**Task 1:**

A graph with blue and orange bars

Description automatically generated

Using a seed value of 42, we randomly picked 25 observations from the diabetes dataset in order to assess the population's representativeness. Next, we contrasted the population statistics for the same variable with the sample's mean and maximum glucose values. Our findings suggested that our sample was typical of the population because the sample's mean and maximum glucose readings matched population statistics, which lead in both mean and max. But we must remember that sampling error is still a possibility when the sample size is small. In conclusion, comparing data using statistical techniques like bar charts is a useful way to confirm representativeness and learn more about the population.

**Task 2**

A blue and orange rectangles

Description automatically generated

We computed the 98th percentile of BMI using a sample of data from the diabetes dataset, with a seed value of 42. To compare the 98th percentile of BMI between the sample and the population visually, we made a bar chart. The bar graph demonstrated that the sample's BMI, at the 98th percentile, was marginally lower than the population's average. This implies that there's a chance the sample isn't totally representative of the population. It is important to remember that the sample size was small, which raises the risk of sampling error. All things considered, comparing statistics with bar charts is a helpful technique to assess a sample's representativeness and learn more about the population.

**TASK 3**

A blue and orange rectangular shapes

Description automatically generated

A blue and orange rectangular shapes

Description automatically generated

A blue and orange squares

Description automatically generated

Using bootstrap sampling with replacement, we were able to generate 500 samples from the population, each with 150 observations. Next, for every bootstrap sample, we determined the BloodPressure variable's mean, standard deviation, and percentile. We then compared these results with the matching population data.

We discovered that the population mean for BloodPressure was 69.105 using a seed value of 42. Additionally, the bootstrap samples' average mean BloodPressure value was 69.176, with a standard deviation of 19.07. This implies that the bootstrap samples' mean BloodPressure values were comparable to the population mean. Additionally, we discovered that the bootstrap samples' BloodPressure 50th and 90th percentiles matched the respective population percentiles.

Bar charts were used to visually compare the population data for the mean, standard deviation, and percentile of BloodPressure with the average bootstrap sample statistics. For each of the three metrics, these charts demonstrated that the bootstrap sample statistics were, on average, quite similar to the population data, suggesting that the bootstrap samples were representative of the population.

To sum up, by employing bootstrap sampling in our study, we were able to generate a sizable sample size from the population and obtain very accurate estimates of the BloodPressure variable's mean, standard deviation, and percentile. The bootstrap samples were shown to be representative of the population when the population statistics and the bootstrap sample statistics were compared. All things considered, bootstrap sampling is a helpful method for calculating statistics and assessing a sample's representativeness.