

Смирнов Иван Сергеевич, 316 группа

Практикум 5 сем, язык R

Для максимальной объективности оценивания данных давафреймы брал те, что обрабатывал на python, просто импортируя их из jupyter notebook в файлы csv, так как приходилось много обрабатывать данные для их исследования.

Весь код на R приложен в отдельных файлах. Здесь только некоторые отрывки.

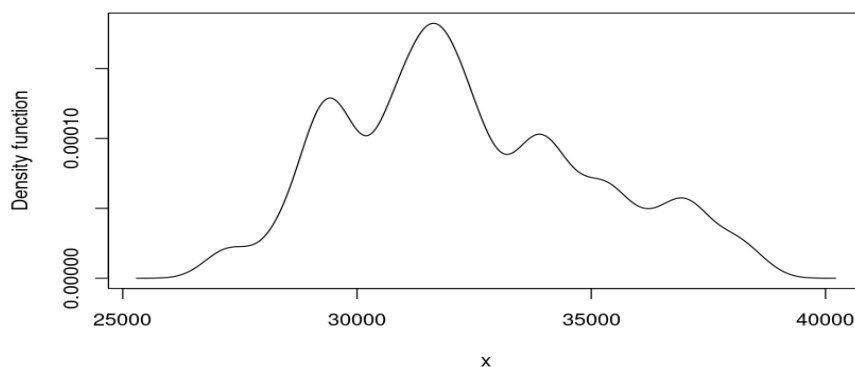
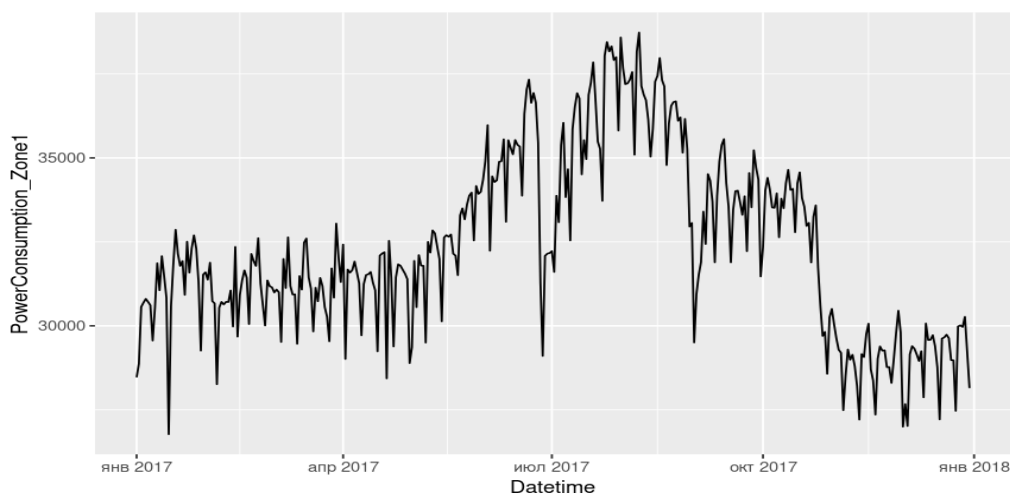
Выводы тоже здесь не приводил, так как они совпадают с теми, что делал на python.

1 Реализовать аппроксимацию распределений данных с помощью ядерных оценок.

```
ggplot(data = data, mapping = aes(x = Datetime, y =  
PowerConsumption_Zone1)) + geom_line()
```

```
dens <- kde(x = data$PowerConsumption_Zone1, h=400)
```

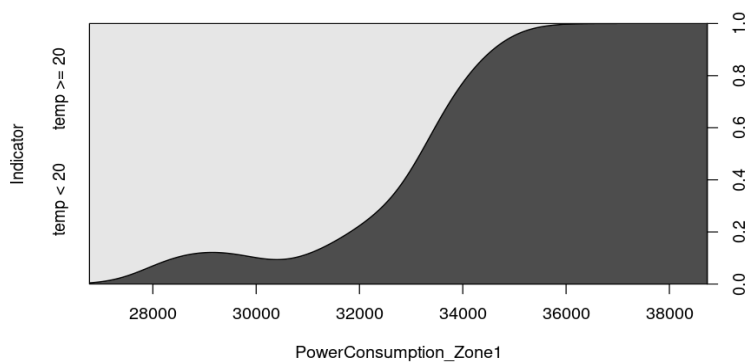
```
plot(dens)
```



2 Реализовать анализ данных с помощью `cdplot`, `dotchart`, `boxplot` и `stripchart`.

