**A 250-500-word Summary of the Project by Emmanuel O Vidal**

**Statistical/Hypothetical question**

As a physician, one of the things that comes to mind is how well and how effective can we treat patients that are seen in the emergency room. In the emergency room all kinds of patients are seen and the physician has a limited time to make a decision on the patient seen. Making the wrong decision can have consequences on the patient life. Makary and Daniel in 2016, said medical error should be considered the third leading cause of death in the United States. The first leading cause is heart disease, hence the importance of formulating a diagnostic criterion in the emergency room which can help to stem the number of deaths recorded from heart disease. It would also help to reduce death due to errors of misdiagnosis of heart diseases. The statistical question was what variables can we use to predict the presence of heart disease in the emergency room.

**Outcome of EDA**

The outcome of the exploratory data analysis showed that the dataset I used had 1025 observations and 14 columns. These 14 columns where variables to help me predict the causes of heart diseases in the emergency room. The analysis and distribution showed that there were some few outliers at the tail end which where normal values and thus were not eliminated but left as part of the dataset. Male were more than females. The mean age was 54 years, mean systolic blood pressure was 131mmHg, mean cholesterol level was 246, mean maximum heart rate was 149. The normal probability plot was fair normal and was described by the normal probability distribution. The minimum cholesterol level and maximum was 126 and 564 respectively which is compatible with life. The covariance was positive signifying a linear relationship. The correlation was however very low. The regression model showed that cholesterol levels, age, resting blood pressure, rest ECG and the slope were major predictors of the presence of heart diseases in the dataset.

**Missed during the Analysis**

I felt that a binary regression would have been more beneficial or a decision tree to better predict the outcome variable.

**Variables that could have helped**

To my knowledge, I’m not sure of any other variable that could have helped. Perhaps the presence of other comorbid state or mental state of the patient.

**Assumptions**

I’m not sure what assumptions could be incorrect.

**Challenges Faced**

How to deal with outliers. I felt that if the outlier were within reasonable normal values, then they should be left alone. I’m not sure though if this was the right decision.

Reference

Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016;353:i2139. Published 2016 May 3. doi:10.1136/bmj.i2139

Link to Github Repository

<https://github.com/Zillah123/DSC_530->