



Universidade Federal de Mato Grosso

# AULA 02 – Estatística

**Prof. Lucas Bianchi**

Cuiabá, 27 de junho de 2016

## Na aula de hoje, estudaremos:

- ✓ Medidas de posição;
- ✓ Medidas de dispersão

# Somatório

Um somatório é um operador matemático que nos permite representar facilmente somas muito grandes ou até infinitas. É representado com a letra grega sigma , e é definido por:

$$\sum = \text{somatório} \quad \Rightarrow \quad \sum_{i=1}^n x_i$$

em que corresponde a soma dos termos  $x_i$ , e o índice  $i$  varia de 1 a  $n$ .

## Somatório de uma constante

Se  $k$  é uma constante, então:

$$\sum_{i=1}^n k = k + k + k + \cdots + k = n \cdot k$$

Somatório do produto de uma constante por uma variável

Se  $k$  é uma constante e  $x_i$  uma variável, então

$$\sum_{i=1}^n kx_i = kx_1 + kx_2 + kx_3 + \cdots + kx_n = k(1 + 2 + 3 + \cdots + n)$$

$$k \cdot \sum_{i=1}^n x_i$$

## Somatório de uma soma algébrica



O somatório de uma soma de variáveis é igual à soma dos somatórios de cada variável

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

se  $a$  e  $b$  são constantes e  $x_i$  uma variável

$$\sum_{i=1}^n (a + bx_i) = \sum_{i=1}^n a + \sum_{i=1}^n bx_i = na + b \sum_{i=1}^n x_i$$

Seja  $X = \{6, 7, 9, 11, 5\}$ , obter:


$$\sum_{i=1}^5 x_i \Rightarrow \sum_{i=1}^5 x_i = 6 + 7 + 9 + 11 + 5 = 38$$

$$\sum_{i=1}^4 2x_i \Rightarrow \sum_{i=1}^4 x_i = 2 \cdot (6 + 7 + 9 + 11) = 66$$

$$\sum_{i=2}^5 3x_i \Rightarrow \sum_{i=2}^5 x_i = 3 \cdot (7 + 9 + 11 + 5) = 96$$

## Medidas de posição ou tendência central

Apesar das tabelas de frequências trazerem muita informação, as vezes queremos resumir ainda mais os dados em valores que representam toda a série. Algumas dessas medidas descrevem a *tendência central*, isto é, a tendência que os dados têm de se agrupar em torno de certos valores. Dentre as medidas de tendência central, destacamos: **Média, Mediana e Moda.**



Descreve resumidamente uma distribuição de frequência (centro de massa de um conjunto dados).

Observações:

- A média é afetada por valores extremos;
- Não utilizável em variáveis categóricas.
- A média é bastante utilizada em distribuições simétricas;

Notação:

$\mu$  é a média populacional.

$\bar{x}$  é chamada média amostral.

# Média

É a soma das observações dividida pelo número delas e ela pode ser obtida de diferentes formas, são algumas delas:

- Média Aritmética Simples;
- Média para dados em distribuição de frequências;
- Média para dados agrupados em classes;
- Média aritmética ponderada.

O cálculo da média é dado pela seguinte expressão:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

onde  $n$  é o tamanho da amostra observada e  $x_i$  é o valor genérico da observação.

**Exemplo:** Valor de compra do dólar americano nos últimos 7 dias:

3,54; 3,56; 3,54; 3,58; 3,61; 3,59; 3,59

$$\bar{x} = \frac{3,54 + 3,56 + 3,54 + 3,58 + 3,61 + 3,59 + 3,59}{7} = \frac{25,01}{7} = 3,5728$$

***Interpretação:***

O custo médio do dólar para compra é de R\$3,57.

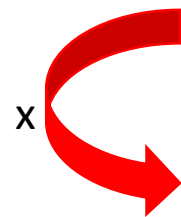
## Média para os dados em distribuição de frequência

O cálculo da média em uma distribuição de frequência é dado pela seguinte expressão:


$$\bar{x} = \sum_{i=1}^n \frac{x_i f_i}{n}$$

$n = \sum_{i=1}^k f_i$  onde  $k$  é o número de valores distintos da tabela.

