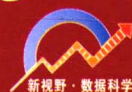


 Springer

*R*语言应用系列



*R*语言在商务分析中的应用

R for Business Analytics

[印] A.奥赫里 著

朱 钰 译

A. Ohri



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

R语言应用系列

R for Business Analytics

R 语言在商务分析中的应用

[印] A. 奥赫里 (A. Ohri) 著
朱 钰 译



西安交通大学出版社

Xi'an Jiaotong University Press

Translation from the English language edition:
R for Business Analytics by A. Ohri
Copyright©2013 Springer New York
Springer New York is a part of Springer Science+Business Media
All Rights Reserved

本书中文简体字版由施普林格科学与商业传媒公司授权西安交通大学出版社独家出版发行。
未经出版者预先书面许可,不得以任何方式复制或发行本书的任何部分。

陕西省版权局著作权合同登记号 图字 25-2013-110 号

图书在版编目(CIP)数据

R 语言在商务分析中的应用/(印)奥赫里著;朱钰译.
—西安:西安交通大学出版社,2016.4
(R 语言应用系列)
书名原文:R for Business Analytics
ISBN 978-7-5605-8303-7

I. ①R… II. ①奥… ②朱… III. ①程序语言-
程序设计-应用-商务-经济分析 IV. ①F7-39

中国版本图书馆 CIP 数据核字(2016)第 032904 号

书 名	R 语言在商务分析中的应用
著 者	[印] A. 奥赫里
译 者	朱 钰
责任编辑	李 颖

出版发行	西安交通大学出版社 (西安市兴庆南路 10 号 邮政编码 710049)
网 址	http://www.xjtupress.com
电 话	(029)82668357 82667874(发行中心) (029)82668315(总编办)
传 真	(029)82668280
印 刷	陕西宝石兰印务有限责任公司

开 本	720mm×1000mm	1/16	印张	18.25
印 数	0001~2000 册		字数	337 千字
版次印次	2016 年 6 月第 1 版	2016 年 6 月第 1 次印刷		
书 号	ISBN 978-7-5605-8303-7/F·587			
定 价	58.00 元			

读者购书、书店添货、如发现印装质量问题,请与本社发行中心联系、调换。

订购热线:(029)82665248 (029)82665249

投稿热线:(029)82665397

读者信箱:banquan1809@126.com

版权所有 侵权必究

译者序

《R 语言在商务分析中的应用》这部书是印度的 A. Ohri 先生根据自己 10 余年的商务分析师的职业生涯经历和经验,倾力打造完成的。书中各章所涉及的数据分析任务都是商务数据分析工作中最常见的任务。A. Ohri 先生还特意添加了多个数据源作为参考文献以飨读者。

作为一部商务分析的实用书籍,其价值体现在两个方面:首先,在商务数据分析方面,本书丰富的实例对于从事商务数据分析(国内也叫业务数据分析)的从业者以及商科的本科生、硕士研究生具有很强的指导性和实用价值。对于经济统计学和应用统计学专业的本科生、硕士研究生也同样具有很高的从业指导价值和学习价值;其次,在 R 语言的应用方面,这部书也提供了足以使读者目不暇接的丰富的应用程序包、插件和链接。

这部书适用于经济管理类几乎所有专业的专科、本科、硕士研究生及统计学类的专科、本科、硕士研究生自修提高,也适合以上提及的各个专业及层次用作专业教材和教参,对于工商管理硕士和统计专业硕士而言,本书更是一部不可多得的教学参考材料。

这部书的翻译工作由西安财经学院师生共同努力完成。西安财经学院朱钰教授对前 7 章进行了初译,统计学硕士研究生朱子项对后 6 章进行了初译,在些基础上朱钰对全书进行了统稿。统计学硕士研究生陈晓茹、王磊、王檬、谢茂茂、张洁、王一维、张娟、林喆、张鹤鸣,本科生曾媛、孙媛、杨欣、杨梅、张蓉、马静、张春秘、王蕾等均参加了翻译辅助工作。

由于译者水平所限,本书在翻译上一定会存在各种问题,在此恳请读者批评指正。

朱 钰

2016 年 1 月于西安财经学院

作者简介

A. Ohri 是决策统计网站 Decisionstats.com 的创始人。他曾在田纳西大学诺克斯维尔分校和印度管理学院勒克瑙分校攻读硕士学位。此外,Ohri 还拥有德里工程学院机构工程学位。他曾访问了 100 多名商务分析从业人员,包括所有分析软件供应商的主要成员。Ohri 在他的博客上写了近 1300 篇文章,除此以外他受邀为具有影响力的分析社区写作。他通过在线教育讲授 R 课程,在过去的十年中一直作为分析顾问在印度工作。Ohri 是印度最早的独立分析顾问,他目前的研究兴趣包括从传播开源分析、分析社交媒体操纵、简单界面到云计算和非正统的加密。

