

图书在版编目(CIP)数据

R语言初学指南 / (美) 丹尼斯 (Dennis, B.) 著 ;
高敬雅, 刘波译. — 北京 : 人民邮电出版社, 2016. 1
ISBN 978-7-115-40787-0

I. ①R… II. ①丹… ②高… ③刘… III. ①程序语
言—程序设计 IV. ①TP312

中国版本图书馆CIP数据核字(2015)第266719号

版权声明

The R Student Companion 1st Edition / by Brian Dennis / ISBN: 978-1-4398-7540-7
Copyright © 2013 by CRC Press.

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。版权所有, 侵权必究。

Posts & Telecommunications 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人民邮电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。

-
- ◆ 著 [美] Brian Dennis
 - 译 高敬雅 刘波
 - 责任编辑 王峰松
 - 责任印制 张佳莹 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
固安县铭成印刷有限公司印刷
 - ◆ 开本: 700×1000 1/16
印张: 18
字数: 368 千字 2016 年 1 月第 1 版
印数: 1-3 000 册 2016 年 1 月河北第 1 次印刷

著作权合同登记号 图字: 01-2013-5568 号

定价: 49.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316
反盗版热线: (010)81055315

内容提要

R 是一个开源、跨平台的科学计算和统计分析软件包，它提供了丰富多样的统计功能和强大的数据分析功能。在大数据和机器学习快速发展的今天，R 已经成为数据分析领域炙手可热的通用语言。

本书的内容涵盖 R 的基础知识，包括创建、运行以及调试 R 脚本，用户自定义 R 函数，用 R 绘制基本图形，R 的循环语句和逻辑控制语句，二次函数、三角函数、指数函数、对数函数以及如何用 R 绘制这些函数图形，矩阵的基本运算和线性方程组的求解，概率分布与模拟，数据的拟合等。这些内容涉及多个领域的应用，有趣、生动、实用。

本书通过大量与科学相关的应用，例如生态学、天文学、化学等学科的例子，深入浅出地介绍了 R 的基本使用方法，以及建立应用模型和求解这些模型的方法。本书是 R 入门的不二选择，读者只要具备高中代数知识，就能顺利读完本书。

献给 Chris、Ariel、Scott 以及 Ellen

你们是我的骄傲

前言

R 是一个用于科学绘图和计算的计算机软件。它由统计学家和科学家编写并维护，以便科学家们在工作中使用。它易于使用，且有强大的功能。R 正在科学和技术领域迅速传播，并为展示科学出版物中的图形数据建立标准。

R 是免费的开源软件，可在大多数计算机（俗称电脑）上安装。它支持 Windows、Mac 和 Unix/Linux 操作系统。人们可从 R 语言网站（<http://www.r-project.org/>）方便地下载并安装 R 软件。

本书是为高中生、大学生，以及其他想要学习怎样使用 R 的人编写的。借助本书，读者可以充分利用计算机来解决应用数学的相关问题，特别是在涉及大量分析、建模、模拟、统计及绘图的学科中，如物理学、生命科学、化学、地球科学、经济学、工程学、商学等。这些定量研究可以用 R 非常直观、明了地展现出来。同样，复杂的数学和统计学概念若用 R 来说明，也会变得非常清晰。

本书将从基础讲起，不要求读者具有计算机编程背景。读者只需具备一定的高中代数知识就能够理解书中用到的数学知识。

R 让具有图形界面的计算器看上去既难以操作又显得过时。图形计算器操作复杂，难以掌握，并且只能解决很少的问题；同时，它显示的图形较小，分辨率低。相比之下，用 R 来计算既直观又有趣。R 可以帮助用户轻而易举地得到符合出版物质量要求的数据图形或（和）公式。用 R 来代替图形计算器，将使那些正在高中和大学学习科学与数学课程的学生从中受益。

本书将要介绍的 R 知识，足以帮助读者在分析自然与社会科学的定量问题时，提出高级的解决方案。但是，由于 R 是一个庞大的软件，本书并不是一本面面俱到的 R 指南。某些入门手册过于复杂，让人误以为学习 R 是一项“艰巨的任务”。用户若想从零开始成为一名专业的 R 使用者，首先应当找到切入点。

