Computational Statistics

Lab 5

Emil K Svensson and Rasmus Holm

2017-03-07

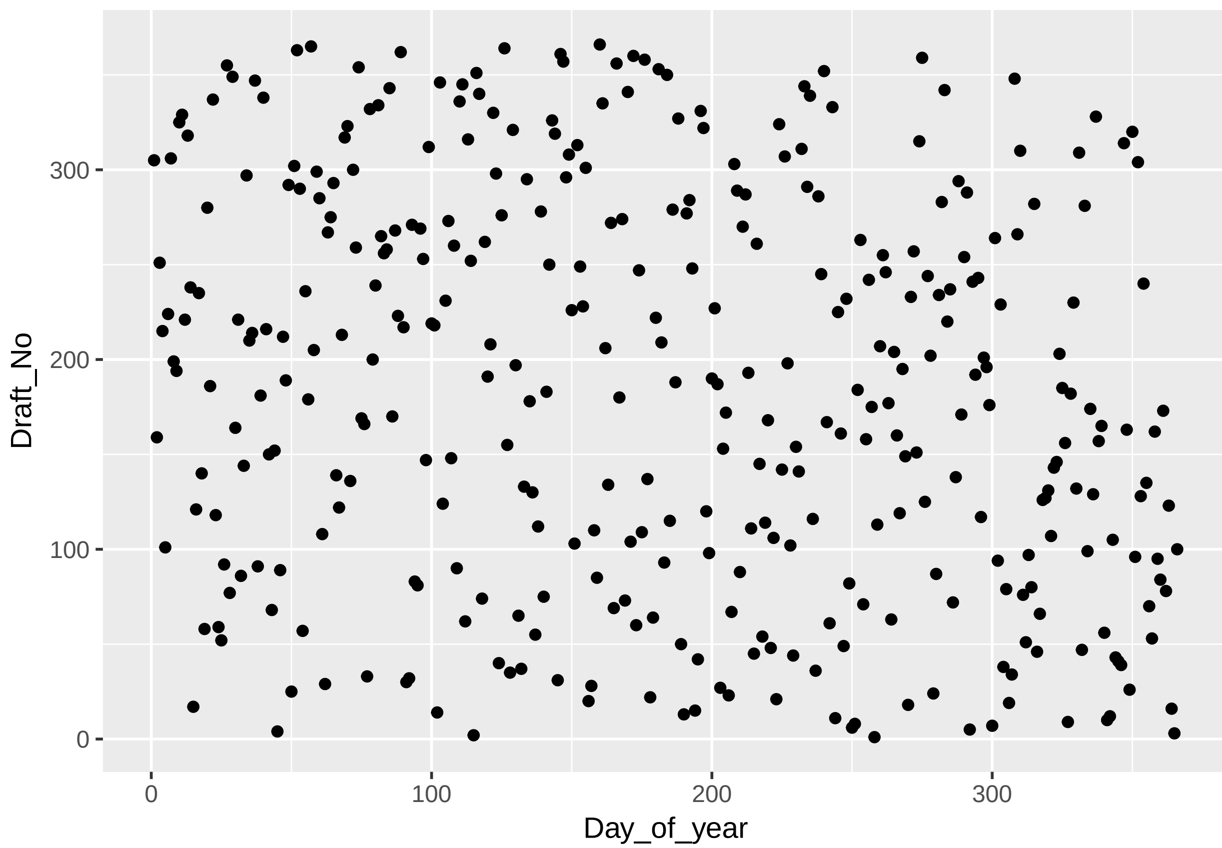
Question 1

1.1

```
library(ggplot2)

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

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



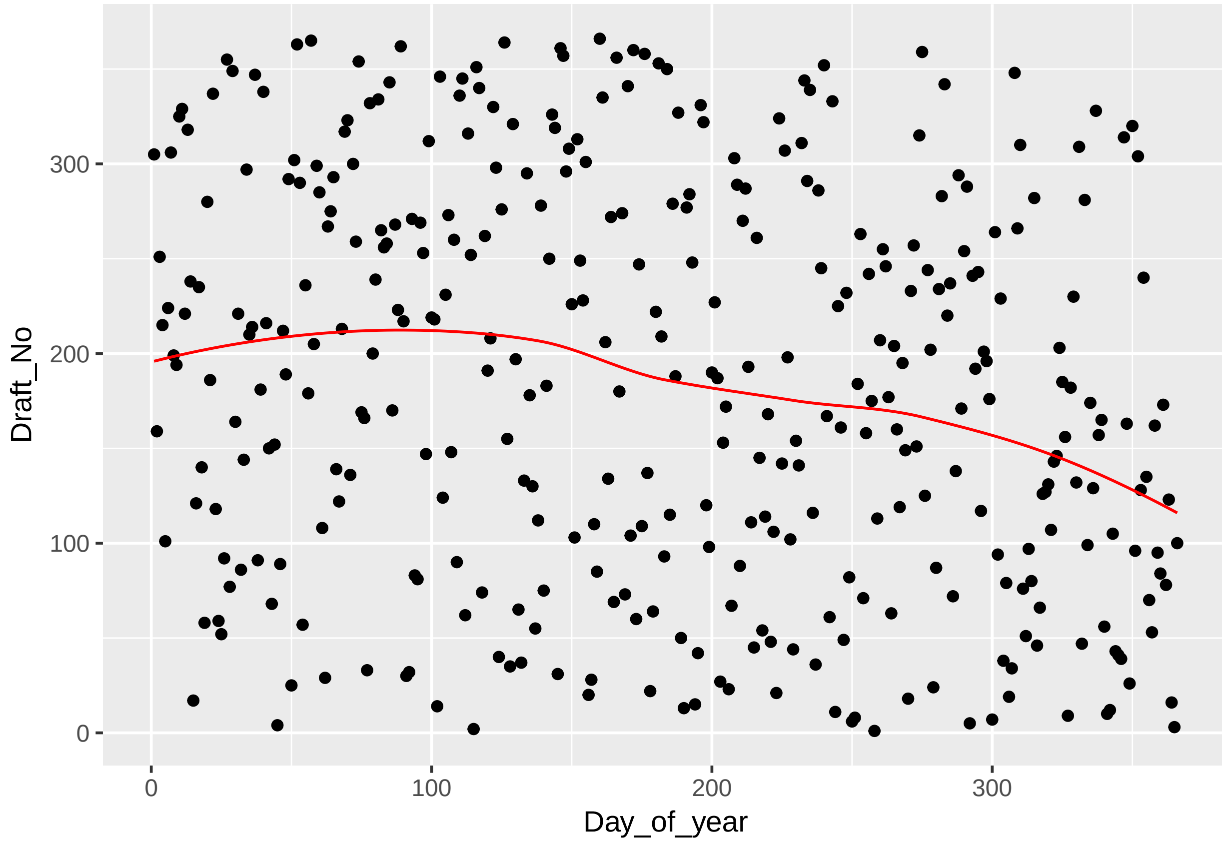

```
data <- data.frame(x=lottery$Day_of_year, y=lottery$Draft_No)
```

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

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

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



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

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

        fit <- model(y ~ x, data)

        ya <- predict(fit, xa)
        yb <- predict(fit, xb)

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

