```
setwd("/Users/duyvo/Desktop/WW Fall 21/Statistics and
probability/Project")

## Use readxl to read excel file
library(readxl)

## import excel data to cs_dataset
cs_dataset=read_excel("Computer_Stats.xlsx")

## check dataset
head(cs_dataset)

##remove cols 3 and 4
cs_dataset <- cs_dataset[,c(-3,-4)]

##remove rows 1 and 2
cs_dataset <- cs_dataset[c(-1,-2),]

# set cols name
colnames(cs_dataset) <- c("AC_2016", "PP_2016", "AC_2017",
"PP_2017")
cs_dataset

summary(cs_dataset)

## convert char to numeric
cs_dataset$AC_2016 <- as.numeric(cs_dataset$AC_2016)
cs_dataset$PP_2016 <- as.numeric(cs_dataset$PP_2016)
cs_dataset$AC_2017 <- as.numeric(cs_dataset$AC_2017)
cs_dataset$PP_2017 <- as.numeric(cs_dataset$PP_2017)

##-----------------------------##
## Question 1
## mean
mean_AC2017 <- mean(cs_dataset$AC_2017)
mean_PP2017 <- mean(cs_dataset$PP_2017)

## median
median_AC2017 <- median(cs_dataset$AC_2017)
median_PP2017 <- median(cs_dataset$PP_2017)

## mode
```

```
mode(cs_dataset$AC_2017)
mode(cs_dataset$PP_2017)

## standara deviation
sd_AC2017 <- sd(cs_dataset$AC_2017)
sd_PP2017 <- sd(cs_dataset$PP_2017)

## Histogram frequency table
par(mfrow = c(1,2))

value_AC2017 <- cs_dataset$AC_2017
value_PP2017 <- cs_dataset$PP_2017
hist(value_AC2017, main = "AC_2017", col = "blue", xlab =
"Assembly cost")
hist(value_PP2017, main = "PP_2017", col = "red", xlab =
"percent of profit")

## boxplot of AC_2017
par(mfrow = c(1,1))
boxplot(value_AC2017, col = " dark green")

## percentage of 36 days AC within +- 1 standard deviation
counter1 <- 0
for (i in value_AC2017){
  if(i > (mean_AC2017 - sd_AC2017) & i < (mean_AC2017 +
sd_AC2017)){
    counter1 = counter1 + 1
  }
}

percentage_AC2017_1d <- counter1/NROW(value_AC2017)*100

## proportion of 36 days PP are 35.4% or more
counter2 <- 0 # hold how many days PP >= 35.4%
for (i in value_PP2017){
  if(i >= 35.4){
    counter2 = counter2 + 1
  }
}

library(MASS)
fractions(counter2/NROW(value_PP2017)) ## actual proportion
```

```
using fraction
proportion_PP2017_1e <- counter2/NROW(value_PP2017)

## percentage PP_2017 >= 35.4% and AC_2017 <= $330
counter3 <- 0
for(i in 1:NROW(value_AC2017)){
  if(value_PP2017[i] >= 35.4 & value_AC2017[i] <= 330){
   counter3 = counter3 + 1
  }
}

percentage_AC2017_1f = (counter3/counter2)*100


##-----------------------------##
## Question 2

## mean , standard deviation
mean_AC2016 <- mean(cs_dataset$AC_2016)
sd_AC2016 <- sd(cs_dataset$AC_2016)

value_AC2016 <- cs_dataset$AC_2016
value_PP2016 <- cs_dataset$PP_2016

counter4 <- 0
for (i in value_AC2016){
  if(i >= 320 & i <= 350){
    counter4 = counter4 + 1
  }
}

fractions(counter4/NROW(value_AC2016)) ## actual proportion
using fraction
proportion_AC2016_2a <- counter4/NROW(value_AC2016)

## scatterplot for 2016
plot(value_AC2016,value_PP2016, main = "Scatterplot of AC
and PP 2016", xlab = "Assembly cost", ylab = "percent of
profit")

## trend line
abline(lm(value_PP2016~value_AC2016),col="red")
```

```
## equation
summary(lm(value_PP2016~value_AC2016)) # y = 27.42308 +
0.02516*x

## percent of profit when assembly cost = 340
percentOfprofit_2d = round(27.42308+0.02516*340)

## assembly cost when percent of profit = 34.7
assemblyCost_2d = round((34.7-27.42308)/0.02516)

## 2e
qqnorm(value_AC2016, ylab = "Assembly cost", col="dark
green")
qqline(value_AC2016, col= "red")

?qqnorm
##------------------------------##
## Question 3

##
# AC_2017 ~ N(335.625,16.69877/sqrt(36))
# n = 36 > 30
# Z-alpha/2 = 1.960
# (335.625-1.960*(16.69877/sqrt(36)), 335.625+1.960*
(16.69877/sqrt(36)))

X <- 335.625-1.960*(16.69877/sqrt(36))
Y <- 335.625+1.960*(16.69877/sqrt(36))

##------------------------------##
## Question 4

# yes, it changed

# null Hypothesis - Ho : mean = 328.7014
# alternative Hypothesis - Ha : mean # 328.7014

# Xbar = mean_AC2016 = 328.7014
# mean = mean_AC2017 = 335.625
# sd = sd_2016 = 22.04782
# n = 36
```

```
# Test Statistic
# Z-Ho = (328.7014 - 335.625)/(22.04782/sqrt(36))

# 2 tail test
# critical Value
# alpha = 0.05
# Z-alpha/2 = 1.960

Z_nullHypo <- (328.7014 - 335.625)/(22.04782/sqrt(36)) #
-1.884159

# conclusion
# (-Z-alpha/2 < Z_nullHypo < Z-alpha/2)
# (-1.960 < -1.884159 < 1.960) (True)
# fail to reject null hypothesis

# p-value = [1-P(Z <= 1.88)]*2
p_value = (1 - pnorm(1.88,0,1))*2 # 0.06010808

# p_value = 0.06 > alpha = 0.05
# => also fail to reject null hypothesis

## there has not been a significant statistical change in
the mean AC since 2014
## as you can see the mean of 36 days AC in 2017 doesn't
have a big difference from the mean of total 36 months AC
since 2014-2016
## as well as the the mean of each month since 2014-2016
```