535_HW2

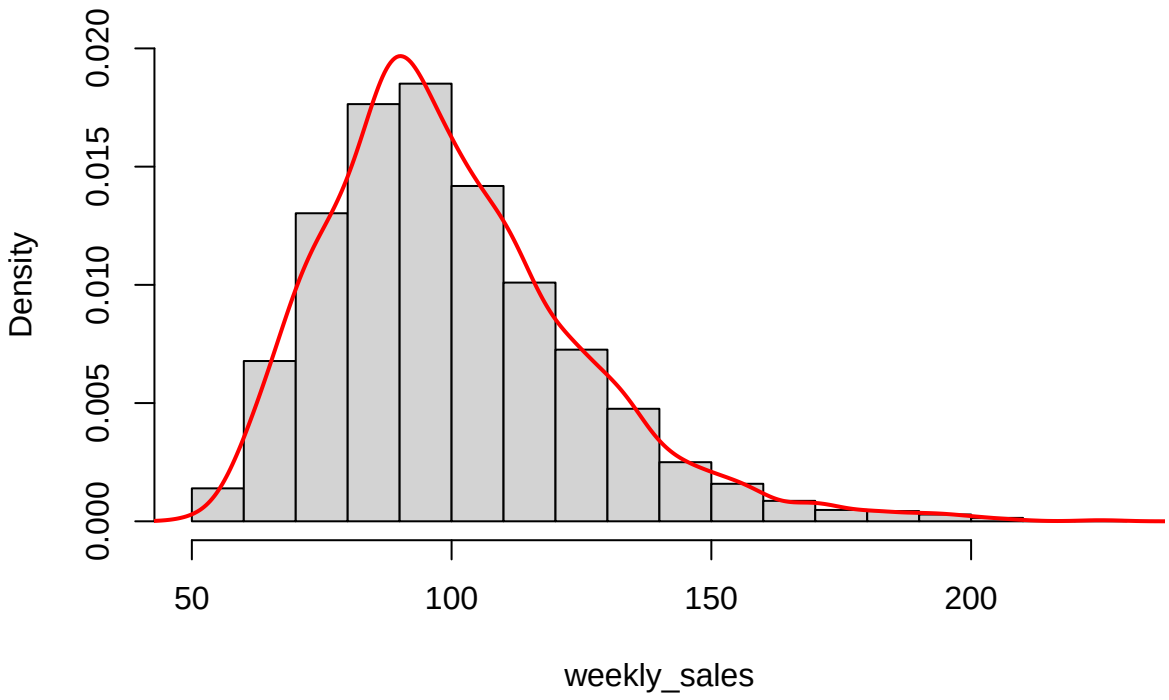
Shijie Gao, USC ID:6037-6293-25

2023-01-18

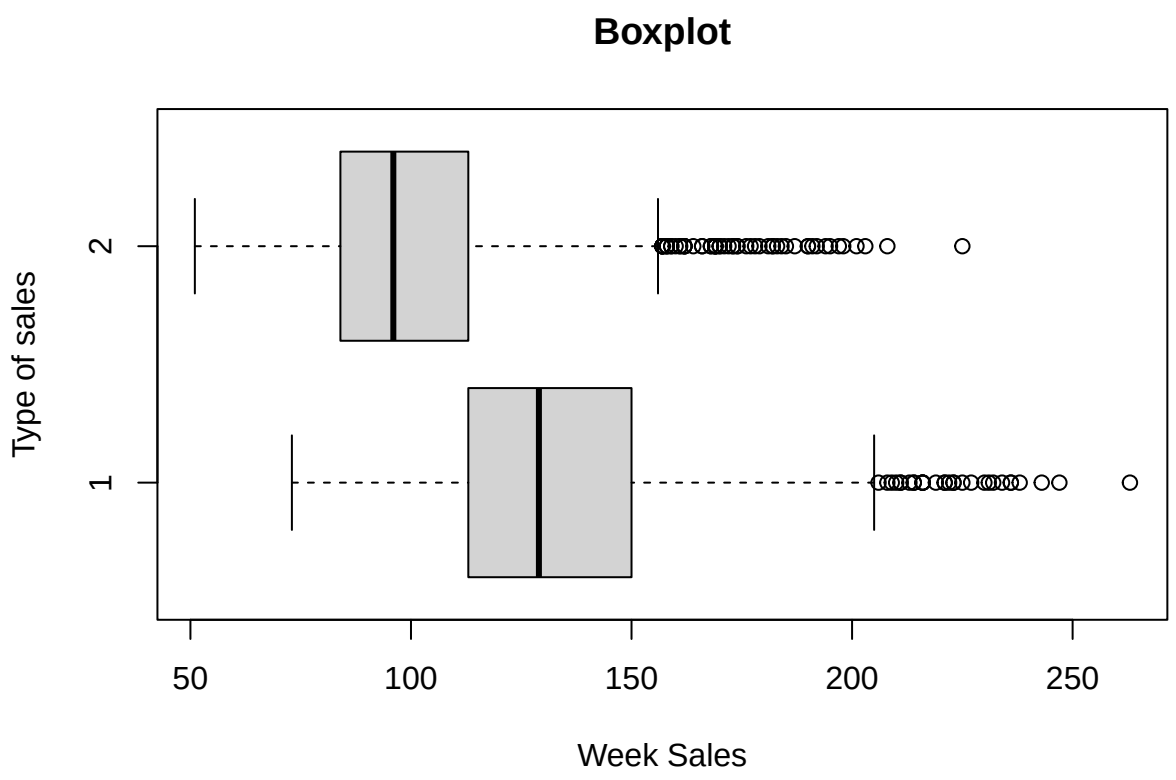
```
setwd("C:/Users/GAOSHIJIE/Desktop")
data = read.csv("store.csv")
#head(data)

#1
library(MASS)
summary(data$p2sales)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    51.0    84.0    96.0   100.2   113.0   225.0
```

