

lab01

Патракова Екатерина, РИ-440004

1 октября 2017 г

Задание 1 Сгенерируйте 200 случайных значений из стандартного экспоненциального распределения и сохраните в вектор exp.1. Найдите среднее и стандартное отклонение из данных этого вектора.

```
exp.1<-rexp(n=200)
exp.1.mean<-mean(exp.1)
exp.1.sd<-sd(exp.1)
```

Повторите, используя следующие значения параметра rate: 0.1, 0.5, 5, 10, и сохраните результаты в векторы: exp.0.1, exp.0.5, exp.5, exp.10.

```
exp.0.1<-rexp(n=200, rate=0.1)
exp.0.1.mean = mean(exp.0.1)
exp.0.1.sd = sd(exp.0.1)

exp.0.5<-rexp(n=200, rate=0.5)
exp.0.5.mean = mean(exp.0.5)
exp.0.5.sd = sd(exp.0.5)

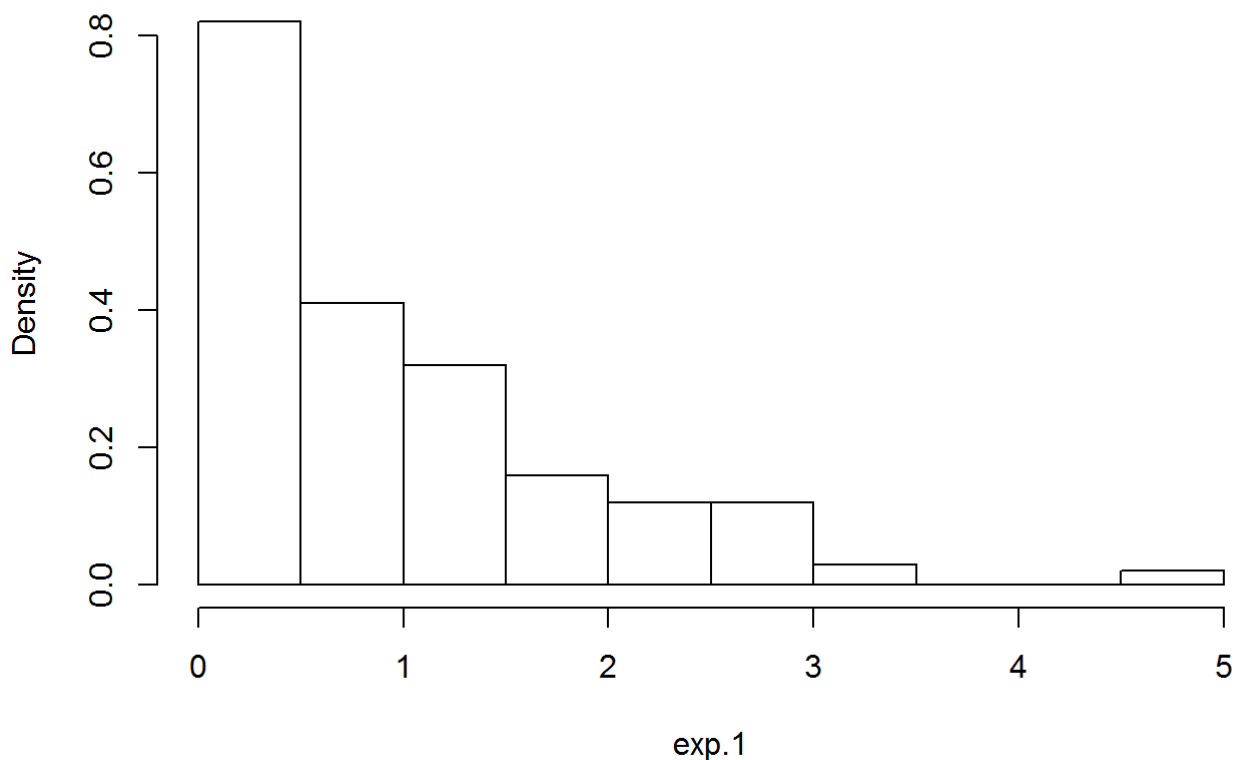
exp.5<-rexp(n=200, rate=5)
exp.5.mean = mean(exp.5)
exp.5.sd = sd(exp.5)

exp.10<-rexp(n=200, rate=10)
exp.10.mean = mean(exp.10)
exp.10.sd = sd(exp.10)
```

Используйте функцию hist() для построения гистограммы полученного ранее стандартного распределения.

```
hist(exp.1, main = "Гистограмма стандартного экс. распределения", prob = TRUE)
```

Гистограмма стандартного экс. распределения



Полученная гистограмма близка к виду стандартного экспоненциального распределения. С увеличением количества наблюдений, распределение становится более точным.

