

# Kiến thức cơ bản về R

## Cheat Sheet

### Tìm kiếm thông tin

Truy cập vào file chứa thông tin

?mean

Tìm thông tin về một hàm.

help.search('weighted mean')

Tìm thông tin về một từ hoặc cụm từ

help(package = 'dplyr')

Tìm thông tin về một package

Đối tượng trong R

str(iris)

Tóm tắt về cấu trúc của đối tượng.

class(iris)

Tìm kiếm class của đối tượng.

### Sử dụng thư viện

install.packages('dplyr')

Tải package từ CRAN.

library(dplyr)

Gọi package để sử dụng

dplyr::select

Sử dụng một hàm cụ thể thuộc một package.

data(iris)

Gọi một tập dữ liệu trong R để sử dụng

### Thư viện làm việc

getwd()

Tìm thư viện làm việc hiện tại

setwd('C://file/path')

Thay đổi thư mục làm việc.

**Lưu ý: Nên sử dụng project trong R để đặt thư viện làm việc**

### Véc-tơ

#### Tạo mới véc-tơ

c(2, 4, 6)	2 4 6	Nhóm các giá trị vào véc-tơ
2:6	2 3 4 5 6	Chuỗi số nguyên
seq(2, 3, by=0.5)	2.0 2.5 3.0	Chuỗi có giá trị từ 2-3, khoảng cách giữa các giá trị là 0.5
rep(1:2, times=3)	1 2 1 2 1 2	Lặp lại các giá trị trong véc-tơ
rep(1:2, each=3)	1 1 1 2 2 2	Lặp lại các giá trị của véc-tơ

#### Hàm với véc-tơ

sort(x)

Sắp xếp thứ tự của x.

rev(x)

Đảo ngược thứ tự của x.

table(x)

Bảng tần xuất giá trị của x.

unique(x)

Các giá trị duy nhất của x

#### Lựa chọn các giá trị của véc-tơ

##### Lựa chọn theo vị trí

x[4] Lựa chọn giá trị thứ 4 trong x.

x[-4] Lựa chọn tất cả trừ vị trí thứ 4.

x[2:4] Lựa chọn các giá trị từ vị trí thứ 2 đến 4.

x[-(2:4)] Lựa chọn tất cả các giá trị trừ vị trí 2 và 4

x[c(1, 5)] Elements one and five.

##### Lựa chọn theo giá trị

x[x == 10] Các giá trị trong x bằng 10

x[x < 0] Các giá trị nhỏ hơn 0

x[x %in% c(1, 2, 5)] Các giá trị thuộc tập {1, 2, 5}

##### Theo tên

x['apple'] Giá trị với tên 'apple'.

### Lập trình

#### Vòng lặp For

```
for (chỉ số chạy){
  Câu lệnh
}
```

##### Ví dụ

```
for (i in 1:4){
  j <- i + 10
  print(j)
}
```

#### Điều kiện If

```
if (điều kiện){
  Câu lệnh 1
} else {
  Câu lệnh 2
}
```

##### Ví dụ

```
if (i > 3){
  print('Yes')
} else {
  print('No')
}
```

#### Vòng lặp While

```
while (điều kiện lặp){
  Câu lệnh
}
```

##### Ví dụ

```
while (i < 5){
  print(i)
  i <- i + 1
}
```

#### Hàm

```
function_name <- function(var){
  Câu lệnh
  return(new_variable)
}
```

##### Ví dụ

```
square <- function(x){
  squared <- x*x
  return(squared)
}
```

### Đọc và ghi dữ liệu

Đầu vào	Đầu ra	Chú giải
df <- read.table('file.txt')	write.table(df, 'file.txt')	Đọc và ghi dữ liệu dạng file txt
df <- read.csv('file.csv')	write.csv(df, 'file.csv')	Đọc và ghi dữ liệu dạng CSV. Đây là trường hợp đặc biệt của hàm read.table/write.table.
load('file.RData')	save(df, file = 'file.Rdata')	Đọc và ghi dữ liệu đặc trưng của R - .Rdata

#### Điều kiện

a == b	Bằng	a > b	Lớn hơn	a >= b	Lớn hơn hoặc bằng	is.na(a)	Kiểm tra dữ liệu bị thiếu
a != b	Không bằng	a < b	Nhỏ hơn	a <= b	Nhỏ hơn hoặc bằng	is.null(a)	Kiểm tra đối tượng có dạng null

## Kiểu dữ liệu

Biến đổi các kiểu dữ liệu thông thường trong R

<b>as.logical</b>	TRUE, FALSE, TRUE	Kiểu logic
<b>as.numeric</b>	1, 0, 1	Kiểu số
<b>as.character</b>	'1', '0', '1'	Kiểu ký tự
<b>as.factor</b>	'1', '0', '1', levels: '1', '0'	Kiểu ký tự có đã chuyển đổi sang "factor", được dùng trong nhiều mô hình phân tích

