**Principles of Data Science (5530-0001)-Assignment 3**

**Name: Ankitha Srirama Reddy**

**ID: 16355578**

**Answers:**

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 program

Description automatically generated

**Results:**

**A graph showing different colored squares

Description automatically generated**

**Analysis:**

The sample mean glucose (113.92) is lower than the population mean glucose (120.89). This suggests that the average glucose levels in the sample are lower than the average glucose levels in the population.

The maximum glucose value (196) for the sample is marginally lower than the maximum value (199) for the population. This implies that the maximum glucose level in the sample is somewhat lower than the highest glucose level in the population.

Our comparison shows that the sample mean and maximum glucose values are almost the same as the population mean and maximum glucose values.

b) 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

**Results:**

**A blue and orange rectangular bars

Description automatically generated**

**Analysis:**

The 98th percentile BMI (44.68) of the sample is lower than the 98th percentile BMI (47.53) of the population. This suggests that the sample's participants, at the 98th percentile, have slightly lower BMIs than the corresponding percentile in the population.

It is evident from the above representation that the population's BMI values and the sample's 98th percentile are nearly identical.

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 screenshot of a computer

Description automatically generated

**Results:**

A graph of different colored bars

Description automatically generated

**Analysis:**

The close proximity of the bootstrap mean (68.97) to the population mean (69.11) indicates that the average BloodPressure readings from the bootstrap samples are consistent with the population mean.

The average bootstrap standard deviation (19.25) and the population standard deviation (19.36) are similar. This suggests that the total variability of the population is comparable to the variation in BloodPressure measurements across bootstrap samples.

The BloodPressure percentiles of the bootstrap samples closely match the corresponding percentiles of the population. This suggests that the distribution of BloodPressure values in the sample is representative of the population as a whole.