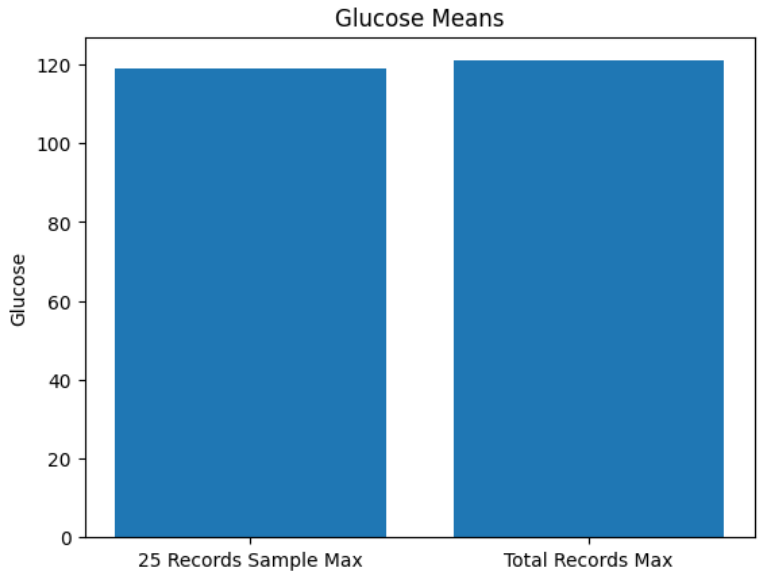
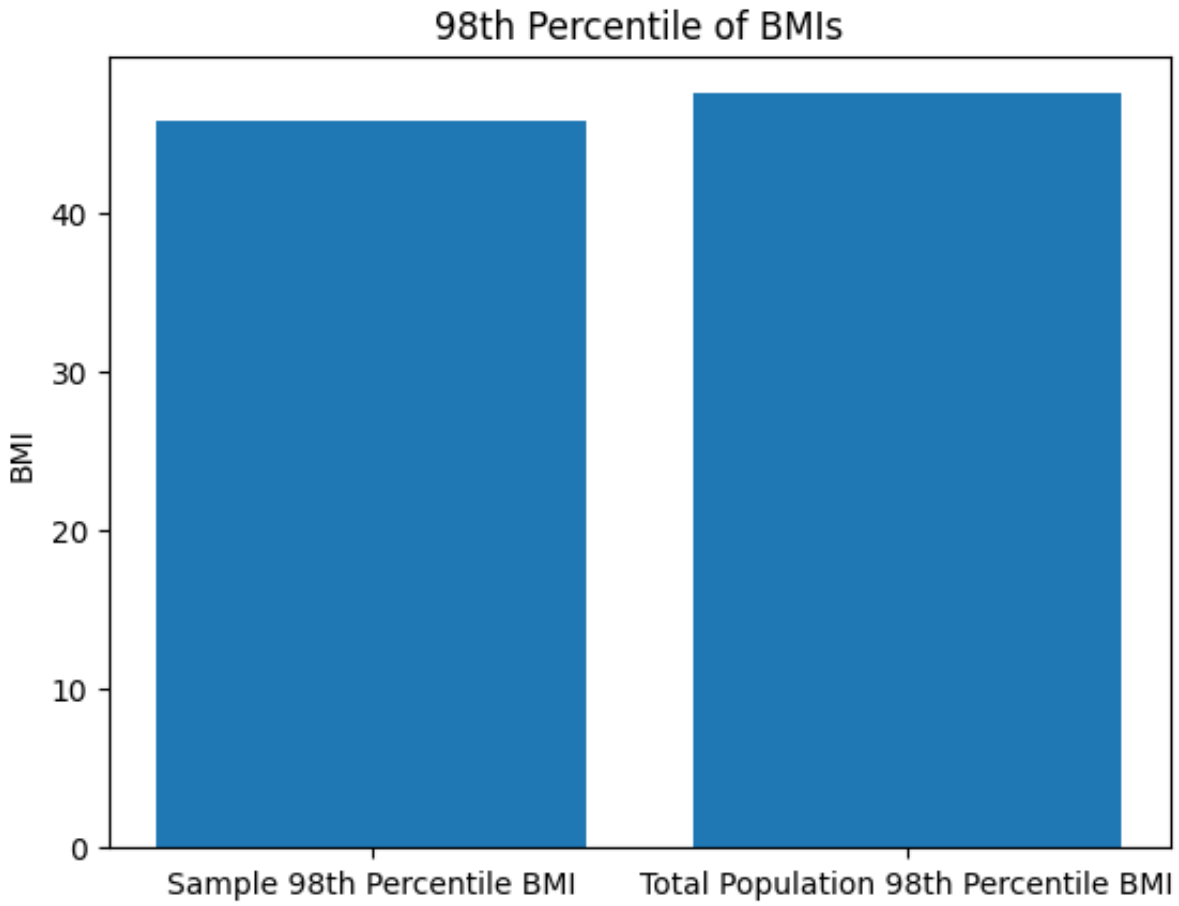
Assignment 2

