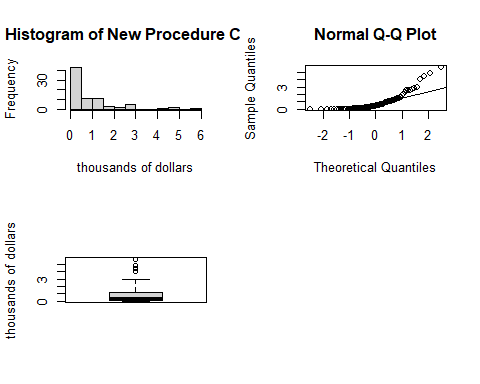
Question 1

Joseph Froelicher

10/15/2020

# Part A

new <- procedure[procedure$Procedure == 2,]  
  
par(mfrow = c(2,2))  
  
hist(new$Cost, main = "Histogram of New Procedure Cost", breaks = 12, xlab = "thousands of dollars")  
qqnorm(new$Cost)  
qqline(new$Cost)  
boxplot(new$Cost, ylab = "thousands of dollars")



# Part B

Our data appear to not be normally distributed. On the Normal QQ plot we can see deviations from the normal quantile line. Additionally, we can see from the histogram there is a right skew to our data.

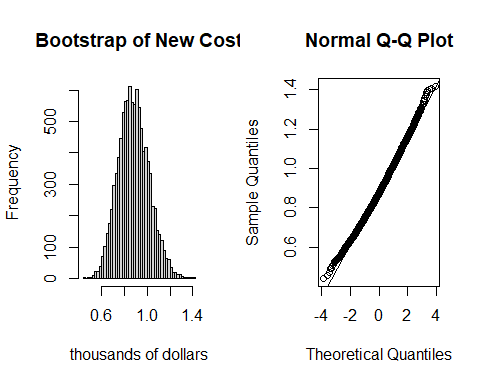
# Part C

mean <- mean(new$Cost)  
min <- min(new$Cost)  
first\_quarter <- quantile(new$Cost, 0.25)  
median <- median(new$Cost)   
third\_quarter <- quantile(new$Cost, 0.75)  
max <- max(new$Cost)  
sd <- sd(new$Cost)  
var <- var(new$Cost)

Our data range from 0 to 5.75, with an inter-quartile range of 0.04 to 1.2075, a median of 0.385, a mean of 0.881625, and a standard deviation of 1.2102416.

# Part D

n <- dim(new)[1]  
B <- 10000  
  
boot\_vec <- vector("double", B)  
  
set.seed(8675309)  
  
for (i in 1:B) {  
 val <- sample(new$Cost, n, replace = TRUE)  
 boot\_vec[i] <- mean(val)  
}  
  
par(mfrow = c(1, 2))  
hist(boot\_vec, main = "Bootstrap of New Cost", breaks = 50, xlab = 'thousands of dollars')  
qqnorm(boot\_vec)  
qqline(boot\_vec)



# Part E

This data is generally bell-shaped as seen in the histogram, with slight deviations from the normal quantile line near the tails on the normal QQ plot.

# Part F

boot\_mean <- mean(boot\_vec)  
boot\_bias <- boot\_mean - mean  
boot\_se <- sd(boot\_vec)

The boostrap data had an estimated mean of 0.8818489, an estimated standard error of 0.1334689, and an estimated bias of 2.2387510^{-4}.

# Part G

alpha = 0.05  
  
normal\_lower <- boot\_mean - qnorm(1 - (alpha / 2)) \* boot\_se  
normal\_upper <- boot\_mean + qnorm(1 - (alpha / 2)) \* boot\_se  
  
coverage\_lower <- sum(boot\_vec < normal\_lower) / B  
coverage\_upper <- sum(boot\_vec > normal\_upper) / B  
  
boot\_lower <- quantile(boot\_vec, 0.025)  
boot\_upper <- quantile(boot\_vec, 0.975)  
  
accuracy <- boot\_bias / boot\_se

Based on our estimates of coverage [0.0171, 0.0302], the 95% normal percentile estimates [0.6202546, 1.1434432] are too low for both the lower and upper bounds. The lower bound has coverage of 1.71%, and the upper bound has coverage of 3.02%, these are not close to our target of 2.5%. The 95% bootstrap confidence interval is [0.636875, 1.1561281]. The accuracy of the bootstrap confidence interval is 0.0016774, which is less than the suggested cutoff of 0.10, indicating good accuracy.