

Министерство образования и науки РФ
ФГБОУ ВО «Кубанский государственный технологический университет»
Институт компьютерных систем и информационной безопасности
Кафедра информатики и вычислительной техники

Отчет

По лабораторной работе № 1
По дисциплине анализ и визуализация данных

Выполнил студент группы 19-КМ-ИБ1:

Кирмасов Б.В.

Преподаватель:

Решетняк М.Г.

Краснодар 2020

Тема: «Основы работы с R. Обработка статистических данных».

Цель работы: «Ознакомиться с интерфейсом RStudio, научиться работать в режиме консоли путем написания скриптов, а также подключать внешние пакеты, изучить основные методы обработки статистических данных».

Отчёт о работе.

Задание

1. Загрузить данные для своего варианта в переменную-вектор.
2. Получить справочную информацию по своим данным, просмотреть их содержимое.
3. Проверить, есть ли среди данных пропуски.
4. Создать новую переменную-вектор, в которой будут 1, если значение в исходном векторе больше среднего, и -1, если значение переменной меньше среднего, и 0, если значение равно среднему.
5. Вывести описательную статистику.
6. Построить графики абсолютных частот и плотности распределения.

```
1. df <- read.csv("train.csv")
2. glimpse(df)
   describe(df)
3. impute(df$Age, mean)
4. df <- within(df, {
  agecat <- NA
  agecat[Age > mean(Age, na.rm = T)] <- 1
  agecat[Age == mean(Age, na.rm = T)] <- 0
  agecat[Age < mean(Age, na.rm = T)] <- -1
})
df[,"agecat"]
```

```
Console Terminal x Jobs x
D:/Магистратура/Анализ и визуализация данных/Ir1/
> df["agecat"]
  agecat
1     -1
2      1
3     -1
4      1
5      1
6     NA
7      1
8     -1
9     -1
10    -1
11    -1
12     1
13    -1
14     1
15    -1
16     1
17    -1
18    NA
19     1
20    NA
21     1
22     1
23    -1
24    -1
25    -1
26     1
27    NA
28    -1
29    NA
30    NA
31     1
```

Рисунок 1 – Содержимое вектора agecat

5. describe(df)

