Ордена Трудового Красного Знамени

Федеральное государственное образовательное бюджетное учреждение

высшего профессионального образования

**МОСКОВСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**

**СВЯЗИ И ИНФОРМАТИКИ**

**Факультет информационных технологий**

Кафедра «Прикладная информатика»

**Отчет по практическому занятию №3**

по дисциплине

**«Теория систем и системный анализ»**

Вариант № 22

Выполнил: студент 1 курса

Хауш А.

Принял:

Доцент Тутова Н.В.

Москва 2021

**Задание**

В выборке представлено число миграций виртуальных машин.

Выберите свой вариант в соответствии с вашим номером в списке группы.

1. Постройте вариационный ряд.
2. Рассчитайте точечные оценки параметров распределения, такие как

* оценка математического ожидания;
* оценка дисперсии;
* и среднеквадратического отклонения;
* оценка моды;
* оценка медианы;
* оценка квартилей;
* оценка коэффициента ассиметрии;
* оценка коэффициента вариации;
* оценка эксцесса;
* оценка абсолютного медианного отклонения.

Какие выводы можно сделать по этим показателям о виде распределения?

1. Постройте гистограмму. Какие выводы можно сделать по виду гистограммы?
2. Пойте ящичковую диаграмму. Какие выводы можно сделать по виду диаграммы?

Расчеты проводить в R.

**Выполнение работы**

В выборке представлено число миграций виртуальных машин.

Данные записаны в виде текстового файла.

Для импорта значений используем команду:

> dataframe = read.table("C:/Users/USER/Downloads/22\_data.txt", quote = "\"", comment.char = "" )

В результате получаем объект типа data.frame. Для создания числового вектора выполняем:

> var\_range = c(dataframe[,1])

Объект выводим на экран:

> var\_range

[1] 145 72 66 48 43 47 35 28 23 30 25 15 34 24 23 17

[17] 32 24 40 35 41 53 27 54 59 35 63 21 27 54 37 40

[33] 33 35 27 39 33 29 47 32 31 26 21 18 26 15 27 47

[49] 42 42 54 41 28 34 42 30 21 26 30 41 26 50 38 44

[65] 28 24 31 28 41 20 66 48 24 18 34 50 25 28 23 22

[81] 34 32 30 28 35 20 44 17 27 13 46 54 30 28 36 12

[97] 24 16 17 19 4 14 13 44 13 62 35 48 19 50 46 47

[113] 24 21 17 40 30 30 28 32 28 28 46 20 44 24 14 18

[129] 12 10 2 11 14 20 40 23 41 15 40 40 17 27 17 20

[145] 36 43 34 21 21 48 31 13 6 17 25 24 28 40 31 29

[161] 30 22 63 43 45 37 16 27 27 54 19 42 39 18 30 45

[177] 19 33 16 30 23 21 25 62 29 52 18 26 15 27 27 29

[193] 21 37 26 19 22 17 13 43 20 46 17 50 15 38 31 29

[209] 22 19 25 23 27 17 16 21 42 27 32 20 17 19 17 22

[225] 49 17 46 16 28 36 19 19 27 36 22 20 29 18 34 21

[241] 28 20 17 32 14 7 4 2 17 42 40 22 45 24 34 30

[257] 32 17 41 15 31 24 39 57 26 34 44 34 21 46 21 41

[273] 18 39 44 32 49 22 29 19 46 20 26 27 20 33 16

Для построения **вариационного ряда** применяем оператор **sort()**, значения записываем в новую переменную.

> sort\_var\_range = sort(var\_range)

> sort\_var\_range

[1] 2 2 4 4 6 7 10 11 12 12 13 13 13 13 13 14

[17] 14 14 14 15 15 15 15 15 15 16 16 16 16 16 16 17

[33] 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 18

[49] 18 18 18 18 18 18 19 19 19 19 19 19 19 19 19 19

[65] 20 20 20 20 20 20 20 20 20 20 20 21 21 21 21 21

[81] 21 21 21 21 21 21 21 22 22 22 22 22 22 22 22 23

[97] 23 23 23 23 23 24 24 24 24 24 24 24 24 24 24 25

[113] 25 25 25 25 26 26 26 26 26 26 26 26 27 27 27 27

[129] 27 27 27 27 27 27 27 27 27 27 28 28 28 28 28 28

[145] 28 28 28 28 28 28 28 29 29 29 29 29 29 29 30 30

[161] 30 30 30 30 30 30 30 30 30 31 31 31 31 31 31 32

[177] 32 32 32 32 32 32 32 33 33 33 33 34 34 34 34 34

[193] 34 34 34 34 35 35 35 35 35 35 36 36 36 36 37 37

[209] 37 38 38 39 39 39 39 40 40 40 40 40 40 40 40 41

[225] 41 41 41 41 41 41 42 42 42 42 42 42 43 43 43 43

[241] 44 44 44 44 44 44 45 45 45 46 46 46 46 46 46 46

[257] 47 47 47 47 48 48 48 48 49 49 50 50 50 50 52 53

[273] 54 54 54 54 54 57 59 62 62 63 63 66 66 72 145

**Оценка математического ожидания** представляется в виде среднего арифметического выбранных значений и рассчитывается с помощью команды **mean().**

> mean(var\_range)

[1] 30.29965

**Дисперсия**, или мера разброса относительно математического ожидания, вычисляется функцией **var().**

> var(var\_range)

[1] 206.2456

**Среднеквадратическое отклонение** представляет собой корень квадратный из дисперсии и используется как показатель рассеивания значений случайной величины.

