

# Estatística Descritiva

## Conceitos Básicos

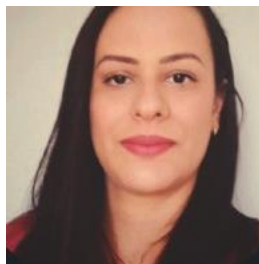
Big Data Science



# Semantix<sup>®</sup>

All about data

# Quem sou eu?



## Talita Costa Dias

Bacharelado e Mestrado em Estatística  
Cientista de dados da Semantix  
Instrutora do Semantix Academy

### Contatos

[talita.costa@semantix.com.br](mailto:talita.costa@semantix.com.br)  
[linkedin.com/in/tcostadias](https://linkedin.com/in/tcostadias)

# Ementa

- Conceitos Básicos
- Medidas Resumo
- Medidas Resumo (Aplicação)
- Análise Gráfica
- Revisão
- Correção de Exercício

# Ementa

- Neste curso, você aprenderá:
- Quais **quantidades numéricas** você pode usar para descrever e resumir seus conjuntos de dados.
- Como obter **estatísticas descritivas** com bibliotecas **Python**.
- Como **visualizar** seus conjuntos de dados.

# Hipóteses

- Dois tratamentos, experimental e controle, tem a mesma eficácia no tratamento do câncer?
- A obesidade é um fator de risco na morte por COVID-19?
- O número de vacinados teve aumento significativo de um mês para o outro?

# Objetivos da Análise de Dados

- Entender o que aconteceu e o que está acontecendo.
- Prever o que vai acontecer no futuro ou circunstâncias que ainda não conhecemos.
- Guiar a tomada de decisão.

# Estatística Descritiva

- Conjunto de técnicas
  - **Objetivo**
    - Descrever, resumir e representar os dados visualmente.
    - Gerar insights sobre o fenômeno em análise.



Fonte: //pixabay.com

**Como escolher as medidas descritivas ou gráficos mais adequado?**

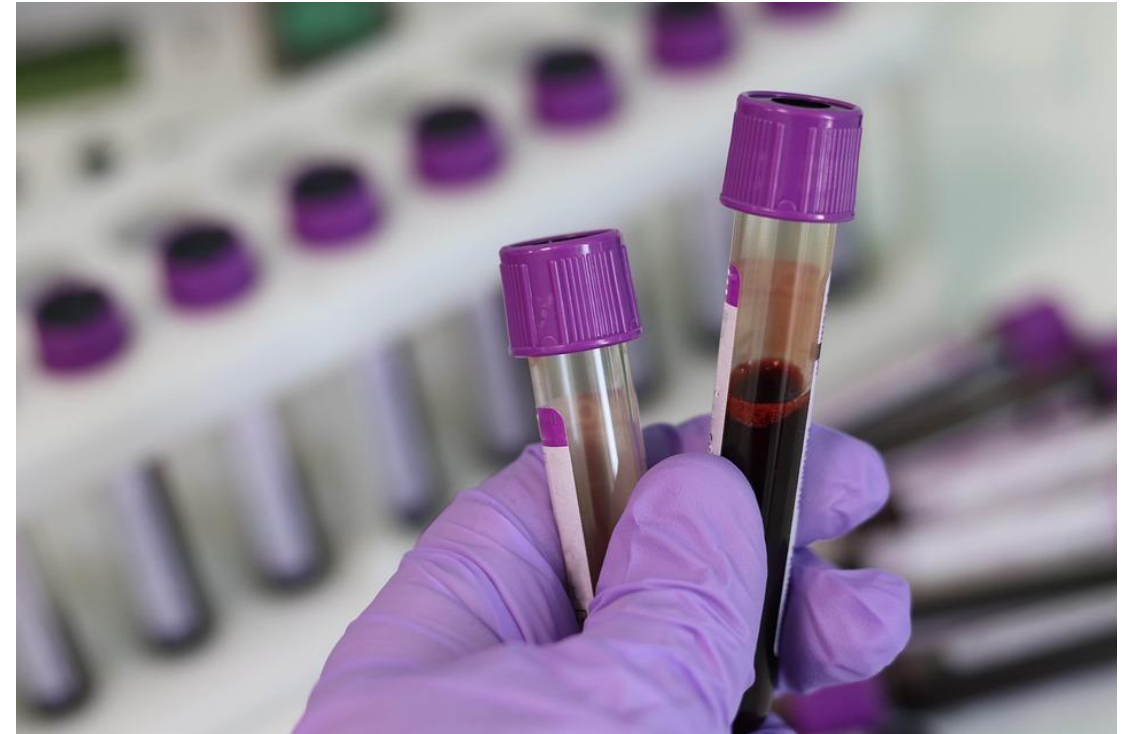
# Conceitos Básicos

## ○ População

- Conjunto de pessoas, objetos, transações... Que possuem características que estamos interessados em estudar.

