Домашно 1 по ВС (практикум) ФН: 45342

Ivo Stratev

```r
fnX = 4
fnY = 2
```

Задача 1.

```r
library("UsingR")
start = fnY * 500
end = start + (1000 - 1)
x = (homedata $ y1970)[start:end]
y = (homedata $ y2000)[start:end]
```

а) Най-евтината и най-скъпата къща съответно за 1970 и 2000 година

```r
min(x)
## [1] 13600
max(x)
## [1] 297200
min(y)
## [1] 39600
max(y)
## [1] 1042000
```

б) 5-те най-скъпи къщи през 1970г. и цената, която имат те през 2000г.

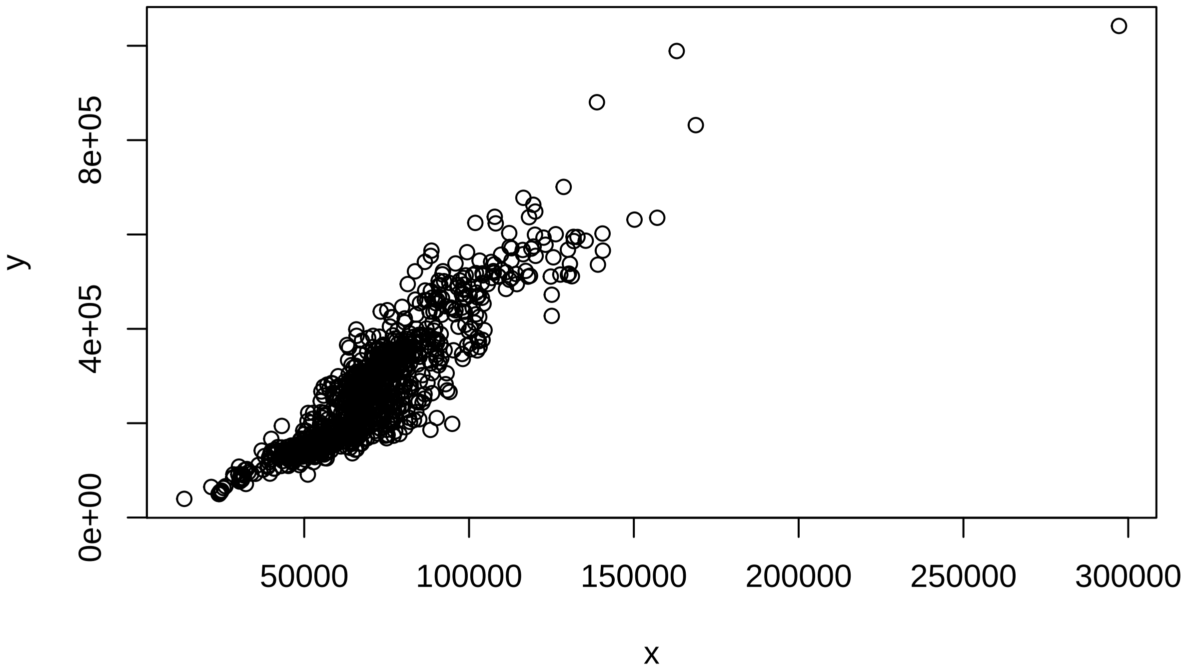
```r
top5 = order(x, decreasing = TRUE)[1:5]
x[top5]
## [1] 297200 168800 163000 157100 150200
y[top5]
## [1] 1042000  831800  988900  635400  631400
```

в) Брой на къщите, чиято цена се е увеличила с по-малко от 40 000$ между 1970г. и 2000г.

```r
price.diff = y - x
sum(price.diff > 0 & price.diff < 40000)
## [1] 11
```

г) Графично представяне на данните

```
plot(x, y)
```



Извод: Съществува линейна зависимост между цената на една къща през 1970г. и 2000г.

д) Линейна зависимост между цената на една къща през 1970г. и 2000г.

```
lr = cor(x, y)
lr
```

```
## [1] 0.8975097
```

```
lr^2
```

```
## [1] 0.8055236
```

```
cor(rank(x), rank(y))
```

```
## [1] 0.8836533
```

Коефициентите на корелация са сравнително близко до 1-ца, следователно има линейна зависимост между данните.

