Computational Statistics

Lab 5

Emil K Svensson and Rasmus Holm

2017-03-06

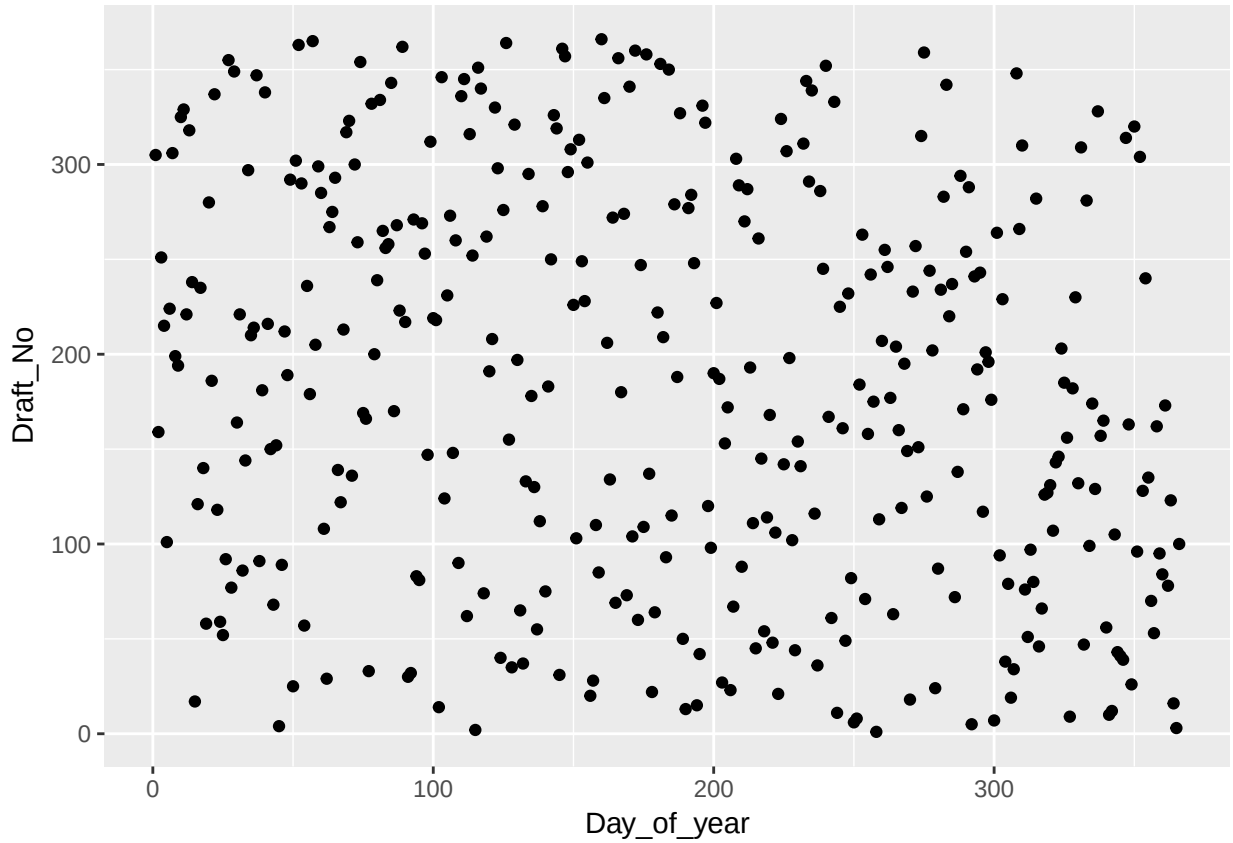
Question 1

1.1

```
library(ggplot2)

lottery <- read.csv2("../data/lottery.csv")

head(lottery)
```

```
##   Day Month Mo.Number Day_of_year Draft_No
## 1   1   Jan         1           1      305
## 2   2   Jan         1           2      159
## 3   3   Jan         1           3      251
## 4   4   Jan         1           4      215
## 5   5   Jan         1           5      101
## 6   6   Jan         1           6      224
```

```
q11 <- ggplot(lottery, aes(x = Day_of_year, y = Draft_No)) + geom_point()
plot(q11)
```

