

R Reference Card

Sae Yeun Lee
CRM Consultant
AMA.COMM
Seoul, South Korea
mandooms@gmail.com

Chel Hee Lee
gnustats@gmail.com

February 18, 2010

By Tom Short, EPRI PEAC, tshort@epri-peac.com 2004-11-07. Granted to the public domain. See www.Rpad.org for the source and the latest version. We appreciate for sharing this wonderful R reference card. This is a simple translation of Tom's original reference card into Korean.

이 문서는 Tom Short (tshort@epri-peac.com) 에 의해서 최초로 작성되어 공개된 문서입니다. 원본소스는 www.Rpad.org 에서 찾을수 있습니다. 이문서는 단순히 Tom의 원본을 번역한 것입니다.

1 Getting help

대부분의 R 함수는 온라인 도움말이 있습니다.

help(topic) 주제와 관련된 문서를 찾습니다.

?topic 위와 같음.

help.search("topic") 도움말 시스템에서 찾습니다.

apropos("topic") 주제와 관련있는 오브젝트를 찾아냅니다.

니다.

help.start() HTML버전의 도움말을 시작합니다.

str(a) R 오브젝트의 내부 구조를 보여줍니다.

summary(a) 일반적으로는 통계량의 요약은 보여주지만, 오브젝트의 특성에 따라 다르게 작동합니다.

ls() 검색경로의 오브젝트를 보여줍니다;

ls.str() 검색경로의 오브젝트에 대한 내부 구조를 보여줍니다.

dir() 현재 사용중인 디렉토리를 보여줍니다.

methods(a) 오브젝트 a에 대한 S3방법을 보여줍니다.

methods(class=class(a)) class a에 대한 사용한 가능한 방법들을 보여줍니다.

2 Input and output

load() 저장해놓은 데이터셋을 불러옵니다.

data(x) 특정 데이터셋을 불러옵니다.

library(x) 추가된 패키지를 불러옵니다.

read.table(file) 테이블 형태나 데이터 프레임 형태로 된 파일을 읽습니다; 기본으로 설정된 구분자(separator)는 공백입니다; **header=TRUE** 는 가장 첫번째 행을 열의 이름으로 인식합니다; **as.is=TRUE**는 문자벡터가 요인변수로 변환되는 것을 방지합니다; **comment.char=" "**를 사용하면, **"#"**이 코멘트로 인식되는 것을 막아줍니다; **skip=n** 은 데이터에서 **n**번째 행을 제외하고 불러옵니다; 도움말을 참조하면 행이름이나 NA처리방법, 기타 다른 옵션에 대해 알 수 있습니다.

read.csv("filename", header=TRUE) id. 하지만 기본 설정은 쉼표로 분리된 파일입니다.

read.delim("filename", header=TRUE) id. 하지만 기본 설정은 탭으로 분리된 파일입니다.

read.fwf(file, widths, header=FALSE, sep=""), as.is=FALSE) 테이블의 가로길이가 정해져있는 파일을 불러옵니다; 가로길이는 정수이어야 하고, 고정된 값으로 주어져야 합니다.

save(file, ...) 특정 오브젝트 (...) 를 XDR 플랫폼의 독립적인 binary 형식으로 저장합니다.

save.image(file) 모든 오브젝트를 저장합니다.

cat(..., file=" ", sep=" ") 요소들을 문자형식으로 바꾸어 출력합니다; **sep**는 요소들을 구분하는 옵션입니다.

print(a, ...) 요소들을 출력합니다; 오브젝트의 형식이 다르면 출력방법도 달라집니다.

format(x,...) 오브젝트를 출력하기에 적절한 형태로 보여줍니다.

write.table(x, file=" ", row.names=TRUE, col.names=TRUE, sep=" ") x를 데이터프레임 형태로 변환한 뒤 출력합니다;

3 Data creation

c(...) 벡터 형식의 결합된 인자들을 생성하는 함수입니다; **recursive=True** 옵션을 사용하면 모든 요소를 하나의 벡터로 **from:to** 수열을 만듭니다; **":"** 기호가 우선권을 갖습니다; **1:4 + 1**은 **"2, 3, 4, 5"**가 됩니다.

seq(from, to) **by=** 로 지정해준 크기만큼 증가하는 수열을 생성합니다; **length=** 를 사용해서 길이를 지정할 수 있습니다.

