R Homework 1

Георги Божинов 81333 курс 3 група 1

```
library(UsingR)
attach(Cars93)
```

Задача 1:

Модел на петте най-скъпи коли

```
mp <- data.frame(Price, Model)
head(mp[order(-mp$Price),], 5)

##    Price    Model
## 59  61.9     300E
## 48  47.9      Q45
## 11  40.1  Seville
## 19  38.0 Corvette
## 4   37.7      100
```

Брой цилиндри на най-мощните спортни коли

```
sport_cars <- Cars93[Type=="Sporty",]
cp <- data.frame(sport_cars$Horsepower, sport_cars$Cylinders)
head(cp[order(-cp$sport_cars.Horsepower),], 3)

##   sport_cars.Horsepower sport_cars.Cylinders
## 2                   300                    8
## 3                   300                    6
## 9                   255               rotary
```

Мили за галон -> Литри за 100 километра Градско:

```
282.48/MPG.city

##  [1] 11.299200 15.693333 14.124000 14.867368 12.840000
12.840000 14.867368
##  [8] 17.655000 14.867368 17.655000 17.655000 11.299200
11.299200 14.867368
## [15] 13.451429 15.693333 18.832000 16.616471 16.616471
14.124000 12.281739
## [22] 14.124000  9.740690 12.281739 12.840000 16.616471
13.451429 15.693333
## [29]  9.740690 14.124000  9.112258 12.281739 12.840000
12.840000 11.770000
## [36] 18.832000 13.451429 15.693333  6.140870  9.416000
11.770000  6.725714
## [43] 11.770000  9.740690 12.840000 10.864615 14.124000
16.616471 15.693333
## [50] 15.693333 16.616471 15.693333  9.740690 10.088571
```

```
10.864615 15.693333
## [57] 16.616471 14.124000 14.867368 12.281739 14.867368
9.740690 15.693333
## [64]  9.740690 11.770000 16.616471 13.451429 11.770000
12.281739 15.693333
## [71] 14.867368 12.281739  9.112258 12.281739 14.867368
14.867368 14.867368
## [78] 14.124000 10.088571  8.560000 11.299200 12.281739
7.243077  8.827500
## [85] 11.299200 12.840000 15.693333 11.299200 16.616471
13.451429 15.693333
## [92] 13.451429 14.124000
```