Используйте функцию `plot()` с тем же самым вектором для построения графика значений в порядке следования.

```
plot(exp.1, main="График значений exp.1", prob = TRUE)
```

```
## Warning in plot.window(...): "prob" -- не графический параметр
```

```
## Warning in plot.xy(xy, type, ...): "prob" -- не графический параметр
```

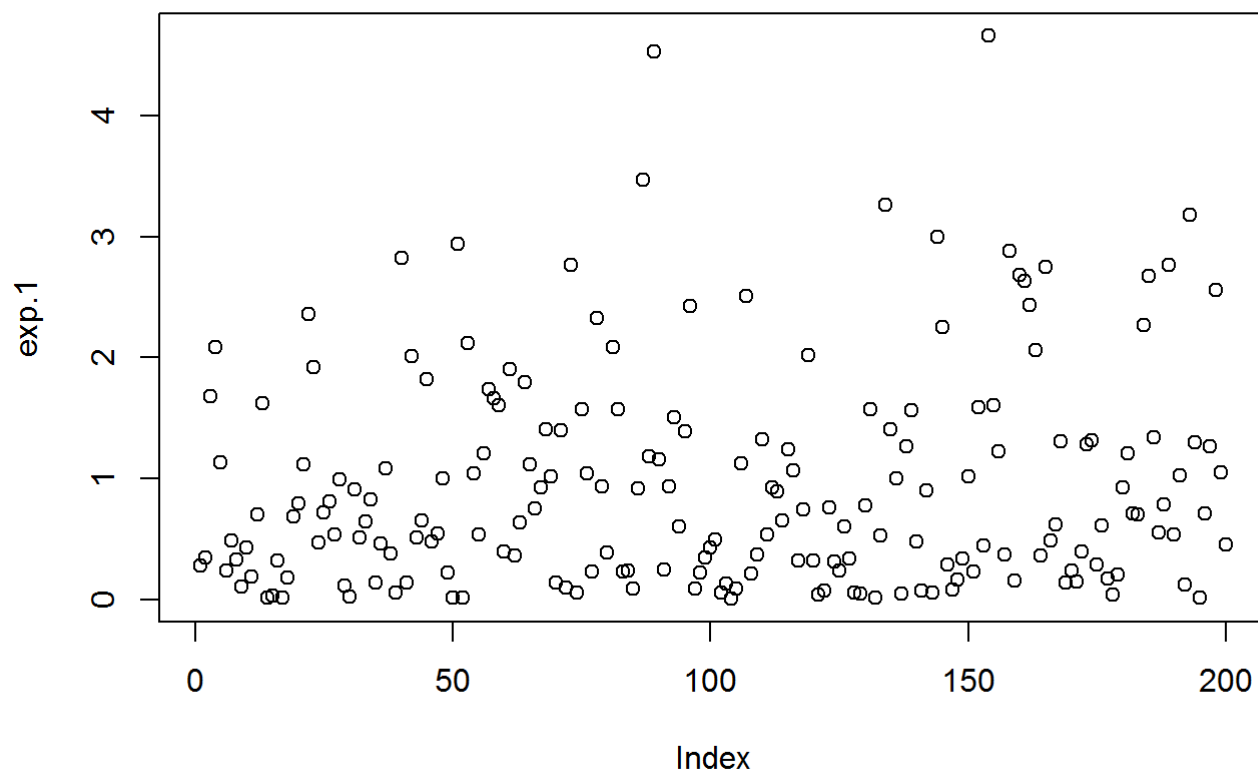
```
## Warning in axis(side = side, at = at, labels = labels, ...): "prob" -- не  
## графический параметр
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "prob" -- не  
## графический параметр
```

```
## Warning in box(...): "prob" -- не графический параметр
```

```
## Warning in title(...): "prob" -- не графический параметр
```

График значений exp.1



Используйте функцию `plot()` с любыми двумя полученными ранее векторами, чтобы получить диаграмму рассеяния (scatterplot) этих векторов относительно друг друга.

```
plot(exp.0.1, exp.10, main="Диаграмма рассеяния exp.0.1 и exp.10", prob = TRUE)
```

```
## Warning in plot.window(...): "prob" -- не графический параметр
```

```
## Warning in plot.xy(xy, type, ...): "prob" -- не графический параметр
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "prob" -- не  
## графический параметр
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "prob" -- не  
## графический параметр
```

```
## Warning in box(...): "prob" -- не графический параметр
```

```
## Warning in title(...): "prob" -- не графический параметр
```