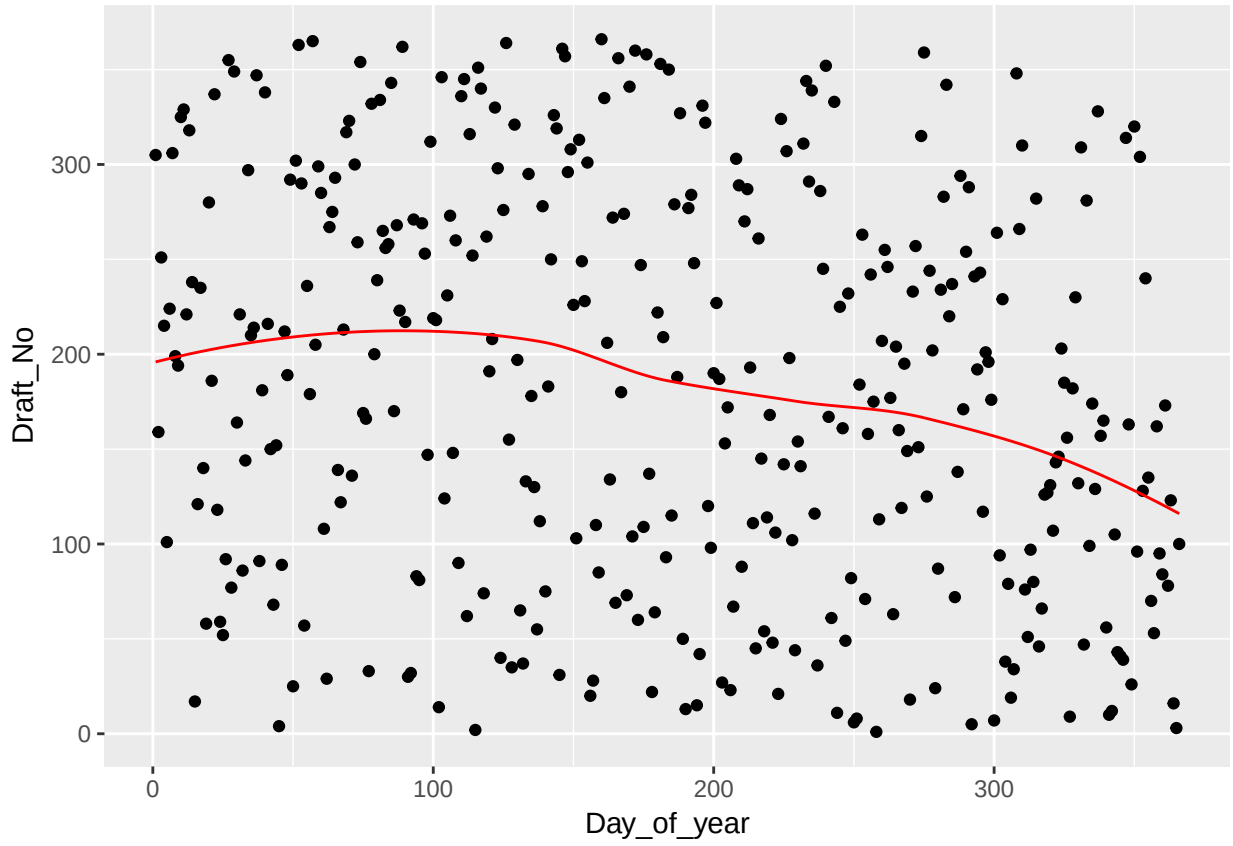



```
data <- data.frame(x=lottery$Day_of_year, y=lottery$Draft_No)
```

The data looks fairly random although there might be some sort of skewness in the right side of the graph were there are a lacking some observations and therefore having a lower probability of beeing selected.

