**Bài tập R**

## Chương I: Cài đặt R

1. Download ngôn ngữ R:

* B1: Truy cập vào trang web <https://www.r-project.org>
* B2: Nhấn vào thanh CRAN bên dưới Download. Sau đó lựa chọn server phù hợp với ngôn ngữ bạn đang sử dụng
* B3: Chọn download R cho các hệ điều hành thích hợp hoặc Linux hoặc Mac OS hoặc Windows -> Chọn base -> chọn download R 4.0.4 for Windows
* B4: Mở file đã download về và cài đặt

1. Download RStudio

* B1: Truy cập trang web <https://rstudio.com/products/rstudio/download>
* B2: Chọn Download bản free -> Lựa chọn phiên bản phù hợp với hệ điều hành đang sử dụng
* B3: Mở file vừa tải và tiến hành cài đặt RStudio

## Chương II: Kiểu dữ liệu trong R

1. **Sum()**

- Ý nghĩa: trả về tổng tất cả các đối số trong hàm

- sum(..., na.rm = FALSE)

- Giải thích các đối số trong hàm:

* … : vector số nguyên, số phức hoặc logic
* na.rm : loại bỏ các giá trị NA

Vd:

sum(1, 2, 3, 4, 5)

[1] 15

1. **Median()**

- Ý nghĩa: trả về trung vị của mẫu (nếu mẫu chưa được sắp xếp thì sẽ tự động sắp xếp theo thứ tự từ bé -> lớn và trả về trung vị của mẫu)

- median(x, na.rm = FALSE, ...)

- Giải thích các đối số trong hàm:

* x: đối số đưa vào có thể là một vector số
* na.rm: loại bỏ giá trị NA

Vd:

a <- c(1, 3, 5, 234, 3, 5, -5, -2, -4)

median(a)

1. 3
2. **Max()**

- Ý nghĩa: Trả về giá trị lớn nhất của mẫu

- max(..., na.rm = FALSE)

- Giải thích các đối số trong hàm:

* … : dữ liệu số hoặc kí tự
* na.rm: loại bỏ giá trị na

Vd:

a <- c(1, 2, 3, 4, 5, 6)

max(a)

1. 6
2. **Min()**

- Ý nghĩa: Trả về giá trị nhỏ nhất của mẫu

- min(..., na.rm = FALSE)

- Giải thích các đối số trong hàm:

* … : dữ liệu số hoặc kí tự
* na.rm: loại bỏ giá trị na

Vd:

a <- c(1, 2, 3, 4, 5, 6)

min(a)

[1] 1

1. **C()**

- Ý nghĩa: Kết hợp tất cả các đối số trong dấu ngoặc tạo thành một vector

- c(...)

- Giải thích các đối số trong hàm:

* … : các đối tượng được nối thành 1 vector

Vd:

c(1, 2, 3, 4, 5, 6)

1. 1 2 3 4 5 6
2. **Cbind()**

- Ý nghĩa: Lấy mỗi chuỗi các đối số vector, ma trận hoặc khung dữ liệu và kết hợp các cột hoặc hàng tương ứng

- cbind(..., deparse.level = 1)

- Giải thích các đối số trong hàm:

* … : vector hoặc ma trận cần được kết hợp
* deparse.level: số nguyên kiểm soát việc xây dựng các nhãn trong trường hợp các đối số không giống ma trận (đối với phương thức mặc định):deparse.level = 0 cấu trúc không có nhãn (mặc định),deparse.level = 1 hoặc 2 tạo nhãn từ tên đối số.

Vd:

a <- c(1, 2, 3, 4, 5)

b <- c("A", "B", "C", "D", "E")

c <- c("ABC", "ac", 1, 222, 333)

d <- cbind(a, b, c)

a b c

[1,] "1" "A" "ABC"

[2,] "2" "B" "ac"

[3,] "3" "C" "1"

[4,] "4" "D" "222"

[5,] "5" "E" "333"

1. **Rbind()**

- Ý nghĩa: Kết hợp các đối số lại với nhau, có thể là 1 df với 1 vector, 2 df với nhau, …

- rbind(..., deparse.level = 1)

- Giải thích các đối số trong hàm:

* … : vector hoặc ma trận cần được kết hợp
* deparse.level: số nguyên kiểm soát việc xây dựng các nhãn trong trường hợp các đối số không giống ma trận (đối với phương thức mặc định):deparse.level = 0 cấu trúc không có nhãn (mặc định),deparse.level = 1 hoặc 2 tạo nhãn từ tên đối số.