Диаграмма рассеяния exp.0.1 и exp.10

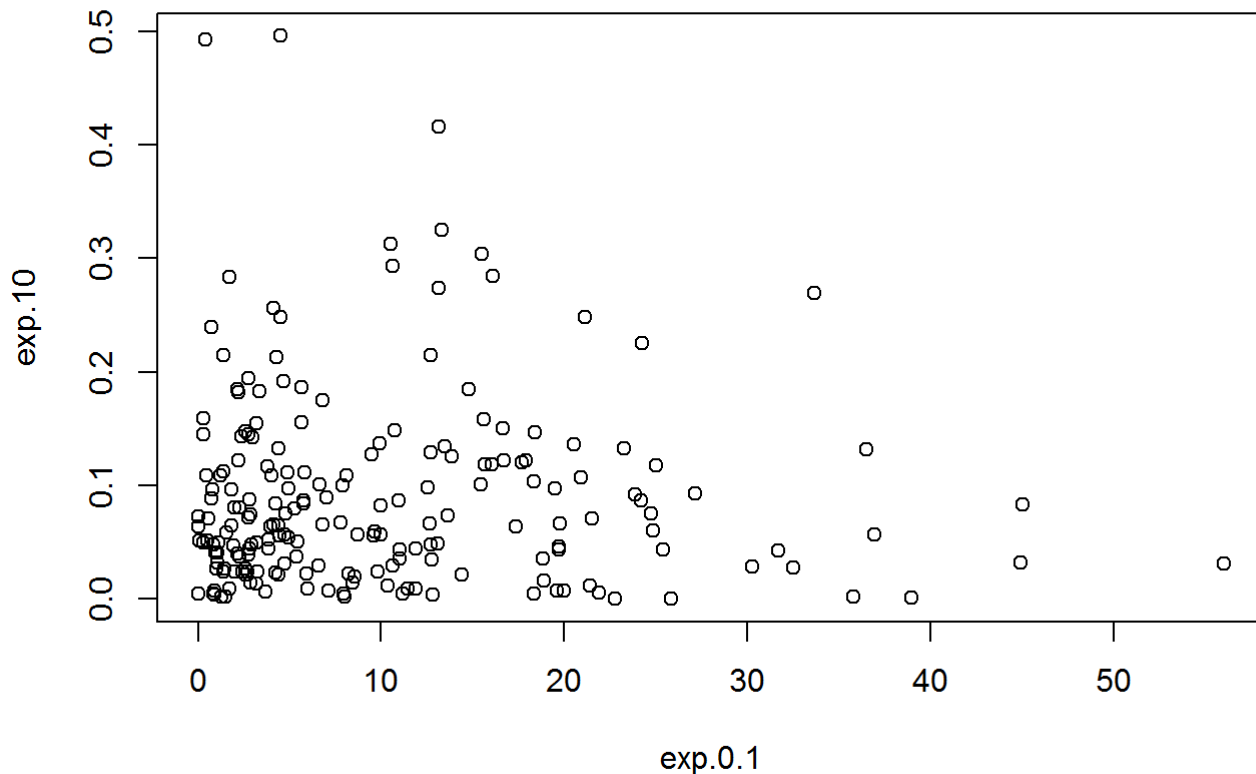


Диаграмма рассеяния показывает, что между рассматриваемыми векторами нет линейной зависимости.

Теперь сравним свойства всех наших векторов. Для начала создадим вектор, содержащий средние значения каждого из 5 полученных ранее распределений в порядке их создания. Полученный вектор сохраним в переменную `exp.means`.

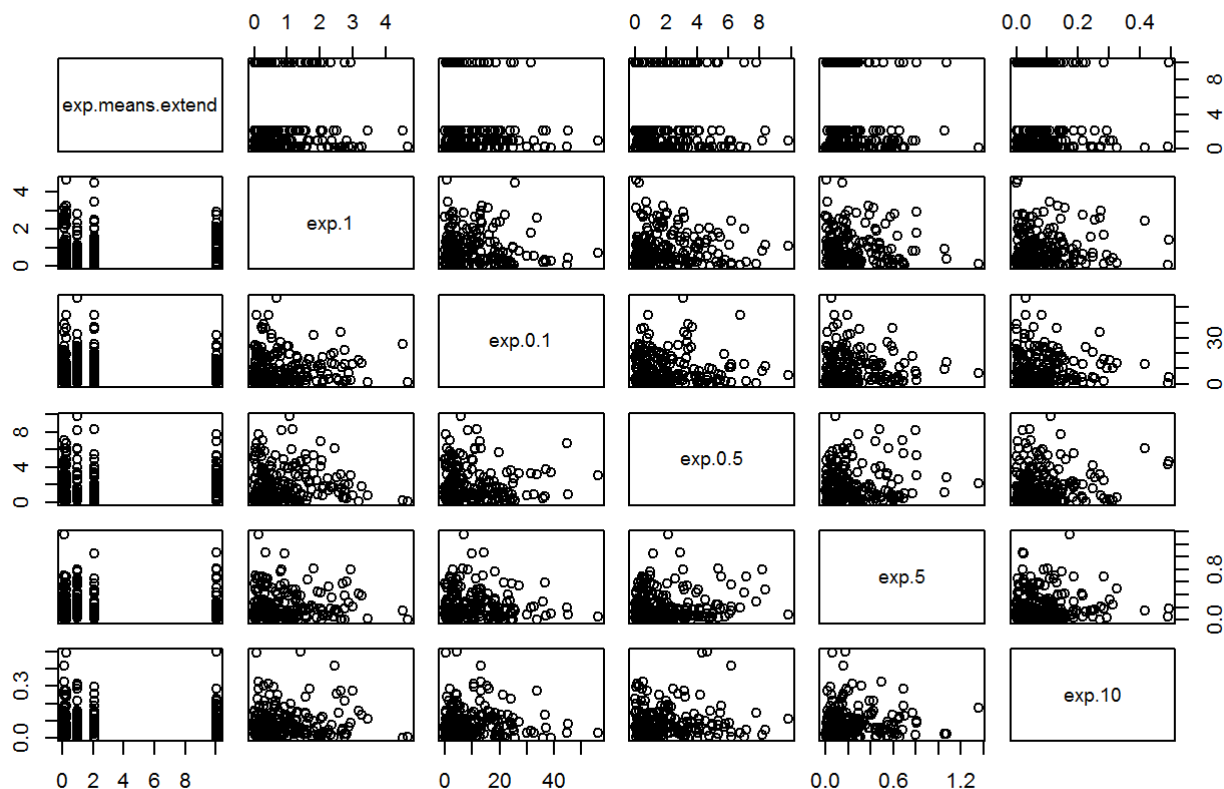
```
exp.means <- c(exp.1.mean, exp.0.1.mean, exp.0.5.mean, exp.5.mean, exp.10.mean)
exp.sds <- c(exp.1.sd, exp.0.1.sd, exp.0.5.sd, exp.5.sd, exp.10.sd)
```

