

生态统计学

沈国春、李勤

2025-08-08

Contents

前言	5
1 R 语言入门准备	7
1.1 学习目标	7
1.2 为什么生态学专业需要学习 R?	7
1.3 R 语言简介与趣味应用	8
1.4 课前准备检查清单	9
2 森林调查数据计算	11
2.1 生态学背景	11
2.2 演示数据	11
2.3 课堂演示过程	11
2.4 R 语言知识点详解	12
2.5 课后练习	12

前言

这是一本关于生态统计学的教材，旨在为生态学者提供实用的数据分析方法。本书结合 R 语言，介绍生态学研究常用的统计技术。

本书特点：

- 面向生态学研究实际问题
- 基于 R 语言实现
- 包含大量生态数据案例
- 循序渐进的教学安排

本书使用以下 R 包：

```
install.packages(c(
  "tidyverse", "vegan", "lme4", "ggplot2",
  "bookdown", "knitr", "rmarkdown"
))
```

0.0.1 课程在线资源

课程简明手册

- 网页版 <https://guochunshen.github.io/ecological-statistics>
- PDF 版 <https://gitee.com/gcshen/ecological-statistics/blob/master/docs/ecological-statistics.pdf>

Chapter 1

R 语言入门准备

1.1 学习目标

- 理解为什么生态学专业需要学习 R 语言
- 掌握 R 和 RStudio 的安装和配置
- 建立良好的项目文件组织习惯
- 了解 R 包管理的基本方法

1.2 为什么生态学专业需要学习 R?

1.2.1 数据分析能力的必要性

现代生态学研究产生海量数据：野外调查数据、实验室测量数据、遥感影像数据、基因序列数据等。传统的 Excel 已无法满足复杂的统计分析需求，而 R 语言提供了完整的数据科学工具链。

```
# 示例：读取大型 CSV 文件
data <- read.csv("large_ecological_data.csv", nrows=100000)
```

1.2.2 可重现研究的科学要求

- **重现性危机**：越来越多的科学研究无法被重现，影响科学可信度
- **R 脚本的优势**：每一步分析都有记录，任何人都可以重现你的分析过程

```
# 示例：可重现的分析流程
# 1. 数据导入
data <- read.csv("field_data.csv")

# 2. 数据清洗
clean_data <- na.omit(data)

# 3. 统计分析
result <- lm(response ~ predictor, data=clean_data)
```

1.2.3 环境设置和配置

1.2.3.1 R 和 RStudio 安装

```
# 检查 R 版本
R.version.string

# 检查已安装的包
installed.packages()
```

1.2.3.2 项目文件夹结构

```
# 创建项目目录结构示例
dir.create("生态学 R 语言课程")
dir.create("生态学 R 语言课程/数据")
dir.create("生态学 R 语言课程/脚本")
dir.create("生态学 R 语言课程/结果")
```

1.2.3.3 包管理

```
# 安装生态学常用包
install.packages(c("vegan", "lme4", "ggplot2", "dplyr"))

# 加载包
library(vegan)
library(ggplot2)
```

1.3 R 语言简介与趣味应用

1.3.1 什么是 R 语言?

- 统计计算语言：专门为统计分析和数据可视化设计的编程语言
- 开源免费：由全球统计学家共同维护发展
- 扩展性强：超过 18,000 个专业扩展包

1.3.2 趣味应用示例

1.3.2.1 生态数据动态可视化

```
# 安装必要包
install.packages(c("ggplot2", "gganimate", "gapminder"))

# 绘制动态变化图
library(ggplot2)
library(gganimate)
ggplot(gapminder, aes(gdpPercap, lifeExp, size = pop, color = continent)) +
  geom_point() +
  scale_size(range = c(2, 12)) +
  scale_x_log10() +
  labs(title = '年份: {frame_time}', x = '人均 GDP', y = '预期寿命') +
  transition_time(year) +
  ease_aes('linear')
```


1.3.2.2 生成音乐

```
install.packages("audio")  
library(audio)  
play(sin(1:10000/10))
```

1.3.3 如何获取帮助?

```
# 查看函数帮助文档  
?plot  
help("plot")  
  
# 搜索帮助文档  
??"regression"  
  
# 在线资源:  
# - R 官方文档: https://cran.r-project.org/manuals.html  
# - RStudio 社区: https://community.rstudio.com
```

1.4 课前准备检查清单

- ☐ R 软件安装成功
- ☐ RStudio 安装成功

- ☐ 能够运行简单的 R 代码
- ☐ 创建了课程文件夹结构
- ☐ 理解了学习 R 语言的重要性
- ☐ 尝试运行了趣味示例代码

Chapter 2

森林调查数据计算

2.1 生态学背景

在野外森林调查中，测量树木胸径（DBH，距地面 1.3 米处的直径）是评估森林生长状况的基本方法。我们需要计算样地内树木的平均胸径来了解林分特征。

2.2 演示数据

```
# 某样地内 10 棵马尾松的胸径测量值（单位：厘米）
tree_dbh <- c(15.2, 18.7, 22.1, 19.5, 16.8, 20.3, 17.9, 21.4, 19.2, 18.6)
```

2.3 课堂演示过程

```
# 1. 创建胸径数据向量
tree_dbh <- c(15.2, 18.7, 22.1, 19.5, 16.8, 20.3, 17.9, 21.4, 19.2, 18.6)

# 2. 查看数据
tree_dbh
```

```
## [1] 15.2 18.7 22.1 19.5 16.8 20.3 17.9 21.4 19.2 18.6
```

```
length(tree_dbh) # 查看测量了多少棵树
```

```
## [1] 10
```

```
# 3. 计算平均胸径
mean_dbh <- mean(tree_dbh)
mean_dbh
```

```
## [1] 18.97
```

```
# 4. 计算总树数
tree_count <- length(tree_dbh)
tree_count
```

```
## [1] 10
```

```
# 5. 简单的数学运算
max_dbh <- max(tree_dbh) # 最大胸径
min_dbh <- min(tree_dbh) # 最小胸径
total_dbh <- sum(tree_dbh) # 胸径总和
```

2.4 R 语言知识点详解

2.4.1 向量创建函数 c()

- 是什么: c() 是 combine 的缩写, 用于将多个值组合成一个向量
- 语法: c(值 1, 值 2, 值 3, ...)
- 重要特点:
 - 向量中的所有元素必须是同一类型 (数值、字符或逻辑)
 - 如果混合不同类型, R 会自动转换为最通用的类型
 - 向量是 R 中最基本的数据结构

2.4.2 基本统计函数

2.4.2.1 mean() - 计算平均值

```
# 示例
mean(c(1, 2, 3, 4, 5))
```

```
## [1] 3
```

2.4.2.2 length() - 计算向量长度

```
# 示例
length(c(1, 2, 3))
```

```
## [1] 3
```

2.5 课后练习

题目: 某湿地调查中测量了 8 棵柳树的树高 (单位: 米): `tree_height <- c(4.2, 5.1, 3.8, 4.7, 5.3, 4.9, 4.1, 4.6)`

请完成: 1. 创建树高向量并查看数据 2. 计算平均树高 3. 找出最高的树有多高 4. 统计总共测量了多少棵树 5. 计算所有树的总高度 6. 计算树高的标准差 (提示: 使用 `sd()` 函数)

```
# 在此处编写你的练习代码
```