Практикум на ЭВМ
МС ВМК 5 семестр
Задание 6
Роор Даниил

1. Ознакомиться с процедурой бутстреп (bootstrap).

```
library(boot)

norm100_1 = rnorm(100, 0, 1)
```

Оценим матожидание по выборочному среднему. Для этого возьмем 100000 подвыборок нашей выборки и для каждой посчитаем среднее

```
Boot <- boot(norm100_1,
             #index - массив индексов выбранных для
             #подвыборки элементов. Берем по этим
             #элементам среднее
             function(sample, index) mean(sample[index]),
             R = 100000)
```

После создания объекта boot поле Boot$t содержит массив из 100000 выборочных средних, посчитанных по подвыборкам нашей выборки. Усредним эти средние:

```
mean(Boot$t)

Output:
0.05258731
```

Посмотрим, насколько отличается этот результат от среднего исходной выборки:

```
mean(Boot$t) - mean(norm100_1)

Output:
-0.0001595471
```

Построим доверительные интервалы для матожидания:

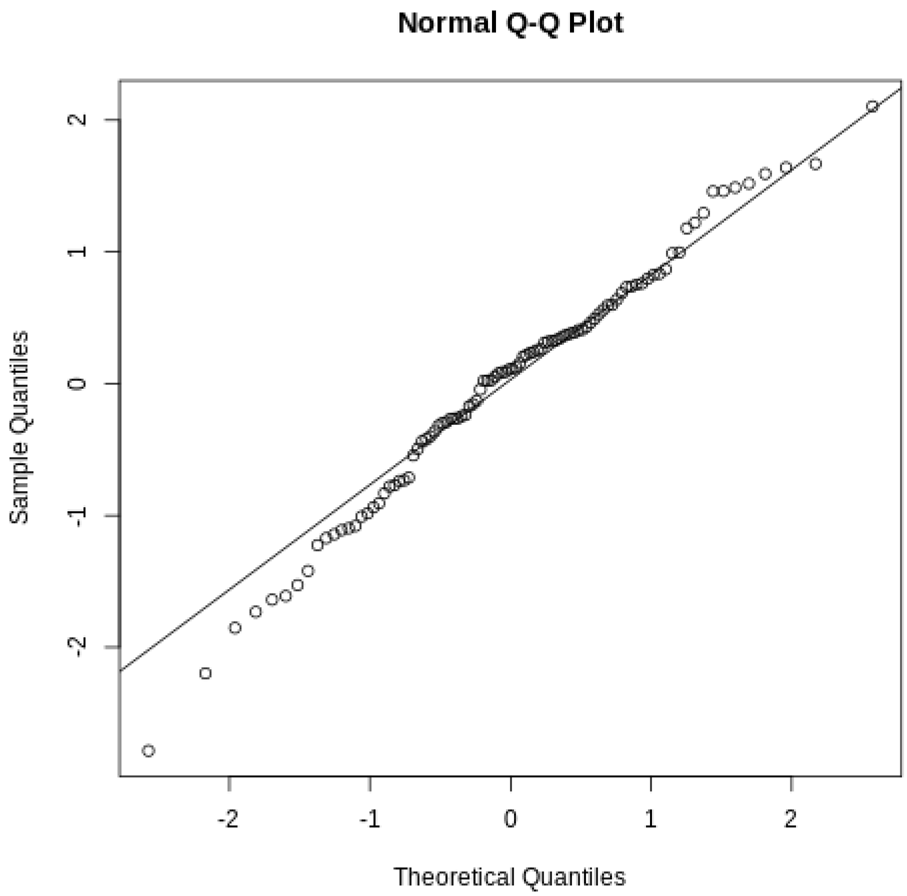
```
boot.ci(Boot, index = 1)

Intervals :
Level      Normal              Basic
95%   (-0.1557,  0.2615 )   (-0.1556,  0.2609 )

Level     Percentile            BCa
95%   (-0.1554,  0.2611 )   (-0.1540,  0.2627 )
```

2. Сгенерировать данные из нормального распределения с различными параметрами и провести анализ с помощью графиков квантилей, метода огибающих, а также стандартных процедур проверки гипотез о нормальности, рассмотренных на семинаре (6 тестов). Рассмотреть выборки малого (не более 50-100 элементов) и умеренного (1000-5000 наблюдений) объемов.