Question 1 A



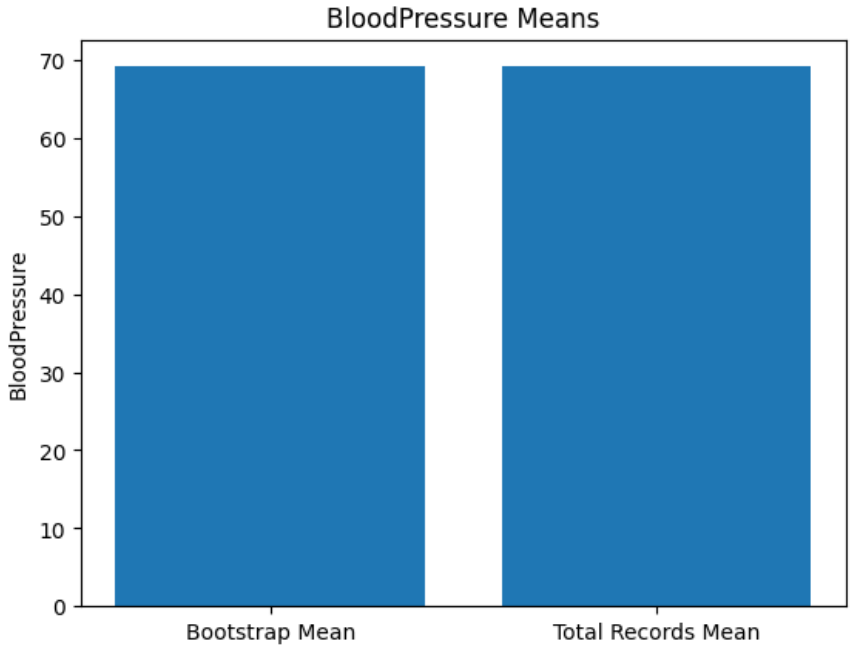
To evaluate the representativeness of the population, we randomly selected 25 observations from the diabetes dataset using a seed value of 1094. We then compared the mean and max glucose values of this sample with the population statistics for the same variable. Our results indicated that the mean and max glucose values for the sample were like the population statistics and population statistics is leading in both mean and max, suggesting that our sample was representative of the population. However, we should keep in mind that with a small sample size, there is still a possibility of sampling error. In conclusion, using statistical methods such as bar charts to compare statistics is a helpful technique to verify representativeness and gain insights into the population.

