Федеральное государственное образовательное бюджетное учреждение высшего профессионального образования «ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

(Финансовый университет)

Департамент математики

Дисциплина «Программирование в среде R»

П.Б. Лукьянов

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНОЙ РАБОТЫ № 3

Векторы и операции с ними

Для студентов, обучающихся по направлению подготовки «Прикладная математика и информатика» (программа подготовки бакалавра)

Для быстрой и эффективной обработки больших массивов информации простейших типов данных (целые и действительные числа, строки, логические значения TRUE / FALSE) оказалось недостаточно. В R определены более сложные структуры данных, для работы с которыми используются специальные функции.

Цель лабораторной работы заключается в изучении самого распространенного контейнера для хранения данных – вектора.

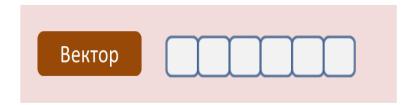
1. Понятие вектора

Вспомним математику. В школьной математике используются две большие категории данных: скалярные величины и векторы. Скаляром называют величину, которая имеет только одну характеристику – численное значение. Все операции с числами – это скалярные операции.

Вектор — математический объект, имеющий кроме величины еще и направление. Если в пространстве задана система координат, то вектор однозначно задаётся набором своих координат. Поэтому в математике, информатике и программировании упорядоченный набор чисел часто тоже называют вектором.

Дадим следующее определение вектору. **Вектор** – это объект, объединяющий элементы одного типа. Вектор предназначен для хранения данных, состоящих из нескольких однотипных значений. Сочетания разных типов данных в векторе не допустимы. Любой вектор характеризуется типом хранимых данных и своей длиной. Тип вектора х проверяется функциями typeof(x), mode(x); длину вектора можно узнать, вызвав функцию length(x) (см. рис. 1).

Например, в одном векторе может храниться возраст всех студентов группы, в другом векторе — их имена и фамилии, в третьем векторе — средний бал успеваемости по каждому студенту, в четвертом векторе — номера их телефонов.



Вектор – именованный одномерный объект, содержащий набор однотипных элементов (или числовые, или логические, или текстовые значения).

Тип вектора x проверяется функциями typeof(x), mode(x)

Длина вектора проверяется функцией length(x)

Элементы вектора проиндексированы, к каждому элементу вектора можно обратиться по его индексу: $\mathbf{x}[\mathbf{i}]$

Рис. 1. Вектор – структура для хранения данных одного типа

В R вообще нет скалярных величин. Следовательно, изначально в R все создаваемые объекты — уже векторы. Таким образом, при изложении предыдущего материала мы рассматривали и создавали векторы, состоящие из одного элемента. Принципиальное решение создателей языка о том, что R будет работать только с векторами, связано с тем, что задачи обработки табличной информации решаются проще, если все данные хранить и представлять векторами.

В соответствии с этим подходом в R были разработаны функции, реализующие быстрые векторные вычисления для обработки данных и представления результатов. Как следствие, традиционные решения программистов с использованием циклов для перебора значений в R не

используют, так как циклы по сравнению с векторными операциями работают гораздо медленнее.

2. Создание векторов

Создадим вектор w, в котором будут храниться четные числа от 1 до 10. Для создания такого вектора вызовем функцию c():

$$w < -c(2, 4, 6, 8, 10)$$

Функция c() (от слова concatenate – связывать) создает объект «вектор». Таким образом, нами был создан вектор w из 5 элементов. Как узнать, какому типу принадлежат элементы w?

Создать вектор можно, используя разные типы данных, но в итоге вектор будет одного, определенного типа, так как все элементы вектора будут приведены к одному типу. Проверьте типы следующих векторов:

Создать вектор z с пустыми (нулевыми) значениями можно с помощью функции vector():

Выясните, какого типа и какой длины был создан вектор z.

Если требуется вектор-заготовка для хранения данных определенного типа и определенного размера, вызывают функцию vector() с заданием нужных параметров, причем при задании значений параметрам функции vector() используется традиционное равенство:

```
z2 <- vector(mode="numeric", length=129)
z3 <- vector(mode="logical", length=22)
z4 <- vector(mode="integer", length=34)
z5 <- vector(mode="character", length=0)
z6 <- vector(mode="double", length=19)</pre>
```

При вызове функции vector() мы можем менять местами параметры функции, ошибки не будет, так как в любом случае значения параметров определяются однозначно. У некоторых параметров есть предопределенное значение, так называемое «значение по умолчанию», в этом случае параметр можно не указывать.

```
z6 <- vector(length=19, mode="double")
z7 <-vector(length=19) # тип данных задается по умолчанию
z8 <- vector(mode="double") # длина вектора задается по умолчанию
```

Например, вектор z4 предназначен для хранения целых значений. Выясните, к каким типам принадлежат вектора z2, z7. Какова длина z8?

Создание векторов определенного типа можно упростить:

```
z2 <- numeric(length=129)
z3 <- logical(length=22)
z4 <- integer(length=34)
z5 <- character(length=0)</pre>
```

Чтобы узнать, что содержится в объекте w, достаточно ввести w и нажать Enter. Проверьте, какие вектора будут созданы при выполнении команд

Вектор из последовательности чисел с шагом 1 можно создать еще проще:

$$w < -4:10$$

Обратите внимание, что границы диапазона значений не обязательно должны быть целочисленными. Какой вектор получится, если задать границы действительными числами?

Для дальнейшего использования векторов нужно выяснить приоритет при выполнении арифметических операций и операции ":". Для этого выполните операции и проанализируйте результат:

```
> 1:10-1
```

> 1: (10-1)

У каких операций приоритет выше?

Если нужен вектор, состоящий из повторяющихся значений, используется функция rep() (от repeat – повторять):

$$w \leftarrow rep(TRUE, 8)$$

Можно дублировать последовательности значений:

$$w2 < -rep(3:-2, 3)$$

Часто возникает задача получения вектора, содержащего некоторую последовательность значений. Функция seq() (от sequence – последовательность) решает эту задачу (см. рис. 2).

В функции seq() у каждого параметра есть значение по умолчанию. Определите эти значения для параметров to, from, by. Еще один параметр функции seq() задает общее количество элементов последовательности, это параметр length.out:

 $w \leftarrow seq(4, from=12, length.out = 10)$

```
> w <- seq(from = -11, to = 5, by = 2.5)

> w

[1] -11.0 -8.5 -6.0 -3.5 -1.0 1.5 4.0

> w <- seq(to = -11, from = 5, by = -2.5)

> w

[1] 5.0 2.5 0.0 -2.5 -5.0 -7.5 -10.0

> |
```

Рис. 2. Разбиение отрезка начинается от значения «from»

Использование при вызове функции всех четырех параметров приведет к ошибке:

$$w < -seq(from = -11, to = 5, by = 2.5, length.out = 22)$$

3. Считывание вектора с клавиатуры или из файла

Порой проще создать нужный вектор, считывая вводимые значения с клавиатуры или из файла данных напрямую. Этот подход реализован в функции scan(). У этой функции более двадцати (!) параметров; если запустить scan(), не задав значения ни одному параметру, будет реализован следующий сценарий:

- активируется режим считывания действительных чисел с клавиатуры
- Пользователь вводит число, нажимает Enter, число заносится в элемент вектора
- Пользователь может набрать несколько действительных чисел в строке, разделяя их произвольным количеством пробелов, все числа последовательно займут соответствующие элементы вектора
- Считывание чисел выполняется до тех пор, пока Пользователь не нажмет Enter два раза (рис. 3).

Обратите внимание, что вместо одного из отсутствующих значений было введено NA; был использован маркер Inf, говорящий об уходе значения за пределы шкалы измерения.

Один из недостатков использования scan() заключается в том, что при вводе данных с помощью этой функции легко сделать ошибку без возможности ее исправления. Для редактирования содержимого вектора предназначена функция fix(), где единственный обязательный параметр – это имя вектора. fix() запускает простейший редактор, в котором можно делать любые исправления, а результат сохранить (рис. 4).

```
> scan()
1: -9.9
3: 223.009
4: 12 7 7
                         Inf 0 5 5 NA -72.007 0 0
15: 19
16: 220000 -4.4
18:
Read 17 items

      -9.900
      0.000
      223.009
      12.000

      7.000
      7.000
      Inf
      0.000

      5.000
      5.000
      NA
      -72.007

      0.000
      19.000
      220000.000

  [1] -9.900
  [5]
 [9]
                5.000
[13]
                0.000
[17]
                -4.400
```

Рис. 3. Ввод данных с использованием функции scan()

Фактически при своем вызове функция fix() вызывает функцию edit() для редактирования переданного в качестве параметра вектора, и можно было бы заменить вызов fix(имя_вектора) вызовом edit(имя_вектора), так как редактор используется один и тот же (рис. 4).

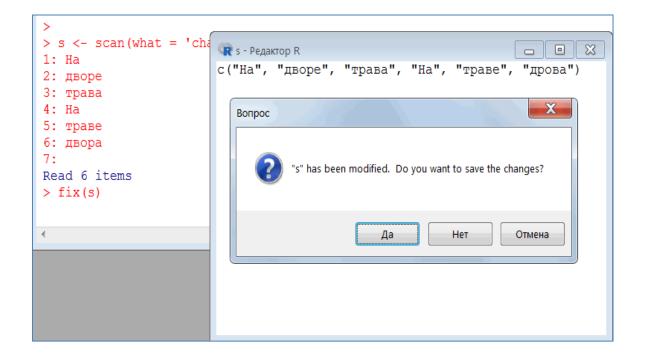


Рис. 4. Коррекция содержимого вектора функцией fix()

4. Генерация случайных значений и сортировка

Частая задача при обработке данных — выполнить сортировку вектора в ту или иную сторону. Рассмотрим, как получить вектор из случайных значений в некотором диапазоне, чтобы затем эти значения отсортировать. Генерация случайных значений действительных чисел выполняется вызовом функции runif() (см. рис. 5). Обязательный параметр — количество случайных чисел п. Если задать n=100, получим вектор из 100 случайных значений.

Если других параметров у runif() нет, случайные числа создаются в диапазоне от 0 до 1. Для задания диапазона, отличного от (0, 1), используется задание фактических значений у параметров min и max.

Рис. 5. Генерация случайных значений функцией runif()

Для получения случайного **целого числа** используется функция sample(), параметры которой задают диапазон (например, x=1:100), количество случайных значений (например, size = 22), запрет или разрешение появления повторных случайных значений (replace=FALSE / TRUE) (см. рис. 6).

Сортировка вектора выполняется функцией sort(). Функция sort(имя_вектора) сортирует элементы вектора по возрастанию, вызов функции sort() со значением параметра decreasing=TRUE сортирует вектор в обратную сторону (рис. 7).

Рис. 6. Генерация случайных значений функцией sample()

```
> x <- runif(-100, n= 6, max = 0)
> x
[1] -61.44684 -45.08358 -13.74772 -43.00052 -58.59973 -75.46657
> y<- sort(x)
> y
[1] -75.46657 -61.44684 -58.59973 -45.08358 -43.00052 -13.74772
> y<- sort(-x)
> y
[1] 13.74772 43.00052 45.08358 58.59973 61.44684 75.46657
> y<- sort(-x, decreasing = TRUE)
> y
[1] 75.46657 61.44684 58.59973 45.08358 43.00052 13.74772
>
```

Рис. 7. Работа функции sort()

5. Манипуляции с векторами

С векторами можно делать множество преобразований, рассмотрим некоторые из них. Как добавить элементы в вектор? Используется функция append():

w<- append(3:12, -2:0) # вектор 3:12 будет увеличен на 3 элемента Формальные параметры у append() следующие:

