ASSIGNMENT-2

a)

The graph shows that sample mean glucose is slightly higher than the sample mean population glucose and the max values are opposite to the mean where population max is slightly higher than the sample data. There is not much difference in the sample and population statistics when we set the seed value to 99.

b)

The 98th percentile of sample is 52.408 and population is 47.526 the 98% of the data will be under this only 2% of the data will be higher than this.

c)

Bootstrap analysis is a resampling technique used to estimate the variability of a statistic or to test a hypothesis by repeatedly sampling from the data with replacement. The code first sets a random seed to ensure reproducibility. Then, it defines a function called calc_stats that takes in a dataset and calculates its mean, standard deviation, and 98th percentile. Next, the code generates 500 bootstrap samples of size 150 each by randomly sampling with replacement from the BloodPressure variable. For each sample, the calc_stats function is called to calculate its mean, standard deviation, and 98th percentile. The bootstrap means, standard deviations, and percentiles are stored in separate lists. After that, the code calculates the population statistics (i.e., mean, standard deviation, and 98th percentile) for the BloodPressure variable using the entire dataset. Finally, the code creates three histograms to compare the distribution of the bootstrap means, standard deviations, and percentiles to the population statistics. Each histogram shows the frequency of the bootstrap statistic values, and a vertical dashed line is drawn to represent the population statistic value.

Based on the plots the 500 samples are generating different values with slight different statistics compared to the population and the values are just around the population diabetes values.

In summary, this code is performing a bootstrap analysis on the BloodPressure variable and visualizing the distribution of the bootstrap statistic values relative to the population statistic values.