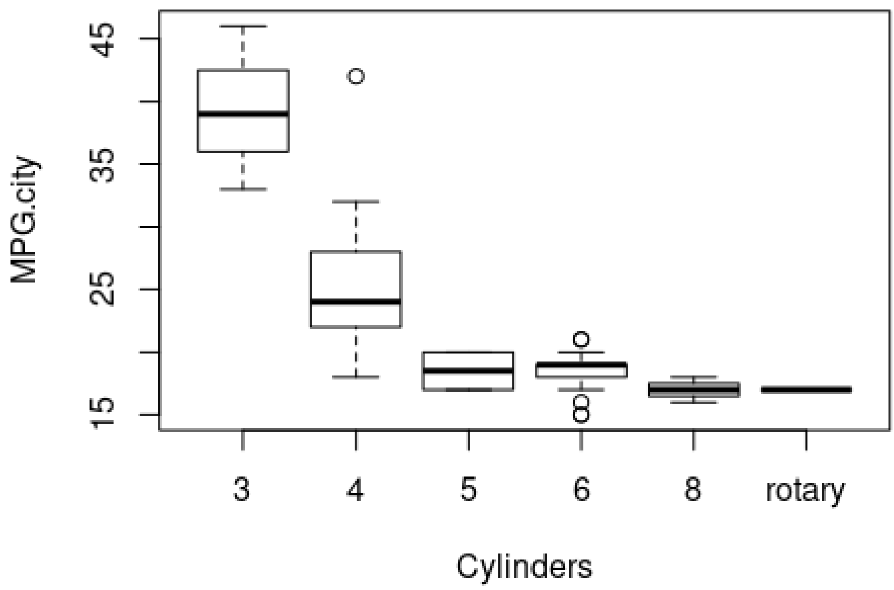
Магистрала:

```
282.48/MPG.highway
```

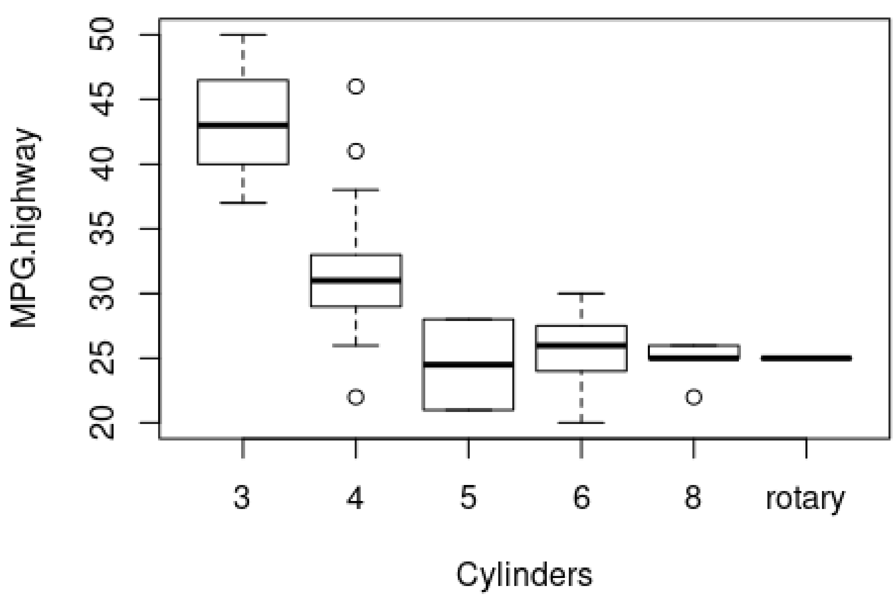
```
##  [1]  9.112258 11.299200 10.864615 10.864615  9.416000
9.112258 10.088571
##  [8] 11.299200 10.462222 11.299200 11.299200  7.846667
8.308235 10.088571
## [15]  9.740690 12.281739 14.124000 10.864615 11.299200
10.088571 10.088571
## [22] 10.864615  8.560000  9.740690 10.462222 13.451429
10.462222 11.770000
## [29]  8.560000 10.088571  8.560000  9.416000 10.462222
9.740690  9.416000
## [36] 14.124000  9.416000 10.864615  5.649600  7.846667
9.112258  6.140870
## [43]  9.112258  8.560000  9.740690  8.308235 10.462222
12.840000 11.770000
## [50] 12.281739 10.864615 10.864615  7.634595  7.846667
8.308235 11.770000
## [57] 11.299200  9.740690 11.299200 10.864615 10.864615
8.560000 11.770000
## [64]  8.560000  9.416000 12.281739 10.864615  9.112258
9.112258 12.281739
## [71] 10.088571  9.416000  6.889756  9.112258 10.088571
10.462222 10.088571
## [78] 10.864615  7.433684  7.634595  9.416000  9.416000
6.569302  7.634595
## [85]  8.827500  9.740690 12.840000  8.560000 13.451429
9.416000 11.299200
## [92] 10.088571 10.088571
```

Връзка между брой цилиндри и мили за галон

```
plot(Cylinders, MPG.city, xlab='Cylinders', ylab='MPG.city')
```



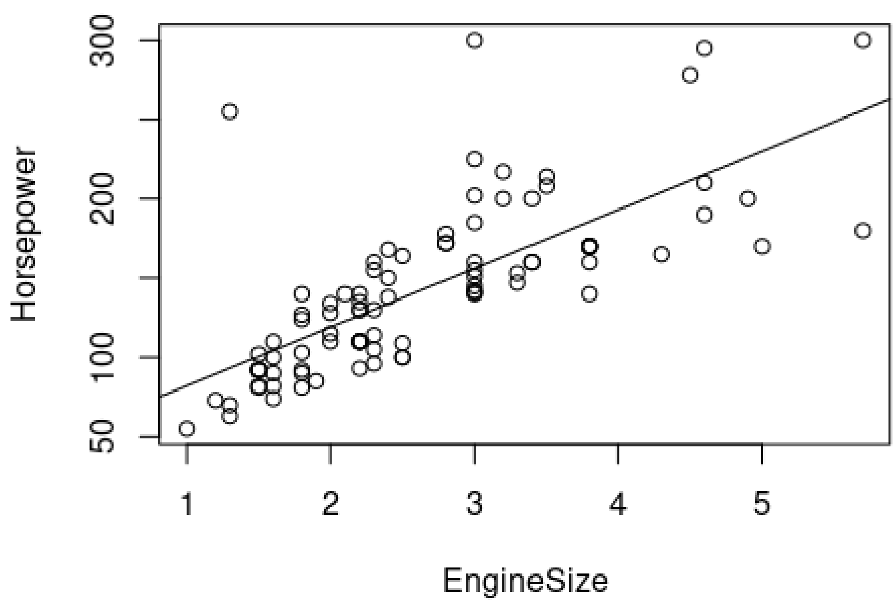

```
plot(Cylinders, MPG.highway, xlab='Cylinders',
ylab='MPG.highway')
```