Vd:

a1 <- c(1, 2, 3)

a2 <- c(2, 3, 4)

a3 <- c(3, 4, 5)

a4 <- c(3, 4, 5, 6)

df <- data.frame(a1, a2, a3)

rbind(df, a4)

a1 a2 a3

1 1 2 3

2 2 3 4

3 3 4 5

4 3 4 5

1. **Vector()**

- Ý nghĩa: Hàm tạo vector có độ dài với mode đã cho

- vector(mode = "logical", length = 0)

- Giải thích các đối số trong hàm:

* mode: chuỗi ký tự đặt tên một mode hoặc "list" hoặc "expression" hoặc "any"
* length: độ dài vector

Vd:

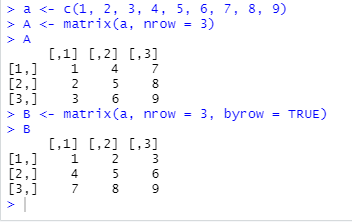
1. **Matrix()**

- Ý nghĩa: Hàm tạo ma trận với kích thước nhập sẵn

- matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)

- Giải thích các đối số trong hàm:

* Data: một vectơ dữ liệu tùy chọn (bao gồm một danh sách hoặc vectơ biểu thức)
* Nrow: số hàng của ma trận
* Ncol: số cột
* Byrow: một cách để tạo một ma trận chuyển vị



1. **Data.frame()**

- Ý nghĩa: Hàm tạo dataframe

- data.frame(..., row.names = NULL, check.rows = FALSE,

check.names = TRUE, fix.empty.names = TRUE,

stringsAsFactors = default.stringsAsFactors())

- Giải thích các đối số trong hàm:

Vd:

emp.data <- data.frame(

emp\_id = c (1:5),

emp\_name = c("Rick","Dan","Michelle","Ryan","Gary"),

salary = c(623.3,515.2,611.0,729.0,843.25),

start\_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15", "2014-05-11",

"2015-03-27")),

stringsAsFactors = FALSE)

print(emp.data)

emp\_id emp\_name salary start\_date

1 1 Rick 623.30 2012-01-01

2 2 Dan 515.20 2013-09-23

3 3 Michelle 611.00 2014-11-15

4 4 Ryan 729.00 2014-05-11

5 5 Gary 843.25 2015-03-27

1. **List()**

- Ý nghĩa: Hàm tạo, ép, kiểm tra một list trong R

- list(...)

Vd:

list\_data <- list("Red", "Green", c(21,32,11), TRUE, 51.23, 119.1)

print(list\_data)

[[1]]

[1] "Red"

[[2]]

[1] "Green"

[[3]]

[1] 21 32 11

[[4]]

[1] TRUE

[[5]]

[1] 51.23

[[6]]

1. 119.1
2. **Rep()**

- Ý nghĩa: Hàm lặp lại với số lần được chỉ định

- rep(x, ...)

- Giải thích các đối số trong hàm:

* X: một vector hoặc factor cần lặp lại
* Times: số lần lặp

Vd:

rep(2, 3)

[1] 2 2 2

1. **Seq()**

- Ý nghĩa: Hàm tạo chuỗi

- seq.int(from, to, by, length.out, along.with, ...)

- Giải thích các đối số trong hàm:

* From: giá trị khởi đầu
* To: giá trị cuối cùng
* By: khoảng cách giữa các giá trị
* Length.out: quyết định độ dài của chuỗi

Vd:

x <- seq(-6,7)

x

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

1. **Dim()**

- Ý nghĩa: Hàm kiểm tra số chiều

- dim(x)

- Giải thích các đối số trong hàm:

* x: Đối số truyền vào cần kiểm tra số chiều

Vd:

dim(Benthic)

1. 45 9
2. **Colnames()**

- Ý nghĩa: Hàm kiểm tra tên cột của đối số được đưa vào

- colnames(x, do.NULL = TRUE, prefix = "col")

- Giải thích các đối số trong hàm:

* x: đối số truyền vào là một ma trận 2 chiều cần kiểm tra

Vd:

colnames(Benthic)

1. "Sample" "Richness" "Exposure" "NAP" "Beach" "grainsize" "humus" "chalk" "sorting1"
2. **Rownames()**

- Ý nghĩa: Hàm kiểm tra tên hàng của đối số được đưa vào

- rownames(x, do.NULL = TRUE, prefix = "row")

- Giải thích các đối số trong hàm:

* x: đối số truyền vào là một ma trận 2 chiều cần kiểm tra

Vd:

rownames(Benthic)

[1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12" "13" "14" "15" "16" "17" "18" "19" "20" "21" "22"

[23] "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33" "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44"

[45] "45"

1. **Setwd()**

- Ý nghĩa: Thiết lập thư mục làm việc hiện hành

- setwd(dir)

