DSC530 T301 Project

Dipika Sharma

Bellevue University

Project Subject Area

Life expectancy means the number of years a person is expected to live based on the statistical average. There is certain variable on which life expectancy depends based on some research. As part of my course project, I am exploring some of these variables to see the correlation between them and to understand how much these variables effect the life expectancy. These data information is from World Health Organization.

Project Outcome

After performing statistical process, I can say that Life expectancy has statistically significant relationship with variables like BMI, Measles, Polio and GDP. Although I found the correlation is week.

Missed during the Analysis

The data regarding Measles and Polio immunization and doses were less for post years 2003. As a result, I think the data is not sufficient.

Variables that could have helped

I tried to find relationship between life expectancy at birth with variables BMI, Measles and Polio. I think I was more focused on these variables if I have spent some time with variable GDP, I would have found strong relationship between the GDP and life expectancy at birth.

Also, I have not used variable life expectancy at age 60 as its distribution is random and do not have normal or skewed distribution.

Assumption in project

I assumed there would be a strong relationship of life expectancy at birth with variables BMI, Measles and Polio but to my surprise I could see only week relationship while visualizing I have noticed that distribution did not follow well defined pattern.

Challenges

There were few challenges that I want to discuss.

This is my first time working with life expectancy data. Although the topic was interesting, but I wish I have more information for the same. As I spent lot of time in understanding columns and its definition.

The second challenge is when we have to compare two scenarios using a PMF. To decide the column and how I want to divide the data to create two scenarios, I have spent lot of time. Later I have chosen year column and created scenarios for year <= 2003 and year > 2003.

Need Help

I think I want to spend more time in scatter plot where we compare two variables to find out the correlation and causation. But for me It was hard to see whether the relationship is linear or nonlinear. I think since the relationship between my target and predictive variable is week that is why it was hard to see the line in graph.