```r
    }
}

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

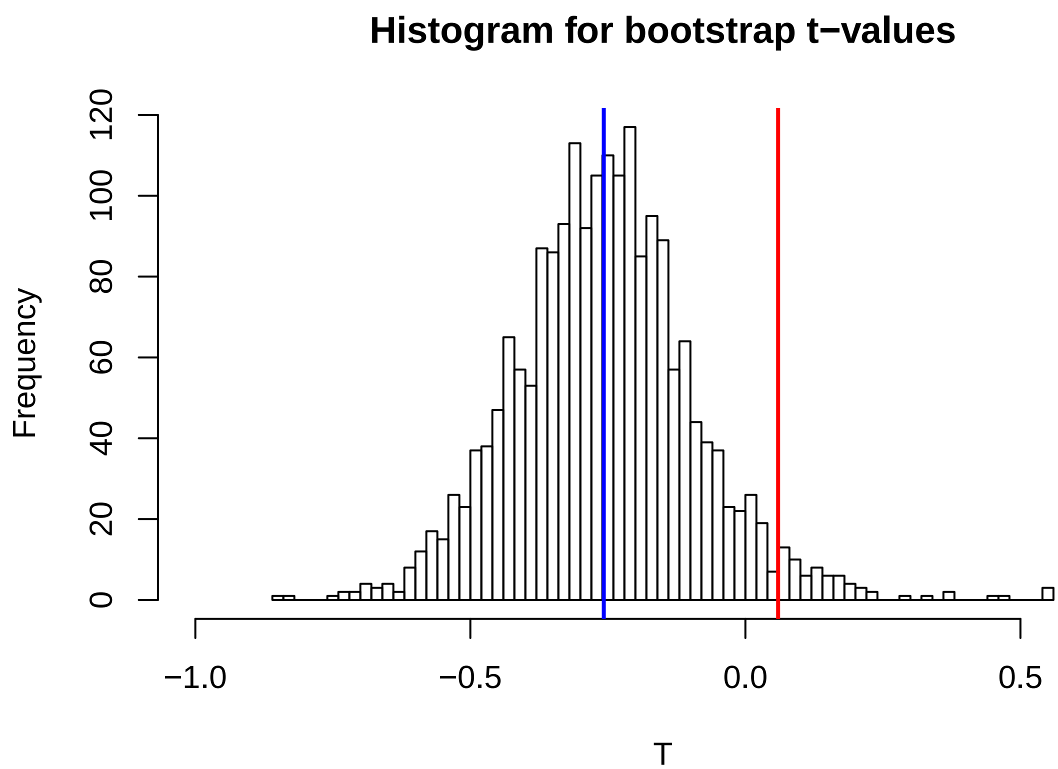
```r
library(boot)

B <- 2000

set.seed(123456)
npboot <- boot(data=data, statistic=teststat_boot, R=B, stat=teststat(model=loess))
pvalue <- sum(npboot$t > 0) / B
pvalue
```

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

```r
hist(npboot$t, xlim = c(-1,0.7), breaks = 50,
     main = "Histogram for bootstrap t-values", xlab ="T")
abline(v=pvalue, col = "red", lwd = 2)
abline(v=mean(npboot$t), col="blue", lwd=2)
```



The p-value was calculated to 0.06 and we can't reject the nullhypothesis at a 0.05 level. We conclude that this test tells us that the lottery is random.

1.4

```r
teststat_permutation<- function(data, B, stat) {
    n <- nrow(data)
    statistics <- rep(0, B)
    newdata <- data.frame(x=data$x, y=sample(data$y, n))

    for (b in 1:B) {
        statistics[b] <- stat(newdata)
        newdata$y <- sample(data$y, n)
    }

    sum(abs(statistics) >= abs(stat(data))) / B
}

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

```
## [1] 0.0925
```

1.5

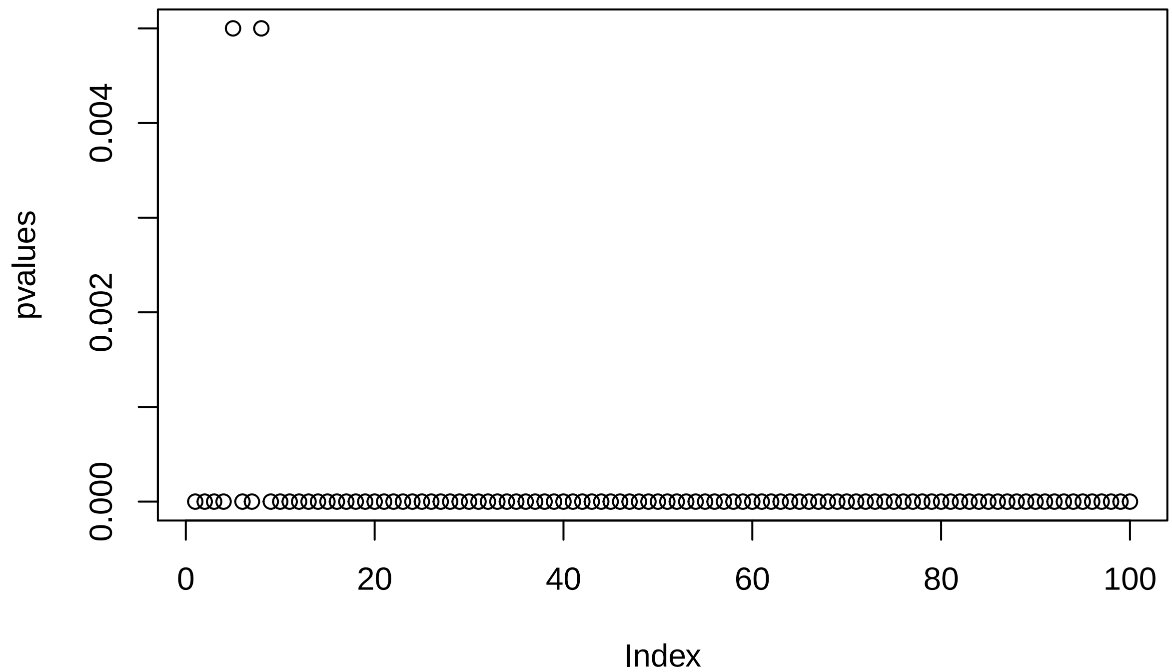
```r
genranddata <- function(x, alpha) {
    data.frame(x=x, y=pmax(0, pmin(alpha * x + rnorm(length(x), mean=183, sd=10), 366)))
}

