Tian Qiu

STAT 350

Lab 6

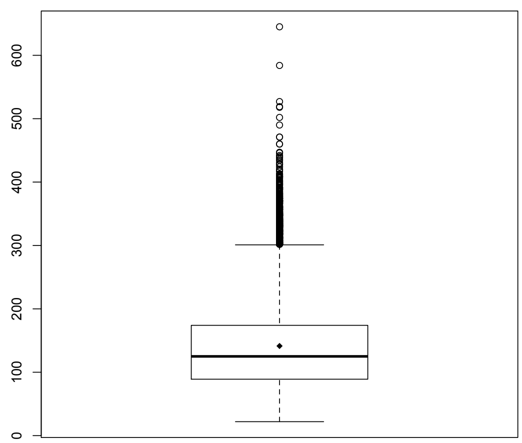
Part A

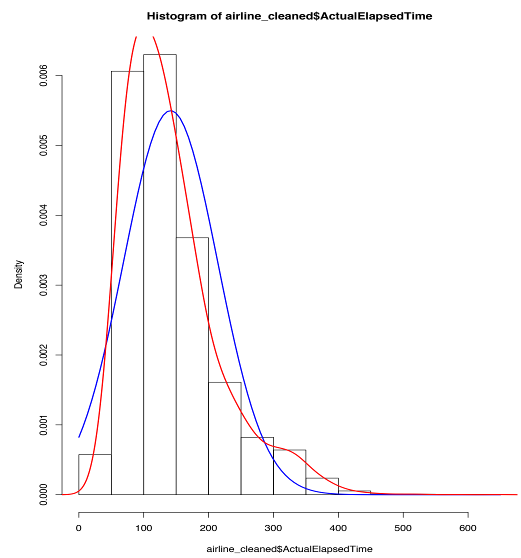
1.

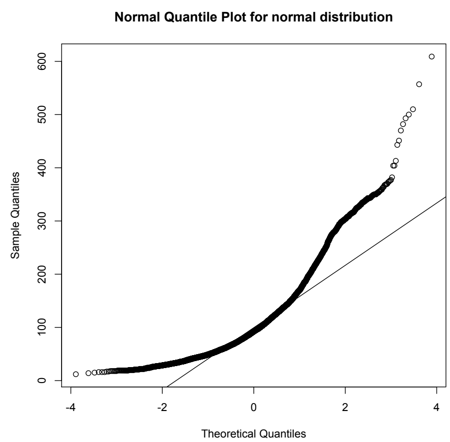
```
setwd("~/Desktop/Purdue/STAT350_R/STAT350/Labs/Lab6")

airline_cleaned                                                    <-

read.delim("~/Desktop/Purdue/STAT350_R/STAT350/Labs/Lab6/airline_cleaned.txt")


attach(airline_cleaned)


ActualElapsedTime          =          log(airline_cleaned$ActualElapsedTime)

# log the data!!!!

airline_cleaned$ActualElapsedTime = log(airline_cleaned$ActualElapsedTime)

mean(ActualElapsedTime)

sd(ActualElapsedTime)

mean(airline_cleaned$ActualElapsedTime)

sd(airline_cleaned$ActualElapsedTime)



print(airline_cleaned$ActualElapsedTime)
```

```
quartz()

hist(airline_cleaned$ActualElapsedTime,freq = FALSE)                    # frequency should
be false to get density curve.....

means <- mean(airline_cleaned$ActualElapsedTime)

print(means)

std <- sd(airline_cleaned$ActualElapsedTime)

curve(dnorm(x, mean=means, sd=std), col="blue", lwd=2, add=TRUE)      # normal
distribution line

lines(density(airline_cleaned$ActualElapsedTime, adjust=2),col = "red", lwd=2)

mean(airline_cleaned$ActualElapsedTime)

sd(airline_cleaned$ActualElapsedTime)


# boxplot

quartz()

boxplot(airline_cleaned$ActualElapsedTime)

points(means, pch = 18)


#    b) Make a Normal quantile plot to confirm that there are no systematic departures
from Normality.

quartz()
```

2.

