



R语言入门介绍

李伟

Welcome to R!



左图是Ross Ihaka，右图是Robert Gentleman, 来自
The University of Auckland.

他们是R语言的发明人！

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

基本知识

R语言官方网站: www.r-project.org(下载R软件，更新程序，下载程序包)

R语言对字母大小写敏感，也就是说x与X是不一样的；

?solve查询函数solve的帮助文档；

example(solve)运行帮助文档中的示例；

基本知识

: 添加注释 ;

新起一行 , 代表新的命令 ;

使用上下箭头来查看当前工作空间的命令 , ctrl + L 清屏 ;

Packages->Install packages在线安装程序包 ;

The background of the slide features a soft-focus photograph of tall, thin grasses or reeds in the foreground, with a warm, golden sunset or sunrise sky in the background. The colors are muted and artistic, with a gradient from light blue at the top to warm orange and yellow near the horizon.

基本知识

亲自动手：

安装R语言及软件包MASS。

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

数据结构

向量：数据类型相同；

矩阵：二维向量，所有数据的类型相同；

数据框：长度相同，类型不同的二维数据；

函数。

数据结构-向量

向量(vector):

```
x <- c(1, 4, 6)
```

```
#建立一个向量，包括1,4,6
```

数据结构-向量

常用的构建向量方式：

1:20或者30:1

seq(from = , to = , by =)

rep(x = , times =)

数据结构-向量

向量相关的命令：

**max(x), min(x), length(x) ,
range(x) = c(min(x), max(x))**

sum(x), prod(x)

mean(x), var(x), sd(x)

sort(x)

数据结构-向量

亲自动手：

构建向量

1. $(1, 5, 12, 19)$

2. $(1, 3, 5, 7, 11, \dots, 997, 999)$

3. $(0, 0, 0, 0, \dots, 0)$ 共有 $1e4$ 个零。

数据结构-矩阵

matrix(data = , ncol = , nrow = , byrow = FALSE, ...)

数据结构-矩阵

例：

```
mdat <- matrix(data = c(1, 2, 3, 11, 12, 18),
```

```
nrow = 2, ncol = 3, byrow = TRUE)
```

```
mdat
```

```
mdat[1, 2]#得到第一行第二列的元素；
```

```
mdat[1, ]#得到第一行的所有元素。
```

数据结构-矩阵

亲自动手：

构建矩阵：

(1) 第一列是(1, 2, 3), 第二列是(6, 5, 4)

(2) 第一行是(1, 2, 3), 第二行是(6, 5, 4)

要求matrix()中的data是相同的。

数据结构-数据框

data.frame(...)

每列的长度要求相等。

数据结构-数据框

```
data1 <-  
data.frame(  
  name = c("Harden", "Wall", "Westbrook"),  
  assists = c(11.2, 10.7, 10.4),  
  team = c("HOU", "WSH", "OKC"))
```


数据结构-数据框

#得到数据框的指定列

data1\$name

data1\$assists

数据结构-数据框

亲自动手：

```
patientID <- c(1, 2, 3, 4)
```

```
age <- c(25, 34, 28, 52)
```

```
Diabetes <-
```

```
c(" Type1 ", " Type2 ", " Type1 ", "Type1")
```

将以上关于糖尿病人的数据变成数据框.

数据结构-函数

**函数名 <- function(参数1 , 参数2 , ...){
表达式 }**

例：

```
mysum <- function(y1, y2) {  
  return(y1 + y2)  
}#计算两数之和
```

数据结构-函数

亲自动手：

写一个函数，其为标准正态分布的概率密度函数。

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

基本图形的绘制-点线图

#基本命令

```
plot(pressure$temperature, pressure$pressure,  
      type = "l")
```

基本图形的绘制-点线图

#图上加点用points

points(pressure\$temperature, pressure\$pressure)

#图上加线用lines

**lines(pressure\$temperature, pressure\$pressure / 2,
col = "red")**

基本图形的绘制-直方图

hist:

hist(mtcars\$mpg)