е) Очаквана цена на една къща ако през 1970г. тя е имала цена 80 000$

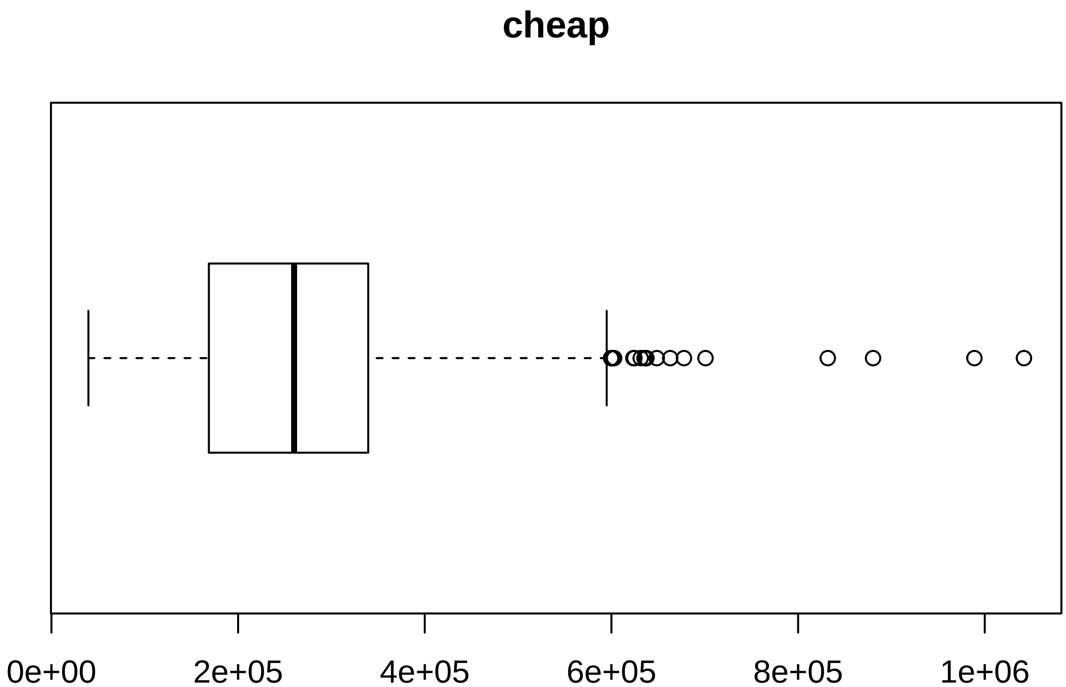
```
predict.lm(lm(y ~ x), data.frame(x = 80000))[[1]]
```

```
## [1] 322771.8
```

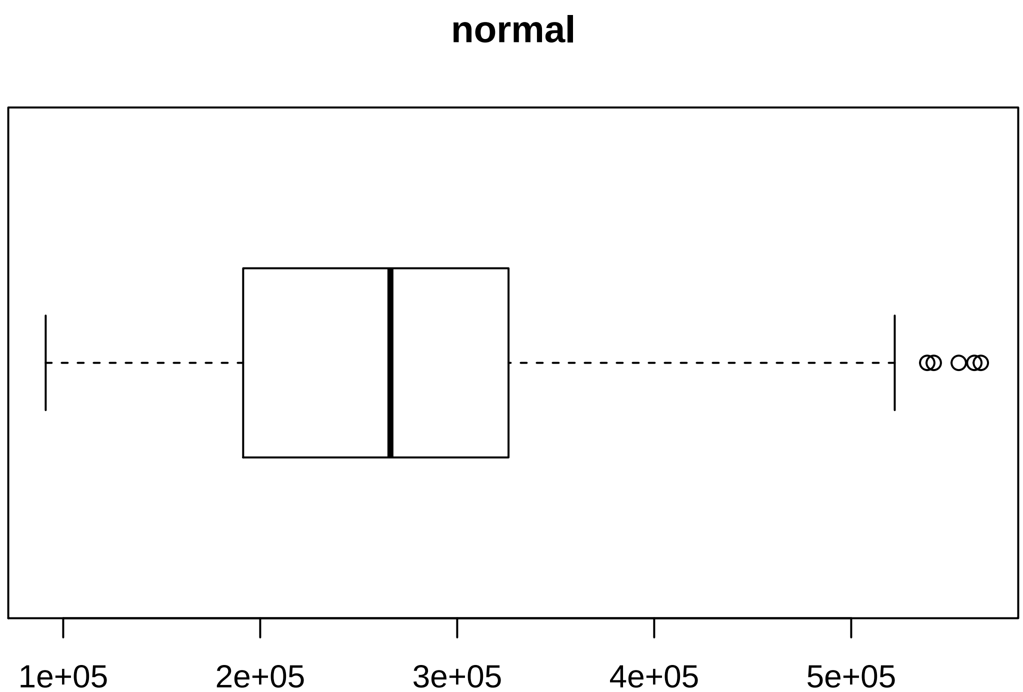
ж) Разделяне на цените на къщите през 1970г. по категории.

```
price.kind = cut(x, breaks = c(0, 50000, 100000, max(x)))
levels(price.kind) = c("cheap", "normal", "expensive")
```

