

Atividade 1 – Projeto Covid19 Brasil

O Brasil continua sofrendo com a pandemia causada pela infecção generalizada pelo vírus Covid-19, também conhecido por Corona vírus. Como você trabalha numa empresa de Data Science voltada a Helthcare, agora seu objetivo é analisar o impacto da doença nos estados brasileiros e correlacionar com suas respectivas características.

Utilizar o arquivo checkpoint2.ipynb para obter os dados do Covid para os estados e carregar os dados do censo de 2010. Faça e responda as seguintes perguntas:

- 1 – Realizar o join com o dataframe “dados_brasil” com o dataframe “dados_censo”.
- 2 – Qual seria o tamanho de amostra para uma pesquisa em todos os municípios do Brasil contidos no dataframe? (Utilizar a variância da renda per capita como base da fórmula de amostragem, $z = 1,96$ e margem de erro = 20).
- 3 – Qual seria o tamanho da amostra para São Paulo? (Utilizar a variância da renda per capita como base da fórmula de amostragem, $z = 1,96$ e margem de erro = 20)
- 4 – Agora que possui os tamanhos de amostra para o Brasil (n_{br}) e para São Paulo (n_{sp}), criar um dataframe para a amostra do Brasil e outro dataframe para amostra de São Paulo. (utilizar o método da amostra aleatória simples)
- 5 – Agora que possui os dataframes de Brasil (dados_amostra_br) e São Paulo (dados_amostra_sp), realizar as seguintes análises:
 - a) Obter as medidas descritivas: média, desvio padrão, primeiro, segundo e terceiro quartil para o número de infectados.
 - b) Obter as medidas descritivas: média, desvio padrão, primeiro, segundo e terceiro quartil para o número de mortes.
 - c) Gerar um histograma par o número de infectados e mortes. O desenho do histograma é similar com alguma das distribuições apresentadas?
 - d) Compare os resultados de Brasil e SP.