1.2

```
loessfit <- loess(y ~ x, data=data)
data$pred <- predict(loessfit, data$x)

linegraf<- predict(loess(Draft_No ~ Day_of_year, data = lottery))

loessmod <- loess(Draft_No ~ Day_of_year, data = lottery)
lottery$pred<- predict(loessmod)

q12 <- q11 + geom_path(data = lottery, aes(y = pred), col = "red")
plot(q12)
```




```
## ggplot(data=data) +
##     geom_point(aes(x=x, y=y), color="red") +
##     geom_line(aes(x=x, y=pred), color="blue")
```

The fit (line) doesn't seem straight and seems to have a decreasing trend which would support previous statements of people born on a days later on in a year has a lower probability of beeing selected.

1.3

```
library(boot)

teststat <- function(model) {
    function(data) {
        xa <- data$x[which.min(data$y)]
        xb <- data$x[which.max(data$y)]

        ya <- predict(model(y ~ x, data), xa)
        yb <- predict(model(y ~ x, data), xb)

        (yb - ya) / (xb - xa)
    }
}

teststat_boot <- function(data, idx, stat) {
    data <- data[idx,]
```

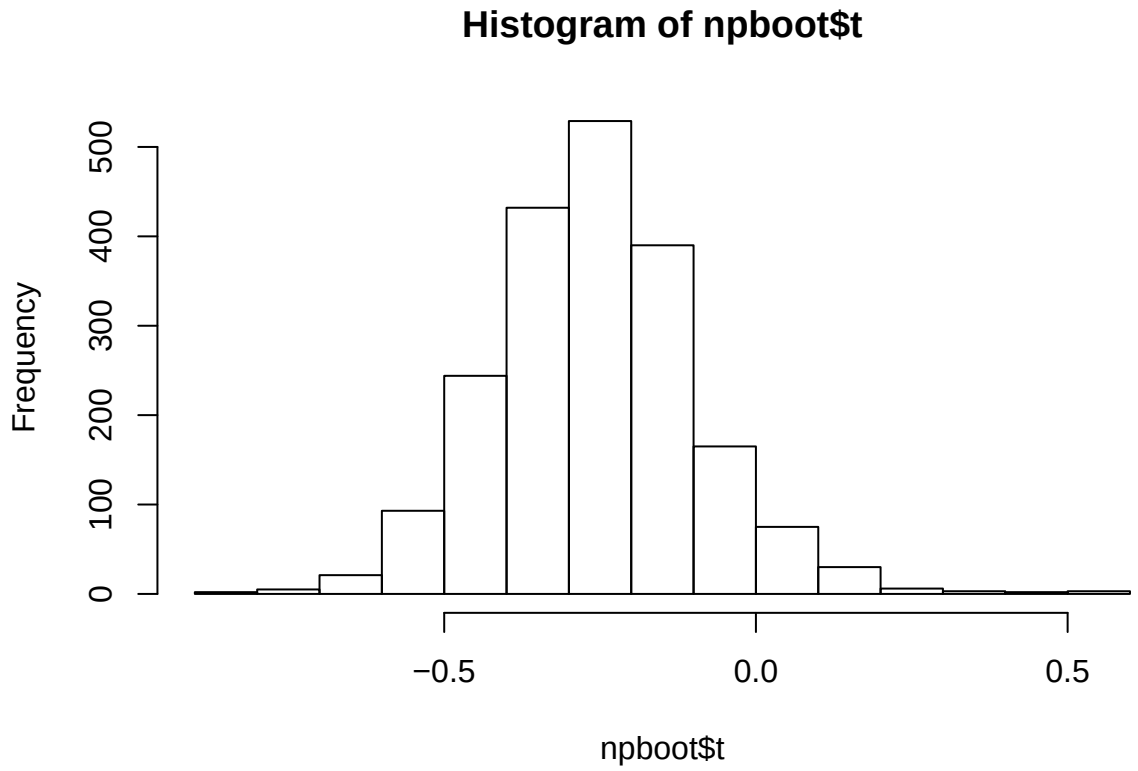
```
    stat(data)
}
```

```
B <- 2000

set.seed(123456)
npboot <- boot(data=data, statistic=teststat_boot, R=B, stat=teststat(model=loess))
```

```
## Two-sided p-value?
sum(npboot$t >= 0) / B
```

