R 参考卡片

原文档由 Tom Short 撰写，你可以在 www.Rpad.org 上得到最新文档。中文版本文档 (已获得 Tom Short 的翻译许可) 结构上同原版类似，局部添加了若干常用命令。后续修订以及维护由刘思喆负责。
项目同步在这里 https://github.com/sunbjt/r_reference
update:2017 年 12 月 3 日

帮助和基础

大部分 R 函数都有在线文档。
help(topic) 关于 topic的文档.
?topic 同上
help.search("topic") 搜索帮助系统
apropos("topic") 返回在搜索路径下包含 (部分) 关键词"topic" 的所有对象名称
help.start() HTML 形式的帮助
demo() R 功能演示
example(f) 运行在线帮助中的例子
str(a) 显示 R 对象的内在属性 (*str*ucture) 或简要说明对象
summary(a) 给出 a的概要, 通常是一个一般性统计概要; 且它对不同属性的 a 有不同的操作方式.
ls() 显示' 搜索路径' 下的对象; 也可按指定条件搜索。
ls.str() str() 搜索路径下的每个变量与其属性
dir() 显示当前目录下的文件
list.files() 同上
getwd() 获得工作路径信息
setwd() 设置工作路径
methods(a) 显示 a的'S3 methods'
methods(class=class(a)) 列表所有可以解决属于对象类的方法
options(...) 设置或检验全局参数; 常用参数有: width, digits, error
install.packages(pkg) 安装 pkg 包
update.packages() 自动对比包版本，并询问更新
library(pkg) 加载 pkg 包
require(x) 同上
library(help=pkg) 展示包 pkg 的信息
attach(x) 将 x 指向 R 的搜索路径; x 可以使一个列表, 数据框, 或者是由 save创建的 R data file. 使用 search()来显示搜索路径.
detach(x) attach的逆过程.
assign(x,value) 将 value赋值给 x , 即"$<-$"
quit() 退出当前 R 会话 (q() 或 Ctrl_z)

输入与输出

load() 加载由 save命令得到的资料集
data(x) 加载指定的数据
edit() 调用文本编辑器修改 R 对象
fix(x) 'fix' 调用'edit' 修改'x'
data.entry(x) 电子数据表形式的录入编辑器
scan(x) 从控制台或文件中读取数据为向量或列表
read.table(file) 读取表格式的文件并将其创建成数据框; 默认分割符 sep=""为任意空白; 使用 header=TRUE 读取第一行作为列标题; 使用 as.is=TRUE防止字符向量变为 factors; 使用 comment.char=""防止"#"被解释为注释; 使用 skip=n 在读数据前跳过 n 行; 详细见帮助关于行命名,NA 处理, 和其他
read.csv("filename",header=TRUE) 同上, 但默认设置为读取 csv 文件 (Comma Separated values)
read.delim("filename",header=TRUE) 同上, 默认设置为读取 tab 分割文件
read.fwf(file,widths,header=F,sep="\t",as.is=FALSE) 以 fixed width formatted 形式读取数据至数据框; widths 是整数向量, 用于设置调整宽度字段
save(file,...) 以不分平台的二进制保存指定的对象
save.image(file) 保存所有的对象
dump("x","...") 将对象 x 保存在'...' 里
cat(..., file="", sep=" ") 强制转化为字符后打印对象的赋值; sep 为对象赋值间的分割符号
print(a, ...) 显示 a的赋值，更一般的，它对于不同的对象可以有不同的表达方式。
format(x,...) 格式化, 更好的显示 R 对象
write.table(x,file="",row.names= T ,col.names= T , sep="") 在把 x 转化为数据框后, 写到文件; 如果 quote 为 TRUE, 字符和因子列就会被 (") 所包围; sep 是字段分隔符; eol 为尾行分割符; na 为缺失值字符串; 使用 col.names=NA 增加列标题以便于和表格输入一致
sink(file) 输出到文件 file, 直到输入命令 sink()

大部分 I/O 函数都有 file 参量. 它经常用一个字符串来命名文件或连接. file="" 意味着标准输入或输出. 连接 (Connections) 可以包含文件 (file), 管道 (pipes), 压缩文件 (zipped files) 或 R 变量
在 windows 操作环境下，数据共享可以通过写字板(clipboard) 的方式. 读取 Excel 表：可以将 Excel 中数据拷贝至写字板(内存) , 使用 x <- read.delim('clipboard') 方式读取数据。
write.table(x, 'clipboard', sep='\t',col.names=NA)可以将数据写入 clipboard, 可直接粘贴到 Excel 中。
数据库方面的交互应用, 请见 RODBC, DBI,RMySQL, RPgSQL, and ROracle包. 读取其他文件格式参考 XML, hdf5, netCDF 等包.