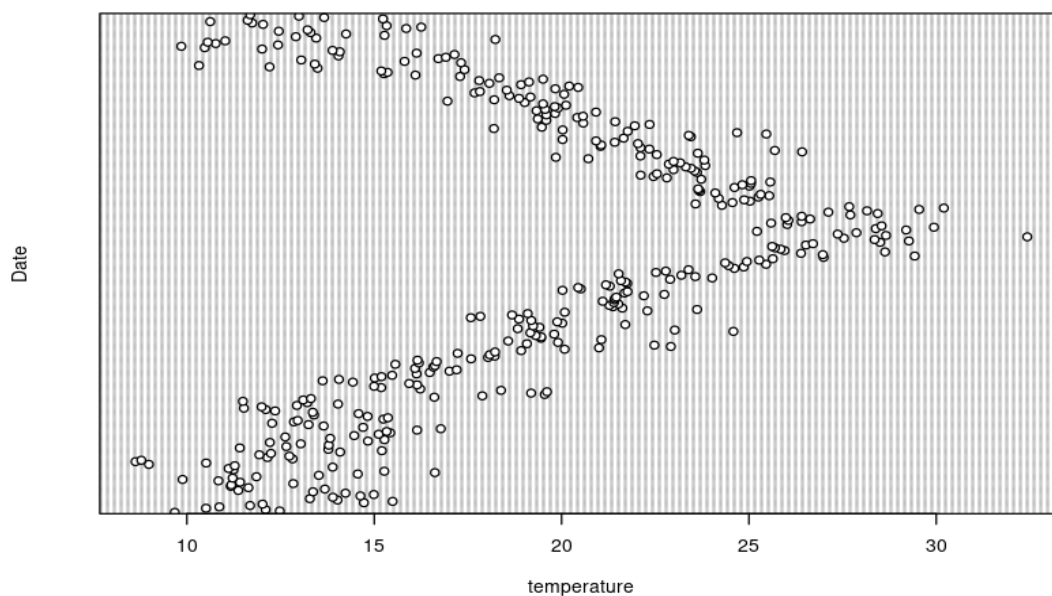
```
# cdplot
```

```
cdplot(Indicator ~ PowerConsumption_Zone1,  
       yaxlabels = c("temp >= 20", " temp < 20"),  
       data = data)
```