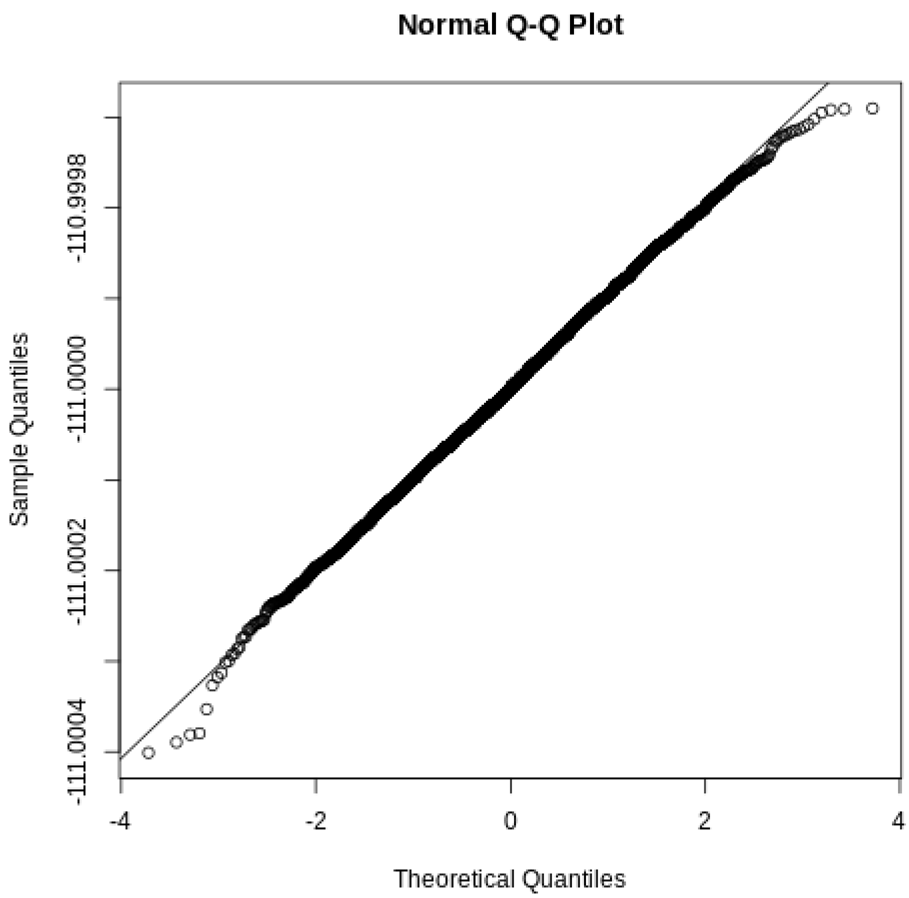
```
norm100_2 = rnorm(100, -10, 7)
norm5000_1 = rnorm(5000, 9, 123)
norm5000_2 = rnorm(5000, -111, 0.0001)
```

Нарисуем график теоретических и выборочных квантилей.

```
png(file = 'qqplot1.png')
qqnorm(norm100_1)
qqline(norm100_1)
dev.off()
```




```
png(file = 'qqplot2.png')
qqnorm(norm5000_2)
qqline(norm5000_2)
dev.off()
```



Получаем по графикам нормальные распределения. Теперь проведем стандартные тесты на нормальность

```
shapiro.test(norm100_1)


Output:

Shapiro-Wilk normality test

data:  norm100_1
W = 0.97338, p-value = 0.04029
```

Отвергаем гипотезу о нормальности

```
ad.test(norm100_1)


Output:
Anderson-Darling normality test

data:  norm100_1
A = 0.41979, p-value = 0.32
```

Принимаем гипотезу о нормальности

```
cvm.test(norm100_1)
```

```
Output:
     Cramer-von Mises normality test

data:  norm100_1
W = 0.045533, p-value = 0.5768
```

Принимаем гипотезу о нормальности

```
lillie.test(norm100_1)
```

```
Output:

     Lilliefors (Kolmogorov-Smirnov) normality test

data:  norm100_1
D = 0.058276, p-value = 0.5525
```

