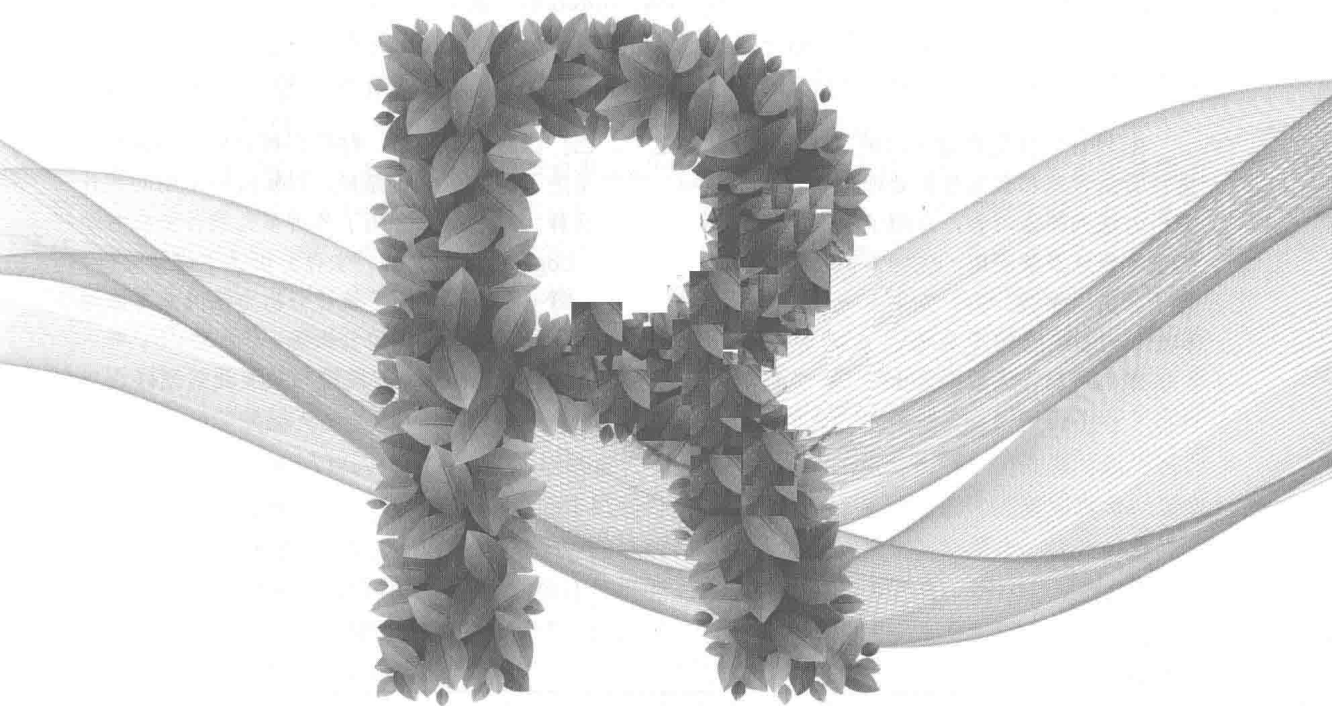


资深专家执笔，结合一线工程实践经验编写而成
快速掌握 R 软件操作及应用该知识点解决实践中的问题
深入剖析数据统计分析应用的流程、细节、难点和技巧
案例涉及心理学、社会学、医学、生物、商业和工程等诸多领域



R 语言统计分析与应用

汪海波 罗 莉 汪海玲◎编著



R语言统计分析与应用

汪海波 罗 莉 汪海玲◎编著

人民邮电出版社
北京

图书在版编目 (CIP) 数据

R语言统计分析与应用 / 汪海波, 罗莉, 汪海玲编著

— 北京: 人民邮电出版社, 2018. 4

ISBN 978-7-115-46982-3

I. ①R… II. ①汪… ②罗… ③汪… III. ①统计分
析—统计程序 IV. ①C819

中国版本图书馆CIP数据核字(2017)第322700号

内 容 提 要

R 统计软件是目前应用最广泛的统计软件之一, 已广泛应用于医学、财经和社会科学等领域中进行数据管理和数据分析处理。本书以 Windows 操作系统下的 R 软件为基础, 以实践中常用的统计分析方法为基本内容, 介绍了 R 语言的编写以及结果解释。本书重点介绍了各种多元统计分析方法的基本原理及其应用, 包括方差分析、多元线性回归、Logistic 回归分析、生存分析、主成分分析、因子分析、聚类分析、判别分析以及典型相关分析等。每一章详细讨论了统计分析方法的基本原理和分析过程, 介绍了 R 语言的使用方法及应用实例说明、结果解释及结论分析等。

本书可以作为数据分析和数据管理人员的参考用书, 也可以作为高等院校相关专业的教材。

◆ 编 著 汪海波 罗莉 汪海玲

责任编辑 张涛

执行编辑 张爽

责任印制 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>

北京鑫正大印刷有限公司印刷

◆ 开本: 787×1092 1/16

印张: 28.5

字数: 751 千字

2018 年 4 月第 1 版

印数: 1—2 400 册

2018 年 4 月北京第 1 次印刷

定价: 79.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字 20170147 号

前 言

R 是世界领先的信息系统，是一个大型规模化的集成应用软件，具有完备的数据存取、管理、分析和显示的功能。R 统计分析软件使用灵活方便、功能齐全，R 语言编程能力强且简单易学，其数据处理和统计分析融为一体，已被广泛应用于医学、财经和社会科学等领域中。R 是一个开源项目，在很多操作系统上都可以免费使用，包括 Windows、Mac OS X 和 Linux。R 还在持续发展中，每天都在纳入新的功能。R 已经成了统计、预测分析和数据可视化的全球通用语言。它提供各种用于分析和理解数据的方法，从最基础的到最前沿的，无所不包。R 以能创建漂亮优雅的图形而闻名，但实际上它可以处理各种统计问题。基本的安装版本就提供了数以百计的数据管理、统计和图形函数。

本书是一本 R 指南，高度概括了该软件和它的强大功能。本书介绍了基本安装中最重要的函数，以及重要扩展包中的函数。整本书都是围绕实际应用展开的，你将学会理解数据，并能够与他人交流对数据的理解。通读本书，你会对 R 的原理和功能有基本的了解，并知道可以从什么地方学习更多的相关知识。你将能用各种技术实现数据可视化，还能解决各种难度的数据分析问题。

本书内容充实、重点突出、实用性强，不仅介绍了常用的统计分析方法，也详细阐述了多元统计分析方法及其应用，例如方差分析、生存分析等，是较完整地覆盖多元统计分析方法的一本参考书，适合不同层次的人员使用。本书对每一种多元统计分析方法的基本原理、分析过程进行了详细的阐述，并介绍了 R 函数的使用方法、应用实例说明、结果解释、结论分析等，使读者可以将统计分析方法与 R 软件应用结合起来学习，全面掌握统计分析方法。统计基础知识比较薄弱者可以忽略计算公式等理论部分，把重点放在对基本原理的理解和利用计算机解决实际问题上。统计基础知识较好者可以精读每一部分内容，使自己的统计分析能力达到更高的一个层次，并为今后进一步学习和研究打下扎实的实践基础。本书适合医学、经济学、农业等多个领域的人员参考。

本书在编写方法上有所创新。目前，许多统计分析教程存在理论和实践操作脱节，或者说计算例题和解决实际问题脱节的现象，使得多元统计在研究上的应用价值不能真正体现出来。本书从讲求实效的原则出发，将统计分析原理和 R 软件的输出结果有机地结合在一起，将理论学习与实际操作融为一体，使得读者有机会把各种统计分析方法的理论知识同解决实际问题的能力结合起来。

本书内容

本书第 1 章介绍了 R 的特点以及结构组成，R 的获取、安装和启动，如何用网上的扩展包增强 R 基本安装的功能，R 菜单操作，用户界面，如何以交互方式和批处理方式运行程序，R 系统的文件管理。第 2 章介绍了 R 编程入门，包括数据集的建立以及各种操作方法等，同时讲解了向 R 中导入数据的诸多方法。这一章的前半部分介绍了如何用 R 来存储数据的数据结构，以及如何用键盘输入数据；后半部分介绍了怎样从文本文件、网页、电子表格、统计软件和数据库向 R 导入数据，涵盖了数据管理中函数（数学函数、统计函数、字符函数）以及常用运算符。第 3 章探讨了基本的数据管理，包括数据集的排序、合并、取子集，以及变量的转换、重编码和删除。第 4 章介绍了样本量和检验效能估计。第 5 章在第 3 章的基础上介绍了高级数据管理方法，涵盖了控制结构（循环、条件执行）的用法，然后介绍如何编写自己的 R

函数，以及如何用不同的方法整合数据。第 6 章概括地总结了定量资料基本统计分析方法的主要内容。从第 7 章开始逐个介绍了工作实践中常用的一些多元统计分析方法，内容包括各种多元统计分析方法的基本原理、计算方法、分析步骤、实例应用、R 函数和使用方法，以及结果解释和结论分析等；介绍了针对一个数值型结果变量与一系列数值型预测变量间的关系进行建模的回归方法，并详细给出了拟合模型的方法、适用性评价和含义解释，包括 t 检验、方差分析、直线回归和相关、多元线性回归与相关、Logistic 回归分析。第 12 章介绍了相对数的概念、率假设检验方法和 R 程序。第 13 章介绍了行 \times 列表分析和交叉表。第 14 章探讨了非参数分析方法，包括配对及单样本秩和检验、两组样本比较的秩和检验、多组样本比较的秩和检验及两两比较、等级分组资料和随机区组设计资料比较的秩和检验及两两比较。第 15 章介绍了生存分析，主要涉及生存率的两种估计方法，乘积极限法和寿命表法，以及 Cox 模型。多元数据分析的一个难点是简化数据。后续章节介绍了如何将大量的相关变量转换成较少的不相关变量（主成分分析），以及如何发现一系列变量中的潜在结构（因子分析）。这些方法涉及许多步骤，每一步都有详细的介绍。第 16~20 章分别阐述了主成分分析、因子分析、聚类分析、判别分析和典型相关分析。从第 21 章开始探讨了 R 软件的拓展应用，包括 ROC 分析和统计图，所涉及的统计图包括条形图、饼图、散点图、折线图、箱线图、直方图、核密度图和点图。另外，本书每一章都给出了使用 R 软件进行数据分析的实例，以帮助读者理解和掌握各章的内容。

本书特色

- 由 R 资深专家执笔。作者深入理解了 R 的内涵和精髓，结合自己丰富的培训经验，以及大量的一线工程实践经验，潜心编写而成。
- 软件版本采用当前最新的 R 版本，在知识点讲解过程中穿插了新功能的讲述与应用。
- 知识全面、系统，科学安排内容的层次架构，由浅入深，循序渐进，适合读者的学习规律。
- 理论与实践应用紧密结合。基础理论知识穿插在知识点的讲述中，言简意赅、目标明确，其目的是使读者知其然，亦知其所以然，达到学以致用目的。
- 知识点+针对每个知识点的小实例+综合实例的讲述方式，可以使读者快速地学习并掌握 R 软件操作及应用该知识点解决实践中的问题。综合实例部分，深入细致地剖析数据统计分析应用的流程、细节、难点、技巧，起到融会贯通的作用。
- 为了让本书内容尽可能接近各个领域的实际情况，作者从心理学、社会学、医学、生物、商业和工程等诸多领域选取了一些例子。所有的这些例子都不需要读者具备这些领域的专业知识。

本书附带所有实例操作的数据和 R 程序。

本书由汪海波、罗莉、汪海玲编著，参与编写的还有郝旭宁、李建鹏、赵伟茗、刘钦、于志伟、张永岗、周世宾、姚志伟、曹文平、张应迁、张洪才、邱洪钢、张青莲、陆绍强、李成。