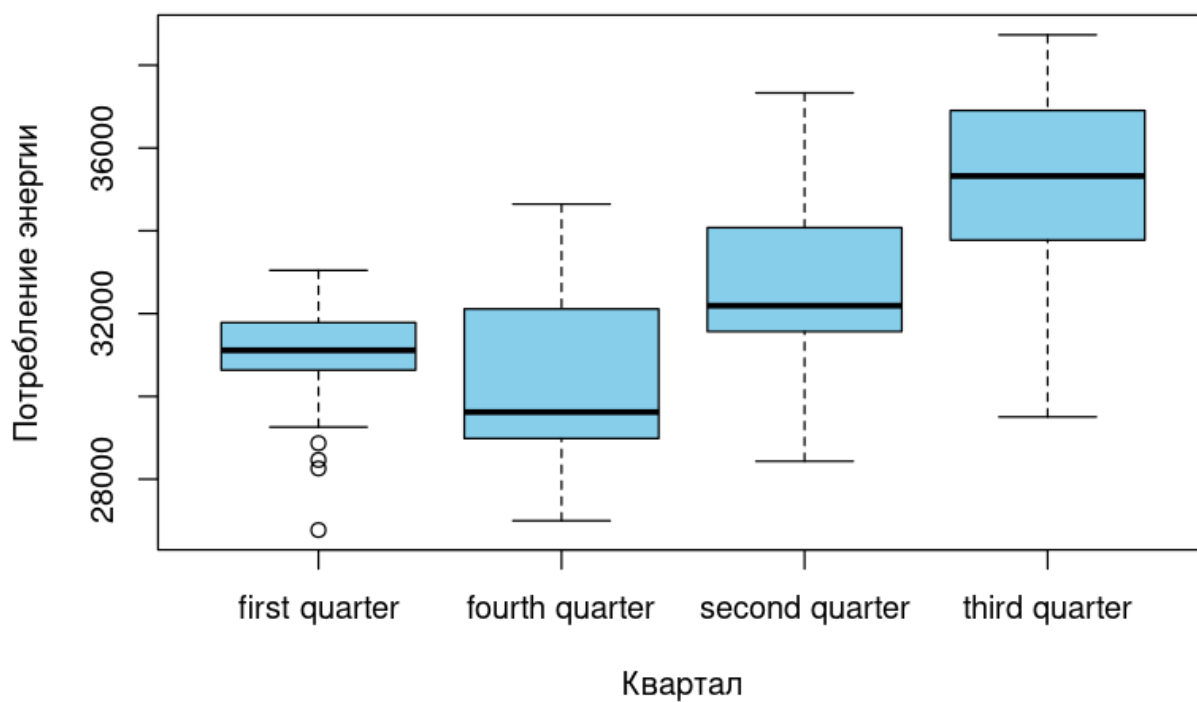
```
# dotchart
```

```
dotchart(data$Temperature,  
         ylab = "Date", xlab="temperature", cex = 0.7)
```



```
#boxplot
```

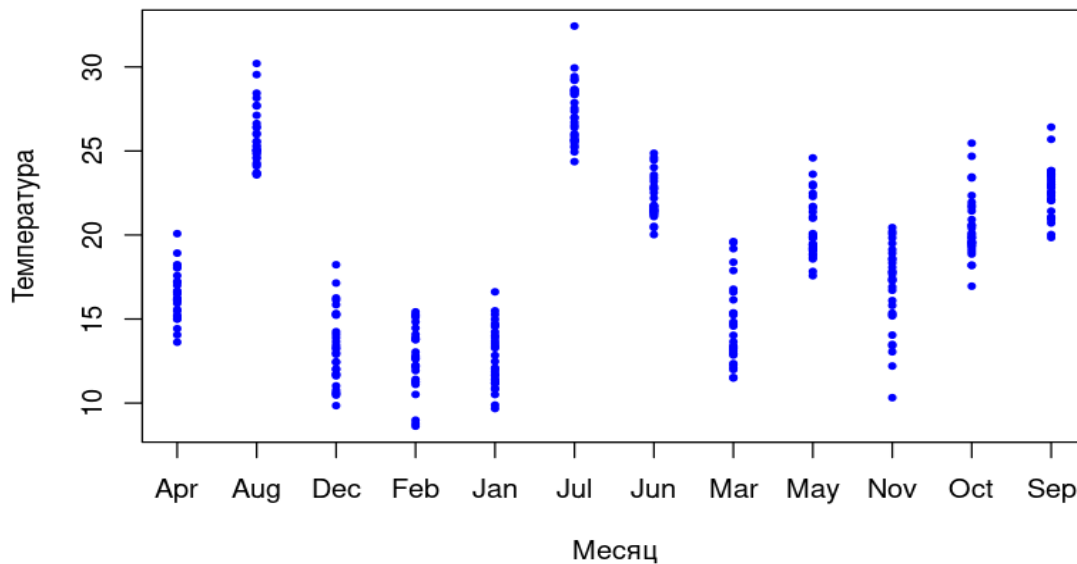
```
boxplot(PowerConsumption_Zone1 ~ quarter,  
        xlab = "Квартал",  
        ylab = "Потребление энергии",  
        col = "skyblue", data = data)
```



```
#stripchart
```

```
stripchart(Temperature ~ Month,  
           data = data,  
           xlab = "Месяц",  
           ylab = "Температура",  
           col = "blue",  
           vertical = TRUE,
```

