

数据科学与大数据技术

不学编程做 R 统计分析： 图形界面 R Commander 官方手册

[加] 约翰·福克斯(John Fox)

著

马国强

马国强

译

清华大学出版社

北 京

北京市版权局著作权合同登记号 图字: 01-2022-4392

Using the R Commander

EISBN: 978-1-4987-4190-3

Copyright© 2017 by Taylor & Francis Group, LLC.

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

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

本书原版由 Taylor & Francis 出版集团旗下 CRC Press 出版公司出版, 并经其授权翻译出版。

本书中文简体翻译版授权由清华大学出版社有限公司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。

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

版权所有, 侵权必究。举报: 010-62782989, beiqinquan@tup.tsinghua.edu.cn。

图书在版编目(CIP)数据

不学编程做R统计分析: 图形界面R Commander官方手册 / (加) 约翰·福克斯(John Fox), 马国强著; 马国强译. —北京: 清华大学出版社, 2023.9

(数据科学与大数据技术)

书名原文: Using the R Commander

ISBN 978-7-302-64564-1

I. ①不… II. ①约… ②马… III. ①统计分析 IV. ①O212.1

中国国家版本馆CIP数据核字(2023)第180128号

责任编辑: 王 军

装帧设计: 孔祥峰

责任校对: 成凤进

责任印制:

出版发行: 清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-83470000 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:

经 销: 全国新华书店

开 本: 148mm×210mm 印 张: 6.625 字 数: 213 千字

版 次: 2023 年 11 月第 1 版 印 次: 2023 年 11 月第 1 次印刷

定 价: 59.80 元

产品编号: 095412-01

作者简介

John Fox 是 McMaster 大学的社会学教授。Fox 教授开发了 R Commander 软件,撰写了许多关于统计方法的文章和书籍,包括与 Sanford Weisberg 合著的 *An R Companion to Applied Regression, Second Edition*。Fox 是 *Journal of Statistical Software* 的副主编,也是专著系列 *Quantitative Applications in the Social Sciences* 的前主编。

马国强是加拿大 Bank of Montreal(蒙特利尔银行)的高级数据库专家, Toronto Metropolitan University(多伦多都市大学)的 MBA 以及企业信息安全、隐私和数据保护专业的硕士。他多年来先后任职于 AEGON Canada、IBM Canada 等公司的数据安全、数据分析及数据库分析和管理高级技术职位,为加拿大联邦政府和安大略省政府的数据分析和数据库系统提供总体设计、实施和技术支持。另外,他还为银行、保险公司的混合云计算环境中活动交易数据、企业存储数据提供数据安全和高可用性 IT 解决方案。

中文版序

R Commander 大幅降低了 R 的使用和入门难度。如果你正在撰写论文，纠结于如何应对统计分析的细节；如果你正在进行一个科研项目，而无法决定到底哪一个模型更契合现有的数据，以进一步支持最终的结论；如果你正在精耕于一个重要的商业项目，筹划用相得益彰的生动插图展示诱人的商业机会以打动重要客户或管理高层——这时，不妨尝试 R Commander。借助我们业已熟悉的菜单和对话框，把所有细节交给它，我们只需要关心最终结果。把宝贵的精力专注于主题——这不正是我们的初心吗？

很多 MBA 学生的经历都给出了同样的反馈，商业统计分析课程的大部分时间都花在了陌生的 R 程序编制开发上，从而轻视了项目筹划本身以及潜在商业规律的发掘和分析。MBA 的学生绝大部分都没有 IT 知识背景，学习程序开发可能还是第一遭。学习、掌握一门语言并且很快应用到实际项目中是一个循序渐进的过程，不是一蹴而就的。

对于有志于使用 R 编程开发进行深度应用的用户，从 R Commander 开始学习也是一个不错的选择。R Commander 的一个典型应用流程是，把原始数据读入系统中，进行数据的初步处理和必要的建模前准备，应用适当的模型作数据统计分析，输出结果洞悉数据背后的内在规律。这一逻辑简洁的流程岂不是也贯穿于 R 的数据应用？