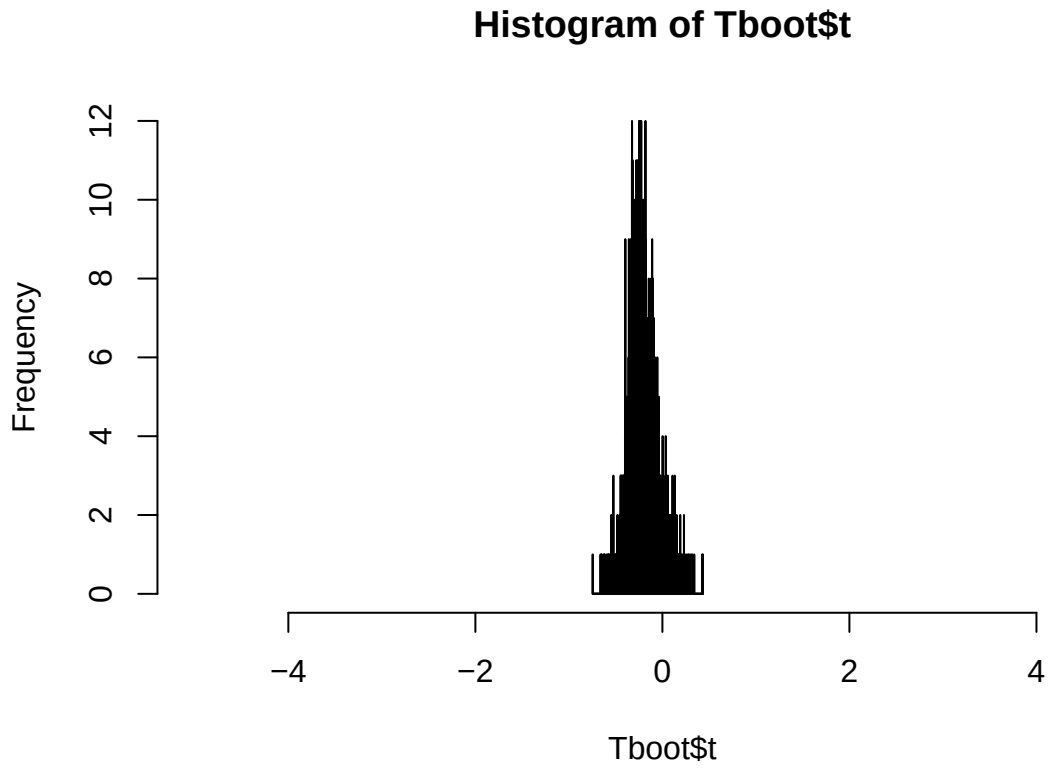
```
## [1] 0.0595
```

```
hist(npboot$t)
```




```
myT <- function(data,ind){
    data <- data[ind,]
    modelloess <-loess(Draft_No ~ Day_of_year, data)

    Xb <- data$Draft_No[which.max(data$Day_of_year)]
    Xa <- data$Draft_No[which.min(data$Day_of_year)]

    YhatXb <- predict(modelloess,newdata = data.frame(Day_of_year = Xb))
    YhatXa <- predict(modelloess,newdata = data.frame(Day_of_year = Xa))

    ##YhatXa <- data$pred[which.min(data$Day_of_year)]
```

```
    return( (YhatXb - YhatXa) / (Xb-Xa))
}

library(boot)

Tboot <- boot(lottery[,c(4,5)],myT, R = 2000)
hist(Tboot$t, xlim = c(-5,5), breaks = 1000)
```




```
pt(Tboot$t0, df = nrow(lottery))
```

```
##    1
## 0.39
```

```
mean(Tboot$t > 0)
```

```
## [1] NA
```