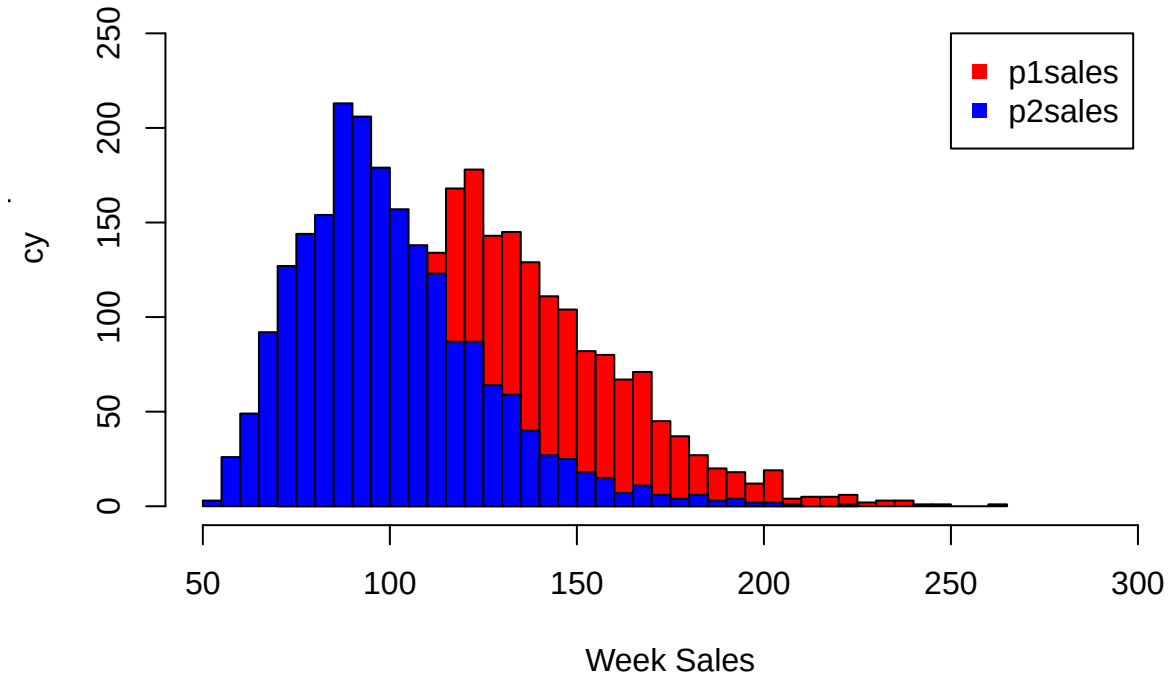
```
weekly_sales = data$p2sales

hist(weekly_sales, freq = F, main = "", ylim = c(0, 0.02))
lines(density(weekly_sales), col = "red", lwd = 2)
```