**Exemplo:** Número de peças defeituosas de 25 máquinas de uma empresa.



Número de defeitos	1	2	3	4	5	6	7	8	Total
$f_i$	1	4	4	4	7	3	1	1	25



$$\bar{x} = \sum_{i=1}^8 \frac{x_i f_i}{n} = \bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + x_3 \cdot f_3 + x_4 \cdot f_4 + x_5 \cdot f_5 + x_6 \cdot f_6 + x_7 \cdot f_7 + x_8 \cdot f_8}{f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8}$$

$$\bar{x} = \frac{1 * 1 + 2 * 4 + 3 * 4 + 5 * 7 + 6 * 3 + 7 * 1 + 8 * 1}{25} = \frac{105}{25} = 4,02 \cong 4$$

O número médio de peças fabricadas com defeitos é de aproximadamente 4 por máquina.

## Média para dados agrupados em classes

O cálculo da média para dados agrupados em classes é dado pela seguinte expressão:

$$\bar{x} = \frac{\sum_{i=1}^n Pm_i f_i}{n}$$

$Pm_i$  é o ponto médio da classe  $i$ .

**Exemplo:** Frequência do tempo (em segundos) até um celular abrir um determinado aplicativo.

Classes	Pmi	fa
5,56 -6,91	6,235	1
6,91 -8,26	7,585	5
8,26 -9,61	8,935	7
9,61 -10,96	10,285	12
10,96 -12,31	11,635	4
12,31 -13,66	12,985	1
Total		30



$$\bar{x} = \frac{\sum_{i=1}^6 Pm_i f_i}{n}$$

$$\bar{x} = \frac{6,235 * 1 + 7,585 * 5 + 8,935 * 7 + 10,285 * 12 + 11,635 * 4 + 12,985 * 1}{30} = \frac{289,65}{30} = 9,655$$

O tempo médio de um celular abrir um determinado aplicativo é 9,655 segundos.

**Exercício 1.** (IBGE/2010 – Modificada) No último mês, Carlos fez 30 ligações de seu celular cujas durações em minutos, estão apresentadas abaixo

14	34	33	13	21
29	24	31	27	27
20	13	28	29	11
17	27	28	18	35
36	24	12	27	26
18	14	27	21	24

- Calcule a média para os dados brutos. Interprete o resultado;
- Disponha os dados em uma tabela de frequências e calcule a média. Houve diferença nos resultados? Justifique.

**Resultado**

$$a) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{30} \sum_{i=1}^{30} (14 + 34 + 33 + \dots + 21 + 24) = \frac{1}{30} * 708 = 23,6$$

b)

Classe	fi	Pmi	fi*Pmi
11 - 16	6	13,5	81
16 - 21	4	18,5	74
21 - 26	5	23,5	117,5
26 - 31	10	28,5	285
31 - 36	4	33,5	134
36 - 41	1	38,5	38,5
Total	30		730

Classe	fi	Pmi	fi*Pmi
11 - 16	0,2000	13,5	81
16 - 21	0,1333	18,5	74
21 - 26	0,1667	23,5	117,5
26 - 31	0,3333	28,5	285
31 - 36	0,1333	33,5	134
36 - 41	0,0333	38,5	38,5
Total			730

$$\frac{730}{30} = 24,33$$

# Mediana

É o termo do meio em um **conjunto ordenado**, ou seja, divide um conjunto ordenado em duas partes iguais. A mediana (**M<sub>d</sub>**) pode ser obtida de duas maneiras:

$$M_d = \begin{cases} X_{\frac{n+1}{2}} & \text{se } n \text{ for ímpar} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} & \text{se } n \text{ for par} \end{cases}$$

**Exemplo:** Considere os dados 6, 3, 7, 13, 9, 15, 10.

**Impar**

Primeiro é necessário ordenar os dados:

3, 6, 7, 9, 10, 13, 15.