seq(along=x) 1, 2, ..., length(along) 의 수열을 생성합니다; **loops**를 사용할 때 유용합니다.

rep(x, times) x를 times만큼 반복합니다; **each=** 옵션은 각각의 요소들을 해당 값만큼 반복합니다;

rep(c(1,2,3),2) 은 1 2 3 1 2 3이 되고, **rep(c(1,2,3), each=2)** 는 1 1 2 2 3 3이 됩니다.

data.frame(...) 이름이 있거나 없는 데이터 프레임을 만듭니다; **data.frame(v=1:4, ch=c("a", "B", "c", "d"), n=10)**; 길이가 맞지 않는 경우에는 긴 쪽에 맞춰서 반복됩니다.

list(...) 이름이 있거나 없는 리스트를 만듭니다; **list(a=c(1,2), b="hi", c=3i)**;

array(x, dim) 데이터 x를 가지고 배열을 생성합니다; 세 부적인 차원은 **dim=c(3,4,2)** 등과 같이 설정할 수 있습니다; x의 길이가 충분하지 않을 때는 x가 반복해서 들어갑니다.

matrix(x, nrow=, ncol=) 행렬; x가 반복되는 행렬을 생성합니다.

factor(x, levels=) 벡터x를 요인변수로 바꾸어 줍니다.

gl(n, k, length=n*k, labels=1:n) 정해진 패턴에 따라 레벨을 만듭니다; n는 레벨의 단위이고, k는 반복될 횟수입니다.

expand.grid() 벡터나 요인의 모든 조합을 데이터 프레임으로 보여줍니다.

rbind(...) 행렬, 데이터 프레임, 혹은 다른 형태의 데이터를 행으로 결합합니다.

cbind(...) 위와 같습니다. 열로 결합합니다.

4 Slicing and extracting data

Indexing vectors

x[n] n번째 관측치

x[-n] n번째 관측치를 제외한 나머지

x[1:n] 처음부터 n번째까지의 관측치

x[-(1:n)] n+1번째 관측치부터 끝까지

x[c(1, 4, 2)] 특정 관측치

x["name"] "name"이란 이름을 가지고 있는 관측치

x[x<3] 3보다 큰 모든 관측치

x[x<3 & x<5] 3보다 크고 5보다 작은 모든 관측치

5 Indexing lists

x[n] n번째 요소를 리스트로 보여줍니다.

x[[n]] 리스트의 n번째 요소를 보여줍니다.

x[["name"]] "name"이란 이름을 가지고 있는 요소를 보여줍니다.

x\$name 위와 같습니다.

6 Indexing matrices

x[i,j] i번째 행, j번째 열을 보여줍니다.

x[i,] i번째 행을 보여줍니다.

x[,j] j번째 열을 보여줍니다.

x[,c(1,3)] 1열과 3열을 보여줍니다.

x["name"] 이름이 "name"인 행을 보여줍니다.

Indexing data frames (matrix indexing plus the following)

x[["name"]] 이름이 "name"인 열을 보여줍니다. **x\$name** 위와 같습니다.

7 Variable conversion

as.array(x), as.data.frame(x), as.numeric(x), as.logical(x), as.complex(x), as.character(x), ... 타입을 변환합니다; 완전한 리스트의 경우에는 **methods(as)**를 사용합니다.

8 Variable information

is.na(x), is.null(x), is.array(x), is.data.frame(x), is.numeric(x), is.complex(x), is.character(x), ... 타입을 확인할 때 사용합니다; 완전한 리스트의 경우에는 **methods(is)**를 사용합니다.

length(x) x의 길이를 확인합니다.

dim(x) 오브젝트의 차원을 설정합니다.

dimnames(x) 오브젝트의 차원의 이름을 설정합니다.

nrow(x) 행의 개수; **NROW(x)** 는 똑같지만 1행으로 구성된 행렬에 적용합니다.

ncol(x), **NCOL(x)** 위와 같음. 열에 대해서 적용합니다.

class(x) x의 클래스를 설정하거나 확인합니다; **class(x)j**-"myclass"

unclass(x) x의 클래스를 삭제합니다.

attr(x, which) x의 어떤 속성을 설정합니다.

attributes(obj) obj의 속성의 리스트를 설정합니다.