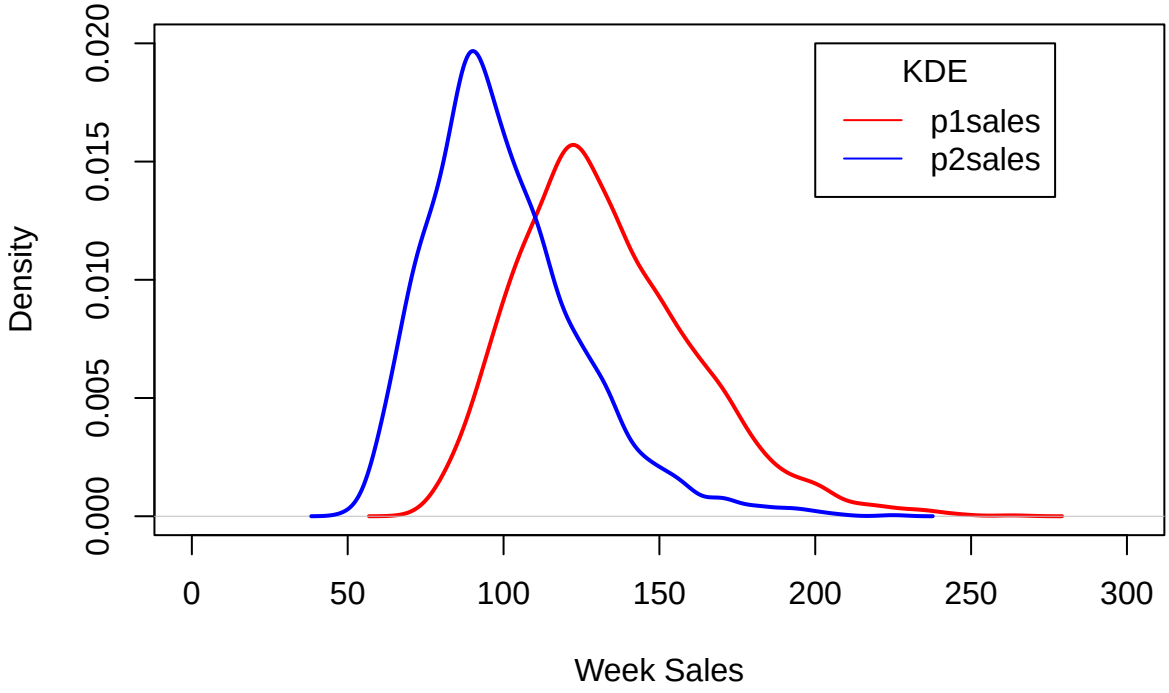
```
#2
#boxplot
boxplot(list(data$p1sales, data$p2sales), horizontal=T, main = "Boxplot",
        xlab = "Week Sales", ylab = "Type of sales")
```



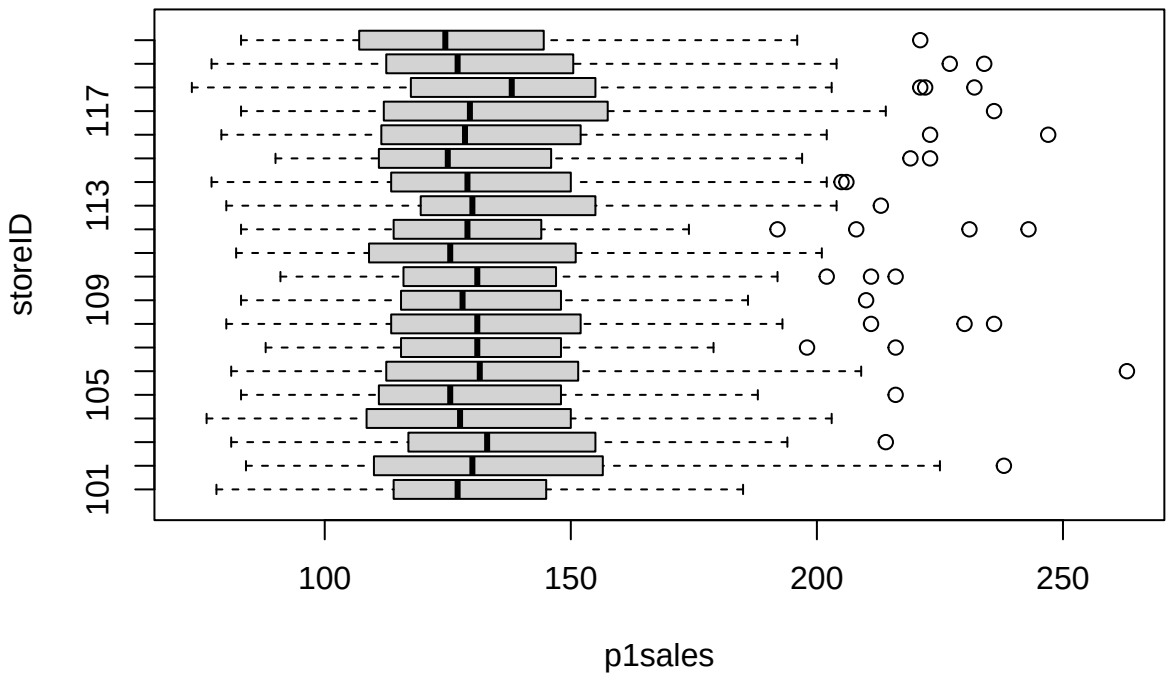

```
#overlap histogram
hist(data$p1sales, freq = T, breaks = 30, col = "red", xlim = c(50, 300),
     ylim = c(0, 250), main ="", xlab = "Week Sales", ylab = "Absolute Frequen
     cy")
hist(data$p2sales, freq = T, breaks = 30, col = "blue", add = TRUE)
legend(250, 250, legend=c("p1sales", "p2sales"), col=c("red", "blue"),
       pch=c(15, 15))
```




```
#overlap KDE
plot(density(data$p1sales), col = "red", lwd = 2 , xlim = c(0, 300),
     ylim = c(0, 0.02), main = "", xlab = "Week Sales")
lines(density(data$p2sales), col = "blue", lwd = 2)
legend(200, 0.020, legend=c("p1sales", "p2sales"), col=c("red", "blue"),
       lty = 1:1, title = "KDE")
```




```
#3
#boxplot
boxplot(p1sales~storeID, data, horizontal=T)
```