数据创建

c(...) 常见的将一系列参量转化为向量的函数; 通过 recursive=TRUE 降序排列列表并组合所有的元素为向量.
from:to 产生一个序列; ':' 有较高级别的优先级; 1:4 + 1 得到'2,3,4,5'
seq(from,to) 产生一个序列 by= 指定间距; length= 指定要求长度
seq(along=x) 产生 1, 2, ..., length(along); 常用在循环上
rep(x,times) 重复 x times次; 使用 each= 来指定元素 x 重复的次数;rep(c(1,2,3),2) 将得到 1 2 3 1 2 3; rep(c(1,2,3),each=2) 将得到 1 1 2 2 3 3
data.frame(...) 创建数据框, 当然变量可能被命名或不被命名; data.frame(v=1:4,ch=c("a","B","c","d"),n=10); 相对较短的向量会被循环填充到最长向量长度
list(...) 创建一个由变量组成的列表, 变量可能被命名或未被命名; list(a=c(1,2),b="hi",c=3i);
array(x,dim=) 产生由 x 组成的数组; 使用类似 dim=c(3,4,2)指定维数; 如果 x 不够长度, 则 x 自动循环
matrix(x,nrow=,ncol=) 矩阵; 同上
factor(x,levels=) 把向量 x 编码成为因子
gl(n,k,length=n*k,labels=1:n) 通过指定水平方式产生水平 (因子); k 为水平的个数; n 为重复的次数
expand.grid() 提供的向量或因子所有组合构成的数据框
rbind(...) 把以行的形式组合矩阵, 数据框, 或其他
cbind(...) 同上. 以列的形式

数据分割和选取

向量索引

x[n]	第 n个元素
x[-n]	除了第 n个元素的 x
x[1:n]	前 n个元素
x[-(1:n)]	第 n+1 至最后的元素
x[c(1,4,2)]	指定元素
x["name"]	名为"name"的元素
x[x > 3]	所有大于 3 的元素
x[x > 3 & x < 5]	区间 (3,5) 的元素
x[x %in% c("a","and","the")]	给定组中的元素

列表索引

x[n]	列表显示元素 n
x[[n]]	列表的第 n个元素
x[["name"]]	名为"name"的元素
x$name	同上.

矩阵索引

x[i,j]	下标为 (i,j) 的元素
x[i,]	第 i行
x[,j]	第 j列
x[,c(1,3)]	第 1 和 3 列
x["name",]	名为"name"的行

数据框索引 (*矩阵索引加下述*)

x[["name"]]	列名为 "name"的列
x$name	同上.

变量变换

as.array(x), as.data.frame(x), as.numeric(x),
as.logical(x), as.complex(x), as.character(x), 等,
转换变量类型; 使用如下命令得到全部列表: methods(as)

变量信息

is.na(x), is.null(x), is.array(x), is.data.frame(x),
is.numeric(x), is.complex(x), is.character(x), ... 检验变量类型; 使用如下命令得到全部列表, methods(is)
dim(x) 重新设置或设置对象的维数; dim(x) <- c(3,2)
dimnames(x) 重新设置或设置对象的名称
nrow(x) 和 **NROW(x)** 返回行的个数 dim(x)[1]
ncol(x) 和 **NCOL(x)** 返回列的个数 dim(x)[2]
class(x) 得到或设置 x 的类;class(x) <- "myclass"
names(x) 查看或设置对象名称 (names)
unname(x) 删除 R 对象的名称 (names) 或维名称 (dimnames)
unlist(x) 将列表 x 转化为向量
attributes(obj) 得到或设置 obj 的属性列表

