Assignment 4

DS-5003/2003 Statistical and Mathematical tools for Data Science (Fall 2021)

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.

Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidemia or already established disease) need early detection and management.

You are required to answer the following questions by using **hypothesis testing**. None of the questions can be answered as Yes or No. You are required to setup and run the experiments to reach a conclusion. Ensure that your conclusions are sound by keeping in mind the Dos and Don'ts for each concept. All the intermediate steps, graphs and experimental results must be included while answering.

- Sex - Gender of patient Male = 1, Female =0
- Age - Age of patient
- Diabetes - 0 = No, 1 = Yes
- Anemia - 0 = No, 1 = Yes
- High blood pressure - 0 = No, 1 = Yes
- Smoking - 0 = No, 1 = Yes
- DEATH_EVENT - 0 = No, 1 = Yes

Questions:

Question 1. Are diabetic peoples being more likely to have Blood Pressure?
Question 2. Are people having anemia are more likely to be from female?
Question 3. Are people whose age is in between 55 and 65 having high Blood Pressure are less likely to survive?
Question 4. Are diabetic peoples who smokes are less likely to survive?
Question 5. Are male people being more likely to get diabetic?

Submission:

You are required to submit a report containing response to each question with supporting graphs, tables or stats. A single R code file containing all the questions must be submitted separately with proper partitioning of questions. Furthermore, you have to mention your definition of the subjective terms used while answering.