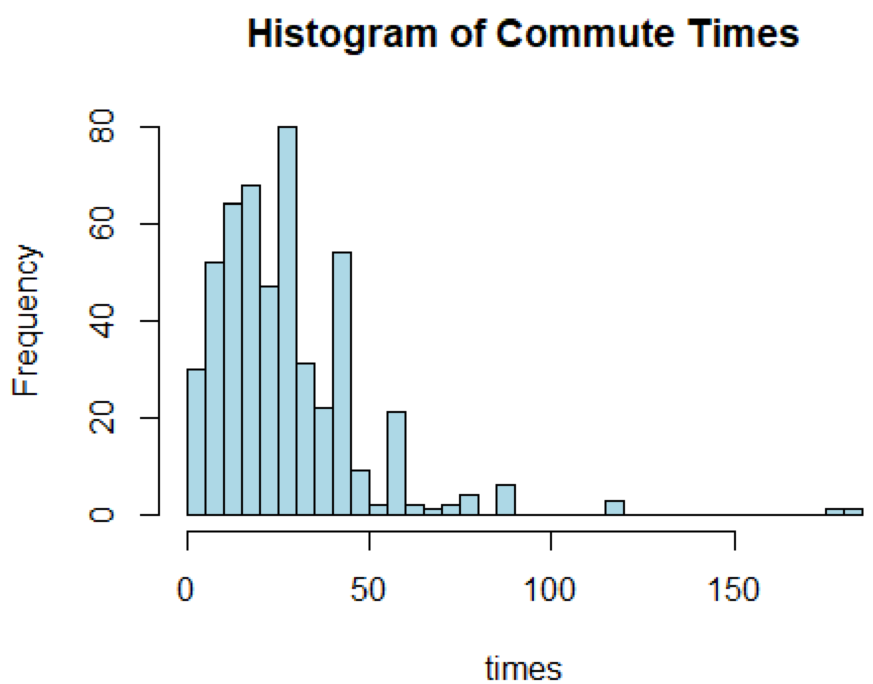
HW3_567

Matthew Stoebe

2024-11-06

#a

```
data(CommuteAtlanta)
times <- CommuteAtlanta$Time

hist(times, breaks = 30, col = "lightblue", main = "Histogram of Commute
Times")
```



#b

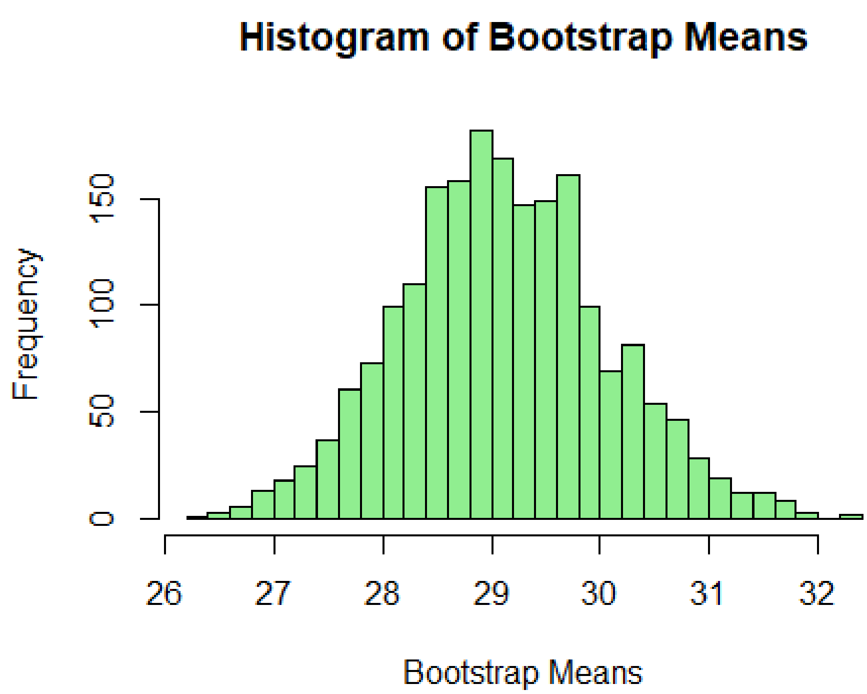
```
## population mean commute time: 29.11
```