Принимаем гипотезу о нормальности

```
sf.test(norm100_1)
```

```
Output:

     Shapiro-Francia normality test

data:  norm100_1
W = 0.97435, p-value = 0.04691
```

Отвергаем гипотезу о нормальности

```
shapiro.test(norm100_2)
```

```
Output:
     Shapiro-Wilk normality test

data:  norm100_2
W = 0.99224, p-value = 0.8381
```

Принимаем гипотезу о нормальности

```
ad.test(norm100_2)
```

```
Output:
     Anderson-Darling normality test

data:  norm100_2
A = 0.2395, p-value = 0.7725
```

Принимаем гипотезу о нормальности

```
cvm.test(norm100_2)
```

```
Output:
     Cramer-von Mises normality test

data:  norm100_2
W = 0.036716, p-value = 0.7384
```

Принимаем гипотезу о нормальности

```
lillie.test(norm100_2)
```

```
Output:
     Lilliefors (Kolmogorov-Smirnov) normality test

data:  norm100_2
D = 0.050024, p-value = 0.7791
```

Принимаем гипотезу о нормальности

```
sf.test(norm100_2)
```

```
Output:

     Shapiro-Francia normality test

data:  norm100_2

W = 0.99281, p-value = 0.7992
```

Принимаем гипотезу о нормальности

```
shapiro.test(norm5000_1)
```

```
Output:
     Shapiro-Wilk normality test

data:  norm5000_1
W = 0.99956, p-value = 0.318
```

Принимаем гипотезу о нормальности

```
ad.test(norm5000_1)
```

```
Output:

     Anderson-Darling normality test

data:  norm5000_1
A = 0.48665, p-value = 0.2249
```

Принимаем гипотезу о нормальности

```
cvm.test(norm5000_1)
```

```
Output:
     Cramer-von Mises normality test

data:  norm5000_1
W = 0.090208, p-value = 0.1529
```

Принимаем гипотезу о нормальности

```
lillie.test(norm5000_1)
```

```
Output:
     Lilliefors (Kolmogorov-Smirnov) normality test

data:  norm5000_1
D = 0.010212, p-value = 0.2345
```

Принимаем гипотезу о нормальности

```
sf.test(norm5000_1)
```

```
Output:
     Shapiro-Francia normality test
```

```
data:  norm5000_1

W = 0.99957, p-value = 0.2768
```

Принимаем гипотезу о нормальности

```
shapiro.test(norm5000_2)
```

```
Output:

     Shapiro-Wilk normality test


data:  norm5000_2

W = 0.99928, p-value = 0.04124
```

Отвергаем гипотезу о нормальности

```
ad.test(norm5000_2)
```

```
Output:

Anderson-Darling normality test


data:  norm5000_2

A = 0.52104, p-value = 0.1851
```

Принимаем гипотезу о нормальности

```
cvm.test(norm5000_2)
```

```
Output:
```

```
	Cramer-von Mises normality test


data:  norm5000_2
W = 0.090694, p-value = 0.1507
```

Принимаем гипотезу о нормальности

```
lillie.test(norm5000_2)
```

```
Output:
	Lilliefors (Kolmogorov-Smirnov) normality
	test


data:  norm5000_2
D = 0.01007, p-value = 0.2528
```

Принимаем гипотезу о нормальности

```
sf.test(norm5000_2)
```

```
Output:
	Shapiro-Francia normality test


data:  norm5000_2
W = 0.99933, p-value = 0.04447
```