## Hàm toán học

<b>log(x)</b>	Log	<b>sum(x)</b>	Tổng
<b>exp(x)</b>	Hàm exp	<b>mean(x)</b>	Giá trị trung bình.
<b>max(x)</b>	Giá trị lớn nhất	<b>median(x)</b>	Trung vị
<b>min(x)</b>	Giá trị nhỏ nhất	<b>quantile(x)</b>	Phân vị x
<b>round(x, n)</b>	Làm tròn đến n chữ số.	<b>rank(x)</b>	Rank các giá trị của x
<b>signif(x, n)</b>	Làm tròn đến n chữ số dạng e	<b>var(x)</b>	Phương sai
<b>cor(x, y)</b>	Hệ số tương quan	<b>sd(x)</b>	Độ lệch tiêu chuẩn

## Gán biến

```
> a <- 'apple'
> a
[1] 'apple'
```

## Môi trường làm việc

<b>ls()</b>	Danh sách các đối tượng trong môi trường làm việc
<b>rm(x)</b>	Xóa đối tượng x
<b>rm(list = ls())</b>	Xóa tất cả đối tượng

Lưu ý: Ta có thể sử dụng tab Environment trong Rstudio để xem danh sách các đối tượng đang có trong môi trường làm việc của R.

## Ma trận

`m <- matrix(x, nrow = 3, ncol = 3)`  
Tạo một ma trận từ x

	<code>m[2, ]</code>	Chọn một dòng	<b>t(m)</b> Ma trận chuyển vị
	<code>m[, 1]</code>	Chọn một cột	<code>m %*% n</code> Phép nhân ma trận
	<code>m[2, 3]</code>	Chọn dòng & cột	<b>solve(m, n)</b> Giải phương trình: $m * x = n$

## Lists

`l <- list(x = 1:5, y = c('a', 'b'))`

List là một tập hợp các đối tượng có thể có các kiểu dữ liệu khác nhau.

<code>l[[2]]</code>	<code>l[1]</code>	<code>l\$x</code>	<code>l[y]</code>
Giá trị thứ 2 của l.	List mới với giá trị đầu tiên	Đối tượng trong list tên là x	Giá trị y trong list

Xem thêm package dplyr

## Data Frames

`df <- data.frame(x = 1:3, y = c('a', 'b', 'c'))`

Trường hợp đặc biệt của x với tất cả các giá trị (véc-tơ) trong list có độ dài bằng nhau.

x	y
1	a
2	b
3	c

Lấy tập con trong data frame

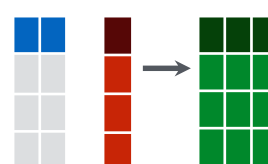
<code>df[, 2]</code>	
<code>df[2, ]</code>	
<code>df[2, 2]</code>	

**nrow(df)**  
Số lượng dòng

**ncol(df)**  
Số lượng cột

**dim(df)**  
Số lượng dòng và cột

**cbind** – Ghép cột.



**rbind** – Ghép dòng.



Hiểu thêm về data frame

**View(df)** Xem data frame

**head(df)** Xem 6 dòng đầu tiên

## Strings

Xem thêm package stringr

<code>paste(x, y, sep = '')</code>	Nối nhiều véc-tơ với nhau.
<code>paste(x, collapse = '')</code>	Nối các giá trị của véc-tơ với nhau
<code>grep(pattern, x)</code>	Tìm chuỗi ký tự trong x
<code>gsub(pattern, replace, x)</code>	Thay thế chuỗi ký tự trong x
<code>toupper(x)</code>	Chuyển sang chữ hoa
<code>tolower(x)</code>	Chuyển sang chữ thường
<code>nchar(x)</code>	Số lượng ký tự trong chuỗi

## Factors

<b>factor(x)</b> Biến đổi véc-tơ sang dạng factor	<b>cut(x, breaks = 4)</b> Biến đổi véc-tơ sang dạng factor nhưng có chia thành các lớp
--	---

## Phân tích thống kê

<b>lm(x ~ y, data=df)</b> Mô hình hồi quy	<b>t.test(x, y)</b> Kiểm định so sánh giá trị trung bình	<b>prop.test</b> Kiểm định tỷ lệ trong tập dữ liệu.
<b>glm(x ~ y, data=df)</b> Mô hình hồi quy tổng quát	<b>pairwise.t.test</b> Kiểm định giá trị trung bình theo nhóm	<b>aov</b> Phân tích phương sai.
<b>summary</b> Xem thông tin kết quả mô hình		

## Hàm phân phối

Phân phối	Biến ngẫu nhiên	Hàm mật độ	Hàm phân phối	Quantile
Normal	<b>rnorm</b>	<b>dnorm</b>	<b>pnorm</b>	<b>qnorm</b>
Poisson	<b>rpois</b>	<b>dpois</b>	<b>ppois</b>	<b>qpois</b>
Binomial	<b>rbinom</b>	<b>dbinom</b>	<b>pbinom</b>	<b>qbinom</b>
Uniform	<b>runif</b>	<b>dunif</b>	<b>punif</b>	<b>qunif</b>

## Biểu đồ

Xem thêm package ggplot2

<b>plot(x)</b> Biểu đồ giá trị của x	<b>plot(x, y)</b> Biểu đồ giá trị của x và y.	<b>hist(x)</b> Histogram của x.
---	--	------------------------------------

## Dữ liệu thời gian

Xem thêm package lubridate.