1.4

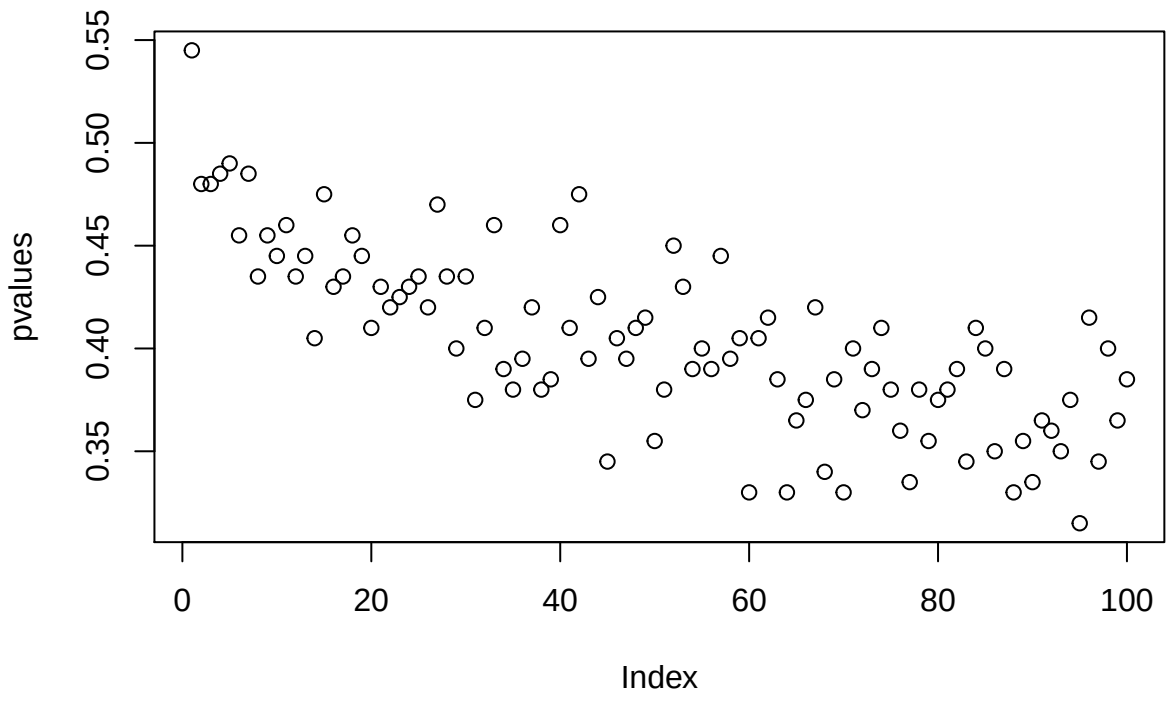
```
teststat_permutation<- function(data, B, stat) {
    n <- nrow(data)

    statistics <- rep(0, B)
    for (b in 1:B) {
        newdata <- data.frame(x=data$x, y=sample(data$y, n))
        statistics[b] <- stat(newdata)
```

```
    }

    sum(statistics >= 0) / B
}

set.seed(123456)
teststat_permutation(data, B, teststat(loess))
```

```
## [1] 0.514
```

1.5

```
genranddata <- function(x, alpha) {
    data.frame(x=x, y=pmax(0, pmin(alpha * x + rnorm(length(x), mean=183, sd=10), 366)))
}

alphas <- seq(0.1, 10, by=0.1)
pvalues <- rep(0, length(alphas))

set.seed(123456)

for (i in 1:length(alphas)) {
    newdata <- genranddata(data$x, alphas[i])
    pvalues[i] <- teststat_permutation(newdata, 200, teststat(loess))
}

plot(pvalues)
```