数据选择和操作

which.max(x) 返回 **x**中最大元素的索引
which.min(x) 返回 **x**中最小元素的索引
rle(x) 返回游程(Runs) 信息
sort(x) 升序排列 x 中的元素; 降序排列使用:rev(sort(x))
cut(x,breaks) 将 x 分割成为几段 (或因子); breaks为分割的段数或分割点向量.
match(x, y) 返回一个和 x 相同长度且和 y 中元素相等的向量,(不等的元素返回 NA)
which(x == a) 如果比较操作为真()`~, 返回向量 x 的索引
choose(n, k) 组合数 $=n!/[(n-k)!k!]$
sign(x) 判断变量是否大于 0, 大于返回“1”, 小于返回“-1”, 等于返回“0”
na.omit(x) 去除缺失值(NA), 如果 x 为矩阵或数据框, 去除相关行
na.fail(x) 返回错误信息如果 x 包含至少一个 NA
unique(x) 如果 x 为向量或数据框, 返回惟一值
duplicated(x) 返回向量或数据框 x 重复元素的逻辑值
table(x) 返回一个由 x 不同值个数组成的表格 (常用于整数或因子), 即频数表
subset(x, ...) 根据条件(...选取 x 中的元素, 如:x$V1 < 10) ; 如果 x 为数据框, 选项 select 通过使用负号的方式保留或去除变量
sample(x, size) 不放回的随机在向量 x 中抽取 size个元素, 选项 replace = TRUE允许放回抽取
prop.table(x,margin=) 根据 margin 使用分数表示表格, 无 margin 时, 所有元素和为 1

数学

+,-,×,÷,^,%%,%/%
< > <= >= ==..!=..
sin,cos,tan,asin,acos,atan,atan2,log,log10,exp
range(x) 返回 c(min(x), max(x))
abs(x) x 的绝对值
sqrt(x) $x^{0.5}$
quantile(x,probs=) 样本分位数, 默认为 0,0.25,0.75,1
IQR(x) 返回数据中间 50% 的区间
weighted.mean(x, w) 加权平均
var(x) or cov(x) 向量 x 的样本方差; 如果 x 是矩阵或数据框, 协方差矩阵将被计算
cor(x) 如果 x 是矩阵或数据框, 返回相关阵 (1 如果 x 为向量)
var(x, y) or cov(x, y) x 和 y 间的协方差; 如果 x ,y 为矩阵或数据框, 返回 x 和 y 各列的协方差
cor(x, y) x 和 y 线性相关系数; 或者相关阵, 如果 x 和 y 为矩阵或数据框
round(x, n) x 的约数, 精确到 n 位
log(x, base) 计算 x 以 base为底的对数,$e = \exp(1)$
scale(x) 如果 x 是一个矩阵, 则中心化和标准化数据; 若只标准化则使用选项 center=FALSE, 若只中心化使用 scale=FALSE (默认 center=TRUE, scale=TRUE)
integrate(f,lower,upper) 函数 f 在区间 (lower,upper) 的面积 (积分)
pmin(x,y,...) x[i], y[i]相比较小者, 组成新的向量
pmax(x,y,...) 同上. 较大者
cumsum(x) 由 x 组成的向量,x[i]=sum{x[1]:x[i]}
cumprod(x) 同上. 连乘
cummin(x) 同上. 最小
cummax(x) 同上. 最大
union(x,y) $x \cup y - x \cap y$
intersect(x,y) $x \cap y$
setdiff(x,y) $x - x \cap y$
setequal(x,y) 返回比较 x , y 是否相同的逻辑值 (x ,y不涉及顺序).
is.element(el,set) 同 x %in% y
Re(x) 复数的实部
Im(x) 虚部
Mod(x) 绝对值 (模); 同 abs(x)
Arg(x) 复数角度 (in radians)
Conj(x) 求 x 的共轭复数
convolve(x,y) 计算两个序列的卷积
fft(x) 排列 (array) 的快速傅立叶变换
mvfft(x) 矩阵各列的快速傅立叶变换
filter(x,filter) 对但变量时间序列或多变量时间序列的单独序列进行线性过滤
大多数学函数使用逻辑参数 na.rm=FALSE 来指定是否移除缺失值 (NA).

