Medidas Estatísticas

NILO FERNANDES VARELA

Tendência Central

- Medidas que orientam quanto aos valores centrais.
- Representam os fenômenos pelos seus valores médios, em torno dos quais tendem a se concentrar os dados.
- Também chamados de **centro da distribuição.**

1. **Média**
2. **Moda**
3. **Mediana**

Média

Média Aritmética

Dados Não Agrupados

Seja Y = (y1, y2, ..., yn) e (N,n) = Quantidade de variáveis em Y, temos que:

$$\text{Parâmetro: } \overline{Y} \; ou \; \mu = \frac{\sum y}{N}$$

$$\text{Estimativa: } \overline{y} \; ou \; m = \frac{\sum y}{n}$$

Exemplo: considerando {3, 7, 8, 10, 11} como uma amostra:

$$\overline{y} = \frac{3+7+8+10+11}{5} = 7,8$$

Média

Média Aritmética

Dados Agrupados

Quando os dados estão agrupados em uma distribuição de frequência temos:

Y = (y1, y2, ..., yn) ponderados pelas respectivas frequências absolutas **F=(F1, F2,...,Fn).**

$$\bar{y} \; ou \; m = \frac{\sum y \cdot F}{\sum F}$$

Média

Média Aritmética

Dados Agrupados

Exemplo:

Y	Fi	Y*Fi
1	1	1
2	3	6
3	5	15
4	1	4
Σ	10	26

$$\bar{y} \; ou \; m = \frac{\sum y \cdot F}{\sum F} = \frac{26}{10} = 2{,}6$$

Média

Média Aritmética

Dados Agrupados

Exemplo:

Idade	Fi	Y	Y*Fi
02 ⊢ 04	5	3	15
04 ⊢ 06	10	5	50
06 ⊢ 08	14	7	98
08 ⊢ 10	8	9	72
10 ⊢ 12	3	11	33
Σ	40		268

$$\bar{y} \text{ ou } m = \frac{\sum y \cdot F}{\sum F} = \frac{268}{40} = 6{,}7$$

Média

Média Geral

- Sejam y1 , y2 , ..., yk as **estimativas das médias aritméticas** de K séries.
- Sejam n1, n2, ..., nk os **números de termos destas séries**, respectivamente.

A média aritmética da série formada pelos termos da K séries é dada pela fórmula:

$$\boxed{\bar{y} \ ou \ m = \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2 + \ldots + n_k \bar{y}_k}{n_1 + n_2 + \ldots + n_k}}$$

Média

Média Geral

Exemplo

{4, 5, 6, 7, 8}	n1 = 5	$\bar{Y}1 = 6$
{1, 2, 3}	n2 = 3	$\bar{Y}2 = 2$
{9, 10, 11, 12, 13}	n3 = 5	$\bar{Y}3 = 11$

$$\overline{Y} = \frac{n_1 \overline{y}_1 + n_2 \overline{y}_2 + \ldots + n_k \overline{y}_k}{n_1 + n_2 + \ldots + n_k} = \frac{5 \cdot 6 + 3 \cdot 2 + 5 \cdot 11}{5 + 3 + 5} = 7$$

Média

Média Geométrica

Usada para médias proporcionais de crescimento quando uma **medida subsequente depende de medidas prévias**.

Sejam y1, y2, ..., yn, valores de Y associados às respectivas frequências absolutas F1, F2, ..., Fn. A média geométrica de Y é definida por:

$$MG \ ou \ mg = \sqrt[n]{y_1^{F_1} \cdot y_2^{F_2} \cdot \ldots \cdot y_n^{F_n}}$$

Média

Média Geométrica

Exemplo

Amostra: {3, 6, 12, 24, 48}

$$\boxed{mg = \sqrt[5]{3 \cdot 6 \cdot 12 \cdot 24 \cdot 48} = \sqrt[5]{248.832} = 12}$$

Média

Média Harmônica

Usada para médias de crescimento e proporções de velocidade.

- Sejam y1, y2, ..., yn, valores de Y
- Sejam F1, F2, ..., Fn, frequências absolutas.

