**Project: Analyzing the Risk of Cardiovascular issues between male/female patients**

**Introduction**

Cardiovascular disease (CVD) is a group of conditions that involve the heart and blood

vessels. Common complications include heart attack, chest pain (angina), or stroke.

The purpose of this project is to predict the effects of different parameters recorded in the data to predict mortality of the patient.

Cardiovascular disease (CVD) is the leading cause of death worldwide, yet important differences exist between men and women.

Men generally develop CVD at a younger age and have a higher risk of coronary heart disease (CHD) than women. Women, in contrast, are at a higher risk of stroke, which often occurs at older age.

**Statistical / Hypothetical Question**

Following hypothesis was tested.

* Resting Heart Rate for Male vs Female is pretty close
* Resting Heart Rate & Fasting blood sugar in Male Patients and Female are closely related

**Outcome of your EDA**

We do see some relation between some parameters/variables on the mortality.

The most important are

1. Exercise induced angina.

2. Maximum heart rate achieved -.Men’s heart rate is higher than women.

3. Fasting blood sugar.

4. Age - Older the patient, higher the risk

The other factors which I observed, did not show much effect.

I was able to support our hypothesis with the tests.

I also was able to generate prediction model (a logistic regression model), and found it be 70% accurate with the current dataset.

**What do you feel was missed during the analysis?**

Another factor which I did not plan earlier was

1. How over the years the heart decease changed in both sexes.

I did not analyze how the impact of changing times, clean eating habits and having access to better medication changed.

**Were there any variables you felt could have helped in the analysis?**

Some of the variables, which I felt should have helped in analysis was the region, country and some more variables on eating habits, cholesterol levels, family history etc. which would have made the study more accurate. Using them would have taken care of region specific as well as habit specific filtering and accurate predictions.

**Were there any assumptions made you felt were incorrect?**

There were some concerns regarding the sample size of the data. Since the data was from a single location, there might be some other factors in play which can be due to habits of people to that specific region or part of the world and may not be reflecting in the data. If we get more geographically separated data, we might come up with better model. This model may be more effective in the region where data comes from but our assumption that it will work for all might be wrong.

**What challenges did you face, what did you not fully understand?**

I had some challenges around the Hypothesis Testing and choice of the test statistic. I chose the difference between the means of the two groups as my main test statistics, however I still think could I use some other comparisons too as test statistics such as standard deviation or chi squared based tests.

I feel still as a novice and learning as I read and practice with different datasets.

**References:**

<https://gh.bmj.com/content/2/2/e000298>

Authors of this Dataset

Ronit

Discover Statistics Using R, Andy Field | Jeremy Miles | Zoe Field

R for Everyone, Jared P Lander

Think Stats, Allen B Downey