Далее необходимо создать следующие диаграммы рассеяния: Вектор с пятью средними относительно каждого из пяти векторов с различными значениями.

```
#Для выравнивания размерностей векторов расширяем вектор exp.means

exp.means.extend <- rep(exp.means,each=40);
pairs(~exp.means.extend+exp.1+exp.0.1+exp.0.5+exp.5+exp.10, main="Матрица рассеяния exp.sd")
```

Матрица рассеяния exp.sd

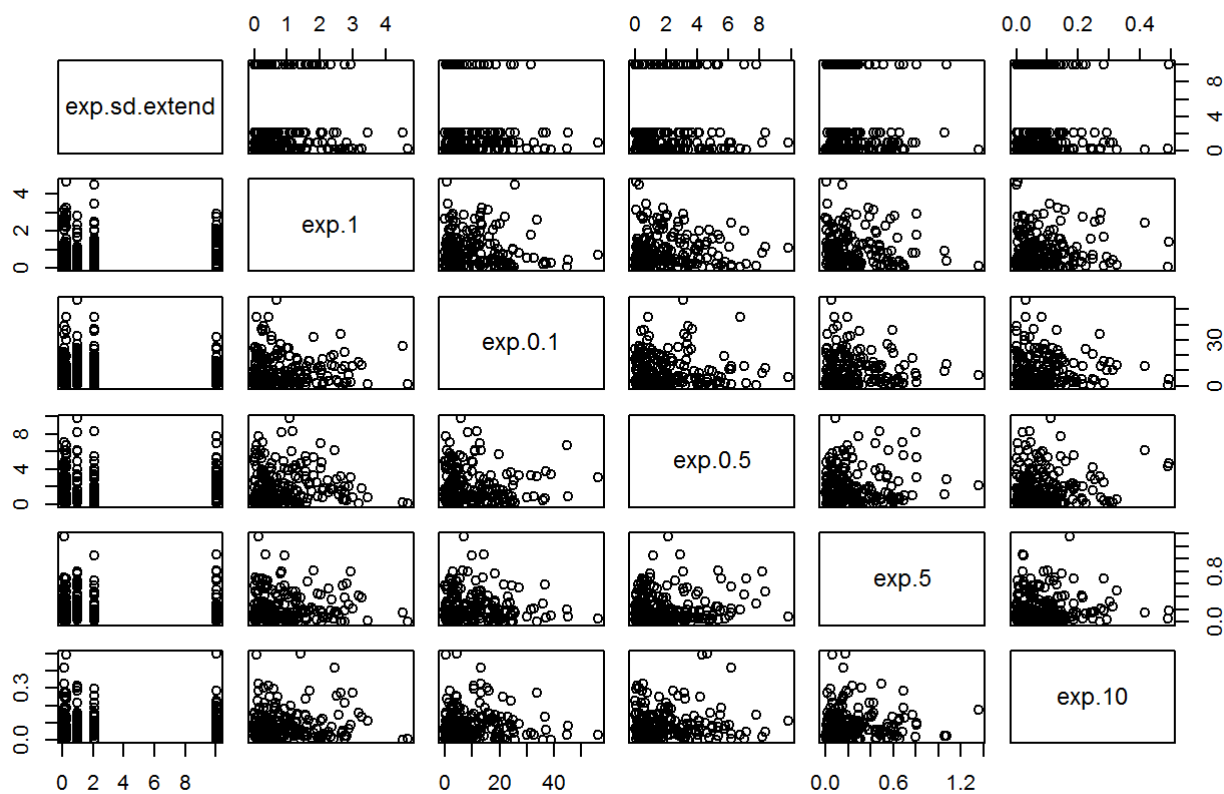


Диаграммы показывают, что между вектором средних и векторами распределений с различными ?? не наблюдается линейная связь.

