Final Project 10.3 Step 3

Moshe Burnstein

`r Sys.Date()`

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

```
__Diabetes and Hypertension__
```

Introduction

According to the CDC, 37.3 million Americans have diabetes. This accounts for roughly 10% of our population. 96 million American adults have prediabetes. The average medical costs of diabetics are more than double the costs of non-diabetics. Diabetes is the 7th leading cause of death in the U.S. One year ago, we celebrated the 100th anniversary of the discovery of insulin. Until this seminal moment, the onset of diabetes was a death sentence. We have yet to discover a cure, and our realistic goal now is to keep the disease under control. There are more diabetics now than ever before, and young people are developing it at unprecedented rates. The internet, society at large, and even the medical profession are full of ideas and strategies to combat the disease. From many varied diets, to lifestyle changes, to biofeedback, everyone has a different idea as to how to deal with this scourge. And this epidemic is worsening by the year.

Problem Statement

The world has yet to tame this disease because there does not yet exist a consensus as to what material things one can do to prevent and treat the disease. I propose exploring hypertension as a possible cause. If this proves to be true, treatments can be tailored to this cause and we would not be flailing in the dark. Data science must be leveraged to unfurl and parse through the wealth of data. I explore four different diabetes datasets, studying different peoples and different observations on the many subjects. I will compare the different populations to extract data to prove or disprove the problem statement.

Analysis and Implications

The Diabetes dataset in the "heplots" package shows diabetics to be slightly overweight, and non-diabetics to be of normal weight. The diabetics' average glucose and insulin resistance were significantly higher than normal. Further parsing of the data shows that relative weight exceeds normal for the chemical diabetics, while the sicker overt diabetics presented with a relative weight below normal. This may indicate that being overweight is not associated with the disease. Or this may point to the fact that overt diabetics produce insufficient insulin. The body needs insulin to extract glucose from the blood and deliver it to the body's cells. The body is forced to burn fat and muscle to produce energy. This would cause weight loss, an absolute side effect of this dreaded disease.

The Syrian refugees in Jordan present with high systolic and diastolic pressure readings. Their BMIs and hba1c averages are likewise high. An average age of 54 rounds out a population with many diabetes factors. Their average BMI of 34 puts them squarely in the category of obese. This dataset may well describe the prototypical diabetic population. In contrast, hypertension did not correlate with diabetes in the Pima dataset. Although the average pressure was 4 mm hg higher for diabetics than non-diabetics, 75 mm hg is well below the 80 mm hg line of hypertension. The partial correlation between blood pressure and diabetic pedigree function is slightly negative. Notwithstanding, the r^2 is 0.91%, which indicates that very little of the variation in pedigree function is due to pressure. Many machine learning models have been created to predict diabetes based on the conglomeration of data, and they also point to the fact that pressure is not independently correlated with a diabetes diagnosis, and therefore pressure is not the cause of diabetes.

Limitations

The NCSU dataset shows that women tend to higher BMIs and higher pressures than men. BMI and blood pressure are positively correlated across genders. This implies that either one causes the other, or that these two factors of cardiovascular disease are part of the general etiology of this and other diseases. This data implies that most strategies to deal with one or more of the factors, like weight loss or strict hba1c control, will not necessarily herald better outcomes. This is why, among other reasons, some experts counsel against focusing on one or even several of the diabetes disease factors studied. Instead, they insist that the best treatment for cardiovascular disease associated with diabetes is lifestyle changes. Studies have shown that reducing many of these factors has not led to better outcomes. Rather, exercising more, eating better and lowering cholesterol have produced better outcomes.

Concluding Remarks

In this great panoply of Type 2 diabetes factors, I have yet to learn the skills to assess all of the different factors with respect to each other. I have proven with high probability, by correlation and partial correlation, however, that hypertension is not the singular cause of diabetes. The best machine models provide prediction in excess of 80% of whether a patient will have a positive diagnosis of diabetes with said factors. From there it is guesswork as to how to treat and prevent the disease. Will reducing triceps skin thickness prevent diabetes? Data Science's Big Data capabilities are constantly expanding. Endocrinologists and laboratories can constantly add and update a central data repository which houses all things Type 2 diabetes. Every blood test result would be found here. Every response on patient surveys about exercise and diet from the initial visit throughout treatment would be culled. An unsupervised machine model can train itself on this enormous and constantly growing database. The model would improve its accuracy daily and would "spit out" to the patient his best treatment plan. What was once considered science fiction is now considered an attainable goal.