A média harmônica de Y é definida por:

$$MH \ ou \ mh = \frac{n}{\frac{F_1}{y_1} + \frac{F_2}{y_2} + \ldots + \frac{F_n}{y_n}} = \frac{n}{\sum_{i=1}^{n} \frac{F_i}{y_i}}$$

Média

Média Harmônica

Exemplo

Amostra: {2,5,8}

$$mh = \frac{3}{\frac{1}{2}+\frac{1}{5}+\frac{1}{8}} = 3,63$$

Média

Vantagens

- Fácil de compreender e calcular
- Utiliza todos os valores da série
- É um valor único
- É fácil de ser incluída em expressões matemáticas
- Pode ser determinada nas escalas: intervalar e proporcional.

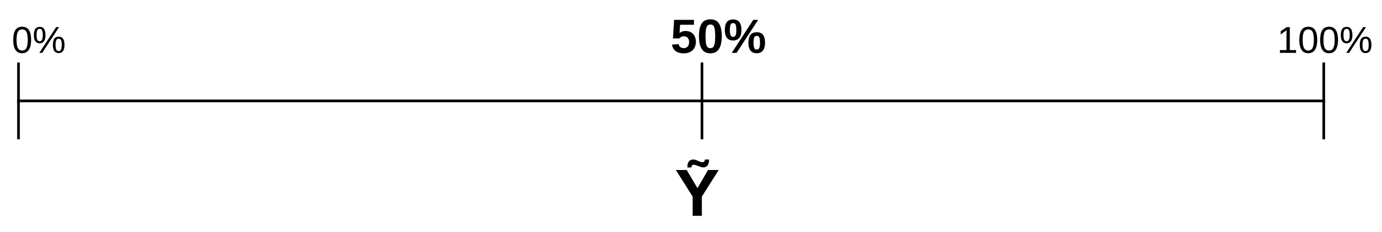
Desvantagens

- Muito afetada por valores extremos
- Necessário conhecer todos os valores da série.

Mediana

Medida de tendência muito usada quando o interesse é a **determinação do valor que separa a série de dados em duas partes iguais**, 50% situados acima e 50% situados abaixo da medida.

Notação adotada: ($\tilde{Y}$ ou MD) para o parâmetro e ($\tilde{y}$ ou md) para a estimativa.



Mediana

Cálculo da mediana para variável discreta

Se n for ímpar:

A mediana é o elemento central.

$$\frac{n+1}{2}$$

Se n for par:

A mediana é a média dos dois elementos centrais.

$$\text{Média}\left[\frac{n}{2}, \frac{n}{2}+1\right]$$

Mediana

Cálculo da mediana para variável discreta

Exemplo

Yi	Fi	Fac
1	1	1
2	3	4
3	5	9
4	2	11
Σ	11	

1	2	3	4	5	6	7	8	9	10	11
1	**2**	**2**	**2**	**3**	**3**	**3**	**3**	**3**	**4**	**4**

n=11

$$\frac{n+1}{2}$$

$$\frac{11+1}{2} = 6º$$

$\tilde{y}$ = **3**

Mediana

Cálculo da mediana para variável discreta

Exemplo

Yi	Fi	Fac
82	5	5
85	10	15
87	15	30
89	8	38
90	4	42
Σ	42	

n=42

média $\left[\frac{n}{2}, \frac{n}{2}+1\right]$

$$\frac{42}{2}, \frac{42}{2}+1$$

(elementos 21º e 22º)

21º=87 22º=87

ỹ = **87**

Mediana

Cálculo da mediana para variável contínua

1. Calcular $\frac{n}{2}$
2. Usar a Fac para identificar a classe que contém a mediana **(classe md).**
3. $\tilde{y} = l_{md} + \frac{\left(\frac{n}{2} - \sum f\right).h}{F_{md}}$

l_{md} = Limite inferior da classe md
n = Tamanho da série
$\sum f$ = Soma das frequências anteriores à classe md
h = Amplitude da classe md
Fmd = Frequência da classe md

Mediana

Cálculo da mediana para variável contínua

Exemplo

Classe	Fi	Fac
35 ⊢ 45	5	5
45 ⊢ 55	12	17
55 ⊢ 65	18	35
65 ⊢ 75	14	49
75 ⊢ 85	6	55
85 ⊢ 95	3	58
Σ	58	268

1. $\frac{58}{2} = 29°$
2. Classe md = 3º

l_{md}= 55
n = 58
$\sum f$ = 17
h = 10
Fmd = 18

3. $\tilde{y} = 55 + \frac{\left(\frac{58}{2} - 17\right).10}{18} = \mathbf{61,67}$

Mediana

Vantagens

- Fácil de compreender e aplicar
- Não é afetada por valores extremos
- É um valor único
- Pode ser determinada nas escalas: ordinal, intervalar e proporcional.

