SANJANA AILENI
PDS ASSIGNMNET 2
STUDENT ID :12649897

Diabetes Dataset: Sample vs Population Comparison

Introduction

This report presents a statistical comparison and visualization of health metrics (Glucose, BMI, Blood Pressure) between a random sample and the overall diabetes population dataset. Key statistics and bootstrap sampling were used to analyze representativeness and variance.

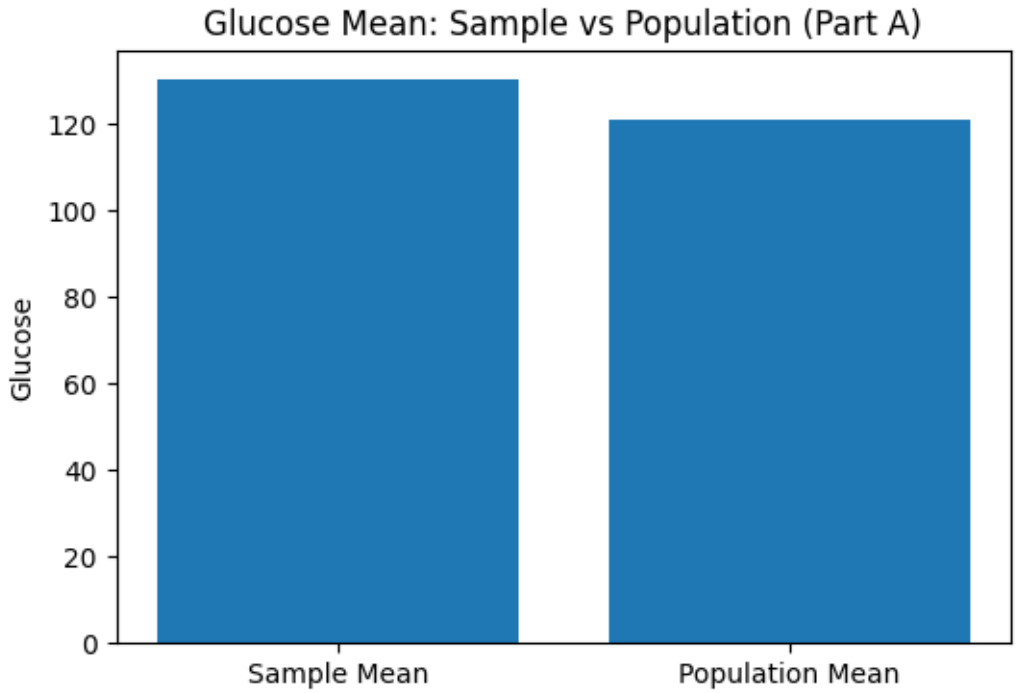
Statistical Comparison Table

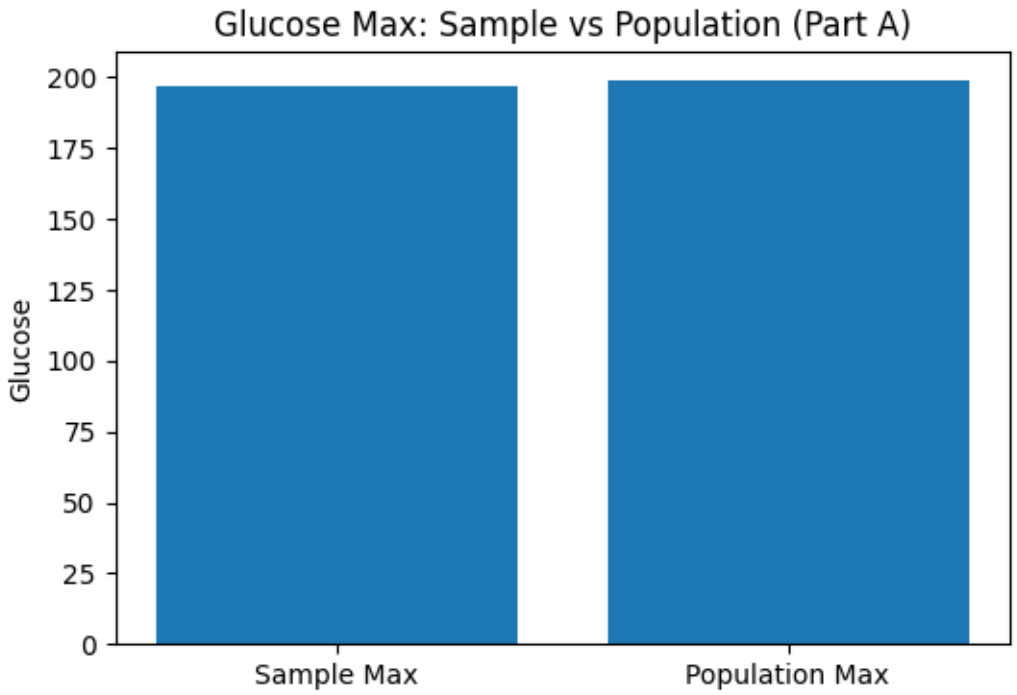
Metric	Sample Value	Population Value
Mean Glucose	130.36	120.89453125
Max Glucose	197	199
98th BMI Percentile	45.26399996	47.52599996
Mean Blood Pressure (bootstrap avg)	69.1522266666	69.10546875
SD Blood Pressure (BP)	19.192792750902836	19.355807170644777
98th BP Percentile (bootstrap avg)	98.03292000003	99.31999999994