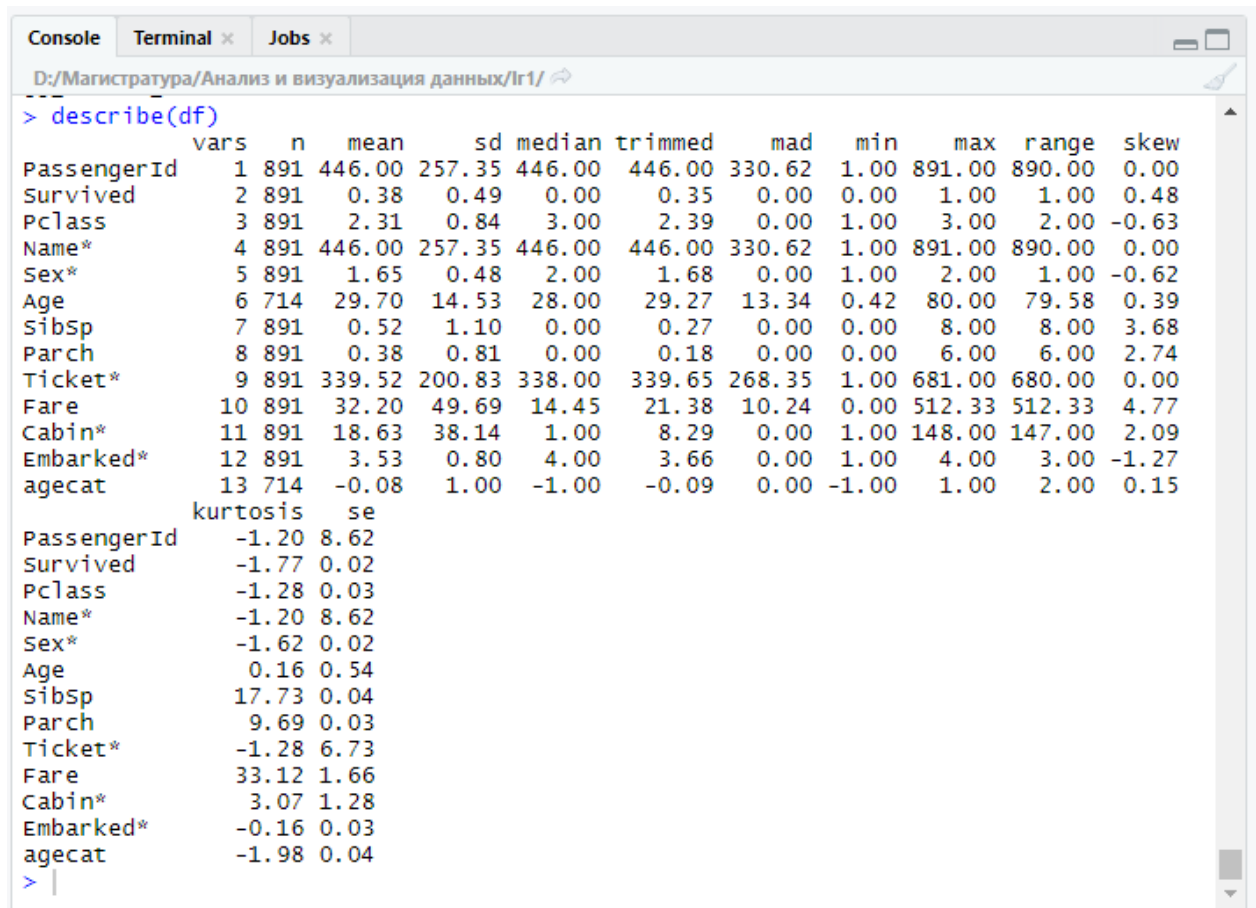


Рисунок 2 – Описательная статистика

6. `qplot(data=df, Age, xlab="Все возрасты", ylab="Количество одинаковых возрастов", main="Данные по возрастам")`