数据统计分析早已渗透到现代生活的许多方面。“回归”一词的诞生不仅讲述着一个非常有趣的故事，更是昭示了研究、应用现代数据统计分析的历史已经发展了一百多年。例如，通过把回归分析和其他现代统计分析方法应用于电影行业，可以发现哪些影片历经了时间的洗礼、历久弥新地影响着怎样的受众，哪些因素影响着票房收入的起伏，以及业界人士应该怎样因势利导、繁荣文化。

当然，现代统计分析也需要理论和应用的不断推进、突破，与时俱进，以更好地服务于人类。例如，现代统计分析应用于席卷全球的重大医学、公共卫生事件——新冠肺炎。首先，需要客观地评估这种疾病的危害，对致死率给出定量的评估。其次，依据已有的数据并纳入病毒变异、免疫获得、疫苗接种、人口变化等因素，准确预测此种疾病的发展趋势。

让我们自豪的是，如今的中国不仅是世界第二大经济体，同时也是在诸多科技领域名列前茅的科技强国。这些日新月异的发展无时无刻不在产生着大量的数据。要想拨云见日，透过纷繁的数据发现潜在的价值，扼住关键因素，促进跳跃式发展，需要得力的统计分析工具。来吧，让我们从 R Commander 做起！

虽然十分用心，书中仍难免有纰漏，希望读者不吝赐教、指正。

致 谢

Michael Friendly、Allison Leanage 和一些匿名审稿人对本书的草稿提出了有益的建议。

在此，我要感谢众多同僚，他们对 **Rcmdr** 包的贡献在包文档中得到体现(请参阅“帮助”|“关于 Rcmdr”菜单)。尤其是，Richard Heiberger 为 R Commander 的早期开发作出了许多贡献，其中最重要的是用于实现 R Commander 插件包自启动的原始代码。Miroslav Ristic 极大地改进了“概率分布”对话框的代码。Milan Bouchet-Valat 作为 **Rcmdr** 包的开发人员在 2013 年加入了我的行列，帮助我对 R Commander 界面进行了现代化的改造，使其更好地适应各种计算平台。

R 核心团队的成员 Peter Dalgaard 将 **tcltk** 包整合到标准 R 发行版中，可以说对 R Commander 间接地作出了重要贡献。同样，Milan 和我使用 Philippe Grosjean 的 **tcltk2** 包增强了 R Commander 的界面，Philippe 一直是 Tcl/Tk 信息的宝贵源泉。R 核心团队的另一位成员 Brian Ripley 一直加班加点，帮助我解决了 R Commander 开发中的各种问题。

还要感谢 Chapman & Hall/CRC 出版社的编辑 John Kimmel、Shashi Kumar(感谢他的 LaTeX 专业知识)，以及 Chapman & Hall/CRC 出版社的所有工作人员，感谢他们给予我的帮助和鼓励。

最后，我在 R Commander 上的一部分研发工作得到了加拿大社会科学和人文研究委员会以及 McMaster 大学社会统计学的参议员 William McMaster 的资助支持。

前言

R Commander 是 R 的点选式图形用户界面，通过熟悉的菜单和对话框提供对 R 统计软件的访问，而不是通过输入难懂的命令。本书解释了如何使用 R Commander，希望让入门级和中级统计课程的学生和教师、想要使用 R 而不必苦恼于编程的科研人员，以及最终要过渡到命令行交互但倾向于入门更容易的读者感兴趣，并对专业工作提供帮助。

在我看来，在基础统计学课程中，中心目标应该是讲解基本的统计学思想——分布、统计关系、估值、样本差异、观测与实验数据、不规则分布等。人们不希望把基础统计课程变成学习如何为统计软件编写命令的练习。最初开发 R Commander 是出于这样的想法：为 R 提供透明、直观、点击式的图形用户界面，使用熟悉的菜单和对话框进行操作，适应所有常用的操作系统(如 Windows、Mac OS X 和 Linux/UNIX)，并作为标准的 R 程序包(称为 **Rcmdr** 包)分发和安装。

尽管最初打算用于基础统计课程，但本书各章中介绍的 R Commander 的当前功能已远远超出基本统计。当前版本的 **Rcmdr** 程序包包含近 15 000 行的 R 代码，这其中不包括注释、空行、文档等。此外，R Commander 就像 R 本身一样，被设计成可通过插件包进行扩展。而且，标准的 R 程序包也可以增加或修改 R Commander 的菜单和对话框(具体内容请参见第 9 章)。

