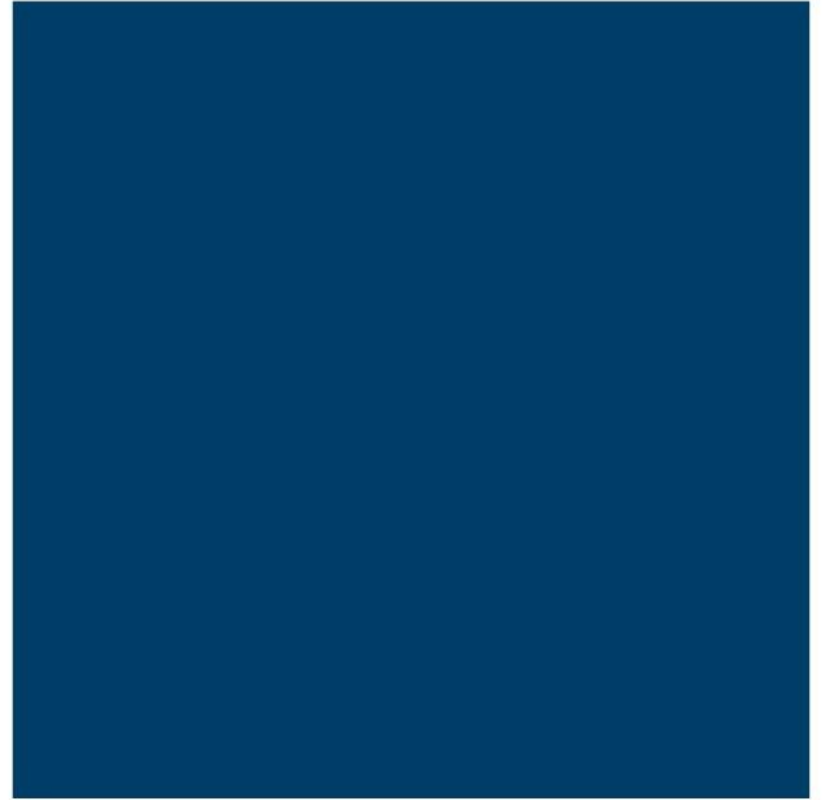


Análise, construção e visualização de dados

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

Vamos construir um conjunto de dados



<https://forms.gle/rn1eZGAatVk6LNpk6>

Análise exploratória

Etapas da estatística descritiva.



Análise descritiva dos dados

Uma descrição óbvia de um conjunto de dados está nos próprios dados.

- Número de observações
- Valores mínimos e máximos
- Ordenação dos valores

Análise descritiva dos dados

Quando temos muitos dados para analisar cada observação, precisamos de novas ferramentas.

- Tendências centrais
- Dispersão
- Correlação

Análise descritiva dos dados

Tendências centrais

- Média
- Mediana

Análise descritiva dos dados

Dispersão

- Amplitude
- Variância
- Desvio-padrão

Análise descritiva dos dados

Correlação

- Covariância
- Correlação

A correlação varia entre -1 (anticorrelação perfeita) e 1 (correlação perfeita)

Pandas e DataFrames

Pandas e Dataframes

Pandas vem de 'Panel Data' (dados em painel um termo da econometria para dados multidimensionais estruturados).

- É adequado para trabalhar com dados não numéricos e numéricos
- Adequado para lidar com dados proveniente de webscrapping
- Lida bem com dados incompletos

Pandas e Dataframes

Vamos usar o Pandas para acessar os nossos dados.

```
import pandas as pd  
file = "meu_arquivo.csv"  
df = pd.read_csv(file)  
print(df)
```

Análise exploratória

Obtendo um sumário descritivo estatístico de um DataFrame inteiro e de colunas selecionadas.

Usando a função `describe()`

É possível utilizar as funções de estatística descritiva diretamente no DataFrame

.

Análise exploratória

~ O Pandas possui funções para facilitar a descrição de dados.

~ `df.describe()`

~ `df.plot()`

Indexação de DataFrames

O DataFrames é organizado em colunas que podem ser acessadas a partir de seu nome.

Para determinar os nomes das colunas, pode-se escrever na tela os primeiros valores com a função `head()`

```
df.head()
```

Indexação de DataFrames

Para selecionar uma coluna de um DataFrame, a sintaxe é a mesma de um dicionário, com o nome da coluna no lugar da chave:

```
df["Name"]
```

Para selecionar elementos, use a indexação de listas:

```
df["Name"][0]
```

Indexação de DataFrames

Slices ou fatias são seleções de mais de um elemento de um objeto ao mesmo tempo.

A faixa de índices é indicada entre o sinal “:”.

Ou seja, para selecionar os dez primeiros elementos, usa-se:

```
df["Name"][0:10]
```




INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA