

ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Высшая математика»

№ 109 (103)

май (2018)

ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ АНАЛИЗА ДААННЫХ

*Методические рекомендации к лабораторным работам
для студентов направления подготовки
15.03.06 «Мехатроника и робототехника»
дневной формы обучения*

Могилев 2018

УДК 51
ББК 22.1

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Высшая математика» «24» апреля 2018 г.,
протокол № 8

Составитель канд. физ.-мат. наук, доц. И. И. Маковецкий

Рецензент канд. техн. наук, доц. И. Д. Камчицкая.

Даны задания для лабораторных работ по дисциплине «Пакеты прикладных программ для анализа данных», а также приведены методические рекомендации по их выполнению, перечень необходимой литературы.

Учебно-методическое издание

ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ АНАЛИЗА ДАННЫХ

Ответственный за выпуск	В. Г. Замураев
Технический редактор	С. Н. Красовская
Компьютерная верстка	Н. П. Полевничая

Подписано в печать . Формат 60x84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. .Уч.-изд. л. .Тираж 56 экз. Заказ №

Издатель и полиграфическое исполнение:
Государственное учреждение высшего профессионального образования
«Белорусско-Российский университет».

Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 24.01.2014.

Пр. Мира, 43, 212000, Могилев

© ГУ ВПО «Белорусско-Российский
университет», 2018

Оглавление

Введение.....	4
1 Лабораторная работа № 1. Знакомство с пакетом обработки данных R. Типы данных в R	5
2 Лабораторная работа № 2. Базовые графические возможности пакета R .	9
3 Лабораторная работа № 3. Точечные параметры распределения	14
4 Лабораторная работа № 4. Моделирование случайных распределений в пакете R. Выбор распределения для исходных данных	16
5 Лабораторная работа № 5. Моделирование нормального распределения с заданными параметрами в пакете R. Сравнение эмпирических данных с нормальным распределением с помощью построенных графиков.....	18
6 Лабораторная работа № 6. Моделирование двумерного нормального распределения	20
7 Лабораторная работа № 7. Проверка статистических гипотез о равенстве выборочных средних, выборочных дисперсий для нормально распределенных выборок в пакете R.....	20
8 Лабораторная работа № 8. Проверка непараметрических гипотез о равенстве выборочных средних и выборочных дисперсий	22
9 Лабораторная работа №9. Проверка гипотезы о нормальном распределении совокупности в пакете R	24
10 Лабораторная работа № 10. Изучение корреляционной зависимости в пакете R	26
11 Лабораторная работа № 11. Оценка зависимости качественных признаков с помощью пакета R	27
12 Лабораторная работа № 12. Расчет числовых характеристик корреляционной зависимости в пакете R	28
13 Лабораторная работа № 13. Множественная линейная регрессия, расчет параметров в пакете R.....	30
14 Лабораторная работа № 14. Подбор модели нелинейной регрессии, расчет ее параметров в пакете R.....	30
15 Лабораторная работа № 15. Однофакторный дисперсионный анализ в пакете R	32
16 Лабораторная работа № 16. Многофакторный дисперсионный анализ в пакете R	33
17 Лабораторная работа № 17. Проведение полноформатного статистического исследования и обработка данных средствами пакета R.....	33
Список литературы	34

Введение

Курс «Пакеты прикладных программ для анализа данных» дает студентам представление о возможностях прикладного пакета R в области статистического анализа данных, знакомит с широкими возможностями, предоставляемыми этим пакетом.

Теоретические знания и практические навыки, полученные студентами при изучении курса, будут востребованы в дальнейшем при изучении таких дисциплин, как «Экспериментальные исследования робототехнических систем», «Методы экспериментальных исследований», а также в научно-исследовательской работе и при выполнении выпускной квалификационной работы. Целью дисциплины является овладение навыками первичной обработки, систематизации, анализа данных, получения на их основании обоснованных выводов.

Лабораторные работы практикума выполняются на ЭВМ в прикладном пакете R. Отчет о выполнении представляет собой скрипт на языке программирования R, запуск которого обеспечивает выполнение задания.

1 Лабораторная работа № 1. Знакомство с пакетом обработки данных R. Типы данных в R

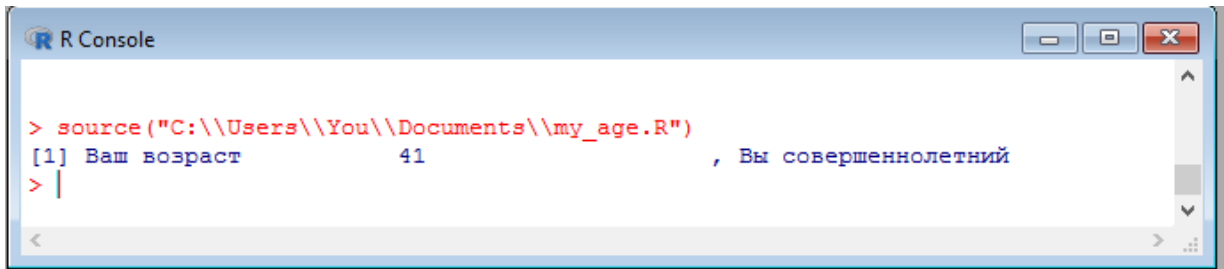
Задание 1

1 Создать переменную `my_age`, присвоить ей числовое значение, соответствующее вашему возрасту.

2 Записать условный оператор, который определяет по значению переменной `my_age`, являетесь ли Вы совершеннолетним. В случае успеха необходимо вывести строку «Ваш возраст ?? лет, Вы совершеннолетний», в противном случае вывести строку «Ваш возраст ??, Вы несовершеннолетний».

3 Создать скрипт R с названием `my_age.r`.

Результат выполнения скрипта представлен на рисунке 1.1.



```

R Console
> source("C:\\Users\\You\\Documents\\my_age.R")
[1] Ваш возраст      41      , Вы совершеннолетний
> |
>
  
```

Рисунок 1.1 – Результат выполнения скрипта `my_age.r`

Задание 2

1 Запустить справочный сервер среды R с командой `help()` из командной строки.

2 Ознакомиться с содержанием справки по команде `scan`.

3 Организовать с помощью команды `scan` ввод с клавиатуры данных в вектор `group_height`, соответствующих росту студентов Вашей группы.

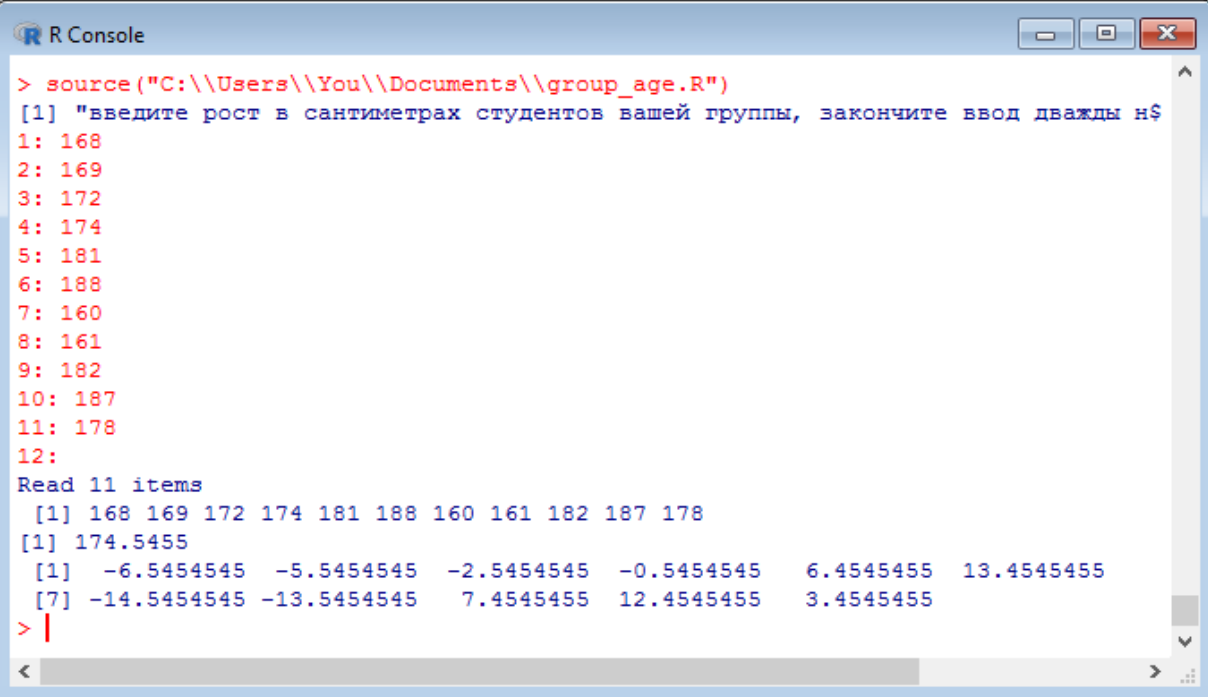
4 Вывести вектор значений `group_height`.

5 Вычислить среднее значение роста по группе с помощью функции `mean()` и присвоить это значение переменной `ave_height`, вывести это значение.

6 Вычислить отклонение значения роста студентов группы от среднего значения, записать полученные значения в вектор `delta`, вывести его на печать.

7 Сохранить действия в скрипт с названием `group_age.r`.

Вариант выполнения задания приведен на рисунке 1.2.



```

R Console
> source("C:\\Users\\You\\Documents\\group_age.R")
[1] "введите рост в сантиметрах студентов вашей группы, закончите ввод дважды н$
1: 168
2: 169
3: 172
4: 174
5: 181
6: 188
7: 160
8: 161
9: 182
10: 187
11: 178
12:
Read 11 items
 [1] 168 169 172 174 181 188 160 161 182 187 178
 [1] 174.5455
 [1] -6.5454545 -5.5454545 -2.5454545 -0.5454545  6.4545455 13.4545455
 [7] -14.5454545 -13.5454545  7.4545455 12.4545455  3.4545455
> |

```

Рисунок 1.2 – Результат выполнения скрипта group_age.r

Задание 3

1 Ознакомиться с разделом справки, связанной с функцией `matrix`.

2 Создать две матричных переменных – a и b , заполнив каждую значениями из последовательных чисел от 1 до 9, при этом матрица a – это матрица-столбец, а матрица b – матрица-строка.

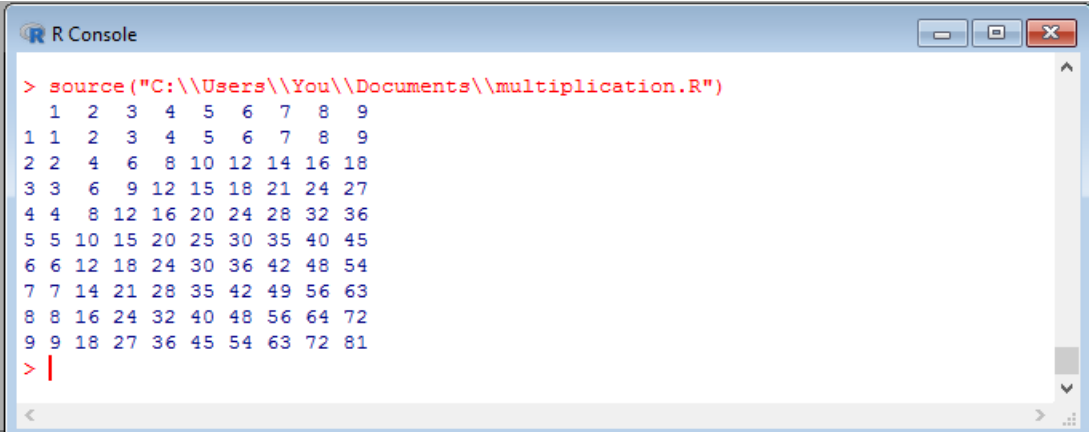
3 Создать матрицу c , полученную произведением a и b .

4 Задать имена строкам и столбцам матрицы c , установив их соответственно последовательными числами от 1 до 9.

5 Вывести результаты счета на экран.

6 Сохранить действия в скрипт с названием `multiplication.r`.

Результат выполнения задания приведен на рисунке 1.3.



```

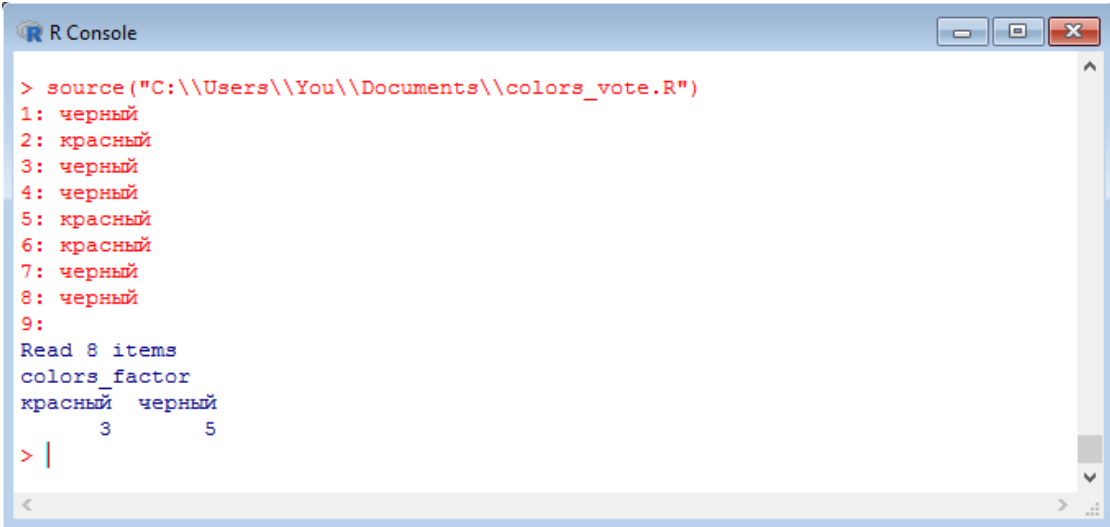
R Console
> source("C:\\Users\\You\\Documents\\multiplication.R")
 1  2  3  4  5  6  7  8  9
1 1  2  3  4  5  6  7  8  9
2 2  4  6  8 10 12 14 16 18
3 3  6  9 12 15 18 21 24 27
4 4  8 12 16 20 24 28 32 36
5 5 10 15 20 25 30 35 40 45
6 6 12 18 24 30 36 42 48 54
7 7 14 21 28 35 42 49 56 63
8 8 16 24 32 40 48 56 64 72
9 9 18 27 36 45 54 63 72 81
> |

```

Рисунок 1.3 – Результат выполнения скрипта multiplication.r

Задание 4

- 1 Ознакомиться с содержанием справки по команде `factor`.
 - 2 С помощью команды `scan()` создать вектор `vote`, в который поместить результаты голосования Вашей учебной группы по поводу любимого цвета из имеющихся двух цветов «красный» и «черный».
 - 3 Создать фактор `colors_factor` с помощью команды `factor`.
 - 4 Сформировать таблицу `colors_stat`, отражающую количество вхождений каждого фактора.
 - 5 Вывести таблицу `colors_stat` на печать.
 - 6 Сохранить действия в скрипт с названием `colors_vote.r`.
- Вариант выполнения задания приведен на рисунке 1.4.



```

R Console
> source("C:\\Users\\You\\Documents\\colors_vote.R")
1: черный
2: красный
3: черный
4: черный
5: красный
6: красный
7: черный
8: черный
9:
Read 8 items
colors_factor
красный  черный
          3      5
> |

```

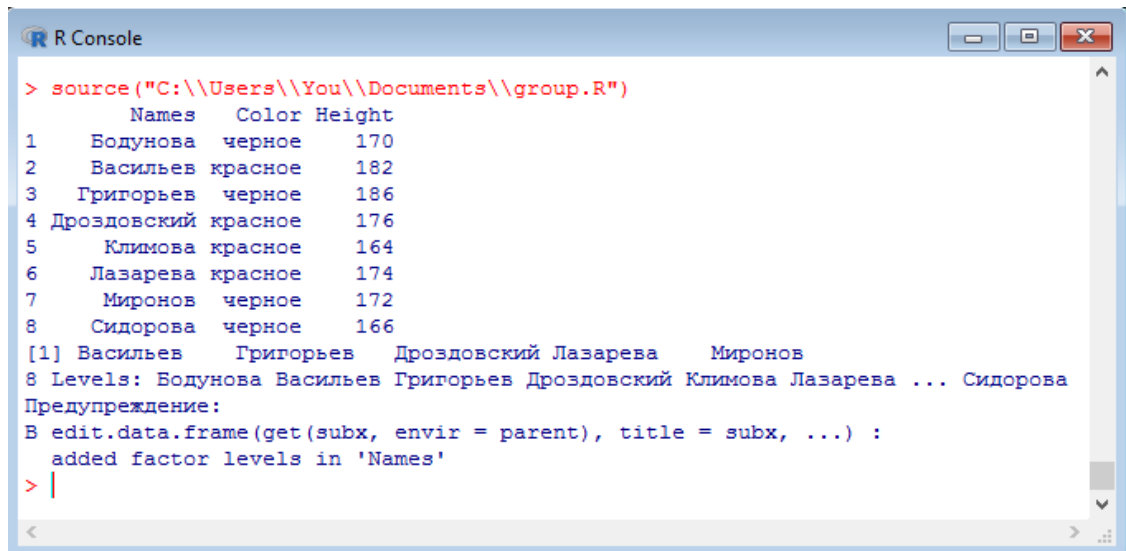
Рисунок 1.4 – Результат выполнения скрипта `colors_vote.r`

Задание 5

- 1 Создать вектор `names`, содержащий фамилии студентов Вашей группы.
- 2 Создать вектор `vote` с результатами голосования о выборе цвета по схеме «красное», «черное».
- 3 Создать вектор `height`, в котором собрана статистика о росте студентов Вашей группы.
- 4 Объединить векторы `names`, `vote`, `height` в список `my_stats` с помощью команды `data.frame()`. Установить названия компонентов соответственно `Names`, `Color`, `Height`.
- 5 С помощью команды `fix()` и встроенного редактора добавить к набору данных фактор, содержащий значения «Сидорова», «черный», 166.
- 6 Вывести на экран набор данных `my_stats`.
- 7 Вывести на экран фамилии людей, рост которых превышает 170 см.

8 Сохранить действия в скрипт с названием group.r.

Пример результата выполнения задания приведен на рисунке 1.5.



```

R Console
> source("C:\\Users\\You\\Documents\\group.R")
  Names    Color Height
1  Бодунова  черное   170
2  Васильев  красное   182
3  Григорьев  черное   186
4  Дроздовский  красное   176
5  Климова  красное   164
6  Лазарева  красное   174
7  Миронов  черное   172
8  Сидорова  черное   166
[1] Васильев  Григорьев  Дроздовский  Лазарева  Миронов
8 Levels: Бодунова Васильев Григорьев Дроздовский Климова Лазарева ... Сидорова
Предупреждение:
В edit.data.frame(get(subx, envir = parent), title = subx, ...) :
  added factor levels in 'Names'
> |

```

Рисунок 1.4 – Результат выполнения скрипта group.r

Задание 6

1 Ознакомиться с разделом справки, связанным с функцией as.POSIXlt().

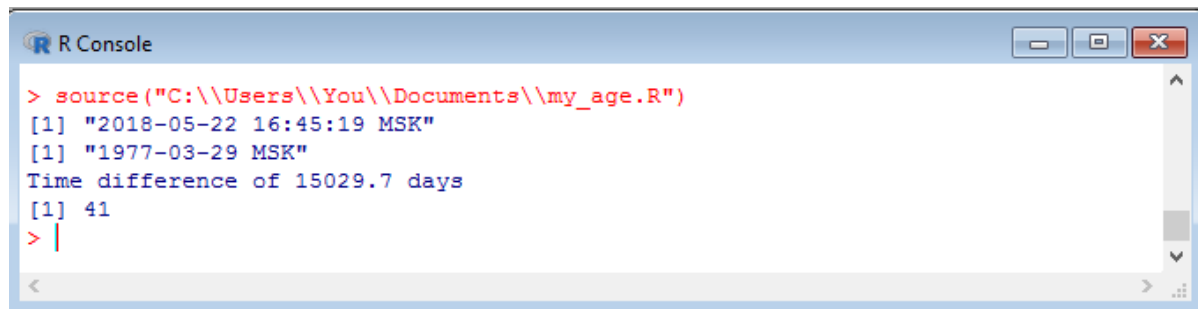
2 Создать переменную *date1*, содержащую информацию о текущей дате.

3 Создать переменную *date2*, содержащую информацию о Вашей дате рождения.

4 Вывести на экран переменные *date1*, *date2*, количество прожитых Вами дней, лет.

5 Сохранить действия в скрипт с названием my_age.r.

Пример результата выполнения задания приведен на рисунке 1.6.



```

R Console
> source("C:\\Users\\You\\Documents\\my_age.R")
[1] "2018-05-22 16:45:19 MSK"
[1] "1977-03-29 MSK"
Time difference of 15029.7 days
[1] 41
> |

```

Рисунок 1.5 – Результат выполнения скрипта my_age.r

Задание 7

1 Ознакомиться с разделом справки, связанным с функцией ts().

2 Создать с помощью редактора данных вектор *temperature*, который содержит информацию о средней температуре за 24 месяца.

3 С помощью функции `ts()` преобразовать вектор `temperature` во временной ряд, начиная с января месяца 2016 г. Результат занести в переменную `temperature.ts`

4 С помощью команды `plot()` построить график среднемесячных температур за 2016 – 2017 г.г.

5 Сохранить действия в скрипт с названием `temp.r`.

Пример результата выполнения задания приведен на рисунке 1.7.

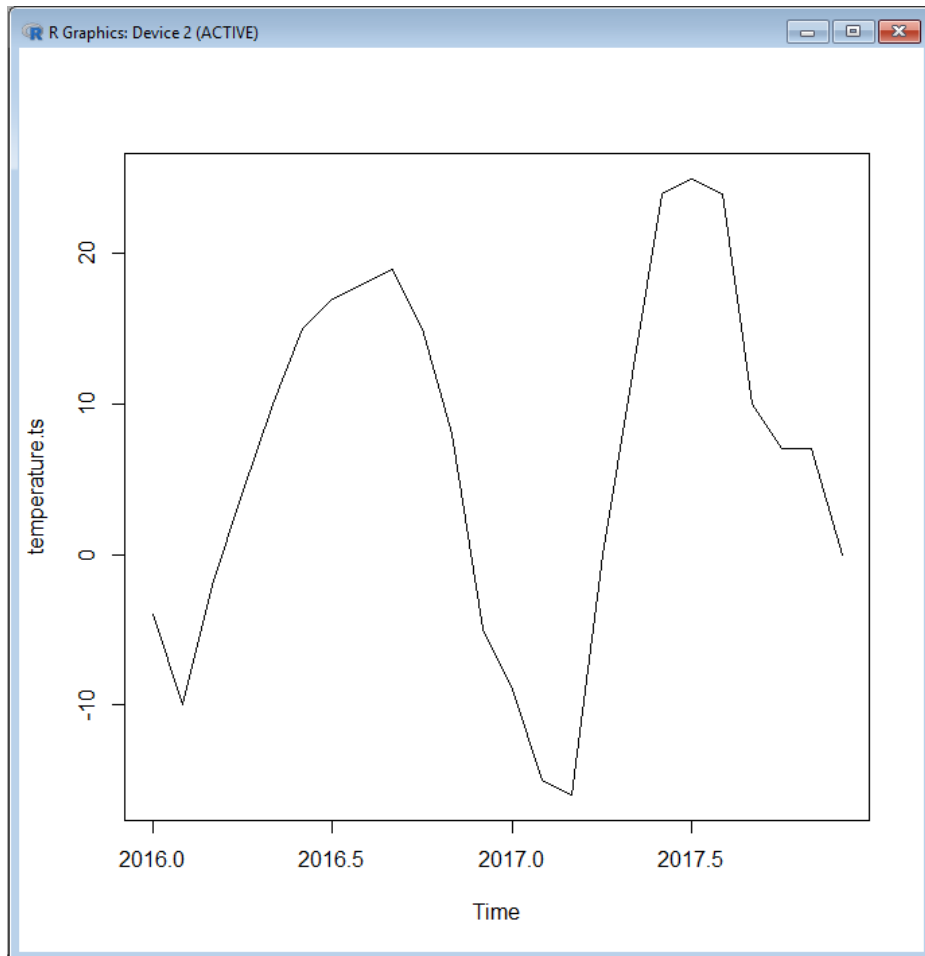


Рисунок 1.6 – Результат выполнения скрипта `temp.r`

2 Лабораторная работа № 2. Базовые графические возможности пакета R

Задание 1

1 Ознакомиться с разделом справки, связанным с функцией `read.csv`.

2 Загрузить из файла `os.csv` данные по использованию в мире операционных систем за последние 12 месяцев. Названия столбцов получить из файла.

3 Построить таблицу, содержащую названия операционных систем, а также средние значения за год по каждой операционной системе. Вывести таблицу на экран.

4 Выяснить, какие три операционные системы занимают лидирующие позиции по использованию в мире.

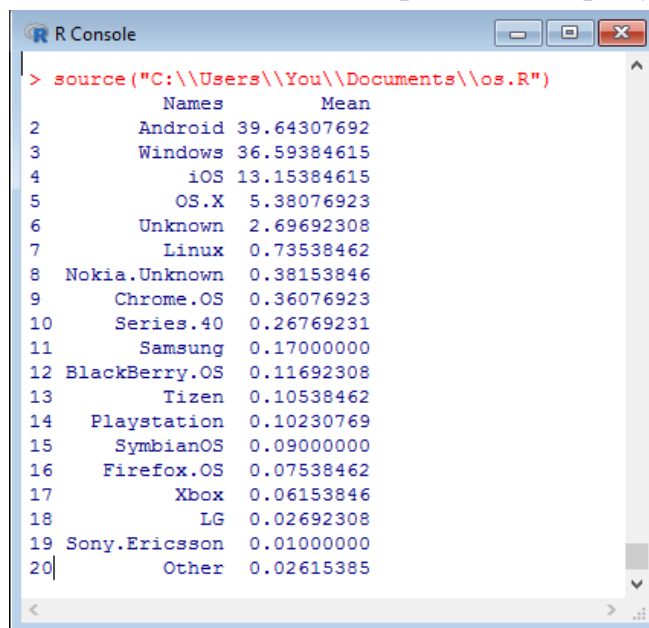
5 Построить для них временные ряды с помощью функции `ts()`.

6 Построить графики динамики использования операционных систем – лидеров рынка на одной иллюстрации.

7 Отформатировать графики таким образом, чтобы на каждом графике различались по форме и размеру маркеры, тип линии, толщина линии, цвет маркера и линии. Изображение должно быть подписано «Динамика использования операционных систем», подписи осей: горизонтальная ось «дата», вертикальная ось «доля рынка».

8 Сохранить действия в скрипт `os.r`.

Пример результата выполнения задания приведен на рисунках 2.1, 2.2.



```
> source("C:\\Users\\You\\Documents\\os.R")
  Names      Mean
2  Android 39.64307692
3  Windows 36.59384615
4    iOS 13.15384615
5   OS.X  5.38076923
6  Unknown 2.69692308
7   Linux  0.73538462
8 Nokia.Unknown 0.38153846
9  Chrome.OS 0.36076923
10 Series.40 0.26769231
11 Samsung 0.17000000
12 BlackBerry.OS 0.11692308
13 Tizen 0.10538462
14 Playstation 0.10230769
15 SymbianOS 0.09000000
16 Firefox.OS 0.07538462
17 Xbox 0.06153846
18 LG 0.02692308
19 Sony.Ericsson 0.01000000
20 Other 0.02615385
```

Рисунок 2.1 – Результат выполнения скрипта `os.r`

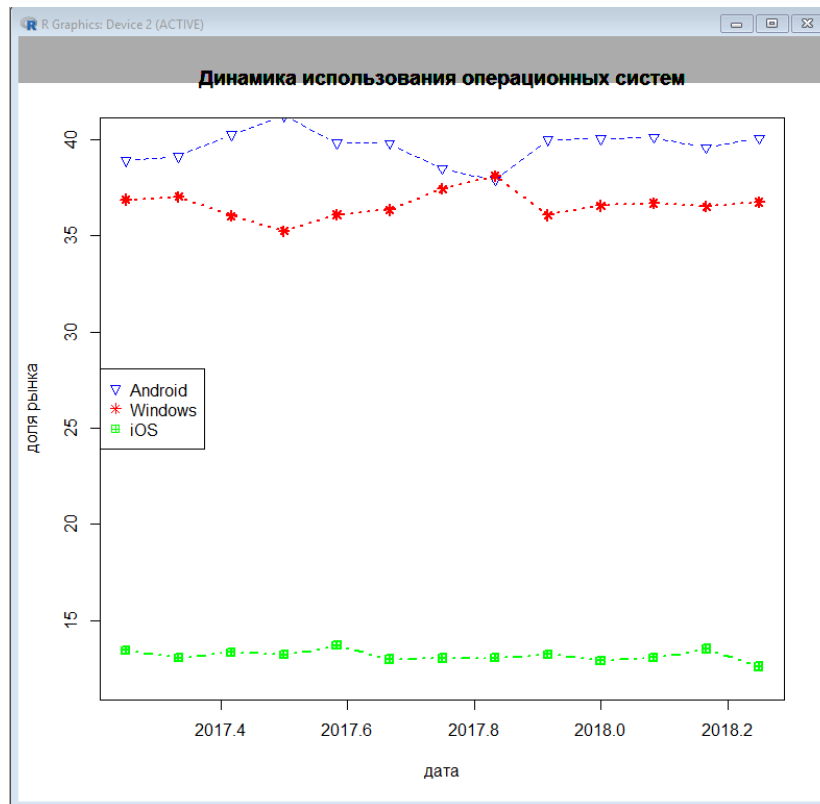


Рисунок 2.2 – Результат выполнения скрипта os.r

Задание 2

1 Создать случайный вектор X , содержащий 100 значений нормально распределенной случайной величины, математическое ожидание которой равно 167, среднее квадратическое отклонение – 20. (Данный вектор моделирует значения роста 100 случайно опрошенных людей).