我在 20 世纪 90 年代后期乘 R 方兴未艾的东风，将其纳入我的教学(应用回归分析和广义线性模型的研究生社会统计学课程)中。2002 年，我出版了一本关于使用 R 和 S-PLUS 进行应用回归分析的书^[20]。¹

我也想用 R 来讲解社会科学研究生和本科生的基础统计学，但感觉 R 的命令行界面是一个障碍。我曾期望有人为 R 引入图形用户界面，但一直

1 该书的第 2 版是我与 Sanford Weisberg 合著的^[29]，并且专注于 R。

没能实现。因此，在 2002 年左右，我决定自己承担这个任务。经过一些前期的探索，最终决定使用 Tcl/Tk GUI 构建器，因为基本的 R 发行版自带 **tcltk** 程序包，它提供了 Tcl/Tk 的 R 接口。这个选择让我能够编写一个 R GUI(即 R Commander)；它可以在 R 支持的所有操作系统上运行，并且完全用 R 编码，以最大限度地减少安装其他不必要的软件。

本书提供了有关 R 和 R Commander 的背景信息，解释了如何在读者的计算机上获取和安装 R 和 R Commander。最后，将展示如何使用 R Commander 执行各种常见的统计任务。

关于参考文献

读者在阅读本书时，会不时遇到参考文献编号，格式是放在方括号的数字。读者可扫描封底二维码下载“参考资料”文件，然后根据编号查看资料信息。

目 录

第 1 章 概述	1
1.1 什么是 R 和 R Commander	1
1.2 简要回顾 R 和 R Commander	3
1.3 章节概要	5
1.4 进一步说明	6
1.5 本书资源网站	7
第 2 章 R 和 R Commander 的安装	9
2.1 获取并安装 R 和 R Commander	9
2.2 在 Microsoft Windows 上安装 R 和 R Commander	10
2.3 在 Mac OS X 上安装 R 和 R Commander	14
2.3.1 为 Mac OS X 安装 X11(XQuartz)窗口系统	15
2.3.2 安装 Rcmdr 程序包及其附属程序包	16
2.3.3 防止 Mac 睡眠模式影响 R	16
2.3.4 故障排除	17
2.4 在运行 Linux 和 UNIX 的计算机上安装 R 和 R Commander	17
2.5 安装可选的辅助软件: Pandoc 和 LaTeX	18
第 3 章 快速浏览 R Commander	21
3.1 启动 R Commander	21
3.2 R Commander 的界面	22
3.3 将数据读入 R Commander 中	25
3.4 检查和重新编码变量	29
3.5 制作列联表	37

3.6	创建报告	39
3.6.1	通过剪切和粘贴创建报告	40
3.6.2	将报告创建为动态文档	41
3.7	编辑命令*	48
3.8	终止 R Commander 会话	51
3.9	自定义 R Commander*	52
3.9.1	使用“Commander 选项”对话框	52
3.9.2	设置 Rcmdr 程序包选项	54
3.9.3	在 R Commander 中管理颜色	55
3.9.4	保存 R Commander 选项	56
第 4 章	数据输入与数据管理	59
4.1	数据管理概述	59
4.2	数据输入	60
4.2.1	在 R Commander 数据编辑器中输入数据	60
4.2.2	从纯文本文件中读取数据	63
4.2.3	从电子表格和其他来源导入数据	65
4.2.4	访问 R 程序包中的数据	70
4.3	从 R Commander 保存和导出数据	72
4.4	修改变量	73
4.4.1	重新编码变量	73
4.4.2	计算新变量	75
4.4.3	变量的其他操作	78
4.5	操作数据集	81
4.5.1	对数据集的特殊操作*	82
4.5.2	选取部分案例	83
4.5.3	合并数据集*	85
第 5 章	数据汇总和图形化	89
5.1	简单的数值总结	89