Desvantagens

- É difícil de ser incluída em expressões matemáticas
- Não usa todos os valores da série.

Moda

Medida de tendência central muito usada quando o interesse é o **valor mais frequente da série**.

Notação adotada: (MO) para o parâmetro e (mo) para a estimativa.

- Série sem moda : Série amodal
- Mais de uma moda : Série multimodal

Moda

Exemplo

Distribuição sem agrupamento de classes

y_i	253	245	248	251	307
F_i	7	17	23	20	8

mo = 248

Moda

Exemplo

Distribuição com agrupamento de classes

1. Ache a classe modal
2. Aplique a fórmula de Czuber

$$\mathbf{m}\boldsymbol{o} = \boldsymbol{l} + \frac{\boldsymbol{\Delta_1}}{\boldsymbol{\Delta_1} + \boldsymbol{\Delta_2}}\ .\boldsymbol{h}$$

l	= Limite inferior da classe mo
Δ_1	= diferença entre a frequência da classe modal e a imediatamente anterior
Δ_2	= diferença entre a frequência da classe modal e a imediatamente posterior
h	= Amplitude da classe

Moda

Exemplo

Distribuição com agrupamento de classes

$Classes$	0 ⊢ 1	1 ⊢ 2	2 ⊢ 3	3 ⊢ 4	4 ⊢ 5	Σ
F_i	3	10	17	8	5	43

1. Classe modal = 3º (2 ⊢ 3)
2. $\mathbf{m}o = \mathbf{2} + \frac{\mathbf{7}}{\mathbf{7+9}}\,.\,\mathbf{1} = \mathbf{2,44}$

Moda

Vantagens

- Fácil de compreender e calcular
- Não é afetada por valores extremos
- Pode ser aplicada em todas as escalas: nominal, ordinal, intervalar e proporcional.

Desvantagens

- Pode estar afastada do centro dos valores
- É difícil de ser incluída em expressões matemáticas
- Não usa todos os valores da série
- A variável pode ter mais de uma moda (bimodal ou multimodal)
- Algumas variáveis não possuem moda.

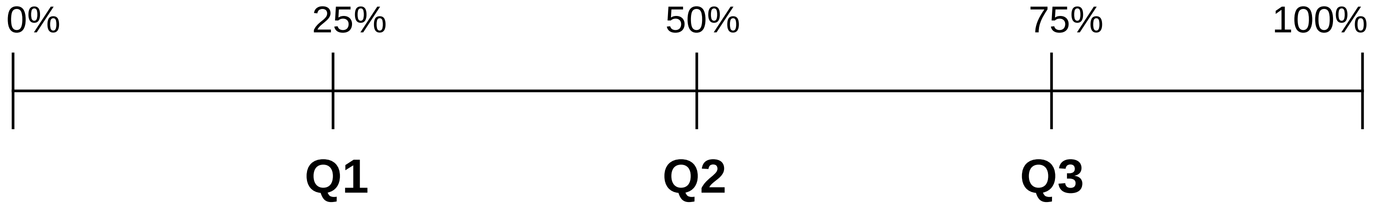
Medidas de Posição ou Separatrizes

- Genericamente denominadas quantis.
- Orientam quanto à posição na distribuição.
- Permitem determinar valores que particionam a série de n observações em partes iguais.

1. Quartis
2. Decis
3. Percentis

Quartis

Dividem uma série em **4 partes iguais**.



Notação adotada: (Q) para o parâmetro e (q) para a estimativa.