Вектор с пятью стандартными отклонениями относительно каждого из пяти векторов с различными значениями.

```
exp.sd.extend <- rep(exp.means,each=40);
pairs(~exp.sd.extend+exp.1+exp.0.1+exp.0.5+exp.5+exp.10,main="Матрица рассеяния exp.sd")
```

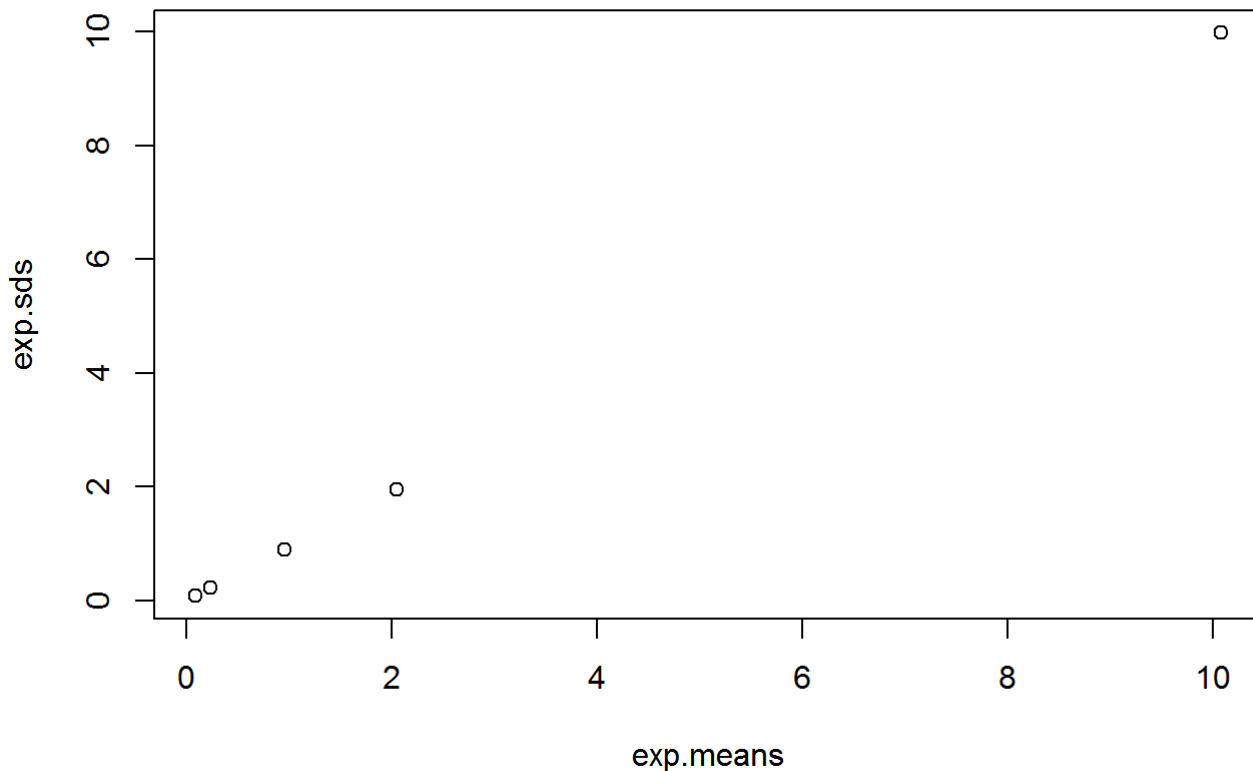
Матрица рассеяния exp.sd



Диаграммы показывают, что между вектором стандартных отклонений и векторами распределений с различными ?? не наблюдается линейная связь.

Вектор с пятью средними относительно вектора с пятью стандартными отклонениями.

```
plot(exp.means, exp.sds)
```



Между вектором средних и вектором стандартных отклонений наблюдается линейная зависимость.

