

17

Phục lục 2

Một số lệnh thông dụng trong R

Lệnh về môi trường vận hành của R

<code>getwd()</code>	Cho biết directory hiện hành là gì
<code>setwd(c:/works)</code>	Chuyển directory vận hành về c:\works (chú ý R dùng “/”)
<code>options(prompt="R>")</code>	Đổi prompt thành R>
<code>options(width=100)</code>	Đổi chiều rộng cửa sổ R thành 100 characters
<code>options(scipen=3)</code>	Đổi số thành 3 số thập phân (thay vì kiểu 1.2E-04)
<code>options()</code>	Cho biết các thông số về môi trường hiện nay của R

Lệnh cơ bản

<code>ls()</code>	Liệt kê các đối tượng (objects) trong bộ nhớ
<code>rm(object)</code>	Xóa bỏ đối tượng
<code>search()</code>	Tìm hướng

Kí hiệu tính toán

<code>+</code>	Cộng
<code>-</code>	Trừ
<code>*</code>	Nhân
<code>/</code>	Chia
<code>^</code>	Lũy thừa
<code>%/%</code>	Chia số nguyên
<code>%%</code>	Số dư từ chia hai số nguyên

Kí hiệu logic

<code>==</code>	Bằng
<code>!=</code>	Không bằng
<code><</code>	Nhỏ hơn
<code>></code>	Lớn hơn
<code><=</code>	Nhỏ hơn hoặc bằng
<code>>=</code>	Lớn hơn hoặc bằng
<code>is.na(x)</code>	Có phải x là biến số missing
<code>&</code>	Và (AND)

	Hoặc (OR)
!	Không là (NOT)

Phát số

numeric(n)	Cho ra n số 0
character(n)	Cho ra n kí tự ""
logical(n)	Cho ra n FALSE
seq(-4, 3, 0.5)	Dãy số -4.0, -3.5, -3.0, ..., 3.0
1:10	Giống như lệnh seq(1, 10, 1)
c(5, 7, 9, 1)	Nhập số 5, 7, 8 và 1
rep(1, 5)	Cho ra 5 số 1: 1, 1, 1, 1, 1.
Gl(3, 2, 12)	Yếu tố 3 bậc, lặp lại 2 lần, tổng cộng 12 số: 1 1 2 2 3 3 1 1 2 2 3 3

Tạo nên số ngẫu nhiên bằng mô phỏng theo các luật phân phối (simulation)

rnorm(n, mean=0, sd=1)	Phân phối chuẩn (normal distribution) với trung bình = 0 và độ lệch chuẩn = 1.
rexp(n, rate=1)	Phân phối mũ (exponential distribution)
rgamma(n, shape, scale=1)	Phân phối gamma
rpois(n, lambda)	Phân phối Poisson
rweibull(n, shape, scale=1)	Phân phối Weibull
rcauchy(n, location=0, scale=1)	Phân phối Cauchy
rbeta(n, shape1, shape2)	Phân phối beta
rt(n, df)	Phân phối t
rchisq(n, df)	Phân phối Chi bình phương
rbinom(n, size, prob)	Phân phối nhị phân (binomial)
rgeom(n, prob)	Phân phối geometric
rhyper(nn, m, n, k)	hypergeometric
rlnorm(n, meanlog=0, sdlog=1)	Phân phối log normal
rlogis(n, location=0, scale=1)	Phân phối logistic
rnbinom(n, size, prob)	Phân phối negative Binomial
runif(n, min=0, max=1)	Phân phối uniform

Biến đổi số thành kí tự và ngược lại

as.numeric(x)	Biến đổi x thành biến số số học để có thể tính toán
as.character(x)	Biến đổi x thành biến số chữ (character) để phân loại
as.logical(x)	Biến đổi x thành biến số logic
factor(x)	Biến đổi x thành biến số yếu tố

Data frames

<code>data.frame(x, y)</code>	Nhập x và y thành một data frame
<code>tuan\$age</code>	Chọn biến số age từ dataframe tuan.
<code>attach(tuan)</code>	Đưa dataframe tuan vào hệ thống R
<code>detach(tuan)</code>	Xóa bỏ dataframe tuan khỏi hệ thống R

Hàm số toán

<code>log(x)</code>	Logarit bậc e
<code>log10(x)</code>	Logarit bậc 10
<code>exp(x)</code>	Số mũ
<code>sin(x)</code>	Sin
<code>cos(x)</code>	Cosin
<code>tan(x)</code>	Tangent
<code>asin(x)</code>	Arcsin (hàm sin đảo)
<code>acos(x)</code>	Arccosin (hàm cosin đảo)
<code>atan(x)</code>	Arctang(hàm tan đảo)

Hàm số thống kê

<code>min(x)</code>	Số nhỏ nhất của biến số x
<code>max(x)</code>	Số lớn nhất của biến số x
<code>which.max(x)</code>	Tìm dòng nào có giá trị lớn nhất của biến số x
<code>which.min(x)</code>	Tìm dòng nào có giá trị nhỏ nhất của biến số x
<code>length(x)</code>	Tổng số yếu tố (elements) trong một biến số (hay số mẫu)
<code>sum(x)</code>	Số tổng của biến số x
<code>range(x)</code>	Khác biệt giữa <code>max(x)</code> và <code>min(x)</code>
<code>mean(x)</code>	Số trung bình của biến số x
<code>median(x)</code>	Số trung vị (median) của biến số x
<code>sd(x)</code>	Độ lệch chuẩn (standard deviation) của biến số x
<code>var(x)</code>	Phương sai (variance) của biến số x
<code>cov(x, y)</code>	Hiệp biến (covariance) giữa hai biến số x và y
<code>cor(x, y)</code>	Hệ số tương quan (coefficient of correlation) giữa biến số x và y.
<code>quantile(x)</code>	Chỉ số của biến số x
<code>cor(x, y)</code>	Hệ số tương quan (correlation coefficient) giữa biến số x và y
<code>is.na(x)</code>	Kiểm tra xem x có phải là số trống không (missing value)
<code>complete.cases(x1, x2, ...)</code>	Kiểm tra nếu tất cả x1, x2, ... đều không có số trống.