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)
```




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

```
## [1] 100
```

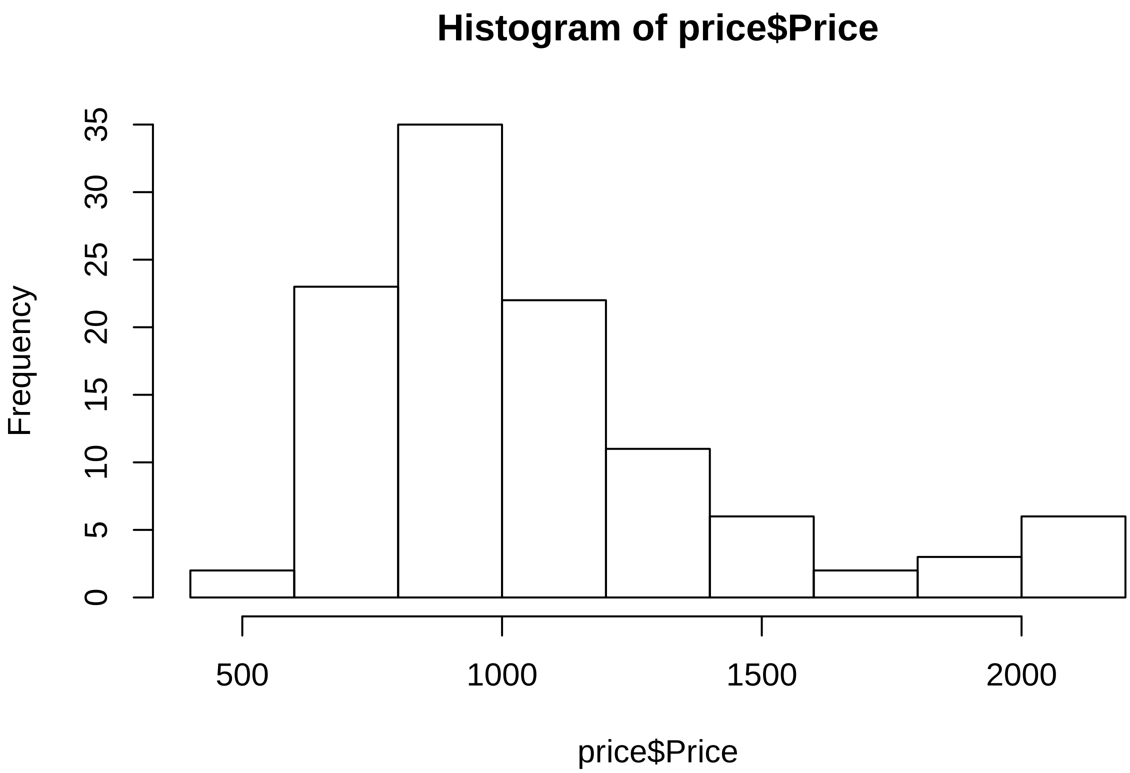
Question 2

2.1

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

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

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



Looks like a Gamma distribution.

2.2

2.3

2.4