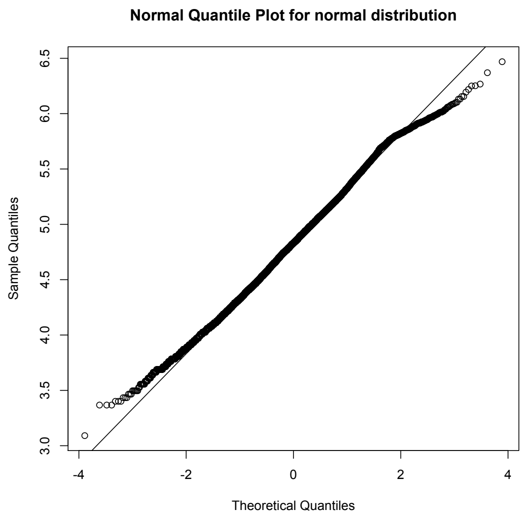


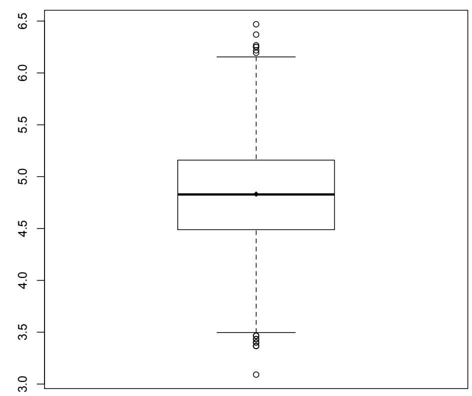


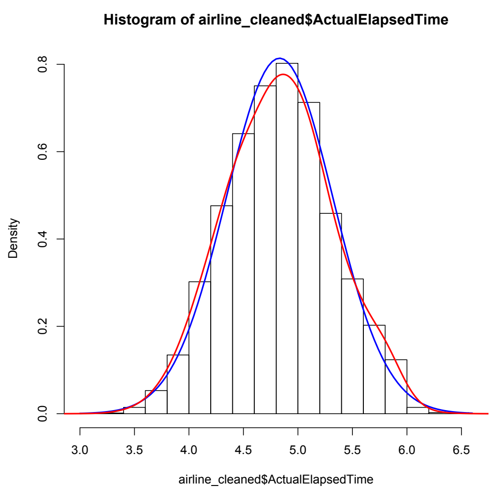


These data are not normally distributed. In the histogram you can see that it is skewed to the right. The normal quantile plot is also not linear. There are a few outliers lying at the upper part of the data (you can find them in both boxplots and the normal quantile plot.

3.







4. It is appropriate to use the t procedure right now because the data are distributed quite normally.

5. Mean = 4.831534 Std = 0.4901595 n = 9997 df = n-1 =9996

t = 2.576

margin of error = t*Std/sqrt(n) = 0.012627403

CI=(mean-MarginofError, mean+MarginofError)=(4.8189, 4.8442)

6. 99 percent confidence interval: (4.818904 4.844164). It is totally the same as I calculated in part 5.

7. We are 99% confident that the true logged mean of actual elapsed time is between 4.818904 and 4.844164.

8.Since our confidence interval contains 4.83, we fail to reject the null hypothesis at alpha = 0.01. We would reject the null hypothesis time = 125.21 since log(125.21) = 2.097, which is not included in our CI.

Part B

1.

```
setwd("~/Desktop/Purdue/STAT350_R/STAT350/Labs/Lab6")
airline_cleaned <-
read.delim("~/Desktop/Purdue/STAT350_R/STAT350/Labs/Lab6/airline_cleaned.
```

```
txt")

attach(airline_cleaned)
AirTime = log(airline_cleaned$AirTime)
# log the data!!!!
airline_cleaned$AirTime = log(airline_cleaned$AirTime)
mean(AirTime)
sd(AirTime)
mean(airline_cleaned$AirTime)
sd(airline_cleaned$AirTime)



print(airline_cleaned$AirTime)
quartz()
hist(airline_cleaned$AirTime,freq = FALSE)         # frequency should be
false to get density curve.....
means <- mean(airline_cleaned$AirTime)
print(means)
std <- sd(airline_cleaned$AirTime)
curve(dnorm(x, mean=means, sd=std), col="blue", lwd=2, add=TRUE)    #
normal distribution line
```

