Report

Part A- The dataset diabetes.csv contains data of 768 patients, including 8 attributes and 1 response variable. The response variable, Outcome, has binary value (1 indicating the outcome is diabetes and 0 means no diabetes). For this analysis, we considered this data as a population. I have set a seed value of 45 to ensure reproducibility and took a random sample of 25 observations from the population. We then calculated the mean and highest Glucose values of the sample, which were 199.24 and 193, respectively. I have created bar charts to visualize the comparison of sample and population mean Glucose values, as well as the comparison of sample and population highest Glucose values.

From the bar charts, it can be seen that the sample mean Glucose value (199.24) was higher than the population mean Glucose value (120.89453125). However, the sample highest Glucose value (193) was less than the population highest Glucose value (199).

Overall, this analysis provides an insight into the Glucose values of the population and a sample of 25 observations.

Part B- I have calculated the 98th percentile of BMI for both the sample and the population, which were 46.233999999999995 and 47.52599999999996, respectively.We compared these percentiles using a bar chart, which showed that the sample 98th percentile of BMI was slightly lower than the population 98th percentile of BMI.This analysis provides an insight into the BMI values of the population and a sample of 25 observations, and how they compare at the 98th percentile

Part C- To perform the bootstrap sampling, Python's numpy library is used. I first loaded the data and extracted the BloodPressure column then calculated the mean, standard deviation, and percentile for this column for the entire population. I then performed the bootstrap sampling by drawing 500 samples of size 150 with replacement. For each sample, I calculated the mean, standard deviation, and percentile for BloodPressure. Finally, I plotted histograms of the bootstrap means and standard deviations, and a boxplot of the bootstrap percentiles, along with the population statistics.

Results

The population statistics for BloodPressure were:

- Mean: 69.10546875
- Standard deviation: 19.355807170644777
- Percentiles: 25th= 62.0, 50th= 72.0, 75th= 80.0

The bootstrap analysis yielded the following results:

- Mean of bootstrap means: 69.110252
- Mean of bootstrap standard deviations: 19.353704
- Mean of bootstrap percentiles:

•25th percentile: 61.0
•50th percentile: 72.0
•75th percentile: 79.0

The histograms of the bootstrap means and standard deviations, and the boxplot of the bootstrap percentiles, along with the population statistics are shown.