сех = 0.8, pch = 20)



3 Проверить, являются ли наблюдения выбросами с точки зрения формальных статистических критериев Граббса и Q-теста Диксона. Визуализировать результаты

#Тест Граббса

```
Grubbs test for one outlier
```

```
data: data$PowerConsumption_Zone1
```

```
G = 2.88346, U = 0.70341, p-value = 0.02792
```

```
alternative hypothesis: lowest value 29103.664459375 is an outlier
```

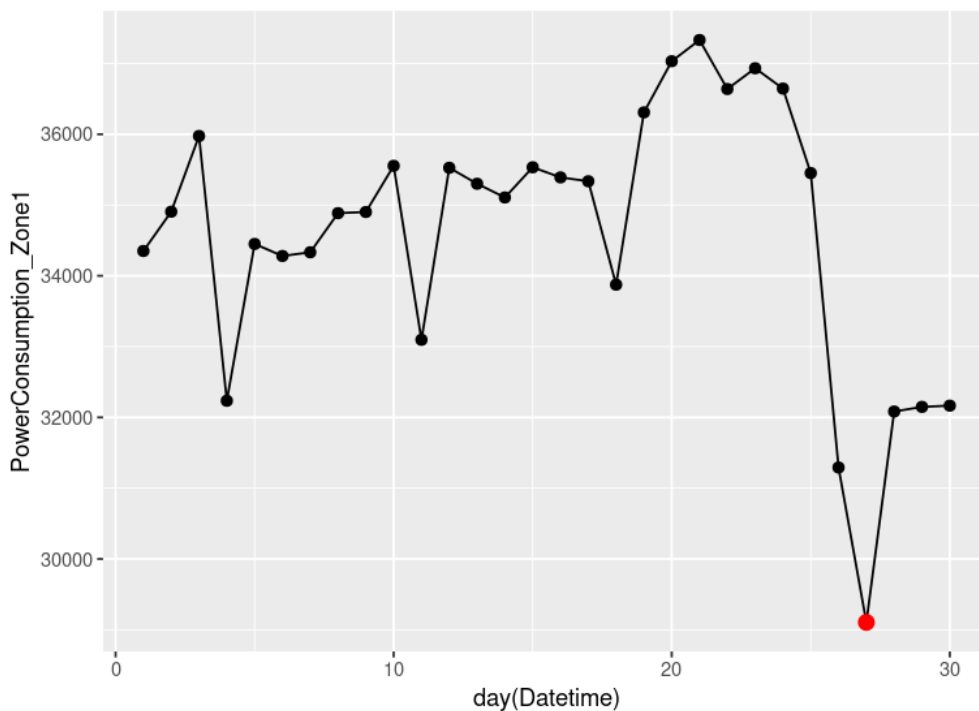
#Тест Диксона

```
Dixon test for outliers
```

```
data: data$PowerConsumption_Zone1
```

```
Q = 0.3805, p-value = 0.09259
```

```
alternative hypothesis: lowest value 29103.664459375 is an outlier
```



4. Воспользоваться инструментами для заполнения пропусков в данных. Пропуски внести вручную и сравнить результаты заполнения с истинными значениями.