Задание 2

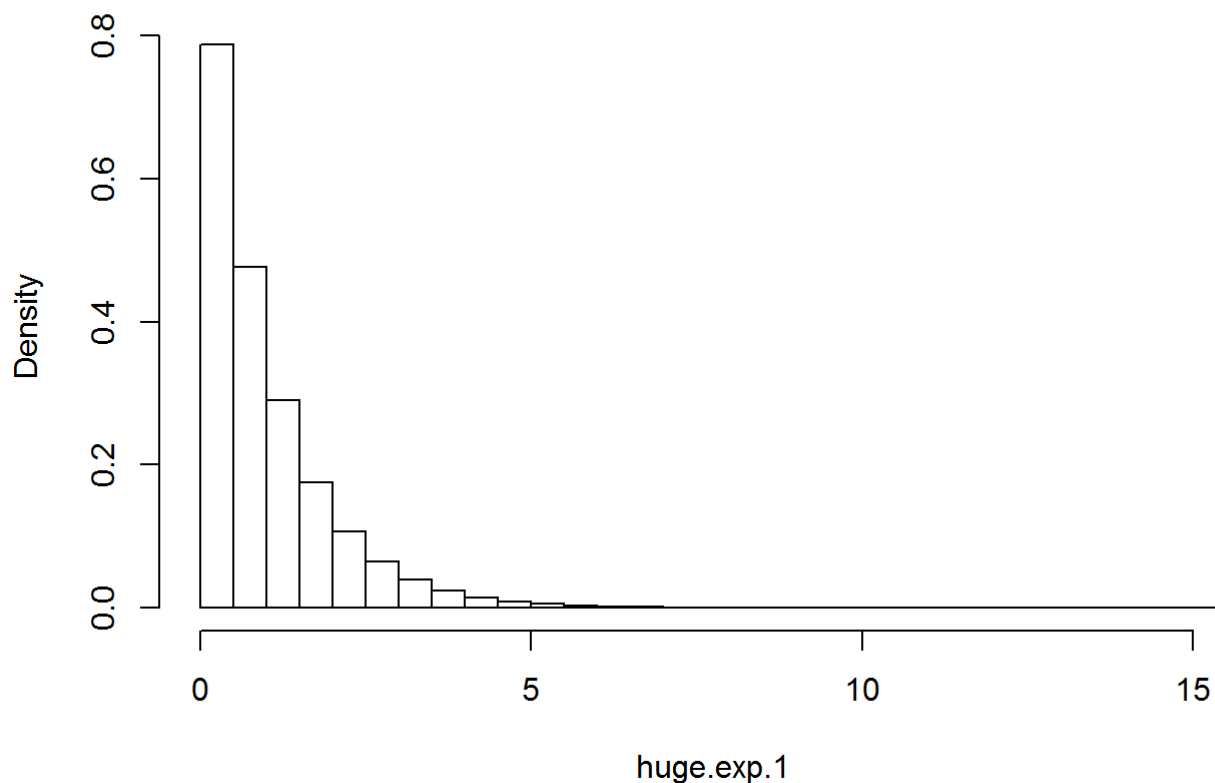
Для демонстрации сгенерируйте 1100000 (1 млн. 100 тысяч) чисел из стандартного экспоненциального распределения и сохраните их в вектор `huge.exp.1`. Получите среднее и стандартное отклонение.

```
huge.exp.1<-rexp(n=1100000)
huge.exp.1.mean<-mean(huge.exp.1)
huge.exp.1.sd<-sd(huge.exp.1)
```

Получите гистограмму `huge.exp.1`. Она соответствует функции $1 - e^{-x}$? Должна ли?

```
hist(huge.exp.1, main = "Гистограмма стандартного экспоненциального распределения", prob = TRUE)
```

Гистограмма стандартного экспоненциального распределения



Получите среднее всех чисел из `huge.exp.1` строго больших 1. Для проверки выборки может потребоваться новый вектор.

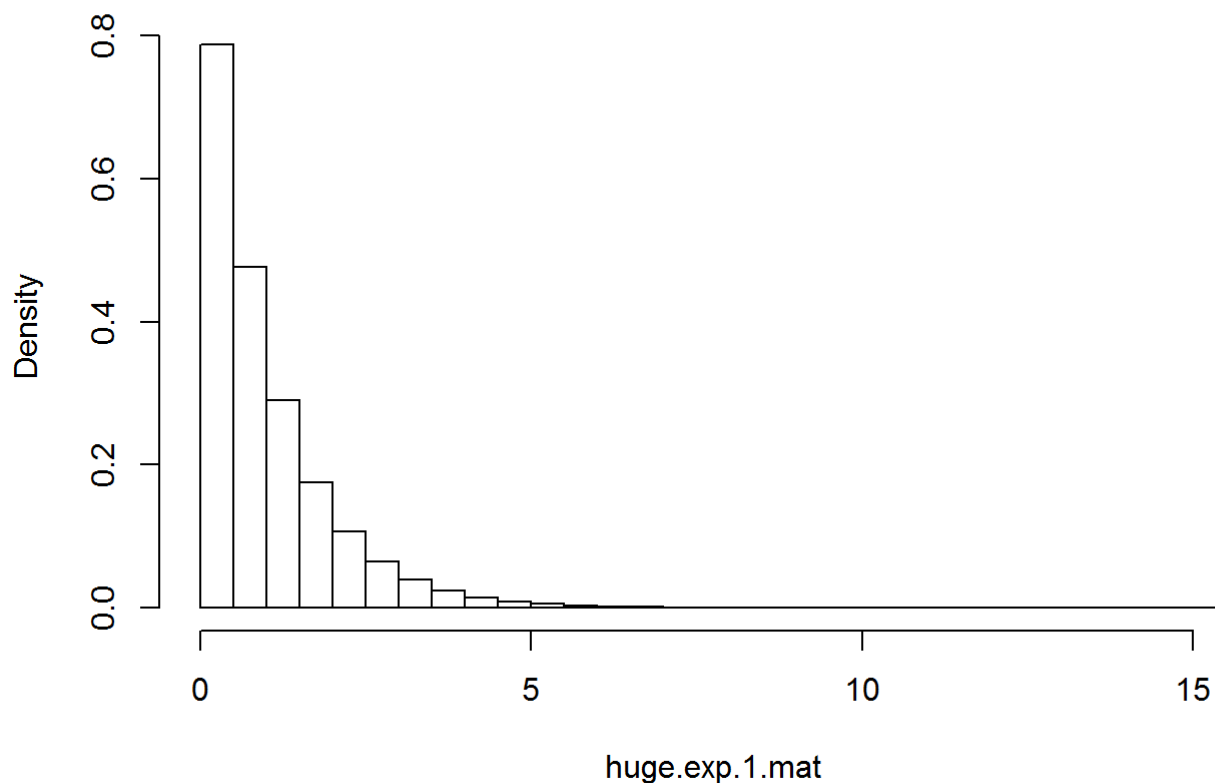
```
huge.exp.more1<-huge.exp.1[huge.exp.1>1]
huge.exp.mean.more1<-mean(huge.exp.more1)
```