- Giải thích các đối số trong hàm:

* Dir: đường dẫn đến file cần truyền

Vd:

setwd("C://Users//PC//Desktop//R//BT R tổng hợp")

1. **Read.table()**

- Ý nghĩa: Đọc dữ liệu của tệp ở dạng txt

- read.table(file, header = FALSE, sep = "", quote = "\"'",

dec = ".", numerals = c("allow.loss", "warn.loss", "no.loss"),

row.names, col.names, as.is = !stringsAsFactors,

na.strings = "NA", colClasses = NA, nrows = -1,

skip = 0, check.names = TRUE, fill = !blank.lines.skip,

strip.white = FALSE, blank.lines.skip = TRUE,

comment.char = "#",

allowEscapes = FALSE, flush = FALSE,

stringsAsFactors = default.stringsAsFactors(),

fileEncoding = "", encoding = "unknown", text, skipNul = FALSE)

- Giải thích các đối số trong hàm:

* File: tên file cần đọc dữ liệu
* Sep: Kí tự phân cách trường dữ liệu
* Header: một giá trị logic cho biết liệu tệp có chứa tên của các biến như dòng đầu tiên của nó hay không. Nếu thiếu, giá trị được xác định từ định dạng tệp: tiêu đề được đặt thành TRUE nếu và chỉ khi hàng đầu tiên chứa một trường ít hơn số cột.

1. **Scan()**

- Ý nghĩa: Đọc dữ liệu thành vectơ hoặc danh sách từ bảng điều khiển hoặc tệp

## Chương III: Truy cập biến, tạo subset data

1. **Write.table**

- Ý nghĩa: in đối số x bắt buộc của nó (sau khi chuyển đổi nó thành khung dữ liệu nếu nó không phải là một hoặc không phải là ma trận) vào một tệp

- write.table(x, file = "", append = FALSE, quote = TRUE, sep = " ",

eol = "\n", na = "NA", dec = ".", row.names = TRUE,

col.names = TRUE, qmethod = c("escape", "double"),

fileEncoding = "")

- Giải thích các đối số trong hàm:

* X: đối tượng cần được viết vào trong file
* File: chuỗi kí tự tên của loại file cần được viết vào

1. **Order**

- Ý nghĩa: Đánh chỉ số cho đối tượng và sắp xếp theo thứ tự tăng dần hoặc giảm dần theo chỉ số của nó

- order(..., na.last = TRUE, decreasing = FALSE,

method = c("auto", "shell", "radix"))

- Giải thích các đối số trong hàm:

* … : một vector số nguyên, số phức, kí tự hoặc biểu thức logic,
* Decreasing: sắp xếp giảm dần nếu TRUE và ngược lại

Vd:

y = c(4,12,6,7,2,9,5)

order(y)

[1] 5 1 7 3 4 6 2

y = c(4,12,6,7,2,9,5)

y[order(y)]

[1] 2 4 5 6 7 9 12

1. **Merge**

- Ý nghĩa: Hợp nhất 2 dataframe theo các cột hoặc tên hàng chung

- merge(x, y, ...)

- Giải thích các đối số trong hàm:

* x, y: dataframe cần nối

Vd:

authors <- data.frame(

surname = c("Tukey", "Venables", "Tierney", "Ripley", "McNeil"),

nationality = c("US", "Australia", "US", "UK", "Australia"),

retired = c("yes", rep("no", 4)))

books <- data.frame(

name = c("Tukey", "Venables", "Tierney", "Ripley", "Ripley", "McNeil"),

title = c("Exploratory Data Analysis",

"Modern Applied Statistics ...",

"LISP-STAT",

"Spatial Statistics", "Stochastic Simulation",

"Interactive Data Analysis"),

other.author = c(NA, "Ripley", NA, NA, NA, NA))

merge(authors, books, by.x="surname", by.y="name")

merge(books, authors, by.x="name", by.y="surname")

surname nationality retired title other.author

1 McNeil Australia no Interactive Data Analysis <NA>

2 Ripley UK no Spatial Statistics <NA>

3 Ripley UK no Stochastic Simulation <NA>

4 Tierney US no LISP-STAT <NA>

5 Tukey US yes Exploratory Data Analysis <NA>

6 Venables Australia no Modern Applied Statistics ... Ripley

1. **Attach**

- Ý nghĩa: Hàm sử dụng để truy cập các biến trong dataframe mà không cần gọi datafram

- attach(what, pos = 2L, name = deparse1(substitute(what), backtick=FALSE),

warn.conflicts = TRUE)

- Giải thích các đối số trong hàm:

* What: là một dataframe hoặc 1 list
* Pos: số nguyên xác định vị trí nơi cần đính kèm.