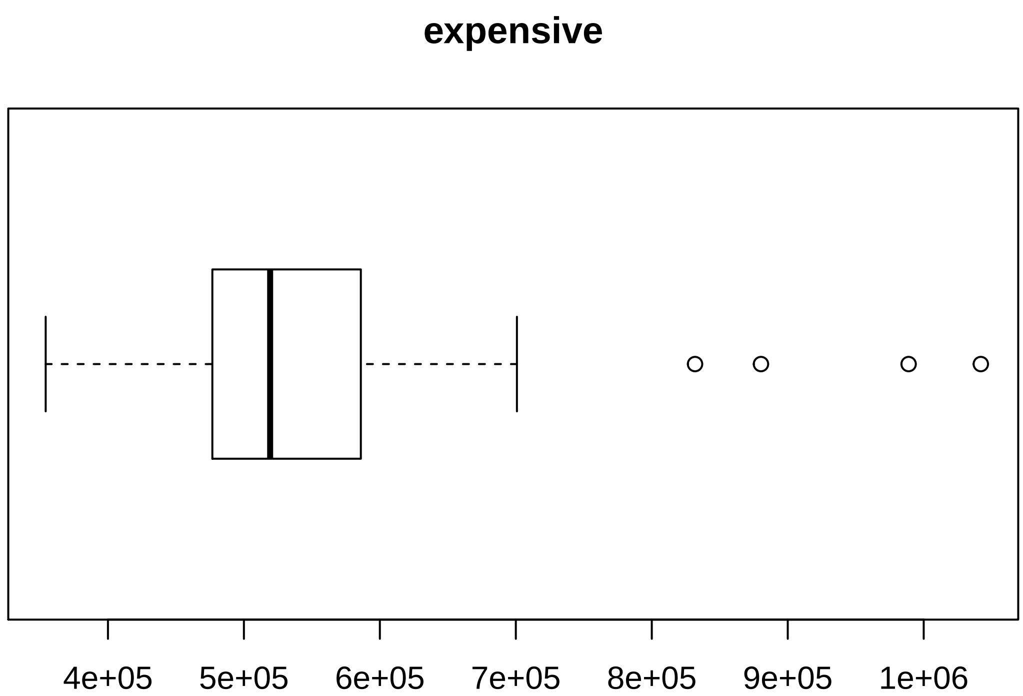
```
boxplot(y, main="cheap", horizontal=TRUE)
```




```
boxplot(y[price.kind == "normal"], main="normal", horizontal=TRUE)
```




```
boxplot(y[price.kind == "expensive"], main="expensive", horizontal=TRUE)
```



з) Средна цената по категории и категоря с най-голямо процентно увеличение

```
mean(x[price.kind == "cheap"])
```

```
## [1] 40351.91
```

```
mean(x[price.kind == "normal"])
```

```
## [1] 70559.44
```

```
mean(x[price.kind == "expensive"])
```

```
## [1] 118242.2
```

```
cats = c("cheap", "normal", "expensive")
cats[which.max(lapply(cats, function(cat) max((y[price.kind == cat] - x[price.kind == cat]) / x[price.kind == cat])))]
```

```
## [1] "normal"
```

Задача 2.

```
next.derangement = function(i, prev) i * prev + if(i %% 2 == 0) 1 else -1

derangements = function(n) {
  ai = 1
  for(i in 1:n) {
    ai = next.derangement(i, ai)
  }
  ai
}

probability.derangements = function(n) derangements(n) / factorial(n)
```

а) Теоретична вероятност никой от (10 + 2) човека да не получи своя подарък при теглене на подарък от шапка е:

```
probability.derangements(10 + fnY)
```

```
## [1] 0.3678794
```