Como se de uma conjunto com  $n = 7$  (ímpar), então:

$$Md = X_{\frac{n+1}{2}} = X_{\frac{7+1}{2}} = X_4$$

Logo a Mediana é igual ao elemento que está na quarta posição do conjunto de dados, assim  $Md = 9$

**Par**

**Exemplo:** Considere os dados 2, 4, 9, 7, 3, 5.

Primeiro é necessário ordenar os dados: 2, 3, 4, 5, 7, 9.

Como se de uma conjunto com  $n = 6$  (par), então:

$$Md = \frac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} = \frac{X_{\frac{6}{2}} + X_{\frac{6+2}{2}}}{2} = \frac{X_{\underline{3}} + X_{\underline{4}}}{2} = \frac{4 + 5}{2} = 4,5$$

Logo a Mediana é igual ao elemento que está na média das posições 3 e 4 do conjunto de dados, assim  $Md = 4,5$

Processo para determinara a mediana:

1. Ordenar os valores;
2. Se o número de dados é ímpar, a mediana é o valor que está no centro da série;
3. Se o número de dados é par, a mediana é a média dos dois valores que estão no centro da série.

## Mediana para dados em distribuição de frequências simples

**Tabela 1:** Número de peças defeituosas de 25 máquinas de uma empresa.

Número de defeitos	1	2	3	4	5	6	7	8	Total
$f_i$	1	4	4	4	7	3	1	1	25
$F_i$	1	5	9	13	20	23	24	25	25

$$n = 25 \quad Md = X_{\left(\frac{n+1}{2}\right)} = X_{\left(\frac{25+1}{2}\right)} = X_{13} = 4$$

Assim, a mediana será dada pelo valor na posição 13, 4 defeitos.

50% das máquinas produzem até 4 peças defeituosas.



## Determinação da mediana de dados agrupados em classes

$$Md = l + h \left( \frac{\frac{n}{2} - F_{ant}}{f_{Md}} \right)$$

onde,

$l$  = limite inferior da classe mediana;

$h$  = amplitude do intervalo da classe;

$n/2$  = posição do elemento mediano;

$F_{ant}$  = frequência acumulada até a classe anterior à classe mediana;

$f_{Md}$  = frequência absoluta simples da classe mediana

**Tabela 2:** Frequência do tempo (em segundos) até um celular abrir um determinado aplicativo.

Classes	Pmi	fi	Fi
5,56 -6,91	6,235	1	1
6,91 -8,26	7,585	5	6
8,26 -9,61	8,935	7	13
9,61 -10,96	10,285	12	25
10,96 -12,31	11,635	4	29
12,31 -13,66	12,985	1	30
Total		30	

$n=30$ , logo a mediana esta entre a média dos valores da posição 15 e 16.

$$Md = l + h \left( \frac{\frac{n}{2} - F_{ant}}{f_{Md}} \right) = 9,61 + 1,35 \left( \frac{\frac{30}{2} - 13}{12} \right) = 9,835$$

# Moda

É o valor que possui maior frequência num conjunto. **Sua vantagem é que pode ser usada para variáveis qualitativas.**

Observações:

- Um conjunto de dados pode apresentar mais de uma moda;
- A moda pode ser calculada para variáveis qualitativas e quantitativas;
- Um conjunto de dados sem moda é chamado Amodal.

Considerando um conjunto ordenado de valores, a moda será o valor predominante, o valor mais frequente desse conjunto. Embora seu significado seja o mais simples possível, nem sempre a moda existe (distribuição amodal) e nem sempre é única. Se apresentar apenas uma moda diremos que é unimodal; se possuir duas modas diremos que é bimodal; se tiver várias modas (mais que duas) diremos que é multimodal.

Exemplos:

$X = \{10,11,12,13,14\}$ : conjunto Amodal;

$X = \{10,11,12,12,13\}$ : conjunto Unimodal, moda = 12;

$X = \{10,11,12,12,13,14,14\}$ : conjunto Bimodal; moda = 12 e 14.

## Moda de Valores Tabulados

**Tabela 3:** Opinião dos clientes entrevistados quanto ao serviço recebido pela empresa contábil

Opinião	Frequência absoluta (fa)
Ruim	6
Regular	4
Bom	7
Ótimo	3
Total	20

Observando os resultados da Tabela 3, conclui-se que a opinião com maior frequência é a resposta “bom”, logo essa será a moda.

