HZ BOOKS
华章科技

R Cookbook

R语言经典实例

O'REILLY®

机械工业出版社
China Machine Press

Paul Teetor 著

李洪成 朱文佳 沈毅诚 译

R语言经典实例

Paul Teetor 著

李洪成 朱文佳 沈毅诚 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc.授权机械工业出版社出版

机械工业出版社

图书在版编目（CIP）数据

R语言经典实例 /（美）蒂特（Teetor, P.）著；李洪成，朱文佳，沈毅诚译.
—北京：机械工业出版社，2013.4
（O'Reilly精品图书系列）

书名原文：R Cookbook
ISBN 978-7-111-42021-7

I. R… II. ①蒂… ②李… ③朱… ④沈… III. 程序语言－程序设计 IV. TP312
中国版本图书馆CIP数据核字（2013）第066465号

北京市版权局著作权合同登记
图字：01-2011-7869号

书　　名/　R语言经典实例
书　　号/　ISBN 978-7-111-42021-7
责任编辑/　盛思源
封面设计/　Karen Montgomery，张健
出版发行/　机械工业出版社
地　　址/　北京市西城区百万庄大街22号（邮政编码 100037）
印　　刷/　藁城市京瑞印刷有限公司
开　　本/　178毫米×233毫米　16开本　26.5印张
版　　次/　2013年5月第1版　2013年5月第1次印刷
定　　价/　79.00元（册）

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：(010)88378991；88361066
购书热线：(010)68326294；88379649；68995259
投稿热线：(010)88379604
读者信箱：hzjsj@hzbook.com

O'Reilly Media, Inc.介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

中文版序

R软件的应用正变得越来越广泛。在雇主招聘需求所要求的技能中，以前只要求具备SPSS技能和SAS技能，现在（2012年6月写这个序言的时候）的招聘需求中对应用R软件的技能更加关注。在美国统计教育学院*Statistics.com*的在线课程中，R课程是增长最快的。R的流行经历了三个阶段。在第一个阶段，R先是由于它的开放源代码和免费而在学术界变得流行；第二个阶段，应用R的相关专业的研究生把R带到他们的工作环境中；第三个阶段，由于研究生把R的火种带到工作中，商业分析领域中非统计专业研究生的其他分析人员也开始试着应用R。从地域上讲，R最早发源于新西兰，然后传播到了美国和欧洲，现在R在中国也逐渐变得广为人知。

本书可以帮助你迅速提高R的技巧。你可以不必按照作者认为重要的冗长主题进行阅读，你可以立即专注于你的工作任务，学习如何用R来解决这些任务。这些任务都在书中有明确的标识，它们组织精炼，有的任务只有一页或者两页，便于查找，方便应用。

如果你寻求解决某个特定任务的方法，你可以在因特网上搜索。那么在本书中寻找解决方案的优势在哪里呢？其一，你可以有信心本书中提供的解决方案是合理的；其二，本书有一致的风格，你可以很容易地找到你需要的内容并应用它们。

大多数烹饪的人是根据菜谱来学习烹饪，或者应用在新情况下获取的知识来学习烹饪，没有人通过阅读食物的历史并学习它们的化学成分来学习烹饪。同样的道理，随着你的R知识的拓展，你会发现本书提供了相当有效的“菜谱”和建议。

——彼得·布如斯

美国统计教育学院*Statistics.com*总裁

译者序

本书的英文版从出版后就在亚马逊美国网站上收到了极高的评价。机械工业出版社以极快的速度引进这本书的中文版，使国内读者在原版出版一年左右的时间里读到中文版，不得不赞扬他们独到的眼光。这本书中文稿完成的时候（2012年6月），其英文版的销量还是排在专业书籍的前列，并稳居生物信息学类书籍的第一位，排在建模和仿真类书籍的第7位。

