**Assignment-3**

**Name: Sowmya Patlolla**

**Std Id: 16338256**

**Step 1**: **Data Collection**

Collected the data from the following link.

Data Link: <https://app.box.com/s/7qv44umhw0vnzgmoe9krfkfkv5kf2atv>

**Step 2: Data Preprocessing**

Loading the raw data and checking the null values in the loaded data and moving the cleaned data to the Data\_Clean folder.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Step 3: Data Analysis**

Loading the cleaned data for data analysis.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**a) set a seed (to ensure work reproducibility) and take a random sample of 25 observations and**

**find the mean Glucose and highest Glucose values of this sample and compare these statistics.**

**with the population statistics of the same variable. You should use charts for this comparison.**

**Importing the required python libraries for data analysis.**

**A screenshot of a computer

Description automatically generated**

**Finding the mean for Glucose column sample.**

sample\_mean=sample\_population["Glucose"].mean()

**Finding max value for Glucose column sample data.**

sample\_max\_value=sample\_population["Glucose"].max()

**Finding mean for Glucose column for whole population.**

population\_mean=df["Glucose"].mean()

**Finding max value for Gl;ucose column for whole population.**

population\_max\_value=df["Glucose"].max()

**Creating list for sample and population mean for visualization.**

keys=["sample","population"]

values=[sample\_mean,population\_mean]

print(values)

**A screenshot of a computer

Description automatically generated**

**Visualizing the sample and population means.**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Generating lists for max glucose values for sample and population data set.**

**A screenshot of a computer

Description automatically generated**

**Visualizing the sample and population max value.**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Find the 98th percentile of BMI of your sample and the population and compare the results using charts.**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Creating 500 samples using bootstrap.**

**Using bootstrap (replace= True), create 500 samples (of 150 observation each) from the population and find the average mean, standard deviation and percentile for Blood Pressure and compare this with these statistics from the population for the same variable. Again, you should create charts for this comparison. Report on your findings.**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

From the above figure we can see that the mean for bootstrap sample and population data set are almost equal.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

From the above figure we can see that standard deviation for bootstrap sample and population dataset are almost equal.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**