

7.R语言基础

戴俊毅

研究员/长聘副教授



简介

闭源

- ❧ R是在贝尔实验室开发的统计软件S的基础上发展而来的开源软件。
- ❧ R一般仍被视作是一款统计软件，但随着开源社区的不断扩大，R的功能一直在不断增强。
- ❧ R的优点在于新的工具包不断涌现，使得它可以很快实现最新出现的数据分析和统计技术，包括贝叶斯数据分析技术。
- ❧ R的缺点在于它的开源性使得其支持文档和错误反馈不够专业。此外，R主要是由统计学家而非计算机科学家开发的，所以其语法结构不尽合理，学习起来比较费力。



安装

✎ R 安装

<https://www.r-project.org/>

✎ R studio 安装

<https://www.rstudio.com/>



帮助和查询功能

🔗 help(FUN)

🔗 ?FUN

?? 可以不是某个函数名

🔗 ??FUN





注释功能

☞ 在R里面，可以使用#字符来对代码进行注释，所有#之后同一行的文本都不会被运行



基本运算

☞ 四则运算: $3 + 2$, $3 - 2$, $3 * 2$, $3 / 2$

☞ 幂运算: 3^2 , $3^{(1/2)}$, $\exp(2)$

☞ 对数运算: $\log(2)$, $\log_2(2)$, $\log_{10}(2)$



等差数列

↪ `seq(start,end,stepsize)`, 例如`seq(1,5,2)`

↪ 步长为1时步长参数可忽略

↪ 当步长为1时，也可使用`start:end`方式构建等差数列

行矢量和列矢量

- ↪ 矢量: `c(1,2,3,4)`
- ↪ 行矢量: `t(c(1,2,3,4))` 或者 `matrix(1:4,nrow=1,ncol=4)`
- ↪ 列矢量: `t(t(c(1,2,3,4)))` 或者 `matrix(1:4,nrow=4,ncol=1)`
- ↪ 矢量或者高维数组的元素可以是任意同一类型的数据, 比如数字型或者逻辑型



矩阵

↻ `A = matrix(1:4, nrow = 2, ncol = 2, byrow = TRUE)`





矢量和矩阵的大小

↻ 矢量: $\text{length}(v)$

↻ 矩阵: $\text{dim}(M)$





矢量和矩阵索引

↪ $v[\text{indices}]$

↪ $M[\text{indices}, \text{indices}]$

↪ $M[:, \text{indices}]$

↪ $M[\text{indices}, :]$



数组

除了1维的矢量和2维的矩阵外，在R里也可以定义更高维的数据结构，称为数组(array)

数组的定义方式和矩阵类似，例如

```
A = array(1:24, dim = c(3,4,2))
```

数组的索引方式也和矢量还有矩阵类似，例如A[1:3, 1:2,]

矢量和矩阵可以看作是1维和2维的数组

列表

- ❧ 列表(list)是对于矢量的拓展，矢量中的元素必须是同类型的，而列表不需要
- ❧ `MyList = list(a = 3, b = "b", c = matrix(1:4, nrow = 2, ncol = 2))`
- ❧ 列表索引方式为：`MyList$a`, `MyList[1]`, 或者 `MyList[[1]]`

数据框架

- ❧ 数据框架(data frame)是对矩阵的拓展，它的每一列内部应该包含相同类型的元素，但是各列之间的元素类型可以不同
- ❧ SPSS或者JASP里的数据就可以看作是数据框架类型的
- ❧ 例如，`d = data.frame(Integers = 1:3, NumberNames = c("one", "two", "three"))`
- ❧ 数据框架既可以像矩阵一样被索引，也可以像列表一样被索引，例如`d[1]`索引第一列，`d[[1]]`或者`d[,1]`索引第一列元素，`d$NumberNames`索引第二列元素



因素型变量

↪ `x = c("high", "medium", "low", "high", "medium")`

↪ `xf = factor(x)`

↪ `as.numeric(xf)`

按元素进行的矩阵运算

↻ $A + B, A - B, A * B, A \wedge B$

- 按元素进行的矩阵运算要求矩阵大小一致
- 若A或者B中有一个是标量，则系统自动将之转换为和另一个矩阵大小相同，且所有元素都等于该标量的矩阵
- 由于矢量是矩阵的特例，所以以上规则也适用于矢量运算

