11,15$

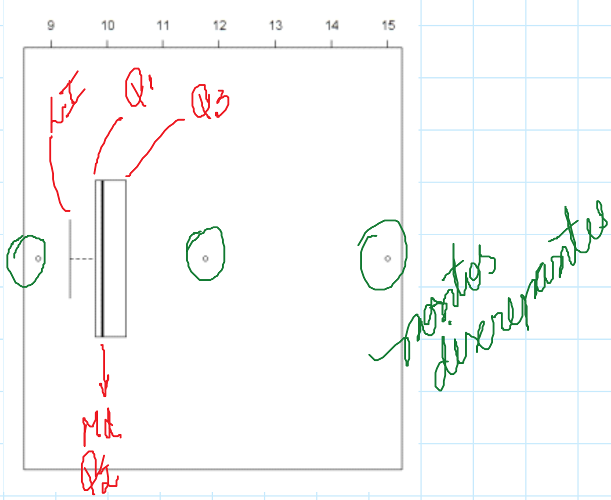
$\left.\begin{matrix} 11,75 \\ 15,00 \end{matrix}\right]$ são discrepantes $\quad LS = 10,34$




Q1
Q3
8
9
10
11
12
13
14
15

Conferindo...


9
10
11
12
13
14
15


9
10
11
12
13
14
15
LI
Q1
Q3
Md
Q2
pontos discrepantes