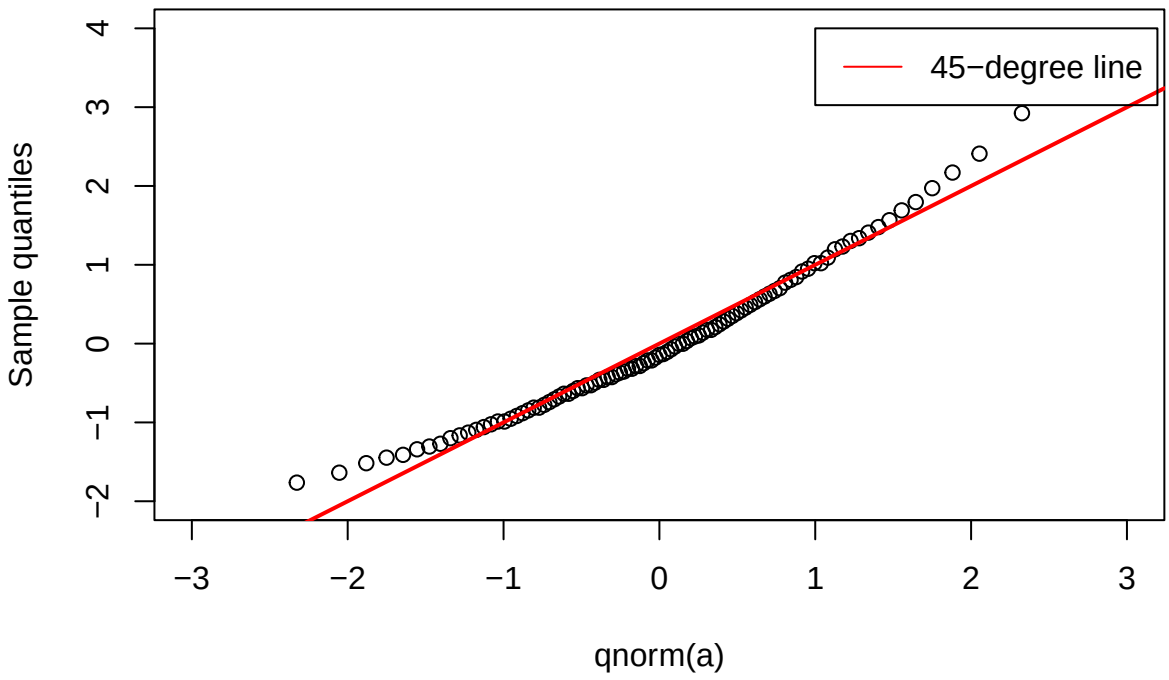

```
#Identify the store with the largest weekly sales of product P1
#Though the boxplot, we can see that storeID=106 is the store with the largest
#weekly sales of product P1


#4
#Skewness
library(e1071)
duration = data$p1sales
skewness(duration)
```

```
## [1] 0.73935
```

```
#kurtosis
kurtosis(duration)
```

```
## [1] 0.656501
```

```
#QQ plot with normal distribution
a = seq(0, 1, 0.01)
x = scale(duration)
plot(qnorm(a), quantile(x, a), xlim = c(-3, 3), ylim = c(-2, 4), ylab =
       "Sample quantiles")
abline(0, 1, lty = 1, col ="red", lwd = 2)
legend(1, 4, legend = "45-degree line", col = "red", lty = 1)
```