9 Data selection and manipulation

which.max(x) x의 요소중에서 가장 큰 값을 불러옵니다.

which.min(x) x의 요소중에서 가장 작은 값을 불러옵니다.

rev(x) x를 거꾸로 배열합니다.

sort(x) 오름차순에 따라서 x를 정렬합니다; 내림차순을 사용하고 싶으면 **rev(sort(x))**를 사용합니다.

cut(x, breaks) x를 간격(요인)에 따라 분리합니다; **breaks**는 분리되는 간격이나 지점의 개수를 의미합니다.

match(x, y) y안에 있는 x의 인자들과 길이가 같은 벡터를 반환합니다.

which(x == a)

choose(n, 1) 조합(combination) 계산을 합니다. $n! / [(n-k)!k!]$

na.omit(x) 결측치(NA)를 표시하지 않습니다. (행렬이나 데이터프레임의 경우에는 결측치에 해당하는 행 전체를 표시하지 않습니다.)

na.fail(x) x가 하나라고 결측치를 가지고 있으면, 에러메시지를 나타냅니다.

unique(x) x가 벡터나 데이터 프레임의 형태이면, 반복되는 값을 제외한 형태로 보여줍니다.

table(x) x의 관측값의 개수를 테이블로 만들어줍니다.

subset(x, ...)

sample(x, size)

prop.table(x, margin=)

10 Math

sin, cos, tan, asin, acos, atan, atan2, log, log10, exp

max(x) x의 최대값을 구합니다.

min(x) x의 최소값을 구합니다.

range(x) x의 범위를 구합니다. **c(min(x), max(x))**

sum(x) x의 구성요소의 합을 구합니다.

diff(x) 벡터x의 관측치 사이의 차이를 계산합니다.

prod(x) x의 구성요소의 곱을 구합니다.

mean(x) x의 평균을 구합니다.

median(x) x의 중간값을 구합니다.

quantile(x, probs=) 주어진 확률에 따른 sample quantile을 구합니다.. (기본설정은 0, .25, .5, .75, 1)

weighted.mean(x, w) 가중치w를 적용한 x의 평균을 구합니다.

rank(x) x의 랭크를 구합니다.

var(x) or **cov(x)** (n-1에서의) x의 분산을 구합니다. 만약에 x가 행렬이나 데이터 프레임이라면 분산-공분산 행렬을 계산합니다.

sd(x) x의 표준편차를 구합니다.

cor(x) x가 행렬이나 데이터 프레임일 경우, 상관계수 행렬을 보여줍니다.

var(x, y) or **cov(x, y)** x와 y의 공분산 혹은 x, y가 행렬이나 데이터 프레임인 경우에는 x의 열과 y의 열의 공분산을 계산합니다.

cor(x, y) x와 y사이의 선형 상관관계를 구하거나 행렬이나 데이터 프레임의 경우에는 상관관계행렬을 보여줍니다.

round(x, n) 소수점 n번째자리까지 반올림을 합니다.

log(x, base) base를 밑으로 하는 로그 계산을 합니다.

scale(x)

11 Matrices

t(x) 전치행렬을 구합니다.

diag(x) 대각원소를 구합니다.

solve(a, b) x에 대한 a

solve(a) a의 역행렬을 구합니다.

rowsum(x) 오브젝트와 마찬가지로 행렬의 행에 대한 합을 구합니다; **rowSums(x)**는 좀 더 빠른 계산을 할 수 있습니다.

colsum(x), **colSums(x)** 위와 같습니다..

rowMeans(x) row means의 빠른 버전입니다.

colMeans(x) 위와 같습니다.

12 Advanced data processing

apply(X, INDEX, FUN=)

13 Strings

paste(...) 벡터를 문자형으로 변환 후 연결합니다;

sep= 는 구분자를 의미합니다(기본은 공백으로 되어 있습니다.);

collapse=

substr(x, start, stop) substrings

14 Dates and Times

Date 클래스는 시간을 제외한 날짜정보를 가지고 있습니다.

POSIXct는 날짜와 지역마다의 시간대를 포함한 시간정보를 가지고 있습니다. 비교문(예를 들어 `<`, `seq()`, `difftime()`) 등이 유용하게 쓰입니다. **Date**는 또한 `+`와 `-`가 가능합니다.

