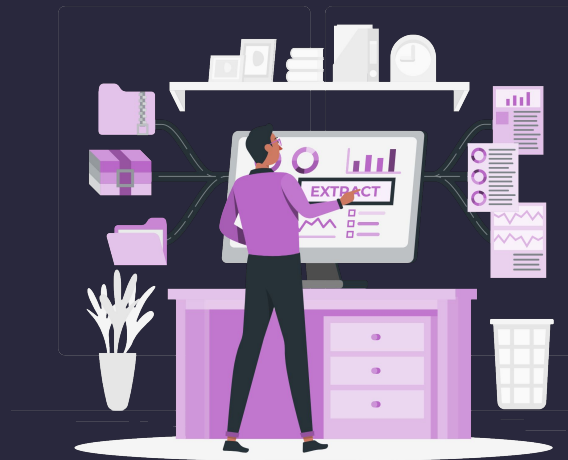




ESTATÍSTICA DESCRITIVA

AULA 01

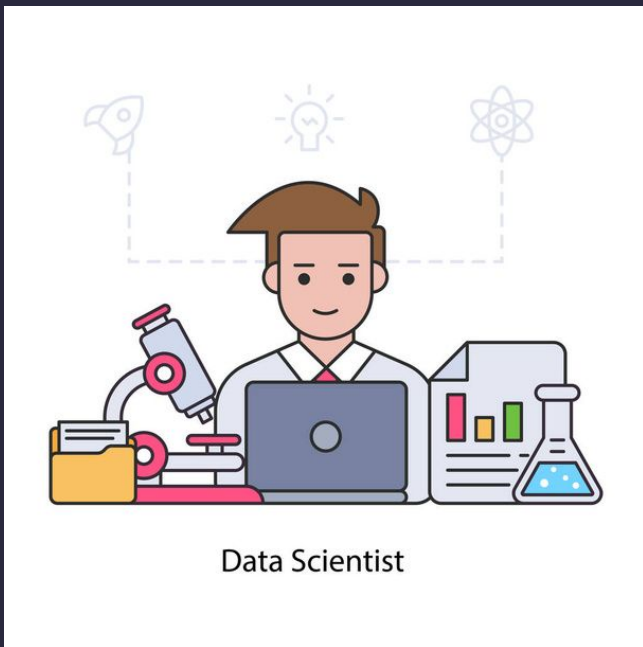


01

ESTATÍSTICA



/CIENTISTA DE DADOS



O Cientista de Dados é um **estatístico** que sabe programar.

O Cientista de Dados é um programador que sabe **estatística**.

CIENTISTA DE DADOS



DEFINIÇÃO DE ESTATÍSTICA

A estatística é um campo da matemática **que se ocupa da coleta, análise, interpretação, apresentação e organização de dados.**

Essa ciência técnicas e métodos para entender e interpretar informações numéricas, permitindo a descrição de conjuntos de dados, a inferência de propriedades da população com base em amostras, a formulação de hipóteses e a previsão de resultados.

CAMPOS DA ESTATÍSTICA

Estatística descritiva

Envolve a descrição, a sumarização e a compreensão dos dados disponíveis. Utiliza medidas de tendência central, medidas de dispersão e técnicas de visualização de dados para apresentar e resumir os dados de maneira fácil de entender.

Estatística inferencial

Utiliza os dados disponíveis para fazer inferências ou generalizações sobre a população maior a partir da qual os dados foram coletados. Isso envolve técnicas como a estimativa de parâmetros, o teste de hipóteses, a análise de regressão e a análise de variância. A estatística inferencial permite que façamos previsões ou tomemos decisões com base nos dados disponíveis.



CAMPOS DA ESTATÍSTICA

Estatística descritiva

- Trata da organização e descrição dos dados.
- Utiliza medidas de tendência central como média, mediana e moda.
- Emprega medidas de dispersão como variação, desvio padrão e amplitude.
- Uso de gráficos e visualizações de dados para entender a distribuição e a estrutura dos dados.
- Facilita a compreensão inicial e a interpretação dos dados.



CAMPOS DA ESTATÍSTICA

Estatística inferencial

- Permite fazer inferências ou generalizações sobre uma população com base em uma amostra.
- Envolve técnicas como estimativa de parâmetros e teste de hipóteses.
- Utiliza análises como regressão e análise de variância para prever e decidir.
- Permite a tomada de decisões e previsões com base nos dados.
- Ajuda a entender as relações, as tendências e os padrões nos dados.



02

DESCRITIVA: MEDIDAS DE TENDÊNCIA CENTRAL



MÉDIA

A média é a soma de todos os valores em um conjunto de dados, dividida pelo número total de valores. Ela é uma medida de tendência central que indica o valor "típico" em um conjunto de dados.

MEDIANA

A mediana é o valor que separa a metade superior da metade inferior em um conjunto de dados quando esses dados são ordenados. Em outras palavras, é o valor no meio do conjunto de dados. Se houver um número par de observações, a mediana será a média dos dois números do meio.

DESVIO PADRÃO

O desvio padrão é uma medida de dispersão ou variabilidade em um conjunto de dados. Ele indica o quanto os valores em um conjunto de dados tendem a se desviar da média. Em outras palavras, o desvio padrão quantifica a variação ou dispersão dos valores em torno da média.

Um desvio padrão baixo significa que os valores estão geralmente próximos da média; um desvio padrão alto significa que os valores estão espalhados por uma faixa mais ampla.

Resumidamente, é uma medida de quanto os dados variam em torno da média do conjunto de dados.

Desvio Padrão (Dp)

$$Dp = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

x_i = valor individual

\bar{x} = média dos valores

n = número de valores



PRECISAMOS CALCULAR TUDO ISSO?



@mrafaelbatista



messiasbatista

www.mrafaelbatista.dev



VAMOS CALCULAR COM PROGRAMAÇÃO





VAMOS CALCULAR COM PROGRAMAÇÃO



BIBLIOTECAS PYTHON PARA ESTATÍSTICA

- NumPy
- SciPy
- Pandas
- StatsModels
- Matplotlib
- Seaborn
- Plotly

```
import numpy as np
import scipy as sp
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sb
```


CALCULANDO A MÉDIA COM PYTHON

```
import pandas as pd  
x = pd.DataFrame({'num_filhos': [2, 3, 4, 5, 10]})  
x.mean()
```

CALCULANDO A **MEDIANA** COM PYTHON

```
import pandas as pd  
x = pd.DataFrame({'num_filhos': [2, 3, 4, 5, 10]})  
x.median()
```

CALCULANDO O DESVIO PADRÃO COM PYTHON

```
import pandas as pd
x = pd.DataFrame({'num_filhos': [2, 3, 4, 5, 10]})
x.std()
```

EXERCÍCIO

Construa o conjunto de dados abaixo e descubra a média, mediana e desvio padrão dos responsáveis:

```
import pandas as pd
```

```
x = pd.DataFrame({  
    'num_filhos': [2, 3, 4, 5, 10],  
    'idade_resp': [29, 32, 40, 32, 30]})
```

EXERCÍCIO

```
import pandas as pd
```

```
x = pd.DataFrame({  
    'num_filhos': [2, 3, 4, 5, 10],  
    'idade_resp': [29, 32, 40, 32, 30]})
```



RESULTADOS

Média:

Mediana:

Desvio Padrão:





ESTATÍSTICA DESCRITIVA

AULA 01

