Certainly! In the given scenario, we are dealing with a dataset named diabetes.csv, which contains information on 768 patients, encompassing attributes such as pregnancies, glucose levels, and blood pressure. The primary objective is to conduct various statistical analyses and make comparisons between the entire population and different randomly selected samples.

Analysis of a Random Sample and Glucose:

Establish a seed to ensure reproducibility.

Draw a random sample comprising 25 observations from the overall population.

Compute and contrast the mean and maximum glucose values between the population and the sample.

Illustrate the distribution of glucose through histograms and boxplots.

Comparison of the 98th Percentile of BMI:

Determine the 98th percentile of BMI for both the complete population and the previously selected sample.

Visualize the cumulative distribution of BMI and compare the 98th percentiles using charts.

Bootstrap Sampling and Blood Pressure Comparison:

Employ bootstrap sampling, generating 500 samples, each with 150 observations, from the population.

Calculate the means, standard deviations, and percentiles for blood pressure within the bootstrap samples.

Contrast these bootstrap statistics with the corresponding figures from the overall population.

Visualize the comparisons using histograms and cumulative distribution plots.

Presentation of Findings:

Display the results, including the mean and maximum values for glucose, the 98th percentile for BMI, and statistics related to blood pressure.

Offer a concise summary highlighting the observed differences and similarities between the population and the various samples.

This analysis aims to facilitate an understanding of how well the randomly chosen samples represent the broader population concerning essential health indicators. The utilization of charts aids in presenting a visually clear depiction of the parallels and distinctions between the overall population and the selected samples.