矩阵

t(x) 转置
diag(x) 对角阵
%*% 矩阵运算
solve(a) 矩阵的逆
eigen(x) 计算矩阵的特征根和特征向量
rowsum(x) 矩阵格式对象行加和; **rowSums(x)** 是一个更快的版本
colsum(x), **colSums(x)** 同上. 列
rowMeans(x) 行平均
colMeans(x) 列平均
dist(x) 计算矩阵 x 行间的距离

高级数据处理

tapply(X,INDEX,FUN=) 根据 x 的索引 (INDEX) 对不完全 (ragged) 的数列应用 FUN
sapply 同 lapply, 比之更友好
by(data,INDEX,FUN) 应用函数 FUN 处理数据框 data 中由 INDEX 定义的子集
merge(a,b) 根据共有的列或行名把两个数据框合并
xtabs(a b,data=x) 从交叉分类因子得到列联表
并且以合适的方式返回结果; by是分组元素列表
stack(x, ...) 将分开列形式的数据框或列表中的数据变量转化为单列
unstack(x, ...) stack()的逆过程
reshape(x, ...) 对'wide' 和'long' 格式对数据框进行改造. 'wide' 格式是根据基准变量横向扩展数据框; 'long' 格式是根据基准变量纵向扩展数据框. 使用 (direction='wide') 或 (direction='long') 参数指定格式.
expression(expr) 创建或检验对象是否为表达 (expression) 形式. 参考 is.expression(x), as.expression(x, ...)
parse(file = "", n = NULL) 以列表形式返回解析过, 但没有经过计算的表达 (expression)
eval(expr) 在指定的环境下计算 R 表达 (expression)

字符

paste(...) 转化为字符后连接向量;sep = 为分割界限 (一个空格为默认); 选择 collapse = 可以分割'collapsed' 结果
substr(x,start,stop) 提取字符向量的子字段; 同样可以赋值, 使用 substr(x, start, stop) <- value
strsplit(x,split) 在 split 的位置分割 x **grep(pattern,x)** pattern条件的匹配和替换; 参见?regex
gsub(pattern,replacement,x) 替换满足正则表达式的字段,sub() 类似, 但只替换第一个出现的字段
tolower(x) 将字母转化为小写
toupper(x) 将字母转化为大写
casefold(x, upper = TRUE) 变化 x 为大写 (TRUE) 或小写 (FALSE)
chartr(old, new, x) 将 x 中的字符 old 变换为字符 new
match(x,table) table 中匹配 x 元素位置组成的向量.
x %in% table 同上. 返回逻辑向量
pmatch(x,table) table 中部分匹配 x 元素
nchar(x) 字符的个数

日期和时间

Date 只包含日期不包含时间.POSIXct 包括日期时间和时区信息. 相比而言 (如.>),seq() 和 difftime() 比较有用.Date 也可以使用 + 和 −. ?DateTimeClasses可以给出更多的信息. 详见 chron包.
as.Date(s)和 **as.POSIXct(s)**转化各自的属性;format(dt)转化为字符表达. 默认的字符格式为'2006-07-24'. 他们接受一个次要表达来指定转化的格式. 一些常见的格式为:

%a, %A 精简和无精简' 星期天'(weekday) 名
%b, %B 精简和无精简月名
%d 月份中的日期 (01--31).
%H 小时 (00--23).
%I 小时 (01--12).
%j 年份中的日期 (001--366).
%m 月份 (01--12).
%M 分钟 (00--59).
%p AM/PM 指示.
%S 十进制的秒 (00--61).
%U 星期 (00--53); 第一个星期天作为第一个星期的第一天.
%w 星期天数 (0--6, 周日为 0).
%W 周 (00--53); 第一个周一作为第一个星期的第一天.
%y 无世纪的年 (00--99). 不要使用.
%Y 有世纪的年.
%z (只输出.) 格林威治补偿;-0800为格林威治西 8 小时.
%Z (只输出.) 时区作为字符串 (无效为空).

weekdays(x) 返回日期 x 的星期几
months(x) 返回日期 x 的月份
quarters(x) 返回日期 x 的季节 (Q1 - Q4)
在输出时会碰到, 显示数字前存在零的问题, 但输入时可以选择性写零或无零. 参见?strftime.