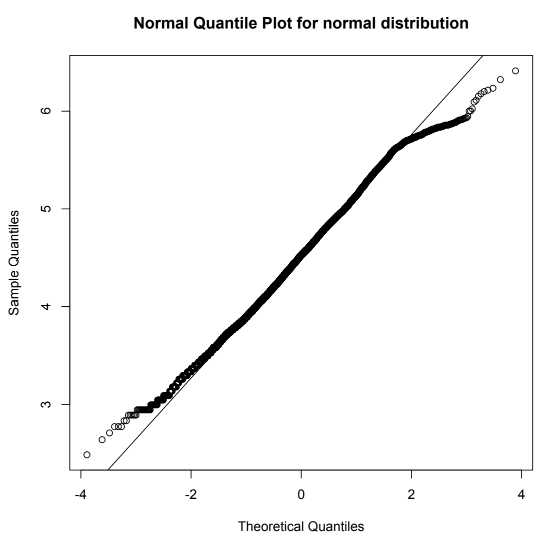
```
lines(density(airline_cleaned$AirTime, adjust=2),col = "red", lwd=2)

mean(airline_cleaned$AirTime)

sd(airline_cleaned$AirTime)



# boxplot

quartz()

boxplot(airline_cleaned$AirTime)

points(means, pch = 18)



# b) Make a Normal quantile plot to confirm that there are no systematic departures from Normality.

quartz()

qqnorm(airline_cleaned$AirTime,main="Normal Quantile Plot for normal distribution")

qqline(airline_cleaned$AirTime)


t.test(airline_cleaned$AirTime, conf.level=0.95, alternative = "two.sided")

library(TeachingDemos)
```

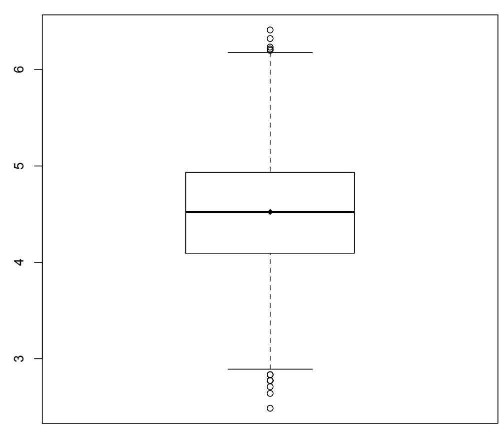
stdev = sd(airline_cleaned$AirTime)

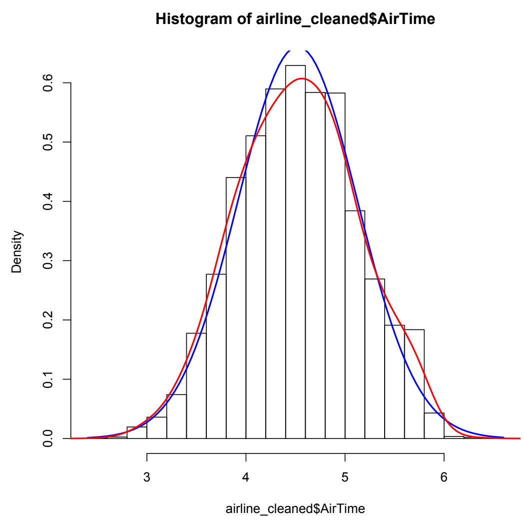
stdev

z.test(airline_cleaned$AirTime, conf.level = 0.95, alternative="two.sided", sd=stdev)

2. From the histogram you can see that it is totally bell-shaped with a really nice normal curve fit. Also the normal quantile plot is approximately linear and the boxplot is symmetric and centered at medium. There is few outlier. Hence, we can conclude that the data is quite normally distributed.







3. It is appropriate to use t procedures since data are normally distributed and we have a really large sample size.
4. Mean = 4.5224 Std = 0.6035763 n =9997 SE = Std/sqrt(n) = 0.0060367
5. t = 1.96 Lower bound = 4.5224 – 1.96*0.0060367 =4.51057

   We are 95% confident that the true log air time is larger than 4.51057.
6. H0: $\mu = 4.1$ Ha: $\mu > 4.1$

   t* = (4.5224-4.1)/0.0060367 = 69.972

   alpha = 0.05 df = 9996 t = 1.645 < 69.972

   Hence we have strong evidence against the null hypothesis.
7. H0: $\mu = 4.8$ Ha: $\mu > 4.8$

   t* = (4.5224-4.8)/0.0060367 = -45.985

alpha = 0.05 df = 9996 t = 1.645 > -45.985

Hence we fail to reject the null hypothesis.

8. They all claim that with a significance level of 5%, the true log time is greater than 4.1 but may not larger than 4.8. However, in part 5 it gives us a more accurate interval.