基本图形的绘制

亲自动手：

下面的数据是不同剂量下药物A与药物B的响应情况。

```
dose <- c (20, 30, 40, 45, 60)
```

```
drugA <- c(16, 20, 27, 40, 60)
```

```
drugB <- c(15, 18, 25, 31, 40)
```

画出不同药物在相同剂量下的对比图。

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

概率分布

Distribution Name	Name in R
beta	beta
binomial	binom
Cauchy	cauchy
chi-squared	chisq
exponential	exp
F	f
gamma	gamma

概率分布

Distribution Name	Name in R
geometric	geom
hypergeometric	hyper
log normal	lnorm
logistic	logis
negative binomial	nbinom
normal	norm
Poisson	pois

概率分布

Distribution Name	Name in R
signed rank	signrank
Student's t	t
uniform	unif
Weibull	weibull
Wilcoxon	wilcox

概率分布

在name in R的前面 ,
加上"d"代表该分布的概率密度函数 ;
加上"p"代表该分布的累积分布函数 ;
加上"q"代表该分布的分位数函数 ;
(代表 $F(x) > q$ 的最小 x)
加上"r"代表产生该分布的随机数。

概率分布的例子

```
den1 <- dnorm(x = seq(from = -4, to = 4, by = 0.001),  
              mean = 0, sd = 1)
```

```
plot(den1)#产生正态分布的概率密度曲线
```

```
pro1 <- pnorm(q = seq(from = -4, to = 4, by = 0.001),  
              mean = 0, sd = 1)
```

```
plot(pro1)#产生正态分布的累积分布曲线
```

概率分布的例子

```
quant1 <- qnorm(p = seq(from = 0, to = 1, by = 0.001),  
               mean = 0, sd = 1)
```

```
plot(quant1)#产生正态分布的分位数曲线
```

```
sample1 <- rnorm(n = 10000, mean = 0, sd = 1)
```

```
#产生正态分布随机数
```

```
hist(sample1)
```


目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

读文件

read.table(...)

读Excel数据：

1.选中你读的数据，并且复制；

2.在命令行写上

```
data <- read.table(file = "clipboard" ,  
                    header = TRUE)
```

(如果第一行是列名的话)

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

循环语句

```
for(i in 1:20){  
  语句B  
}
```

循环语句

例：

```
mark <- 0
```

```
for(i in 1:20){
```

```
  mark[i] <- 10 + 20 * sqrt(i)
```

```
}
```

循环语句

其它：while,repeat,break,...

注：在R语言中用for的机会比C中会少很多。

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

代码编写习惯

- 文件名有意义，且以 “.R” 结尾；
- 变量与函数命名建议全部使用小写单词，并且不同的单词用 “_” 分开；
- 在运算符与赋值符的两侧加上空格，比如 =, +, -, <-, *, / 等；唯一的例外是 “:”

代码编写习惯

- ,的后面加上空格 ;

- 大段的注释使用

```
#####
```

```
#download data#
```

```
#####
```

- 小段的注释使用

```
#placing before or after a line of code
```

目录

1. 基本知识
2. 数据结构
3. 基本图形的绘制
4. 常用分布的相关函数
5. 读文件
6. R语言中的循环语句
7. 代码编写习惯
8. 参考文献

参考文献

- 1. An Introduction to R(R语言自带文档)**
- 2. R Graphics Cookbook BY Winston Chang**
- 3. Data Wrangling with R BY Bradley Boehmke**

作业

1. 产生1个长度为100的向量，
并且将其变成一个5行20列的矩阵；
2. 写一个函数，其为 $[0,1]$ 上的均匀分布的概率密度函数，
只可以使用基本的数学运算，不能调用R的函数；

作业

3. 产生服从均值为1，方差为4的正态分布的随机数1000个，并且绘制相应的直方图；
4. 利用`read.table(...)`去读一个EXCEL文件中的数据。