Гистограмма должна соответствовать не функции $1 - e^{-x}$ (функции распределения), а функции плотности распределения $x \cdot e^{-x}$.

Получите матрицу `huge.exp.1.mat`, содержащую 1100 строк и 1000 колонок. Постройте гистограмму, используя полученную матрицу. Что можно сказать о полученных данных?

```
huge.exp.1.mat<- matrix( huge.exp.1,nrow=1100)
hist(huge.exp.1.mat, prob = TRUE, main = "Гистограмма стандартного экспоненциального распределения")
```


Гистограмма стандартного экспоненциального распределения



Полученная гистограмма более близка к виду стандартного экспоненциального распределения, чем гистограмма для 200 значений. С увеличением количества наблюдений, распределение становится более точным.

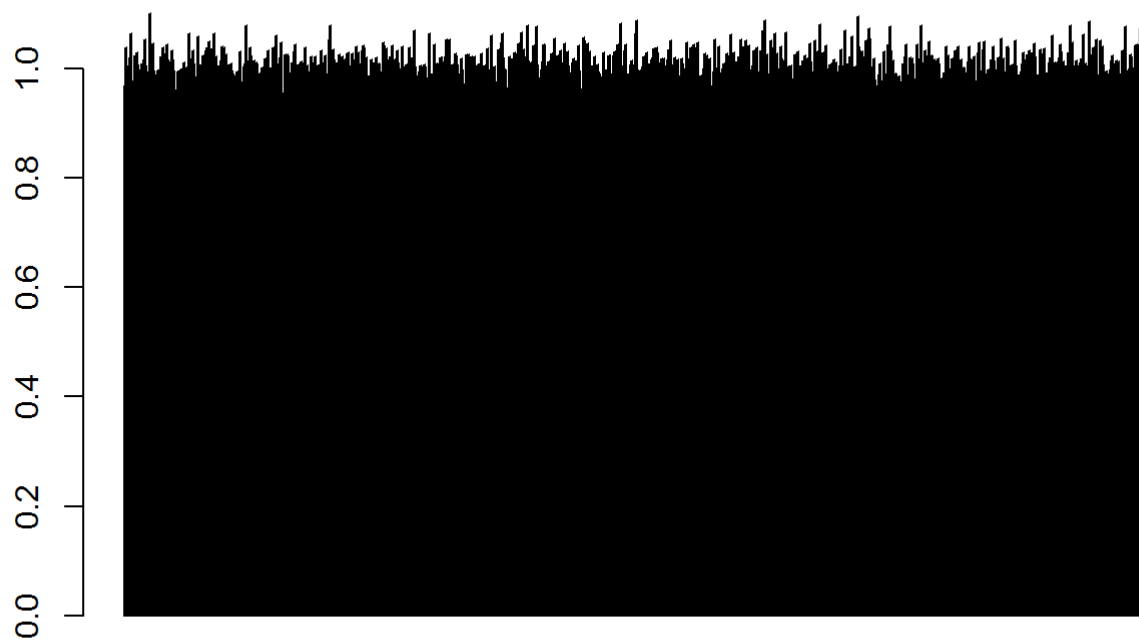
Рассчитайте среднее 137-й колонки из huge.exp.1.mat.

```
mean.137<-mean(huge.exp.1.mat[, 137])
```

Найдите среднее всей 1000 колонок и сохраните их в вектор. Отобразите диаграмму средних значений колонок. Объясните почему форма диаграммы отличается от виденных ранее.

