

# CIÊNCIA DE DADOS (BIG DATA)

## ANÁLISE ESTATÍSTICA

**Professor curador:** Mário Olímpio de Menezes



Mackenzie



# TRILHA 1

## PARTE B – INTRODUÇÃO À ANÁLISE ESTATÍSTICA

# PARTE B – INTRODUÇÃO À ANÁLISE ESTATÍSTICA

# MÉTODOS NUMÉRICOS PARA DESCRIÇÃO DE DADOS

Utilizamos algumas medidas para descrever um conjunto de dados, para sumarizar a variabilidade numericamente:

- Medidas que descrevem o centro dos dados.
- Medidas que descrevem o espalhamento dos dados.
- Medidas que indicam posições (áreas) dos dados.
- Medidas de Assimetria e Curtose.

# MEDIDAS DE POSIÇÃO CENTRAL

É representativa do conjunto das observações.

Descreve, aproximadamente, onde os dados estão centralizados.

- Média (Aritmética, Ponderada, Harmônica e Geométrica).
- Mediana.
- Moda.

# MEDIDAS DE DISPERSÃO

Busca quantificar o quanto os valores do conjunto de dados estão afastados ou dispersos em relação a uma medida de posição central (média, mediana etc.).

- Amplitude total.
- Desvio padrão.
- Variância.
- Amplitude interquartílica.

# MEDIDAS DE POSIÇÃO

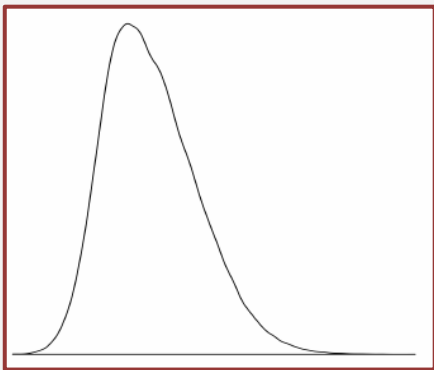
Medidas que dividem a área de uma distribuição (valores do conjunto de dados) em regiões de áreas (quantidades) iguais.

- Quartil – pedaços de 25% cada.
- Percentil – pedaços de 1% cada.

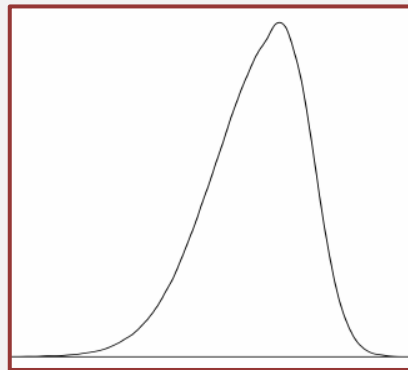
# ASSIMETRIA (SKEWNESS)

Pode ser negativa, positiva ou indefinida.

POSITIVA



NEGATIVA





# CURTOSE

Está relacionada ao grau de achatamento da distribuição.

- Mesocúrtica – pouco achatamento.
- Leptocúrtica – pico mais pronunciado.
- Platocúrtica – pico mais achatado.

# COLETA DE DADOS

# COLETA DE DADOS

- A Coleta de Dados é uma etapa crítica de um Estudo Estatístico.
- A qualidade dos dados disponíveis e, portanto, a qualidade dos resultados das análises dependem de como foi realizada a coleta dos dados.
- Planejamento é essencial para uma boa coleta de dados.

# AMOSTRAGEM

- Nem sempre é possível se estudar uma população inteira.
- Muitos estudos são realizados com uma porção da população, chamada de **Amostra**.
- Para podermos generalizar os resultados da amostra para a população, é importante que a amostra seja **representativa** da população.

# VIÉS NA AMOSTRAGEM

## VIÉS DE SELEÇÃO

- Exclusão sistemática de alguma parte da população.

## VIÉS DE MEDIDA OU RESPOSTA

- Método de observação com problema.

## VIÉS DE NÃO-RESPOSTA

- Alguns indivíduos são omitidos, mesmo tendo sido selecionados.

# MÉTODOS DE AMOSTRAGEM

## AMOSTRAGEM ALEATÓRIA

- Simples
  - Sem substituição
  - Com substituição

## AMOSTRAGEM ALEATÓRIA

- Extratificada
- Por agrupamento
- Sistemática

## R e RStudio



é um software estatístico aberto, grátis, desenvolvido e mantido por uma comunidade de entusiastas.

12k+ pacotes



Studio é uma IDE para trabalhar com o R. Ambiente de produtividade, com integração de ferramentas.