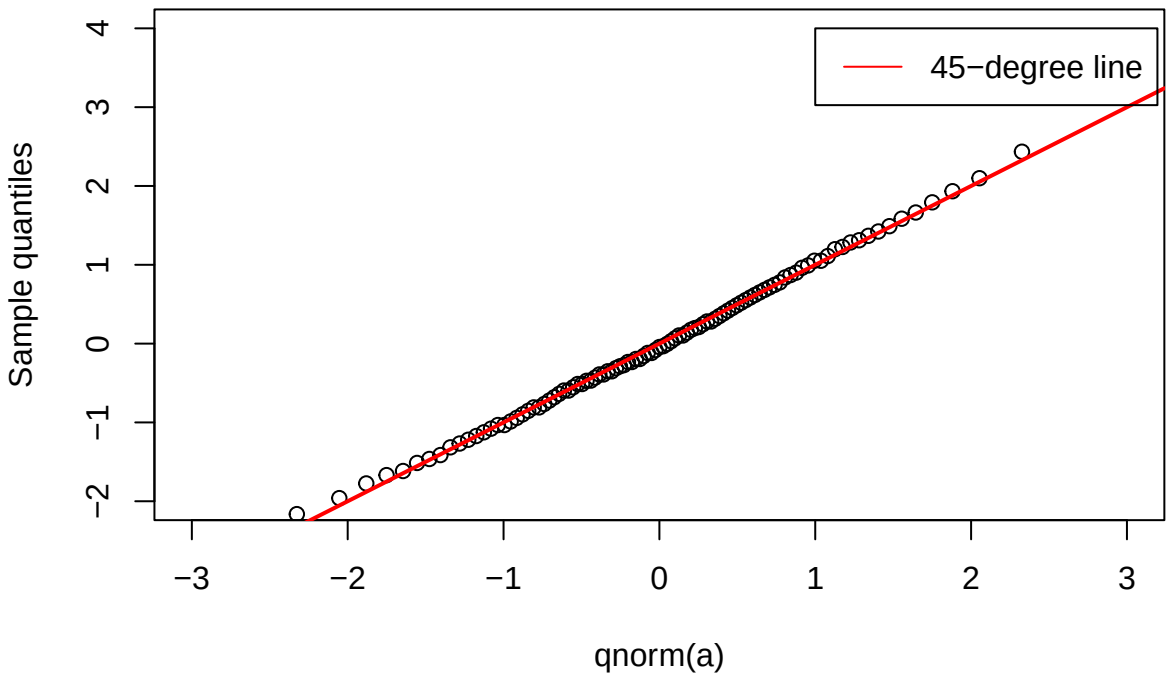

```
#ln(data$p1sales)
#Skewness
duration = log(data$p1sales)
skewness(duration)
```

```
## [1] 0.1601015
```

```
#kurtosis
kurtosis(duration)
```

```
## [1] -0.2026867
```

```
#QQ plot with normal distribution
a = seq(0, 1, 0.01)
x = scale(duration)
plot(qnorm(a), quantile(x, a), xlim = c(-3, 3), ylim = c(-2, 4), ylab =
       "Sample quantiles")
abline(0, 1, lty = 1, col ="red", lwd = 2)
legend(1, 4, legend = "45-degree line", col = "red", lty = 1)

#So normal distribution fits well to p1$sales.


#5
library(sp)
```




```
library(rworldmap)
```

```
## ### Welcome to rworldmap ###
```

```
## For a short introduction type :   vignette('rworldmap')
```

```
library(RColorBrewer)
p1sales_sum = aggregate(p1sales~country, data, sum)
# create map object
#
p1sales.map = joinCountryData2Map(p1sales_sum,
                                  joinCode = "ISO2",
                                  nameJoinColumn = "country")
```

```
## 7 codes from your data successfully matched countries in the map
## 0 codes from your data failed to match with a country code in the map
## 235 codes from the map weren't represented in your data
```

```
# display the map
#
mapCountryData(p1sales.map, nameColumnToPlot="p1sales",
               mapTitle="Sales by Country",
               colourPalette=brewer.pal(7, "Reds"),
               catMethod="fixedWidth")
```

Sales by Country