

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. O objetivo geral é analisar a relação entre o Produto Interno Bruto (PIB) e a taxa de desemprego no Brasil. Para isso, são utilizados dados do PIB e da taxa de desemprego do Brasil de 2011 a 2020. O objetivos específicos são:

- Estimar a relação entre o PIB e a taxa de desemprego no Brasil;
- Analisar a influência do crescimento (ou declínio) do PIB na taxa de desemprego.

BIG DATA E REGRESSÃO LINEAR

Big data é um conjunto de dados muito grande, complexo e heterogêneo, que pode ser difícil de analisar com métodos tradicionais. **Análise de dados** em big data é o processo de encontrar padrões e tendências em conjuntos de dados grandes, complexos e heterogêneos. **Regressão linear** é um modelo matemático que descreve a relação linear entre duas variáveis.

Exemplo: A altura pode ser usada para prever o peso.

- Aplicações: Previsão, análise de risco e otimização de recursos.
- Vantagens: Pode identificar padrões e tendências, fazer previsões e otimizar processos.
- Desvantagens: Baseia-se em suposições e é sensível a outliers

METODOLOGIA

Para estimar a relação entre o PIB e a taxa de desemprego no Brasil, foi utilizada uma regressão linear simples. Os dados foram obtidos do IBGE, Google Bard e analisados em Python, usando as bibliotecas pandas, numpy e matplotlib.

As etapas do projeto foram: Coleção de dados, preparação dos dados, cálculo da regressão linear e interpretação dos resultados obtidos para estimar a relação entre as duas variáveis (PIB e taxa de desemprego, respectivamente)

RESULTADOS DA REGRESSAO LINEAR

Primeiramente, importamos todas bibliotecas necessárias. Após isso, colocamos nossos dados em arrays utilizando a biblioteca NumPy, onde o ‘X’ é a variável indepedente(PIB) e ‘Y’ e a variável dependente(Taxa de desemprego). Depois acontece o cálculo da regressão linear com a função ‘np.polyfit’. Em seguida, ocorre a previsão de ‘Y’. Agora, usamos a biblioteca ‘matplotlib’ para montarmos o gráfico da regressão linear. Finalizando o código, vamos apenas usar a estatística descritiva.