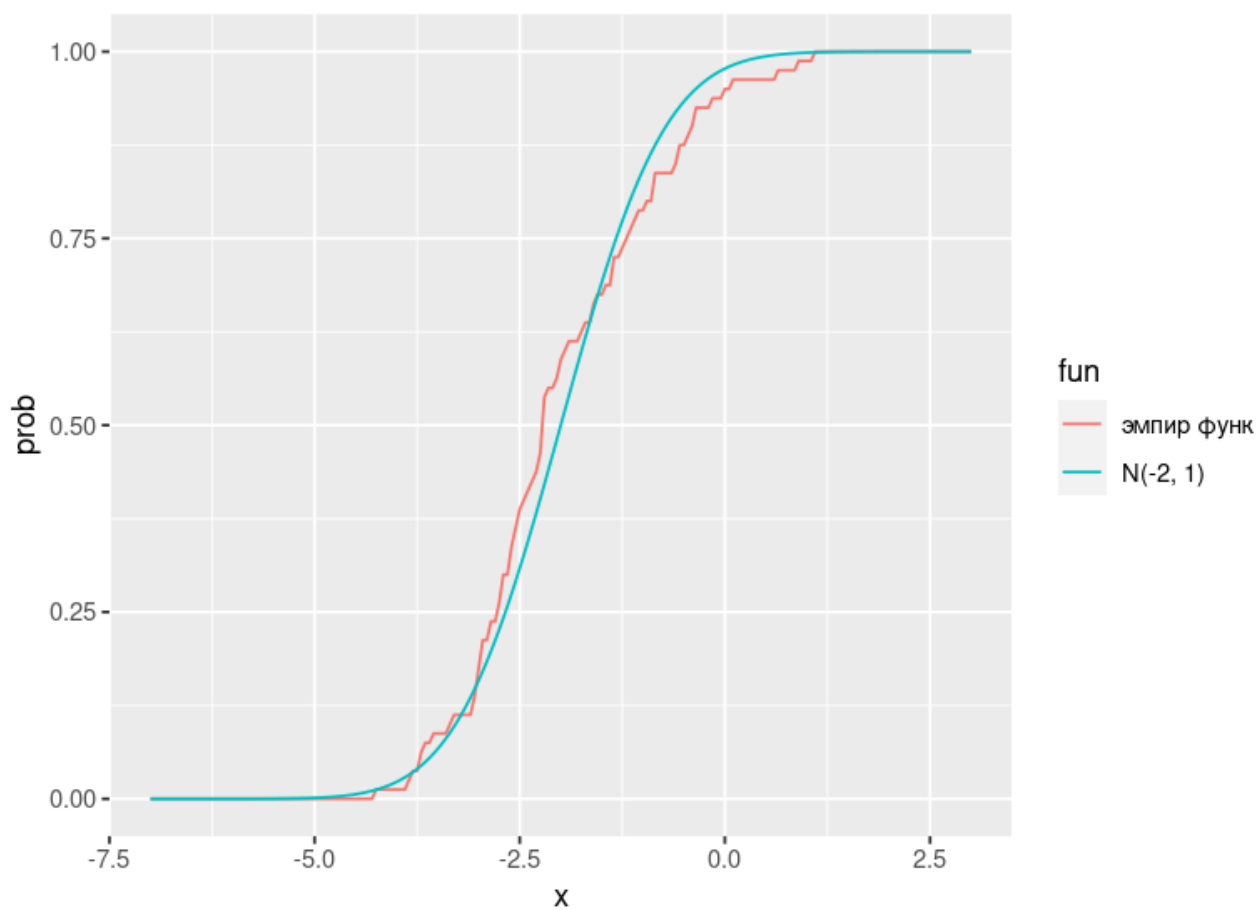
#Заполнение пропусков выборочной медианой

```
meanPC    meanTemp
1 -0.08623068 -0.03371896
```

