**Principles of Data Science (5530-0001)-Assignment 2**

**Name: Srujan Nagothu**

**ID: 16341576**

**Answers:**

1. 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.

**Code:**

Text

Description automatically generated

**Representation: Chart, bar chart

Description automatically generated**

**Analysis: when we compared the sample mean and max values with the population mean and max values of Glucose, we found that the values are almost similar.**

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

**Code:**

**Graphical user interface, text, application, email

Description automatically generated**

**Representation:**

**Chart, histogram

Description automatically generated**

**Analysis: From the above representation it is clearly shows that the 98th percentile of sample and population values of BMI are almost similar**

1. 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.

Code:

**Graphical user interface, text, application, email

Description automatically generated**

**Representation:**

**Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Analysis: From the above representation, the analysis clearly shows that the bootstrap samples mean, standard deviation and percentile (50th,75th & 90th) are all almost similar and nearer when compared to population values. So, we can infer that the bootstrap samples adequately reflect the characteristics of the population, and thus we can use them to make reliable estimates of Blood Pressure statistics.**