б) Емперична вероятност никой да не получи своя подарък на база 10 000 опита:

```
has.fix.point = function(perm) {
  n = length(perm)
  for(i in 1:n) {
    if(perm[i] == i)
      return(TRUE)
  }
  return(FALSE)
}

emperical.probability.derangements = function(n) {
  nofix = 0
  for(k in 1:10000) {
    if(!has.fix.point(sample(n)))
      nofix = nofix + 1
  }
  nofix / 10000
}
emperical.probability.derangements(10 + fnY)
```

```
## [1] 0.3687
```

в) Математическо очакване за броя хора получили своя подарък

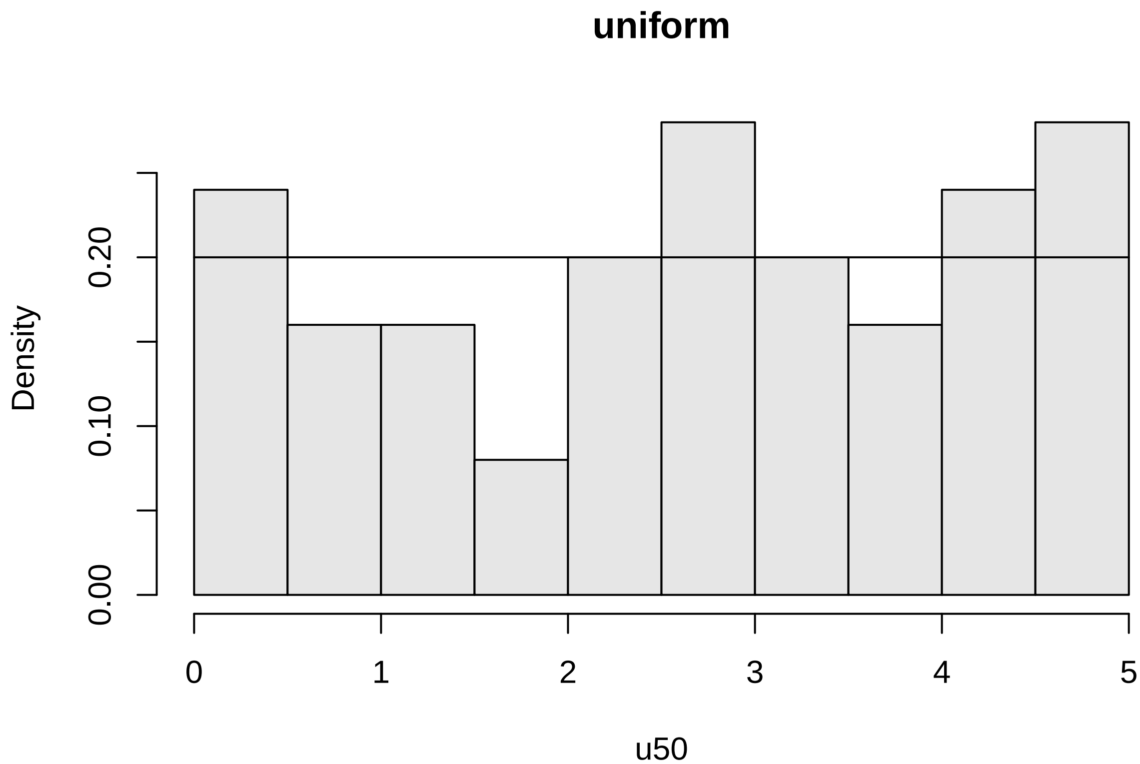
```
expectation.derangements = function(n) {
  s = 0
  for(k in 0:n) {
    s = s + (k * 1 / factorial(k) * probability.derangements(n - k))
  }
  s
}

expectation.derangements(10 + fnY)
```

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

Задача 3.

