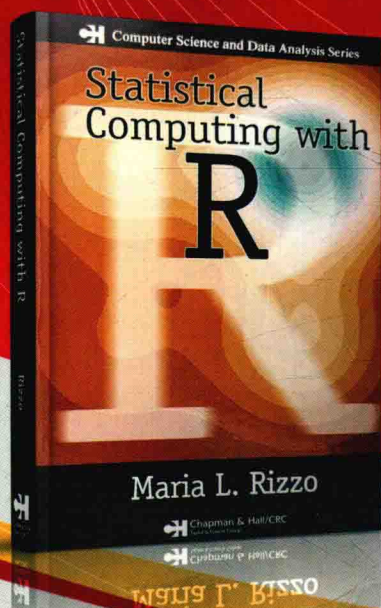


国外实用统计丛书

统计计算 使用R

Statistical Computing with R

[美] 玛利亚 L. 里佐 (Maria L. Rizzo) 著
胡锐 李义 译



 **CRC Press**
Taylor & Francis Group

 **机械工业出版社**
CHINA MACHINE PRESS

国外实用统计丛书

统计计算 使用 R

Statistical Computing with R

[美] 玛利亚 L. 里佐 (Maria L. Rizzo) 著

胡 锐 李 义 译



机械工业出版社

本书是统计计算或者计算统计学教材。书中包含了计算统计学的传统核心问题：概率分布模拟随机变量、蒙特卡罗积分和方差缩减法、蒙特卡罗法和马尔可夫链蒙特卡罗方法、自助法和水手刀法、密度估计和多元数据可视化等内容。

本书包含大量实例和练习，所有实例中的代码都可以在网站上下载。实现实例所使用的数据也都是 R 中的公开数据或者模拟数据。

本书既可作为高年级本科生和研究生的教材，也可作为相关科研人员 and 统计爱好者的参考书。

Statistical Computing with R by Maria L. Rizzo/ISBN: 9781584885450

Copyright © 2008 by CRC Press.

CRC Press is an imprint of Taylor & Francis Group, an Informa business

Authorized translation from English language edition published by CRC Press, part of Taylor & Francis Group LLC; All rights reserved.

本书原版由 Taylor & Francis 出版集团旗下，CRC 出版公司出版，并经其授权翻译出版，版权所有，侵权必究。

China Machine Press is authorized to publish and distribute exclusively the Chinese (Simplified Characters) language edition. This edition is authorized for sale throughout Mainland of China. No part of the publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

本书中文简体翻译版授权由机械工业出版社独家出版并限在中国大陆地区销售。未经出版者书面许可，不得以任何方式复制或发行本书的任何部分。

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal.

本书封面贴有 Taylor & Francis 公司防伪标签，无标签者不得销售。

北京市版权局著作权合同登记 图字：01-2012-5198 号。

图书在版编目 (CIP) 数据

统计计算：使用 R/(美) 玛利亚 L. 里佐 (Maria L. Rizzo) 著；胡锐，李义译。—北京：机械工业出版社，2016. 12

(国外实用统计丛书)

书名原文：Statistical Computing with R (Chapman Hall/CRC The R Series)
ISBN 978-7-111-55362-5

I. ①统… II. ①玛…②胡…③李… III. ①概率统计算法—高等学校—教材 IV. ①O242. 28

中国版本图书馆 CIP 数据核字 (2016) 第 273011 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑：韩效杰 责任编辑：韩效杰