## ○ Amostra

- Subconjunto representativo de uma população.



Fonte: //pixabay.com



# Conceitos Básicos

## ○ Unidade Amostral

- Objeto sobre o qual coletamos informações.

## ○ Variável

- Característica de uma unidade amostral.

## ○ Outliers ou Valores Atípicos

- Ponto de dados que difere significativamente da maioria.
  - **Variação natural** nos dados
  - **Mudança** no comportamento do sistema observado
  - **Erros** na coleta de dados

# Natureza das Variáveis

- **Qualitativa**

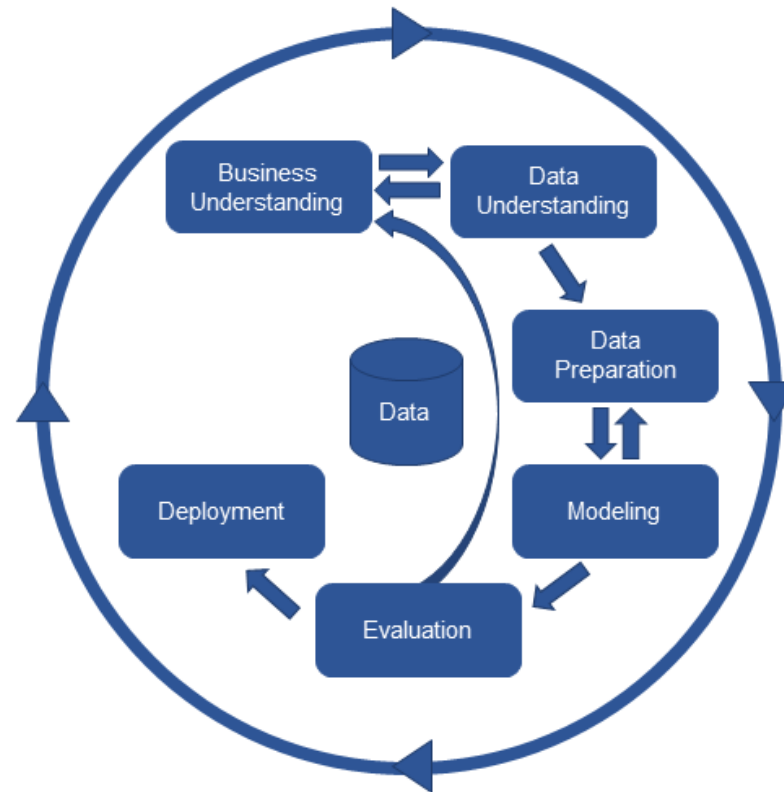
- Descreve qualidades.
  - **Nominal** – Sem ordenação.
    - Região geográfica
    - Estado civil
  - **Ordinal** – Com ordenação.
    - Grau de instrução
    - Estágio de uma doença

# Natureza das Variáveis

## ○ Quantitativa

- Descreve quantidades.
  - **Discreta – Processos de contagem** - Valores finitos em um intervalo.
    - # Sementes germinadas/área
    - # Pacientes que se recuperam/semana
  - **Contínua – Processos de medição** - Valores infinitos em um intervalo.
    - Rendimento/mês de investimento
    - Consumo/mês de energia

# CRISP - DM (Cross Industry Standard Process for Data Mining)



Fonte: [//medium.com/matgonz/crisp-dm-na-prática](https://medium.com/matgonz/crisp-dm-na-pr%C3%A1tica)



# Semantix<sup>®</sup>

All about data

[talita.costa@semantix.com.br](mailto:talita.costa@semantix.com.br)

[www.semantix.com.br](http://www.semantix.com.br)