



I EXPOTECH

Ênfase em Análise de Dados em *Big Data*

07-09 Novembro 23



Introdução

Este trabalho de extensão, é apresentado a disciplina extensionista de Tópicos de Big Data, tem por objetivo apresentar uma análise de dados de Regressão Simples, bem como uma análise descritiva sobre a relação do número dos casos de Covid em relação com o número de óbitos por Covid, no estado do Rio de Janeiro em outubro de 2021, utilizando os dados apurados em cada município.



Big Data e Regressão Linear

Big Data é o termo dado a um massivo número de dados coletados ou criados a partir de informações obtidas, seja de empresas, pesquisas ou até redes sociais. Como é um número de dados muito grande, é difícil de ser analisado com formas tradicionais, o trabalho da análise de Big Data é filtrar, e disponibilizar as informações mais significativas e de maior interesse.

Um bom exemplo disso, é a Regressão Linear, que utiliza a análise do comportamento de uma variável relacionada a uma ou mais variáveis, reproduzindo um gráfico e gerando a linha de Regressão Linear, com o objetivo de analisar a relação entre x e y.



Metodologia

Para realizar a análise dos dados estudados, utilizamos em Python (Linguagem de Programação de Alto Nível e código aberto) as bibliotecas Pandas e Numpy, que facilitam o uso da ferramenta, tornando mais fácil a análise de dados através da manipulação de listas e DataFrames, que foi o que utilizamos.

Através do site do Brasil.io, conseguimos uma tabela com os dados de casos e óbitos por covid de toda a região do Brasil, sendo num quadro geral, ou por cada estado. Como nosso interesse era mais centrado no estado do Rio de Janeiro, apuramos apenas esses dados mais específicos.

Com a linguagem de programação Python, fomos capazes de transformar a tabela com os dados do estado do Rio de Janeiro, informando separadamente por município em um DataFrame, que nos possibilitou manipular os dados, e então definir os dados de Óbitos por Covid como a variável independente, os de Casos Confirmados como a variável dependente, podendo assim definir a Regressão Linear.

ANÁLISE DE DADOS

CASOS DE COVID NO RIO DE JANEIRO

João Victor e Heloís Racibi

I Exposição de Projetos de Extensão de Tecnologia da Informação do Campus UNESA - Niterói



Resultados da Regressão Linear com Python

```
# - Regressão Linear dos Casos de Covid -  
coef_angular, intercepto = np.polyfit(df['obitos'], df['confirmados'], 1)  
y_pred = coef_angular * df['obitos'] + intercepto
```