责任校对：王延 封面设计：张静

责任印制：孙炜

保定市中国画美凯印刷有限公司印刷

2019 年 2 月第 1 版第 1 次印刷

184mm × 260mm · 24 印张 · 580 千字

标准书号：ISBN 978-7-111-55362-5

定价：79.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

网络服务

服务咨询热线：010-88379833

机工官网：www.cmpbook.com

读者购书热线：010-88379649

机工官博：weibo.com/cmp1952

教育服务网：www.cmpedu.com

封面无防伪标均为盗版

金书网：www.golden-book.com

译者序

本书不但是美国统计计算方面的畅销教科书，而且在欧洲和澳大利亚的销量也很好，同时也已出版日文版。本书的最大特点是，不但给出了关于统计计算算法的简单解释，而且还通过相应简单且高效的R程序去实现。阅读本书时强烈建议读者运行相应的R程序，读者可以通过微调一些参数的大小，帮助自己更好地理解书中的概念和算法。把书中的程序稍加改变，便可将其应用在自己的学习、工作和科研中。其他关于统计计算的书籍，虽然对统计计算的概念和算法也有很好的论述，但是往往缺少与之相对应的程序去实现书中的算法。对于第一次学习统计计算的读者，如果只有关于算法的论述，往往只能获得对算法的模糊认识。本书把讲解和程序相结合，这也正是其在美国成为畅销教科书的一个重要原因。本书不但对学习统计计算很重要，而且对学习其他统计课程有很高的参考价值。

前 言

本书是对统计计算和计算统计学的介绍。计算统计学是统计研究和应用中发展非常迅速的一个领域。它包含了统计学中需要进行大量计算的方法，比如蒙特卡罗方法、自助法、马尔可夫链蒙特卡罗方法、密度估计、非参数回归、分类和聚类，以及多元数据可视化等。Gentle^[113]和Wegman^[295]将计算统计学称为统计学中的计算密集型方法。统计计算主要专注于统计量的数值算法，至少传统上是这样的（参见Thisted^[269]）。通常一本书的书名只含有两个名词中的一个，比如Givens和Hoeting的《计算统计学》(Computational Statistics^[121])，该书包含了最优化、数值积分、密度估计和光滑，以及计算统计学中的蒙特卡罗方法和马尔可夫链蒙特卡罗方法中经典的统计计算问题。我们将本书的书名定为《统计计算 使用R》，它既是指统计计算，也是指计算统计学。而且虽然没有在书名中指出，但是本书着重强调了蒙特卡罗方法和重抽样。

R是一个基于S语言的统计学计算环境。相关软件在自由软件基金会(Free Software Foundation)的GNU通用公共许可协议(GNU General Public License)条款下是免费的。它在很多操作系统中都可以使用，比如Linux，Windows，以及Mac OS。具体说明可以参见<http://www.r-project.org/>。本书中所有的例子都是在R中实现的。

本书面向的读者是研究生，以及学习了微积分、线性代数、概率论和数理统计课程的高年级本科生。本书可以作为计算统计学的入门教材，也可以用来自学。此外，由于相关内容的计算特性，本书也是一本很好的R语言使用指南，书中通过大量的例子在实际计算问题的背景下阐明编程概念。本书并不要求读者熟悉任何编程语言。

相对于理论，本书更注重实现过程，但是会解释清楚与数学思想以及理论基础之间的联系。第1章对计算统计学做了一个综述，对R统计计算环境做了简单的介绍。第2章对概率和经典统计推断中的基本概念做了总结和回顾。余下部分每一章都涵盖了一个计算统计学的主题。

本书的主题包含了计算统计学的传统核心内容：由概率分布模拟随机变量、蒙特卡罗积分和方差缩减方法、蒙特卡罗和马尔可夫链蒙特卡罗方法、自助法和水手刀法、密度估计以及多元数据可视化。尽管R提供了常见概率分布的随机生成程序，但是生成这些分布在算法学习方面还是有指导意义的。在研究问题中经常遇到非标准的、推广的或未实现的分布。书中也介绍了生成混合变量和多元数据的方法。本书在最后一章介绍了R中的数值方法。

本书包含了大量的例子和练习. 所有的例子都可以在R统计计算环境中彻底实现, 例子的代码可以在作者的个人网站personal.bgsu.edu/~mrizzo上下载. 为了保证资料的独立完整, 所有的例子和练习所使用的都是R (基本程序包和推荐程序包) 中的数据集或者模拟数据. 此外也使用了一些CRAN提供的程序包中的函数和数据集, 它们可以通过R中的函数来安装.