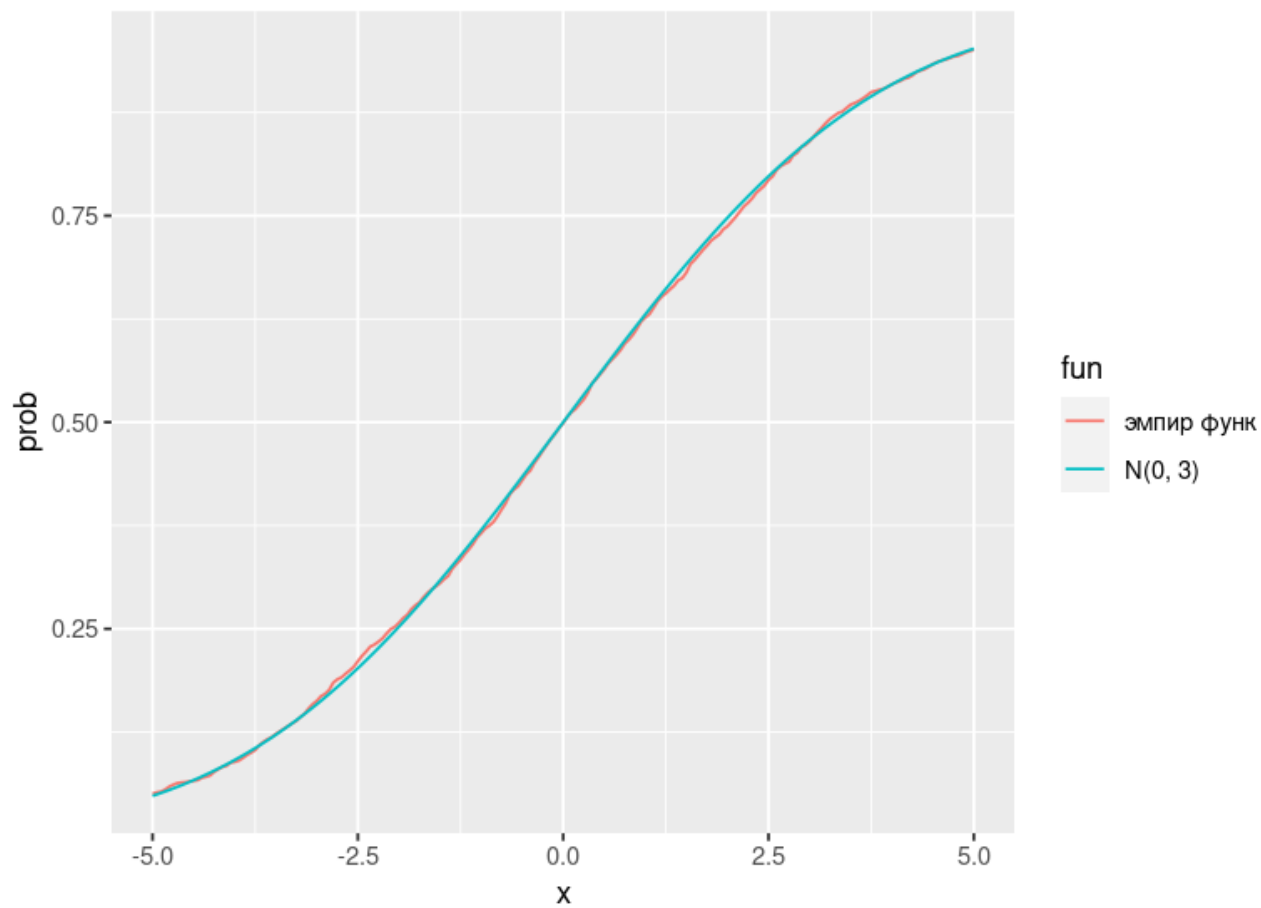
5. Сгенерировать данные из нормального распределения с различными параметрами и провести анализ с помощью графиков эмпирических функций распределений, квантилей, метода огибающих, а также стандартных процедур проверки гипотез о нормальности (критерии Колмогорова-Смирнова, Шапиро-Уилка, Андерсона-Дарлинга, Крамера фон Мизеса, Колмогорова-Смирнова в модификации Лиллиефорса и Шапиро-Франсия).

