

## Aula Prática 02 Estatística Descritiva

**Objetivo:** avaliar dados numéricos com base em cálculos estatísticos

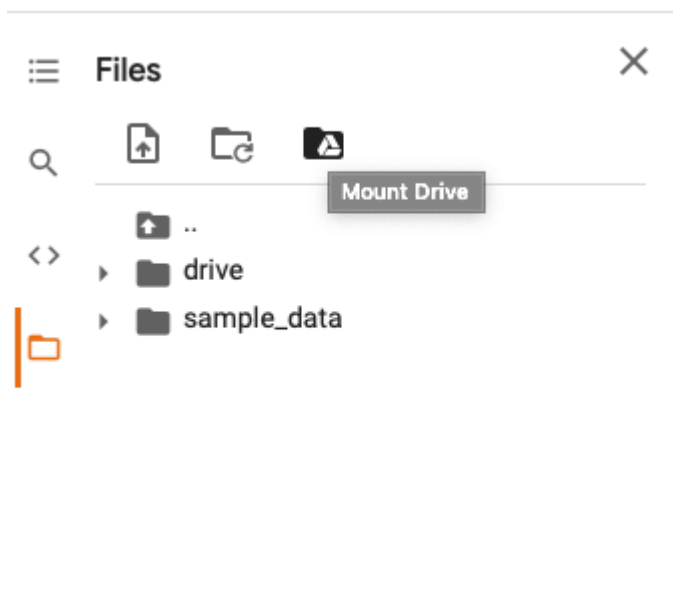
**Pré-requisitos:** linguagem de programação Python, estatística básica

**Meta:** ao final da prática, o aluno será capaz de utilizar ferramentas de análise de dados para calcular indicadores estatísticos e comparar valores.

### Roteiro:

Ler dados de arquivo (arquivo *series.csv*)

- Se estiver utilizando o Google Colab, você deve montar a sua pasta do Drive



- Para testar, rode o comando abaixo para listar os arquivos no seu Drive.

```
!ls "/content/drive/My Drive/"
```

- Coloque o arquivo "series.csv" em uma pasta no seu Drive, e então você poderá ler de lá.

```
path="/content/drive/My Drive/Colab Notebooks"
```

```
data = pd.read_csv(path+"/series.csv",index_col=False,header=None,squeeze=True)  
display(data)
```

```
0      13  
1       3  
2       5  
3       6  
4       7  
5       9  
6       5  
7      33  
8      67  
9     432  
10      5  
11      7  
12     35  
13     67  
14     83  
15     57  
16     88  
Name: 0, dtype: int64
```

- Explorar os dados com base em estatísticas descritivas

```
# Mínimo  
print("O mínimo é: ",data.min())  
  
#Máximo  
print("O máximo é: ",data.max())  
  
#Média  
print("A média é: ",format(data.mean(),".2f"))  
  
#Desvio Padrão  
print("O desvio padrão é: ",format(data.std(),".2f"))  
  
#Moda  
print("A moda é: ",data.mode())
```

```
O mínimo é: 3  
O máximo é: 432  
A média é: 54.24  
O desvio padrão é: 101.94  
A moda é: 0 5  
dtype: int64
```

- Calcular os percentis

```
# Percentil 25 (1o quartil)
print("Percentil 25: ",format(data.quantile(.25),".2f"))

# Percentil 50 (2o quartil ou mediana)
print("Percentil 50: ",format(data.quantile(.50),".2f"))

# Percentil 75 (3o quartil)
print("Percentil 75: ",format(data.quantile(.75),".2f"))

# Percentil 95
print("Percentil 95: ",format(data.quantile(.95),".2f"))
```

```
Percentil 25: 6.00
Percentil 50: 13.00
Percentil 75: 67.00
Percentil 95: 156.80
```

- Calcular a tabela de frequências

```
data.value_counts()
```

```
5      3
67     2
7      2
57     1
88     1
83     1
432    1
13     1
9      1
6      1
3      1
33     1
35     1
Name: 0, dtype: int64
```

### Atividade (Entregar via PVANet o em formato .ipynb):

1. Faça um código para ler os arquivos *altura\_homens.csv* e *altura\_mulheres.csv*. Esses arquivos contém as alturas (em cm) de 1000 homens e 1000 mulheres, respectivamente. Em seguida, responda às seguintes perguntas:

- a) Qual a altura mínima e máxima dos homens e das mulheres dessas amostras?
- b) Qual a média de altura dos homens e das mulheres? E qual a mediana dessas alturas?
- c) Qual o desvio padrão da altura dos homens e das mulheres?
- d) Qual o percentual de homens com altura menor que 160cm?
- e) Qual o percentual de mulheres com altura maior que 180cm?
- f) Um homem com altura 185cm está em qual percentil? (pesquise sobre a função *percentileofscore* do pacote *scipy*)
- g) Uma mulher com altura 150cm está em qual percentil?
- h) Quais as três alturas de homens que são as mais frequentes? Quantos homens possuem essas alturas?
- i) Quais as três alturas de mulheres que são as mais frequentes? Quantas mulheres possuem essas alturas?
- j) Um homem com altura 185cm está distante quantos desvios padrões da média dos homens?
- k) Um homem com altura 145cm está distante quantos desvios padrões da média dos homens?
- l) Uma mulher com altura 185cm está distante quantos desvios padrões da média das mulheres?
- m) Uma mulher com altura 145cm está distante quantos desvios padrões da média das mulheres?
- n) É possível afirmar com determinado grau de confiança que uma pessoa com altura 150cm é um homem ou uma mulher?
- o) E uma pessoa com altura 190cm?
- p) E uma pessoa com altura 165cm?
- q) As alturas dos homens e mulheres seguem uma distribuição Normal?

2. Para que serve a função “*describe()*” de uma Series?

3. Para que serve a função “*unique()*” de uma Series?

### Referência:

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.Series.html>

(Seção *Computations / Descriptive Stats*)