> sqrt(var(var\_range))

[1] 14.36125

**Мода** определяет самое часто встречающееся в выборке значение, для ее расчета применяем функцию **table()**, которая подсчитывает количество уникальных значений

>table\_var\_range = table(var\_range)

>table\_var\_range

2 4 6 7 10 11 12 13 14 15 16 17 18 19 20 21 22 23

2 2 1 1 1 1 2 5 4 6 6 16 7 10 11 12 8 6

24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41

10 5 8 14 13 7 11 6 8 4 9 6 4 3 2 4 8 7

42 43 44 45 46 47 48 49 50 52 53 54 57 59 62 63 66 72

6 4 6 3 7 4 4 2 4 1 1 5 1 1 2 2 2 1

145

1

> MODA = table\_var\_range[which.max(table\_var\_range)]

>MODA

17

16

В нашем случае значение «17» встречается в выборке наиболее часто (16 раз), т.е. можно сделать вывод о одномодальном распределении.

**Медиана** представляет собой значение, которое отсекает половину упорядоченной выборки и рассчитывается функцией **median()**

>median(mass\_peremen)

[1] 28

Т.к. медиана слегка меньше средней (30,29) можно сделать вывод о том, что данное распределение стремится к нормальному.

**Оценка квартилей** проводится с помощью функций **IQR()**, которая рассчитывает межквартильный разброс и функции **quantile()**

> IQR(var\_range)

[1] 19.5

> quantile(var\_range)

0% 25% 50% 75% 100%

2.0 20.0 28.0 39.5 145.0

Разброс значений между 3 и 4 квартилем свидетельствует о наличии правого «хвоста»

**Коэффициент ассиметрии** показывает, какой из «хвостов» распределения длиннее и определяется с помощью функции **skew()**

> skew(var\_range)

[1] 2.113863

**Коэффициент вариации** показывает, какую долю составляет средний разброс от среднего значения величины и рассчитывается, как

>sd(var\_range)/mean(var\_range)\*100

[1] 47.39742

**Коэффициент эксцесса** характеризует степень остроты пика распределения и толщину хвостов и рассчитывается функцией **kurtosi()**

> kurtosi(var\_range)

[1] 13.01466

**Абсолютное медианное отклонение** используется вместо среднего отклонения, когда крайние значения из области отклонений должны оказывать меньшее влияние на величину отклонения и рассчитывается с помощью **mad()**

>mad(var\_range)

[1] 13.3434

**Гистограмма** распределения создается с помощью функции **hist()** и представлена на Рис.1. По ее виду можно сделать вывод о том, что большинство значений величины лежит в пределах от 0 до 60, пик приходится на интервал 10-20, а само распределение имеет правый хвост.

> hist(var\_range)

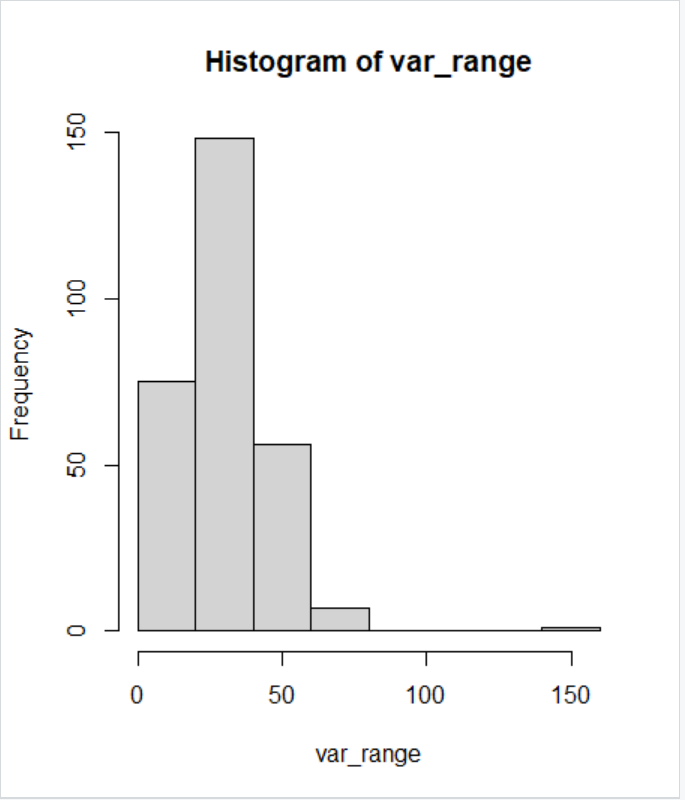


Рис. 1

**Ящичковая диаграмма** строится функцией **boxplot()**и представлена на Рис.2.

«Ящик» сформирован нижним и верхним квартилем, утолщенной линией обозначена медиана. Выбросы расположены выше значения верхнего уса.

