

## 函数

```
install.packages() 安装包  
library() 查看已经安装的包  
library(help = "MASS") 看 MASS 包的 help  
  
setwd("C:/Users/Administrator/Desktop")
```

`d1=read.csv("prices.csv")` 读数据 (先把数据读到一个变量中再操作)

```
head(d1) 看头六行数据  
class(d1) 看数据表类型  
str(d1) 看数据表的结构, 如果是 data frame 的话, 看每列数据的类型  
d1$Date = as.Date(d1$Date) 比如: as.numeric() 把数据转换成数值类型,  
as.Date() 把数据转换成日期类型 这个操作的关键在于赋值与否, 不重新赋值根本  
不会变  
range() 看取值范围  
summary(d1[,c(2:5)]) 看一个 data frame 的属性, 最大最小中位数……  
apply(d2,2,mean) 2 指的是 by column, 1 是 by row, 最后是应用的函数, 此  
为按列求取中位数的一个 vector
```

```
set.seed() 设计随机数的种子  
sample(x, size) 生成 x 以下 size 个数字, x 也可使 vector
```

## 操作

`x = c(1,2,4,8)` `x=1:10` `x=seq(from=1,to=3,by=0.5)` 赋值

```
c0 = c(2,4,1,5,3,0,2,4,6,7,8,11,9,7,1)
```

```
rep(c(0,1,2),each=3) 000111222333  
rep(c(0,5,10),times=3) 051005100510
```

`mat0 = matrix(c0,nrow=3,ncol=5)` 创建矩阵

```
matrix(c0,nrow=3,ncol=5,byrow=T) 默认创建按列排序, byrow 按行  
dim, nrow, and ncol 看矩阵的行列大小  
length(d3) 查看 vector 的长度
```

```
c1 = c(2,4)
```

```
c2 = c(1,3)
```

`mat1 = cbind(c1,c2)` 将两个矩阵按列组合, `rbind()` 按行组合

矩阵相乘是 `%*%`, `*` 只是矩阵中对应数的乘积

`d2 = d1[,c(2,4,9)]` `d2 = d1[c(1:10),]` 任意选行选列 前行后列 **部分选取**

## 取

`n = nrow(d1)` 看d1一共有几列

`d1[(n-8):n,]` 选最后八行

可以发现中括号表示的是选取条件，比如 `d3=d2[d2>670]` 表示

`d2 = data.frame(d1$AAPL,d1$GS,d1$XOM)`

`d2 = subset(d1,select=c(AAPL,GS,XOM))` 两个都是按名字选取，这个是真身  
`subset(x, subset, select)` `select` 是判断式，逻辑量；`subset` 是要选取的列

`order()` **排序**，生成一个 vector，是被排序 vector 从小到大的序号

`d2 = d1[order(d1$Weight),]`

`index <- order(d1$Weight,d1$Width)` `weight`为主要，`width`为次要

`d2 = d1[index,]`

## 变量

Vector

Matrix

`array(0,dim=c(4,5,2))` 四行五列的两个 table(matrix)

**data frame** 每个 column 只能储存相同属性变量，不同 column 的变量类别可以不同

**list** a list may contain a vector of characters, a matrix, a dataframe, and, another list

`d` (probability or density function)

`dbino(x,n,p)` 贝努利试验，二项分布

`p` (distribution function)

`q` (quantile function)

`r` (function to generate observations randomly from the given random variable)

## 画图

`plot(d1$AAPL,type="l")` `l`代表折线图，默认为点图

`grid()` 画网格线

`library(xts)`

`d1x=xts(d1[, -1],order.by=d1$Date)` 按日期排序

`plot(d1x$AAPL)`

`curve(log(x),0.05,5)` 画函数，后两个表示 `x` 的范围

```
ex :
curve(dchisq(x,3),0,20,ylab="",main="Chi-squared pdfs")
curve(dchisq(x,13),col="red",add=T)
legend(15,0.2,c("df 3","df
13"),cex=0.6,lty=c(1,1),lwd=c(2.5,2.5),
col=c(1,"red"))
grid()
```

hist() 画柱状图

plot() 画箱图

```
h1 = hist(d1)
str(h1) 所谓改图利器 (breaks 每个柱子的左边界 mids counts density 这四个属性)
```

```
h3=hist(d1, breaks=seq(17,35,by=3))
```

```
hh <- hist(d1)
hh$counts = hh$counts/sum(hh$counts)
plot(hh) 改相对数值
plot(hh,xlim=c(15,35),ylim=c(0,0.2),ylab="relative
frequencies",main="") lim边界数值 lab标签 main主题
```

```
library(MASS)
str(geyser)
waiting = geyser$waiting
hist(waiting, freq = F, col="gray") # histogram
lines(density(waiting)) # kernel density estimate
grid()
```

```
d3 = data.frame(x=p1$x,y=p1$y)
aux = max(d3$y)
d3[d3$y==aux,] 提取某一行数据
```

