

Сложни структури от данни и извикване на функции

Статистическа обработка на данни с R

Пламен Петров и Тодор Балабанов

Център за обучение
Институт по информационни и комуникационни технологии
Българската академия на науките
p.petrov@iit.bas.bg todorb@iinf.bas.bg

05.V.2020

Acknowledgments

These teaching materials are funded by Velbazhd Software LLC and it is partially supported by the Bulgarian Ministry of Education and Science (contract D01–205/23.11.2018) under the National Scientific Program “Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)”, approved by DCM # 577/17.08.2018.

Съдържание

- 1 Извикване на функции
 - Обща информация за функциите
 - Търсене на функции
- 2 Линейни контейнери за данни
 - Вектори
 - Списъци
- 3 Многомерни контейнери за данни
 - Матрици
 - Масиви
 - Липсващи стойности
- 4 Структури подходящи за статистическа обработка
 - Рамкирани данни
 - Тип данни фактор
- 5 Заключение
 - Дискусия

Извикване на функции

Групи от команди

Извикване на функции

```
x <- c(1, 2, 3, 5, 6, 7, 8, 9)
[1] 1 2 3 5 6 7 8 9
mean( x )
[1] 5.125
median( x )
[1] 5.5
sd( x )
[1] 2.900123
```

Помощна информация

Документация за функциите

? mean
 ?? mean
 ? median
 ?? median
 ? sd
 ?? sd

Помощна информация

Документация за операции

? '+'

? '-'

? '*'

? '/'

Възможности за намиране на функция

Частично търсене

```
apropos( "med")  
[1] "elNamedelNamed<-medianmedian.default"  
[5] "medpolishrunmed"
```


Линейни контейнери за данни

Подредено множество от стойности

Вектор от числа и вектор от символни низове

```
v1 <- c(1, 3, 2, 1, 5)
v2 <- c("Peter "Ivan "Geroge")
```

Векторни пресмятания

Аритметични операции над вектори

```
x <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
[1] 1 2 3 4 5 6 7 8 9 10
x * 5
[1] 5 10 15 20 25 30 35 40 45 50
x + 3
[1] 4 5 6 7 8 9 10 11 12 13
x - 4
[1] -3 -2 -1 0 1 2 3 4 5 6
x / 5
[1] 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0
```

Векторни пресмятания

Степенуване на вектори

x^3

[1] 1 8 27 64 125 216 343 512 729 1000

$\text{sqrt}(x)$

[1] 1.000000 1.414214 1.732051 2.000000 2.236068 2.449490

2.645751 2.828427

[9] 3.000000 3.162278

Задаване на стойности

Алтернативен синтаксис за създаване на вектори

1:10

[1] 1 2 3 4 5 6 7 8 9 10

10:1

[1] 10 9 8 7 6 5 4 3 2 1

-2:3

[1] -2 -1 0 1 2 3

5:-7

[1] 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6 -7



```
x <- 1:10
y <- -10:-1
nchar( x )
[1] 1 1 1 1 1 1 1 1 1 2
nchar( y )
[1] 3 2 2 2 2 2 2 2 2 2
x + y
[1] -9 -7 -5 -3 -1 1 3 5 7 9
x - y
[1] 11 11 11 11 11 11 11 11 11 11
```

При вектори с еднаква дължина

Умножение и деление

$x * y$

[1] -10 -18 -24 -28 -30 -30 -28 -24 -18 -10

x / y

[1] -0.1000000 -0.2222222 -0.3750000 -0.5714286 -0.8333333

-1.2000000

[7] -1.7500000 -2.6666667 -4.5000000 -10.0000000

При вектори с еднаква дължина

Степенуване и сравнение

x^y

[1] 1.000000e+00 1.953125e-03 1.524158e-04 6.103516e-05
6.400000e-05

[6] 1.286008e-04 4.164931e-04 1.953125e-03 1.234568e-02
1.000000e-01

$x > y$

[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE

Контейнери с обекти от различен тип

Създаване на списък

```
l1 <- list(1, 2, 3, 4, 5) [[1]]
```

```
[1] 1
```

```
[[2]]
```

```
[1] 2
```

```
[[3]]
```

```
[1] 3
```

```
[[4]]
```

```
[1] 4
```

```
[[5]]
```

```
[1] 5
```