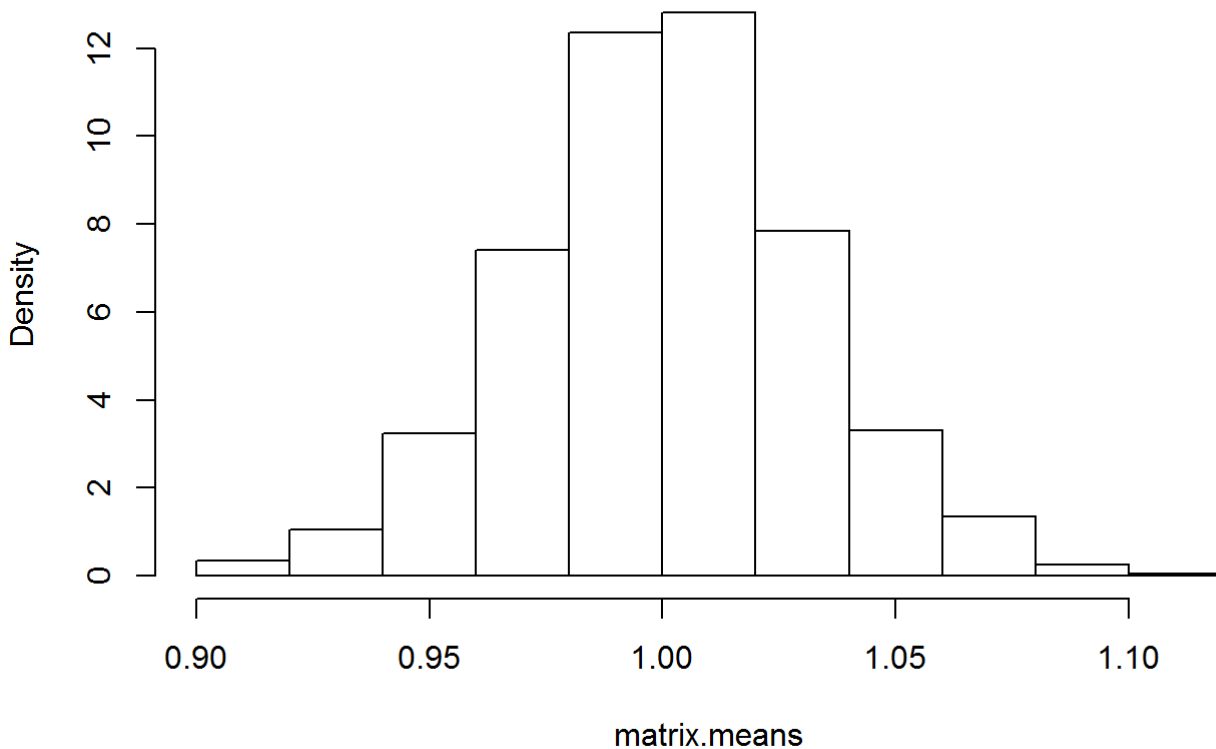
Среднее значение экспоненциального распределения равно $1/\lambda$, в данном случае $\lambda=1$, следовательно среднее значение каждой колонки стремится к 1.

```
matrix.means<-colMeans(huge.exp.1.mat)  
barplot(matrix.means)
```



```
hist(matrix.means, prob = TRUE, main = "Гистограмма средних значений")
```

Гистограмма средних значений



Полученная гистограмма близка к нормальному распределению, что объясняется при помощи Центральной предельной теоремы. ЦПТ утверждает, что сумма большого количества независимых случайных величин имеет распределение, близкое к нормальному. Так как при нахождении среднего значения по колонкам применяется суммирование всех элементов этих колонок, полученные средние значения удовлетворяют данной теореме и образуют вектор нормально распределенных случайных величин.

Рассчитайте квадрат каждого значения из `huge.exp.1` и найдите среднее и стандартное отклонение полученного вектора квадратов. Объясните отличие от соответствующих показателей вектора `huge.exp.1`. Вам может помочь формула, по которой R рассчитывает стандартное отклонение: $\sqrt{\text{sum}((x - \text{mean}(x))^2) / (n - 1))}$.

```
huge.exp.1.sqrt<-sqrt(huge.exp.1)
huge.exp.sqrt.mean<-mean(huge.exp.1.sqrt)
huge.exp.sqrt.sd<-sd(huge.exp.1.sqrt)
```

Среднее значение экспоненциального распределения равняется $1/\lambda$, а в данном случае $\lambda=1$. При распределении квадратов исходных значений среднее значение увеличивается почти в два раза. В формуле среднего сумма исходных значений находится в числителе, то есть при увеличении исходных λ значений возрастает и среднее. Стандартное отклонение тоже увеличивается, причем соотношение стандартных отклонений квадратов исходных значений и самих этих значений уже больше, чем 2. Это связано с тем, что присутствует возведение в квадрат.