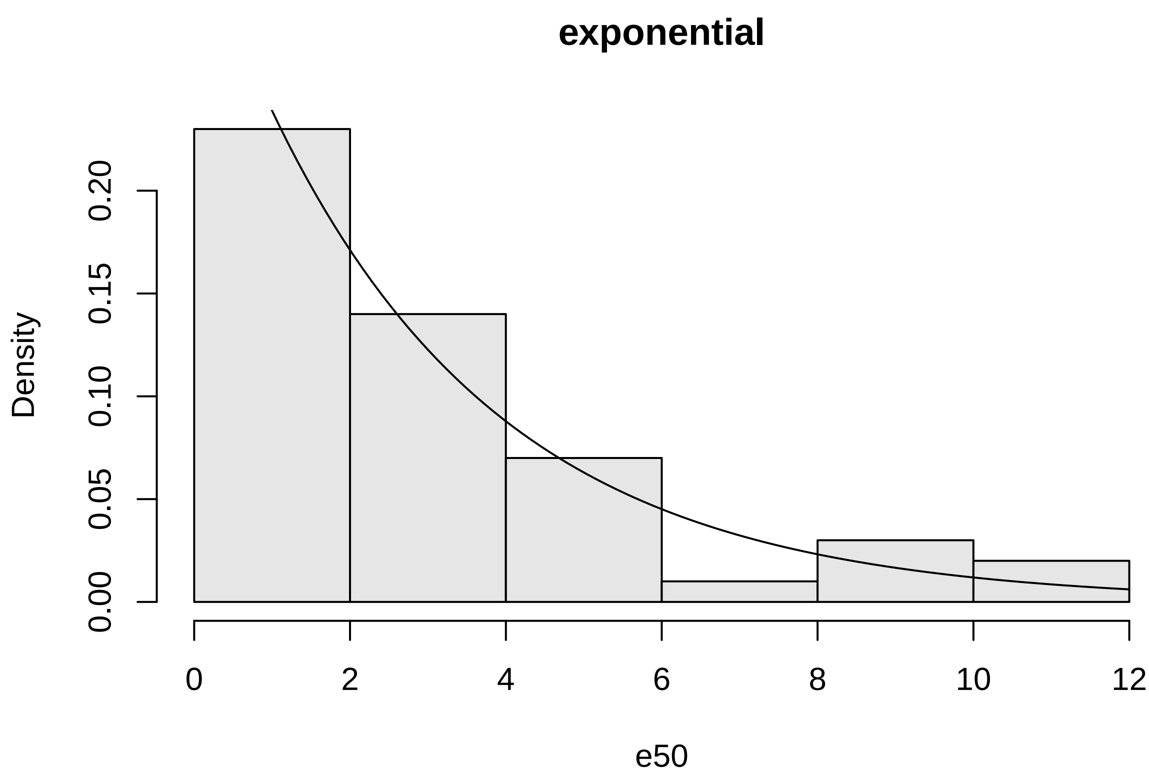
50 случайни наблюдения с равномерно разпределение в интервала [0, 4 + 1]:

```
u50 = runif(50, 0, fnX + 1)
hist(u50, probability=TRUE, col=gray(.9), main="uniform")
curve(dunif(x, 0, fnX + 1), add=TRUE)
```



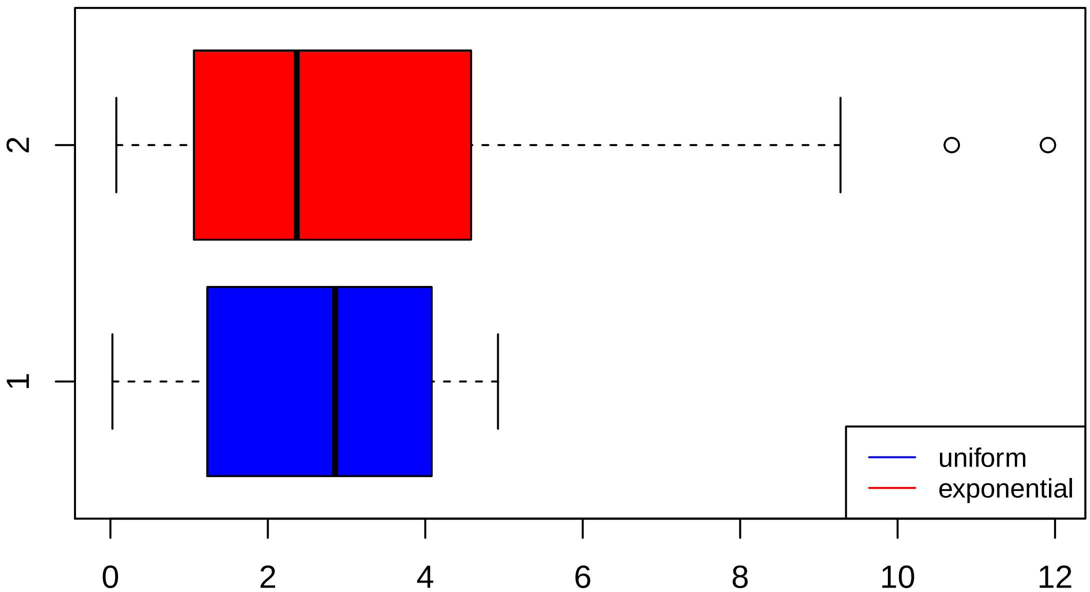
50 случайни наблюдения генерирани с експоненциално разпределение с параметър $1 / (2 + 1)$:

```
e50 = rexp(50, 1 / (fnY + 1))
hist(e50, probability=TRUE, col=gray(.9), main="exponential")
curve(dexp(x, 1 / (fnY + 1)), add=TRUE)
```



