18
Phục lục 2
Một số lệnh thông dụng trong R

Lệnh về môi trường vận hành của R

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

Lệnh cơ bản

`ls()`	Liệt kê các đối tượng (objects) trong bộ nhớ
`rm(object)`	Xóa bỏ đối tượng
`seach()`	Tìm hướng

Kí hiệu tính toán

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

Kí hiệu logic

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

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

Hàm số toán

`log(x)`	Logarít bậc e
`log10(x)`	Logarít bậc 10
`exp(x)`	Số mũ
`sin(x)`	Sin
`cos(x)`	Cosin
`tan(x)`	Tangent
`asin(x)`	Arcsin (hàm sin đảo)
`acos(x)`	Arccosin (hàm cosin đảo)
`atan(x)`	Arctang(hàm tan đảo)

Hàm số thống kê

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

Chỉ số ma trận

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

Nhập dữ liệu

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

Phần phụ trong read.table

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

Phân phối thống kê

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

Phân tích thống kê

`t.test`	Kiểm định t
`pairwise.t.test`	Kiểm định t cho paired design
`cor.test`	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 Khi 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
`survdiff(s~g)`	Kiểm định Log-rank giữa hai nhóm g
`coxph(s ~ x1+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 = *p*lotting *ch*aracters)
`mfrow, mfcol`	Tạo ra nhiều cửa sổ để vẽ nhiều đồ thị cùng một lúc (*m*ulti*fr*ame)
`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