读者对象

本书可以作为从事数据分析和数据管理工作人员的参考用书，以及有关科研人员和科研管理者进一步提高统计分析水平的参考书，也可以作为高等院校师生的统计教材。

因编者水平有限，不妥之处敬请读者提出宝贵意见，联系邮箱为 zhangtao@ptpress.com.cn。

编者

目 录

第一篇 R 基础与入门篇

第 1 章 R 入门	2	3.2.3 用 : 运算符创建向量	37
1.1 R 简介	2	3.2.4 使用 seq() 函数创建向量	38
1.1.1 R 特点	2	3.3 处理数据对象的实用函数	38
1.1.2 R 支持资料	3	3.4 变量的重编码	39
1.2 R 的获取、安装和启动	4	3.5 变量的重命名	40
1.2.1 R 的获取	4	3.6 缺失值	41
1.2.2 R 的安装	5	3.7 日期值	42
1.2.3 R 的启动	7	3.8 类型转换	44
1.3 R 菜单操作	7	3.9 数据排序	45
1.4 工作空间	10	3.10 数据集的合并	45
1.5 程序包	11	3.11 数据集取子集	46
1.5.1 什么是程序包	11	3.11.1 选入观测	46
1.5.2 安装程序包	11	3.11.2 选入变量	47
1.6 R 使用以及图形界面	12	3.11.3 剔除变量	48
1.7 本章小结	13	3.11.4 subset() 函数	49
第 2 章 R 编程入门	14	3.12 本章小结	49
2.1 R 语言	14	第 4 章 样本量和检验效能估计	50
2.1.1 数据集的概念	14	4.1 样本量估算以及 R 程序包	50
2.1.2 R 运算符	21	4.1.1 样本量影响因素	50
2.2 R 常用函数及其应用	23	4.1.2 检验效能分析 pwr 包	52
2.2.1 数学函数	24	4.2 t 检验	53
2.2.2 样本统计函数	26	4.2.1 单样本与已知总体检验时样本量的估计及 R 程序	53
2.2.3 概率函数	27	4.2.2 两总体均数比较样本量的估计及 R 程序	54
2.2.4 字符处理函数	28	4.2.3 配对设计两样本均数比较样本量的估计及 R 程序	55
2.2.5 其他实用函数	30	4.3 方差分析	56
2.3 数据的输入	31	4.4 相关分析	57
2.3.1 使用键盘输入数据	31	4.5 线性模型	58
2.3.2 数据集的导入	32	4.6 分类资料的样本量估计	59
2.4 本章小结	34	4.6.1 单样本与已知总体检验时样本量的估计及 R 程序	59
第 3 章 基本数据管理	35	4.6.2 两样本率比较样本量的估计及 R 程序	60
3.1 创建新变量	36		
3.2 向量运算	37		
3.2.1 添加或删除向量元素	37		
3.2.2 向量运算和逻辑运算	37		

4.6.3 配对设计总体率比较样本量的估计及 R 程序.....61	5.1.2 条件执行.....65
4.7 本章小结.....62	5.2 数据处理综合实例.....67
第 5 章 高级数据管理.....63	5.3 转置与整合.....70
5.1 控制语句.....63	5.3.1 转置.....70
5.1.1 重复和循环.....63	5.3.2 整合数据.....71
	5.4 本章小结.....72

