**REPORT**

**Introduction**:

The purpose of this report is to compare the statistics of BloodPressure variable obtained from a bootstrap sample and the population. To achieve this, I generated 500 samples, each containing 150 observations, from the population using the bootstrap method with replacement. We then calculated the average mean, standard deviation, and percentile for BloodPressure for both the bootstrap sample and population and compared the results.

**Results**:

The average mean of BloodPressure for the bootstrap sample was found to be 72.2886, which is very close to the population mean of 72.2728. The standard deviation for the bootstrap sample was 0.9897, which is also quite similar to the population standard deviation of 11.9009. However, the 98th percentile for the bootstrap sample was 98, while for the population, it was 91.01. This difference might suggest that the BloodPressure values in the bootstrap sample are skewed towards higher values compared to the population.

**Charts**:

The following charts compare the statistics of BloodPressure for the bootstrap sample and the population.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

BloodPressure Statistics Comparison for Bootstrap Sample and Population

**Conclusion**:

In conclusion, I found that the BloodPressure statistics for the bootstrap sample and the population were quite similar, except for the 98th percentile. This difference in percentile might indicate a slight skewness towards higher values in the bootstrap sample. Overall, my analysis suggests that the bootstrap method is a reliable way to estimate the statistics of BloodPressure variable for the population.

**Report for a & b:**

The data file diabetes.csv contains data of 768 patients, where there are 8 attributes and 1 response variable (Outcome). For this assignment purposes, we have considered this data as a population. The objective was to perform two tasks.

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.

To perform this task, I set a seed and took a random sample of 25 observations. I then calculated the mean glucose and maximum glucose values of the sample, which were found to be 133.56 and 197, respectively. I also calculated the mean glucose and maximum glucose values of the population, which were found to be 121.68 and 199, respectively.

To compare these statistics, I created a bar chart using ggplot2 library, which clearly shows the comparison between mean and maximum glucose values of the sample and population. As can be seen from the chart, 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.

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

To perform this task, I calculated the 98th percentile of BMI for both the sample and the population, which were found to be 42.912 and 46.8, respectively. I then created a bar chart using ggplot2 library, which clearly shows the comparison between the 98th percentile of BMI of the sample and population. As can be seen from the chart, the 98th percentile of BMI of the sample is lower than that of the population.

In conclusion, the analysis of the sample data indicates 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. The 98th percentile of BMI of the sample is lower than that of the population. These findings can be useful in understanding the diabetes population and can help in the development of better treatments and management strategies.