Chỉ số ma trận

<code>x[1]</code>	Số đầu tiên của biến số <code>x</code>
<code>x[1:5]</code>	Năm số đầu tiên của biến số <code>x</code>
<code>x[y<=30]</code>	Chọn <code>x</code> sao cho <code>y</code> nhỏ hơn hoặc bằng 30
<code>x[sex=="male"]</code>	Chọn <code>x</code> sao cho <code>sex</code> bằng <code>male</code>

Nhập dữ liệu

<code>data(name)</code>	Xây dựng một kho dữ liệu
<code>read.table("name")</code>	Đọc / nhập số liệu từ file <code>name</code>
<code>read.csv("name")</code>	Đọc / nhập số liệu dạng excel (cách nhau bằng “;”) từ file <code>name</code>
<code>read.delim("name")</code>	Đọc / nhập số liệu dạng tab delimited
<code>read.delim2("name")</code>	Đọc / nhập số liệu dạng tab delimited, cách nhau bằng “;” và số thập phân là “.”
<code>read.csv2("name")</code>	Đọc / nhập số liệu dạng csv, cách nhau bằng “;” và số thập phân là “.”

Phần phụ trong read.table

<code>header=TRUE</code>	Hàng đầu tiên của dữ liệu là tên của biến số
<code>sep=","</code>	Số liệu ngăn cách bằng dấu hiệu “;”
<code>dec="."</code>	Số thập phân là “.” (để phân biệt với “;”)
<code>na.strings="."</code>	Số liệu trống (missing value) là “.”

Phân phối thống kê

<code>pnorm(x, mean, sd)</code>	Phân phối chuẩn
<code>plnorm(x, mean, sd)</code>	Phân phối chuẩn logarit
<code>pt(x, df)</code>	Phân phối t
<code>pf(x, n1, n2)</code>	Phân phối F
<code>pchisq(x, df)</code>	Phân phối Chi bình phương
<code>ppois(x, lambda)</code>	Phân phối Poisson
<code>punif(x, min, max)</code>	Phân phối uniform (đồng dạng)
<code>pexp(x, rate)</code>	Phân phối hàm mũ
<code>pgamma(x, shape, scale)</code>	Phân phối gamma
<code>pbeta(x, a, b)</code>	Phân phối beta

Phân tích thống kê

<code>t.test</code>	Kiểm định t
<code>pairwise.t.test</code>	Kiểm định t cho paired design
<code>cor.test</code>	Kiểm định hệ số tương quan

	method = "kendall"
	method = "spearman"
var.test	Kiểm định phương sai
bartlett.test	Kiểm định nhiều phương sai
wilcoxon.test	Kiểm định Wilcoxon
kruskal.test	Kiểm định Kruskal
friedman.test	Kiểm định Friedman
lm(y ~ x)	Phân tích hồi qui tuyến tính (linear regression)
lm(y ~ factor)	Phân tích phương sai 1 chiều (1-way analysis of variance)
lm(y ~ factor+x)	Phân tích hiệp biến (analysis of covariance)
lm(y ~ x1+x2+x3)	Phân tích hồi qui tuyến tính đa biến số (multiple linear regression)
binom.test	Kiểm định nhị phân (Binomial test)
prop.test	Kiểm định so sánh nhiều tỉ số
prop.trend.test	Kiểm định so sánh nhiều tỉ số theo xu hướng
fisher.test	Kiểm định Fisher
chisq.test	Kiểm định Chi bình phương
glm(y~x1+x2+x+x3)	Phân tích hồi qui logistic
s<-Surv(time,event)	Phân tích survival
survfit(s)	Biểu đồ Kaplan-Meier
survdifff(s~g)	Kiểm định Log-rank giữa hai nhóm g
coxph(s ~ x`+x2)	Phân tích hồi qui Cox

Đồ thị

plot(y~x)	Vẽ đồ thị y và x (scatter plot)
hist(x)	Vẽ đồ thị y và x (scatter plot)
plot(y ~ x z)	Vẽ hai biểu đồ x và y theo từng nhóm của z
pie(x)	Vẽ đồ thị tròn
boxplot(x)	Vẽ đồ thị theo dạng hình hộp
qqnorm(x)	Vẽ phân phối quantile của biến số x
qqplot(x, y)	Vẽ phân phối quantile của biến số y theo x
barplot(x)	Vẽ biểu đồ hình khối cho biến số x
hist(x)	Vẽ histogram cho biến số x
stars(x)	Vẽ biểu đồ sao cho biến số x
abline(a, b)	Vẽ đường thẳng với intercept=a và slope=b
abline(h=y)	Vẽ đường thẳng ngang
abline(v=x)	Vẽ đường thẳng đứng
abline(lm.object)	Vẽ đồ thị theo mô hình tuyến tính

Một số thông số cho đồ thị

pch	Kí hiệu để vẽ đồ thị (pch = <i>plotting characters</i>)
mfrow, mfcol	Tạo ra nhiều cửa sổ để vẽ nhiều đồ thị cùng một lúc (<i>multiframe</i>)
xlim, ylim	Cho giới hạn của trục hoành và trục tung
xlab, ylab	Viết tên trục hoành và trục tung
lty, lwd	Dạng và kích thước của đường biểu diễn
cex, mex	Kích thước và khoảng cách giữa các kí tự.
col	Màu sắc