х – исходный вектор

values – вектор, который должен быть добавлен

after – номер индекса, после которого будет добавлен values

Если параметр after не указан, добавление выполняется после последнего индекса, х и values — обязательные параметры. Если мы работаем с вектором х и хотим добавить в него несколько элементов, вызов append() может быть таким:

 $x \leftarrow append(x=x, values=x[4:6], after=length(x)-2)$

В приведенном примере важно понимать, что означает каждое использование символа х:

- х в крайней левой позиции задает имя итогового вектора, который получится в результате добавления элементов в исходный вектор х, т.е. новому вектору дается имя старого вектора
- х в левой части равенства x=x это имя формального параметра функции append(), задающего исходный вектор, к которому выполняется добавление
- х в правой части равенства x=x это имя исходного вектора, к которому добавляются новые элементы
- x[4:6] в выражении values=x[4:6] означает, что добавляться будут три элемента исходного вектора x, расположенные в ячейках с 4-й по 6-ю

• length(x) в выражении after=length(x)-2 означает длину исходного вектора x. Добавляться три новых элемента будут после элемента исходного вектора x c номером length(x)-2

Если к вектору прибавить число, то оно прибавится ко всем элементам вектора:

```
x < -6 + c(1, 5:7) # получим (7, 11, 12, 13)
```

Аналогично, умножение, деление и все математические формулы с векторами применяются ко всем элементам вектора (см. рис. 8).

Рис. 8. Пример использования вектора в формулах

Именно в таком подходе заключается основное преимущество R – если одни и те же действия нужно выполнить над тысячами значений, эти значения помещаются в вектор, а затем пишется одна формула его обработки.

Два вектора можно объединить в третий вектор (см. рис. 9). Обратите внимание, что при объединении векторов разного типа преобразование типов выполняется автоматически. Какими функциями проверяется тип вектора?

```
> x1<- 1:5

> x2<--4:-6

> x3<-c(x2, x1)

> x3

[1] -4 -5 -6 1 2 3 4 5

> y1<-c('Когда','прилетит', 'вертолет?')

> y2<-12.40

> y3<-c(y1, y2)

> y3

[1] "Когда" "прилетит" "вертолет?" "12.4"
```

Рис. 9. Примеры объединения векторов

Как убрать лишние элементы вектора? Как получить подмножество элементов вектора? Для этого используются квадратные скобки с указанием нужного диапазона:

```
x < -x[3:5]
```

z < -c(x[3:5], x[7], x[1:4]) # некоторые элементы войдут в z два раза

Функцию конкатенации можно помещать внутрь индексации, выделяя отдельные элементы векторов:

```
z <- x[c(2,4,6)]

z <- c(x[c(2,4,6)], y[c(length(y)-3:length(y))])
```

Для удаления элемента вектора перед его индексом ставят минус:

z < -x[c(-3, -2, -5)] # будут удалены 2-й, 3-й и 5-й элементы вектора х Для фильтрации значений элементов вектора можно использовать логическое условие:

z <- x[c(x<=1, x>400)] # будут выбраны элементы, значения которых # меньше или равны 1 и больше 400

Значения векторов можно менять поэлементно:

$$x[3] < -44.6$$

```
x[2:6] < -c(17:19, sample(1:10,2))
```

В последнем примере меняются значения вектора x, расположенные в индексах с 2 по 6. Чему будут равны значения x[2], ..., x[6] ? По какому правилу выполняется заполнение значений элементов вектора?

Если в манипуляциях с индексами выйти за длину вектора, ошибки не произойдет, длина вектора увеличится, а все неопределенные элементы получат значение NA (см. рис. 10).

