

# Princípios básicos da Análise Exploratória de Dados UFCD: 10807

CARGA HORÁRIA: 25 horas

ALEXANDRA CAMPOS VIDAL DE SOUZA  
FORMADORA



# Outras medidas de Posição

**Partindo da intuição da Mediana...**

**Uma vez que ordenamos os dados, por que não dividi-los em MAIS de 2 partes?**

Uma vez que nós ordenamos os dados...

...isso dá origem a outras medidas de posição:

- **Quartis:** Dividimos os dados em 4 quantidades iguais.
- **Quintis:** Dividimos os dados em 5 quantidades iguais.
- **Decis:** Dividimos os dados em 10 quantidades iguais.
- **Vintis:** Dividimos os dados em 20 quantidades iguais.
- **Percentis:** Dividimos os dados em 100 quantidades iguais.



# Aplicação de Medidas de Dispersão a partir da Média

## Quanto erramos ao usar a Média?

No caso das idades calculamos a média (além da mediana e dos quartis), mas o quanto estamos errando ao estimar a idade de todos os alunos com a média?

Para calcular o erro, precisamos de alguns passos



# Aplicação de Medidas de Dispersão a partir da Média

**Primeira coisa a fazer é calcular o erro**

Idade – Média

Para calcular o erro, precisamos pegar a idade e diminuir da media.





# Aplicação de Medidas de Dispersão a partir da Média

As medidas de dispersão são estatísticas que quantificam a dispersão, a variabilidade ou a dispersão dos valores em um conjunto de dados. Elas ajudam a entender o quão dispersos estão os valores em torno da medida central (como a média). As duas medidas de dispersão mais comuns são a **variância e o desvio padrão**.



# Variância

A variância é uma medida que indica o quanto os valores em um conjunto de dados variam em torno da média. Uma variância maior indica uma maior dispersão dos valores, enquanto uma variância menor sugere que os valores estão mais próximos da média. A variância é calculada como a média dos quadrados das diferenças entre cada valor e a média do conjunto de dados.



- **Exemplo Prático**
- Suponha que temos os seguintes preços de um bem ao longo de 5 meses:
- {100,110,105,115,120}
- **Calcular a média:**

$$\bar{x} = \frac{100 + 110 + 105 + 115 + 120}{5} = 110$$





- **Exemplo Prático**
- Suponha que temos os seguintes preços de um bem ao longo de 5 meses:
- {100,110,105,115,120}

**1. Calcular a média:**

$$\bar{x} = \frac{100 + 110 + 105 + 115 + 120}{5} = 110$$



## 2. Calcular a diferença em relação à média e elevar ao quadrado:

Valor ( $x_i$ )	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
100	$100 - 110 = -10$	100
110	$110 - 110 = 0$	0
105	$105 - 110 = -5$	25
115	$115 - 110 = 5$	25
120	$120 - 110 = 10$	100

## 3. Somar os valores das diferenças ao quadrado:

$$100 + 0 + 25 + 25 + 100 = 250$$



- **Calcular a variância:**
- Se for **populacional:**

$$\sigma^2 = \frac{250}{5} = 50$$

- Se for **amostral:**

$$s^2 = \frac{250}{5 - 1} = \frac{250}{4} = 62.5$$



# Desvio padrão

O desvio padrão **é a raiz quadrada da variância** e também mede a dispersão dos valores em um conjunto de dados em relação a média.

Ele é expresso na mesma unidade de medida dos valores originais, o que o torna mais fácil de interpretar em comparação com a variância. Um desvio padrão maior indica maior variabilidade dos valores, enquanto um desvio padrão menor sugere que os valores estão mais próximos da média.



# Desvio padrão

A variância e o desvio padrão são medidas de dispersão que ajudam a **quantificar a dispersão** ou a variabilidade dos valores em um **conjunto de dados em torno da média**. Essas medidas são úteis para entender a consistência dos dados e para comparar a dispersão entre diferentes conjuntos de dados ou variáveis.



# Coeficiente de Variação

O coeficiente de variação (CV) é uma forma de medir quão espalhados estão os dados em relação à média. Ele compara o desvio padrão (uma medida de dispersão) com a média e mostra essa relação em percentagem.

Ele ajuda a comparar a variabilidade de diferentes grupos de dados, mesmo que tenham unidades ou tamanhos diferentes.

Fórmula para calcular o coeficiente de variação

$$\text{(CV): CV} = (\text{Desvio Padrão} / \text{Média}) \times 100$$

$$\text{CVPop} = (7,07 / 50) \times 100 = 14,14\%$$

$$\text{CVamost} = (7,91 / 50) \times 100 = 15,82\%$$



# Coeficiente de Variação

Imagine que temos dois vendedores:

**O Vendedor A** vende uma média de 100 produtos por mês, mas às vezes vende 90 ou 110 (pouca variação).

**O Vendedor B** vende uma média de 10 produtos por mês, mas pode vender 5 ou 15 (muita variação).

Em resumo CV Mostra a dispersão em termos relativos (percentagem)



# Exemplo Prático:

Medidas de Posição  
de Erros de Modelos  
Considerando...

- As medidas de posição estudadas e mais as medidas de dispersão estudadas...
- O conjunto de dados reais e 2 modelos que estimam este dado real:





# Obrigado!