本书的出版周期很长, 而软件却在不断演变. 当读者拿到这本书时可能已经出现了一个或者多个R的新版本. 笔者在现有R版本下已尽了最大努力来检查代码示例, 欢迎读者的批评、建议和指正.

致谢

R是非常出色的统计计算软件包, 而这在一定程度上也促成了本书的出现, 笔者对开发小组不断支持和改进这个软件表示感谢.

一些专家给出了非常有价值的建议和意见, 我在此表示感谢, 尤其是Jim Albert、Hua Fang、Herb McGrath、Xiaoping Shen和Gábor Székely. 我还想对我的学生们所做的贡献表示感谢, 他们在俄亥俄大学使用了本书的初稿并提供了很多有益的反馈, 特别感谢Roxana Hritcu、Nihar Shah和Jinfei Zhang. 在整个出版过程中, Taylor&Francis/CRC出版社的编辑Bob Stern、项目编辑Marsha Hecht、项目助理Amber Donley给予了很大的帮助. 最后, 我想感谢我的家人, 谢谢他们一直以来的支持和鼓励.

Maria L. Rizzo

博林格林州立大学数学与统计系

目 录

译者序

前 言

第1章 引 言	1
1.1 计算统计和统计计算	1
1.2 R环境	2
1.3 第一次如何使用R	3
1.4 使用R在线帮助	4
1.5 函数	7
1.6 数组 数据框 表	8
1.7 工作区和文档	14
1.8 使用脚本	16
1.9 使用软件包	17
1.10 图形	17
第2章 概率和统计回顾	20
2.1 随机变量和概率	20
2.2 一些离散分布	24
2.3 一些连续分布	27
2.4 多元正态分布	30
2.5 极限定理	31
2.6 统计学	32
2.7 贝叶斯定理和贝叶斯统计	36
2.8 马尔可夫链	37
第3章 随机变量生成方法	41
3.1 引言	41
3.2 逆变换法	43

3.3 接受拒绝法	49
3.4 其他变换方法	51
3.5 求和变换与混合	55
3.6 多元分布	62
3.7 随机过程	74
练习	84
第4章 多元数据可视化	87
4.1 引言	87
4.2 平面显示	88
4.3 曲面图和三维散点图	89
4.4 等高线图	96
4.5 数据的其他二维表现	99
4.6 数据可视化的其他方法	103
练习	105
第5章 蒙特卡罗积分和方差缩减	107
5.1 引言	107
5.2 蒙特卡罗积分法	107
5.3 方差缩减	114
5.4 对偶变量法	115
5.5 控制变量法	119
5.6 重要抽样法	126
5.7 分层抽样法	130
5.8 分层重要抽样法	133
练习	136
R代码	137
第6章 统计推断中的蒙特卡罗方法	139
6.1 引言	139
6.2 估计中的蒙特卡罗方法	139
6.3 假设检验中的蒙特卡罗方法	147
6.4 应用：“Count Five”等方差检验	158
练习	164
习题	165

第7章 自助法和水手刀法	167
7.1 自助法	167
7.2 水手刀法	173
7.3 基于自助法的水手刀法	177
7.4 自助法置信区间	179
7.5 更好的自助法置信区间	185
7.6 应用：交叉验证	189
练习	193
习题	194
第8章 置换检验	195
8.1 引言	195
8.2 同分布检验	198
8.3 多元同分布检验	201
8.4 应用：距离相关性	214
练习	221
习题	222
第9章 马尔可夫链蒙特卡罗方法	223
9.1 引言	223
9.2 Metropolis-Hastings算法	225
9.3 Gibbs样本生成器	238
9.4 收敛性监测	241
9.5 应用：变点分析	247
练习	253
R代码	254
第10章 概率密度估计	256
10.1 一元密度估计	256
10.2 核密度估计	269
10.3 二元和多元密度估计	278
10.4 密度估计的其他方法	286
练习	288
R代码	289
第11章 R中的数值方法	292
11.1 引言	292
11.2 一维中的求根法	298

