

R 数据处理

谢益辉

中国人民大学统计学院



生物多样性数据分析和生物统计培训班
2008年11月17至20日于中国科学院植物研究所

1 数据输入输出

- 数据输入
- 数据输出
- 数据库操作

2 初步统计分析

- 常见描述统计
- 分类描述统计

3 数据模拟

- 随机数
- 模拟过程

目录

1 数据输入输出

- 数据输入
- 数据输出
- 数据库操作

2 初步统计分析

- 常见描述统计
- 分类描述统计

3 数据模拟

- 随机数
- 模拟过程

简易扫描数据

- `scan()`

从文件读入

- `read.table()`
- `read.csv()`
- `read.delim()`
- `readLines()`



从连接读入

- 网络数据
- 压缩包数据
- Socket 数据



输出文本文件

- `writeLines()`
- `sink()`



生成数据表

- `write.table()`
- `write.csv()`



ODBC 简介

- RODB 包



目录

1 数据输入输出

- 数据输入
- 数据输出
- 数据库操作

2 初步统计分析

- 常见描述统计
- 分类描述统计

3 数据模拟

- 随机数
- 模拟过程

一元描述统计

- 均值、方差
- 频数
- 极值
- 分位数



多元描述统计

- 相关

分类描述统计

- *apply() 系列
- by()、aggregate()

目录

1 数据输入输出

- 数据输入
- 数据输出
- 数据库操作

2 初步统计分析

- 常见描述统计
- 分类描述统计

3 数据模拟

- 随机数
- 模拟过程

生成随机数

r*() 系列函数

```
> rnorm(30)
```

```
[1] -0.899992652 -1.053719503  0.570052714 -0.202587846  
[5] -1.860768444  0.160340966  0.838635607  1.271956644  
[9] -0.800247598 -1.686782821  0.077350413 -1.182930209  
[13] 0.165737271 -0.702731065 -0.152410304  1.601906270  
[17] 1.505812706  0.988383436  0.276114242 -0.549306324  
[21] 0.059629062 -0.001818236 -0.356716750 -0.804922784  
[25] 1.196394327 -0.436788143 -0.535638067  0.886055724  
[29] 1.469004038 -0.706854248
```

硬币示例

真假硬币问题

(A)

```
11110000000000100000101100000100000101001111001100
01111110010110110101101001111001100011011101100000
10001001111110100100001011001011101101110001010010
01100111111100011100101000101001110011100010100111
```

(B)

```
01110010010010100010011110010100010011010111001110
01111011010111101101001000111001101011010101101001
00101001110110100100001110101101101001110101100110
01110011110110001110011010111001110011110010100111
```