**R code集锦**

* 如果想要用copy paste的方式读取Excel中的数据，可以在R中用一下code

***x <- read.delim(pipe(“pbpaste”))*** 大意就是读取剪切板上的data[[1]](#footnote-1)

> getwd()

[1] "/Users/Maggie"

>setwd("Documents/R Programming/specdata")

> getwd()

[1] "/Users/Maggie/Documents/R Programming/specdata"

* ***list.files()*** 🡪 就可以把当前directory 中的所有文件都读了，（）里不用再加要读的directory的名字了。
* ***vector***[[2]](#footnote-2)***（mode= “logical”, length = 0）***vector()产生一个向量然后给予它长度和mode.这个Mode 是用来命名atomic mode or “list” or “expression” or (except for vector) “any”. The atomic modes are "logical", "integer", "numeric" (synonym "double"), "complex", "character" and "raw".
* Let's take the output of list.files() and store it:

files <- list.files("diet\_data")

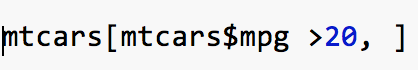
files

Knowing that 'files' is now a list of the contents of 'diet\_data' in alphabetical order, we can call a specific file by subsetting it:

files[1]

files[2]

files[3:5]

* 例子：数据名字是mtcars,现在要读取里面mpg这一列的所有数值大于20的数据
* 
* mtcars$mpg > 20 是一个logical的判断，会把所有MPG的数据都用True, false来判断一遍
* 如果我们想查看其中的两列

mtcars[mtcars$mpg > 20, c(“mpg”, “dimp”)]

* **Drop** 
  1. V[-1:-3] drop the first three
  2. V[-length(v)] drop the last element
* K

[Basic knowledge for R](http://blog.sciencenet.cn/blog-474707-421675.html)  
2011-3-10  
相关的函数记录与整理  
1、source("文件名.r")：调取主程序的文件，在程序结构复杂的时候很有用，可以将一部分复杂的运算主程序放入其中。  
2、install.packages("fields")：安装程序包  
3、library(fields)：导入程序包  
4、t(x)转置函数，对于csv中横排的转置很有用  
5、dev.off()：中断函数  
6、a <- as.character(b):因子型转化为字符型函数  
7、position <- regexpr('\_',a)：regexpr()函数对字符的定位很有用，返回值position为特定字符，如字符串a中’\_’的位置  
8、结合定位函数，对字符串如x345\_xbt,进行拆分，利用函数substring(要拆分的字符串，开始的字符位置，结束的字符位置)  
namecol1 <- substring(a, 2, position - 1)  
namecol2 <- substring(a, position + 1, nchar(a))  
结合regexpr()函数，这两个命令返回的值为，namecol1<-345；namecol2<-xbt;  
9、合并向量data.frame(vetor1, vetor2, vetor3)  
cbind(vetor1, vetor2, vetor3)  
10、取名字相同的交集unique（）函数  
例如对包含行名的向量R1、R2、 R3取名字相同的行，组成新的向量。  
nam1 <- rownames(R1)  
nam2 <- rownames(R2)  
nam3 <- rownames(R3)  
tnam <- unique(c(nam1,nam2,nam3))返回结果为只剩下名字相同的行的数值和rownames  
或者取一个向量中唯一一个值的数据，合并重复数据。unique(x, incomparables = FALSE, ...)在R中三个点…,表示可传递参数  
11、对程序包里面的具体的函数源代码，通过安装包后直接输入函数名回车，可以看到函数；注意找到对应的子函数。也可以在R镜像网页中的packages中，下载package的数据包，减压后，看文件夹得R函数中，这个包含程序注释，更好。  
12、t检验函数  
t.test(x, ...)  
## Default S3 method:  
t.test(x, y = NULL,alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95, ...)  
## S3 method for class 'formula'  
t.test(formula, data, subset, na.action, ...)  
13、对于一个向量x中选择某个符合条件的数值出来(大于某个数或者是某个条件limit)，直接利用表达式y<-x[,1][判断语句或者limit函数参数，在外部设定limit值]  
14、判断两个向量的交集部分选择%in%  
1:10 %in% c(1,3,5,9)  
15、帮助已经加载的程序包有哪些函数，可以用主界面的help>html帮助  
16、regression with damped exponential correlation回归的函数为rdec(),需要的函数包为RDEC  
17、相关常用的R运算   
计算  
log(x)   
log10(x)   
exp(x)   
sin(x)  
cos(x)   
tan(x)   
asin(x)  
acos(x)  
min(x)  
max(x)  
range(x)   
length(x)  
 统计检验  
mean(x)  
sd(x)  
var(x)  
median(x)  
quantile(x,p)  
cor(x,y)  
t.test()  
lm(y ~ x)  
wilcox.test()  
kruskal.test()  
 统计检验  
lm(y ~ f+x)  
lm(y ~ x1+x2+x3)  
bartlett.test  
binom.test  
fisher.test  
chisq.test  
glm(y ~ x1+x2+x3, binomial)  
friedman.test  
...     
    