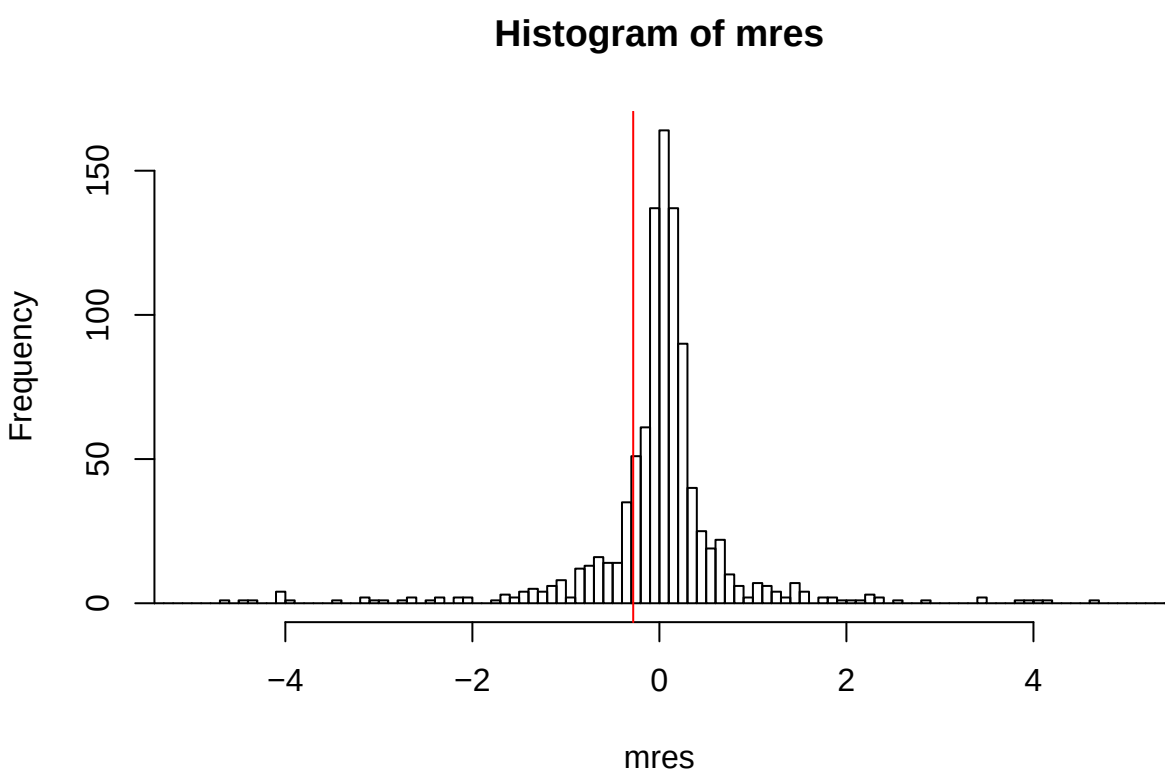

```
print(sum(pvalues <= 0.05))
```

```
## [1] 0
```

```
for (aha in seq(0.1,10, by= 0.1)){

    YandNum<- cbind(lottery[,4]*aha + rnorm(1,183,10),366)
    Ymin <-apply(YandNum, MARGIN = 1, FUN = min)
    Yx <- apply(cbind(0,Ymin), MARGIN = 1, FUN = max)
    newLotto <- data.frame(cbind(lottery[,4],Yx))
    colnames(newLotto) <- colnames(lottery[,c(4,5)])

    newres<-perm_test(newLotto, B = 200)

    ## hist(newres,breaks = 200, xlim = c(-5,5))
}
```

```
## perm_test <- function(data){
## #Assuming the data first column is labels and second is data.
##   lables<- data[,1]
##   data[,1]<- sample(lables, size = nrow(data), replace = FALSE)
##   return(aggregate(data, by = list(unique(data[,1])), FUN = mean  ))
## }

perm_test <- function(data,B){

    myPermT <- function(data){
```

```
        data[,1]<- sample(data[,1], size = nrow(data), replace = FALSE)
        data$pred <- predict(loess(Draft_No ~ Day_of_year, data))

        Xb <- data$Draft_No[which.max(data$Day_of_year)]
        Xa <- data$Draft_No[which.min(data$Day_of_year)]

        YhatXb <- data$pred[which.max(data$Day_of_year)]
        YhatXa <- data$pred[which.min(data$Day_of_year)]

        return( (YhatXb - YhatXa) / (Xb-Xa))
    }

    res <- c()

    for (i in 1:B){
        res[i] <- myPermT(data)
    }

    return(res)
}


mres<-perm_test(lottery[,c(4,5)], B = 1000)
hist(mres,breaks = 1000, xlim = c(-5,5))

abline(v = myT(data = lottery[,c(4,5)]), col = "red")
```