Para dados agrupados a moda se localiza na classe de maior frequência (classe modal) e é obtida por meio da expressão (Moda de Czuber):

$$Mo = l + h \left( \frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$$

$l$  é o limite inferior da classe modal;

$h$  é a amplitude da classe modal;

$\Delta_1$  é a diferença da frequência da classe modal e a frequência da classe imediatamente anterior;

$\Delta_2$  é a diferença da frequência da classe modal e a frequência da classe imediatamente posterior.

**Tabela 4:** Frequência do tempo (em segundos) até um celular abrir um determinado aplicativo.

Classes	Pmi	fi	Fi
5,56 -6,91	6,235	1	1
6,91 -8,26	7,585	5	6
8,26 -9,61	8,935	7	13
9,61 -10,96	10,285	12	25
10,96 -12,31	11,635	4	29
12,31 -13,66	12,985	1	30
Total		30	

$$Mo = l + h \left( \frac{\Delta_1}{\Delta_1 + \Delta_2} \right) = 9,61 + 1,35 \left( \frac{12 - 7}{12 - 7 + 12 - 4} \right) = 10,1292$$

50% dos celulares demoram até 10,12 segundos para abrir o app.

**Exercício 2.** (IBGE/2010 – Modificada) No último mês, Carlos fez 30 ligações de seu celular cujas durações em minutos, estão apresentadas abaixo

14	34	33	13	21
29	24	31	27	27
20	13	28	29	11
17	27	28	18	35
36	24	12	27	26
18	14	27	21	24

- Calcule a mediana para os dados brutos. Interprete o resultado;
- Calcule a moda para os dados brutos. Interprete o resultado.



## Comparação: Média x Mediana x Moda

**Média:** Soma de todos os valores divididos pelo total de elementos do conjunto;

**Vantagens:** Centro de massa da distribuição;

**Limitações:** É influenciada por valores extremos;

Quando usar:

- Deseja-se obter a medida de posição que possui a maior estabilidade;
- Houver necessidade de um tratamento algébrico posterior.

## Comparação: Média x Mediana x Moda

**Mediana:** Valor que divide o conjunto em duas partes iguais;

**Vantagens:** Menos sensível a valores extremos que a média;

**Limitações:** Difícil de determinar para grande quantidade de dados.

Quando usar:

1. Deseja-se obter o ponto que divide o conjunto em partes iguais;
2. Há valores extremos que afetam de maneira acentuada a média

## Comparação: Média x Mediana x Moda

**Moda:** Definição: Valor mais frequente;

**Vantagens:** Valor “típico”; Maior quantidade de valores concentrados neste ponto;

**Limitações:** Pode não haver moda para certos conjuntos de dados.

Quando usar:

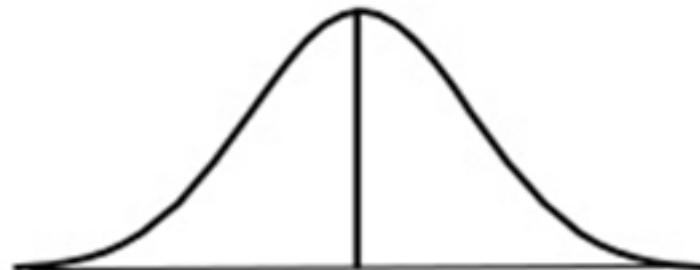
- Deseja-se obter uma medida rápida e aproximada da posição;
- A medida de posição deve ser o valor mais típico da distribuição.

# Simetria

A determinação das medidas de posição permite discutir sobre a simetria da distribuição dos dados.

