R Homework 11

Fisher Ankney

November 27th, 2018

Statistics 5193

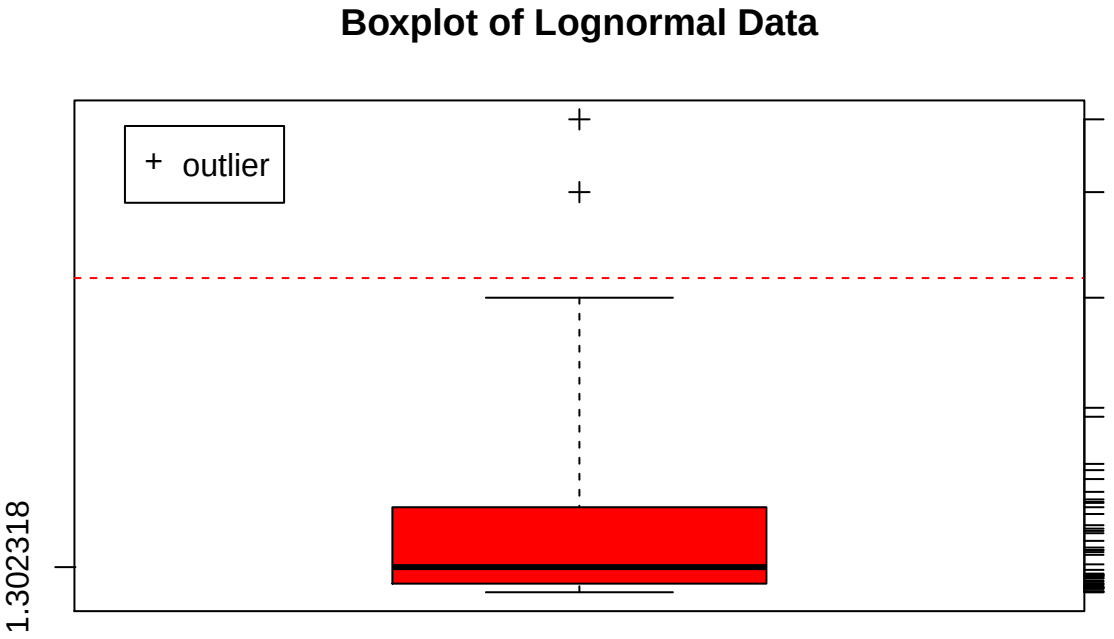
*Note: this document was created using R Markdown.

Question 1

Generate data using the code "set.seed(1)" and "x<-rlnorm(50, meanlog=0, sdlog=2)". Construct a boxplot of the data that is

- red and has title "Boxplot of lognormal data"
- has a tick mark for the median with the value of the median printed on the left side of the y axis.
- Identifies outliers as being greater than Q3+ 3*(Q3 – Q1), plots outliers with a "+", and depicts the outlier cutoff with a dashed red line.
- Has a legend that reads "+ outlier"
- Draws a tick mark for each observation on the right axis.

```
set.seed(1)

x <- rlnorm(50, meanlog = 0,sdlog = 2)

boxplot(x,col = "red",
        main="Boxplot of Lognormal Data",
        range=3,
        yaxt='n',
        pch=3)

summary <-boxplot.stats(x)

axis(2,at=summary$stats[3])

outlier<-summary$stats[4] + 3*(summary$stats[4] - summary$stats[2])

abline(h=outlier, col ="red", lty =2)

legend("topleft", legend="outlier", pch ="+", inset=0.05 )

axis(4,at=x, label=FALSE)
```


