**Principles of Data Science**

**Assignment - 3**

**Task (a): Glucose Statistics (Sample vs Population)**

A random sample of 25 patients was taken to compare mean and maximumGlucose levels with those of the full population.

* **Sample Mean Glucose:** 116.64
* **Population Mean Glucose:** 120.89
* **Sample Max Glucose:** 183
* **Population Max Glucose:** 199

**Observation:**  
The sample values are slightly lower than the population values, which is expected due to the small sample size. The variation illustrates the effect of sampling variability in estimating population parameters.

**Task (b): 98th Percentile of BMI**

The 98th percentile of **BMI** was calculated for both the sample and the population.

* **Sample 98th Percentile BMI:** 40.25
* **Population 98th Percentile BMI:** 47.53

**Observation:**  
The 98th percentile in the sample is noticeably lower than that of the population. This is again due to the small sample size not capturing the most extreme values present in the full dataset.

**Task (c): Bootstrap Analysis of BloodPressure**

Using 500bootstrapsamples (n=150), we estimated average statistics for BloodPressure and compared them to the population values.

| **Statistic** | **Bootstrap Avg** | **Population** |
| --- | --- | --- |
| Mean | 69.18 | 69.11 |
| Standard Deviation | 19.08 | 19.36 |
| 98th Percentile | 97.92 | 99.32 |

**Observation:**  
The bootstrap averages are very close to the population statistics. This demonstrates the power of bootstrap sampling in approximating the population distribution — especially for the mean and standard deviation. The 98th percentile is slightly underestimated, likely due to variability in tail values.

**Overall Conclusion:**

* **Random sampling** can yield results close to the population but may miss extremes in small samples.
* **Bootstrap methods** provide reliable approximations of population statistics.
* **Visual comparisons** (charts) help highlight differences between samples and full data effectively.