SANJANA AILENI
PDS ASSIGNMNET 2
STUDENT ID :12649897

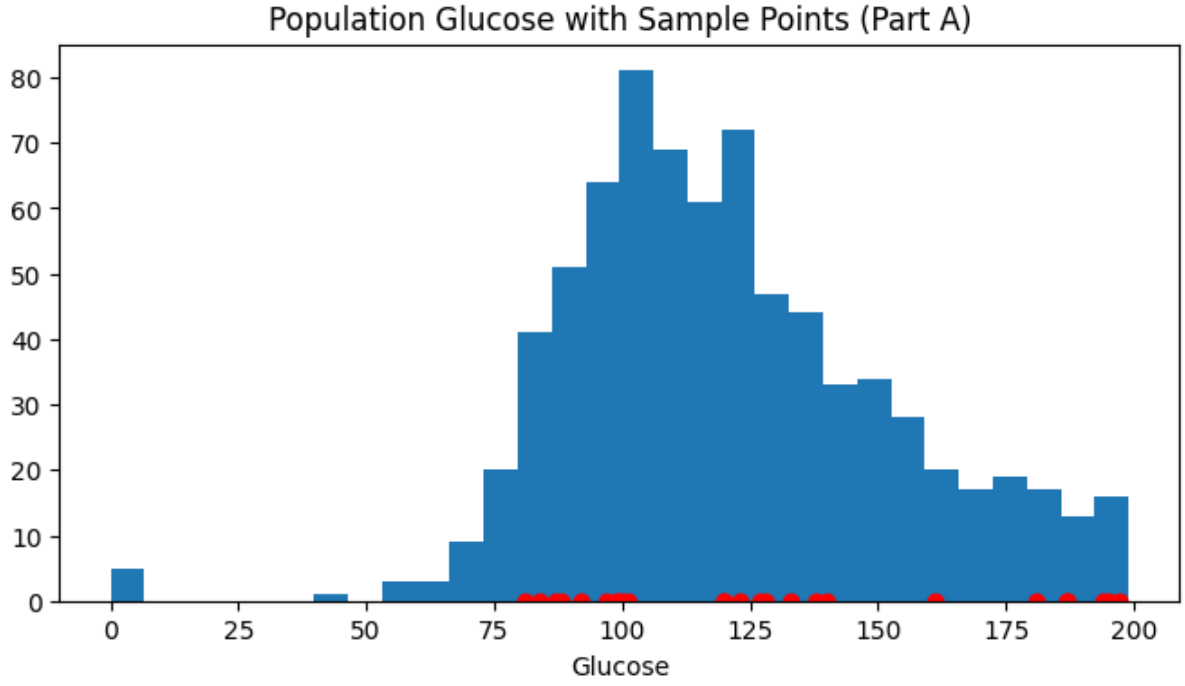
Comparison Bar Graphs

Glucose Mean and Max:



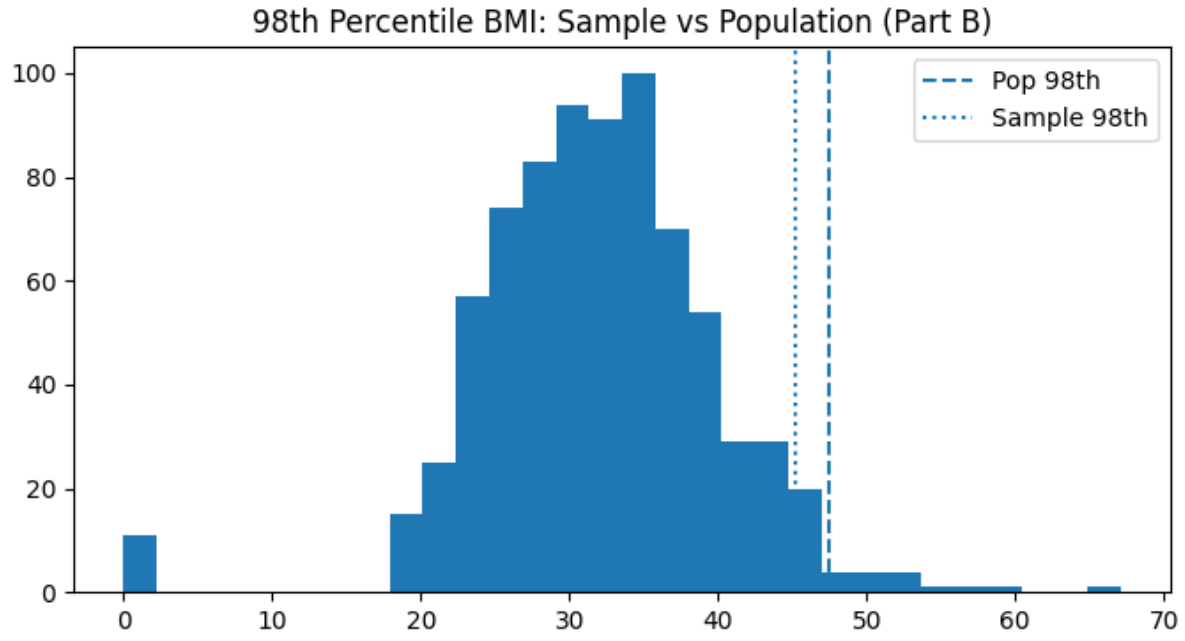


SANJANA AILENI
PDS ASSIGNMNET 2
STUDENT ID :12649897

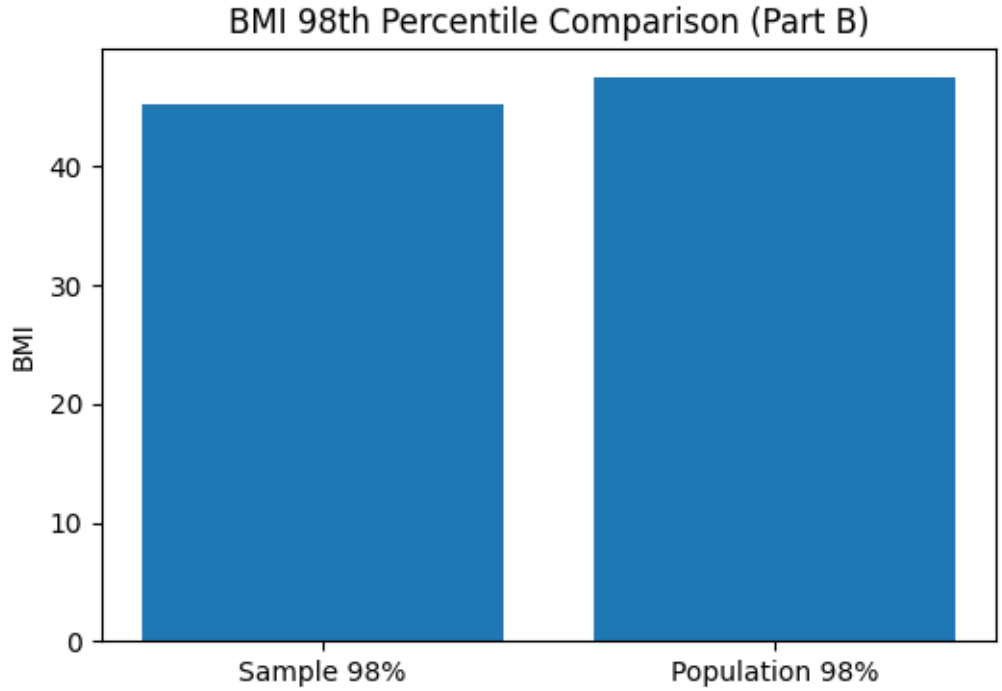


Interpretation:
The sample mean and maximum are close to population values, confirming that the random sample is representative for this metric.