Boxplot of Lognormal Data
outlier
1.302318

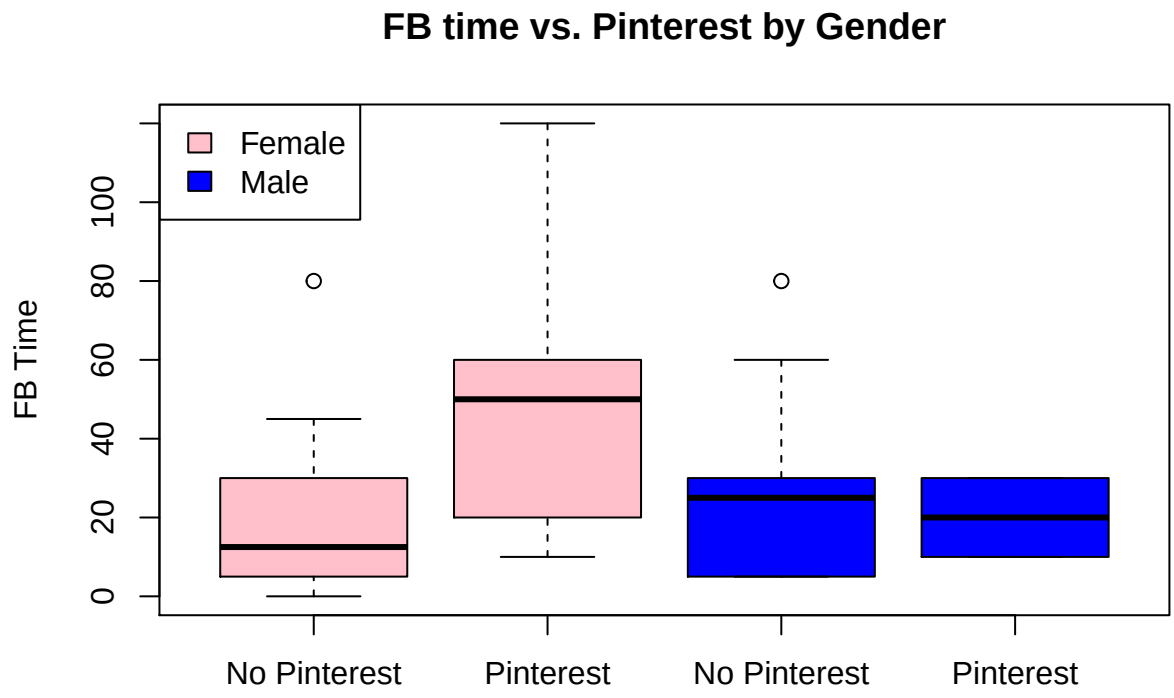
Question 2

For the student data, make side by side boxplots of FBtime by Pinterest and Gender. There should be four boxes – the first 2 are females and the second 2 are males. Specific details are:

- The female boxes should be pink and the male boxes blue
- Each box should be labeled either "Pinterest" or "No Pinterest"
- You must have the y axis labeled and have a title
- There should be a legend that indicates Female and Male with a pink and blue box, respectively. Hint: see the fill option

```
library(readxl)
StudentData <- read_excel("/Users/fisher/Documents/data_science/r_stat_5193/data/StudentData.xlsx")
```

```
boxplot(StudentData$Fbtime ~ StudentData$Pinterest*StudentData$Gender,
        col = c('pink', 'pink', 'blue', 'blue'),
        names = c('No Pinterest', 'Pinterest', 'No Pinterest', 'Pinterest'),
        ylab = 'FB Time',
        main = 'FB time vs. Pinterest by Gender'
)

legend('topleft', legend = c('Female', 'Male'), fill = c('pink', 'blue'))
```



Question 3

Get the medians of the 4 groups above and the 5% trimmed means of the four groups above using aggregate function.

```
aggregate(StudentData$Fbtime,
          by = list(pinterest = StudentData$Pinterest,
                    gender = StudentData$Gender),
          FUN = median)
```

```
##   pinterest gender    x
## 1         N      F 12.5
## 2         Y      F 50.0
## 3         N      M 25.0
## 4         Y      M 20.0
```

```
ag_mean <- function(x) {
  mean(x, trim = 0.05)
}

aggregate(StudentData$Fbtime,
          by = list(pinterest = StudentData$Pinterest,
                    gender = StudentData$Gender),
          FUN = ag_mean)
```

```
##   pinterest gender        x
## 1         N      F 21.00000
## 2         Y      F 51.15385
## 3         N      M 28.50000
## 4         Y      M 20.00000
```

Question 4

Create a dotplot of murder rates by state using the USArrests data. Make sure to order the data by murder rate and provide a title and state names.

```
ordered <- USArrests[order(USArrests$Murder),]

dotchart(ordered$Murder, labels = row.names(ordered),
         main = "Murder Rate by State")
```