5.2	列联表	93
5.3	绘制变量分布图	95
5.3.1	图形化数值数据	95
5.3.2	图形化分类数据	99
5.4	绘制关系图	102
5.4.1	简单的散点图	102
5.4.2	增强型散点图*	104
5.4.3	散点图矩阵	105
5.4.4	散点图和散点图矩阵中的点识别	107
5.4.5	3D 散点图*	108
5.4.6	绘制平均数	110
第 6 章	简单的统计检验	113
6.1	平均数检验	113
6.1.1	独立样本的平均数差异 T 检验	113
6.1.2	单因素(因子)方差分析	116
6.1.3	双因素(因子)和多因素(因子)方差分析	121
6.2	比例检验	123
6.3	方差检验	126
6.4	非参数检验	127
6.5	其他简单检验*	128
第 7 章	拟合线性模型和广义线性模型	131
7.1	线性回归模型	131
7.2	带有因子的线性模型*	134
7.2.1	线性模型公式	135
7.2.2	边际原则	136
7.2.3	使用加拿大职业声望数据的例子	137
7.2.4	虚拟变量和因子变量的其他对比方式	141

7.3	拟合回归样条和多项式*	144
7.3.1	多项式项	144
7.3.2	回归样条	146
7.4	广义线性模型*	147
7.5	其他回归模型*	151
7.6	可视化线性模型和广义线性模型*	151
7.7	系数的置信区间和假设检验	156
7.7.1	置信区间	157
7.7.2	方差分析和偏差表分析*	157
7.7.3	检验比较模型*	159
7.7.4	检验线性假设*	161
7.8	回归模型诊断*	162
7.9	模型选择*	166
第 8 章	概率分布与模拟	169
8.1	运用概率分布	169
8.1.1	连续型分布	169
8.1.2	离散型分布	172
8.2	绘制概率分布	173
8.3	简单随机模拟	175
8.3.1	设置 R 伪随机数发生器的种子	176
8.3.2	中心极限定理的简单模拟示例	177
第 9 章	使用 R Commander 插件包	181
9.1	获取和加载插件	181
9.2	使用 RcmdrPlugin.TeachingDemos 包	183
9.3	使用 RcmdrPlugin.survival 包进行生存分析*	185
附录 A	R Commander 菜单指南	195

第 1 章

概述

本章将介绍 R 和 R Commander，解释它们的含义和由来；还将概述全书的内容，展示如何访问本书的网站。

1.1 什么是 R 和 R Commander

本书的主题 R Commander 是 R 的点选式图形用户界面(GUI)，它允许通过平常熟悉的菜单和对话框来使用 R 统计软件，而无须输入命令。本书假定读者已经熟悉书中所使用的统计方法，或者说读者正通过课堂或阅读相关材料学习它们。本书的目的是展示如何在 R Commander 中运用常见的统计方法来进行数据分析，而不是讲解统计方法。

这意味着读者可以跳过本书中那些采用不熟悉的统计方法的章节。例如，第 7 章的大部分篇幅(介绍在 R Commander 中使用统计模型)超出了一般性的基础统计课程的范围。书中对讨论高级或困难内容的章节标记了星号。

R 是功能强大、免费、开源的统计软件。虽然很难确切地说出使用 R 的具体人数，但若通过 Internet 流量来判断，R 可能是世界上最流行的统计软件。无论如何，R 的使用非常广泛，并且它的使用量正在迅速增长。

R 吸收了一种非常适合统计应用程序开发的编程语言。它起源于 S 编程语言，后者最初由 John Chambers 领导的贝尔实验室的统计学家和计算机科学家在 20 世纪 80 年代开发而成^[6]。事实上，R 可以被视为 S 的方言。由于最终被并

入了称为 S-PLUS 的商业产品中, 因此在 R 被开发出来之前, S 在统计学家中广受欢迎。目前, 免费的、开源的 R 已完全使收费的 S-PLUS 黯然失色。

用 Richard Stallman¹的话讲, R 是自由软件, 其中包含两层含义^[41]。首先, 很明显, 用无须花钱这一含义来衡量, 它是自由软件; 但从更深层的意义上讲, 用户可以自由地修改和分发 R。具体来说, R 已得到自由软件基金会的通用公共许可证(GPL)授权, 该授权可防止个人或公司限制用户进一步修改和重新分发 R 的自由。自由的第二层含义实际意味着 R 是开源的, 即 R 不仅能以可执行程序的形式分发, 而且 R 的源代码(用各种编程语言编写, 包括用 R 本身编写)也可供感兴趣的用户使用。有关 R 的更多信息, 可访问 R 网站 <https://www.r-project.org/>。R Commander 也是在 GPL 授权下分发的免费开源软件。

