Results

Name : Lohitha Bandarupalli

Student ID : 16355819

2) The data file diabetes.csv contains data of 768 patients. In this data there are 8 attributes (Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, and Age) and 1 response variable (Outcome). The response variable, Outcome, has binary value (1 indicating the outcome is diabetes and 0 means no diabetes). For this assignment purposes we will consider this data as a population. Use this data to perform the following:

a) set a seed (to ensure work reproducibility) and take a random sample of 25 observations and find the mean Glucose and highest Glucose values of this sample and compare these statistics with the population statistics of the same variable. You should use charts for this comparison.(5 points)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

From the above, the sample mean glucose level is slightly lower than the population mean, and the sample's maximum glucose level is also lower than the population's maximum which indicates some variation.

b) Find the 98th percentile of BMI of your sample and the population and compare the results using charts.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The 98th percentile of BMI in the sample is slightly lower than in the population, which suggests a slight variation in the distribution of extreme BMI values between the sample and the overall population.

c) Using bootstrap (replace= True), create 500 samples (of 150 observation each) from the

population and find the average mean, standard deviation and percentile for BloodPressure and compare this with these statistics from the population for the same variable. Again, you should create charts for this comparison. Report on your findings.

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

From the above, the population mean blood pressure is approximately 69.11 mmHg with a standard deviation of around 19.36 mmHg, which indicates the average and spread of blood pressure values within the population. The 95th percentile blood pressure is 90.0 mmHg, which indicates that 95% of the population has a blood pressure below this value.