本书将帮助读者从零基础开始学习 R 的使用。本书认为，读者首先应当扎实掌握 R 的基本技巧。这套技巧可以用来解决高中和大学科学学科中的大多数定量研究的问题，同时它也是探索更深奥的问题的基础。独立发现 R 中蕴藏的巨大资源，也是使用 R 的乐趣和回报的一部分。

根据作者的经验，对于那些学习 R 有困难的学生，实际的难点往往在于建模分析过程中所包含的各种数学概念。本书仅仅要求读者具有一定的高中代数知识。部分章节会探索在科学研究中非常有用的一些代数概念，例如二次函数、线性方程组、三角函数及指数函数。每章都会提供一份代数概念的复习指导，以及一份在 R 中进行计算和绘图的操作指南。各章都会描述一些来自原创科学出版物的实际案例，然后介绍如何用 R 来得出这些科学结论。R 为高中生和大学的研究带来了最现代、最前沿的定

量分析方法。

R 是统计学分析软件中当之无愧的佼佼者。但是，R 并不仅限于统计学。它内置了许多高级的数学工具、绘图工具以及模拟工具，是一个可以解决各种科学计算问题的综合性软件包。尽管本书涉及了一些基础的统计学方法，但关注点还是 R 在各个领域更广泛的使用，即将 R 作为一个全面的科学计算和绘图工具。

导致人们认为 R 非常难学的另一个原因在于：现有关于 R 的书籍和网站还涉及统计学和数据分析。然而，使用本书并不需要统计学基础，并且书中会尽量避免使用 R 中内置的统计学程序。统计推断的概念很有挑战性。本书会介绍一些概率的计算与模拟、统计学概论与数据图以及曲线拟合问题，但不会去研究统计推断的概念。事实上，那些已经掌握了本书关于 R 知识的学生，在初次接触统计推断时就会对统计推断有更深刻的认识。

本书不要求读者预先掌握统计学和微积分。很少有学生会在大一之前学过微积分，而在大二、大三之前学过统计学的学生则更少。因此，本书将重点介绍 R 在科学和数学中关于微积分、统计学的先修课程上的使用。在科学、数学以及其他定量分析课程中，所有要使用计算器的地方，R 都可以更好地替代它们。此外，通过使用 R，学生还可以处理更为复杂的科学问题。

使用 R 的学生将在科学课程上获益良多。R 让科学计算和绘图变得既有趣又简单易行。学生在使用 R 时，可专注于科学和数学上的概念，而不用去钻研计算器手册上那些令人望而却步的按键说明。人们没有办法精通并牢记计算器的使用说明，而且这些说明还会随着新计算器的推出而不断变化。相比之下，R 的使用技巧是在不断使用中获得的，这些技巧同样适用于更高级的课程。学生们可以像世界上各个实验室的科学家们一样分析数据和发现规律。

R 鼓励团队合作。就像科学家分享他们的应用程序一样，学生可以小组的形式来开展 R 项目、构建 R 脚本、改进每个人的工作，并收集、累积、清晰地展示经典的图形分析。在小组讨论时，计算机屏幕上显示的结果比计算器上显示的结果更容易阅读，而且在合作中对 R 脚本的保存和修改也比操作计算器的按键要简单许多。在这里，学生也可相互发送最新的 R 脚本来进行在线合作。每个新的班级都可以看到之前班级的工作，并可在此基础上继续改进以取得新的成果。R 自身也在改进中不断发展。

在大学和学院中，R 的使用急速增长，而且一些具有远见的技术公司最近也开始使用 R（参见“Data Analysts Captivated by R’s Power,” by Ashlee Vance, *The New York Times*, January 6, 2009）。通过简单的在线搜索就可以揭示出 R 对顶级大学课程的影响力。对学生们来说，不仅 R 的技术可以伴随他们的整个学习过程，而且 R 的知识也是具有实际价值的专业科学凭证。

网络上有大量使用 R 的在线资源。科学家和教授们会随时分享他们在科学研究中所积累的 R 知识。科学家和课程的主页上到处都有免费的网络课程、入门读物以及参考书。除此之外，读者还可以在关于 R 的论坛里获得对问题的解答、分享 R 的脚本、再现最新科技论文中的计算和图表。但这些资源往往只针对高级且专业的大学课程。