图形装置 (Graphics Devices)

x11(), windows() 打开一个绘图窗口
dev.list() 图形窗口列表
dev.set() 指定图形窗口
plot.new() 为绘制新图形结束当前图形窗口
savePlot(file,type) 将当前窗口中的绘图保存到文件 file , 保存文件类型 type 可以选择'wmf', 'mf', 'png', 'jpeg', 'jpg', 'bmp', 'ps', 'eps', 'pdf'
postscript(file) 为创建 PostScript 图形开启图形装置驱动; 使用 horizontal = FALSE, onefile =FALSE, paper = "special" 指定 EPS 格式文件; family=指定字体 (AvantGarde, Bookman, Courier, Helvetica, Helvetica-Narrow, NewCenturySchoolbook, Palatino, Times, or ComputerModern); width= 和 height=指定以 inches 为单位的区域大小; paper=指定纸张类型.
ps.options() 辅助函数, 设置或查看 (如果没有参数)postscript参数的缺省值
pdf, png, jpeg, bitmap, xfig, pictex; 参看 ?Devices
dev.off() 关闭指定 (默认当前) 图形装置; 也可以参考 dev.cur, dev.set
layout() 根据用户给定的 matrix (绘制顺序) 和参数 width=,height= 将图形装置分割

绘图

plot(x) 在 x 轴上顺次地绘制 x 值 (y 轴上)
plot(x, y) 双变量绘图 (散点图)
barplot(x) x 的条形图; 使用 horiz=FALSE 改变绘图水平或垂直
dotchart(x) 如果 x 为数据框, 绘制 Cleveland dot 图
pie(x) 饼图
boxplot(x) 箱线图
curve(expr, from, to ,add = FALSE) 根据给定函数或表达在区间'[from,to]' 上绘制曲线
sunflowerplot(x, y) 是以相似坐标的点作为花朵，其花瓣数目为点的个数
coplot(x~y | z) 根据 z 值或值间隔绘制 x 和 y 的双变量图
interaction.plot (f1, f2, y) 如果 f1 和 f2 是因子, 作 y 的均值图, 以 f1 的不同值作为 x 轴，而 f2 的不同值对应不同曲线: 可以用选项 fun 指定 y 的其他的统计量 (缺省计算均值,fun=mean)
matplot(x,y) 二元图, 其中 x 的第一列对应 y的第一列, x 的第二列对应 y 的第二列, 依次类推。
fourfoldplot(x) 用四个四分之一圆显示 2×2 列联表情况 (x 必须是 dim=c(2, 2,k)的数组, 或者是 dim=c(2, 2) 的矩阵，如果 $k = 1$)
assocplot(x) Cohen--Friendly 图, 显示在二维列联表中行, 列变量偏离独立性的程度
mosaicplot(x) 列联表的对数线性回归残差的马赛克图
plot.ts(x) 如果 x 是类 ts 的对象, 作 x 的时间序列曲线,x 可以是多元的, 但是序列必须有相同的频率和时间
ts.plot(x) 同上, 但如果 x 是多元的，序列可有不同的时间但须有相同的频率
qqnorm(x) 正态分位数图
contour(x, y, z) 绘制轮廓图 (画曲线时使用内插替换补充空白的值) , x 和 y 必须为向量,z 必须为矩阵, 使得 dim(z)=c(length(x),length(y)) (x 和 y 可以省略)
filled.contour(x, y, z) 同上, 等高线之间的区域是彩色的, 并且绘制彩色对应的值的图例
image(x, y, z) 同上, 但是实际数据大小用不同色彩表示
persp(x, y, z) 同上, 但为透视图
stars(x) 如果 x 是矩阵或者数据框, 用星形和线段画出, 星代表 x 的每一行线段代表列的长度.
symbols(x, y, ...) 在由 x 和 y 给定坐标画符号 (圆，正方形，长方形，星，温度计式或者盒形图), 符号的类型、大小、颜色等由另外的变量指定
termplot(mod.obj) 绘制回归模型 (mod.obj) 的 (偏) 影响图
下面的参数经常用于一般绘图函数
add=FALSE 如果 TRUE , 在前一个图上 (如果存在) 添加绘图
axes=TRUE 如果 FALSE , 不绘出坐标轴和盒子
type="p" 指定绘制图的类型,"p": 点,"l": 线,"b" 用线连接的点, "o": 同上. 但线穿过点,"h": 垂直的线, "s": 阶梯, 但数据由垂直线的顶端代表, "S": 阶梯, 但数据由垂直线的底端代表
xlim=, ylim= 指定坐标轴的最小和最大限制
xlab=, ylab= 注释坐标轴
main= 主标题
sub= 副标题 (小号字体)

