Diabetes

Report

2023-04-12

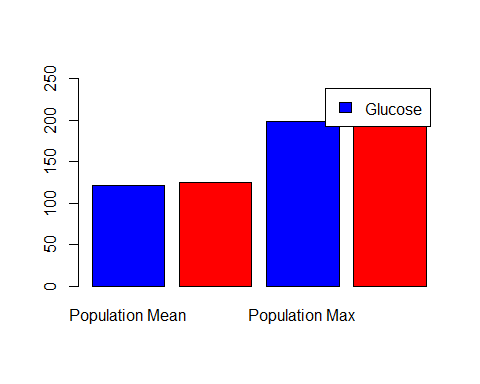
# Load data  
df <- read.csv('C:/Users/fvcds/Desktop/Project/diabetes.csv')

**Question#1(a)**

# Set seed and take a random sample of 25 observations  
set.seed(1234)  
sample <- df[sample(nrow(df), 25), ]

# Calculate mean and maximum glucose for sample and population  
pop\_mean\_glucose <- mean(df$Glucose)  
pop\_max\_glucose <- max(df$Glucose)  
  
sample\_mean\_glucose <- mean(sample$Glucose)  
sample\_max\_glucose <- max(sample$Glucose)  
  
# Create bar chart to compare mean and maximum glucose  
barplot(c(pop\_mean\_glucose, sample\_mean\_glucose, pop\_max\_glucose, sample\_max\_glucose),   
 names.arg = c('Population Mean', 'Sample Mean', 'Population Max', 'Sample Max'),   
 col = c('blue', 'red', 'blue', 'red'),   
 legend.text = 'Glucose',   
 ylim = c(0, 250))

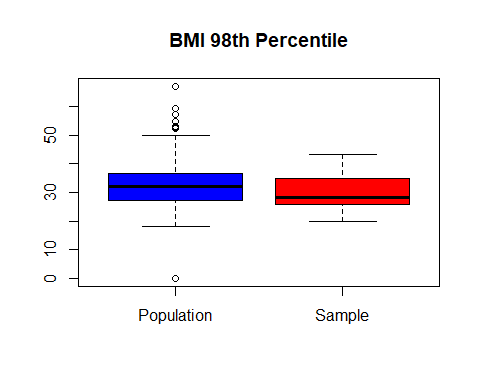
**Output**



**Question#1 (b)**

# Find the 98th percentile of BMI for sample and population  
pop\_98th\_percentile\_bmi <- quantile(df$BMI, 0.98)  
sample\_98th\_percentile\_bmi <- quantile(sample$BMI, 0.98)  
  
# Create boxplot to compare 98th percentile of BMI  
boxplot(df$BMI, sample$BMI,   
 names = c('Population', 'Sample'),   
 main = 'BMI 98th Percentile',   
 col = c('blue', 'red'))

**Output**



**Question#1 (c)**

# Create 500 bootstrap samples of 150 observations each and compare the statistics with the population statistics for BloodPressure  
n\_samples <- 500  
sample\_size <- 150  
  
bp\_pop\_mean <- mean(df$BloodPressure)  
bp\_pop\_sd <- sd(df$BloodPressure)  
bp\_pop\_25th\_percentile <- quantile(df$BloodPressure, 0.25)  
bp\_pop\_75th\_percentile <- quantile(df$BloodPressure, 0.75)  
  
bp\_sample\_means <- numeric(n\_samples)  
bp\_sample\_sds <- numeric(n\_samples)  
bp\_sample\_25th\_percentiles <- numeric(n\_samples)  
bp\_sample\_75th\_percentiles <- numeric(n\_samples)  
  
for (i in 1:n\_samples) {  
 sample <- sample(df$BloodPressure, size = sample\_size, replace = TRUE)  
 bp\_sample\_means[i] <- mean(sample)  
 bp\_sample\_sds[i] <- sd(sample)  
 bp\_sample\_25th\_percentiles[i] <- quantile(sample, 0.25)  
 bp\_sample\_75th\_percentiles[i] <- quantile(sample, 0.75)  
}

# Create plots to compare the statistics of the 500 bootstrap samples with the population statistics  
par(mfrow = c(2, 2))  
  
hist(bp\_sample\_means, main = 'Blood Pressure Mean', xlab = 'Mean', col = 'red')  
abline(v = bp\_pop\_mean, col = 'blue')  
  
hist(bp\_sample\_sds, main = 'Blood Pressure Standard Deviation', xlab = 'Standard Deviation', col = 'red')  
abline(v = bp\_pop\_sd, col = 'blue')  
  
hist(bp\_sample\_25th\_percentiles, main = 'Blood Pressure 25th Percentile', xlab = '25th Percentile', col = 'red')  
abline(v = bp\_pop\_25th\_percentile, col = 'blue')  
  
hist(bp\_sample\_75th\_percentiles, main = 'Blood Pressure 75th Percentile', xlab = '75th Percentile', col = 'red')  
abline(v = bp\_pop\_75th\_percentile, col = 'blue')

**Outputs**