前言

我基本上是按照我将近 10 年的商务分析师的职业生涯中最常执行的任务来组织这部书的。因此每章都是最常见的任务,并且我添加了多个数据源作为参考文献以便在更深的层次上帮助读者探索一个特定的主题。我再一次强调,这是一部关于商务分析的书,而不是关于统计的书,我自己作为一个 MBA 的经历和 MBA 在商务分析(尤其是 R)方面的文献指引我做出了这样的选择。因此这部书是为商务分析师量身定制的,而不是为统计人员编写的。它不会帮助你在你的研究生论文上取得更好的成绩,但它肯定会帮助你获得或留住一份分析的工作。如果你是一名正在学习 R 的学生,它可以帮助你更快地完成课外作业。

我相信,在当前的商业环境,焦点将转回分析师而不是软件工具,并且拥有多个平台的技能,尤其是既高端又兼具低成本的分析平台,对于用户会有好处。

这本书将明确地关注图形用户接口、技巧、秘诀、技术和快捷方式,并将精力集中于商务分析师在日常生活中面临的最常见任务的案例研究。内容安排将会尽可能的简单但又不是过于简单。每一章都将有一个案例研究、教程或示例问题。每一章末尾列出所用的函数和程序包以帮助读者复习。有时候可能有一些函数会被重复或被再次强调,这更多的是由于他们在分析中的用途和必要性。对创造者、作者和开发人员简短访谈的目的是便于商务分析师更好地掌握 R 的相应内容。现有的 R 资源使用的案例旨在帮助分析社区中尚未采用 R 的决策者给 R 一个机会。

这本书有一个务实的目的,是针对那些已经或希望在企业的商务环境中使用 R 的读者。本书将提供足够的参考文献,在理论或先进水平方面给读者以帮助,并在他或她的 R 之旅中以其他可用的资源予以支持。

我们鼓励读者跳过那些没有直接相关性的章节,直接跳到那些对他们的

目的有最大效用的章节。我面对的一个问题是,在我写此书的过程中,R项目发布了几近4个新版本,所以如果有任何不准确的地方或错误请告知我。

本书内容安排

第1章:为什么要用R?讲述在商务分析中使用R的理由。在这一章中,我们讨论选择R作为分析平台而不仅是统计计算平台的原因,与其他分析软件的比较及在商业环境中使用R的一些一般性的成本和收益。它列举了业务分析师应该选择学习R的各种原因和其他分析平台所没有的R的特性。R的益处被分成三大类:商务分析、数据挖掘和商务智能/数据可视化。

第2章:R的基础配置:讲述如何设置R分析的基础配置。在这一章中,我们讨论如何实际地在基于R的环境中设置一个分析环境,包括硬件、操作系统、额外的软件、必要的预算以及培训和软件支持的需要。我们讨论各种操作系统、硬件选择以及各种基于R的解决方案的提供者。内容涉及R的基本安装、R的程序包库、R的更新,直到综合性的用户帮助和启动R的基本教程。

第3章:R界面:根据你的需要使用R Analytics的方法。在这一章中,我们对照比较了各种与R分析平台互动的方式。可以通过命令行、图形用户界面(GUIs)使用R,也可以通过网络接口(包括云计算接口)使用。这一章概述了基于GUI方法的优缺点,还比较了九种GUI的特性,包括一个比较优势和劣势的汇总表。本章还讨论了从其他软件和web接口使用R的方式。最后,介绍了一些在亚马逊云上运行R的实用教程。

第4章:数据处理:在R中取得数据。这一章讨论在R中获得数据的不同方式,包括基本的语句。它专门针对数据库中的数据,因为商务数据通常都是这样的。这一章展示了用户怎样可以连接到MySQL和Pentaho数据库,这是两个最先使用的开源数据库。某些节致力于和R一起使用SQLite,并奉献给商务智能从业人员。我们简要地考察一下Jaspersoft和Pentaho——两个在BI领域领先的开源解决方案,以及它们如何与R互动。这一章在简要提到处理大数据集的额外资源的同时,它也道出了一系列常见的分析任务(人们期望分析师对任何数据都要进行的操作),以帮助一些人更好地过渡到R。

第5章:探索数据:数据可视化中新崛起的业务。这一章讨论在R中使用视觉和图形手段探索数据。它讲述了R中基本的图形和一系列高级的图形,可以很容易地由一个学习R的普通程序员在很短的时间创建出来。它引入了专门用来进行数据探索的GUI,grapheR和Deducer,还包含对ggplot2创造者

哈德利·威克姆(Hadley Wickham)的访谈。制图方面包含了代码、容易重现的例子和互动图示的信息。本章旨在为那些主要使用电子表格程序创建图示的读者揭开神秘的、有时令人生畏的数据可视化艺术的面纱。

第6章:构建回归模型。回归模型是商务分析行业中统计的主力。他们或许因为固有的与业务客户沟通的简单性而过度使用回归模型。我们学习如何建立线性和逻辑回归模型,研究一些风险模型和记分卡,讨论 PMML 作为模型实施的方法。为了保持本书关注商务分析的目的,本书仅涉及在 R 中简化构建逻辑斯蒂模型流程的一个简短的案例研究,并没有介绍回归理论。

第7章:使用 R 进行数据挖掘。使用 R 进行数据挖掘采用 Rattle GUI 简化并加速了数据建模的过程。然而,它从给读者介绍信息阶梯和各种数据挖掘方法的概念入手,简要地说,包括 CRISP-DM, SEMMA 和 KDD。它还提供与两位作者的简短访谈的摘录,这两位都曾写过使用 R 进行数据挖掘方面的书。文本挖掘、网络挖掘、谷歌预测应用程序接口也作为特殊的情况包含在本章内。

第8章:聚类和数据分割。数据分割在这本书中主要涉及聚类分析,我们会讨论各种类型的聚类。在这里增加聚类是因为大数据商务环境中对于数据压缩技术的固有的、日益增长的需求而且数据集的大小也在飞速地增长。我们再次借助 Rattle GUI,但仍简要讨论了 R 中的其他用于聚类的 GUI。对关于 Revolution R 在大量数据的聚类中的应用提供了一个小案例。

第9章:预测和时间序列模型。企业使用商业智能来了解其过去和现在的经营情况,他们的重点是改进对未来的决策。时间序列和预测很强大,但在很多企业都未能得到充分应用,这一章对这个问题进行了讨论并提供了一个使用 R 命令行 GUI 程序包的相应案例。我努力使这一章成为一个实用的章节,用来帮助你的商业团队在所有部门中利用现有的初期数据做更多的预测。

第10章:数据导出和输出。获得分析的数据只是工作的一部分。输出结果应该以一种能够启发决策者制定可行的决策的方式来展示。R 提供了许多灵活的方式来生成和嵌入输出,这一章都有介绍。

第11章:优化 R 代码。现在,你已经学会了如何在商务分析中使用 R 的功能,下一步是要了解如何利用其强大的灵活性,而不要被淹没在其庞大的可用库中。这一章讨论技巧、变通和工具,包括使用代码编辑器来帮助你更好、更快地编写代码。

第 12 章:更多的培训文献。这一章对于有兴趣扩展关于 R 的知识和完整了解 R 文档环境的读者来说是必要的。

第 13 章:R 应用案例研究。这一章提供了基于各种商务应用的编码的案例研究,包括网页分析,旨在帮助读者在他们自己的经营活动环境中使用 R 进行商务分析时找到一个现成的参考资料。

序 言

在 R 项目网站上有一个包含 115 部书的清单,在那个清单最近更新以后又有很多部书出版,如果你想知道为什么还需要再增添一部,好吧,我来说说为什么。我在商务分析领域工作已有近 10 年了。你可能知道统计计算、数据挖掘、商务智能,或者,最近的数据科学等领域。我通常更喜欢决策科学这个术语,但总的来说,该领域是指使用数据、统计技术和专门的工具来协助在政府、研究和工业领域的决策者提高洞察力,以达到积极成果最大化和成本最小化。

从职业生涯的角度看,我发现商务分析领域既非常愉快也有利可图。曾经对同样的人销售信用卡和贷款并获得更多的收入非常有意思,至少在早些年对我来说是那样。我发现商务分析有趣的原因是,它结合了调查学科和带有商业洞察力的勤奋思考。

我还发现商务分析领域有点混乱。有两种类型的人们在怎样应用商务分析上持几乎相反的看法:技术为主的人们(如计算机科学和统计学专业)目标在于获得稳健的结果,而商业为主的人们则瞄准了收入以及更快的结果。在两派偶尔的交火中我被夹在中间。

