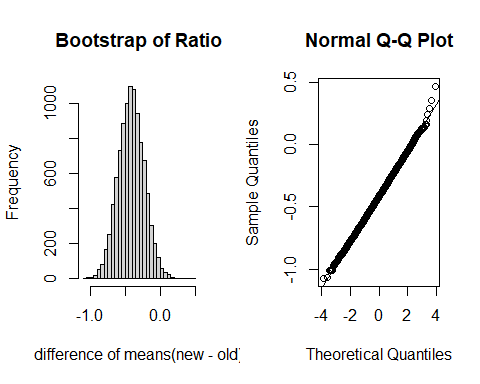
Question 2

Joseph Froelicher

10/15/2020

# Part A

n1 <- dim(new)[1]  
n2 <- dim(old)[1]  
  
B <- 10000  
  
boot\_ratio <- vector("double", B)  
  
set.seed(8675309)  
  
for (i in 1:B) {  
 val1 <- sample(new$Cost, n1, replace = TRUE)  
 val2 <- sample(old$Cost, n2, replace = TRUE)  
   
 boot\_ratio[i] <- mean(val1) - mean(val2)  
}  
  
par(mfrow = c(1, 2))  
hist(boot\_ratio, main = "Bootstrap of Ratio", breaks = 50, xlab = 'difference of means(new - old)')  
qqnorm(boot\_ratio)  
qqline(boot\_ratio)



# Part B

The histogram of the bootstrap data seem to be bell-shaped, with very few outliers from the normal quantile line on the normal QQ plot.

# Part C

boot\_mean <- mean(boot\_ratio)  
boot\_bias <- boot\_mean - (mean(new$Cost) - mean(old$Cost))  
boot\_se <- sd(boot\_ratio)

# Part D

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\_ratio < normal\_lower) / B  
coverage\_upper <- sum(boot\_ratio > normal\_upper) / B  
  
boot\_lower <- quantile(boot\_ratio, 0.025)  
boot\_upper <- quantile(boot\_ratio, 0.975)  
  
accuracy <- boot\_bias / boot\_se

Based on our estimates of coverage [0.0235, 0.0254], the 95% normal percentile estimates [-0.7729894, -0.0540617] are pretty good for both the lower and upper bounds. The lower bound has coverage of 2.35%, and the upper bound has coverage of 2.54%, these are close to our target of 2.5%. The 95% bootstrap confidence interval is [-0.768501, -0.0536615]. The accuracy of the bootstrap confidence interval is -0.0103628, which is less than the suggested cutoff of 0.10, indicating good accuracy.