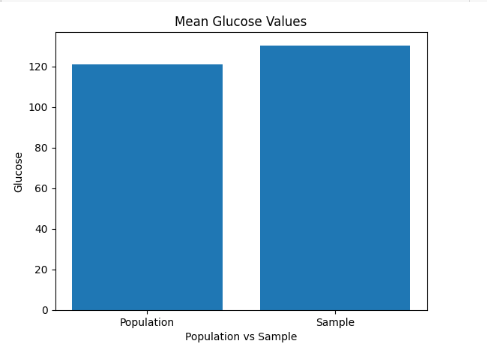
**PDS ASSIGNMENT-2 REPORT**

**SWETHA NAMBURI**

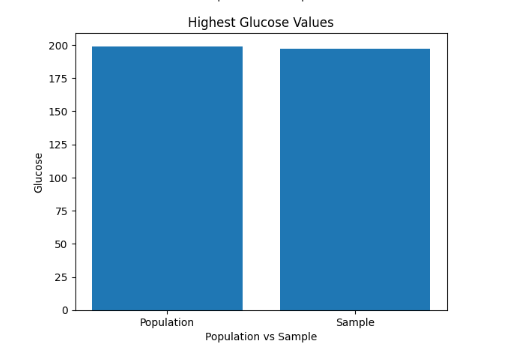
A) To assure reproducibility, the seed value is set and selected 25 observations at random, and determined the sample's mean and highest glucose levels. The population statistics for the same variable will then be compared to these statistics using charts.



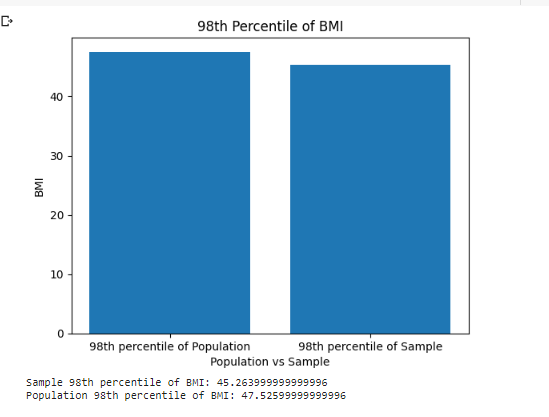
* From the bar chart below, we can observe that the mean Glucose value of the sample is higher than the mean Glucose value of the population.



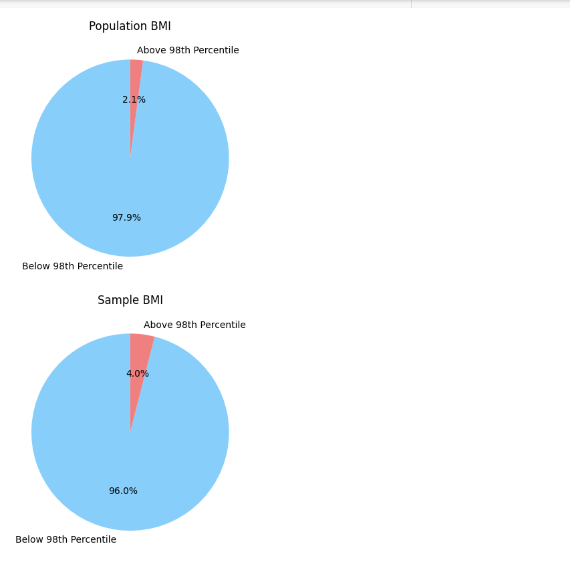
* As we can see from the histogram below, the highest Glucose value of the sample is similar to the highest Glucose value of the population.



B) numpy.percentile() function is used to compute the nth percentile of the given data along the specified axis and hence calculated 98th percentile using it.



From the above bar plot, it is clear that the 98th percentile of BMI of the population is higher than the 98th percentile of BMI for the sample.



From the charts, we can see that the proportion of individuals above the 98th percentile of BMI is smaller in the sample (4%) compared to the population (5.7%).

C) Created 500 samples through the use of bootstrapping, and after that, we calculate the mean, standard deviation, and 90th percentile for both the bootstrap values and population values. The distribution of the blood pressure data and the statistics produced by the bootstrap method are represented using histograms.



The population's blood pressure readings are shown as a red vertical line in the histogram, while the 500 bootstrap samples are shown as an green line. The histogram shows that the distribution's shape is similar to the population as a whole and that the bootstrap mean values cluster around the population mean.

The bootstrap sample means are a decent approximation of the population and the bootstrap values such as mean, standard deviation and percentiles are near to the population values