第二篇 统计方法与 R 分析实例

第 6 章 定量资料的统计描述.....74	第 8 章 方差分析.....108
6.1 统计描述基础理论知识.....74	8.1 方差分析及 ANOVA 模型拟合概述.....108
6.1.1 集中趋势描述.....75	8.1.1 方差分析的基本思想.....108
6.1.2 离散趋势描述.....77	8.1.2 方差分析基本术语.....110
6.1.3 正态分布.....79	8.1.3 ANOVA 模型拟合.....111
6.2 统计描述分析实例.....81	8.2 完全随机设计资料的方差分析.....112
6.2.1 summary()函数分析实例.....81	8.2.1 单因子方差分析介绍.....113
6.2.2 sapply()函数分析实例.....83	8.2.2 单因子方差分析的 R 程序实例.....113
6.2.3 describe()函数分析实例.....85	8.3 随机区组设计资料的方差分析.....118
6.2.4 stat.desc()函数分析实例.....89	8.3.1 随机区组方差分析介绍.....119
6.2.5 分组计算描述性统计量.....91	8.3.2 随机区组方差分析的 R 程序实例.....121
6.2.6 对数正态分布资料的统计描述.....94	8.4 拉丁方设计资料的方差分析.....126
6.3 本章小结.....95	8.4.1 拉丁方方法介绍.....126
第 7 章 <i>t</i> 检验.....96	8.4.2 拉丁方分析的 R 程序实例.....128
7.1 单样本 <i>t</i> 检验.....96	8.5 析因设计资料的方差分析.....131
7.1.1 单样本 <i>t</i> 检验的基础理论.....96	8.5.1 析因设计方法介绍.....131
7.1.2 单样本 <i>t</i> 检验分析实例.....97	8.5.2 析因方差分析的 R 程序实例.....134
7.1.3 无原始数据的单样本 <i>t</i> 检验 R 程序.....98	8.6 正交试验设计资料的方差分析.....136
7.2 配对设计资料的 <i>t</i> 检验.....98	8.6.1 正交试验设计方法介绍.....136
7.2.1 配对设计资料 <i>t</i> 检验的基础理论.....98	8.6.2 正交试验设计资料分析的 R 程序实例.....138
7.2.2 配对 <i>t</i> 检验实例.....100	8.7 重复测量资料的方差分析.....139
7.2.3 无原始数据的配对设计的 <i>t</i> 检验分析实例.....102	8.7.1 重复测量设计方法介绍.....140
7.3 两独立样本的 <i>t</i> 检验.....103	8.7.2 重复测量资料分析的 R 程序实例.....141
7.3.1 两独立样本 <i>t</i> 检验的基础理论.....103	8.8 协方差分析.....144
7.3.2 独立样本 <i>t</i> 检验分析实例.....105	8.8.1 协方差分析方法介绍.....144
7.3.3 无原始数据的两独立样本 <i>t</i> 检验分析实例.....107	8.8.2 协方差分析的 R 程序实例.....145
7.4 本章小结.....107	8.9 本章小结.....148
	第 9 章 直线回归与相关.....149
	9.1 直线相关分析.....149
	9.1.1 直线相关分析介绍.....149

9.1.2 直线相关分析的 R 实例	151	13.2.2 四格表配对卡方检验的 R 程序	225
9.2 直线回归分析	154	13.3 列变量为顺序变量的行均分检验	226
9.2.1 直线回归分析介绍	155	13.3.1 行均分检验介绍	227
9.2.2 直线回归分析的 R 程序实例	157	13.3.2 行均分检验的 R 程序	227
9.3 本章小结	162	13.4 行列均为顺序变量的相关检验	230
第 10 章 多元线性回归与相关	163	13.4.1 行列均为顺序变量的相关 检验介绍	230
10.1 多元线性回归与相关的基础理论	163	13.4.2 行列均为顺序变量的相关 检验的 R 程序	231
10.1.1 多元线性回归	163	13.5 分层行列列表的分析	235
10.1.2 复相关系数与偏相关系数	176	13.5.1 分层行列列表的分析简介	235
10.2 分析实例	178	13.5.2 分层行列列表的分析的 R 程序	236
10.2.1 多元线性回归方程的建立	178	13.6 趋势卡方检验	239
10.2.2 复相关系数与偏相关系数的 R 程序实例	183	13.6.1 趋势卡方检验简介	239
10.3 本章小结	185	13.6.2 趋势卡方检验的 R 程序	239
第 11 章 Logistic 回归分析	186	13.7 卡方分割与卡方合并	241
11.1 非条件 Logistic 回归	186	13.7.1 卡方的分割与合并简介	241
11.1.1 非条件 Logistic 回归介绍	187	13.7.2 卡方分割与卡方合并分析 实例	241
11.1.2 非条件 Logistic 回归模型的 建立和检验	188	13.8 本章小结	243
11.1.3 非条件 Logistic 回归的 R 程序	190	第 14 章 非参数统计	244
11.2 条件 Logistic 回归	205	14.1 单样本资料与已知总体参数的非 参数检验	245
11.2.1 条件 Logistic 回归介绍	205	14.1.1 单组资料的符号及符号秩和 检验	245
11.2.2 条件 Logistic 回归的 R 程序	206	14.1.2 单组资料的非参数检验 R 程序	247
11.3 本章小结	207	14.2 配对设计资料的非参数检验	248
第 12 章 相对数	208	14.2.1 配对设计资料的符号及符号 秩和检验	248
12.1 相对数简介	208	14.2.2 配对设计资料的非参数检验 R 程序	249
12.1.1 率的标准化	210	14.3 两组定量资料的非参数检验	250
12.1.2 率的假设检验	212	14.3.1 两组定量资料的非参数检验 方法概述	251
12.2 R 分析实例	214	14.3.2 两组定量资料非参数检验的 R 程序	252
12.2.1 率的标准化 R 程序	214	14.4 多组定量资料的非参数检验	253
12.2.2 率的 Z (U) 检验的 R 程序	215	14.4.1 多组定量资料的非参数检验 方法概述	253
12.3 本章小结	216	14.4.2 多组定量资料非参数检验的 R 程序	255
第 13 章 行×列表分析	217		
13.1 四格表资料	217		
13.1.1 四格表卡方检验介绍	218		
13.1.2 四格表卡方检验的 R 程序	220		
13.2 配对计数资料的卡方检验	224		
13.2.1 四格表配对卡方检验介绍	224		