BMI 98th Percentile:

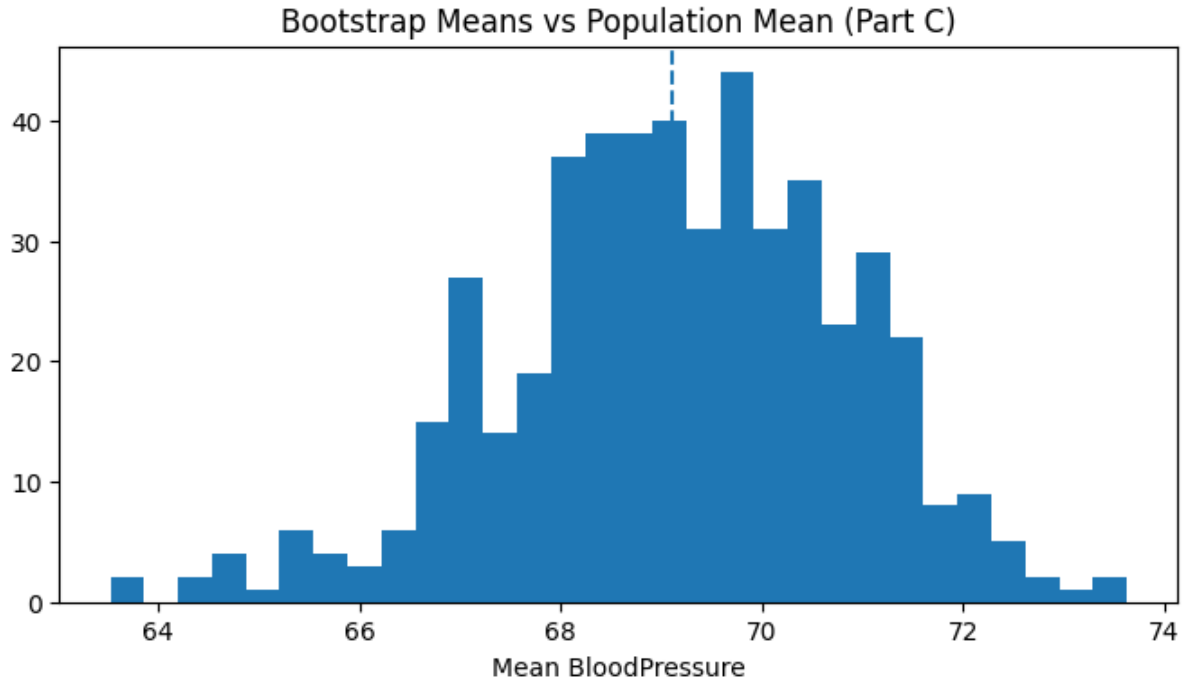


SANJANA AILENI
PDS ASSIGNMNET 2
STUDENT ID :12649897

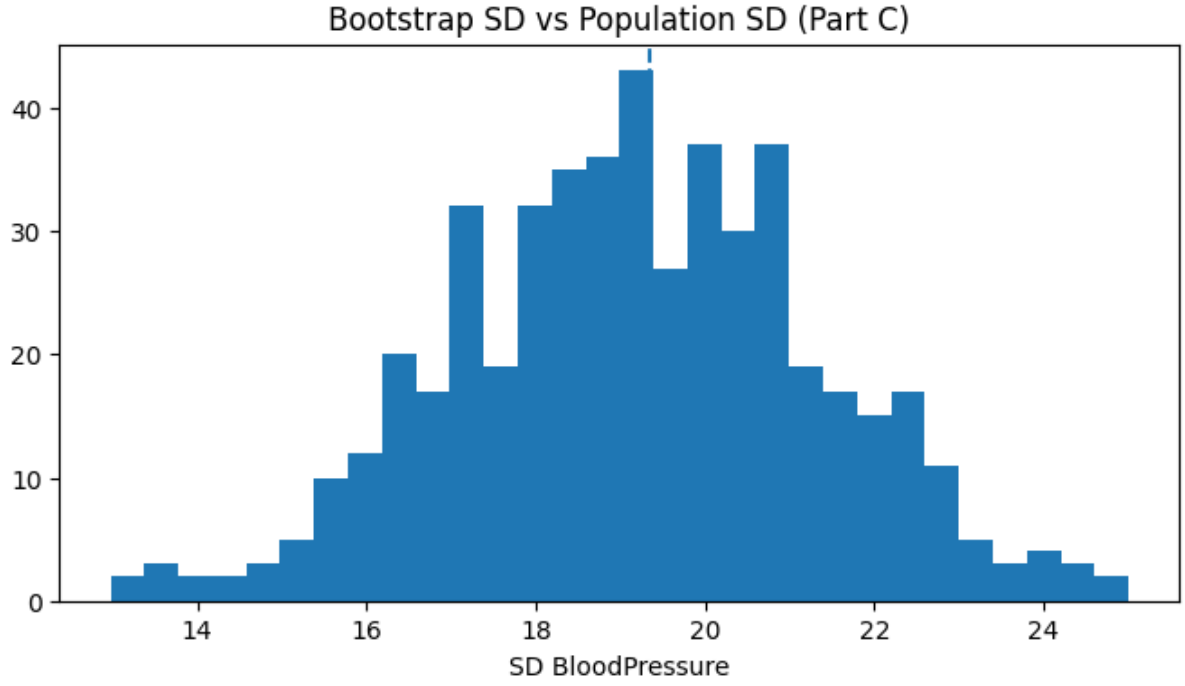


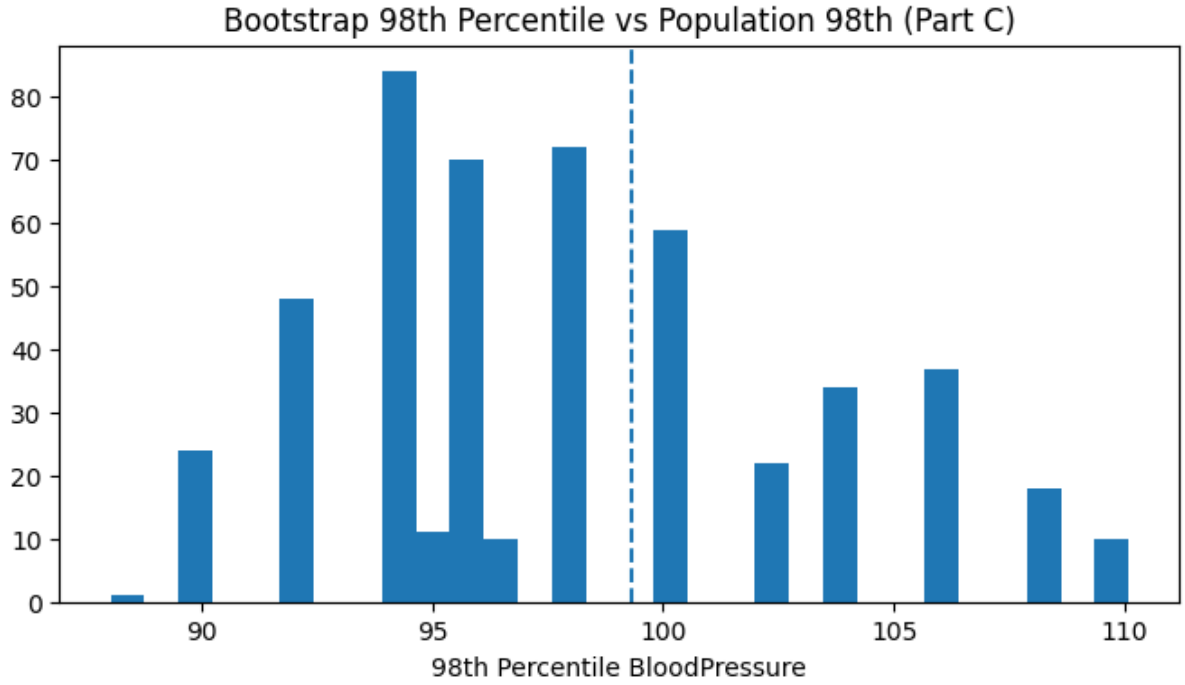
Interpretation:
This chart shows if the highest BMI values in the sample align with those in the whole dataset, which is important for assessing extremes.

Bootstrap Analysis of Blood Pressure:



SANJANA AILENI
PDS ASSIGNMNET 2
STUDENT ID :12649897





Interpretation:
The bootstrap histograms show that the most frequent (central) estimate from repeated sampling matches closely with the population value, showing low estimation bias.

Summary and Conclusion

- The random sample statistics are broadly similar to the full population for major health metrics, confirming effective sampling.

SANJANA AILENI
PDS ASSIGNMNET 2
STUDENT ID :12649897

- The comparative bar graphs clearly display differences and similarities, visually supporting your numerical findings.
- Bootstrap analysis adds reliability to the sample estimates, as shown by the agreement between bootstrap means and population statistics.
- The approach and visuals fulfill the assignment's goals of comparative analysis using both numerical and graphical evidence.