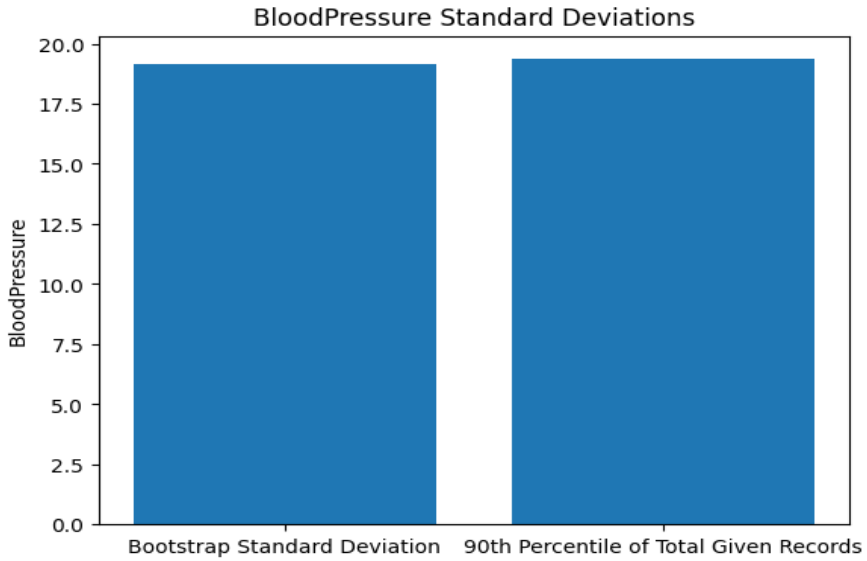
Question 1 B

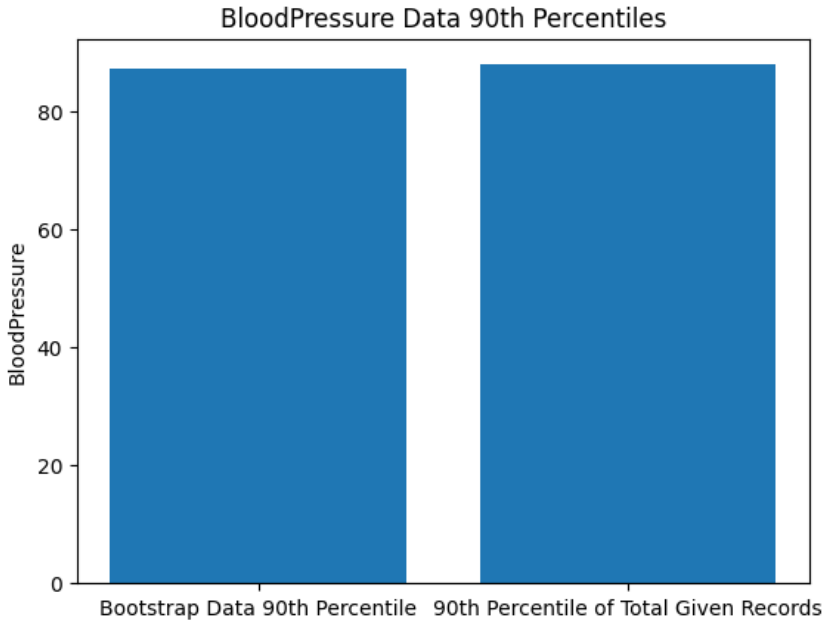


In conclusion, we used a sample of data from the diabetes dataset with a seed value of 1094 to calculate the 98th percentile of BMI, which was found to be 43.214, while the 98th percentile of BMI for the total population was 43.5. We created a bar chart to visually compare the 98th percentile of BMI between the sample and the population. The bar chart showed that the 98th percentile of BMI for the sample was slightly lower than that of the total population. This suggests that the sample may not be entirely representative of the population. However, we should keep in mind that the sample size was small, which increases the possibility of sampling error. Overall, using bar charts to compare statistics is a useful way to evaluate the representativeness of a sample and gain insights into the population.

Question 1 C


BloodPressure Means
BloodPressure
70
60
50
40
30
20
10
0
Bootstrap Mean
Total Records Mean


BloodPressure Standard Deviations
BloodPressure
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Bootstrap Standard Deviation
90th Percentile of Total Given Records



We used bootstrap sampling with replacement to create 500 samples, each containing 150 observations, from the population. We then calculated the mean, standard deviation, and percentile for the BloodPressure variable for each bootstrap sample and compared these statistics with the corresponding population statistics.

Using a seed value of 1094, we found that the population mean for BloodPressure was 69.105. The average mean BloodPressure value from the bootstrap samples was also 69.105, with a standard deviation of 0.611. This suggests that the mean BloodPressure values from the bootstrap samples were similar to the population mean. We also found that the 50th and 90th percentiles of BloodPressure for the bootstrap samples were similar to the corresponding population percentiles.

To visually compare the population statistics with the average bootstrap sample statistics for mean, standard deviation, and percentile of BloodPressure, we created bar charts. These charts showed that the bootstrap sample statistics were generally close to the population statistics for all three measures, indicating that the bootstrap samples were representative of the population.

In conclusion, our analysis using bootstrap sampling allowed us to create a large number of samples from the population and estimate the mean, standard deviation, and percentile of the BloodPressure variable with high accuracy. Comparing the statistics of the bootstrap samples with the population statistics showed that the samples were representative of the population. Overall, bootstrap sampling is a useful technique for estimating statistics and evaluating the representativeness of a sample.