Distribuição simétrica  $\rightarrow \bar{x} = Md = Mo$

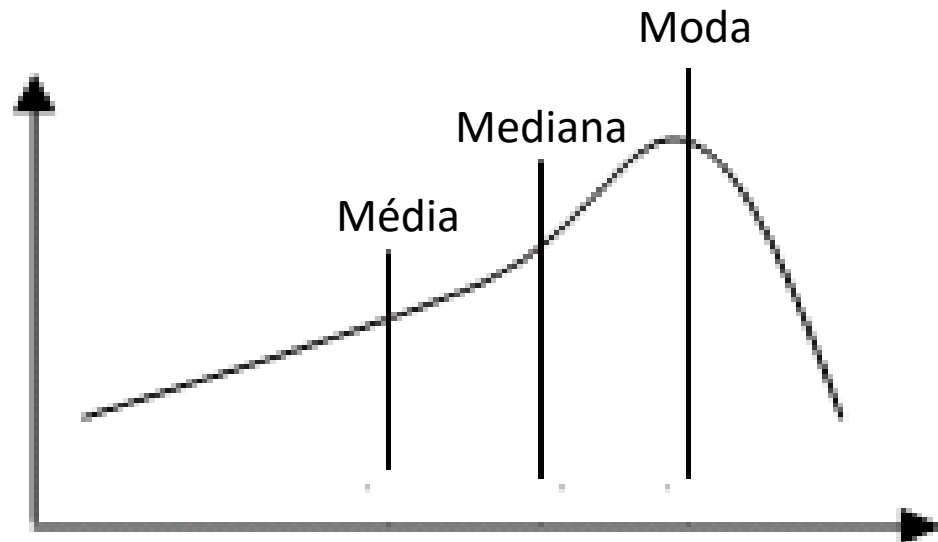
**Média = Mediana = Moda**



# Simetria

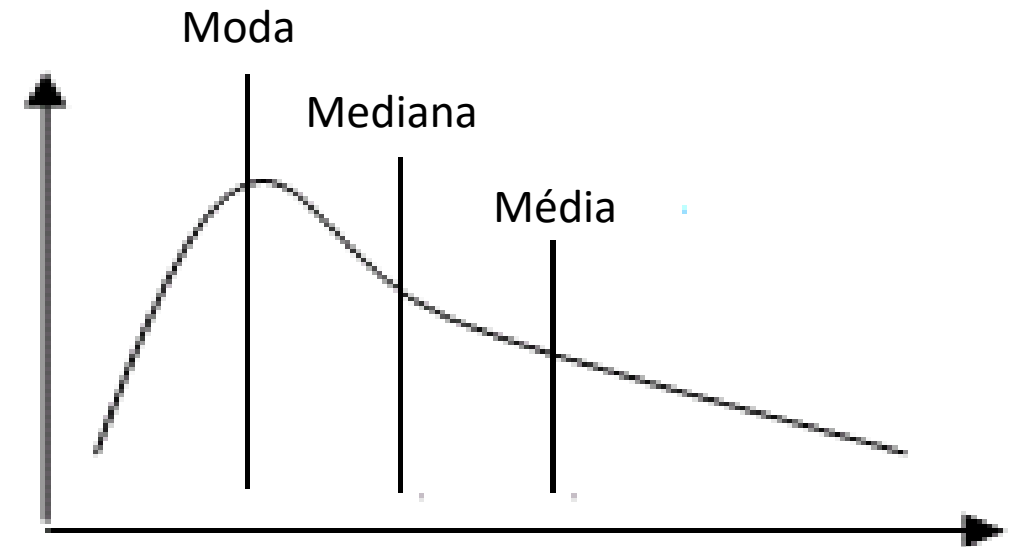
Distribuição assimétrica - ocorrem diferenças entre os valores da média, mediana e moda. A assimetria pode ser:

**Assimétrica à esquerda ou Negativa**



À esquerda  $\rightarrow \bar{x} \leq Md \leq Mo$

**Assimetria à direita ou Positiva**



À direita  $\rightarrow Mo \leq Md \leq \bar{x}$

# Coeficiente de assimetria baseados nas medidas de tendência central

O Coeficiente de Assimetria de Pearson ( $A_p$ ) baseia-se na posição relativa das medidas de tendência central de acordo com o tipo de assimetria dos dados, é definido como:

$$A_p = \frac{\bar{x} - M_o}{S}$$

Assim temos que:

Distribuição simétrica  $\rightarrow \bar{x} = Md = M_o \rightarrow$  se  $A_p = 0$

Distribuição assimétrica positiva  $\rightarrow \bar{x} > Md > M_o \rightarrow$  se  $A_p > 0$

Distribuição assimétrica negativa  $\rightarrow \bar{x} < Md < M_o \rightarrow$  se  $A_p < 0$

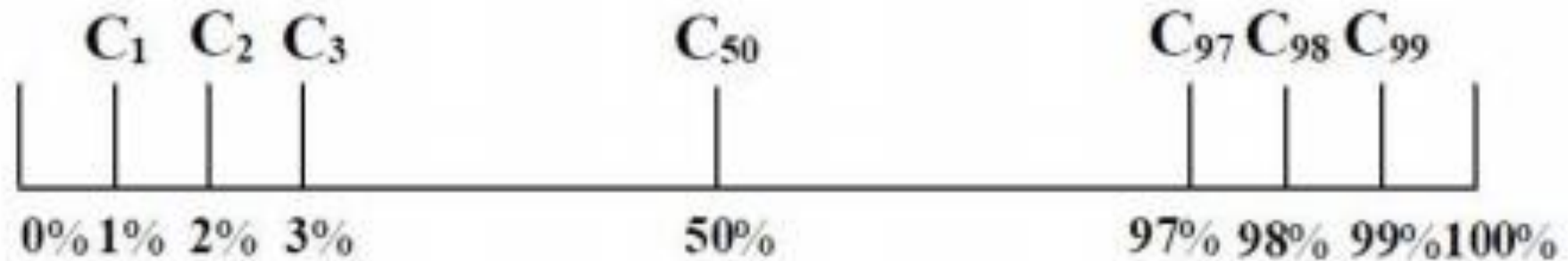
## Separatrizes

Além das medidas de posição que estudamos, há outras que, consideradas individualmente, não são medidas de tendência central, mas possuem característica de separar a série em partes que apresentam o mesmo número de valores, ou seja, grupos onde cada um irá conter o mesmo percentual dos valores da sequência. São elas:

- Percentis - Dividem a amostra em 100 partes iguais;
- Decis - Dividem a amostra em 10 partes iguais;
- Quartis - Dividem a amostra em 4 partes iguais;

# Percentis

São as medidas que dividem a amostra em 100 partes iguais. Assim:





Calculando o percentil:

1. Ordene os dados (Rol)

2. Calcule  $i$ , onde  $i = \left(\frac{p}{100}\right)n$ ,

onde  $p$  é o percentil de interesse e  $n$  é o número de observações.

3. A) Se  $i$  não for inteiro, arredonde para cima. O próximo inteiro maior que  $i$  denota a posição do  $p$ -ésimo percentil.

B) Se  $i$  é um inteiro, o  $p$ -ésimo percentil é a média dos valores das posições  $i$  e  $i+1$ .

**Exemplo:** Salário mensal inicial para uma amostra de 12 graduados recém formados. Determinar o 85º percentil.

2350	2450	2550	2380
2255	2210	2390	2630
2440	2825	2420	2380

O primeiro passo é ordenar os dados (Rol): 2210, 2255, 2350, 2380, 2380, 2390, 2420, 2440, 2450, 2550, 2630, 2825.

$$i = \left(\frac{p}{100}\right)n = \left(\frac{85}{100}\right)12 = 10,2$$

Como  $i$  não é inteiro, arredonde para cima, obtendo a posição 11, percentil 2630.

**Exercício 3.** Calcule o 50º percentil.

2350	2450	2550	2380
2255	2210	2390	2630
2440	2825	2420	2380

# Quartis

1º quartil (Q1): Deixa 25% dos elementos antes do seu valor;

2º quartil (Q2): Deixa 50% dos elementos antes do seu valor. Coincide com a mediana;

3º quartil (Q3): Deixa 75% dos elementos antes do seu valor. (Consequentemente, 25% dos elementos acima do seu valor).