?DateTimeClasses 명령어를 사용하면 좀 더 많은 정보를 확인할 수 있습니다. **chron** 패키지를 참고하세요. **as.Date(s)** 그리고 **as.POSIXct(s)** 각각의 클래스로 변환합니다; **format(dt)** 는 문자열로 바꿔줍니다. 기본적인 형식은 "2001-02-21" 입니다. These accept a second argument to specify a format for conversion. 몇 가지 형식은 아래와 같습니다.

15 최적화, 모델링

optim(par, fn, method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN")) 일반적인 목적의 최적화 함수; **par**는 초기값, **fn**은 최적화를 위한 함수 (normally minimize)

nlm(f,p) Newton-type 알고리즘을 이용해서 함수 **f**를 최소화시켜줍니다. **P**는 초기값입니다.

lm(formula) 선형모형을 적합시킨다; 종속변수의 기본적인 형태는 termA + termB + ... 입니다; $I(x*y) + I(x^2)$ 이런 형식을 사용하면 비선형 모델을 만들 수 있습니다.

glm(formula,family=) fit generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution; family is a description of the error distribution and link function to be used in the model; ?family를 참고하세요.

nls(formula) 비선형모델에 대한 추정을 합니다.

approx(x,y=) 주어진 자료를 선형으로 interpolate 합니다. x can be an xy plotting

16 structure

spline(x,y=) cubic spline interpolation

loess(formula) fit a polynomial surface using local fitting
Many of the formula-based modeling functions have several common arguments: data= the data frame for the formula variables, subset= a subset of variables used in the fit, na.action= action for missing values: "na.fail", "na.omit", or a function. 다음은 모델을 적합시킬 때 자주 사용되는 일반적인 함수들입니다:

predict(fit,...) 적합모델에 대한 예측을 합니다

df.residual(fit) 잔차에 대한 자유도를 구합니다

coef(fit) 추정계수를 반환합니다(표준오차와 함께 나오는 경우도 있습니다)

residuals(fit) 잔차를 반환합니다

deviance(fit) 편차를 반환합니다

fitted(fit) fitted value를 반환합니다

logLik(fit) computes the logarithm of the likelihood and the number of

parameters

AIC(fit) AIC를 계산합니다

17 Statistics

aov(formula) 분산분석을 합니다

anova(fit,...) 하나 혹은 그 이상의 모델에 대한 분산분석 테이블을 만듭니다

density(x) x에 대한 kernel density 를 추정합니다

binom.test(), **pairwise.t.test()**, **power.t.test()**, **prop.test()**, **t.test()**, ... **help.search("test")** 참고하세요.

분포

rnorm(n, mean=0, sd=1) 가우시안분포(정규)

rexp(n, rate=1) 지수분포

rgamma(n, shape, scale=1) 감마분포

rpois(n, lambda) 포아송분포

rweibull(n, shape, scale=1) 웨이블분포

rcauchy(n, location=0, scale=1) 코시분포

rbeta(n, shape1, shape2) 베타분포

rt(n, df) 'Student' (t)

rf(n, df1, df2) Fisher?Snedecor (F) (c2)

rchisq(n, df) Pearson

rbinom(n, size, prob) 이항분포

rgeom(n, prob) 기하분포
rhyper(nn, m, n, k) 초기하분포
rlogis(n, location=0, scale=1) 로지스틱분포
rlnorm(n, meanlog=0, sdlog=1) 로그-정규분포
rnbinom(n, size, prob) 음이항분포
runif(n, min=0, max=1) 균일분포
rwilcox(nn, m, n), rsignrank(nn, n) Wilcoxon 통계량
 이 함수들은 모두 r을 d, p, q 등으로 바꿔서 사용할 수 있는데, 각각 확률밀도(dfunc(x, ...)), 누적확률밀도(pfunc(x,...)), 분위수(qfunc(p, ...), with $0 \leq p \leq 1$)를 나타냅니다

18 Programming

function(arglist) expr 함수 정의
return(value)
if(cond) expr
if(cond) cons.expr else alt.expr
for(var in seq) expr
while(cond) expr
repeat expr
break
next
 를 이용하여 묶어줘야 합니다
ifelse(test, yes, no) a value with the same shape as test filled with elements from either yes or no
do.call(funname, args) executes a function call from the

name of the function and a list of arguments to be passed to it

19 참여

잘 못된 내용을 찾으셨거나, 보완, 수정 그리고 더 좋은 레퍼런스 카드를 만들기 위해서 함께 하실 분은 연락주세요.