Histogram of mres
Frequency
0
50
100
150
−4
−2
0
2
4
mres

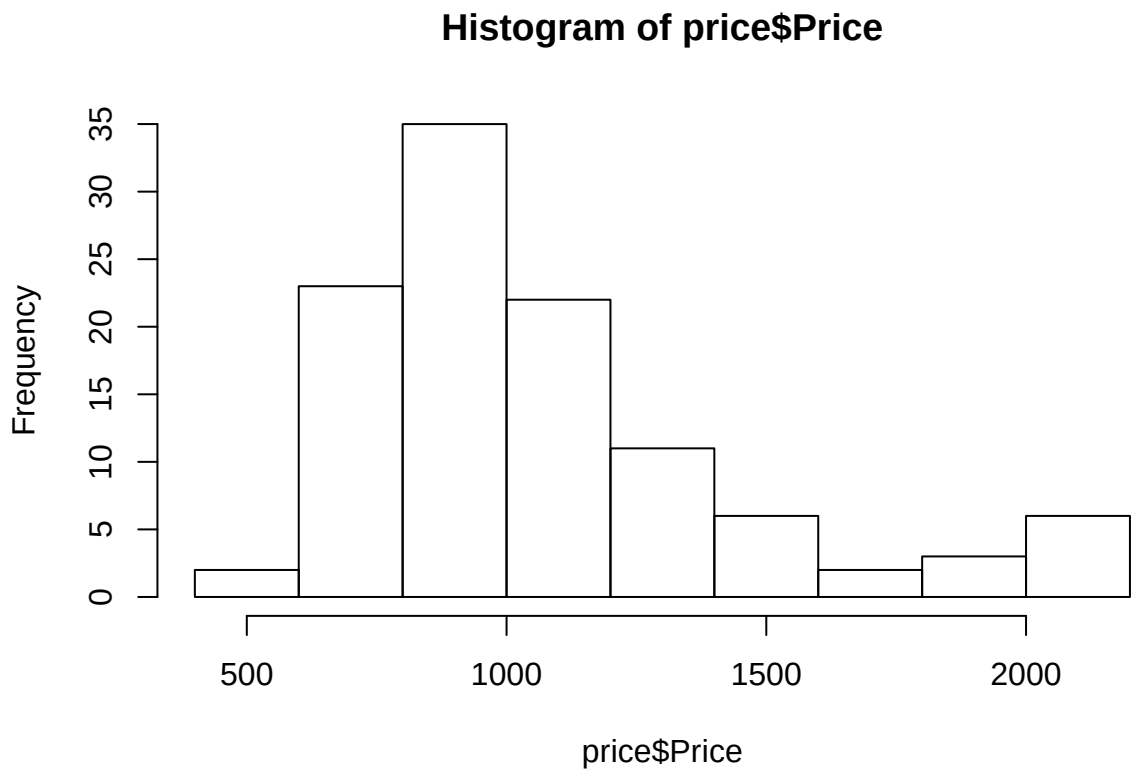
Question 2

2.1

```
price <- read.csv("../data/prices1.csv", sep=";")
mean(price$Price)
```

```
## [1] 1080
```

```
hist(price$Price)
```



Looks like a Gamma distribution.

2.2

2.3

2.4