

## Практическая работа №8.

### Описательные статистики в R

**Цель:** освоить навыки применения языка программирования **R**, основные и специальные функции, пакеты **R** для расчета параметров описательной статистики, научиться строить график частотного распределения.

### Задания

1. Создать пять векторов, в каждом из которых 17 числовых значений. Тему выбрать самостоятельно из области бизнеса.

Например, 5 детских магазинов и оценки 17 покупателей в десятибальной системе:

```
DetskiyMir<-c(8,5,10,8,9,6,9,10,9,10,8,9,9,10,9,9,9)
```

```
DochkiSyinochki<-c(6,7,10,10,9,7,10,8,9,8,10,10,8,8,6,9,10)
```

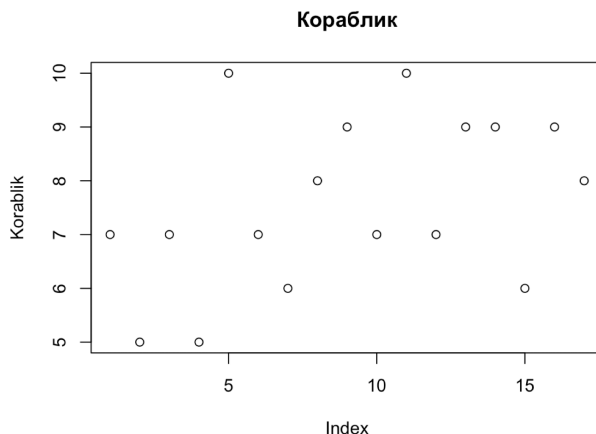
```
BabyKid<-c(6,8,9,10,7,10,9,9,6,10,8,9,7,10,8,9,9)
```

```
Olant<-c(9,8,10,7,9,8,9,10,8,10,8,9,10,8,6,9,6)
```

```
Korablik<- c(7,5,7,5,10,7,6,8,9,7,10,7,9,9,6,9,8)
```

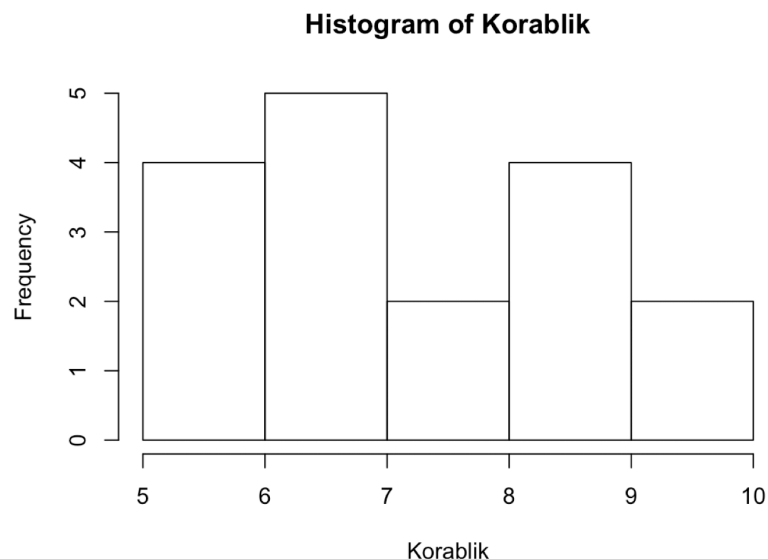
2. Вычислить меры центральной тенденции для всех созданных переменных (среднее арифметическое, моду, медиану) и общую статистику с использованием функции **summary()**.
3. По результатам полученных данных сделать выводы.
1. Построить графики, которые иллюстрируют разброс значений переменных, используя функцию **plot()**.

```
plot(Korablik)  
title(main="Кораблик")
```



2. Построить гистограммы переменных.

```
hist(Korablik)
```



3. Объединить пять созданных переменных в таблицу **Tab** с помощью функции **data.frame()**, открыть таблицу в консоли **RStudio**.

```
KidShops <- data.frame (DetskiyMir=DetskiyMir,
                        DochkiSyinochki=DochkiSyinochki, BabyKid = BabyKid,
                        Olant = Olant, Korablik = Korablik)
```

KidShops

	DetskiyMir	DochkiSyinochki	Babykid	Olant	Korablik
1	8	6	6	9	7
2	5	7	8	8	5
3	10	10	9	10	7
4	8	10	10	7	5
5	9	9	7	9	10
6	6	7	10	8	7
7	9	10	9	9	6
8	10	8	9	10	8
9	9	9	6	8	9
10	10	8	10	10	7
11	8	10	8	8	10
12	9	10	9	9	7
13	9	8	7	10	9
14	10	8	10	8	9
15	9	6	8	6	6
16	9	9	9	9	9
17	9	10	9	6	8

4. Рассчитать показатели описательной статистики данных таблицы **Tab**.

### **summary(KidShops)**

DetskiyMir	DochkiSyinochki	BabyKid	olant
Min. : 5.000	Min. : 6.000	Min. : 6.000	Min. :
1st Qu.: 8.000	1st Qu.: 8.000	1st Qu.: 8.000	1st Qu.:
Median : 9.000	Median : 9.000	Median : 9.000	Median :
Mean : 8.647	Mean : 8.529	Mean : 8.471	Mean :
3rd Qu.: 9.000	3rd Qu.:10.000	3rd Qu.: 9.000	3rd Qu.:
Max. :10.000	Max. :10.000	Max. :10.000	Max. :

korablik

Min. : 5.000

1st Qu.: 7.000

Median : 7.000

Mean : 7.588

3rd Qu.: 9.000

Max. :10.000

5. Сравнить результаты п.7 с результатами, полученными в п.2.
6. Рассчитать показатели изменчивости для всех переменных таблицы **Tab**.

**Пример.** Дисперсия для всех переменных набора данных **KidShops**.

### **var(KidShops)**

	DetskiyMir	DochkiSyinochki	BabyKid	olant	korablik
DetskiyMir	1.8676471	0.6360294	0.1764706	0.5514706	0.8455882
DochkiSyinochki	0.6360294	2.0147059	0.4852941	0.1102941	0.4816176
BabyKid	0.1764706	0.4852941	1.7647059	-0.1102941	-0.6691176
olant	0.5514706	0.1102941	-0.1102941	1.6397059	0.4558824
korablik	0.8455882	0.4816176	-0.6691176	0.4558824	2.5073529

- 6.1.Стандартное отклонение для переменной DetskiyMir таблицы **KidShops**.

### **sd(KidShops\$DetskiyMir)**

[1] 1.366619

7. Рассчитать квартили, процентиля (25%, 90%) для всех переменных таблицы **Tab**.

### **quantile(KidShops\$DetskiyMir)**

0%	25%	50%	75%	100%
5	8	9	9	10

**quantile(KidShops\$DetskiyMir,probs=0.1) #10%-ный**

10%

7.2

8. Вычислить минимальное и максимальное значения для третьей переменной таблицы **Tab**.
9. Определить порядковый номер элементов второго столбца переменной **Tab**, обладающих минимальным, максимальным значениями.
10. Загрузить в новую переменную **Avto** таблицу данных **avto.xlsx**. Используя функцию **tapply()** найти среднюю стоимость (**cost**) автомобилей разного класса (**class**).
11. Найти среднюю стоимость (**cost**) автомобилей разного класса (**class**) и потребляющих разный вид топлива (**fuel**) из таблицы данных **Avto**.
12. Придумайте произвольную функцию, создайте её в **R**.
13. Найдите значение функции из п.15 для переменной стоимость (**cost**) автомобилей разного класса (**class**).
14. Установить и загрузить в **R** пакет **moments**.
15. Рассчитать коэффициенты эксцесса, асимметрии для переменной стоимости автомобилей (**cost**).
16. Используя функцию **aggregate()** разбить переменную **body** (кузов) по классу автомобилей (**class**). Рассчитать среднее значение по этой переменной в группах.
17. Заменить названия "Group.1", "x" на "body", "class".
18. С помощью функции **aggregate()** сделать разбиение переменной **cost** по двум переменным (**class**, **fuel**) двумя способами (стандартной формой, с помощью формулы). Сравнить результаты, полученные двумя способами и сделать выводы.
19. Установить и загрузить пакет **psych()**.
20. Используя функцию **describe()**, найти статистики для всех переменных таблицы данных **Avto**.
21. Исключить из таблицы данных **Avto** все нечисловые переменные и посмотреть статистики оставшихся переменных.
22. С помощью команды **describeBy()** найти описательные статистики, рассчитанные по группам **class** и **fuel**. Результат получить в двух

вариантах: в виде списка и матрицы. Исключить из расчетов все нечисловые переменные.

### Контрольные вопросы

1. Перечислите основные статистические характеристики и дайте краткую характеристику каждой из них.
2. Какая команда вычисляет арифметическую среднюю?
3. Чем отличаются абсолютная, относительная и процентная частоты?
4. Дайте понятие медианы.
5. Дайте определение моды. Как найти моду в R?
6. Для чего предназначена команда **tapply**?
7. Как вычислить статистические характеристики данных, имеющих пропущенные значения?
8. Для чего предназначена команда **attach**?
9. Какой функцией можно выполнить расчет основных параметров описательной статистики?
10. Назовите предназначение пакета **psych**.
11. Какое назначение у функции **aggregate()**?
12. Перечислите основные аргументы команды **describeBy()**.
13. Какими способами можно записать команду **aggregate()**?

### Домашнее задание

1. Оформить отчет по практической работе (шаблон отчета взять из практической работы №3).
2. Вычислить основные описательные статистики для переменных таблицы данных, взятых с сайта **data.gov.ru** (см. практическую работу №2).