18、ls() 列出工作空间中的对象；rm() 删除工作空间中的对象  
19、对象类型转换  
as.numeric() #转换为数值型  
as.logical()  
as.charactor()  
as.matrix()  
as.dataframe()

某人的R学习笔记

<http://www.cnblogs.com/holbrook/archive/2013/05/09/3068613.html>

某博客<http://blog.sina.com.cn/s/blog_8f5b2a2e0101fmiq.html>

1、一元线性回归

lm() #计算beta0,beta1参数

summary() # 提取lm()参数信息

anovn() #方差分析

predict() # 根据给出自变量预测因变量的值

例：

a=lm(y~1+x,data=…) #对x,y回归分析

summary(a)  # 提取回归分析结果

new=data.frame(x=0.16) # 当x=0.16时对y进行预测，必须为数据框格式

lm.new=predict(a,new,interval=”prediction”,level=0.95)#利用a得到的回归分析，对new进行预测

abline(a)#绘制回归直线

b=residuals(a) # 计算回归方程残差

plot(b) #绘制残差散点图

提取模型lm()信息的通用函数

anova(a)#计算方差分析表aov()

coefficients()# 提取模型系数

deviance()#计算残差平方和

formula()#提取模型公式

plot()#绘制模型诊断图

predict()#预测

print()# 显示

residuals()#计算残差

step()#做逐步回归

summary()#提取模型资料

2、多元回归分析

update()修正拟合模型

例：

fm5=lm(y~x1+x2+x3+x4+x5,data=…)

fm6=update(fm5, .~.+x6)

smf6=update(fm6,sqrt(.)~.)

lm(y~x1+I(x2^2))

add1()

drop1()

step()

3、逐步回归

step(object,scope,

direction=c(“both”,”backward”,”forward”)

trace=1,keep=Null,steps=1000,k=2,…)

4、回归诊断(1.误差项：独立性 等方差性 正态性

2.选择线性模型是否合适

3.是否存在异常样本

4.回归分析结果是否对某个样本依赖过重

5.自变量之间是否存在高度相关)

作图

残差

residuals()

Shapiro.test() #残差正态性检验

rstandard(model,infl=lm.influence(model,do.coef=FALSE)

sd=sqrt(deviance(model)/df.residual(model)),…) #标准化残差

rstudent()#学生化残差

回归值Y与残差的散点图

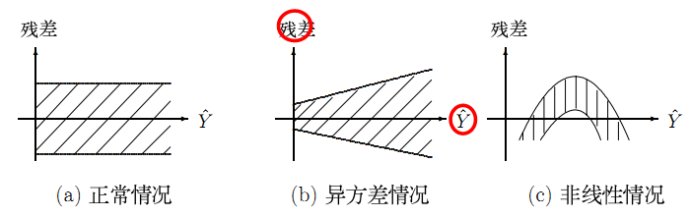
y.res=resid(lm.sol)

yfit=predict(lm.sol)

plot(y.res~y.fit)#残差图

y.rst=rstandard(lm.sol)

plot(y.rst~yfit)#标准化残差图

[](http://photo.blog.sina.com.cn/showpic.html#blogid=8f5b2a2e0101fmiq&url=http://album.sina.com.cn/pic/002CLCF0gy6FZoDXVxj4d)

以自变量为横坐标的残差图

y.res=resid(lm.sol)

plot(y.res~x1)

plot(y.res~x2)

强影响点（异常值）

hatvalues()

hat()

dffits()

cooks.distance()

covratio()

influence.measures(model)#回归诊断的总括

多重共线

kappa() #计算矩阵的条件数

eigen() # 计算矩阵特征值

5、广义线性模型

glm(formula, # 公式

family=c(binomial,Gaussian,gamma,inverse,poisson,quasi)# 分布族

data=…)# 数据

非线性回归

lm(y~1+x+I(x^2),data=…)

poly(x,degree=1,codfs=NULL)正交多项回归 degree多项式阶数

lm(y~1+poly(x,2))

nls()#求非线性最小二乘问题

nlm()#非线性模型参数估计

6、方差检验

单因素方差分析

aov(x~A)# x要分析变量A变量因子水平

均值的多重比较

pairwise.t.test(x,g) # x相应变量 g因子向量

P值修正

P.adjust(p,method=…)

方差齐性检验

  误差正态性检验

  shapiro.test(X[A==1])

  方差齐性检验

  bartlett.test()

通过秩统计量进行方差分析(分析各水平之间是否存在差异)

K-W秩和检验

kruskal.test()

Friedman秩和检验

friedman.test()

双因素方差分析

aov(Y~A+B+A:B)#对A B两个因素进行分析 并考虑了交互作用

介绍一下函数predict，用法如下：

predict(object, newdata, se.fit = FALSE, scale = NULL, df = Inf,

interval = c("none", "confidence", "prediction"),

level = 0.95, type = c("response", "terms"),

terms = NULL, na.action = na.pass,

pred.var = res.var/weights, weights = 1, ...)

说明一下，newdata的数据结构是一个数据框。

           这里还值得一提的参数时interval，他有三个选项：none代表不作区间预测，仅给出响应变量的估计；confidence是给出E（Y|X=x）的置信区间；prediction是给出真实Y的置信区间，运行代码你就会发现两者的差别。出于稳健性考虑，给出自变量取值时，预测Y值通常会采用prediction参数，但是对于X=x时Y的均值的预测就应该用confidence参数（因为回归模型是Y=βX+e，所以自变量相同，响应变量也未必一样）

1. R Programming Tutorial Lesson 2: Copying Data from Excel <https://www.youtube.com/watch?v=U69k3hjDlM0&list=PLvhG5FRoq78otMp0WbWuJsFOarEAchMek&index=2> [↑](#footnote-ref-1)
2. <http://www.inside-r.org/r-doc/base/as.vector> [↑](#footnote-ref-2)