# PDS ASSIGNMENT 3&4

**MANOJ KUMAR SUGGALA**

**16357169**

**PART-2**

The provided question states that the dataset consist of diabetes records of 768 patients. It contains the following 8 attributes: (Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, and Age) and 1 Outcome which tells whether a particular patient has diabetes or not, [0 states no diabetes, 1 indicates the person is suffering from diabetes]. The three tasks are performed as following:

After reading the dataset from csv file we can see it as:

A screenshot of a computer

Description automatically generated

1. Here, we are going to take a random 25 records and we are finding out the mean Glucose and highest Glucose values of this sample and compare these statistics with the population statistics of the same variable, and also visualizing the same.

A graph of a graph with red and black bars

Description automatically generated

2) 98th percentile of BMI of your sample and the population and compare the results using charts:

The below bar graph clearly states that the Sample’s BMI is bit lower when compared to complete average. It's critical to keep in mind that the small sample size increases the possibility of sampling mistakes. In summary, using bar charts to compare statistics with data is a useful method for determining the representativeness of a sample and gaining additional insight into the population.

A red rectangle with black squares

Description automatically generated

3) Using bootstrap (replace= True), create 500 samples (of 150 observation each) from the population and find the average mean, standard deviation and percentile for BloodPressure and compare this with these statistics from the population:

A red and black rectangles

Description automatically generatedA red and black rectangular object

Description automatically generated

The above graphs states that the standard deviation of blood\_pressure of bootstarp samples and complete population and also Average blood\_pressure of bootstarp samples and complete population respectively. You can clearly see the difference between them.

A red and black rectangles

Description automatically generated

Now, above graph states that the 98th percentile of BloodPressure of both bootstrap data and the complete\_population.

with 42 as the seed value. Furthermore, with a standard deviation of 19.07, the average mean BloodPressure value for the bootstrap samples was 69.176. This suggests that the average blood pressure in the bootstrap samples was similar to the average in the population.

Finally, by using bootstrap sampling in our study, we were able to generate a sizable sample size from the population and obtain very accurate estimates of the Blood Pressure mean, standard deviation, and percentile, hence we can estimate that bootstrap sampling is useful for calculating the statistics and representing samples representatives.