低级绘图命令

points(x, y) 添加点 (选项 type= 可以使用)
lines(x, y) 同上. 但用线
text(x, y, labels, ...) 在坐标点 (x , y) 加入文字;
典型的使用方法:plot(x, y, type="n"); text(x, y, names)
mtext(text, side = 3, line = 0, ...) 在指定的 side 添加文字 (参考 axis); line指定添加文字的绘图区域
segments(x0, y0, x1, y1) 从点 (x0,y0) 划线至点 (x1,y1)
arrows(x0, y0, x1, y1, angle = 30, code = 2) 同上.
当 code=2以点 (x0,y0) 为基原点的箭头, 当 code=1以点 (x1,y1) 为原点的箭头, 当 code=3双箭头; angle 控制箭头张开的角度
abline(a,b) 以截距为 a 斜率为 b 的斜线
abline(h = y) 在 y点的垂线
abline(lm.obj) 根据 lm.obj做出回归线
rect(x1, y1, x2, y2) 做出左, 右, 底, 高限制为 x1, x2, y1,y2 的四边形
polygon(x, y) 多边形作图
legend(x, y, legend) 在点 (x ,y) 添加图例
title() 添加标题
axis(side, at) 添加坐标轴, 底部 (side=1), 左侧 (2), 顶部 (3) 或右侧 (4); 可选参数 at 指定画刻度线的位置坐标
box() 在当前图形周围加一个盒子
rug(x) 在 x-轴上添加 _rug_(1-d plot), 来代表数据
locator(n, type="n", ...) 在用户使用鼠标在图上点击 n 次后返回 n 次点击的坐标 (x,y)；并可以在点击处绘制符号 (type="p") 或线 (type="l") ，缺省情况下不画符号或连线 (type="n")