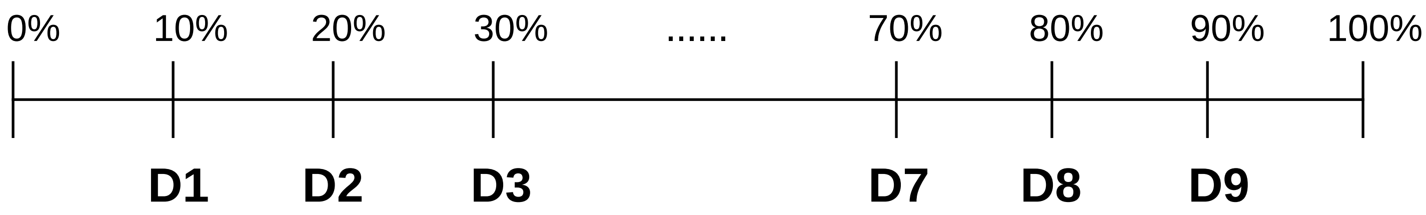
Quartis

$$\boldsymbol{q_i = l_{q_i} + \frac{\left(\frac{i.n}{4} - \sum f\right).h}{F_{q_i}}}$$

l_{q_i} = limite inferior da classe q_i (i = 1, ..., 3)
i = 1 para q_1, ..., 3 para q_3
n = tamanho da série
Σ f = soma das frequências anteriores à classe q_i
h = amplitude da classe q_i
F_{q_i} = frequência da classe q_i

Decis

Dividem uma série em **10 partes iguais**.



Notação adotada: (D) para o parâmetro e (d) para a estimativa.

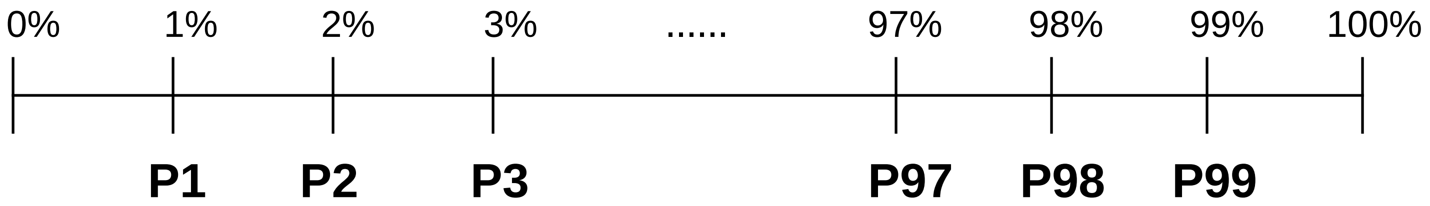
Decis

$$\boldsymbol{d_i = l_{d_i} + \frac{\left(\frac{i.n}{10} - \sum f\right).h}{F_{d_i}}}$$

l_{d_i} = limite inferior da classe q_i (i = 1, ..., 9)
i = 1 para d_1, ..., 9 para d_9
n = tamanho da série
Σ f = soma das frequências anteriores à classe d_i
h = amplitude da classe d_i
F_{d_i} = frequência da classe d_i

Percentis

Dividem uma série em **100 partes iguais**.



Notação adotada: (P) para o parâmetro e (p) para a estimativa.

Percentis

$$\boldsymbol{p_i = l_{p_i} + \frac{\left(\frac{i.n}{100} - \sum f\right).h}{F_{p_i}}}$$

l_{p_i} = limite inferior da classe q_i (i = 1, ..., 99)
i = 1 para p_1, ..., 99 para p_{99}
n = tamanho da série
Σ f = soma das frequências anteriores à classe p_i
h = amplitude da classe p_i
F_{p_i} = frequência da classe p_i

Medidas de Dispersão

- Usadas para avaliar o **grau de variabilidade** ou dispersão **dos valores da série em torno da média**.

1. Amplitude
2. Desvio médio
3. Desvio quadrático médio
4. Variância
5. Desvio padrão

Amplitude Total

Medida da dispersão dos dados definida como a diferença entre o maior e o menor dos valores da série.

Notação: (AT) para o parâmetro (at) para a estimativa.

- Quantifica a dispersão dos dados.
- Permite distinguir séries de dados em relação à homogeneidade:
  - Séries homogêneas: menor valor da amplitude total
  - Séries heterogêneas: maior valor da amplitude total

Amplitude Total

$$\textbf{At ou at} = \ y_{max} - y_{min}$$

Ex:

{10, 12, 20, 22, 25, 33, 38}

At = 38 – 10 = 28

Desvio

Diferença entre o valor e a média.