Отвергаем гипотезу о нормальности

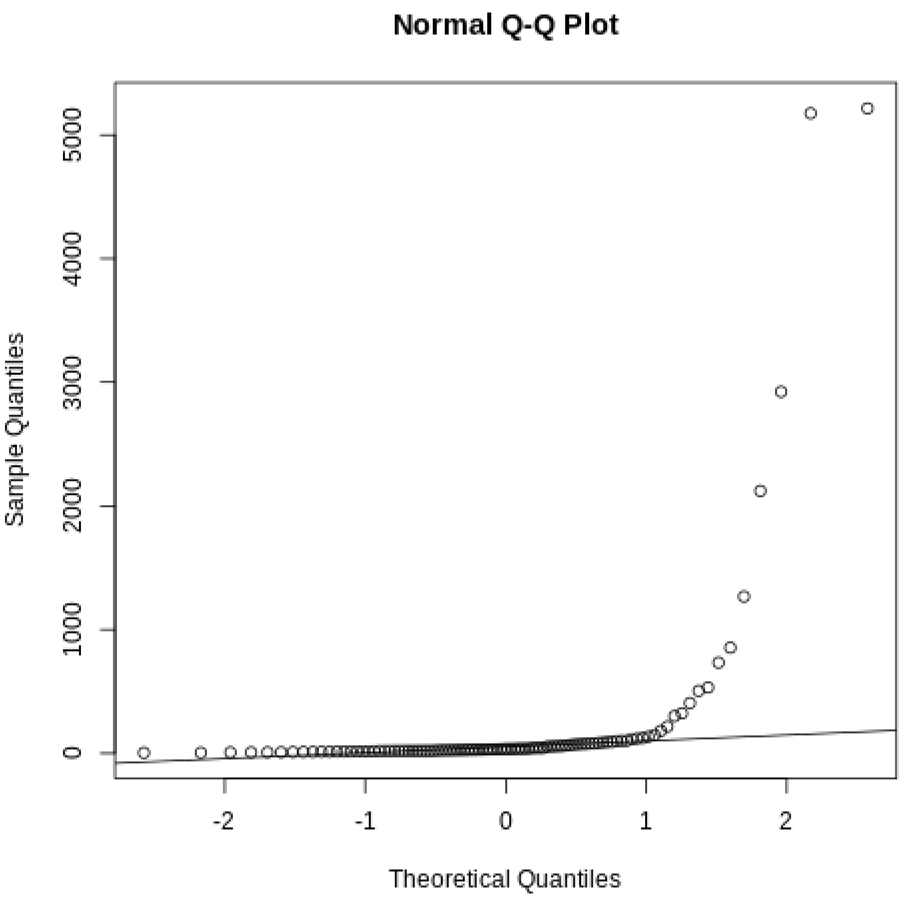
Критерии Шапиро-Уилка и Шапиро-Франчиа оказались слабее остальных, так как дали ошибку на двух выборках. Остальные критерии правильно определили нормальность.

3. Сгенерировать данные из комбинаций, реализованных в R распределений, а затем провести анализ с помощью графиков квантилей, метода огибающих, а также стандартных процедур проверки гипотез о нормальности. Рассмотреть выборки малого и умеренного объемов. Сравнить эффективность методов.

Создадим большую и малую выборки из гамма-распределения со случайными параметрами из экспоненциального и биномиального распределений.

```
rand_rate = rexp(100,5)

rand_shape = rbinom(5000, 10, 0.2)

gamma_small = rgamma(100, 5, rand_rate)

gamma_large = rgamma(100, rand_shape, 0.5)
```

```
png(file = 'qqplot3.png')

qqnorm(gamma_small)

qqline(gamma_small)

dev.off()
```



По графику распределение не нормально.

```
shapiro.test(gamma_small)


Output:

	Shapiro-Wilk normality test


data:  gamma_small

W = 0.39789, p-value < 2.2e-16
```

Отвергаем гипотезу о нормальном распределении

```
ad.test(gamma_small)


Output:

	Anderson-Darling normality test


data:  gamma_small

A = 20.484, p-value < 2.2e-16
```

Отвергаем гипотезу о нормальном распределении

```
cvm.test(gamma_small)


Output:

	Cramer-von Mises normality test


data:  gamma_small

W = 4.1331, p-value = 7.37e-10
```

Отвергаем гипотезу о нормальном распределении

```
lillie.test(gamma_small)
```

```
Output:
	Lilliefors (Kolmogorov-Smirnov) normality test

data:  gamma_small
D = 0.34042, p-value < 2.2e-16
```

