Sophia Haass, Samson Tessema, Robbie Durham, Aditi Kishore

Probability and Statistics Proposal

Modeling Health Care Costs Based on Smoking, BMI, Age, Gender, Number of Children, and Region

Why we are interested:

Through the entire Democratic Primaries, health care has been the center of attention. Rising health care costs and lack of coverage for millions of Americans is turning health care into a luxury item. Bernie Sanders, Elizabeth Warren, and others have constantly mentioned Medicare for All, while their opposition will argue that costs will be way too high for this to be feasible. While we cannot predict what these costs will be, we can predict things that are in our control, like BMI, whether we smoke, and how many children we have. We want to predict how health care costs can change for each person and show how to limit an individual's health care spending.

Model:

We want to model the data in Python using linear regression and any other models that we can find. We know from our very first class that using linear regression can lead to skewed results, like 120% obesity in the United States by 2030, so we are hoping that comparing with other models will lead to better analysis of the data. We want to use machine learning to be able to predict what a person can expect their health costs to be based on the dataset. We plan specifically on calculating weights for each category using the mean, standard deviation, and final healthcare cost (combined with each person's other characteristics).

Sources of data: Kaggle https://www.kaggle.com/mirichoi0218/insurance

Expected Outcome:

We assume younger people will have lower healthcare costs due to high metabolism, strong immune systems, and lower risk for diseases. Additionally, men will more likely have lower healthcare costs than women due to increasing breast cancer and cervical cancer cases and numerous pregnancy

complications. Non-smokers will have lower risk of lung cancer, loss of eyesight, and cardiovascular disease, so their health care costs would be less than people who smoke. We predict medical costs will be lower for a smaller number of children. Also, people with healthy BMI scores will have lower healthcare costs because BMI is an indicator of obesity and high body fat so having a healthy BMI score will reduce health care costs. We predict that people with low BMI scores will also have increased health care costs relative to individuals with healthy BMI scores, but lower than people with high BMI scores. This is because being underweight is still a health risk, but not as likely to cause diabetes or heart disease, which are expensive.

Printing:

We will print the poster in Latrobe during the last week of class because the civil and mechanical engineering department allows us to print there.