#c

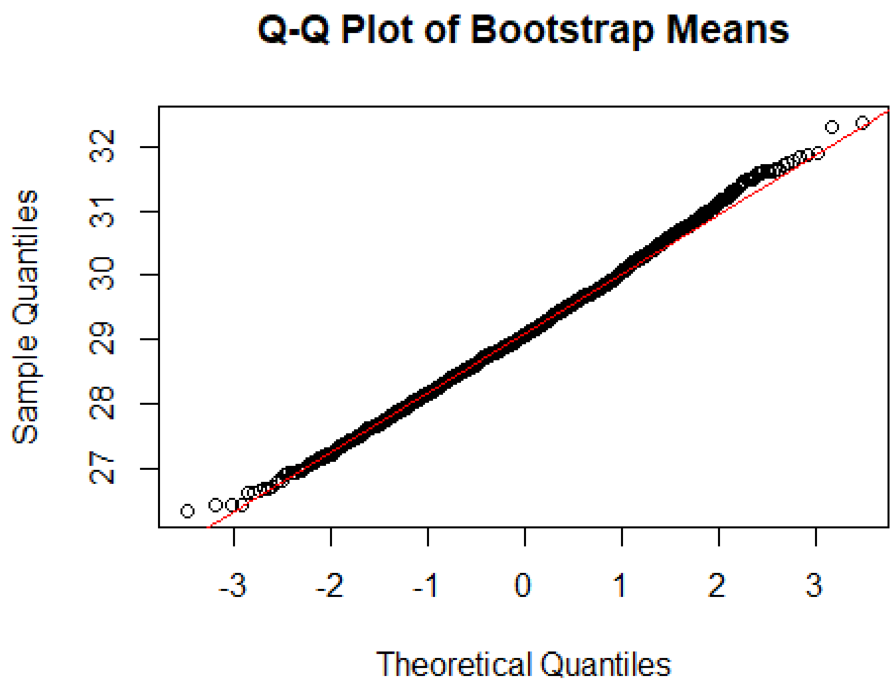
```
t_test_result <- t.test(times)

print(t_test_result$conf.int)

## [1] 27.28958 30.93042
## attr(,"conf.level")
## [1] 0.95
```

#d

```
bootstrap_mean <- function(data, R) {
  set.seed(12345)
  n <- length(data)
  boot_means <- numeric(R)
  for (i in 1:R) {
    sample_i <- sample(data, n, replace = TRUE)
    boot_means[i] <- mean(sample_i)
  }
  return(boot_means)
}

# Run the bootstrap
R <- 2000
boot_means <- bootstrap_mean(times, R)

# Histogram of bootstrap means
hist(boot_means, breaks = 30, col = "lightgreen", main = "Histogram of 
Bootstrap Means", xlab = "Bootstrap Means")
```




```
# Q-Q plot of bootstrap means
qqnorm(boot_means, main = "Q-Q Plot of Bootstrap Means")
qqline(boot_means, col = "red")
```



#e

The histogram of means does appear to be bell shaped and symmetric. This makes sense as we know that according to the central limit theorem, the sampling distribution of the sample mean will be approximately normal when the sample size is large enough. With our sample size of 2000, the CLT should hold.

#f

```
mean_orig <- mean(times)
mean_bootstrap <- mean(boot_means)

bias <- mean_bootstrap - mean_orig
variance <- var(boot_means)
MSE <- variance + bias^2


cat("bias: ", bias, "\n")

## bias:  -0.004001

cat("variance: ", variance, "\n")

## variance:  0.9052314

cat("MSE: ", MSE, "\n")

## MSE:  0.9052474
```

#g

```
# Calculate the percentile confidence interval
alpha <- 0.05
CI_percentile <- quantile(boot_means, probs = c(alpha/2, 1 - alpha/2))
CI_percentile

##     2.5%    97.5%
## 27.25535 31.05810
```

#h

```
# Calculate the basic confidence interval
CI_basic_lower <- 2 * mean_orig - quantile(boot_means, probs = 1 - alpha/2)
CI_basic_upper <- 2 * mean_orig - quantile(boot_means, probs = alpha/2)
CI_basic <- c(CI_basic_lower, CI_basic_upper)
CI_basic

##    97.5%     2.5%
## 27.16190 30.96465
```

#i

```
n <- length(times)
mean_orig <- mean(times)
sd_orig <- sd(times)
SE_orig <- sd_orig / sqrt(n)

set.seed(12345)  # For reproducibility
R <- 2000
t_boot <- numeric(R)

for (i in 1:R) {
  sample_i <- sample(times, n, replace = TRUE)
  mean_i <- mean(sample_i)
  sd_i <- sd(sample_i)
  SE_i <- sd_i / sqrt(n)
  t_boot[i] <- (mean_i - mean_orig) / SE_i
}

alpha <- 0.05
t_lower <- quantile(t_boot, alpha/2)
t_upper <- quantile(t_boot, 1 - alpha/2)

CI_lower <- mean_orig - t_upper * SE_orig
CI_upper <- mean_orig - t_lower * SE_orig
CI_studentized <- c(CI_lower, CI_upper)
CI_studentized

##    97.5%     2.5%
## 27.35427 31.22359
```