```
#Assginment 2

#a
install.packages("readxl")
library(readxl)
setwd("C:/Users/yijia/Desktop/Courses/Core Statistics for MS Students Using R/
assign/assign2")
df <- read_excel("DeliveryTimes.xlsx")
mean <- mean(df$DeliveryTime) = 45.106
sd <- sd(df$DeliveryTime) = 2.470
var <- var(df$DeliveryTime) = 6.100

#b
#If the delivery time is X~U[30,60]
#mean is (30+60)/2 = 45, approximately the same as the sample mean 45.11
#Variance is (60-30)^2/12 = 75, which is very different from sample variance
6.10
#So the findings do not support managers' claim

#c
#Under the normal distribution assumption
#--95% Prediction interval for the whole population distribution of delivery
times

PI_LB_95 <- qnorm(0.025, mean = 45.106, sd = 2.470)
PI_UP_95 <- qnorm(0.975, mean = 45.106, sd = 2.470)
PI_95 <- c(40.265, 49.947)

#--80% Prediction interval for the whole population distribution of delivery
times

PI_LB_80 <- qnorm(0.1, mean = 45.106, sd = 2.470)
PI_UP_80 <- qnorm(0.9, mean = 45.106, sd = 2.470)
PI_80 <- c(41.941, 48.271)

#The 80% PI is narrower than 95% PI


#Under the uniform distribution assumption
#--95%
PI_LB_95_U <- qunif(0.025, min = 30, max = 60)
PI_UP_95_U <- qunif(0.975, min = 30, max = 60)
PI_95_U <- c(30.75, 59.25)

#--80%
PI_LB_80_U <- qunif(0.1, min = 30, max = 60)
PI_UP_80_U <- qunif(0.9, min = 30, max = 60)
PI_80_U <- c(33.00, 57.00)


#d
#delivered in 45 minutes or less <- P(X <= 45) = F(45)
#normal
pnorm(45, mean = 45.106, sd = 2.470) = 0.483
#uniform
punif(45, min = 30, max = 60) = 0.5
```

```
#delivered in 40 minutes or less <- P(X <= 40) = F(40)
#normal
pnorm(40, mean = 45.106, sd = 2.470) = 0.019
#uniform
punif(40, min = 30, max = 60) = 0.333

#delivered in 50 minutes or more <- P(X >= 50) = 1 - F(50)
#normal
1 - pnorm(50, mean = 45.106, sd = 2.470) = 0.024
#uniform
1 - punif(50, min = 30, max = 60) = 0.333


#e
#Yes, it matters. The probability is very different under different
distribution assumptions.

#f
n = nrow(df)
std_err <- sd/sqrt(n)
CI_LB_95 <- mean - 1.96*std_err = 44.977
CI_UP_95 <- mean + 1.96*std_err = 45.235
CI_95 <- c(44.977,45.235)
#They are different, since CI is constructed for the population mean, while PI
is for the
#whole population.

#g
#No. No matter what the population distribution is, the sample mean still
follows normal
#distribution by CLT, and we can still construct CIs by using the same
formula, so the CI will
#not change.

#h
#1% level
t_value <- (mean-50)/std_err = -74.142
#|t_value| > 2.576, so the null at 1% level can be rejected. The average
delivery time is
#significantly different from 50 mins.

#5% level
t_value <- (mean-50)/std_err = -74.142
#|t_value| > 1.96, so the null at 5% level can be rejected. The average
delivery time is
#significantly different from 50 mins.
```