Option	Description
<code>pch</code>	point character ( <code>pch=1, 2, ...</code> )
<code>lty</code>	line type ( <code>lty=1, 2, ...</code> )
<code>lwd</code>	line thickness ( <code>lwd= 1, 2,...</code> )
<code>col</code>	color ( <code>col="red", "blue",...</code> )
<code>xlim</code>	x-axis limits: <code>xlim=c(min,max)</code>
<code>ylim</code>	y-axis limits
<code>xlab</code>	x-axis label: <code>xlab="my label"</code>
<code>ylab</code>	y-axis label
<code>main</code>	main title
<code>sub</code>	sub title

## Curve

To plot smooth curves, use the `curve` command. The `_rst` argument must be an expression in terms of `x`:

```
curve(x^2, from = 0, to = 2)
```

```
curve(cos(x), from = 0, to = pi)
```

```
curve(cos(x), from = 0, to = pi, lty = 4, col = "red")
```

### 1.2 Plot

```
plot(Poverty ~ Unemp, data = States03, xlab = "Unemployment", ylab = "Poverty")
```

```
abline(3, 5) adds the straight line  $y = 3 + 5x$ 
```

```
abline(v = 2) adds the vertical line,  $x = 2$ 
```

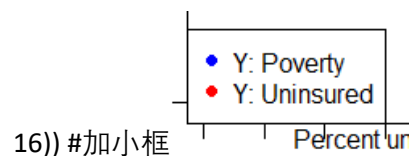
```
abline(h = 0) adds the horizontal line,  $y = 0$ 
```

```
text(30, 18, "mean unemployment rate") #text at (30, 18)
```

```
title("Data from 2003")
```

```
mtext("Percent uninsured", side = 4) #在图表每一边上加文本
```

```
legend("bottomleft", legend = c("Y: Poverty", "Y: Uninsured"), col = c("blue", "red"), pch = c(16,
```



## Probability Distributions

Distributions	root
beta	beta
Cauchy	cauchy
chi-square	chisq
exponential	exp
F	f
gamma	gamma
normal	norm
student's t	t
uniform	unif
Weibull	weibull

**droot** returns the density, **proot** a cumulative probability, **qroot** a quantile (分位数, 比如 `qnorm(0.5)` = 0, 返回x) , **rroot** a random number.

`pnorm(1.25)` #累计概率  $P(X \leq 1.25) = \Phi(1.25)$ , that is, the amount of area under the standard normal density curve to the left of  $x = 1.25$

`pnorm(2.8, 2, 3)` #  $X \sim N(\mu = 2, \sigma = 3)$

`curve(dnorm(x), from = -3, to = 3)` #probability distribution function

`curve(pnorm(x), from = -3, to = 3)` #cumulative distribution function

## Graph

Option	Description
<code>pch</code>	point character ( <code>pch=1, 2, ...</code> )
<code>lty</code>	line type ( <code>lty=1, 2, ...</code> )
<code>lwd</code>	line thickness ( <code>lwd= 1, 2,...</code> )
<code>col</code>	color ( <code>col="red", "blue",...</code> )
<code>xlim</code>	x-axis limits: <code>xlim=c(min,max)</code>
<code>ylim</code>	y-axis limits
<code>xlab</code>	x-axis label: <code>xlab="my label"</code>
<code>ylab</code>	y-axis label
<code>main</code>	main title
<code>sub</code>	sub title

## Curve

To plot smooth curves, use the `curve` command. The `_rst` argument must be an expression in terms of `x`:

```
curve(x^2, from = 0, to = 2)
```

```
curve(cos(x), from = 0, to = pi)
```

```
curve(cos(x), from = 0, to = pi, lty = 4, col = "red")
```

### 1.2 Plot

```
plot(Poverty ~ Unemp, data = States03, xlab = "Unemployment", ylab = "Poverty")
```

```
abline(3, 5) adds the straight line  $y = 3 + 5x$ 
```

```
abline(v = 2) adds the vertical line,  $x = 2$ 
```

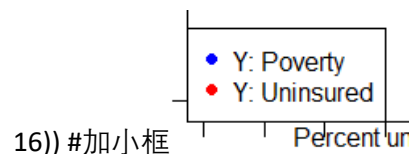
```
abline(h = 0) adds the horizontal line,  $y = 0$ 
```

```
text(30, 18, "mean unemployment rate") #text at (30, 18)
```

```
title("Data from 2003")
```

```
mtext("Percent uninsured", side = 4) #在图表每一边上加文本
```

```
legend("bottomleft", legend = c("Y: Poverty", "Y: Uninsured"), col = c("blue", "red"), pch = c(16,
```



## Probability Distributions

Distributions	root
beta	beta
Cauchy	cauchy
chi-square	chisq
exponential	exp
F	f
gamma	gamma
normal	norm
student's t	t
uniform	unif
Weibull	weibull

**droot** returns the density, **proot** a cumulative probability, **qroot** a quantile (分位数, 比如 `qnorm(0.5)` = 0, 返回x) , **rroot** a random number.

`pnorm(1.25)` #累计概率  $P(X \leq 1.25) = \Phi(1.25)$ , that is, the amount of area under the standard normal density curve to the left of  $x = 1.25$

`pnorm(2.8, 2, 3)` #  $X \sim N(\mu = 2, \sigma = 3)$

`curve(dnorm(x), from = -3, to = 3)` #probability distribution function

`curve(pnorm(x), from = -3, to = 3)` #cumulative distribution function