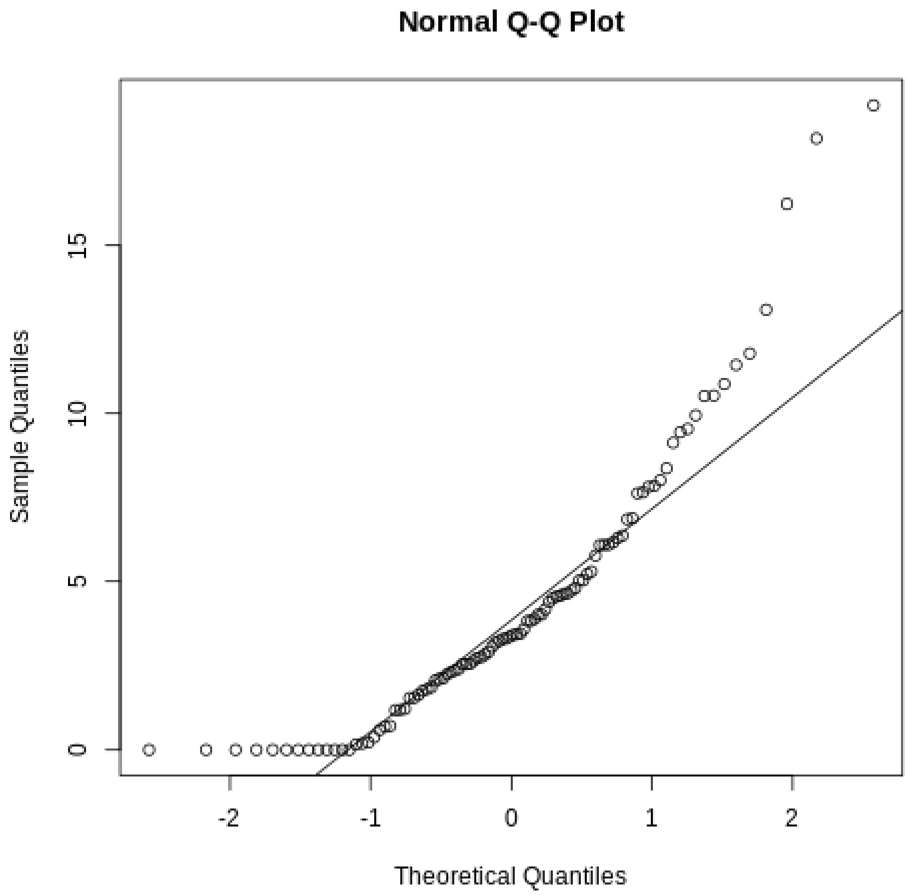
Отвергаем гипотезу о нормальном распределении

```
sf.test(gamma_small)
```

```
Output:
	Shapiro-Francia normality test

data:  gamma_small
W = 0.38608, p-value = 9.225e-16
```

Отвергаем гипотезу о нормальном распределении

```
png(file = 'qqplot4.png')
qqnorm(gamma_large)
qqline(gamma_large)
dev.off()
```




```
shapiro.test(gamma_large)
```

```
Output:

	Shapiro-Wilk normality test


data:  gamma_large
W = 0.88179, p-value = 2.163e-07
```

Отвергаем гипотезу о нормальном распределении

```
ad.test(gamma_large)
```

```
Output:

	Anderson-Darling normality test
```

```
data:  gamma_large

A = 2.8239, p-value = 3.729e-07
```

Отвергаем гипотезу о нормальном распределении

```
cvm.test(gamma_large)
```

```
Output:

     Cramer-von Mises normality test


data:  gamma_large

W = 0.44931, p-value = 8.01e-06
```

Отвергаем гипотезу о нормальном распределении

```
lillie.test(gamma_large)
```

```
Output:

     Lilliefors (Kolmogorov-Smirnov) normality test


data:  gamma_large

D = 0.14723, p-value = 1.423e-05
```

Отвергаем гипотезу о нормальном распределении

```
sf.test(gamma_large)
```

```
Output:

     Shapiro-Francia normality test
```

```
data:  gamma_large

W = 0.8804, p-value = 1.189e-06
```

Отвергаем гипотезу о нормальном распределении

Все критерии правильно определили ненормальность. Исходя из результатов эксперимента можно предположить: в критериях Шапиро-Уилка и Шапиро-Франчиа ошибки первого рода встречаются чаще, чем в других критериях. Ошибки второго рода встречаются с примерно одинаковой частотой.