**REPORT**

**Overall Analysis:**

Data Loading and Sample Selection:

The code starts by importing necessary libraries and loading the diabetes dataset using Pandas. It then selects a random sample of 25 observations from the dataset.

Sample Statistics:

The code calculates the mean and the highest glucose value of the selected sample using the "mean" and "max" methods of Pandas. It also calculates the 98th percentile of BMI for the sample using the "percentile" method of NumPy.

Population Statistics:

The code calculates the mean and the highest glucose value of the entire population using the same methods as for the sample. It also calculates the 98th percentile of BMI for the entire population using the "percentile" method of NumPy.

Visualization:

The code creates bar charts to compare the mean and the highest glucose values of the sample and the population, as well as the 98th percentile of BMI for both. It uses Matplotlib library to create these charts.

Bootstrap Sampling:

The code defines a function "generate\_bootstrap\_samples" to generate bootstrap samples from the population using NumPy's "choice" method. It generates 500 bootstrap samples of 150 observations each from the "BloodPressure" column of the diabetes dataset.

Bootstrap Statistics:

The code calculates the mean, standard deviation, and percentile of each bootstrap sample using the "mean", "std", and "percentile" methods of NumPy, respectively. It also calculates the same statistics for the entire population using the same methods.

Bootstrap Visualization:

The code creates histograms to visualize the bootstrap percentiles, means, and standard deviations compared to the corresponding population statistics. It uses Matplotlib library to create these histograms.

Overall, the code performs statistical analysis on the diabetes dataset, compares the sample statistics to the population statistics, and uses bootstrap sampling to estimate the variability of the population statistics. It also visualizes the results using bar charts and histograms.

**Report for C:**

The objective of this report is to compare the statistical properties of the variable 'BloodPressure' obtained from a bootstrap sample to those of the population. The bootstrap method was used with a sample size of 150, and 500 samples were generated. The mean, standard deviation, and percentile of the 'BloodPressure' variable were calculated for both the bootstrap sample and the population, and the results were compared.

The average mean of BloodPressure for the bootstrap sample was found to be 69.152, which is very close to the population mean of 69.10. The standard deviation for the bootstrap sample was 19.128, which is also quite similar to the population standard deviation of 19.355. Finally, the 98th percentile for the bootstrap sample was 98.032, while for the population, it was 99.319.

Overall, my findings suggest that the bootstrap method is a trustworthy approach for estimating the statistics of the BloodPressure variable for the population.

Chart, bar chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

**Report for a and b:**

The diabetes.csv file contains data on 768 patients, with 8 attributes and 1 response variable (Outcome). For this assignment, we considered this data as a population, and the objective was to perform two tasks.

a) In the first task, I set a seed to ensure reproducibility and randomly sampled 25 observations from the population. I then calculated the mean glucose and maximum glucose values of the sample, which were 130.36 and 197, respectively.

I also calculated the mean glucose and maximum glucose values of the population, which were 120.89and 199, respectively. To compare these statistics, I created a bar chart which clearly shows that the mean glucose value of the sample is slightly higher than that of the population, while the maximum glucose value of the sample is the same as that of the population.

b) In the second task, I calculated the 98th percentile of BMI for both the sample and the population, which were 45.26 and 47.52, respectively. I then created a bar chart which clearly shows that the 98th percentile of BMI of the sample is lower than that of the population.