Project Proposal

Bhavya Bhoomireddy

ELEN 520

Topic:

Using Multi Linear Regression to predict heart failure using various predictors such as age, platelets in the blood, whether the patient has high blood pressure, male or female, follow up period(no of days), and so on.

Introduction:

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide.

Heart failure is a common event caused by CVDs and this dataset contains 12 features that can be used to predict mortality by heart failure.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, hyperlipidemia or already established disease) need early detection and management wherein a machine learning model can be of great help.

Hypothesis Testing:

Response: If the patient deceased during the follow-up period (DEATH_EVENT).

Predictors: Age, anaemia, creatinine_phosphokinase, diabetes, ejection_fraction, high_blood_pressure, platelets, serum_creatinine, serum_sodium, sex, smoking and time.

Null Hypothesis:
There is no relationship between the response variable and the 12 predictors.

Alternative Hypothesis:
There is a relationship between the response variable and the 12 predictors.

Dataset:

Dataset used is from Kaggle Website.
Link: https://www.kaggle.com/andrewmvd/heart-failure-clinical-data

Citation:
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://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5)