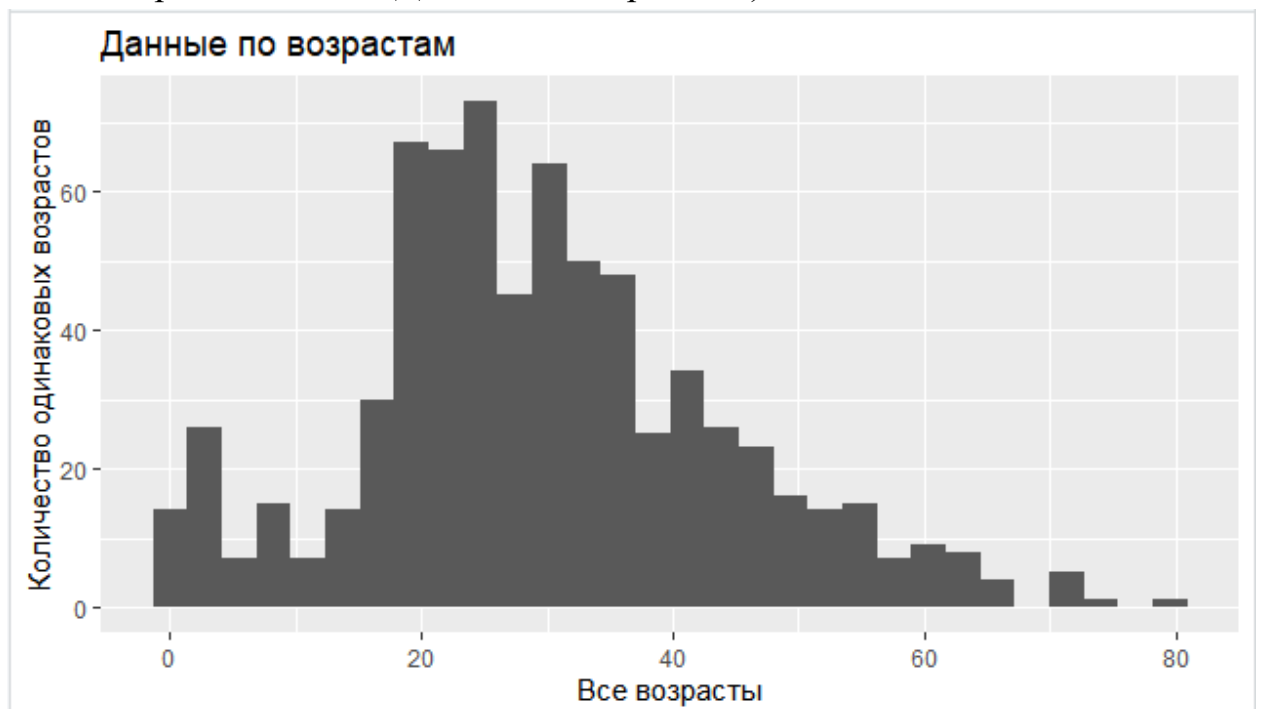


Рисунок 3 – Гистограмма абсолютных частот для переменной Age

```
hist(df$Age, probability = TRUE, col="grey")
```

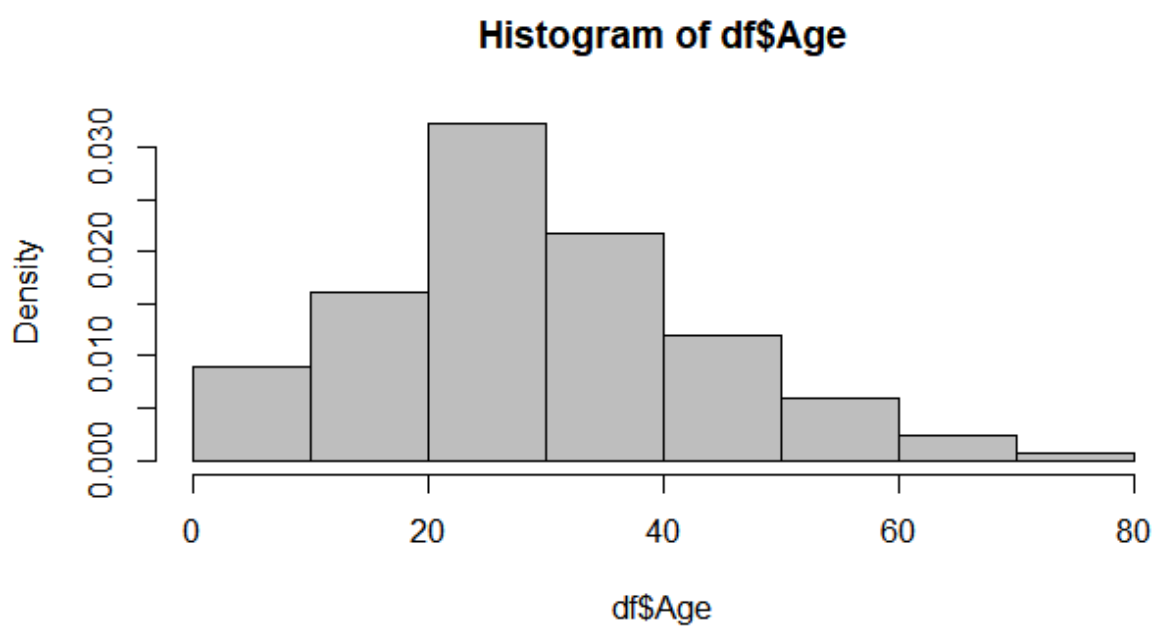


Рисунок 4 – Гистограмма плотности распределения для переменной Age

Вывод: Мы ознакомились с интерфейсом RStudio, научились работать в режиме консоли путем написания скриптов, а также подключили внешние пакеты, изучили основные методы обработки статистических данных.