#Эмпирические функции распределения

1) малая выборка на 80 значений



2) умеренная выборка на 2500 значений



```

> #Критерий Колмогорова-Смирнова
> mapply(ks.test, samples, rep("pnorm", 2), means, sds)
      [,1]      [,2]
statistic 0.1270617 0.01207191
p.value   0.1384257 0.8594863
alternative "two-sided" "two-sided"
method     "One-sample Kolmogorov-Smirnov test" "One-sample Kolmogorov-Smirnov test"
data.name  "dots[[1L]][[1L]]" "dots[[1L]][[2L]]"
> #критерий Шапиро-Уилка
> mapply(shapiro.test, samples)
      [,1]      [,2]
statistic 0.9719027 0.9991841
p.value   0.07506157 0.331674
method     "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name  "dots[[1L]][[1L]]" "dots[[1L]][[2L]]"
> #Критерий Андерсона-Дарлинга
> mapply(ad.test, samples)
      [,1]      [,2]
statistic 0.7651095 0.3570257
p.value   0.04479339 0.4554441
method     "Anderson-Darling normality test" "Anderson-Darling normality test"
data.name  "dots[[1L]][[1L]]" "dots[[1L]][[2L]]"
> #Критерий Крамера фон Мизеса
> mapply(cvm.test, samples)
      [,1]      [,2]
statistic 0.1445342 0.04670974
p.value   0.02738801 0.5605177
method     "Cramer-von Mises normality test" "Cramer-von Mises normality test"
data.name  "dots[[1L]][[1L]]" "dots[[1L]][[2L]]"

> #Критерий Лиллиефорса
> mapply(lillie.test, samples)
      [,1]      [,2]
statistic 0.1211607 0.01230387
p.value   0.005454448 0.4759015
method     "Lilliefors (Kolmogorov-Smirnov) normality test" "Lilliefors (Kolmogorov-Smirnov) normality test"
data.name  "dots[[1L]][[1L]]" "dots[[1L]][[2L]]"
> #Критерий Шапиро-Франция
> mapply(sf.test, samples)
      [,1]      [,2]
statistic 0.9741649 0.9990905
p.value   0.09514126 0.2065434
method     "Shapiro-Francia normality test" "Shapiro-Francia normality test"
data.name  "dots[[1L]][[1L]]" "dots[[1L]][[2L]]"
> |

```

6. Продемонстрировать пример анализа данных с помощью графиков квантилей, метода огибающих, а также стандартных процедур проверки гипотез о нормальности. Рассмотреть выборки малого и умеренного объемов

1) выборка малого размера

Shapiro-Wilk normality test

```
data: data$X0  
W = 0.92496, p-value = 0.001221
```

```
> #Критерий Андерсона-Дарлинга  
> ad.test(data$X0)
```

Anderson-Darling normality test

```
data: data$X0  
A = 1.8743, p-value = 7.634e-05
```

```
> #Критерий Крамера фон Мизеса  
> cvm.test(data$X0)
```

Cramer-von Mises normality test

```
data: data$X0  
W = 0.34557, p-value = 9.49e-05
```

```
> #Критерий Лиллиефорса  
> lillie.test(data$X0)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: data$X0  
D = 0.17553, p-value = 8.555e-05
```

```
> #Критерий Шапиро-Франсия  
> sf.test(data$X0)
```

Shapiro-Francia normality test

```
data: data$X0  
W = 0.92559, p-value = 0.002075
```


2) выборка умеренная

Shapiro-Wilk normality test

```
data: data$X0  
W = 0.95826, p-value < 2.2e-16
```

```
> #Критерий Андерсона-Дарлинга  
> ad.test(data$X0)
```

Anderson-Darling normality test

```
data: data$X0  
A = 31.8, p-value < 2.2e-16
```

```
> #Критерий Крамера фон Мизеса  
> cvm.test(data$X0)
```

Cramer-von Mises normality test

```
data: data$X0  
W = 4.3571, p-value = 7.37e-10
```

Предупреждение:

```
В cvm.test(data$X0) :  
  p-value is smaller than 7.37e-10, cannot be computed more acc  
> #Критерий Лиллиефорса  
> lillie.test(data$X0)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: data$X0  
D = 0.090896, p-value < 2.2e-16
```

```
> #Критерий Шапиро-Франсия  
> sf.test(data$X0)
```

Shapiro-Francia normality test

```
data: data$X0  
W = 0.95881, p-value < 2.2e-16
```