2 Постройте на одном графике гистограмму плотности вероятности распределения случайного вектора X , сглаженный график его плотности вероятности, график его плотности вероятности без сглаживания. Отформатируйте график в соответствии с рисунком 2.3: цвет гистограммы светло-голубой, цвет графика плотности вероятности – красный, цвет несглаженного графика плотности вероятности – синий, толщина графиков – 2.

3 Сохранить действия в скрипт graph.r.

Вариант выполнения задания приведен на рисунке 2.3.

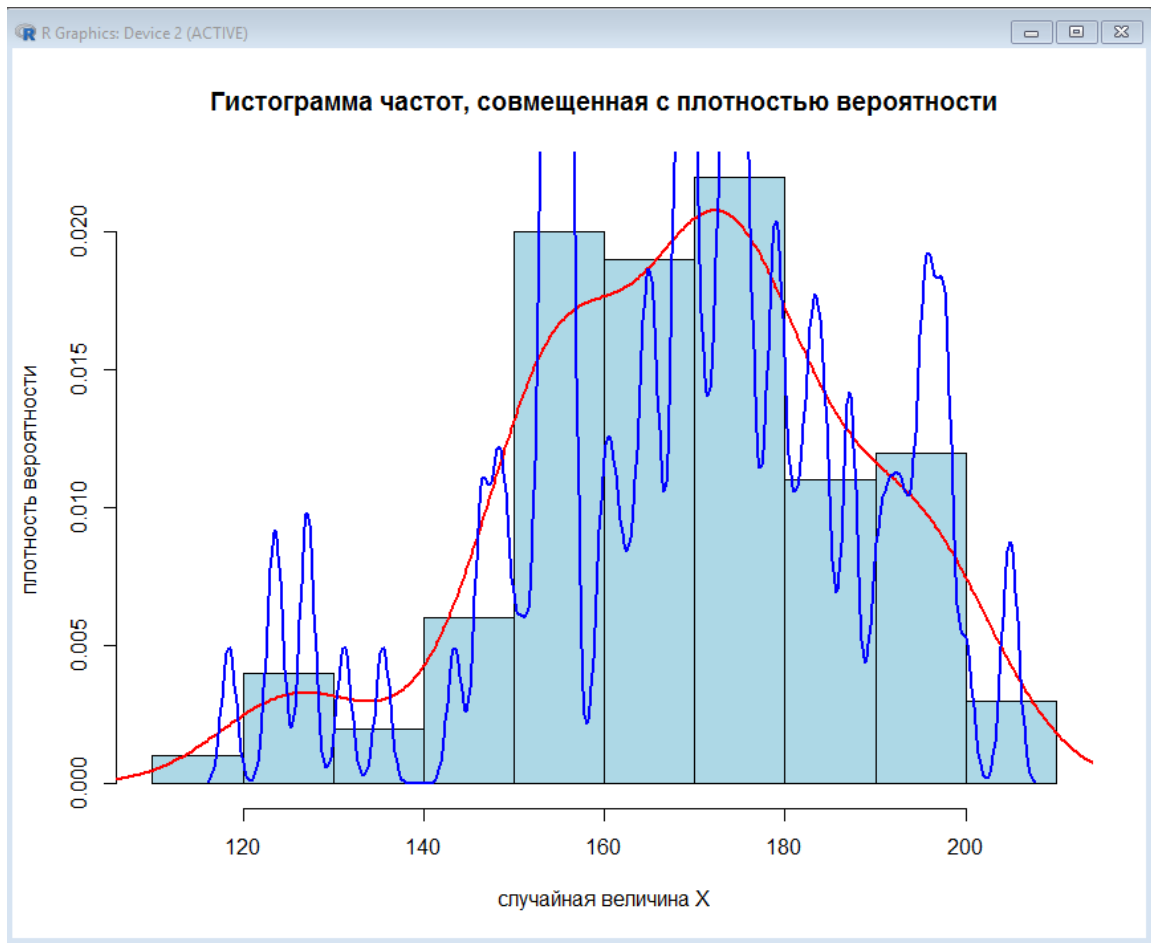


Рисунок 2.3 – Результат выполнения скрипта graph.r

Задание 3

1 Загрузить данные о курсе Биткойна за предыдущие 100 дней из файла BTC.csv.

2 На одном изображении построить график динамики курса криптовалюты и диаграмму «коробка с усами». По «коробке с усами» определить значения курса, являющиеся выбросами, среднее значение курса за 100 дней.

3 Сохранить действия в скрипт bitcoin.r.

Вариант выполнения задания представлен на рисунке 2.4.

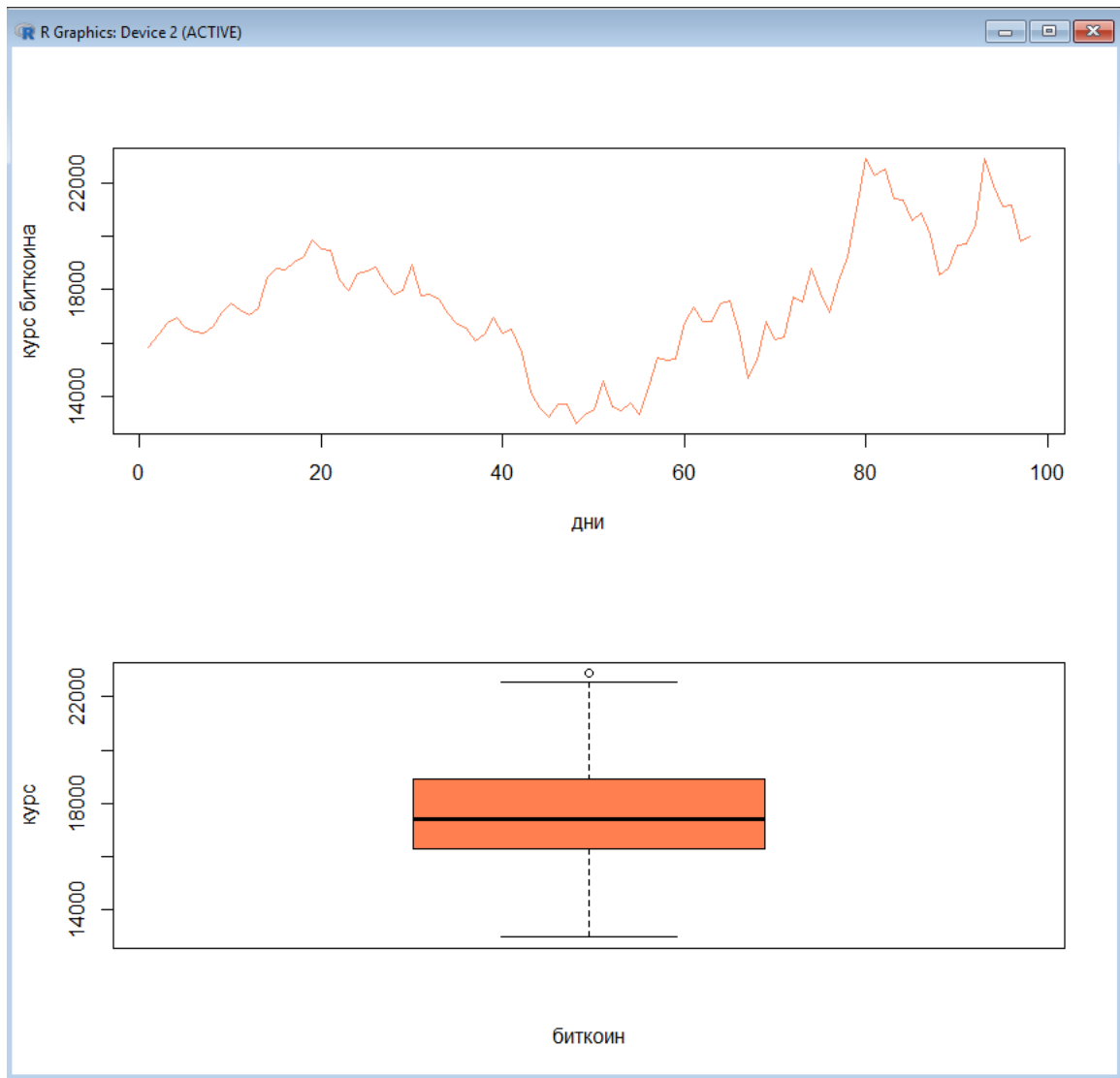


Рисунок 2.4 – Результат выполнения скрипта bitcoin.r

Задание 4

- 1 Создать вектор `vote`, состоящий из результатов голосования в текстовой форме (да, нет). Размерность `vote` – не менее 20 измерений.
 - 2 Построить круговую диаграмму, отражающую результаты голосования.
 - 3 Вывести таблицу результатов голосования.
 - 4 Сохранить действия в скрипт `pie_diagram.r`.
- Вариант выполнения задания представлен на рисунке 2.5.

