**Name: Bhushan Suryawanshi   
Date: 2020-08-05  
Title: Heart Failure Analysis**

**Statistical/Hypothetical Question**

My hypothesis for this study was -

H1 = Ejection fraction impacts Heart Failure.

H0 = There is not effect of Ejection fraction on heart failure.

**Outcome of your EDA**

There were 299 observation in the dataset. The independent variables in the dataset are -

age, anaemia, creatinine\_phosphokinase, diabetes, ejection\_fraction, high\_blood\_pressure, platelets, serum\_creatinine, serum\_sodium, sex, smoking, time and dependent variable DEATH\_EVENT. I used ejection fraction for analyzing impact on death event.

It has been observed in the EDA analysis that Ejection fraction with p-value less than significance level of 0.05. Which means Ejection fraction has impact on heart failure.

Hence we reject null hypothesis H0.

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

There is lot can be done in the analysis. Since its first time I was doing the EDA it took time in cleaning data, coming up with right graphs and analysis. Spending more time and looking carefully at the variables could have improved analysis. Also it would have helped in getting more insights.

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

With a limited domain knowledge in this area I can say that if “blood pressure values” or “cholesterol levels” would have been good addition for this study. Race/Ethinicity which actually contribute to life style details can also be good addition to this study. It could have helped to classify results based on different life style groups. Also the dataset doesn’t mention if subjects under observation were on any medication or treatment.

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

I am using ejection fraction for my analysis and assumed it has normal distribution. When plotted normality graph it was found that the data is approximately normal. The variable used “creatinine\_phosphokinase” was assumed to be having normal distribution but normal probability plot shows its not normal.

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

There are lot of challenges in terms of domain. Also this is first time I was doing EDA and hence getting correct plots, analysis was time consuming. We can do EDA without domain. However to proceed forward in better way the domain knowledge will guide on assumptions or relations between different variables. In terms of EDA, what I still need to explore more is effectively applying survival analysis. I know the current EDA for heart failure is perfect case for survival analysis because it has death event. I will explore more about it in the coming 3 week vacation till I start with Fall Session.

Reference -

1. Davide Chicco, Giuseppe Jurman: Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making 20, 16 (2020). (<https://www.kaggle.com/andrewmvd/heart-failure-clinical-data> )