14.5 等级分组资料的非参数检验.....	260	第 18 章 聚类分析.....	338
14.5.1 等级分组资料的非参数检验 方法概述.....	260	18.1 聚类分析简介.....	338
14.5.2 等级分组资料非参数检验的 R 程序.....	261	18.2 聚类分析及 R 实例.....	344
14.6 随机区组资料的非参数检验.....	264	18.2.1 varclus()函数.....	344
14.6.1 随机区组资料的非参数检验 方法概述.....	264	18.2.2 kmean()函数.....	348
14.6.2 随机区组资料非参数检验的 R 程序.....	265	18.2.3 hclust()函数实例.....	352
14.7 等级相关(秩相关).....	266	18.3 本章小结.....	355
14.7.1 秩相关概述.....	266	第 19 章 判别分析.....	356
14.7.2 spearman 秩相关的 R 程序.....	267	19.1 判别分析简介.....	357
14.8 本章小结.....	268	19.2 判别分析及 R 实例.....	362
第 15 章 生存分析.....	269	19.3 本章小结.....	386
15.1 生存分析简介.....	269	第 20 章 典型相关分析.....	388
15.1.1 生存数据.....	269	20.1 典型相关简介.....	388
15.1.2 生存时间函数.....	270	20.1.1 典型相关分析的理论架构及 基本假设.....	390
15.1.3 均数、中位数和半数 生存期.....	271	20.1.2 冗余分析.....	391
15.1.4 生存分析的基本方法.....	271	20.1.3 典型相关系数的假设检验.....	392
15.2 生存曲线.....	272	20.2 cancel()函数实例.....	392
15.2.1 寿命表法及 R 分析实例.....	273	20.3 本章小结.....	400
15.2.2 乘积极限法(Kaplan-Meier)及 R 分析实例.....	278	第 21 章 诊断试验的 ROC 分析.....	401
15.2.3 Cox 回归及 R 分析实例.....	280	21.1 诊断试验简介.....	401
15.3 本章小结.....	285	21.1.1 诊断试验介绍.....	401
第 16 章 主成分分析.....	286	21.1.2 诊断试验评价指标.....	402
16.1 主成分分析简介.....	287	21.1.3 ROC 分析资料收集与整理.....	404
16.1.1 主成分分析的数学模型.....	287	21.1.4 ROC 曲线构建.....	405
16.1.2 主成分分析的方法步骤.....	288	21.2 ROC 分析及 R 分析实例.....	406
16.1.3 主成分分析的应用.....	290	21.3 本章小结.....	423
16.2 R 中的主成分分析实例.....	291	第 22 章 统计图.....	425
16.3 本章小结.....	307	22.1 条形图.....	425
第 17 章 因子分析.....	308	22.2 饼图.....	429
17.1 因子分析简介.....	308	22.3 散点图.....	431
17.2 主成分分析与因子分析比较.....	317	22.4 折线图.....	433
17.3 因子分析及 R 实例.....	318	22.5 箱线图.....	434
17.4 本章小结.....	337	22.6 直方图.....	437
		22.7 核密度图.....	442
		22.8 点图.....	442
		22.9 本章小结.....	444
		参考文献.....	445