当我开始学习这一切时——2001 年——主要的平台是 SPSS 和 SAS 语言。2007 年,在我的第一个儿子出生之后,我发现 R 编程语言几乎是保持我的羽翼未丰的咨询公司能够开下去的一个必要工具,因为我需要一个可靠的分析平台但年费又不能太高。在 30 岁的年纪开始启动一个分析事业,我发现我负担不起那些我一直在使用的、一些世界上最大的公司(并且发现,在印度没有单独为小企业的折扣)开发的工具。所以我转到了 R,因为它是免费的,我就是这样想的。

但学习 R 花了我大量的时间,时间可不是免费的——我,或是这个语言,或是整个世界出现了一点问题。与此同时,我开始写一个分析方面的博客,称

为 DecisionStats(<http://decisionstats.com>)以便与其他商务分析师联络,每个月有近 20 000 次浏览和近 100 个与分析从业者领军人物的访谈,我逐渐获得了对于商务分析领域更多的实际见解。

经过一段时间,我慢慢地沉浸于 R,我发现它像所有的语言,都有自己的技巧和技术用于更快地完成任务,我确实不需要记忆大量的代码,也不必被那些能够记住的人吓倒。有几本好书帮我,包括 Bob Muenchen 所著的《SAS 和 SPSS 用户怎样使用 R》,还有诸如 David Smith 和 Tal G. Overall 的极好的博客,我以比我最初的预期更快的速度学会了 R,多亏了创造了 R Commander 的 John Fox 教授,创造了 Rattle 的 Graham Williams 教授,和许多其他为 R 的发展作出贡献的那些人。R 社区的规模、可接受性和组织都在极速扩张,我不再是单纯以分析为目的而孤立作战去学习 R 的一个人。

在 2012 年,R 以我在 2007 年时无法想象的速度发展。我受宠若惊而又幸运地得知 R 是最主要的统计计算语言,由领先的技术公司,包括微软、SAS 研究所、甲骨文(Oracle)、谷歌和其他机构使用并完全支持。没有一个研发部门可以与全世界几乎所有的统计系及免费为此项目工作的教授们相匹敌。

如果你是一个决策者,正在考虑将 R 与你现有的分析基础设施结合使用,你会发现 R 的各种合作伙伴和贡献者的简短访谈是很有启发性、很有助益的。在本书中添加它们是专门用来增加本书的可读性的,方便那些喜欢文字而非编码的商务分析领域的读者学习、理解。

我要感谢 JMP 的 Anne Milley, Baseline Consulting 公司的 Jill Dyche, 田纳西大学的 Bob Muenchen, Rexer Analytics 公司的 Karl Rexer 和 KD Nuggets 公司的 Gregory Shapiro, 感谢他们在我的漫漫分析之路上对我的指导。我还想感谢施普林格美国代表处的 Marc Strauss 和 Susan Westendorf, 还有 Leo Augustine 的帮助。

关注 Decisionstats 博客的网友非常耐心,深深感动了我,我同时感谢 R 和更加广泛的分析社区对作者表现出的关心。

印度德里
A. Ohri

目 录

译者序

作者简介

前言

序言

第 1 章 为什么要用 R	(1)
1.1 为什么把 R 划分为一个完整的分析环境	(1)
1.2 与其他分析程序包相比 R 的其他优势	(2)
1.3 R 作为统计语言和分析平台的差异	(2)
1.4 使用 R 的成本和收益	(3)
1.5 同时使用 SAS 和 R	(4)
1.6 简短的采访:用 JMP 实现 R	(5)
第 2 章 R 的基础配置	(8)
2.1 为商务分析设置 R 的一些选择	(8)
2.2 下载和安装 R	(12)
2.3 安装 R 程序包	(12)
2.4 在 R 中启动教程	(16)
2.5 R 的数据类型	(17)
2.6 对约翰·福克斯——R 的 Rcmdr GUI 的创造者的简短访谈	(18)
2.7 本章用过的命令小结	(19)
第 3 章 R 界面	(21)
3.1 R 统计语言的界面	(21)
3.2 R 基础知识	(21)
3.3 R 图形用户界面的优点和局限性	(23)
3.4 GUI 清单	(23)
3.5 R GUIs 总结	(29)
3.6 在其他软件中使用 R	(29)