绘图参数

可以使用 **par(...)** 来永久性改变绘图参数; 很多参数也可以作为绘图命令的选项.
adj 控制文字对齐方式 (0 左对齐, 0.5 居中对齐, 1 右对齐)
bg 指定背景颜色 (如: bg="red", bg="blue", ... 用 colors()可以显示 657 种可用颜色名)
bty 控制图形边框形状, 可使用的值为:"o", "l", "7","c", "u" 或 "]" (边框和字符想像); 如果 bty="n" 则不绘制边框
cex 控制缺省状态下符号和文字大小的值; 下面的参数有同样的功能: cex.axis, 坐标轴刻度, cex.lab, 坐标轴标签, cex.main, 标题, cex.sub, 副标题
col 控制符号和连线的颜色; 使用颜色名称: "red", "blue" 参考 colors() 或作为 "#RRGGBB"; 参考 rgb(), hsv(), gray(), 和 rainbow(); 同参数 cex 类似: col.axis, col.lab, col.main, col.sub
font 控制文本字体的整数 (1: 正常, 2: 斜体, 3: 粗体, 4: 斜粗体); 还可以使用 font.axis, font.lab, font.main, font.sub
las 控制坐标轴刻度数字标记方向的整数 (0: 平行于轴, 1: 横排, 2: 垂直于轴, 3: 竖排)
lty 控制连线的类型, 可以是整数或字符 (1: "solid", 2: "dashed", 3: "dotted", 4: "dotdash", 5: "longdash", 6: "twodash"), 或不超过 8 个字符的字符串 ("0"至"9"间的数) 交替指定线和空白的长度), 单位为磅 ("points") 和像素, 如 lty="44" 和 lty=2 一样
lwd 控制连线宽度的数字, 默认 1
mar 控制图形边空的有 4 个值的向量 c(bottom, left, top, right), 默认值为 c(5.1, 4.1, 4.1, 2.1)
mex 声明图形同边缘协调程度的字符大小的附加变量。注意，它并不改变字体的大小。
mfcol 用 c(nr,nc) 向量分割绘图窗口为 nr 行和 nc 列, 按列使用子窗口
mfrow 同上. 但按行使用子窗口
pch 控制符号的类型, 可以是 1 至 25 的整数, 或是 "" 里的单个字符
1○ 2△ 3＋ 4× 5◇ 6▽ 7⊠ 8✳ 9⊕ 10⊕ 11⛝ 12⊞ 13⊠ 14⛋ 15■ 16● 17▲ 18◆ 19● 20● 21○ 22□ 23◇ 24△ 25▽ * ⋆ . · X X a a ? ?
ps 控制文字大小的整数，单位为磅 (points)
pty 指定绘图区域类型的字符，"s": 正方形, "m": 最大利用
tck 指定轴上刻度长度的值，单位是百分比，以图形宽、高中最小一个作为基数；如果 tck=1 则绘制 grid
tcl 同上. 但以文本行的高度为基数 (默认为 tcl=-0.5)
xaxt 如果 xaxt="n" 则设置 x-轴但不显示 (有助于和 axis(side=1, ...)一起使用)
yaxt 同上. y-轴

网格 (Lattice) 绘图

xyplot(y~x) 双变量散点图
dotplot(y~x) Cleveland 点图 (逐行逐列累加图)
densityplot(~x) 密度函数图
bwplot(y~x) 箱线图
qqmath(~x) x 关于某理论分布的分位数 -分位数图
stripplot(y~x) 一维图,x 必须是数值型, y可以是因子
qq(y~x) 比较两个分布的分位数, x 必须是数值型, y可以是数值, 字符或者是因子, 但必须是两个' 水平'

splom(~x) 二维图矩阵
parallel(~x) 平行坐标图
levelplot(z~x*y|g1*g2) 在 x , y 坐标点的 z 值的彩色等值线图(x , y和 z等长)
wireframe(z~x*y|g1*g2) 3d 透视图 (面)
cloud(z~x*y|g1*g2) 3d 透视图 (点)
在一般性 Lattice 公式中, y~x|g1*g2 有可选择条件变量 g1 和 g2 组合绘制在单独的'panels' 上. Lattice 函数使用了很多相同的参量作为基础附加绘图, 如 data=,subset=. 使用 panel= 来定义定制'panel' 函数 (参考 apropos("panel")和 ?llines). Lattice 函数返回一个 trellis 类型的对象并且是'print-ed' 来生成图形. 内部使用 print(xyplot(...))函数时, 自动绘图并无效果. 使用 lattice.theme 和 lset 来改变 Lattice 默认设置.

模型拟和

optim(par, fn, method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN") 用于求多元函数的最值. 基于 Nelder-Mead, quasi-Newton and conjugate-gradient 算法. 同时, 也可以求区间内的最值.par 为函数初值, fn 是求最值的函数 (通常为最小)
nlm(f,p) 根据初始值通过使用牛顿 (Newton-type) 算法的最小化函数
lm(formula) 拟和线性模型; formula的典型形式为 response ~ termA + termB + ...; 使用 I(x*y) + I(x^2) 来构成非线性成分
glm(formula,family=) 通过指定线性预测模型和残差分布来拟和广义线性模型; family为残差分布的描述且同模型整合; 详见?family
nls(formula) 非线性最小二乘估计
approx(x,y=) 线性插值;
approxfun(x,y) 线性插值函数
spline(x,y=) 立方 (曲线) 插值
splinefun(x,y) 立方 (曲线) 插值函数
loess(formula) 局部近似回归。利用局部加权回归进行一个非参回归。这种回归对显示一组凌乱数据的趋势和描述大数据集的整体情况非常有用。
很多以公式为基础的模型函数有很多通用的参量: data= 公式变量的数据框, subset= 满足条件的子集; na.action= 缺失值处理方式:"na.fail","na.omit", 或一个函数. 下面常用于模型拟和函数:
predict(fit,...) 通过拟和模型 fit 计算预测值
df.residual(fit) 返回残差的自由度
coef(fit) 返回被估计的系数(有时候还包括他们的标准差)
residuals(fit) 返回残差值
deviance(fit) 返回方差
fitted(fit) 返回拟和值
logLik(fit) 计算对数似然值和参数数目
AIC(fit) 计算 Akaike 信息准则(Akaike information criterion or AIC)