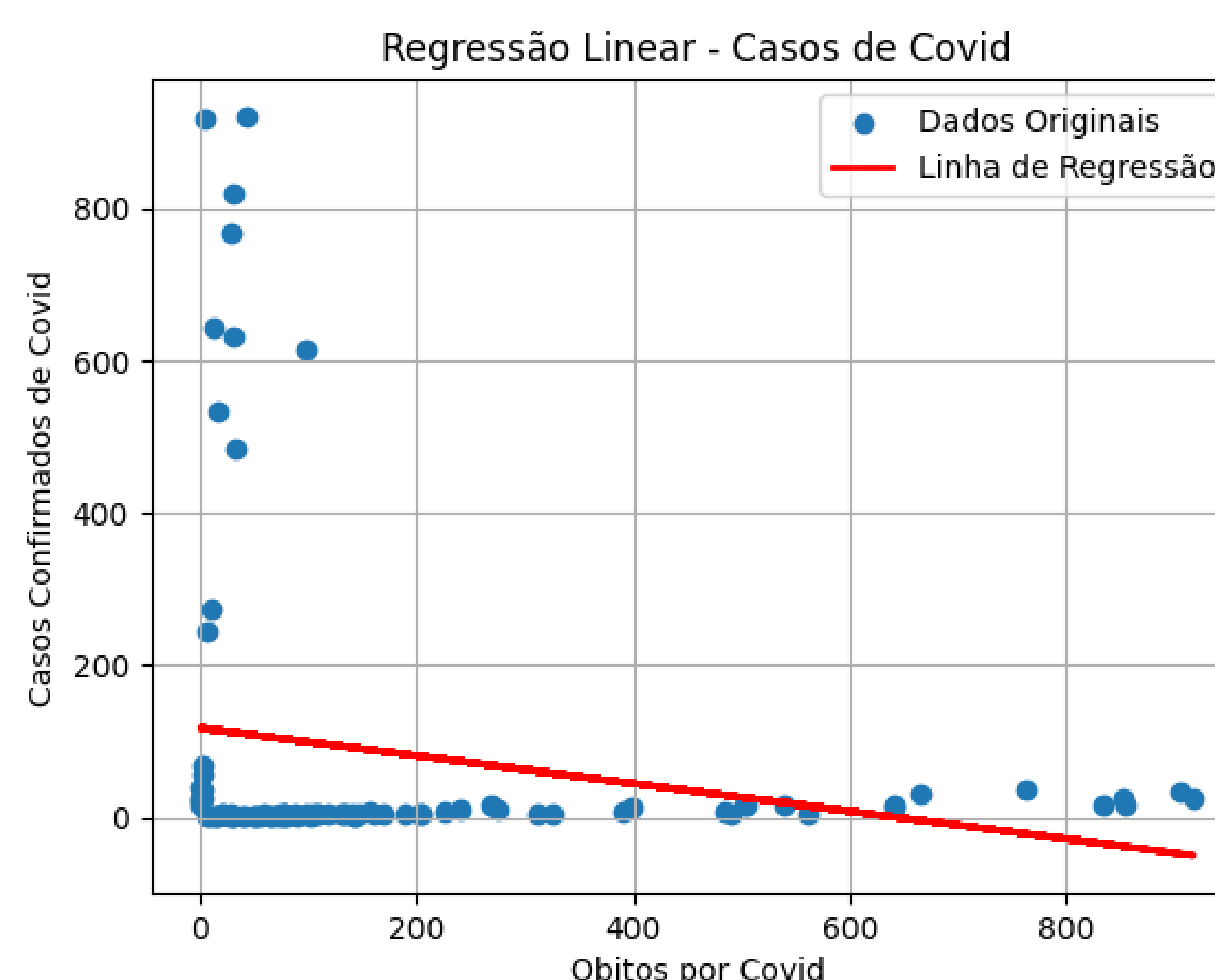
No Código acima, fazemos o cálculo da Regressão Linear utilizando o Número de Óbitos em relação com os Casos Confirmados.

```
plt.scatter(df['obitos'], df['confirmados'], label='Dados Originais')  
plt.plot(df['obitos'], y_pred, color='red', linewidth=2, label='Linha de Regressão')  
plt.xlabel("Óbitos por Covid")  
plt.ylabel("Casos Confirmados de Covid")  
plt.legend()  
plt.title("Regressão Linear - Casos de Covid")  
plt.grid(True)  
plt.show()
```

Como mostrado anteriormente, plotamos o gráfico de óbitos (x) e casos confirmados (y) com a biblioteca *matplotlib.pyplot*.

```
# Analise estatísticas da coluna Óbitos por covid  
Omedia = df['obitos'].mean()  
Omoda = df['obitos'].mode()  
Omediana = df['obitos'].median()  
Odesvio = df['obitos'].std()  
  
# Analise estatísticas da coluna Casos confirmados de Covid  
Cmedia = df['confirmados'].mean()  
Cmoda = df['confirmados'].mode()  
Cmediana = df['confirmados'].median()  
Cdesvio = df['confirmados'].std()
```

Ao lado, fizemos a análise descritiva dos dados de cada tabela, que são o cálculo da média, mediana, moda e desvio padrão.



Análise Descritiva dos Óbitos
Média: 186.29708695652175
Mediana: 79.5
Desvio Padrão: 242.01946123580944
moda: 0 29.0
1 31.0
Name: obitos, dtype: float64

Análise Descritiva dos Casos Confirmados
Média: 83.23114130434782
Mediana: 4.96
Desvio Padrão: 214.55741743543203
moda: 0 3.465
Name: confirmados, dtype: float64

Como pode ser visto no gráfico, a Regressão Linear é do tipo Correlação Negativa. Embora a linha indique uma queda, os pontos não são regulares, e isso é devido a vários fatores externos que variam de município para município, como o acesso a informação, a conscientização da população, acesso e índice de vacinação, entre outros.



Referências

- [1] Dados de Casos Confirmados, e Óbitos por Covid:
- <https://brasil.io/covid19>
- [2] Fonte de estudo dos códigos:
- <https://numpy.org/doc/stable/user/index.html>
- https://pandas.pydata.org/docs/user_guide/index.html



Estácio