3.7	R 的 Web 界面	(32)
3.8	访谈:将 R 用作基于 web 的应用程序	(35)
3.9	使用 R 进行云计算	(35)
3.10	谷歌和 R	(41)
3.11	访谈:在谷歌中使用 R	(45)
3.12	访谈:在 cloudnumbers.com 通过云计算使用 R	(47)
3.13	本章用过的命令小结	(48)
第 4 章	数据处理	(51)
4.1	分析式数据处理挑战	(51)
4.2	较小数据集的读取方法	(53)
4.3	一些常见的分析任务	(58)
4.4	使用 R 进行一个简单的分析	(67)
4.5	比较 R 图形用户界面的数据输入	(70)
4.6	与数据库及商业智能系统一起使用 R	(71)
4.7	本章用过的命令小结	(80)
4.8	引文和参考文献	(81)
4.9	更多的资源	(82)
第 5 章	探索数据	(94)
5.1	商务指标	(94)
5.2	数据可视化	(94)
5.3	图示的参数	(95)
5.4	在 R 中创建图	(96)
5.5	使用 ggplot2 画更好的图	(139)
5.6	交互式的图	(140)
5.7	GrapheR:绘制简单图的 R GUI	(141)
5.8	Deducer:高级数据可视化 GUI	(142)
5.9	调色板	(147)
5.10	访谈:哈德利·威克姆(Hadley Wickham), <i>ggplot 2: Elegant Graphics for Data Analysis</i> 的作者	(149)
5.11	本章用过的命令小结	(150)
第 6 章	构建回归模型	(155)
6.1	线性回归	(155)
6.2	逻辑斯蒂回归	(156)
6.3	风险模型	(156)

6.4	评分卡	(156)
6.5	在 R 中构建回归模型的一些有用的函数	(157)
6.6	使用 R Cmdr 建立回归模型	(158)
6.7	回归模型的其他程序包	(167)
6.8	PMML	(167)
6.9	本章用过的命令小结	(169)
第 7 章	使用 R 进行数据挖掘	(171)
7.1	定义	(171)
7.2	Rattle: R 中的一个数据挖掘 GUI	(176)
7.3	访谈: 格雷厄姆·威廉姆斯(Graham Williams), <i>Data Mining with Rattle and R</i> 的作者	(187)
7.4	使用 R 进行文本挖掘分析	(188)
7.5	谷歌预测 API	(194)
7.6	数据挖掘者的数据隐私	(196)
7.7	本章用过的命令小结	(196)
第 8 章	聚类和数据分割	(198)
8.1	什么时候使用数据分割和聚类	(198)
8.2	R 对聚类的支持	(198)
8.3	使用 RevoScaleR 进行循环分析	(199)
8.4	一个称为 Playwith 的 GUI	(200)
8.5	使用 Rattle 进行聚类分析	(202)
8.6	本章用过的命令小结	(210)
第 9 章	预测和时间序列模型	(212)
9.1	时间序列简介	(212)
9.2	时间序列和预测方法	(212)
9.3	时间序列模型的种类	(217)
9.4	处理日期-时间数据	(218)
9.5	使用 R commander GUI 的 epack 插件	(219)
9.6	本章用过的命令小结	(226)
第 10 章	数据导出和输出	(229)
10.1	本章用过的命令小结	(232)
第 11 章	优化 R 代码	(233)
11.1	有效编码的例子	(233)

11.2	定制 R 软件启动	(235)
11.3	代码编辑器	(236)
11.4	增强的代码编辑器的优点	(240)
11.5	访谈:J. J. Allaire, RStudio 的创造者	(240)
11.6	Revolution R——一个高效的环境	(241)
11.7	评估代码效率	(242)
11.8	使用 system.time 来评估编码效率	(244)
11.9	使用 GUIs 学习并更快编写 R 代码	(245)
11.10	并行编程	(245)
11.11	使用硬件解决方案	(246)
11.12	本章用过的命令小结	(246)
第 12 章	更多的培训文献	(248)
12.1	Cran Views	(248)
12.2	阅读材料	(249)
12.3	R 中使用的其他 GUIs	(250)
12.4	本章用过的命令小结	(255)
第 13 章	R 应用案例研究	(256)
13.1	使用 R 进行网络分析	(256)
13.2	使用 R 进行社交媒体分析	(259)
13.3	使用 R 进行 RFM 分析	(262)
13.4	使用 R 建立倾向模型	(263)
13.5	使用 R 建立金融风险模型	(264)
13.6	使用 R 进行药物分析	(264)
13.7	分析方面的阅读文选	(264)
13.8	商务分析师为什么应该学习 R 语言	(267)
13.9	分析生涯	(268)
13.10	本章用过的命令小结	(270)
索引		(271)