Виждаме, че колкото повече цилиндри има кола, толкова по-малко мили за галон може да измине. И съответно толкова повече литри изхабява за 100 километра.

Връзка между обем на двигателя и мощността на колата.
Намираме и outliers с функция identify.

```
plot(EngineSize, Horsepower)
abline(lm(Horsepower~EngineSize))
p <- identify(EngineSize, Horsepower)
```



Корелация с outliers, изглежда като линеен модел, тъй като корелацията е > 0.5.

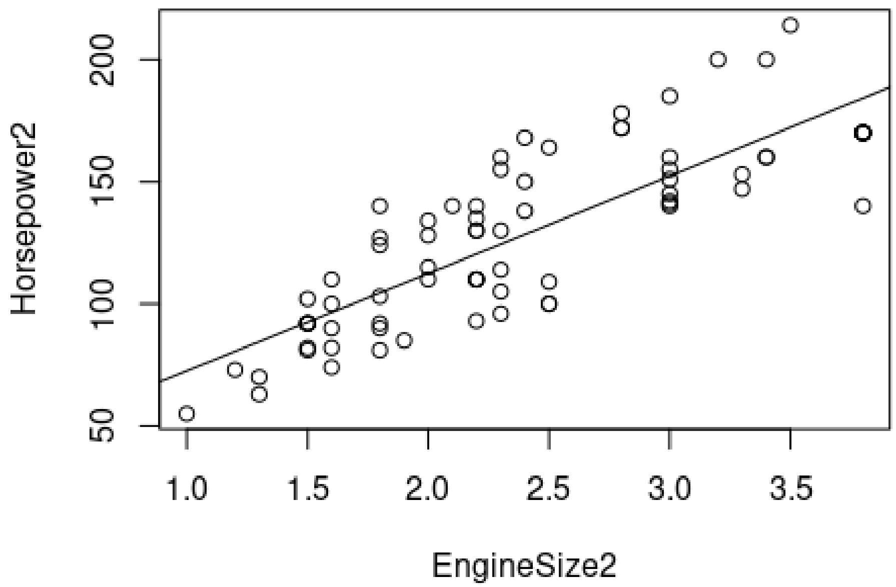
```
cor(Horsepower, EngineSize)

## [1] 0.7321197
```

Махаме outliers (емпирично).

```
p <- c(5, 6, 8, 10, 11, 17, 18, 19, 28, 38, 48, 50, 51, 52, 57,
59, 63)
Horsepower2 <- Horsepower[-c(p)]
EngineSize2 <- EngineSize[-c(p)]
```

Графика без outliers.

```r
plot(EngineSize2, Horsepower2)
abline(lm(Horsepower2~EngineSize2))
```



Доближаваме се до линеен модел с по-добра корелация.