Рис. 10. Автоматическое увеличение длины вектора с инициализацией неопределенных элементов значениями NA

6. Сравнение векторов

Как узнать, какие элементы вектора больше или меньше какого-либо значения? Пусть имеется вектор с массами сотрудников:

```
mas < c(55, 71, 84, 90, 77, 60, 58, 94, 49, 53, 81)
```

Есть ли сотрудники с массой более 90 кг? 100 кг? Напишем условия:

$$m100 <- mas >= 100$$

$$m90 < -mas > = 90$$

Выполните код, проверьте результаты. Если хотим узнать, какие значения масс превышают или равны 90 кг, напишем другую формулу:

$$m90 <- mas[mas>=90]$$

Можно сравнивать вектора между собой (см. рис. 11). В примере на рис. 11 вектора не только имеют разный тип, но и разный размер. При сравнении векторов было выдано предупреждение о том, что длины векторов не совпадают, но сравнение все равно было выполнено. Разберем по шагам, какие действия были сделаны средой R перед сравнением:

- Было выдано предупреждение о несовпадении размеров векторов
- Длина вектора b была увеличена на 5 элементов для того, чтобы векторы а и b были одного размера
- В эти пять новых элементов вектора b были скопированы 5 первых элементов этого же вектора
- Было произведено поэлементное сравнение векторов а и b, результат записан в вектор z

```
> a<- sample(1:100, size = 11)
> a
[1] 15 49 59 21 98 95 85 57 34 76 41
> b<- runif(n=6, min=1, max=100)
> b
[1] 59.192064 10.632442 88.130070 7.324676 83.313045 61.953748
> z<- a<b/>
Предупреждение:
В a < b:
    длина большего объекта не является произведением длины меньшего объекта
> z
[1] TRUE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE TRUE
> |
```

Рис. 11. Сравнение двух векторов разного типа и разной длины

На этом примере показан общий подход, реализуемый в R при операциях с векторами, имеющими разную длину: вектор, имеющий меньшую длину, дополняется до вектора с большей длиной. Новые элементы вектора заполняются значениями этого же вектора, начиная с его значения по первому индексу и далее по порядку, в цикле, пока все поля не

заполнятся. Только после этого начинают выполняться действия с векторами.

Если под этим углом посмотреть на выражения, где выполняются действия между скаляром (единичным вектором) и другим вектором, то становится понятно, что на самом деле действия совершаются между векторами одинаковой длины:

$$10 + c(1:5) # (10 + 1, 10 + 2, 10 + 3, 10 + 4, 10 + 5)$$

$$18.8 / c(3,5,7,9) # (18.8 / 3, 18.8 / 5, 18.8 / 7, 18.8 / 9)$$

Сформулируем два общих вопроса, связанные с анализом результата проверки вектора на некоторое условие:

- 1. Удовлетворяет ли ХОТЯ БЫ ОДИН элемент вектора заданному условию? (логическое ИЛИ)
- 2. Удовлетворяют ли ВСЕ элементы вектора заданному условию? (логическое И)

Ответ на первый вопрос дает функция any(), где параметрами функции являются логические вектора. Проверим, есть ли среди сотрудников те, у кого масс меньше 50 и больше 100 кг?

res вернет TRUE, так как в векторе mas есть элементы, меньшие 50, первое условие истинно.

Не обязательно в логических параметрах должны быть условия, связанные с одним вектором, рассмотрим пример:

Для проверки истинности всех логических условий предназначена функция all(), где параметры функции такие же, как и у any(). Проверим ее работу:

res
$$<$$
- all(mas ==55, height > 190)

res = FALSE, так как не все элементы mas равны 55, не все значения из height больше 190. Даже по отдельности эти условия дадут FALSE:

```
res <- all(mas ==55)
res <- all(height > 190)
```

all() вернет TRUE, если все элементы mas и все элементы height будут удовлетворять условию, например:

```
res <- all(mas >=49, height < 200)
```

Контрольные вопросы и задания

Все задания выполнять в одном скрипте с форматом имени Группа Фамилия Лаб3.

- 1. Исследовать и сформулировать отличия в работе функций, определяющих тип вектора: typeof(), mode(), str()
- 2. Какого типа вектора w1, w2, w3, w4, w5, w6 на стр. 4? Чему равна их длина? Почему векторы w3 и w4 одного типа?
- 3. Выяснить, что будут содержать итоговые вектора w1, w2, w3 при различных параметрах функции rep(). Объяснить полученный результат.

```
w1 \leftarrow rep(c(0, -1, 1:3), times = 3)

w2 \leftarrow rep(c(0, -1, 1:3), each = 3)

w3 \leftarrow rep(c(0, -1, 1:3), each = 3, times = 2)
```

4. Создать вектор через вызов функции seq(), одновременно использовать параметры by и length.out. Объяснить правило формирования вектора в этом случае.