第 1 章

为什么要用 R

本章提要:在这一章中把 R 介绍给读者,讨论选择 R 作为分析工具而不仅是 P.1
作为统计计算平台的原因,并与其他分析软件相比较,向读者展示了在商业环境中
使用 R 的各种成本和收益。

由于 R 基本上是 S 语言的一个开源的衍生物和后代语言,所以 R 也被称为
GNU S(具有通用公共许可的 S)。R 已经以各种形式存在了近二十年,它的专业
化的数据可视化、数据分析和数据处理软件库也一直在扩大。R 拥有大约二百万
用户,它有着最大的统计算法和软件包库。

虽然 R 最初是统计计算语言,然而在 2012 年,你可以称它为一个完整的分析
环境。

1.1 为什么把 R 划分为一个完整的分析环境

出于以下原因,R 可被当做是一个完整的分析环境。

- 多平台多接口输入命令:从命令行到大量的专用图形用户接口(GUIs)(见第 2
章)R 有多个接口提供桌面工作环境。对于群集计算、云计算和远程服务器环
境,R 现在有大量的软件包,如 SNOW, RApache, RMpi, R Web 和 Rserve。
- 软件兼容性:众多的商业供应商,包括软件公司都已经开发出了官方商业接口,
他们之前都把 R 看作是在分析空间中的一个挑战者(见第 4 章)。Oracle,
ODBC, 微软 Excel, PostgreSQL, MySQL, SPSS, Oracle 数据挖掘器, SAS/
IML, JMP, Pentaho Kettle, 以及 Jaspersoft BI 只不过是能够和 R 应用相兼容 P.2
的众多商业软件中的个别例子。按照基本的 SAS 语言,一个 WPS 软件经销商
提供了一个称为“通往 R 之桥”的插件。分析力革命公司(Revolution
Analytics)提供了一个以分析为主、以 R 语言的许可为依托的产品,但其他一
些小公司已经成功地开发了商业的 R 应用和 R 程序包。

- 数据的可互操作性:来自不同数据库和不同文件格式的数据可通过一个程序包的连接,在 R 中直接使用,或可缩减为一个中间格式以便导入到 R 中(见第 2 章)。
- 广泛的数据可视化能力:这包括比其他软件更好的动画和图形(见第 5 章)。
- 最大、发展最快的开源统计库:当前统计软件包的数量和新软件包持续升级的增长速度确保 R 具备了分析问题的长期解决方案的连续性。
- 包含从 R 程序包的统计库、分析库、数据挖掘库、操作界面库、数据可视化库到在线应用的广泛的解决方案,使其成为在实际中分析领域最宽广的平台。

1.2 与其他分析程序包相比 R 的其他优势

那么 R 还有别的什么优势吗?下面列出了 R 一些额外的特性,这些特性使它优于其他分析软件。

- R 的源代码是为了确保完成定制的解决方案和植入某个特定的应用程序而设计的。开源的代码具有在期刊和科学文献中被广泛的同行评议的优势。这意味着程序漏洞会被发现,其信息会被共享,而解决方案也会被透明地传送。
 - 关于 R 分析平台有广泛的书籍形式的培训材料(见第 12 章)。
 - R 在其分析软件中提供了最好的数据可视化工具(除了最新版的 Tableau 软件之外)。在 R 包中广泛的数据可视化包含各种各样的可定制的图形和动画。第三方软件开始创建 R 接口的主要原因是 R 程序包中的图形库更高级,并具有日益增多的功能。
 - R 许可证是免费的,这使得 R 对于学者和小型、大型分析团队来说在预算上显得很友好。
 - R 为你的数据环境提供灵活的编程。这包括能够确保兼容 Java,Python,C++ 等的程序包。
 - 从其他分析平台可以很容易地迁移到 R 平台。非 R 平台的用户转换到 R 平台相对容易,并且因为源代码的 GPL 性质以及开放社区,因而不存在被供应商锁死的危险,其 GPL 可在 <http://www.gnu.org/copyleft/gpl.html> 找到。
- P.3 • R 中有最新的、范围最广泛的统计算法。这归因于 R 的程序包结构,它使开发者比在其他任何类似的分析平台上更容易创建新的程序包。

1.3 R 作为统计语言和分析平台的差异

有时确实能够将统计计算和分析区分开来。虽然统计学是一种基于工具和技术的方法,分析则更关心业务的目标。统计数据基本上是数字,可以昭示(通过描