11.3 数值积分	302
11.4 极大似然问题	306
11.5 一维最优化	309
11.6 二维最优化	312
11.7 期望最大化算法	316
11.8 线性规划——单纯形法	318
11.9 应用：博弈论	319
练习	322
附录	
附录 A 符号	325
附录 B 处理数据框和数组	327
B.1 重抽样和数据划分	327
B.2 构造子集和重塑数据	329
B.3 数据输入和数据分析	334
参考文献	345
索引	363

表 格 目 录

1.1 R语法和常用的操作符	5
1.2 常用函数	5
1.3 关于向量和矩阵的R语法和函数	6
1.4 R (“graphics”包)和其他包中的一些基本图形函数	18
3.1 R中可用的单变量概率函数	43
3.2 概率分布	84
4.1 R (“graphics”包)和其他包中的一些基本图形函数(续)	87
6.1 例6.3中第 k 层切尾均值的均方误差估计($n = 20$)	144
6.2 例6.11中与污染正态选择相比,三种正态性检验的经验功效($n = 30, \alpha = 0.1, se \leq 0.01$)	158
8.1 二元正态位置选择 $F_1 = N_2((0, 0)^T, \mathbf{I}_2)$ 和 $F_2 = N_2((0, \delta)^T, \mathbf{I}_2)$ 的显著性检验(在 $\alpha = 0.1, se \leq 0.5\%$ 下的最接近整体百分比)	214
8.2 例8.14中在 $\alpha = 0.1(se \leq 0.5\%)$ 下 $Y = X\varepsilon$ 的独立性显著检验的百分比	221
9.1 例9.4中目标分布和生成链条的分位数	233
10.1 按照直方图的三种准则对模拟数据进行估计而得到的最佳组区间数	264
10.2 密度估计的核函数	272
11.1 猜拳游戏的支付矩阵	320

插图目录

3.1	例3.2中使用逆变换法得到的随机样本的概率直方图，其中叠加了理论密度函数 $f(x) = 3x^2$	44
3.2	对Beta(3, 2)分布和例3.8中用伽马比值法模拟的随机样本进行对比得到的QQ图	53
3.3	左图为例3.11中模拟Gamma(2, 2)和Gamma(2, 4)随机变量的卷积的直方图，右图为模拟Gamma(2, 2)和Gamma(2, 4)随机变量的50%混合变量的直方图	57
3.4	例3.12中的密度估计：多个伽马密度（细线条）的混合分布（粗线条）	59
3.5	例3.14中的密度：多个伽马密度（细线条）的混合分布（粗线条）	60
3.6	例3.16中具有均值向量 $\mathbf{0}$ 、方差 $\sigma_1^2 = \sigma_2^2 = 1$ 和相关系数 $\rho = 0.9$ 的随机二元正态样本的散点图	65
3.7	例3.18中模拟的多元正态随机样本的二元边缘分布对比图（参数和iris virginica 数据的均值和方差吻合）	67
3.8	例3.20中多元正态位置混合变换的边缘分布的直方图	71
3.9	例3.21中由200个服从双变量分布 (X_1, X_2) 的点构成的随机样本，它们在单位圆上是均匀分布的	73
3.10	例3.25中模拟的更新过程的样本均值序列，参考线代表均值为 $\lambda t = t/4$ 的齐次泊松过程	80
3.11	例3.26中对称随机游动的部分实现	81
4.1	例4.1中比较virginica种鸢尾花的四种尺寸的散点图矩阵	89
4.2	例4.1中比较setosa种（圆圈）、versicolor种（三角）和virginica种（十字）鸢尾花的四种尺寸的散点图矩阵	90
4.3	例4.2中的标准二元正态密度的透视图	91
4.4	例4.3中，通过函数“persp”返回的视点变换添加了元素的标准二元正态密度透视图	93
4.5	例4.5中由cloud(lattice)生成的鸢尾花数据的三维散点图，其中不同的绘图符号表示不同的种类	95