> boxplot(var\_range)

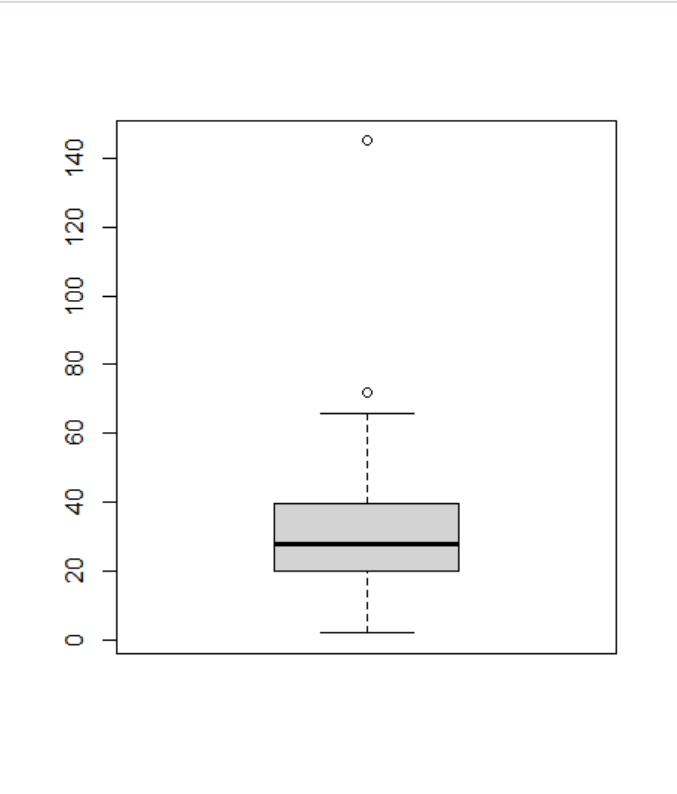


Рис. 2

Для дополнительной визуализации плотности распределения можно применить следующую функцию (Рис. 3):

> plot(density(var\_range))

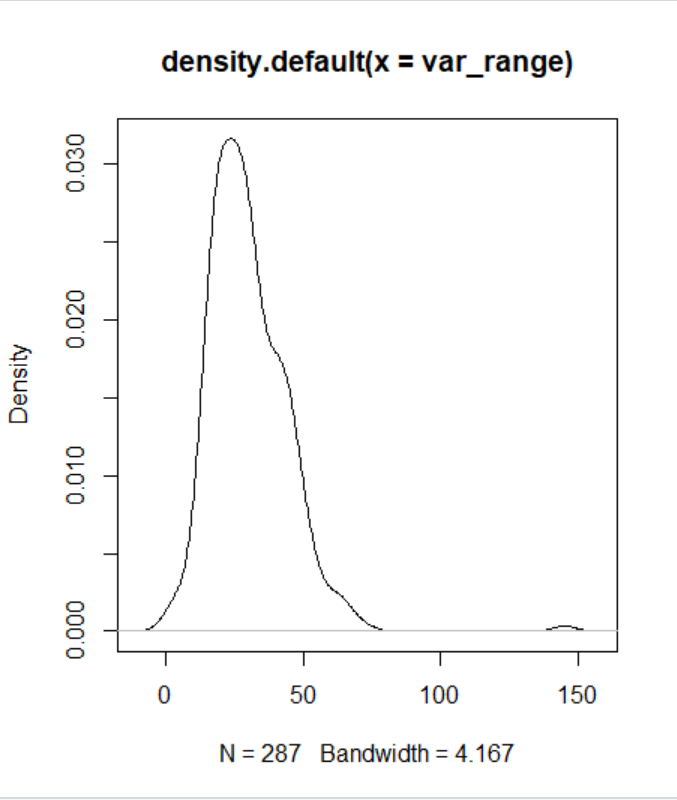


Рис. 3

Проверка гипотезы о нормальности распределения числа атак критерием **Пирсона и Колмогорова-Смирнова** осуществляется функцией **ks.test()**

>ks.test(var\_range, "rnorm")

One-sample Kolmogorov-Smirnov test

data: var\_range

D = 3.1338, p-value < 2.2e-16

alternative hypothesis: two-sided

Т.к. p-value меньше критического значения (0,05), то нулевая гипотеза не подтверждается.

Выводы: в работе проведена оценка числовых характеристик случайной величины, показано, что гипотезу о нормальном распределении случайной величины принять нельзя.