



# Estatística

# Aula 2

Objetivo da aula:

- Conceitos Básicos;
- Medidas de dispersão e Medidas de posição.

## Medidas de Posição:

As medidas de posição são o valor médio, a mediana e a moda.

Estes indicadores estatísticos são de extrema importância para a análise de dados, tanto os qualitativos como quantitativos.

É de extrema importância para o analista saber quando usar estes indicadores estatísticos a partir dos dados coletados.

Um erro de interpretação no uso de um indicador deste colocará toda a análise estatística em descrédito.

# Exemplificando a importância do uso correto de um indicador estatístico:

Exemplo: Uma pesquisa foi realizada em 20 apartamentos de um prédio. Em um item da pesquisa foi perguntado ao morador quantos aparelhos smartphones haviam na sua residência. A tabela abaixo mostra os dados da pesquisa:

101	102	103	104	201	202	203	204	301	302	303	304	401	402	403	404	501	502	503	504
2	3	3	2	4	4	3	3	2	3	3	3	2	105	3	2	4	3	2	4

## Continuação da exemplificação:

A partir dos dados coletados, o analista toma a decisão de calcular a média aritmética dos dados. Este então apresenta o seguinte resultado:

Na média, cada apartamento do prédio tem 8 aparelhos smartphones.

Ai está o problema na metodologia adotada, pois não foi observado que o morador do ap. 402 apresenta um dado muito discrepante dos outros moradores.

## Continuação da exemplificação:

Provavelmente, o morador do ap. 402 é um técnico que conserta smartphones no seu ap. ou ele pode ser um colecionador de smartphones.

Então, apresentar o valor médio como indicador estatístico pode não ser o mais indicado como neste caso.

# Medidas de Posição: Valor Médio.

O valor médio ( $\mu$ ) é calculado como a média aritmética de todos os  $n$  valores ( $x_i$ ).

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

Por exemplo: O valor médio para o conjunto de idade de 5 pessoas cujos valores são: 22, 43, 35, 30, 20 é 30.

# Medidas de Posição: Mediana.

A mediana (*Med.*) de um conjunto de valores (colocados em Rol) é o valor situado no centro do conjunto de dados.

O conjunto pode ter um valor par ou ímpar de dados, então a determinação da mediana deverá seguir uma regra.

A mediana é muito bem utilizada quando temos um conjunto de dados com muitos '*outliers*' que são dados extremamente discrepantes no conjunto.



# Medidas de Posição: Mediana.

Conjunto de dados:  
n ímpar.

Neste caso o valor central do conjunto (em Rol) é a mediana.

Exemplo: A mediana da série: 1, 1, 2, 2, 3, 3, 4 é 2.  
(1, 1, 2), 2, (3, 3, 4)

# Medidas de Posição: Mediana.

Conjunto de dados:  
n par.

Neste caso o valor central do conjunto (em Rol) é dado pela média aritmética dos dois valores centrais.

Exemplo: A mediana da série: 2, 10, 12, 20, 22, 31, 34, 40 é 21.  
(2, 10, 12), 20, 22, (31, 34, 40)  
ou nominais.

## Medidas de Posição: Moda.

A moda (Mod.) de uma série de valores é o valor de maior frequência absoluta, ou seja, o valor que aparece o maior número de vezes no conjunto.

Exemplo: A moda da série de dados: 1, 22, 33, 41, 33, 23 é 33.

A moda é utilizada para preencher dados tipo '*missing values*' em variáveis categóricas. Por exemplo se numa tabela estiver faltando dados na coluna sexo onde lê-se M=0 e F=1 opta-se pela moda para preencher os dados faltantes.

# Medidas de Dispersão: Variância.

A variância e o desvio padrão são indicadores de precisão no conjunto de dados.

A variância ( $\sigma^2$ ) é dada pela fórmula:

$$var: \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 / (n - 1).$$

Observação: A variância é um indicador estatístico da precisão dos dados em torno do valor médio.

# Medidas de Dispersão: Desvio Padrão.

O desvio padrão ( $\sigma$ ) é dado pela fórmula:

$$dp: \sigma = \sqrt{\sum_{i=1}^n (x_i - \mu)^2 / (n - 1)}.$$

Ou seja, o desvio padrão é a raiz quadrada da variância.

O desvio padrão é o mais utilizado que a variância, pois o desvio padrão está na mesma unidade de medida do valor médio.

## Medidas de Dispersão: Variância e Desvio Padrão.

Exemplo: Determinar o valor médio, a variância e o desvio padrão para a série de 5 medidas de idade a seguir:

17	18	21	25	19
----	----	----	----	----

Os resultados são:

Valor médio: 20 anos

Variância: 10 anos<sup>2</sup>

Desvio Padrão: 3,16 anos

# Memória de Cálculo:

Valor médio=20, Variância=10.

n	$x_i$	$(x_i - \mu)$
1	17	$17 - 20 = -3$
2	18	$18 - 20 = -2$
3	21	$21 - 20 = 1$
4	25	$25 - 20 = 5$
5	19	$19 - 20 = -1$
	$\mu = 20$	

## Exercício de Revisão:

Determine o valor médio, moda, mediana para o conjunto de dados de idades em anos:

a) 3, 2, 4, 4, 5, 7, 6

b) 3, 2, 4, 5, 4, 6, 7, 8



## Exercício de Revisão:

O resultado de uma medida de idade é  $I(\text{anos})=12\pm 2$ .  
Qual o valor médio, desvio padrão e variância?