标准矩阵乘法

↪ $A \%* \% B$

- 标准矩阵乘法要求A矩阵的列数=B矩阵的行数
- 设 $C = A*B$, A为 $m*n$ 矩阵, B为 $n*p$ 矩阵, 则C为 $m*p$ 矩阵, 且

$$C_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$$



逻辑运算

⌘ 逻辑与: $\&\&$ (标量), $\&$ (矢量)

⌘ 逻辑或: $\|$ (标量), $|$ (矢量)

⌘ 逻辑非: $!$





关系运算

☞ 相等==

☞ 不相等!=

☞ 小于<

☞ 大于>



概率分布函数

❧ 随机数函数: `rnorm(n, mu, sigma)`

❧ 概率密度函数: `dnorm(x, mu, sigma)`

❧ 累积分布函数: `pnorm(q, mu, sigma)`

❧ 逆累积分布函数: `qnorm(p, mu, sigma)`

❧ 其他分布:

`binom` 二项分布 `unif` 均匀分布 `exp` 指数分布 `gamma` 伽玛分布



复制函数

↻ `rep(x, times, each, length)`, 例如

`rep(1:3, times = 3): 1 2 3 1 2 3 1 2 3`

`rep(1:3, each = 3): 1 1 1 2 2 2 3 3 3`

`ABC = "ABC"`

`rep(ABC, each = 2, length = 8): "A" "A" "B" "B" "C" "C" "A" "A"`



实用函数

↪ min()

↪ max()

↪ summary()

↪ table()

↪ apply()



绘图函数

☞ windows()

☞ plot()

☞ hist()

☞ par()

☞ abline()

☞ points()

A decorative image featuring a basket of white flowers, likely tulips, with yellow centers. The basket is made of light-colored wicker and sits on a wooden surface. The background is a soft, out-of-focus white, creating a clean and elegant aesthetic.

判断结构

```
if (cond){  
    statements  
} else  
{  
    statements  
}
```



循环结构

```
for (i in 1:10){  
    statements  
}
```



A decorative image featuring a basket of white flowers, possibly tulips, with yellow centers. The basket is made of light-colored wicker and sits on a wooden surface. The background is a soft, out-of-focus light green and white.

数据读写函数

❧ 读文件: `read.table`, `read.csv`

❧ 写文件: `write.table`, `write.csv`



定义函数

```
↻ myfun = function( $x_1, \dots, x_M$ ) {  
    expr  
    return(value)  
}
```

↻ myfun: 函数名

↻ x_1, \dots, x_M : 函数输入

↻ 函数输出由value给定

↻ 函数输出也可以通过将value作为expr最后一行给定




A decorative image of white flowers in a basket, serving as a background for the slide. The flowers are white with yellow centers, and the basket is made of light-colored wicker. The background is a soft, out-of-focus white.

定义函数示例

```
stat = function(x){  
  n = length(x)  
  m = sum(x)/n  
  s = sqrt(sum((x-m)^2/n))  
  return(c(m,s))  
}
```




调用函数

- ❧ 函数定义无需先保存为文件
 - ❧ 可以在代码文件或者命令行中调用
 - ❧ 例如`stat(1:10)`或者`a = stat(1:10)`
- 

援引文件内容

- ❧ 当需要使用其他文件中的代码时，需要先使用source函数援引该文件。
- ❧ 例如，如果我们在一个名为stat.R的文件中定义了以上的stat函数，那么我们可以在Rstudio的命令窗口里，通过写入source('stat.R')，援引该函数定义。
- ❧ 随后，我们就可以在命令窗口里，使用该函数了



代码示例

- ✧ ExamplesOfR.R
- ✧ BernBetaExample.R
- ✧ BernGridExample.R
- ✧ HtWtDataGenerator.R
- ✧ SimpleGraph.R