对于初学者来说，这些高深浩繁的材料让人不知道从哪里开始学起。大多数 R 的使用者为了找到自己想要的知识，不得不先处理和过滤一堆杂乱无章的入门材料。

本书在设计时充分考虑了初学者的需求。通过本书，读者可以掌握一系列 R 技术，并可使用这些技术来完成基础科学和应用数学课程上的大多数计算和绘图项目。这些技术都可以在本书中找到，也很容易理解。在学完本书后，读者会对学习更加高级的技术充满信心。

如何使用本书：本书的实践性很强。首先，需要在电脑上安装 R，或是在实验室找到一台已经安装了 R 的电脑。然后，请打开本书并跟随本书讲解的 R 命令来进行操作。读者可将所做的工作保存在电脑硬盘或移动硬盘的一个文件夹中。在学完每章之后，可通过完成一个或几个计算任务来检查自己所学的知识。这些任务不仅仅是练习，它们还是一些微型项目，不论是在创造力上还是在解决问题的能力上，都可以帮助读者建立起使用 R 的思维模式。如果在课堂上使用本书，老师可以把每章后面的任务分配给每个学生或每组学生，让他们将自己的结果分享或汇报给全班同学。本书在最后的附录部分还给出了一些关于 R 的建议，包括如何安装 R，如何获得关于 R 技术的帮助信息以及一个关于 R 的常用命令和选项的清单。

本书中所用到的程序脚本、绘图脚本以及数据集都可以在下面这个网站中找到：
<http://webpages.uidaho.edu/~brian/rsc/RStudentCompanion.html>。

读、写、算，这就是 R!

Brian Dennis
爱达荷州，莫斯科

作者介绍

布莱恩·丹尼斯 (Brian Dennis) 是爱达荷大学鱼类与野生动物科学系、统计科学系的合聘教授。他在宾夕法尼亚州立大学获得统计学硕士学位和生态学博士学位。他曾撰写超过 70 篇科学论文，这些论文涉及统计学和数学建模在生态学和自然资源管理中的应用。在十几年的教学和科研工作中，他一直使用 R。

译者介绍

高敬雅，首都经济贸易大学统计学院硕士毕业，目前为北京师范大学-香港浸会大学联合国际学院统计学教师，主要研究兴趣为金融时间序列分析、多元统计分析及临床实验设计等。在校期间多次参与北京市自然科学基金项目课题，亦曾多次获得国家级和省部级数学建模竞赛奖励和优秀论文奖。

刘波，重庆大学计算机学院博士毕业，目前为重庆工商大学计算科学与信息工程学院教师，主要从事机器学习、计算机视觉、最优化技术以及 Spark 平台下分布式计算的研究，同时爱好 Linux 平台的编程和 Oracle 数据库的开发。现已发表论文 10 余篇，翻译 3 本与计算机相关的书，编写 Oracle 教材 1 本，承担国家自然科学基金青年基金项目 1 项，承担重庆市科委和重庆市教委研究项目 3 项。

目录

第 1 章 介绍：开始使用 R	1
1.1 R 教程	1
1.2 向量	3
1.3 图形	6
1.4 实际案例	7
1.5 本章小结	10
1.6 计算任务	12
1.7 参考文献	15
第 2 章 R 脚本	17
2.1 创建与保存 R 脚本	17
2.2 运行 R 脚本	18
2.3 找到 R 脚本中的错误	19
2.4 利用注释使脚本明了	21
2.5 实际案例	22
2.6 本章小结	25
2.7 计算任务	30
2.8 参考文献	33
第 3 章 函数	35
3.1 在 R 中建立新函数	37
3.2 关于 R 中自定义函数的更多内容	38
3.3 实际案例	39
3.4 本章小结	41
3.5 计算任务	42
3.6 补充说明：案例短评	43
3.7 参考文献	44
第 4 章 基本绘图	45
4.1 实际案例	45
4.2 单变量绘图	49
4.2.1 带状图	49