- 5. При вызове функции seq() в некоторых случаях используется другая функция, а именно, seq_len(). Исследовать параметры и поведение seq_len() и выяснить, в каком случае вызов seq() приведет к вызову и исполнению seq_len().
- 6. С какими параметрами нужно запустить scan(), чтобы был реализован ввод с клавиатуры элементов логического вектора? Привести пример.
- 7. Как реализовать ввод с клавиатуры вектора, состоящего из строк? Привести пример.
- 8. Сравнить редактирование с помощью fix() и с помощью edit() и выяснить, в чем заключаются различия. Написать пояснение в виде комментария.
- 9. Даны два числовых вектора разной длины:

$$a < -c(7:4, 0)$$

$$b < -c(8, 10.5, 0, -2, 9)$$

Написать программу, выполняющую следующие действия над векторами и выводящую результат:

- 1) Сложение векторов а + b
- 2) Умножение векторов а * b
- 3) Деление векторов а / b
- 4) Нахождение среднего арифметического для каждого вектора
- 5) Нахождение суммы элементов каждого вектора

Указание. В программах использовать функции print(), paste(), paste(), sum(), mean().

- 10. Пусть вектор w создан из значений разного типа. Выяснить правило, по которому выполняется преобразование данных разных типов, хранящихся в векторе. Для этого последовательно создать вектора с данными одного типа, двух типов, трех типов и т.д. Перебрать все возможные сочетания типов.
- 11. Проверить работу арифметических операторов и результат для векторов разного типа. Последовательно использовать векторы:
 - 1) одного типа, одного размера
 - 2) одного типа, разных размеров
 - 3) разных типов, одного размера
 - 4) разных типов, разного размера

Обобщить результаты расчетов в виде правил вычисления результирующего вектора для векторов произвольной длины и произвольного типа.

- 12. Создать 5 векторов разных типов и длины и написать с ними 2-3 формулы с использованием логических операций. Объяснить полученный результат, добавив в скрипт комментарии.
- 13. Написать формулу получения 10 случайных целых чисел из диапазона (min = -17, max = 28) с помощью функции runif(), не используя при вызове runif() параметров для задания минимального и максимального значений.
- 14. Написать формулу получения 20 случайных действительных чисел из диапазона (min = -7, max = -2) с помощью функции sample(), используя при вызове sample() задание диапазона от 1 до 1000.

- 15. Исследовать назначение и использование параметров функции scan(). Для получения информации использовать подсказку ?scan. Привести примеры использования scan() с различными параметрами.
- 16. Дано два вектора: w1 < -3.3:4.67 и w2 < rep(c(1:-2), 3). Вычислить вектор w3, равный вектору w1 в степени w2. Объяснить полученный результат в виде короткого комментария.
- 17. Написать программу получения вектора из N случайных символов русского алфавита. Использовать функцию sample().
- 18. Для сортировки векторов кроме функции sort() может использоваться функция order(). Выяснить отличия в работе этих функций. Привести примеры.
- 19. Вводом с клавиатуры получить два целочисленных вектора а и b. Сформировать логический вектор z, в котором z[i] должен равняться TRUE, если соответствующие значения элементов а и b кратны друг другу, FALSE в обратном случае.
- 20. Создать два вектора h1 и h2 типа double длиной 20 через задание соответствующих параметров функции vector(). При создании h1 для задания параметров использовать традиционный знак равенства, при создании h2 вместо традиционного равенства использовать знак '<-'. Проверить наличие, тип и значение переменных mode и length в первом и втором случаях. В случае необходимости для удаления переменных использовать функцию rm(). Найти и объяснить найденные отличия в создании векторов.