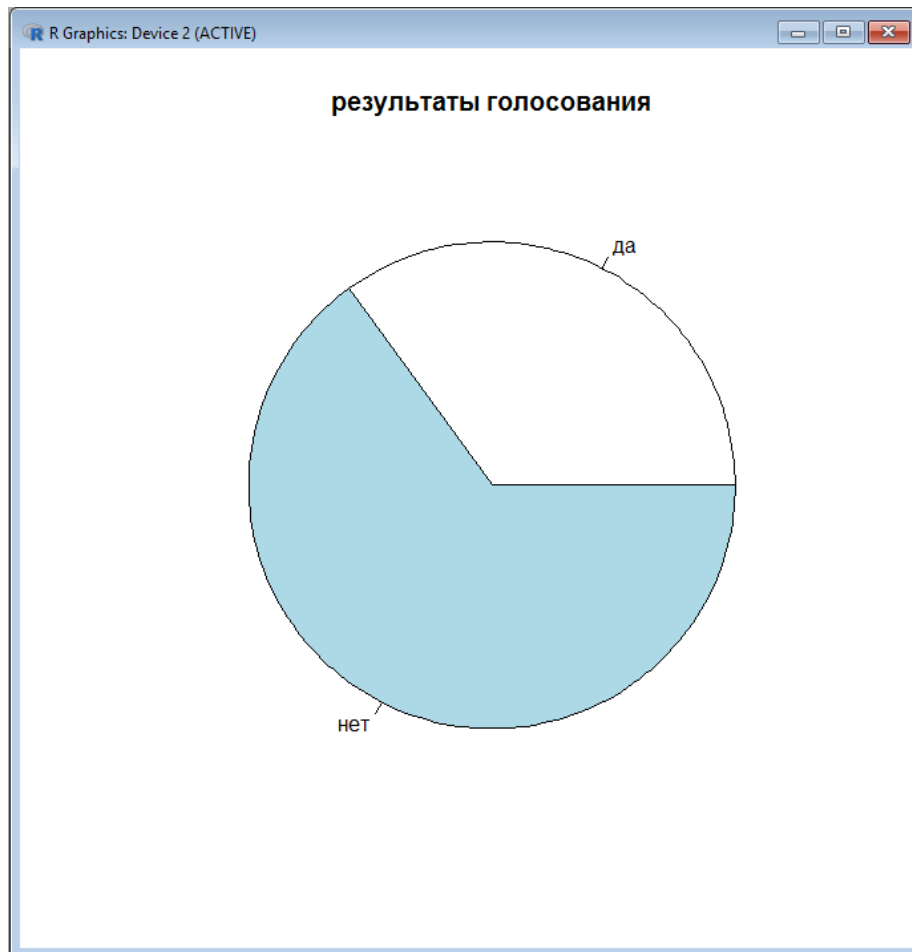


Рисунок 2.5 – Результат выполнения скрипта `pie_diagram.r`

3 Лабораторная работа № 3. Точечные параметры распределения

Задание 1

1 Получить данные из файла `VAZ.csv`, разместить их в переменную `VAZ`, учесть наличие неполных данных.

2 Вывести название моделей, обладающих наибольшим объемом двигателя, наименьшим временем разгона до 100 км/ч, наибольшей мощностью.

3 Вывести название автомобилей, имеющих наибольший объем двигателя, наибольшую мощность, наименьшее время разгона до 100 км/ч, наименьший расход топлива, имеющих одинаковый признак «тип топлива».

4 С помощью функции `tapply()` получить следующие данные об автомобилях, классифицированных по виду топлива: средний объем двигателя, средняя мощность двигателя, средний расход топлива.

5 Получить сводную статистику по данным, представленным в `VAZ`.

6 Получить точечные оценки статистик, являющихся компонентами `VAZ`.

7 Оценить визуально с помощью команды `boxplot()` взаимосвязь расхода топлива и объема двигателя, а также взаимосвязь расхода топлива и мощности двигателя.

8 Сохранить действия в скрипт `VAZ.r`.

Пример выполнения задания приведен на рис. 3.1 и 3.2.

```

> source("C:\\Users\\You\\Documents\\VAZ.R")
[1] "модель с наибольшим объемом двигателя 2120-22 1.8 , ее объем двигателя 1774 куб. см."
[1] "модель с наибольшей мощностью 2115-91 1.3 Wankel , ее мощность 135 л.с."
[1] "модель с наименьшим временем разгона до 100 км/ч 2115-91 1.3 Wankel , она разгоняется до 100 за 9 секунд"
[1] "среди автомобилей на топливе M-80 наибольший объем двигателя имеет 21035 1.2 , объем двигателя 1198"
[1] "среди автомобилей на топливе M-92 наибольший объем двигателя имеет 2120-22 1.8 , объем двигателя 1774"
[1] "среди автомобилей на топливе M-95 наибольший объем двигателя имеет 2131 1.7 i , объем двигателя 1774"
[1] "среди автомобилей на топливе дизельное топливо наибольший объем двигателя имеет 21045 1.5 D , объем двигателя 1452"
[1] "среди автомобилей на топливе M-80 наибольшую мощность двигателя имеет 21035 1.2 , мощность 64"
[1] "среди автомобилей на топливе M-92 наибольшую мощность двигателя имеет 2115-91 1.3 Wankel , мощность 135"
[1] "среди автомобилей на топливе M-95 наибольшую мощность двигателя имеет 2172 Priora Хэтчбек 21723 (1.6 i 16V (98 Hp)) , мощность 98"
[1] "среди автомобилей на топливе дизельное топливо наибольшую мощность двигателя имеет 21045 1.5 D , мощность 59"
[1] "среди автомобилей на топливе M-80 быстрее всего до 100 км/ч разгоняется 21035 1.2 , время разгона до 100 км/ч 9 секунд"
[1] "среди автомобилей на топливе M-92 быстрее всего до 100 км/ч разгоняется 2115-91 1.3 Wankel , время разгона до 100 км/ч 9 секунд"
[1] "среди автомобилей на топливе M-95 быстрее всего до 100 км/ч разгоняется 1118 1.6 i 16V , время разгона до 100 км/ч 11.5 секунд"
[1] "среди автомобилей на топливе дизельное топливо быстрее всего до 100 км/ч разгоняется 21045 1.5 D , время разгона до 100 км/ч 25 секунд"
[1] "среди автомобилей на топливе M-80 наименьший расход топлива имеет 1111 0.65 , расход составляет 6 литров"
[1] "среди автомобилей на топливе M-92 наименьший расход топлива имеет 21073 1.7 i , расход составляет 8.9 литров"
[1] "среди автомобилей на топливе M-95 наименьший расход топлива имеет , расход составляет литров"
[1] "средний объем двигателя автомобилей по типу топлива"
      AM-80      AM-92      AM-95 дизельное топливо
1198.000    1397.909    1565.643    1452.000
[1] "средняя мощность двигателя автомобилей по типу топлива"
      AM-80      AM-92      AM-95 дизельное топливо
64.00000     70.78409     85.75000     53.00000
[1] "средний расход топлива автомобилей по типу топлива"
      AM-80      AM-92      AM-95 дизельное топливо
NaN          9.625000    9.814286    NaN
      Car_type  Volume  Fuel  Power  Time100  Gas_mileage
1111 0.65     : 1  Min. : 649  AM-80   : 1  Min. : 29.00  Min. : 9.00  Min. : 6.000
11113 0.75     : 1  1st Qu.:1328 AM-92   :44  1st Qu.: 70.00  1st Qu.:13.00  1st Qu.: 9.125
1117 Kallina Универсал 1.4i: 1  Median :1500 AM-95   :28  Median : 78.00  Median :16.00  Median : 9.800
1117 Kallina Универсал 1.6i: 1  Mean :1459   дизельное топливо: 1  Mean : 76.11  Mean :16.58  Mean : 9.678
1118 1.6 i     : 1  3rd Qu.:1596   : 1  3rd Qu.:16.25  3rd Qu.:10.000
1118 1.6 i 16V : 1  Max. :1774   : 1  Max. :30.00  Max. :12.500
(Other)      :68
[1] "дисперсия объема двигателя 201.780018487442"
[1] "дисперсия мощности двигателя 14.2863444947469"
[1] "дисперсия расхода топлива 1.11963733195265"
>

```

Рисунок 3.1 – Результат выполнения скрипта `VAZ.r`

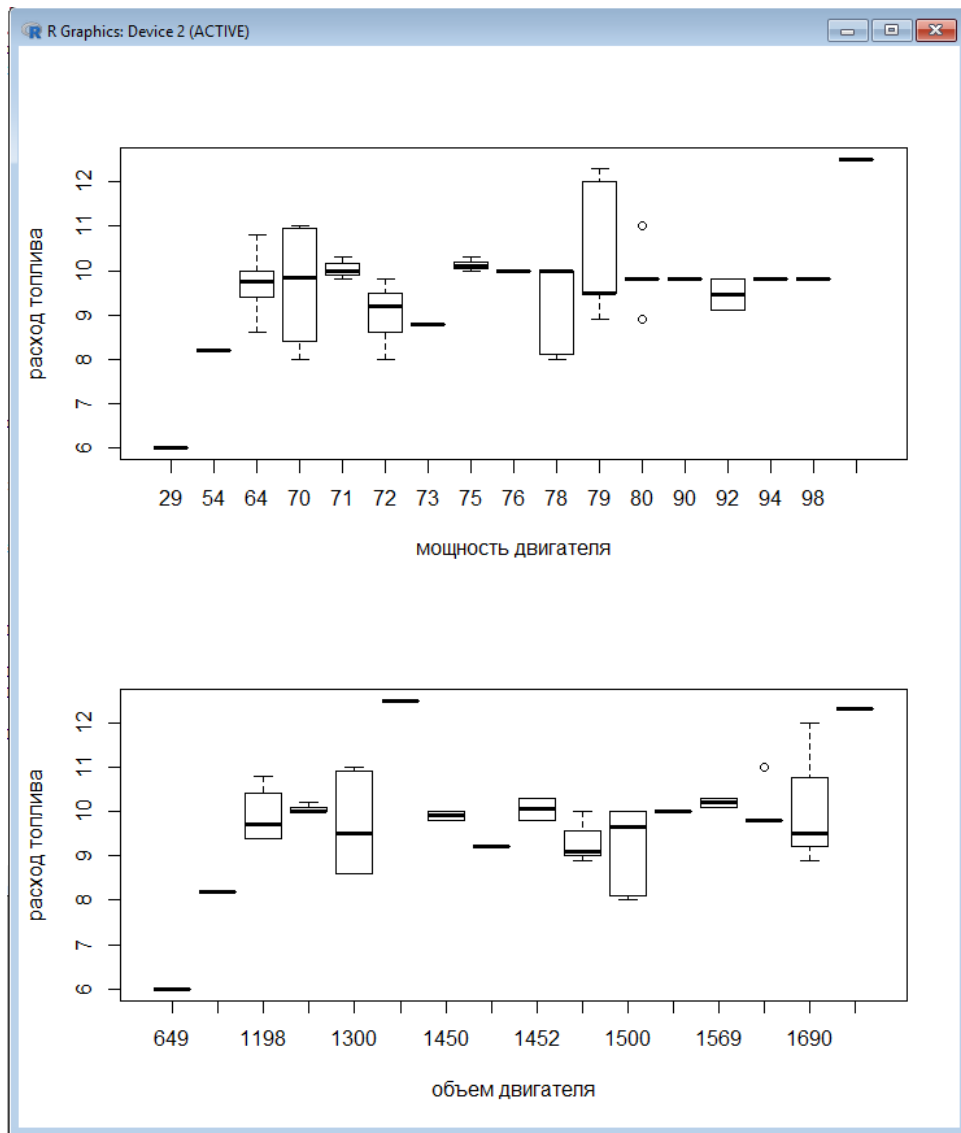


Рисунок 3.2 – Результат выполнения скрипта VAZ.r

4 Лабораторная работа № 4. Моделирование случайных распределений в пакете R. Выбор распределения для исходных данных