4.2.2	直方图	49
4.2.3	茎叶图	51
4.2.4	箱线图	51
4.2.5	时序图	52
4.3	双变量绘图	53
4.3.1	散点图	53
4.3.2	并列箱线图	55
4.3.3	条形图与饼图	55
4.3.4	条形图与饼图的数据展示	56
4.4	本章小结	59
4.5	计算任务	61
4.6	补充说明	62
第 5 章	数据输入与输出	63
5.1	R 中的数据框	66
5.2	本章小结	71
5.3	计算任务	73
5.4	补充说明	74
第 6 章	循环	75
6.1	建立 for 循环	76
6.2	检查循环	77
6.3	好吧, 斐波那契先生……那又怎样呢?	77
6.4	实际案例	78
6.5	本章小结	82
6.6	计算任务	82
6.7	参考文献	84
第 7 章	逻辑与控制	85
7.1	逻辑比较运算及逻辑向量	85
7.2	布尔运算	86
7.3	缺失数据	88
7.4	索引及其相关内容	89
7.5	条件语句	91
7.6	实际案例	95
7.7	本章小结	100
7.8	计算任务	103

7.9	补充说明	103
7.10	参考文献	104
第 8 章	二次函数	105
8.1	实际案例	110
8.2	本章小结	113
8.3	计算任务	116
8.4	参考文献	116
第 9 章	三角函数	117
9.1	直角三角形	117
9.2	三角函数	118
9.3	直角三角形, 圆形与弧	119
9.4	三角函数的特性	123
9.5	极坐标	124
9.6	距离的三角测量	126
9.7	实际案例	127
9.7.1	太阳系附近恒星的距离	127
9.7.2	抛体运动	127
9.7.3	天体轨道	129
9.8	本章小结	130
9.9	计算任务	132
9.10	补充说明	132
第 10 章	指数函数与对数函数	133
10.1	实数指数幂	133
10.2	特殊的数字 e	135
10.3	数字 e 的应用	137
10.4	指数函数	138
10.5	指数增长	139
10.6	对数函数	140
10.7	对数尺度	143
10.7.1	里氏震级	143
10.7.2	pH 值	143
10.7.3	恒星等级	144
10.8	实际案例	145
10.8.1	放射性衰变	145

4 R 语言初学指南

10.8.2	种群增长的极限	147
10.8.3	石油顶峰	150
10.9	本章小结	151
10.10	计算与代数任务	153
10.11	参考文献	156
第 11 章	矩阵运算	157
11.1	向量相乘的另一种方式	157
11.2	矩阵乘法	158
11.3	矩阵的加减运算	161
11.4	将数据文件读取为矩阵	162
11.5	实际案例	162
11.6	本章小结	165
11.7	计算任务	166
11.8	补充说明	167
11.9	参考文献	167
第 12 章	线性方程组	169
12.1	矩阵表示	169
12.2	矩阵的逆	170
12.3	R 中的矩阵求逆和方程组的解	172
12.4	现实中的例子	174
12.4.1	老忠实泉	174
12.4.2	一个不远的星系	180
12.5	本章小结	183
12.6	计算任务	184
12.7	补充说明	186
12.8	参考文献	187
第 13 章	高级绘图	189
13.1	绘制二维图形	189
13.2	符号、线与坐标轴风格的选项	190
13.2.1	数据符号类型	190
13.2.2	连接线类型	190
13.2.3	曲线类型	191
13.2.4	坐标轴的限制	191
13.2.5	刻度线	191

13.2.6	坐标轴标签	191
13.2.7	不显示坐标轴	191
13.2.8	符号与标签的大小、线与坐标轴的宽度	192
13.3	其他自定义功能	192
13.3.1	添加点	192
13.3.2	添加线	192
13.3.3	增加文本	193
13.3.4	标题和副标题	194
13.3.5	图例	194
13.3.6	新图形窗口	194
13.3.7	全局性与局部性	195
13.4	多屏图	195
13.5	三维图	197
13.6	颜色	200
13.7	本章小结	201
13.8	计算任务	202
13.9	参考文献	203
第 14 章	概率与模拟	205
14.1	随机变量	205
14.2	概率	206
14.3	离散概率分布	208
14.4	连续概率分布	212
14.4.1	均匀分布	212
14.4.2	正态分布	214
14.5	实际案例	218
14.6	计算任务	223
14.7	补充说明	224
14.8	参考文献	224
第 15 章	拟合数据模型	225
15.1	随机变量	225
15.2	多元预测变量	228
15.3	非线性统计方法	230
15.4	本章小结	235
15.5	计算任务	239
15.6	补充说明	241