# Medidas de dispersão

Para avaliar o grau de variabilidade ou dispersão dos valores de um conjunto de números utilizamos as chamadas medidas de dispersão. Essas nos proporcionarão um conhecimento mais completo do fenômeno a ser analisado, permitindo estabelecer comparações entre fenômenos da mesma natureza e mostrando até que ponto os valores se distribuem acima ou abaixo da medida de tendência central.

Observe as sequências:

X : 13; 13; 13; 13; 13; 13; 13; 13; 13; 13:

Y : 12; 13; 13; 14; 12; 14; 12; 14; 13; 13:

Z : 10; 1; 18; 20; 35; 3; 7; 15; 11; 10:

- Na sequência X não há variabilidade dos dados.
- Na sequência Y, a média 13 representa bem a série, mas existem elementos da série levemente diferenciados da média= 13.
- Na sequência Z os elementos estão bem diferenciados da média= 13

## Tipos de medidas de dispersão:

- Amplitude Total;
- Desvio-Padrão ( $S$ );
- Variância ( $S^2$ );
- Coeficiente de variação (CV).

# Amplitude Total

Representa a diferença entre o maior e o menor valor de um conjunto de dados. Ela mostra a dispersão dos valores de uma série. Se a amplitude for um número elevado, então os valores da série estão distribuídos afastados; se a amplitude for um número baixo, então, os valores na série estão próximos uns dos outros. Esta medida será simbolizada por “H” e é calculada da seguinte maneira:

$$H = l_s - l_i$$



# Variância

É a medida de dispersão mais usada e mais importante. Mede a concentração dos dados em torno da média. É dada pela soma dos quadrados dos desvios divididos pelo número total de observações.

A notação  $S^2$  e é usada para representar a variância amostral.

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

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1}$$

# Desvio-padrão

É a raiz quadrada da variância. Deixa a medida de variabilidade na **mesma unidade** de medida dos dados, diferente da variância.

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

## Coeficiente de variação (CV)

É uma medida útil para **comparações**, do grau de concentração em torno da média, de séries distintas.

$$CV = \frac{\sigma}{\mu} 100 \text{ ou } CV = \frac{S}{\bar{x}} 100$$

Algumas observações sobre o CV:

- Uma média muito próxima de zero pode inflacionar o CV;
- Coeficiente de variação superior a 50% sugere alta dispersão;
- Quanto maior for este valor, menos representativa será a média.

**Exemplo:** Os dados seguintes representam altura (cm) de sete atletas do curso de Engenharia Elétrica - UFMT:

169 171 173 175 175 177 179 181

Calcule as medidas de dispersão:

- a) Amplitude Total;
- b) Variância ( $S^2$ );
- c) Desvio-Padrão ( $S$ );
- d) Coeficiente de variação (CV)

**Resultado**

$$H = ls - li \Rightarrow H = 181 - 169 = \underline{12}$$

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1} = \frac{112}{8 - 1} = 16$$

$$S = \sqrt{16} = 4$$

$$CV = \frac{S}{\bar{x}} = \frac{4}{175} = 0,0229 * 100\% \Rightarrow 2,29\%$$

$X_i$
169
171
173
175
175
177
179
181
TOTAL

**Exemplo:** Tomemos os resultados das medidas das estaturas e dos pesos de um mesmo grupo de indivíduos:

	$\bar{X}$	S
Estatura	175 cm	4,0 cm
Peso	68 kg	2,0 kg

Tem-se:

$$CV_E = \frac{4}{175} 100\% = 2,29\% \quad CV_P = \frac{2}{68} 100\% = 2,94\%$$

Os pesos apresentam maior grau de dispersão que as estaturas.

**Exercício 3:** Os dados seguintes representam 20 observações relativas ao índice pluviométrico em determinados municípios do Estado:

144 152 159 160 160 151 157 146 154 145  
150 142 146 142 141 141 150 143 158 141

- Construir a tabela de frequências (absoluta e acumulada) e os pontos médios;
- Calcular as medidas de posição (Média, Mediana e Moda) para os dados brutos e para a tabela
- Calcular as medidas de variação (Amplitude, Variância e Desvio-padrão) para os dados brutos e para a tabela