Seja **Y = (y1, y2, ..., yn)** e **Ȳ** a média aritmética de Y, o desvio ***i*** é dado por:

$$D_i = (y_i - \bar{Y})$$
$$d_i = (y_i - \bar{y})$$

Desvio

Ex: Y = {3,4,5,6,7}
$\bar{Y} = 5$
$D_4 = 6 - 5 = 1$

Obs: $\sum Di = \sum di = 0$

Desvio Médio

Média dos desvios absolutos em relação à média aritmética.

Notação: (DM) para o parâmetro e (dm) para a estimativa.

- Quantifica a dispersão dos dados.
- Permite distinguir séries de dados em relação à homogeneidade:
  - Séries homogêneas: menor valor de desvio médio
  - Séries heterogêneas: maior valor de desvio médio

Desvio Médio

$$\boldsymbol{DM = \frac{\sum |D|}{N}}$$

$$\boldsymbol{dm = \frac{\sum |d|}{n}}$$

Desvio Médio

Exemplo

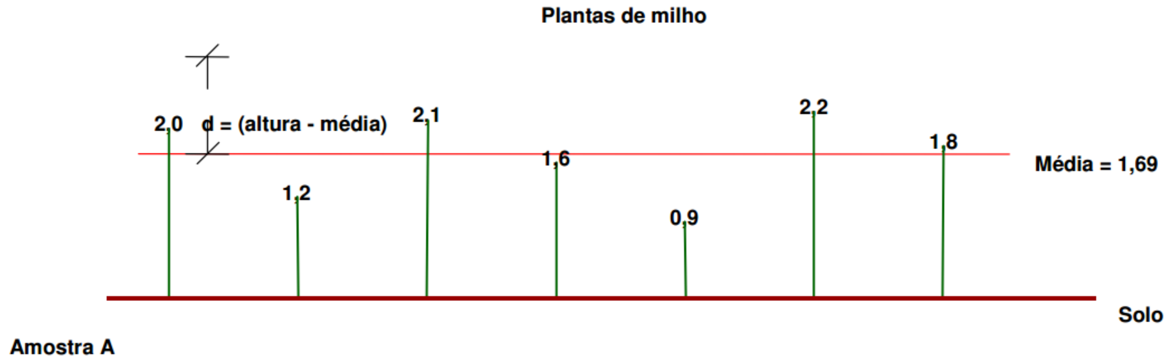


Figura 5.2 – Ilustração de uma amostra de plantas de milho.

$$\boldsymbol{dm = \frac{|2,0-1,69|+\cdots+|1,8-1,69|}{7} = 0,39m}$$

Desvio Quadrático Médio

Média dos quadrados dos desvios em relação à média aritmética.

Notação: (DQM) para o parâmetro e (dqm) para a estimativa.

- Quantifica a dispersão dos dados.
- Permite distinguir séries de dados em relação à homogeneidade:
  - Séries homogêneas: menor valor de desvio quadrático médio
  - Séries heterogêneas: maior valor de desvio quadrático médio