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 x = np.array([4.2, 4.5, 4.9, 5.5, 4.9, 4.4, 4.7, 5.2, 5.5, 4.6]) #em trilhões de R$
6 y = np.array([6.8, 6.9, 7.5, 7.6, 8.5, 9.8, 10.3, 11.1, 11.9, 13.5]) #em porcentagem
7
8 coeficiente_angular, intercepto = np.polyfit(x, y, 1)
9
10 y_pred = coeficiente_angular * x + intercepto
11
12 plt.scatter(x, y, label = 'Dados originais')
13 plt.plot(x, y_pred, color = 'black', linewidth = 2, label = 'Linha de Regressao')
14 plt.xlabel('PIB em trilhoes')
15 plt.ylabel('Desemprego em porcentagem')
16 plt.legend()
17 plt.title('Relação entre PIB e taxa de desemprego')
18 plt.grid(True)
19 plt.show()
20
21 x = pd.DataFrame({'Valores': [4.2, 4.5, 4.9, 5.5, 4.9, 4.4, 4.7, 5.2, 5.5, 4.6]})
22 y = pd.DataFrame({'Valores': [6.8, 6.9, 7.5, 7.6, 8.5, 9.8, 10.3, 11.1, 11.9, 13.5]})
23
24 x_media = x['Valores'].mean()
25 x_mediana = x['Valores'].median()
26 x_moda = x['Valores'].mode()
27 x_std = x['Valores'].std()
28
29 y_media = y['Valores'].mean()
30 y_mediana = y['Valores'].median()
31 y_moda = y['Valores'].mode()
32 y_std = y['Valores'].std()
```

O gráfico mostra que, em geral, quanto maior o PIB, menor a taxa de desemprego. Isso significa que o crescimento econômico está associado a uma redução do desemprego.

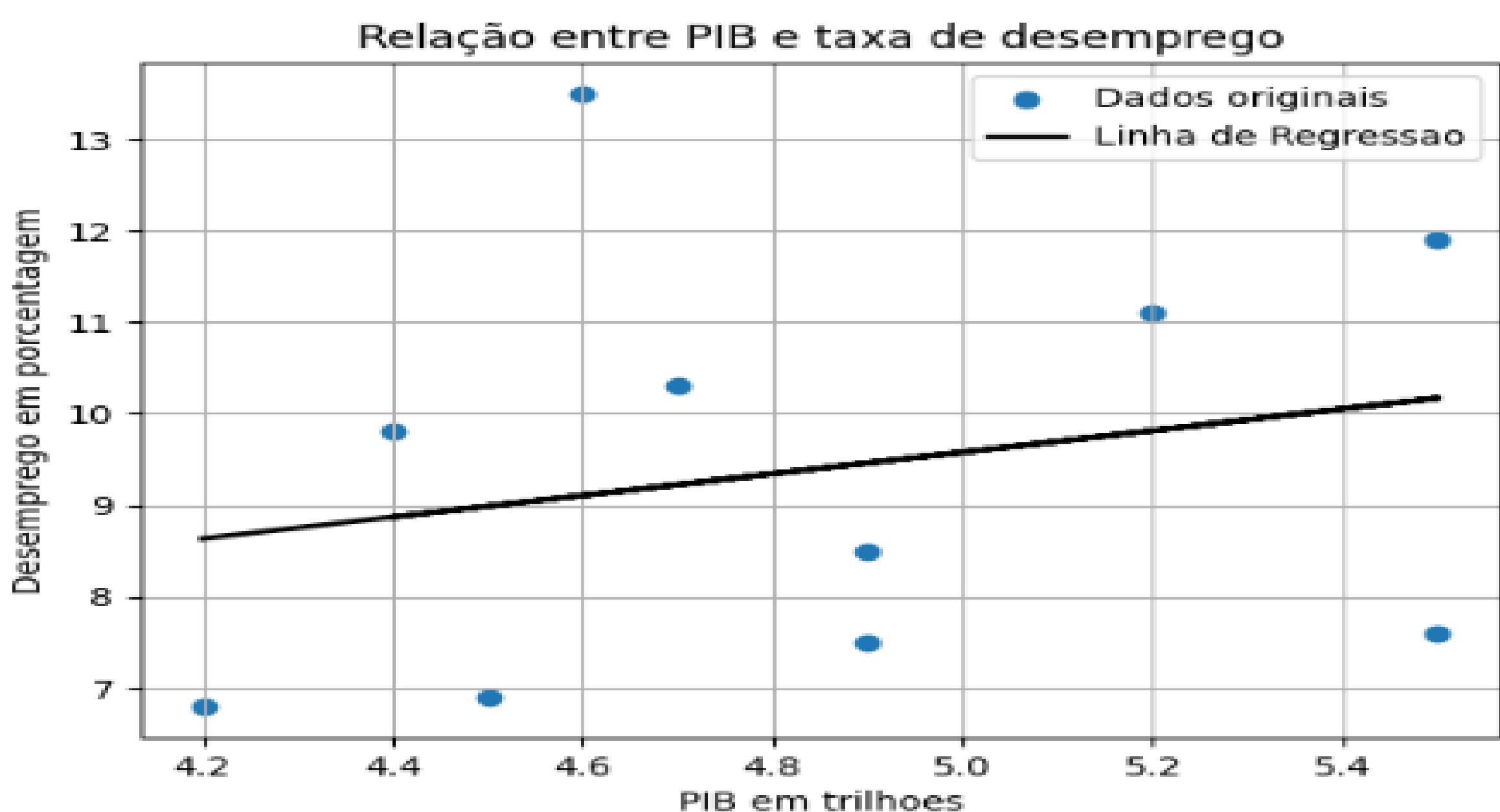


Figura 1. Gráfico de Regressão Linear resultante.

REFERÊNCIAS

[1] referência: <https://numpy.org/doc/>

[2] dados: <https://github.com/mathcute/data-repository>