

Verificando a correlação entre o IDH dos países e o número de mortes por milhão causadas pela COVID-19

Introdução

Desde 2020 o mundo passa por uma situação crítica: a pandemia causa pelo vírus COVID-19. Desde seu início, cientistas vêm estudando e tentando entender como este micro-organismo funciona e como se espalhou tão rapidamente por todo o planeta. Este trabalho, desenvolvido como método de avaliação da disciplina de MAT013 (probabilidade e estatística) da professora Hevilla Nobre Cezar, tem como fim analisar a correlação entre o número de mortes por milhão causadas pela COVID-19 e o IDH de todos os países, verificando se existe uma possível relação entre o nível de desenvolvimento de um país e a quantidade relativa de mortes acarretadas pelo chamado “coronavírus”.

Materiais e Métodos

Para realizar a análise, inicialmente, foi necessário encontrar uma base de dados confiável. A plataforma “Our World in Data” possui uma série de APIs *open access* e *open source* confiáveis que são utilizadas por importantes sites de notícias brasileiros, como G1, CNN Brasil e UOL Notícias, e foi escolhida para fornecer os dados necessários para este trabalho. Todos os dados utilizados podem ser consultados no site [Coronavirus \(COVID-19\) Deaths - Statistics and Research - Our World in Data](#).

Para trabalhar com a API, a linguagem de programação *Python* foi escolhida por sua sintaxe simples e por conter uma série de pacotes que facilitam a manipulação de dados. Utilizando o editor de código fonte *Visual Studio Code*, a primeira fase do código foi criar uma função para separar da API apenas os dados que são relevantes para o objetivo do trabalho em um arquivo de extensão “.csv”: país, população, IDH, mortes por milhão e a data de atualização dos dados. Isso foi feito para que se pudesse ter um maior controle e visualização dos dados, visto que, desta forma, os dados poderiam ser facilmente abertos por um editor de planilhas.

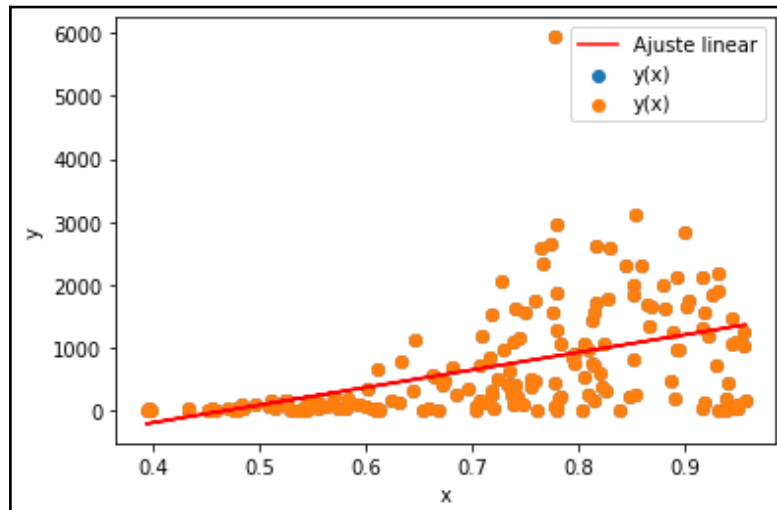
Com o coeficiente de correlação de Pearson, torna-se mais simples a análise de quão forte as variáveis são. A segunda fase do código foi criar uma função que fizesse os cálculos e devolvesse o coeficiente, pois é inviável manusear manualmente uma quantidade tão grande de dados. A última fase foi a função responsável por encontrar a reta que melhor se ajusta aos dados e seus desvios.

Resultados

Primeiramente, é importante ressaltar que nem todos os países possuíam os dados necessários para esta análise. Dos 195 países que existem no mundo, foi possível analisar 179, ou seja, esta análise possui cerca de 91% de confiabilidade. O valor do coeficiente de correlação gerado foi 0,488, e a reta que melhor se ajustou aos dados é dada pelo coeficiente angular 2799,75 e coeficiente linear -1318,34. Sendo assim, a equação da reta ficou parecida com a equação I. O desvio encontrado foi 759,90 e o erro médio 511,82. O Gráfico I mostra como ficou o ajuste linear com os dados obtidos, lembrando que o eixo X refere-se ao IDH e o eixo Y refere-se ao número de mortes por milhão.

$$2799,75x - 1318,34 \text{ Equação I}$$

Gráfico I - IDH vs número de mortes por milhão nos países



Conclusão

Apesar da relação entre as variáveis em destaque parecerem ter uma óbvia ligação, a análise neste trabalho acabou mostrando o contrário. O coeficiente de correlação mostrou que, apesar de apresentar uma certa ligação, não existe uma forte relação entre o desenvolvimento do país e a quantidade de mortes causadas pela COVID-19. Ademais, pode-se confirmar, analisando o gráfico, que certamente estas variáveis estão ligadas até certo ponto, e que provavelmente um dos dados está puxando bastante a média do restante das informações. Mas, seria necessário uma segunda análise para seguir mais a fundo nessa questão.