

DSC 530

Term Project

Astrid Fuentes

March 6th, 2021

Term Project Summary

For this project, I wanted to work with Covid-19 data. Unfortunately, the first dataset I selected did not contain at least five quantitative variables. Therefore, I had to select a new one and start over. My new data set contained the following variables: 'Date', 'Confirmed', 'Deaths', 'Recovered', 'Active', 'New cases', 'New deaths', 'New recovered', 'No. of countries', and a few other variables that I decided not to keep. Given that I have a reported date, I wanted to see if there is a significant difference between the number of deaths and the number of countries affected between the first and second semesters of the pandemic.

As I went through the different stages of the project, I realized that my dataset was not the best as all the variables were strongly related to each other and I did not feel I was able to extract meaningful information from it. I was able to build a test a hypothesis based on the variables I have, and I was able to go through all the topics as needed. However, If I had a data set more similar to the American Survey one used in the book, I would have probably been able to show more relevant plots and conclusions. One of my initial ideas included to be able to compare the distributions of Male and Female deaths to show that Male covid19 patients die more than females, for example. I was unable to do that because I did not have that data.

I do not think that any incorrect assumptions were made but I did face some challenges when working with the Probability Mass Function and the Cumulative Distribution Function. In this section, I tried to plot different variables and it just didn't look right. In the end, I plotted the PMF and CDF for the number of countries affected and I still was unable to confidently interpret what I was seeing. I do believe I need to review this topic and try this again with a more meaningful data set.

Overall, I think I was able to cover the different topics and I feel I would be able to complete this again with another data set if I had more time. This class and project have given me the tools I need in order to be able to successfully perform exploratory analysis and to be able to extract meaningful information from quantitative variables.

Link to repository:

https://github.com/asfuentes01/DSC530_FuentesA.git