Перед выполнением заданий необходимо ознакомиться с разделами справки, связанными с функциями `dbeta`, `dbinom`, `dcauchy`, `dchisq`, `dexp`, `df`, `dgamma`, `dgeom`, `dhyper`, `dlnorm`, `dmultinom`, `dnbinom`, `dnorm`, `dpois`, `dt`, `dunif`, `dweibull`.

Задание 1

1 Построить графики плотности вероятности и функции распределения для случайной величины, имеющей бета-распределение со следующими наборами

параметров: $\alpha = 1, \beta = 1$; $\alpha = 0.5, \beta = 0.5$, $\alpha = 2, \beta = 2$, $\alpha = 5, \beta = 1$, $\alpha = 2, \beta = 5$.
 Разместить графики на одном рисунке, изобразив их разными цветами.

2 Сохранить действия в скрипт beta.r.

Результат выполнения задания представлен на рисунке 4.1.

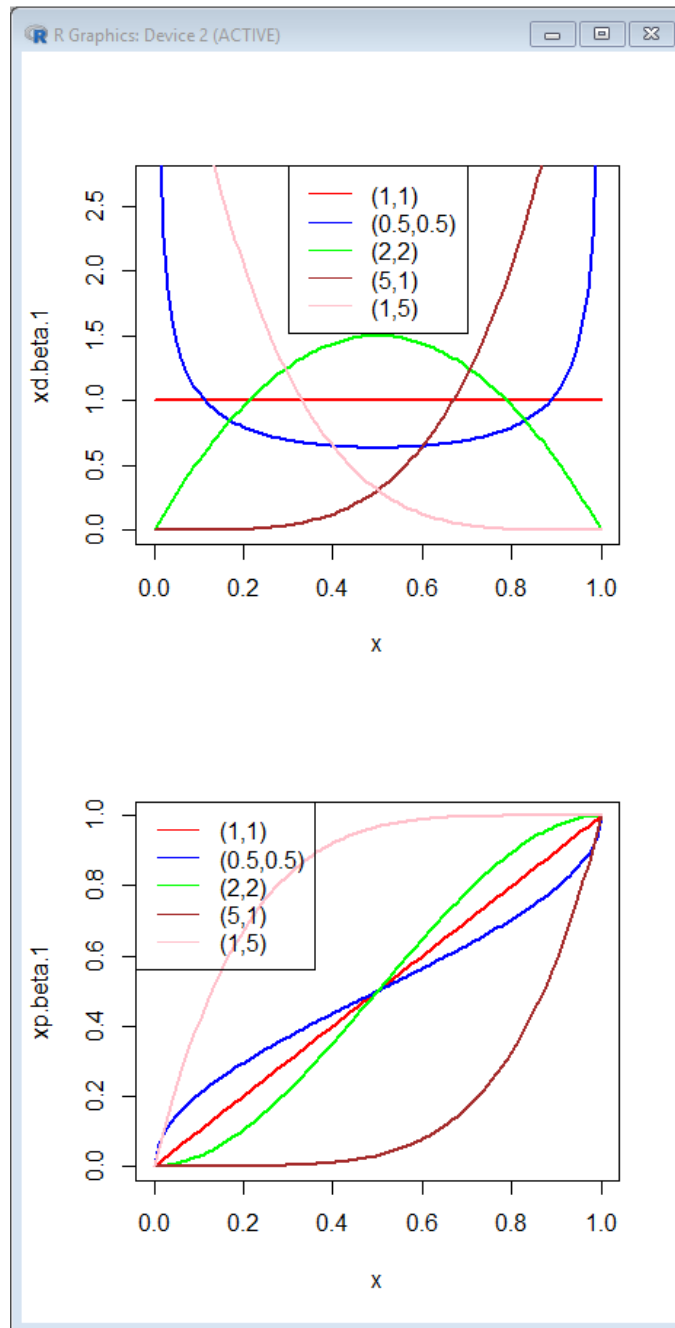


Рисунок 4.1 – Результат выполнения скрипта beta.r

Задание 2

1 Случайная величина X – число выпадений четного числа очков при десяти подбрасываниях игральной кости. Построить график плотности вероятности и функции распределения случайной величины X .

2 Сохранить действия в скрипт binom.r.

Задание 3

Построить график плотности вероятности стандартной нормально распределенной случайной величины. Масштаб подобрать так, чтобы по оси Ox размер графика соответствовал правилу «трех сигм». Вывести вторым графиком на этом же рисунке график функции распределения стандартной нормально распределенной случайной величины.

Задание 4

Вывести на одном рисунке графики плотности вероятности нормально распределенных случайных величин $N(-10,10)$, $N(0,10)$, $N(10,10)$, на интервале $(-20; 20)$. На втором рисунке вывести их функции распределения.

Задание 5

Вывести на одном рисунке графики плотности вероятности нормально распределенных случайных величин $N(0,1)$, $N(0,0.5)$, $N(0,1.5)$. На втором рисунке вывести графики функций распределения этих случайных величин.

5 Лабораторная работа № 5. Моделирование нормального распределения с заданными параметрами в пакете R. Сравнение эмпирических данных с нормальным распределением с помощью построенных графиков

Задание 1

Результаты измерений некоторой статистики заданы таблицей. Найти точечные параметры распределения, построить модельное нормальное распределение по заданным параметрам. Определить с помощью графиков, согласуются ли эмпирические данные с построенным нормальным распределением.

3,12	4,9	4,14	3,06	4,23	3,97	4,03	4,04	3,56	4,31
3,07	4,76	3,27	3,21	3,49	3,79	3,69	4,59	4,26	3,8
3,74	3,51	3,35	3,6	3,93	3,33	3,15	3,33	4,41	3,6
3,97	3,22	3,36	4,32	3,06	4,21	4,26	4,14	4,93	4,32
4,12	4,41	3,87	3,21	4,53	3,74	4,5	3,68	3,25	3,92
4,74	3,86	3,58	3,17	3,32	4,39	4,14	4,03	3,52	3,11
4,12	4,14	3,96	3,51	4,06	3,5	4,16	4,87	3,24	4,42
3,4	3,3	4,13	3,88	3,04	4,24	3,85	4,8	4,64	4,47

4,3 4,61 3,89 4,63 4,26 3,16 3,31 4,75 3,01 4,53
 3,44 3,5 3,77 4,06 4,13 4,55 3,78 3,16 3,53 4,45

Задание 2

Сгенерировать случайную выборку, содержащую 100 элементов. Выборка должна иметь нормальное распределение с параметрами $N(100,10)$. Сравнить с помощью графиков распределение полученной выборки с модельным нормальным распределением с аналогичными параметрами.

Задание 3

Проверить с помощью графиков, согласуется ли распределение выборки, представленной в таблице 5.1, с нормальным распределением с аналогичными параметрами.

Таблица 5.1 – Данные для задания 3

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Задание 4

С помощью графиков проверить, согласуется ли распределение выборки, представленной в таблице 5.2, с нормальным распределением.

Таблица 5.2 – Данные для задания 4

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Задание 5

С помощью графиков проверить, согласуется ли распределение выборки, представленной в таблице 5.3, с нормальным распределением.

Таблица 5.1 – Данные для задания 5

интервал h_i	[3;8)	[8;13)	[13;18)	[18;23)	[23;28)	[28;33)	[33;38)
n_i	6	8	15	40	16	8	7

Задание 6

С помощью графиков проверить, согласуется ли распределение выборки, представленной в таблице 5.4, с нормальным распределением.

Таблица 5.2 – Данные для задания 6

интервал h_i	[-20;-10)	[-10;0)	[0;10)	[10;20)	[20;30)	[30;40)	[40;50)
n_i	20	47	80	89	40	16	8

6 Лабораторная работа № 6. Моделирование двумерного нормального распределения

Перед выполнением лабораторной работы необходимо установить пакет MASS, выполнив команду `library(MASS)`.

Задание 1

- 1 Определить матрицу, содержащую центр распределения μ .
- 2 Определить квадратную ковариационную матрицу, характеризующую разброс двумерной случайной величины σ .
- 3 С помощью команды `mvrnorm` сгенерировать вектор `b1vn`, содержащий 5000 значений, образующих нормально распределенную случайную двумерную величину.
- 4 Вывести на экран несколько первых значений из вектора `b1vn`
- 5 С помощью команды `bve2d` построить плотность вероятности двумерной нормально распределенной случайной величины. Значения двумерной плотности вероятности сохранить в матрицу `b1vn.kde`.
- 6 Визуализировать полученную двумерную плотность вероятности с помощью команды `image`. Для удобства чтения изображения нанести дополнительно горизонтали с помощью команды `contour`.
- 7 Построить трехмерное изображение построенного распределения с помощью функции `perspective`.
- 8 Сохранить действия в скрипте `binormal.r`.