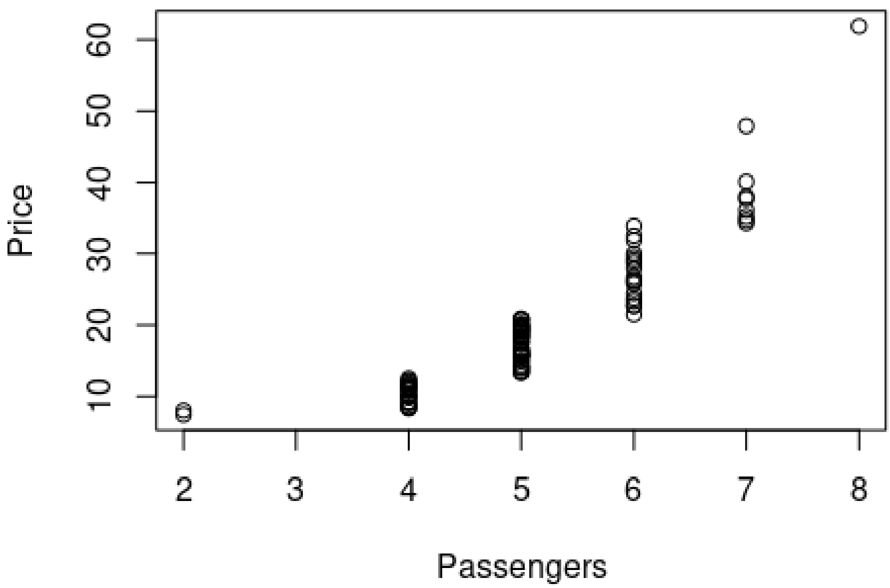
```r
cor(Horsepower2, EngineSize2)

## [1] 0.827873
```

Извод: Махането на outliers увеличава корелацията. Моделът се доближава все повече до линеен.

```r
cor(Horsepower, EngineSize, method = "spearman")

## [1] 0.8141756
```

Сравняваме корелацията с метод, който изпълнява по-строга корелация. Почти същата е като “емпиричната”.

```r
qqplot(Passengers, Price)
```

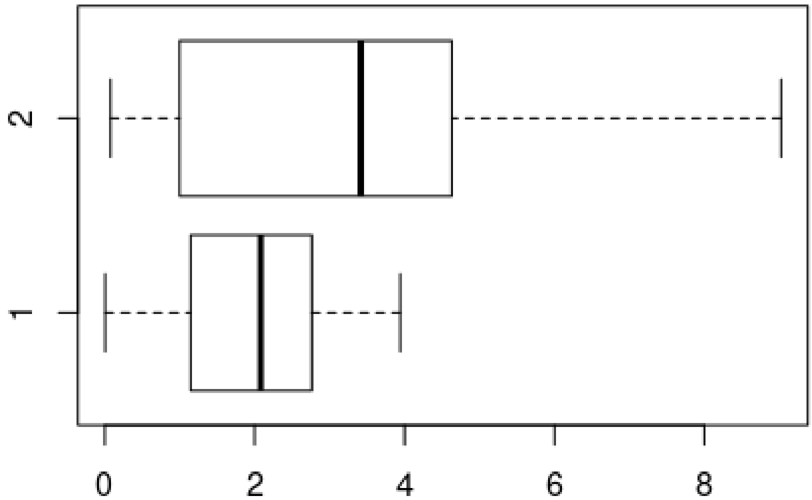


Извод: Пътниците са управляваща група. От графиката виждаме за кои ценови интервали по колко пътника може да побере колата.

Задача 2:

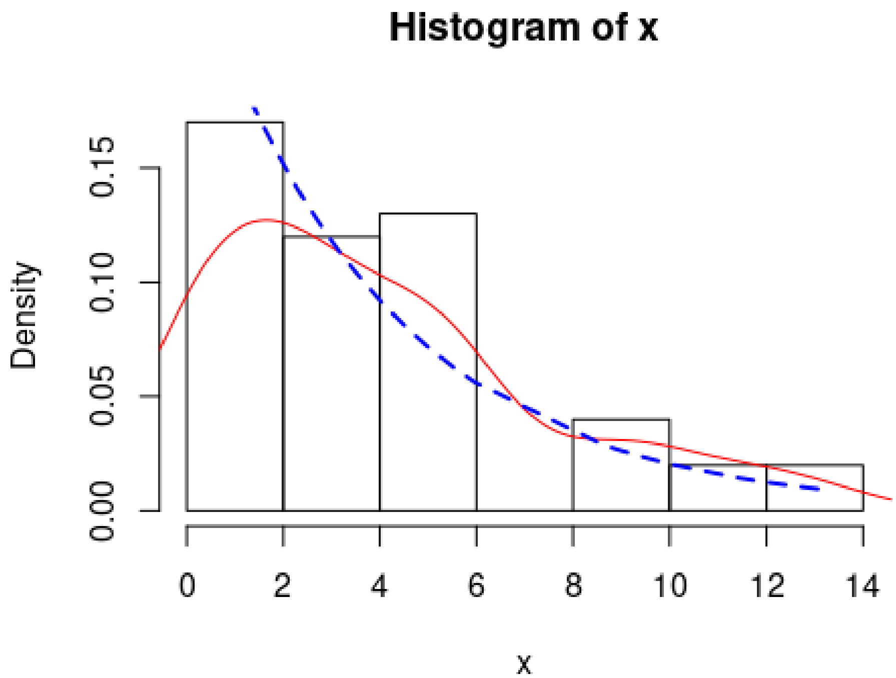
Фак. номер 81333

```
x <- runif(50, 0, 4)
y <- rexp(50, 1/4)

boxplot(x, y, horizontal = T)
```