统计

aov(formula) 方差分析
anova(fit,...) 一个或多个模型对象的方差表(或残差平方和表) 分析
kmeans(x) k 均值聚类
hclust(d, method = "complete") 层次聚类分析, d 由函数 dist 构造,method 可参考?hclust
prcomp(x, ...) 主成分分析
factanal(x,factors,data) 因子分析
cancor(x, y, xcenter = TRUE, ycenter = TRUE) 典型相关分析 (canonical correlations)

检验

t.test() t 检验
wilcox.test() Wilcoxon 检验
prop.test(x,n,p) n 次试验中, 出现的 x 的概率是否以概率 p 出现的假设检验
binom.test(x,n) 贝努力试验检验
chisq.test(x,p) χ^2 检验
fisher.test(x ,y = NULL) Fisher 精确性检验
ks.test(x,y="name",) Kolmogorov-Smirnov 检验, 检验向量数据是否服从"name" 分布
shapiro.test(x) Shapiro-Wilk 正态分布检验
PP.test(x, lshort = TRUE) PP(Phillips-Perron) 检验
quada.test(x) quade 检验
friedman.test(x) Friedman 秩和检验
pairwise.t.test(), **power.t.test()**
help.search("test")

分布

rnorm(n, mean=0, sd=1) 高斯 (正态)
rexp(n, rate=1) 指数
rgamma(n, shape, scale=1) γ 分布
rpois(n, lambda) Poisson 分布
rweibull(n, shape, scale=1) Weibull 分布
rcauchy(n, location=0, scale=1) Cauchy 分布
rbeta(n, shape1, shape2) β 分布
rt(n, df) t 分布
rf(n, df1, df2) F 分布
rchisq(n, df) χ^2 分布
rbinom(n, size, prob) 二项
rgeom(n, prob) 几何
rhyper(nn, m, n, k) 超几何
rlogis(n, location=0, scale=1) logistic 分布
rlnorm(n, meanlog=0, sdlog=1) 对数正态
rnbinom(n, size, prob) 负二项分布
runif(n, min=0, max=1) 均匀分布
rwilcox(nn, m, n),rsignrank(nn, n) Wilcoxon 分布
所有的函数都可以使用 d,p 或 q 来替换 r 分别得到概率密度 (dfunc(x, ...)), 累积概率密度 (pfunc(x, ...)), 分位数 (qfunc(p, ...), $0 < p < 1$).

编程

function(arglist) expr 定义函数
return(value)
if(cond) expr
if(cond) cons.expr else alt.expr
for(var in seq) expr
while(cond) expr
repeat expr
break
next
使用表达 (statements) 使用大括号 {}
ifelse(test, yes, no) 如果满足条件 test返回 yes, 反之返回 no
do.call(funname, args) 根据函数名和表达式 (arguments) 执行调用函数.

R 内嵌常数

letters 返回 26 个小写英文字母
LETTERS 同上 (大写)
month.abb 返回 3 个字母缩写的月份名
month.name 返回月份名
pi π

其他

sessionInfo() 显示关于 R 的版本信息和关联的 Packages
all.equal(x,y) 检验两个对象是否 (渐进) 相等, 相等返回 TRUE, 否则返回 $abs(x-y)/x$
identical(x,y) 严格检验对象是否相等
memory.size() 返回当前使用的内存大小
RSiteSearch() 搜索http://search.r-project.org 上的结果, 包括邮件列表, 手册和帮助页