7 Лабораторная работа № 7. Проверка статистических гипотез о равенстве выборочных средних, выборочных дисперсий для нормально распределенных выборок в пакете R

Задание 1

Результаты измерений температуры высокоточным термометром у 15 человек: 36,5; 36,8; 36,2; 36,2; 36,4; 36,5; 36,9; 36,3; 36,4; 36,4; 36,6; 36,8; 36,7; 36,1; 36,1. Как известно, нормой температуры здорового человека является значение 36,6. Исходя из предположения, что генеральная совокупность распределена по нормальному закону, выяснить, отличается ли выборочная средняя от заданного эталонного значения.

Задание 2

Автоматизированная линия производит рассверливание отверстий, диаметр которых в соответствии с технической документацией должен быть 10 мм. Отдел технического контроля произвел измерение диаметров отверстий у случайно отобранных 10 деталей. Результаты измерений диаметров: 10,05, 10,03, 10,00, 9,98, 9,99, 10,01, 9,97, 10,03, 9,99, 10,01. Исходя из предположения, что генеральная совокупность распределена по нормальному закону, выяснить, отличается ли выборочная средняя от заданного проектного значения диаметра.

Задание 3

В результате эксперимента 10 раз измерено значение исследуемой величины, результаты измерений следующие: 4,24; 4,5; 4,46; 4,31; 4,38; 4,35; 4,52; 4,23; 4,35; 4,38. Из теории известно, что данная величина должна иметь значение, равное 4,35. Проверить, согласуются ли результаты эксперимента с теоретическим значением при предположении, что выборка и генеральная совокупность распределены по нормальному закону.

Задание 4

Двумя независимо работающими приборами произведены по 10 замеров одного и того же параметра при одинаковых условиях измерения. Исходя из предположения о нормальном распределении выборок, проверить, согласуются ли средние значения выборок.

Результаты измерения первого прибора: 12,5; 12,6; 12,42; 12,44; 12,67; 12,55; 12,31; 12,59; 12,46; 12,34.

Результаты измерений второго прибора: 12,36; 12,38; 12,41; 12,44; 12,40; 12,51; 12,54; 12,49; 12,44; 12,26.

Задание 5

Автомат выполняет измерения одного и того же показателя. Имеются данные 10 замеров до наладки датчика и 10 замеров после наладки датчика. Выяснить, повлияла ли наладка датчика на результаты измерений (сравнить выборочные средние, исходя из предположения о нормальном распределении выборок).

Результаты измерений до наладки: 12, 14, 14, 13, 15, 10, 11, 10, 13, 12.

Результаты измерений после наладки: 12, 12, 14, 15, 14, 13, 13, 14, 12, 11.

Задание 6

У 15 пациентов взят анализ крови на уровень содержания сахара до приема пищи и через 2 ч после приема пищи. Сравнить среднее содержание сахара в крови до и после приема пищи, исходя из предположения, что этот параметр имеет нормальное распределение. Данные представлены в таблице 7.1.

Таблица 7.1 – Данные для задания 6

До	4,1	5,2	5,9	5,6	5,8	4,8	6,1	4,9	4,4	5,6	5,3	5,2	5,0	4,9	5,1
После	7,4	7,5	7,2	7,9	8,0	8,1	8,4	7,9	7,8	8,0	8,2	10,0	9,8	8,5	8,2

Задание 7

Двумя независимо работающими приборами произведено по 10 замеров одного и того же параметра при одинаковых условиях измерения. Исходя из предположения о нормальном распределении выборок, выяснить, обладают ли приборы одинаковой точностью измерений.

Результаты измерения первого прибора: 12,5; 12,6; 12,42; 12,44; 12,67; 12,55; 12,31; 12,59; 12,46; 12,34.

Результаты измерений второго прибора: 12,36; 12,38; 12,41; 12,44; 12,40; 12,51; 12,54; 12,49; 12,44; 12,26.

Задание 8

Автомат выполняет измерения одного и того же показателя. Имеются данные 10 замеров до наладки датчика и 10 замеров после наладки датчика. Выяснить, повлияла ли наладка датчика на результаты измерений (сравнить выборочные дисперсии, исходя из предположения о нормальном распределении выборок).

Результаты измерений до наладки: 12, 14, 14, 13, 15, 10, 11, 10, 13, 12.

Результаты измерений после наладки: 12, 12, 14, 15, 14, 13, 13, 14, 12, 11.

8 Лабораторная работа № 8. Проверка непараметрических гипотез о равенстве выборочных средних и выборочных дисперсий

Задание 1

Результаты измерений температуры высокоточным термометром у 15 человек: 36,5; 36,8; 36,2; 36,2; 36,4; 36,5; 36,9; 36,3; 36,4; 36,4; 36,6; 36,8; 36,7; 36,1; 36,1. Как известно, нормой температур здорового человека является значение 36,6. Исходя из предположения, что о распределении генеральной

совокупности ничего не известно, выяснить, отличается ли выборочная средняя от заданного эталонного значения.

Задание 2

Автоматизированная линия производит рассверливание отверстий, диаметр которых в соответствии с технической документацией должен быть 10 мм. Отдел технического контроля произвел измерение диаметров отверстий у случайно отобранных 10 деталей. Результаты измерений диаметров: 10,05, 10,03, 10,00, 9,98, 9,99, 10,01, 9,97, 10,03, 9,99, 10,01. О распределении генеральной совокупности ничего не известно. Выяснить, отличается ли выборочная средняя от заданного проектного значения диаметра.

Задание 3

В результате эксперимента 10 раз измерено значение исследуемой величины, результаты измерений следующие: 4,24; 4,5; 4,46; 4,31; 4,38; 4,35; 4,52; 4,23; 4,35; 4,38. Из теории известно, что данная величина должна иметь значение, равное 4,35. Проверить, согласуются ли результаты эксперимента с теоретическим значением при предположении, что нет данных о распределении выборки и генеральной совокупности.

Задание 4

Двумя независимо работающими приборами произведено по 10 замеров одного и того же параметра при одинаковых условиях измерения. О распределении выборок ничего не известно. Проверить, согласуются ли средние значения выборок.

Результаты измерения первого прибора: 12,5; 12,6; 12,42; 12,44; 12,67; 12,55; 12,31; 12,59; 12,46; 12,34.

Результаты измерений второго прибора: 12,36; 12,38; 12,41; 12,44; 12,40; 12,51; 12,54; 12,49; 12,44; 12,26.

Задание 5

Автомат выполняет измерения одного и того же показателя. Имеются данные 10 замеров до наладки датчика и 10 замеров после наладки датчика. Выяснить, повлияла ли наладка датчика на результаты измерений (сравнить выборочные средние, если распределение выборок неизвестно).

Результаты измерений до наладки: 12, 14, 14, 13, 15, 10, 11, 10, 13, 12.

Результаты измерений после наладки: 12, 12, 14, 15, 14, 13, 13, 14, 12, 11.

Задание 6

У 15 пациентов взят анализ крови на уровень содержания сахара до приема пищи и через два часа после приема пищи. Сравнить среднее содержание сахара в крови до и после приема пищи, исходя из предположения, что о распределении этого параметра ничего не известно. Результаты измерений представлены в таблице 8.1.

Таблица 8.1 – Данные для задания 6

До	4.1	5.2	5.9	5.6	5.8	4.8	6.1	4.9	4.4	5.6	5.3	5.2	5.0	4.9	5.1
После	7.4	7.5	7.2	7.9	8.0	8.1	8.4	7.9	7.8	8.0	8.2	10.0	9.8	8.5	8.2

Задание 7

Двумя независимо работающими приборами произведено по 10 замеров одного и того же параметра при одинаковых условиях измерения. О распределении выборок и генеральных совокупностей нет дополнительной информации. Выяснить, обладают ли приборы одинаковой точностью измерений.

Результаты измерения первого прибора: 12,5; 12,6; 12,42; 12,44; 12,67; 12,55; 12,31; 12,59; 12,46; 12,34.

Результаты измерений второго прибора: 12,36; 12,38; 12,41; 12,44; 12,40; 12,51; 12,54; 12,49; 12,44; 12,26.

Задание 8

Автомат выполняет измерения одного и того же показателя. Имеются данные 10 замеров до наладки датчика и 10 замеров после наладки датчика. Выяснить, повлияла ли наладка датчика на результаты измерений (сравнить выборочные дисперсии, если о распределении выборок нет дополнительной информации).

Результаты измерений до наладки: 12, 14, 14, 13, 15, 10, 11, 10, 13, 12.

Результаты измерений после наладки: 12, 12, 14, 15, 14, 13, 13, 14, 12, 11.

9 Лабораторная работа №9. Проверка гипотезы о нормальном распределении совокупности в пакете R

Перед выполнением лабораторной работы необходимо подключить библиотеку `nortest`.

Для анализа применить тесты Колмогорова-Смирнова, Пирсона, Лиллифорса, Крамера-фон Мизеса, Андерсона – Дарлинга, Шапиро-Франсиа.

Задание 1

Используя средства пакета R, проверить, имеет ли выборка, представленная в таблице 9.1, нормальное распределение.

Таблица 9.1 – Данные для задания 1

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Задание 2

Опросом группы получить числовые значения из интервала [10;17]. Проверить полученную выборку на нормальное распределение.

Задание 3

Проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки объема $n = 200$, представленной в таблице 9.2:

Таблица 9.2 – Данные для задания 3

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Задание 4

Проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки, представленной в таблице 9.3:

Таблица 3.3 – Данные для задания 4

интервал h_i	[3;8)	[8;13)	[13;18)	[18;23)	[23;28)	[28;33)	[33;38)
n_i	6	8	15	40	16	8	7

Задание 5

Проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки, представленной в таблице 9.4:

Таблица 9.4 – Данные для задания 5

интервал h_i	[-20;-10)	[-10;0)	[0;10)	[10;20)	[20;30)	[30;40)	[40;50)
n_i	20	47	80	89	40	16	8

Задание 6

Сгенерировать случайную выборку, распределение которой близко к нормальному. Проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки.

10 Лабораторная работа № 10. Изучение корреляционной зависимости в пакете R

Задание 1

По данным таблицы 10.1 построить корреляционное поле, вычислить коэффициент корреляции, ковариацию и коэффициент вариации.