Vd:

data <- data.frame(x1 = c(9, 8, 3, 4, 8),

x2 = c(5, 4, 7, 1, 1),

x3 = c(1, 2, 3, 4, 5))

data

attach(data)

x1

x2

1. 9 8 3 4 8
2. 5 4 7 1 1
3. **Str**

- Ý nghĩa: Hiển thị cấu trúc của từng cột trong dataframe truyền vào

Vd:

Benthic <- read.csv("RIKZ2.txt", sep = "\t", header = TRUE)

str(Benthic)

'data.frame': 45 obs. of 9 variables:

$ Sample : int 1 2 3 4 5 6 7 8 9 10 ...

$ Richness : int 11 10 13 11 10 8 9 8 19 17 ...

$ Exposure : int 10 10 10 10 10 8 8 8 8 8 ...

$ NAP : num 0.045 -1.036 -1.336 0.616 -0.684 ...

$ Beach : int 1 1 1 1 1 2 2 2 2 2 ...

$ grainsize: num 222 200 194 221 202 ...

$ humus : num 0.05 0.3 0.1 0.15 0.05 0.1 0.1 0.1 0.15 0 ...

$ chalk : num 2.05 2.5 3.45 1.6 2.45 2.5 1.85 1.7 2.3 2.6 ...

$ sorting1 : num 69.8 59 59.2 67.8 57.8 …

1. **Factor**

- Ý nghĩa: Dùng để mã hóa vector

Vd:

factor(Benthic$Richness)

[1] 11 10 13 11 10 8 9 8 19 17 6 1 4 3 3 1 3 3 1 4 3 22 6 0 6 5 4 1 6 4 2 1 1 3 4 3 5

[38] 7 5 0 7 11 3 0 2

## Chương IV: Hàm đơn giản

|  |  |
| --- | --- |
| **Tên hàm** | **Ý nghĩa** |
| tapply | Áp dụng hàm cho mỗi ô của mảng |
| sapply | Nhận df làm đối tượng và trả về hàm yêu cầu |
| sd | Hàm tính độ lệch chuẩn của đối số x đưa vào |
| length | Lấy hoặc đặt độ dài của vector |
| summary | Trả về bảng tóm tắt thông kê đơn giản |
| table | Trả về một bảng các giá trị yêu cầu |

## Chương V: Plot

|  |  |
| --- | --- |
| **Tên hàm** | **Ý nghĩa** |
| plot | Hàm vẽ biều đồ |
| lines | Hàm vẽ biểu đồ, các tọa độ được nối với nhau bằng đường |
| order | Đánh chỉ số cho đối tượng và sắp xếp theo thứ tự tăng dần hoặc giảm dần theo chỉ số của nó |
| loess | Điều chỉnh bề mặt đa thức được xác định bởi một hoặc nhiều công cụ dự đoán số, sử dụng tính năng phù hợp |
| fitted |  |

## Chương VI: Học loop function

|  |  |
| --- | --- |
| **Tên hàm** | **Ý nghĩa** |
| jpeg | Thiết bị đồ họa cho các tệp bitmap định dạng BMP, JPEG, PNG và TIFF. |
| Dev.off | Hàm cung cấp quyền kiểm soát trên nhiều thiết bị đồ họa |
| function | Tạo hàm |
| paste | Nối các vector sau khi chuyển đổi thành kí tự |
| if | Câu lệnh điều kiện |
| ifelse | Câu lệnh điều kiện |
| If elseif | Câu lệnh điều kiện |

## Chương VII: Type Plot

|  |  |
| --- | --- |
| **Tên hàm** | **Ý nghĩa** |
| pie | Vẽ biểu đồ hình tròn |
| Pie3D | Vẽ biểu đồ hình tròn 3D |
| par | Thiết lập hoặc truy vấn biểu đồ |
| barplot | Vẽ biểu đồ thanh |
| arrows | Vẽ mũi tên giữa các điểm |
| box | Tạo trục số cho biểu đồ |
| boxplot | Vẽ boxplot |
| text |  |
| points | là một hàm tổng quát để vẽ một chuỗi các điểm tại các tọa độ xác định. (Các) ký tự được chỉ định được vẽ, căn giữa tại các tọa độ. |
| legend | Thêm chú thích cho biểu đồ |
| title | Thêm tiêu đề cho biểu đồ |
| expression | Tạo hoặc kiểm tra các đối tượng của mode "biểu thức". |
| pairs | Hàm vẽ nhiều biểu đồ |
| coplot |  |
| layout | chia thiết bị thành nhiều hàng và cột như có trong ma trận, với chiều rộng cột và chiều cao hàng được chỉ định trong các đối số tương ứng. |