**Name: Aishwarya Musuku**

**Id:16344653**

**Task 1:**

A graph of a comparison of glucose

Description automatically generated

We randomly selected 25 observations from the diabetes dataset using a seed value of 42 in order to compare the population statistics. The sample's mean and maximum glucose values were then compared to the population statistics for the same variable. The analysis of our sample's mean and maximum glucose readings revealed that it was identical to the population as a whole because the population statistics resulted in both mean and max are equal to that of sample's mean and max. However, it's important to keep in mind that small sample sizes can still result in sampling error. To sum up, comparing data using statistical methods such as bar charts can be a helpful way to verify representativeness and gain additional insights into the population.

**Task 2**

A blue and green rectangular bars

Description automatically generated

Using a sample of data from the diabetes dataset and a seed value of 42, we calculated the 98th percentile of BMI. We created a bar chart to visually compare the 98th percentile of BMI between the sample and the population. The bar graph showed that, at the 98th percentile, the sample's BMI was slightly below than the 98th percentile of  population. This suggests that there's a possibility the sample may not be entirely representative of the community. It's critical to keep in mind that the small sample size increases the possibility of sampling error. In summary, utilizing bar charts to compare statistics with data is a useful method for determining the representativeness of a sample and gaining additional insight into the population.

**TASK 3**

A yellow and grey rectangular squares

Description automatically generated with medium confidence

A yellow and grey rectangular object

Description automatically generated with medium confidence

A yellow and grey squares

Description automatically generated

We extracted 500 samples with 150 observations each from the population using bootstrap sampling with replacement. Next, we calculated the mean, standard deviation, and percentile of the BloodPressure variable for each bootstrap sample. Next, we contrasted these findings with the corresponding population data.   
Using a seed value of 42, we found that the population mean for BloodPressure was 69.105. Furthermore, the average mean BloodPressure value of the bootstrap samples was found to be 69.176, with a standard deviation of 19.07. This suggests that the mean BloodPressure values in the bootstrap samples were similar to the population mean. Furthermore, we found that the BloodPressure at the 90th percentile of the bootstrap samples matched the corresponding population percentiles.

The population data for BloodPressure's mean, standard deviation, and percentile were visually compared with the average bootstrap sample statistics using bar charts. These charts showed that the bootstrap sample statistics were, on average, fairly similar to the population data for each of the three metrics, indicating that the bootstrap samples were likely representative of the population.

In conclusion, we were able to extract a large sample size from the population and derive highly precise estimates of the mean, standard deviation, and percentile of the BloodPressure variable by utilizing bootstrap sampling in our research. When the statistics from the bootstrap samples and the population were compared, it was found that the samples were representative of the population.