Таблица 10.1 – Данные для задания 1

Y\X	20	25	30	35	40
16	4	6			
26		8	10		
36			32	3	9
46			4	12	6
56				1	5

Задание 2

По данным таблицы 10.2 построить корреляционное поле, вычислить коэффициент корреляции, ковариацию и коэффициент вариации.

Таблица 4 – Данные для задания 2

Y\X	5	10	15	20	25	30	35	40
100	2	1						
120	3	4	3					
140			5	10	8			
160				1		6	1	1
180							4	1

Задание 3

По данным таблицы 10.3 построить корреляционное поле, вычислить коэффициент корреляции, ковариацию и коэффициент вариации.

Таблица 5 – Данные для задания 3

Y\X	18	23	28	33	38	43	48
125		1					
150	1	2	5				
175		3	2	12			
200			1	8	7		
225					3	3	
250						1	1

Задание 4

По данным таблицы 10.4 построить корреляционное поле, вычислить коэффициент корреляции, ковариацию и коэффициент вариации.

Таблица 6 – Данные для задания 4

Y\X	5	10	15	20	25	30	35
100						6	1
120						4	2
140			8	10	5		
160	3	4	3				
180	2	1		1			

Задание 5

Используя данные из файла население.csv, выяснить, существует ли корреляционная зависимость между рядами данных.

Выяснить, существует ли корреляционная зависимость между численностью сельского населения Республики Беларусь и количеством походов в театры, цирки.

Выяснить, существует ли корреляционная зависимость между количеством обучающихся в профессионально-технических учреждениях образования и заболеваемостью алкоголизмом.

Задание 6

Используя данные из файла население.csv, выяснить, между какими рядами данных существует корреляционная зависимость, сделать выводы. Построить корреляционные поля для величин, имеющих сильную зависимость.

11 Лабораторная работа № 11. Оценка зависимости качественных признаков с помощью пакета R

Задание 1

На основе данных, представленных в файле квартиры.csv, сделать вывод о наличии зависимости между расположением квартиры в доме и ее стоимостью.

Задание 2

На основе данных, представленных в файле квартиры.csv, установить, имеется ли зависимость между стоимостью квартиры и годом постройки дома.

Задание 3

На основе данных, представленных в файле авто.csv, установить, имеется ли зависимость между стоимостью автомобиля и обивкой салона.

Задание 4

На основе данных, представленных в файле авто.csv, установить, имеется ли зависимость между стоимостью автомобиля и его маркой.

Задание 5

Исследовательская группа проводит эксперимент по изучению свойств стирального порошка. Пятьдесят восемь образцов пятен однородного происхождения подвергаются стирке с помощью изучаемого стирального порошка и хозяйственного мыла (по 29 образцов на каждый способ). Результаты эксперимента приведены в таблице 11.1:

Таблица 11.1 – Данные для задания 5

	отстиралось	не отстиралось
с порошком	19	10
со стиральным мылом	14	15

Выяснить, имеется ли связь результатов стирки с использованием нового вида порошка.

12 Лабораторная работа № 12. Расчет числовых характеристик корреляционной зависимости в пакете R

Задание 1

Предполагая, что между данными, представленными в таблице 12.1, имеет место линейная зависимость, найти параметры уравнения регрессии, построить корреляционное поле и график регрессии.

Таблица 12.1 – Данные для задания 1

Y/X	20	25	30	35	40
16	4	6			
26		8	10		
36			32	3	9
46			4	12	6
56				1	5

Задание 2

Предполагая, что между данными, представленными в таблице 12.2, имеет место линейная зависимость, найти параметры уравнения регрессии, построить корреляционное поле и график регрессии.

Таблица 12.2 – Данные для задания 2

Y\X	5	10	15	20	25	30	35	40
100	2	1						
120	3	4	3					
140			5	10	8			
160				1		6	1	1
180							4	1

Задание 3

Предполагая, что между данными, представленными в таблице 12.3, имеет место линейная зависимость, найти параметры уравнения регрессии, построить корреляционное поле и график регрессии.

Таблица 12.3 – Данные для задания 3

Y\X	18	23	28	33	38	43	48
125		1					
150	1	2	5				
175		3	2	12			
200			1	8	7		
225					3	3	
250						1	1

Задание 4

Предполагая, что между данными, представленными в таблице 12.4, имеет место линейная зависимость, найти параметры уравнения регрессии, построить корреляционное поле и график регрессии.

Таблица 12.4 – Данные для задания 4

Y\X	5	10	15	20	25	30	35
100						6	1
120						4	2
140			8	10	5		
160	3	4	3				
180	2	1		1			

13 Лабораторная работа № 13. Множественная линейная регрессия, расчет параметров в пакете R

Задание 1

Построить модель множественной линейной регрессии, используя данные из файла квартиры.csv, определить степень влияния каждого из имеющихся параметров на стоимость квартиры.

Задание 2

Построить модель множественной линейной регрессии, используя данные из файла население.csv, определить степень влияния каждого из имеющихся параметров на численность населения Республики Беларусь.

14 Лабораторная работа № 14. Подбор модели нелинейной регрессии, расчет ее параметров в пакете R

Задание 1

Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ по данным, приведенным в корреляционной таблице 14.1.

Таблица 7

$y_i \backslash x_i$	2	3	5
25	20		
45		30	1
110		1	48

Задание 2

Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице 14.2.

Таблица 14.2 – Данные для задания 2

$y_i \backslash x_i$	0	1	2	3	4
0	18	1	1		
3	1	20			
5	3	5	10	2	
10			7	12	
17					20

Задание 3

Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице 14.3.

Таблица 14.3 – Данные для задания 3

$y_i \backslash x_i$	0	4	6	7	10
7	19	1	1		
13	2	14			
40		3	22	2	
80				15	
200					21

Задание 4

Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице 14.4.

Таблица 14.4 – Данные для задания 4

$y_i \backslash x_i$	0	4	5
1	50	5	1
35		44	
50		5	45

Задание 5

Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице 14.5.

Таблица 14.5 – Данные для задания 5

$y_i \backslash x_i$	0	1	2	3	4
10	20	5			
11	7	15	3	1	
20		3	17	4	
35			8	13	7
50				5	42

Задание 6

Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице 14.6.

Таблица 14.6 – Данные для задания 6

$y_i \backslash x_i$	7	8	9
200	41	7	
300	1	52	1
400		8	40

15 Лабораторная работа № 15. Однофакторный дисперсионный анализ в пакете R

Задание 1

Используя данные из файла население.csv, определить степень влияния численности населения Республики Беларусь на количество посещений учреждений культуры.

Задание 2

Используя данные из файла население.csv, определить степень влияния заболеваемости алкоголизмом на численность населения Республики Беларусь.

Задание 3

Используя данные из файла население.csv, определить степень влияния заболеваемости алкоголизмом на количество преступлений.

Задание 4

Используя данные из файла население.csv, определить степень влияния численности обучающихся в учреждениях высшего образования на численность обучающихся в профессионально-технических учреждениях образования.

Задание 5

Используя данные из файла квартиры.csv, определить степень влияния площади квартиры на ее стоимость.

Задание 6

Используя данные из файла квартиры.csv, определить степень влияния года постройки дома на стоимость квартиры.

Задание 7

Используя данные из файла квартиры.csv, определить степень влияния этажа, на котором расположена квартира, на ее стоимость.

Задание 8

Используя данные из файла авто.csv, определить степень влияния типа салона на стоимость автомобиля.

16 Лабораторная работа № 16. Многофакторный дисперсионный анализ в пакете R

Задание 1

Проанализировать параметры множественной регрессии по данным, представленным в файле квартиры.csv. В качестве результирующего параметра принять стоимость квартиры.

Задание 2

Проанализировать параметры множественной регрессии по данным, представленным в файле население.csv. В качестве результирующего параметра принять численность населения.

Задание 3

Проанализировать параметры множественной регрессии по данным, представленным в файле авто.csv. В качестве результирующего параметра принять стоимость автомобиля.

17 Лабораторная работа № 17. Проведение полноформатного статистического исследования и обработка данных средствами пакета R

Задание 1

Обработать данные, полученные в результате исследования, по нижеприведённой схеме.

- 1 Создать файл, содержащие статистический материал.
- 2 Импортировать данные в пакет R.
- 3 Произвести предварительную обработку и анализ данных.

- 4 Вывести графики плотности вероятности, функции распределения, гистограммы относительных частот для исследуемого признака.
- 5 Определить, присутствуют ли в выборке выбросы, выяснить причину их появления.
- 6 Вычислить числовые характеристики выборки.
- 7 Выяснить, какое распределение могут иметь генеральная совокупность и выборка.
- 8 Смоделировать данное распределение.
- 9 Сравнить эмпирическое распределение с теоретическим графически.
- 10 Проверить гипотезу о распределении выборки.
- 11 Проверить гипотезы о значимости выборочного среднего и дисперсии.
- 12 Определить, насколько сильно взаимосвязаны с собой исследуемые факторы, сделать выводы.
- 13 При наличии связи найти коэффициент корреляции, построить уравнение регрессии.

Список литературы

1. **Вентцель, Е. С.** Теория вероятностей и ее инженерные приложения: учебное пособие / Е. С. Вентцель. - 5-е изд., стер. - Москва : КНОРУС, 2013. – 480 с.
2. **Белько, И. В.** Теория вероятностей, математическая статистика, математическое программирование: учебное пособие / И. В. Белько. - Минск ; Москва : Новое знание; ИНФРА-М, 2016. – 299 с.
3. **Барботько, А. И.** Статистические алгоритмы обработки результатов экспериментальных исследований в машиностроении: учебное пособие / А. И. Барботько. - Старый Оскол: ТНТ, 2017. – 404 с.
4. **Матальцкий, М. А.** Теория вероятностей и математическая статистика : учебник для вузов / М. А. Матальцкий. - Минск: Вышэйшая школа, 2017. – 591 с.
5. **Карманов, Ф. И.** Статистические методы обработки экспериментальных данных с использованием пакета MathCad: учебное пособие / Ф. И. Карманов. - Москва: КУРС: ИНФРА-М, 2018. – 208 с.