4.6	例4.6和例4.7中volcano数据的等高线图和等值面图	97
4.7	例4.8中函数“hexbin”生成的二元正态数据的具有六边形格子的平面密度直方图	98
4.8	例4.9中的安德鲁斯曲线, 该曲线表示了纬度17.1处的“leafshape17”(“DAAG”)数据的叶长、叶宽和叶柄尺寸. 不同的曲线代表不同的叶结构	101
4.9	例4.10中对“crabs”(“MASS”)数据的一个子集生成的平行坐标图. a)从总体上看, 种类(B代表蓝色, O代表橙色)和性别(M代表雄性, F代表雌性)之间的差异很大程度上被大幅变化掩盖了; b)根据螃蟹的大小调整它的尺寸后, 群与群之间的差异变得明显了	102
4.10	例4.11中对“crabs”(“MASS”)数据中的雄性数据的一个子集生成的分段图. 根据每一个螃蟹的总体大小调整了尺寸. 两个种类分别是蓝色(B)和橙色(O)	104
5.1	例5.10中的重要函数: a)函数 f_0, \dots, f_4 (曲线0到4)和函数 $g(x)$; b)比例 $g(x)/f(x)$	129
6.1	例6.9中 t 检验 $H_0: \mu = 500, H_1: \mu > 500$ 的经验功效 $\hat{\pi}(\theta) \pm \widehat{se}(\hat{\pi}(\theta))$	153
6.2	例6.10中, 对 ε 污染正态比例混合备择假设的偏度正态性检验的经验功效 $\hat{\pi}(\varepsilon) \pm \widehat{se}(\hat{\pi}(\varepsilon))$	155
6.3	例6.11中与污染正态选择相比, 三种正态性检验的经验功效($n = 30, \alpha = 0.1, se \leq 0.01$)	159
6.4	显示例6.12中Count Five统计量端点的箱形图	159
7.1	例7.2中法学院数据的自助法重复试验	170
7.2	例7.17基于ironslag数据提出的4个模型	191
7.3	例7.17中ironslag 数据二次模型的残差图	193
8.1	例8.1(左)和例8.2(右)中重复试验的置换分布	199
8.2	例8.6中 $T_{n,3}$ 的置换分布	208
8.3	例8.8中双样本 e 统计量重复试验的置换分布	213
8.4	例8.13中dCov的置换重复试验	220
8.5	例8.14中dCov和Wilks Lambda W 距离协方差检验的经验功效比较	221
9.1	例9.1中Rayleigh 分布的Metropolis-Hastings 样本生成器生成的链条的一部分	228
9.2	例9.1中叠加了目标Rayleigh密度的直方图和Metropolis-Hastings链条的QQ图	229
9.3	例9.3中不同方差的建议分布生成的随机游动Metropolis链条	232
9.4	例9.5中 β 的随机游动Metropolis链条	235

