Problem 5: Giving Data the Boot

Rakesh Vishwabrahmana

Question 1

Solution:

The sample mean, median, standard deviation, first quartile, geometric mean, harmonic mean, and median absolute deviation of given measurements are given in the summary.

```
x <- c(207,202,169,211,191,212,108,92,186,203,126,184,206,177,164,53,190)
n <- length(x)
x_mean <- mean(x)
x_med <- median(x)
x_sd <- sd(x)
q1 <- function(x) as.vector(quantile(x, 0.25))
x_q1 <- q1(x)
gm <- function(x) exp(mean(log(x)))
x_gm <- gm(x)
hm <- function(x) 1/mean(1/x)
x_hm <- hm(x)
x_mad <- mad(x)
sum_stat <- rbind(x_mean,x_med,x_sd,x_q1,x_gm,x_hm,x_mad)
rownames(sum_stat) <- c("Mean","Median","Standard Deviation","First
Quartile",
                      "Geometric Mean","Harmonic Mean","Median Absolute
Deviation")
colnames(sum_stat) <- "Estimate"
sum_stat
```

```
##                            Estimate
## Mean                      169.47059
## Median                    186.00000
## Standard Deviation         46.87099
## First Quartile            164.00000
## Geometric Mean            160.65023
## Harmonic Mean             147.92154
## Median Absolute Deviation  29.65200
```

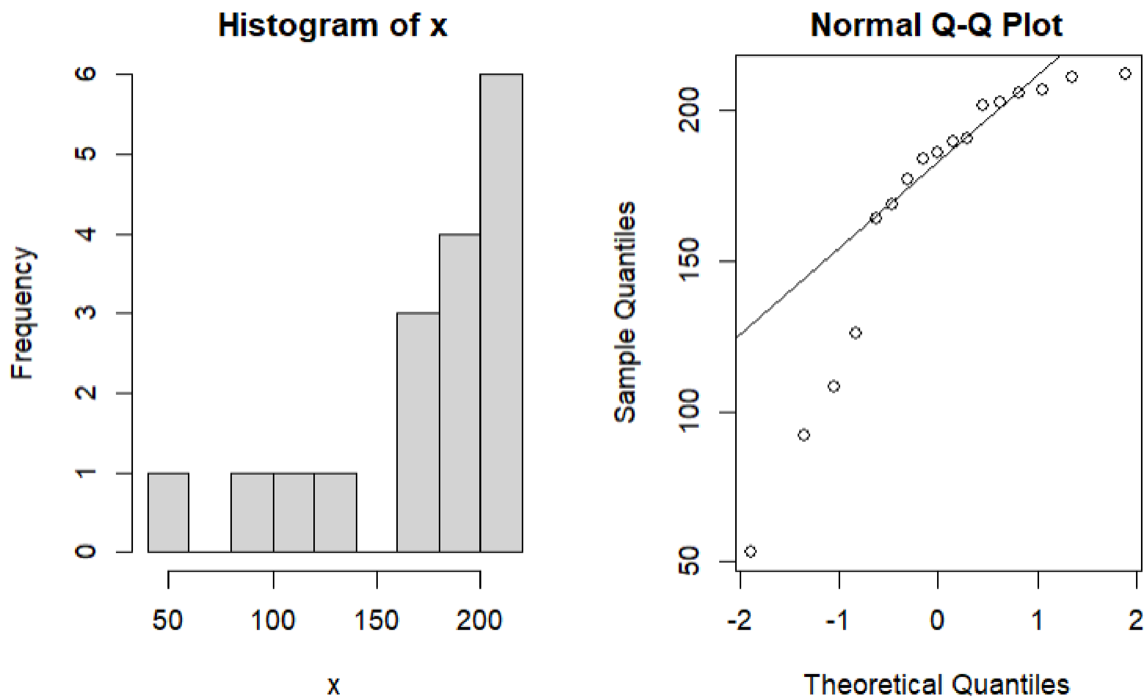
Question 2

Solution:

The histogram of the data and a normal quantile-quantile plot is given below.

```
par(mfrow=c(1,2))
hist(x)
```

```
qqnorm(x)
qqline(x)
```



Question 3

Solution:

In order to ascertain whether the given data show or not a serious deviation from normality, we used Shapiro-Wilk's method.

```
norm_test <- shapiro.test(x)
norm_test

##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.82029, p-value = 0.003903
```

The *p-value* of the test is 0.0039. Since $p_{value} < 0.05$ which implies that the data are significantly differ from normal distribution.

- We observed from the above figure, four data points (values <150) might be an outlier and therefore resulting distribution of the given data deviates from normal.

Question 4

Solution:

The 95% confidence intervals on the population mean, median,standard deviation, first quartile, geometric mean, harmonic mean, and median absolute deviation using a Studentized bootstrap is given Table 2.

```
boot_CI <- function(x, stat_fun, n_boot, percnt_CI) {
stat <- stat_fun(x)
n <- length(x)
boot_x <- sample(x, n*n_boot, replace = TRUE)
boot_mat <- matrix(boot_x, nrow = n, ncol = n_boot)
boot_stat <- apply(boot_mat, 2, stat_fun)
del_star <- boot_stat - stat
pct_diff <- (1 - percnt_CI) / 2
del <- as.vector(quantile(del_star, c(pct_diff, 1-pct_diff)))
ci <- stat - c(del[2], del[1])
return(ci)
}
boot_mean <- boot_CI(x, stat_fun = mean, n_boot = 1000, percnt_CI = 0.95)
boot_med <- boot_CI(x, stat_fun = median, n_boot = 1000, percnt_CI = 0.95)
boot_sd <- boot_CI(x, stat_fun = sd, n_boot = 1000, percnt_CI = 0.95)
boot_q1 <- boot_CI(x, stat_fun = q1, n_boot = 1000, percnt_CI = 0.95)
boot_gm <- boot_CI(x, stat_fun = gm, n_boot = 1000, percnt_CI = 0.95)
boot_hm <- boot_CI(x, stat_fun = hm, n_boot = 1000, percnt_CI = 0.95)
boot_mad <- boot_CI(x, stat_fun = mad, n_boot = 1000, percnt_CI = 0.95)
pop_ci <- rbind(boot_mean, boot_med, boot_sd, boot_q1, boot_gm, boot_hm,
boot_mad)
colnames(pop_ci) <- c("$L_{CI}$","$U_{CI}$")
rownames(pop_ci) <- c("Mean","Median","Standard Deviation","First Quartile",
                      "Geometric Mean","Harmonic Mean","Median Absolute
Deviation")
pop_ci

##                             $L_{CI}$  $U_{CI}$
## Mean                       148.76471 192.23824
## Median                     169.00000 208.00000
## Standard Deviation          32.87222  69.74516
## First Quartile             142.00000 236.00000
## Geometric Mean             134.66805 188.83432
## Harmonic Mean              111.22658 179.43505
## Median Absolute Deviation   -2.96520  48.96286
```