**Principles of Data Science - Assignment 4**

**Name: Praveen Kumar Reddy Kadapala**

**ID: 16355737**

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.

A screenshot of a computer

Description automatically generated

**A graph showing different colored squares

Description automatically generated**

The mean glucose values of the sample is less than the mean glucose values of the population. This implies that the sample's average glucose levels are lower than the population's average glucose levels. The sample's maximum glucose value is slightly less than the population's maximum value. This suggests that the sample's maximum glucose level is marginally lower than the population's maximum glucose level.

We can see from our comparison that the population mean and maximum glucose values are nearly identical to those of the sample.

b) Find the 98th percentile of BMI of your sample and the population and compare the results

using charts.

A screenshot of a computer program

Description automatically generated

**A blue and orange rectangular shapes

Description automatically generated**

The sample's 98th percentile BMI is lower than the population's 98th percentile BMI. This implies that the individuals in the sample, at the 98th percentile, have slightly lower BMIs than the population's corresponding percentile.   
The graph shown above makes it clear that the 98th percentile of the sample and the BMI values of the population are quite similar.

c) 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 for the same variable. Again, you should create charts for this comparison. Report on your findings .A screenshot of a computer program

Description automatically generated

A graph of different colored rectangular shapes

Description automatically generated

It is clear from the graph that the average blood Pressure readings from the bootstrap samples are consistent with the population mean because the bootstrap mean and population mean are close to each other.   
The population standard deviation is similar to the average bootstrap standard deviation. This implies that blood pressure measurement variation across bootstrap samples is similar to the population's overall variability.   
Both the bootstrap sample blood Pressure percentiles and the population percentiles closely correspond to each other. This implies that blood Pressure values in the bootstrap sample have a distribution that is same as of the entire population.