От резултатите виждаме, че 50 наблюдения не винаги стигат, за да се проявят характеристиките на разпределението. Равномерното разпределение не винаги изглежда съвсем симетрично, а експоненциалното не винаги е с голяма дясна опашка. Функциите се държат добре в повечето случаи, но 100 наблюдения биха били по-добър критерий.

```
x <- rexp(50, 1/4)
hist(x, probability = T)
# Empirical density
lines(density(x), col="red")
# Theoretical density
lines(sort(x), dexp(sort(x), rate=1/4), col="blue", lty=2, lwd=2)
```

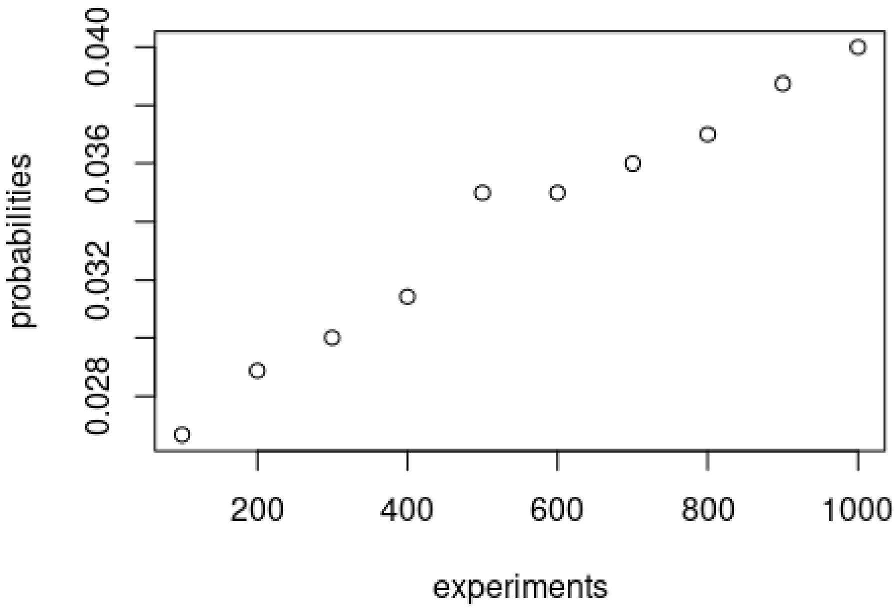


Тук отново се вижда, че наблюденията са малко. Има голямо разминаваме между плътностите.

Задача 3:

```
prob_of_sum <- function(k, t, n = 100) {
  k_count <- 0
  for (i in 1:n) {
    x <- sample(0:t, 4, replace=T)
    if(x[1] + 2*x[2] + 3*x[3] + x[4] == k) {
      k_count <- k_count + 1
    }
  }

  return(k_count/n)
}


experiments <- c()
probabilities <- c()

t <- sample(1:10, 1)
k <- sample(10:30, 1)

for (i in seq(100, 1000, 100)) {
  experiments <- append(experiments, i)
```

```
  prob <- prob_of_sum(k, t, i)

  probabilities <- append(probabilities, prob)
}

qqplot(experiments, probabilities)
```



Правим 10 последователни експерименти от 100 до 1000 със стъпка 100 и чертаем графика за вероятността да получим съответната сума. Сравняваме го с теоретично разпределение чрез qqplot.