15.7 参考文献	242
第 16 章 结论——你不需要成为一名火箭学家	243
16.1 真实的太阳系的例子	243
16.2 问题	243
16.3 概念	244
16.4 速度的变化	245
16.5 移动地球	246
16.6 组织整理	247
16.7 计算地球轨迹的 R 脚本构思	247
16.8 R 脚本	249
16.9 计算任务	251
16.10 补充说明	252
16.10.1 微积分和圆锥曲线	252
16.10.2 Feynman 最后的讲座	252
16.10.3 三体问题	253
16.10.4 海王星	253
16.10.5 误差的传播	253
16.10.6 阿波菲斯	253
16.10.7 混乱的冥王星轨道	254
16.10.8 忽略质量	254
16.10.9 水星运动轨道和广义相对论	254
16.10.10 测量单位	254
16.11 参考文献	255
附录 A 安装 R	257
附录 B 获得帮助	259
附录 C R 的常用命令和选项	261
译后记	273

第 1 章

介绍：开始使用 R

R 是一种用来进行科学绘图和计算的计算机程序。

R 由科学家编写，供他们在工作中使用。

R 具有非常强大的功能，且十分易于使用。

R 是免费的。

记得之前提到过 R 是免费的吗？

R 有不同的版本，适用于各种操作系统，比如，Windows, Mac，甚至包括 Unix/Linux 系统。如果读者在家有台能联网的电脑，便可以从这个网站 <http://www.r-project.org/> 下载并安装 R。

R 的安装比电脑游戏还要简单（若需安装帮助，可以参考附录 A）。

安装 R 之后，会在电脑桌面上或程序菜单中看到一个蓝色字母“R”的图标，双击它就可打开 R 程序，其控制台窗口“R console”就会弹出来了。

在控制台上，会看到命令提示符“>”和其后面闪烁的光标。现在就可以在 R 中执行命令了！只需按照本书的教程，在提示符后面输入命令即可，或输入老师在投影仪上展示的命令。在这之后，可独自或以团队的形式一起完成本章最后给出的计算任务。你可将其结果展示给全班同学。准备好了吗？那就开始吧。

1.1 R 教程

使用 R 的最简单方式就是把它当成一个强大的计算器。在提示符处输入 $5+7$ ，然后敲击回车键：

```
> 5+7  
[1] 12
```

这个结果的第一部分是 12。稍后会看到有些结果会包含许多个部分。所以 R 在给出结果时会对其中的某些部分编号。

下面来试一下减法运算。每次在提示符处输入字符后，敲击回车键，就会在下一行得到相应的答案：

```
> 5-7  
[1] -2
```

2 R 语言初学指南

R 可以识别负数。例如：

```
> 5+-2
[1] 3
```

这里用星号 “*” 可以进行乘法运算：

```
> 5*7
[1] 35
```

除法运算的符号是斜线 “/”，所以 5 除以 7 为：

```
> 5/7
[1] 0.7142857
```

跟计算器一样，R 可以进行小数运算。

做幂运算：回忆一下“5 的 7 次方”可以写为 5^7 ，表示 $5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5$ 。R 可将其计算出来。这里用脱字符 “^” 表示乘方运算。因此，5 的 7 次方为：

```
> 5^7
[1] 78125
```

可把一系列运算放在同一个命令中。乘法运算和除法运算的优先级要高于加法运算和减法运算，在计算同级别运算时遵循从左到右的原则：

```
> 5+7*3-12/4-6
[1] 17
```

读者可手算一下，看得到的结果是否和上面一样。并且，幂运算会优先进行，它的优先级要高于乘法运算和除法运算：

```
> 1+4*3^2
[1] 37
```

可使用括号来改变运算顺序：

```
> (5+7)*3-(12/4-6)
[1] 39
```

当括号内部还有括号时，程序会优先计算最内层的括号！在输入程序时，一定要确保每个左括号 “(” 都有一个对应的右括号 “)”：

```
> (5+7)*3-(12/(4-6))
[1] 42
```

R 可将运算结果保存到用户命名的变量中：

```
> sally=5+7
> ralph=4-2
> sally-ralph
[1] 10
```

在命名变量时需注意，R 会区分大小写。在 R 中，sally 和 Sally 是不一样的。

可看一下变量 ralph 和 sally 的值：