a) Boxplot на двете извадки

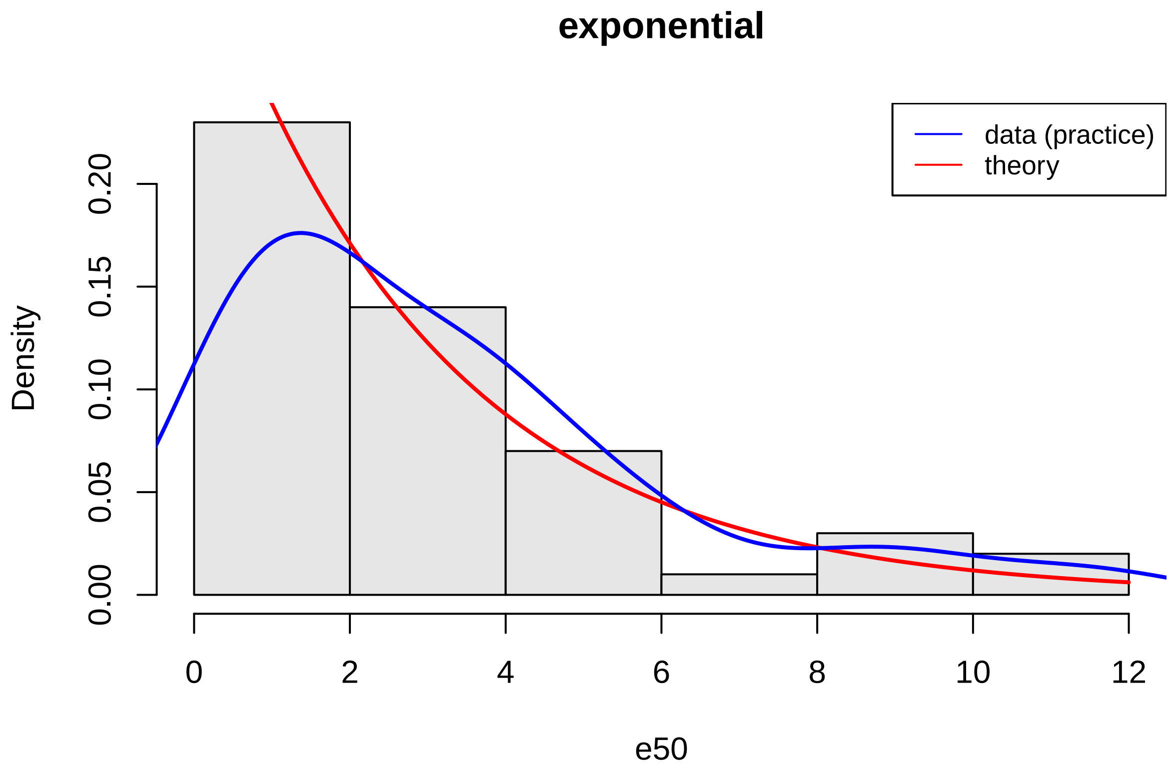
```
boxplot(u50, e50, horizontal=TRUE, col=c("blue","red"))
legend("bottomright", legend=c("uniform","exponential"), lty=1, col=c("blue","red"), cex=0.8)
```



Извод: Стандартното отклонение за равномерно разпределената извадка е по-малко от това на експоненциалната, също така във втората има “outlier” наблюдения, докато при първата няма.

б) Хистограма на експоненциалната извадка с теоретичната плътност и плътността построена по данните

```
hist(e50, probability=TRUE, col=gray(.9), main="exponential")
curve(dexp(x, 1 / (fnY + 1)), col="red", lwd=2, add=TRUE)
lines(density(e50), col="blue", lwd=2)
legend("topright", legend=c("data (practice)","theory"), lty=1, col=c("blue","red"), cex=0.8)
```



Теоретичната плътност е в червено, а тази на данните в синьо.