使用 R 分析数据并不一定需要用 R 语言编写程序, 因为基本的 R 发行版本具有令人惊叹的内置统计功能。不过, 截至目前, 标准 R 发行版的功能已扩展了近 8000 个由用户提供的 R 附加程序包, 这些程序包可以通过 Internet 上的 Comprehensive R Archive Network(缩写为 CRAN, CRAN 的发音为 kran 或 see-ran; 见 <https://cran.r-project.org/>)免费获得。此外, 可以通过密切相关的 Bioconductor Project(<http://bioconductor.org/>)获得大约 1000 个其他 R 程序包, 该项目主要为生物信息学(基因组学)开发软件。

无论是编写自己的 R 程序, 还是使用预先打包的程序, R 中的标准数据分析都是由输入 R 语言命令构成。举一个简单的例子, 要计算变量 **income** 的平均值, 可以输入命令 `mean(income)` 去调用 R 的 **mean** 函数(程序)。类似地, 要计算 **income** 对 **education** 和 **experience** 的线性最小二乘回归, 可以使用 **lm**(线性模型)函数, 输入命令 `lm(income~education + experience)`。学习编写这样的 R 命令是一项重要技能, 并且最终是使用 R 的最有效方法(参见 1.4 节), 但它可能会对 R 的新用户或临时用户带来巨大的障碍。

1 Richard Stallman 是自由软件基金会的创始人, R Project for Statistical Computing 与该基金会有关联。

1.2 简要回顾 R 和 R Commander

R 始于 1990 年左右，当时是 Robert Gentleman 和 Ross Ihaka(新西兰奥克兰大学的两位统计学家)的个人项目^[32]。Gentleman 和 Ihaka 实际上是将预先存在的统计编程语言 S 的语法嫁接到 Lisp 的 Scheme 分支上，后者是一种通常与人工智能工作相关的编程语言。事实证明，这是一个不错的选择，因为如上所述，S 语言已被统计学家广泛使用。

最终，Ihaka 和 Gentleman 在 Internet 上宣传了他们的成果，吸引了其他开发人员加入该项目，其中包括 S 的主要开发人员 John Chambers。然后，在 1997 年，由 9 名开发人员组成的核心团队正式确定了 R Project for Statistical Computing 计划。该团队至今已增加到 20 名成员，其中许多是统计计算领域的重要人物。R 核心团队负责基本 R 发行版的持续开发和维护。

正如前面所解释的，R 是在自由软件通用公共许可证下分发的。它的版权归 R 基金会所有，R 基金会的组成人员包括 R 核心团队的成员以及其他大约 12 人，其中本人(John Fox)也是 R 基金会的一员。

R 的增长速度令人惊讶。例如，图 1.1 显示了在能够获得数据的 14 年间 CRAN 中 R 程序包存档的扩展情况。¹图的横轴记录了 R 的版本和相应的日期，而纵轴以对数刻度显示了 CRAN 程序包的数量，因此线性趋势表示了其呈指数级增长。²图上的线通过最小二乘回归拟合³。从图 1.1 中可以看到，尽管 CRAN 最初几乎以指数级增长，但近年来它的增长速度有所放慢。最小二乘线的斜率表明，在此期间，CRAN 的平均年增长率约为 35%。

1 此图更新自作者的一篇文章^[23]，文中讨论了 R Project 的社会组织和沿革。

2 如果读者不熟悉对数，则不必担心。要点是：随着程序包数量的增加，刻度的压缩比会更高，例如对数刻度上 100~200 个程序包之间的距离等同于 200~400 之间的距离，以及 400~800 之间的距离，因此所有这些相等的距离表示程序包数量的加倍。