Desvio Quadrático Médio

$$\boldsymbol{DQM} = \frac{\sum(\boldsymbol{D})^2}{\boldsymbol{N}}$$

$$\boldsymbol{dqm} = \frac{\sum(\boldsymbol{d})^2}{\boldsymbol{n}}$$

Desvio Quadrático Médio

Exemplo

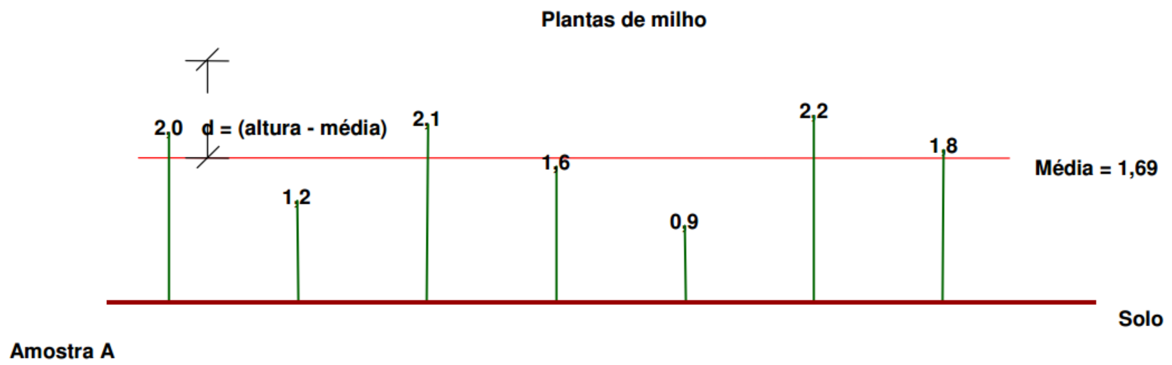


Figura 5.2 – Ilustração de uma amostra de plantas de milho.

$$\boldsymbol{dqm = \frac{(2,0-1,69)^2+\cdots+(1,8-1,69)^2}{7} = 0,19m^2}$$

Variância

Razão entre a soma de quadrados dos desvios de cada valor em relação à média aritmética, $\boldsymbol{\Sigma d^2}$, e o número de elementos da série, **N para populações ou n-1 para amostras.**

Notação: (σ^2) para o parâmetro e (s^2) para a estimativa.

- Quantifica a dispersão dos dados.
- Medida da totalidade dos desvios em relação à média.

- Permite distinguir séries de dados em relação à homogeneidade:
  - Séries homogêneas: menor valor da variância
  - Séries heterogêneas: maior valor da variância

Variância

Parâmetro: $$\boldsymbol{\sigma^2 = \frac{\sum D^2}{N}}$$

Estimativa:

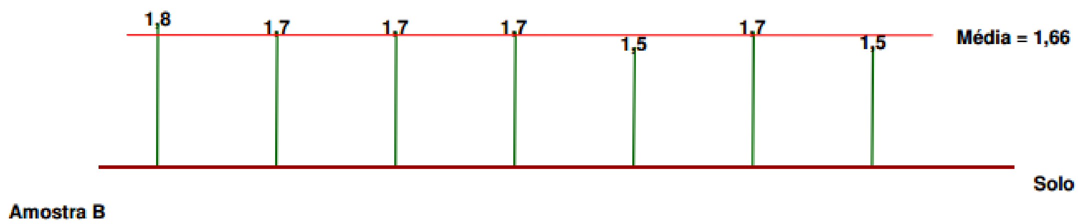
Quando $\bar{Y}$ é conhecido(caso raro): $$\boldsymbol{s^2 = \frac{\sum D^2}{n}}$$

Quando $\bar{Y}$ não é conhecido(caso comum): $$\boldsymbol{s^2 = \frac{\sum d^2}{n-1}}$$

Variância

Exemplo:

$$s_A^2 = \frac{(2,0 - 1,69)^2 + \cdots + (1,8 - 1,69)^2}{7 - 1} = 0,23m^2$$



$$s_B^2 = \frac{(1,8 - 1,66)^2 + \cdots + (1,5 - 1,66)^2}{7 - 1} = 0,01m^2$$

Desvio Padrão

Raiz quadrada da variância

Notação: (σ) para o parâmetro e (s) para a estimativa.

- Quantifica a dispersão dos dados.
- Junto a variância, o desvio padrão é uma das medidas mais usadas para quantificar a dispersão dos dados em torno da média.

Parâmetro: $\sigma = \sqrt{\sigma^2}$

Estimativa: $s = \sqrt{s^2}$

Desvio Padrão Relativo e Coeficiente de Variação

Razão entre o desvio padrão e a média aritmética.

Notação: (DPR e CV) para os parâmetros e (dpr e cv) para as estimativas.

- Quantificam a dispersão relativa dos dados em relação à média aritmética.

Parâmetro: $DPR = \frac{\sigma}{\bar{\bar{Y}}}$ $\quad$ $CV = \frac{\sigma}{\bar{\bar{Y}}}.100$

Estimativa: $dpr = \frac{s}{\bar{\bar{y}}}$ $\quad$ $cv = \frac{s}{\bar{\bar{y}}}.100$