



Sardar Patel Institute of Technology, Mumbai
Department of Electronics and Telecommunication Engineering
B.E. Sem-VII (2021-2022)
EC344 - Machine Learning and AI

Name: Astik Jadhav

Roll No. 2018110019

Date: 30/11/2022

Subject : Data Analytics

Objective: Getting Familiar with SAS Studio and Perform Exploratory Data Analysis using SAS

System Requirements: SAS Studio

DataSet: The Dataset considered in this experiment is the inbuilt dataset from SASHELP. It is the data about the medical information about people with attributes like weight BP, Cholesterol, Smoking, Status(Alive or Dead), Cause of death.

Code: Refer SAS Code File.

Output: Please Consider the SAS Output File

Interpretation:

1. First we found out what types of data are there in the dataset. It consisted of Numerical data(weight, Height, Age) and Categorical data(Sex, BP_Status)
2. We found out the mean of the numerical data. The mean age at which the people getting diagnosed is 63 years. This is due to the fact that older people are more likely to get affected by CHD. The Average and median BP of CHD patients shows an abnormally high figure, this clearly indicates that CHD Patients have high BP.
3. By looking closely at the attribute of smoking the people who are found to have CHD smokes on average 9 cigarettes per day.
4. There were some of the values missing also. After getting the correlation of the attributes, there is no correlation among the features so we can say that the features are independent of each other.
5. However there lies some correlation of order 60% in the feature weight and height.
6. Next getting to know about the outliers we found that the people who died were having higher diastolic and systolic pressure than the people alive. This clearly indicates that the people who died had reached the critical stage of CHD.
7. Similarly when we see outliers for the Cholesterol ranges there were no outliers in the borderline since the range of the borderline is very small. For the desirable cholesterol we found that the outlier exists in the downward side this indicates that people are trying to have as much lower cholesterol levels as possible. For high cholesterol we found that due CHD the cholesterol level shows large variation of the higher side

Conclusion:

- We got introduced to SAS
- We understood how to load dataset in SAS, do computation and plot graphs.