Трансформация на контейнери

Вектор към списък

```
l2 <- list( c(1, 2, 3, 4, 5) )
[[1]]
[1] 1 2 3 4 5
```

Контейнер с разнородни типове данни

Названия на елементите в списъка

```
df2 <- data.frame(First=x, Second=y, Sport=q)
rownames( df2 ) <- c("One "Two "Three "Four "Five")
l3 <- list( df2, 1:5, l1 )
names( l3 ) <- c("Frame "Vector "Element")
```

Достъп до елементите

Достъп до елементите на списъка

l3[1]

l3[" Frame"]

\$Frame

First Second Sport

One 4 0 Football

Two 1 -2 Basketball

Three 3 1 Volleyball

Four 2 -1 Handball

Five 5 2 Rugby

Достъп до елементите

Вложено позоваване

```
l3[[1]]
```

```
First Second Sport
```

```
One 4 0 Football
```

```
Two 1 -2 Basketball
```

```
Three 3 1 Volleyball
```

```
Four 2 -1 Handball
```

```
Five 5 2 Rugby
```

```
l3[["Frame"]]$Sport
```

```
[1] Football Basketball Volleyball Handball Rugby
```

```
Levels: Basketball Football Handball Rugby Volleyball
```



Запис на данни

Добавяне на елемент

```
l3[ 4 ] <- "Games"
```

```
[[4]]
```

```
[1] "Games"
```

Многомерни контейнери за данни

Многомерни данни

Създаване на матрици

```
m1 <- matrix(1:6, nrow=3)
m2 <- matrix(7:12, nrow=3)
m3 <- matrix(7:18, nrow=2)
```


Пресмятания при повече измерения

Операции с матрици

```
nrow( m1 )
ncol( m1 )
dim( m1 )
m1 + m2
m1 * m2
m1 == m2
```

Пресмятания при повече измерения

Матрично умножение

m1 %*% t(m2)

[,1] [,2] [,3]

[1,] 47 52 57

[2,] 64 71 78

[3,] 81 90 99

Преименуване

Имена на колоните и редовете

```
colnames( m1 ) <- c("First", "Second")
rownames( m1 ) <- c("One", "Two", "Three")
colnames( m2 ) <- c("Left", "Right")
rownames( m2 ) <- c("1st", "2nd", "3rd")
```

Многомерни данни

Работа с масиви

```
a1 <- array(1:12, dim = c(2, 3, 2))
a1[1, , 1]
[1] 1 3 5
a1[, , 1]
[,1] [,2] [,3]
[1,] 1 3 5
[2,] 2 4 6
```

Проблеми при отчитане на данните

Липсващи стойности

```
x <- c(1, NA, 3, NA, 5)
[1] 1 NA 3 NA 5
is.na( x )
[1] FALSE TRUE FALSE TRUE FALSE
mean( x )
[1] NA
median( x )
[1] NA
sd( x )
[1] NA
```

Проблеми при отчитане на данните

Аварийна стойност

```
y <- c(1, NULL, 3, NULL, 5)
[1] 1 3 5
is.na( y )
[1] FALSE FALSE FALSE
mean( y )
[1] 3
median( y )
[1] 3
sd( y )
[1] 2
is.null( y )
[1] FALSE
```

Структури подходящи за статистическа обработка

Атрибути на рамкираните данни

Имена на колоните и размери

```
df2 <- data.frame(First=x, Second=y, Sport=q)
rownames( df2 ) <- c("One", "Two", "Three", "Four", "Five")
nrow( df1 )
ncol( df1 )
dim( df1 )
names( df2 )
rownames( df2 )
```

Инспекция на рамкираните данни

Тип, начало и край

```
class( df1 )
head(df1, n=3)
tail(df1, n=3)
```

Символни низове в рамковите данни

Фактори в рамковите данни

```
class( df2$Sport )
[1] "factor"
df2$Sport
[1] Football Basketball Volleyball Handball Rugby
Levels: Basketball Football Handball Rugby Volleyball
```


Заклучение

Въпроси и отговори

Благодаря за вниманието!