Data Types and Data Structures in R

1

Data types in R

Data types are linked with measurement scale, such as nominal, ordinal, interval and ratio

- Character
- Complex (not commonly used in public health research)
- Numeric (Integer, Double)
- Logical

2

Data types in R (Example)

Creating various types of data objects and check that

```
cVec <- c("Cricket", "Football", "Basketball", "Rugby")
cVec <- c(game1="Cricket", game2="Football", game3="Basketball",
          game4="Rugby")
nVec <- c(1:10)
Lvec <- c(TRUE, FALSE, FALSE, TRUE)
nVec2 <- runif( n=5, min=10, max=20)
cVec2 <- sample(x=letters, size= 5, replace = F)
Lvec2 <- nVec2>=13
mixVec <- c(1,3, "Cricket", "Football", "Basketball", "Rugby")
```

3

Data Structure in R

The way of organizing data in R

- Vector (contains only one type of data)
- Matrix (contains one type of data arranged in rows and columns)
- Array (contains one type of data arranged in rows columns and beyond)
- Data frame (Arranged like matrix but contains more than one type of data)
- List (contains everything in it)

4

Creating a Single Variable: Vector

The default function `c()`, is the most convenient way to create a single variable

It takes the element of the vector as input separated by a comma as : `c(item1, item2, item3)`

To assign a variable name (also called object in R), an assignment operator "<–"

Example: `a <- c(3,5)`, this tells R that the value 3 and 5 is assigned to an object called "a"

5

5

Matrix in R

- A matrix in R is a two-dimensional representation of a dataset
- Each column **must** be same length
- Each column **must contain same types of data** either all numeric, or all character, or all logical or all complex number
- You cannot store different types of data in different columns

```
x <- c(13, 21, 19, 18, 21, 16, 21, 24, 17, 18,
       12, 18, 29, 17, 18, 11, 13, 20, 25, 18,
       15, 19, 21, 21, 7, 12, 23, 31, 16, 19,
       23, 15, 25, 19, 15, 25, 25, 16, 29, 15,
       26, 29, 23, 24, 20, 19, 14, 27, 22, 26)

xmat <- matrix(data = x, nrow = 10, ncol = 5,
               byrow = TRUE)
```

6

6

Matrix in R

```
> xmat <- matrix(data = x, nrow = 10, ncol = 5, byrow = TRUE)
> xmat
      [,1] [,2] [,3] [,4] [,5]
 [1,]   13   21   19   18   21
 [2,]   16    9   24   17   18
 [3,]   12   18   29   17   18
 [4,]   11   13   20   25   18
 [5,]   15   19   21   21    7
 [6,]   12   23   31   16   19
 [7,]   23   15   25   19   15
 [8,]   25   25   16   29   15
 [9,]   26   29   23   24   20
[10,]   19   14   27   22   26
> |
```

- We had a vector "x" and now it is converted into a matrix,
- The question is:
  - ✓ How the elements will be organized?
  - ✓ Where to place which value?

7

7

Data frame in R

A data frame is also a two-dimensional representation of data

Each column **must** be same length

The columns of a data frame could be a mix of numeric, character, logical and complex number

8

8

```
> arrayA <- array(1:16, dim=c(2,2,4))
> arrayA
, , 1

     [,1] [,2]
[1,]    1    3
[2,]    2    4

, , 2

     [,1] [,2]
[1,]    5    7
[2,]    6    8

, , 3

     [,1] [,2]
[1,]    9   11
[2,]   10   12
```

Array in R

- The way to organize data with more than two dimension
- Each component **must** same types of data
- The number of dimension could be any

9

9

List in R

- List is the natural generalization of a data frame
- It can contain heterogeneous data, e.g., mix of numeric and character with different length
- It can contain a vector, matrix, data frame and a list itself

```
cVec <- c("Cricket", "Football", "Basketball", "Rugby")
nVec <- c(1:10)
Lvec <- c(TRUE, FALSE, FALSE, TRUE)
matA <- matrix(1, nrow=2, ncol=2)
datA <- data.frame(ID = 1:5, hourSpentOnInternet = c(5,3,4,1,2),
GENDER = c("M", "F", "F", "M", "F"))
arrayA <- array(1:16, dim=c(2,2,4))


listB <- list(
            vector1 = cVec,
            vector2 = nVec,
            vector3 = Lvec,
            matrix1 = matA,
            data1 = datA,
            array1 = arrayA
)
```

10

Data structure in R (Intuitive analogy)

Vector
Milk chocolate
with different shapes

Matrix
All milk chocolates with different ***colour*** and ***shapes*** and possibly from different brand

Data Frame
Along with ***milk chocolate*** it could contain ***dark chocolates***, ***nuts*** and other things from variety of brands

List
It contains everything

11