本书的作者Paul Teetor是一位统计学专家和计算机专家，他同时也是一位量化投资分析和风险管理专家。他和大芝加哥地区的对冲基金、投资组合经理一起工作，因此有丰富的投资量化分析经验。他把R软件用户在学习和应用R软件中常见的问题进行系统整理和总结，汇集成了本书。书中涵盖了从R软件的基础知识（安装、帮助系统、解决实际问题的途径、R数据结构、R编程、R的输入和输出等）到用R进行数据分析的具体方法（数据变换、概率和统计基础、R绘图、回归分析和方差分析、R常用技巧、高级数据分析方法和时间序列分析等）。全书以问题、解决方案和对解决方案的讨论与拓展为主线来组织内容。读者既可以把本书作为学习R的一本优秀教材，也可以根据自己的需要参考书中的某些具体方法，找到自己实际问题的解决方案。

R本身是一款十分优秀的统计分析软件，R的书籍和文档也是相当多的。但是缺乏一本适合R初学者的书籍，尤其是针对那些对R不甚精通但是急切需要用R来解决问题的R用户。本书以问题和解决方案的形式组织内容，脉络清晰，读者很容易找到自己需要的内容。不管是R初学者，还是熟练的R用户都能从书中找到对自己有用的内容。

本书的译者从2003年年初接触R，那时国内几乎没有多少R用户。现在情况发生了极大的变化，R成为数据分析从业者谈论最多的软件之一，许多学校的统计系在教学中也广

泛地应用R。从2008年开始，国内的R用户每年召开R用户大会，大力传播和推广R。从2011年开始国内出版社也开始出版和引进R书籍，这对国内R的推广都起到了极大的作用。

在本书的翻译过程中，得到了原作者Paul Teetor的大力帮助，他解答了译者在翻译过程中遇到的各种问题。美国统计教育学院*Statistics.com*是目前世界上开设R在线课程最多的机构，十分感谢他们的总裁彼得•布如斯先生欣然为本书中文版作序。另外，十分感谢机械工业出版社的王春华老师的大力支持和帮助，感谢盛思源编辑对本书一丝不苟的校对与检查。本书的翻译工作由李洪成、朱文佳和沈毅诚共同完成，由于水平所限，中间可能会有翻译不当之处，希望读者多加指正。

——李洪成

目录

前言

R软件是进行统计分析、绘图和统计编程的强大工具。现在成千上万的人用它来进行日常的重要统计分析。R软件是一个自由、开源的软件平台，它是许多聪明、勤奋工作的人们集体工作的成果。R软件有超过两千多个软件包插件。R软件是其他所有商业统计软件包的强劲竞争对手。

但是，开始使用R软件可能感到无从下手。对于许多任务，即便是一些基本的任务，R的实现也不是很明显。当了解了R的使用方法后，简单的问题自然能得心应手地解决，但学习“如何”使用R的过程有时会让人感到发狂。

本书介绍了如何使用R软件的一些方法，其中每一个方法对应解决某个特定的问题。介绍这些方法的途径是这样的：首先给出待解决的问题，然后给出解决方案的简单介绍，之后再给出对解决方案的讨论，深入剖析解决方案，给出该方案的原理。我知道这些方法有效实用，我也知道这些方法可行，因为我本人也使用它们。

这些方法所涉及的范围较为广泛。首先从基本的任务开始介绍，然后介绍数据的输入和输出、基础统计、绘图以及线性回归。与R有关的工作都将或多或少地涉及本书介绍的方法。

通过本书的讲解，初学者能快速地了解R并获得提高。如果你对R软件有一定的了解，那么本书也能帮助你巩固已学的知识，拓宽你的思维（例如，“下一次我应该怎么使用K-S检验”）。

从严格意义上来说，本书并不是一本关于R软件的教程，但你将会从中学习到许多R软件的应用技巧。本书也不是一本关于R的参考手册，但它确实包含了许多实用的内容。本书也不是一本R软件的编程指南，但书中很多方法都可以应用到R的编程脚本中。