9.5	例9.6中 p 的建议分布为Beta(1, 1)的独立性样本生成器链条的分布 (去掉了长度为100的训练样本)	238
9.6	例9.6中 p 的建议分布分别为Beta(1, 1) (左) 和Beta(5, 2) (右) 时独立性样本生成器生成的链条	239
9.7	例9.7中Gibbs样本生成器生成的二元正态链条	242
9.8	例9.8中4个Metropolis-Hastings链条的移动均值 ψ 的序列	246
9.9	例9.8中4个Metropolis-Hastings链条的Gelman-Rubin- \hat{R} 序列: a) $\sigma = 0.2$; b) $\sigma = 2$	247
9.10	例9.9中矿难的年次数	248
9.11	例9.9中Gibbs样本生成器的输出	252
9.12	例9.9中矿难变点分析的 μ 、 λ 和 k 的分布	252
10.1	例10.1中: a) 大小为25; b) 大小为1000的样本的直方图正态密度估计 (附加了标准正态密度曲线)	259
10.2	例10.3中Old Faithful (老忠实泉) 等待时间密度的直方图估计: a) Scott准则给出了9个分组; b) 对分点使用“pretty”函数之后, 具有参数“breaks=“scott””的“hist”函数只使用了7个分组	263
10.3	例10.4中Old Faithful (老忠实泉) 等待时间密度的频率多边形估计	266
10.4	一个正态样本在相同组宽、不同分组起点下的直方图估计以及标准正态密度曲线	266
10.5	例10.6中Old Faithful (老忠实泉) 等待时间的ASH密度估计	270
10.6	使用带宽为 h 的高斯核的核密度估计	271
10.7	密度估计的核函数	272
10.8	例10.7中density在不同带宽下得到的Old Faithful (老忠实泉) 等待时间的高斯核密度估计	275
10.9	例10.8中density在不同带宽下得到的降水数据的核密度估计	276
10.10	例10.10中一个指数密度的高斯核密度估计 (实线) 以及真实密度 (虚线). 在第二幅图中, 对相同的数据使用了反射边界方法	277
10.11	例10.13中二元正态数据的密度多边形, 使用了正态参考准则 (Sturges准则) 来决定组宽	281
10.12	例10.14中二元正态数据的二元ASH密度估计	284
10.13	例10.15中二元正态混合数据的乘积核估计 (左边是正态参考准则)	287
11.1	例11.8($n = 10, r = 0.5, \rho = 0.2$): a) 被积函数的图形; b) 将积分值看成 ρ 的函数	304
11.2	样本大小为10时的相关性统计量密度	306
11.3	例11.11中的函数 $f(x)$	310

11.4 例11.12中Gamma($r = 5, \lambda = 2$)随机变量的似然函数通过数值最优化得到的极大似然估计重复试验 312

第1章 引言

1.1 计算统计和统计计算

计算统计和统计计算是统计的两个领域，可宽泛地描述为用计算、图形和数值方法去解决统计问题。统计计算传统上把重点放在数值方法和算法，例如最优化和随机数产生，而计算统计可能包括诸如探索数据分析、蒙特卡罗方法和数据分割的主题。然而，大多数使用计算密集方法的研究人员会同时使用统计计算和计算统计的方法；统计计算和计算统计有比较大的重叠，在不同的上下文和不同的学科中，术语的使用有所不同。Gentle^[113]、Givens和Hoeting^[121]使用计算统计去涵盖所有与之相关的并且应该包含在现代入门教科书中的主题，这样统计计算在某种程度上被包含于这个更宽泛的计算统计的定义中。另一方面，期刊和专业组织却使用这两个术语去包含相同的领域，这样的组织有国际统计计算协会、国际统计协会和美国统计协会的统计计算分会。

这本书包含了统计计算和计算统计的部分内容，因为一本统计计算方法的入门书必须同时包括两者。现将部分本书所涵盖的主题表述如下：

蒙特卡罗方法是指在需要使用模拟技术的统计推断和数值分析中的一个宽泛的集合。许多统计问题可以通过某种形式的蒙特卡罗积分来解决。在参数自举中，用产生自给定概率分布的样本去计算概率，获得诸如偏差和样本误差之类的信息来评估在统计推断中各方法的性能，并比较基于同一问题不同方法的性能。诸如普通自助法和水手刀法等重复抽样方法属于非参数方法。当随机变量的分布或是产生变量分布的方法无法得到时，非参数方法可以被使用。对蒙特卡罗分析的需求之所以不断增长，是因为在许多问题中，渐进估计值是令人不满意的或是困难的。收敛到极限分布的速度很慢，或是我们需要基于有限样本的结果；或是渐进分布有未知参数。5、6、7、8和9章是讲述蒙特卡罗方法的。模拟分析的第一工具是产生伪随机样本；这些方法将在第3章中讲述。