3 同样，如果不熟悉最小二乘法，也不要担心，因为几乎可以肯定会在基础统计学课程中学到该主题。基本思想是该线在某种意义上会尽可能平均地接近这些数据点。

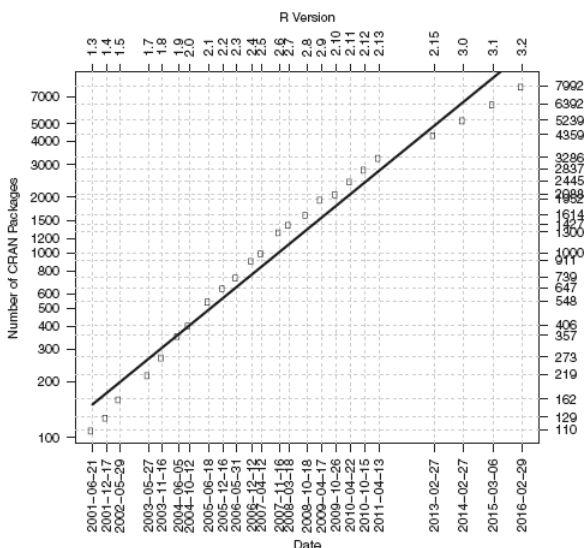


图 1.1 CRAN 的增长情况。纵轴以对数刻度显示 CRAN 程序包的数量，而日期和相应的 R 发行版本分别显示在图的底部和顶部。图中的线是对数据点的最小二乘回归拟合。图中省略了两个 R 版本(1.6 和 2.14)，因为它们的记录日期与之前版本的日期非常接近(数据来源：2016 年 3 月 3 日从 <https://svn.r-project.org/R/branches/> 下载)

正如在本书前言中提到的那样，我从 2002 年左右开始从事 R Commander 相关的工作，并于 2003 年 5 月向 CRAN 提交了 **Rcmdr** 程序包的 0.8-2 版本。第一个非 beta 版本 1.0-0 于两年后问世，我在 *Journal of Statistical Software*(该杂志是美国统计协会的在线期刊)的一篇论文中对其作了介绍^[21]。2016 年 3 月，当撰写本章时，尽管已过去十多年，但该论文已被下载近 14 万次。

在此期间，我仍继续开发 R Commander：1.1-1 版本发布于 2005 年，它引入了将 R Commander 界面翻译成其他语言的功能(这是 R 本身支持的功能)，目前提供 18 种这样的翻译(中文和简体中文被视为两种翻译)。2007 年，1.3-0 版本首先提供 R Commander 插件包，目前 CRAN 上有大约 40 个这样的插件。2013 年，Milan Bouchet-Valat 作为 R Commander 的开发人员加入了开发工作，同年发布的 2.0-0 版本具有经过改进的更加一致的界面，例如带有选项卡式的对话框。

1.3 章节概要

第 2 章介绍如何从 Internet 下载 R 并在 Windows、Mac OS X 和 Linux/UNIX 系统上安装它和 R Commander。如果读者已成功安装 R 和 R Commander，可随时跳过这一章。不过，该章有一些故障排除信息，如果遇到问题，可以参考这些信息。

第 3 章通过解决一个简单问题来演示 R Commander 图形用户界面的使用：构造一个列联表以检查两个分类变量之间的关系。在研究示例时，将说明如何启动 R Commander，描述 R Commander 界面的结构，并显示如何将数据读入 R Commander，如何修改数据以准备进行分析、如何绘制图形、如何计算数据的数值总结、如何创建工作的输出报告、如何编辑和重新执行 R Commander 生成的命令，及如何终止 R 和 R Commander 会话。简而言之，将介绍使用 R Commander 进行数据分析的典型工作流程，同时将说明如何自定义 R Commander 界面。

第 4 章介绍如何从多种来源将数据送入 R Commander 中，包括直接在键盘上输入数据、从纯文本文件中读取数据、访问存储在 R 程序包中的数据、从 Excel 或电子表格以及其他统计软件中导入数据。该章还将解释如何在 R Commander 中保存和导出 R 数据集以及如何修改数据，例如如何创建新变量以及如何创建当前数据集的子集。

第 5 章介绍如何使用 R Commander 计算数据的简单数值总结、构造/分析列联表以及绘制常见的统计图。尽管一般性的基础统计学课程涵盖了该章的大部分统计知识内容，但是一些主题(例如分位数比较图和平滑散点图)则较为高级。

第 6 章介绍如何执行平均数、比例和方差的简单的统计假设检验(并计算置信区间)，以及简单的非参数检验、正态性检验和相关性检验。大部分检验通常会在基础统计课中遇到，尤其是通常使用平均数和比例的检验和置信区间来引入统计推断。

第 7 章介绍如何在 R Commander 中处理线性和广义线性回归模型，以及一旦对数据进行拟合将如何对回归模型执行其他计算。

第 8 章介绍如何使用 R Commander 对概率分布进行计算、绘制概率分布图, 以及进行简单的随机模拟。

第 9 章介绍如何使用 R Commander 插件包。CRAN 上提供的许多插件包大幅增强了 R Commander 的功能。插件是 R 程序包, 用于向 R Commander 添加菜单、菜单项和对话框。该章将向读者展示如何安装插件程序包, 并以 **RcmdrPlugin.TeachingDemos** 程序包和 **RcmdrPlugin.survival** 程序包为例说明 R Commander 插件的应用。

本书最后的附录显示了完整的 R Commander 菜单集。

1.4 进一步说明

如果你经常使用 R, 则可能不需要 R Commander 的帮助来编写自己的 R 命令以及需要的 R 程序。优先使用命令行界面而非图形界面(如 R Commander)有以下理由。

- R Commander 图形界面仅提供对 R 功能以及 CRAN 上众多可用的 R 程序包的一小部分应用。因此, 要充分利用 R, 将必须学习命令编码。
- 即使满足于 R Commander 及其各种插件中的功能, 但经常使用 R 的用户会发现命令行界面更有效。一旦记住了各种命令及其参数, 将体会到在命令行中比在图形界面中工作效率更高。
- 编程开发对完成项目大有帮助。例如, 编写脚本和程序通常是完成数据管理任务最快、最直接的方法。

如果决定通过命令行界面学习使用 R, 那么可以从很多书籍和其他资源得到帮助。例如, 作者和 Sanford Weisberg 撰写了一本书^[29], 该书在应用回归分析的背景下介绍了 R, 包括 R 编程。许多有关备用资源(包括免费资源)的信息请参见 R 主页(<https://www.r-project.org>)上的 Documentation 链接。

R Commander 旨在促进向 R 的命令行使用过渡: R Commander 生成的命令在“R 语法文件”选项卡中可见。“R 语法文件”选项卡的内容可以保存到文件中, 并可以在 R Commander 或 R 编程编辑器中重新使用。同样, 可以独立于 R

Commander 保存、编辑和执行 R Commander 的 R Markdown 选项卡中生成的动态文档。第 3 章将简要讨论这些功能。

R 的 Windows 和 Mac OS X 安装都带有简单的编程编辑器，但是强烈建议将 RStudio 交互式开发环境(IDE)用于 R 的命令行操作。RStudio 包含功能强大的编程编辑器，是在 R 中进行常规数据分析和编程的理想选择，包括 R 程序包的开发，并且它支持 R Markdown 文档。与 R 和 R Commander 一样，RStudio 也是免费的开源软件：可访问 RStudio 网站(网址为 <https://www.rstudio.com/products/rstudio/>)以获取详细信息，包括大量文档。

1.5 本书资源网站

作者为读者设立了一个网站(<https://tinyurl.com/RcmdrBook>)，提供各种资源。

- 本书示例中使用的所有数据文件。
- 详细(并且可能会更新)的安装说明，包括超出第 2 章中说明的故障排除信息。
- 本书出版后有关 R Commander 的重大更新信息以及对书中错误更正的勘误表。
- R Commander 插件包作者手册。

关于软件版本的说明：本书是基于目前的 R 版本 4.1.0 和 **Rcmdr** 程序包版本 2.7-1。未来 R Commander 的重大更改或将对 R Commander 有重大影响的更改都将在本书的网站上披露。

Chapman & Hall/CRC 出版社维护了本书的英文网站链接，网址是 <https://www.crcpress.com/Using-the-R-Commander-A-Point-and-Click-Interface-for-R/Fox/p/book/9781498741903>，当然也可以从 <http://tinyurl.com/RcmdrBook> 进行访问。