马尔可夫链蒙特卡罗（Markov Chain Monte Carlo, MCMC）方法是基于样本产生自特定目标概率分布的算法，而该分布是马尔可夫链的平稳分布。这些方法在贝叶斯分析、计算物理和计算金融中被广泛使用。第9章讲述马尔可夫链蒙特卡罗方法。

在计算密集型方法中，一些主题也同样值得介绍。第10章密度估计讲述了密度的非参数估计，在探索数据分析、聚类分析和其他方面有着很多用途。计算方法是多维数据可视化和维数缩减的核心。随着人们对海量数据和流数据的兴趣日渐浓厚，来自工程和

生物领域的高维数据也在日益增多, 例如, 在多变量分析及可视化方面, 就迫切需要一种改进的或全新的算法. 第4章介绍了多维数据的可视化. 第11章讲述了在最优化和数值积分中的一些方法.

一些好的参考文献被推荐. Gentle^[113]和Gentle负责编写的计算统计手册详尽地涵盖了计算统计的内容. Givens 和Hoeting^[121]是一本关于计算统计和统计计算比较新的研究生教材. Martinez和Martinez^[192]是计算统计的入门书, 书中的例子代码是用MATLAB写的. 计算统计的书籍包括Kennedy和Gentle^[161]以及Thisted^[269]. Kundu或Basu^[165]包含了对统计计算比较新的概述. Lange^[168]或Monahan^[202]则讲述了数值分析的统计应用. 涵盖蒙特卡罗方法或是重复抽样的书籍包括Davison和Hinkley^[63], Efron和Tibshirani^[84], Hjorth^[143], Liu^[179], 以及Robert 和Casella^[228]. Scott^[244]和Silverman讲述了密度估计.

1.2 R环境

R环境是基于S的用来数据分析和可视化的一套软件和编程语言. “什么是R?” 这是一个经常被问到的问题, 这个问题也包括在R的在线文档中. 以下是从R FAQ中的一段摘录.

R是统计计算和图形软件. 它包含一种语言、图形运行环境、编译器, 以及特定系统函数的入口, 它能运行存储在脚本文档中的程序.

R软件的主页是<http://www.r-project.org/>, 当前可用的R版本和说明文档存储于R 综合典藏网(CRAN). CRAN的主网站设在奥地利的维也纳技术大学(TU Wien, Austria). 可通过<http://cran.R-project.org/>访问. R版本包括基本和推荐的包含文档的包. 帮助系统和几个参考手册将和软件一起被安装.

R是基于S语言的. R-FAQ (常见问题)^[147]给出了R和S的一些详细区别. Venables和Ripley是在应用统计中如何使用S、Splus和R的一本好书. 其他关于S的资料有参考文献[24, 41, 42, 277].

一本非常好的入门书是《R入门手册》^[279]. 一些入门书籍包括 Dalgaard^[62]和Verzani^[280]. 关于编程方法参见Chambers^[41], 以及Venables和Ripley^[277,278]. 其他关于Splus、S和(或)R的文献同样也有帮助(参见Crawley^[57]或Everitt和Hothorn^[88]). Albert^[5]是贝叶斯统计计算的入门书. 关于统计建模, 参见Faraway^[90,91], Fox^[97], Harrell^[131], 以及Pinheiro和Bates^[211]. 更多的参考书可以在R软件主页上找到.

在本书中编程只在需要的章节才讨论. 本书的“R笔记”解释新的函数或编程方法. 本书鼓励读者查询R帮助文档和手册^[147,279,217]. 关于安装和使用图形用户界面, 最好的参考书是R手册^[218]和网站www.r-project.org上的信息.

在余下部分, 我们讲述一些基本信息来帮助新用户使用R. 其主题包括基本语法、如何使用在线帮助、数据集、文档、脚本和各种包, 同时也简述了各种图形函数. 关于数据框, 参照附录B.