第一篇 R 基础与入门篇

第1章 R 入门

学习目标

- 了解 R 系统的特点以及结构组成。
- 熟悉 R 的安装和启动。
- 掌握 R 菜单操作。
- 掌握 R 系统的文件管理。

内容概要

R 是允许用户编辑算法并使用其他可编程工具的一种计算机语言，具有完备的数据存取、管理、分析和显示等功能，是现今最受欢迎的数据分析和可视化平台之一，是一种针对统计分析和数据科学的功能全面的开源统计语言。它是自由的开源软件，并同时提供 Windows、Mac OS X 和 Linux 系统的各种版本。它在商业、工业、政府部门、医药和科研等涉及数据分析的领域都有广泛的应用。

本章对 R 软件和 R 系统的特点进行了较详细的介绍，描述了 R 软件的要求和环境支持。希望大家对 R 系统能够有较全面的了解。首先简要介绍了 R，以及它作为数据分析平台的诸多特性。本章主要介绍了 R 的获取，以及如何用网上的扩展包增强 R 基本安装的功能。另外，本章还介绍了用户界面，以及如何以交互方式和批处理方式运行程序。同时简要地介绍了 R 的安装和启动，对 R 菜单操作也进行了详细的讲解，尤其是那些 R 操作中经常应用的菜单。R 虽然是编程软件，但是菜单操作仍有非常重要的作用。

1.1 R 简介

R 语言是从 S 统计绘图语言演变而来的，S 语言在 20 世纪 70 年代诞生于贝尔实验室，由 Rick Becker、John Chambers 和 Allan Wilks 开发。基于 S 语言开发的商业软件 Splus，可以方便地编写函数、建立模型，具有良好的扩展性，在国外学术界应用很广。1995 年，新西兰奥克兰大学统计系的 Robert Gentleman 和 Ross Ihaka 基于 S 语言的源代码，编写了一个能执行 S 语言的软件，并将该软件的源代码全部公开，这就是 R 软件，其命令统称为 R 语言。

1.1.1 R 特点

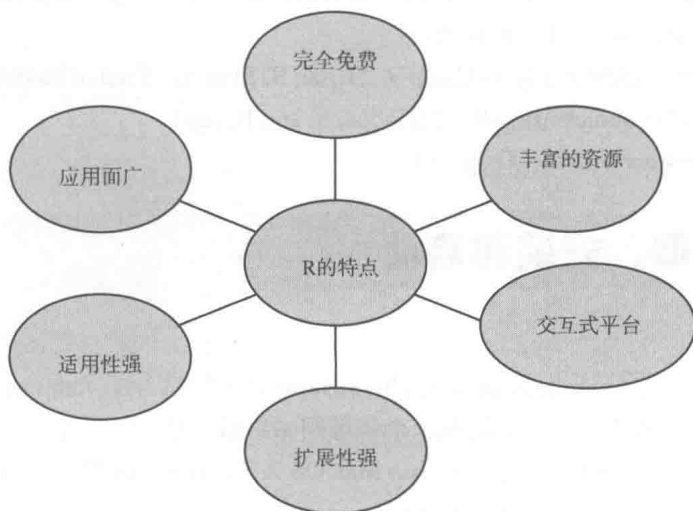
与起源于贝尔实验室的 S 语言类似，R 也是一种为统计计算和绘图而生的语言和环境，它是一套开源的数据分析解决方案，由一个庞大且活跃的全国性研究型社区维护。R 的特点如图 1-1 所示。

具体来说，R 有以下 7 个方面的特点。

(1) 完全免费，软件本身及程序包的源代码公开。多数商业统计软件价格不菲，投入成千上万美元都是有可能的。

(2) 丰富的资源，涵盖了多种行业数据分析中几乎所有的办法。R 是一个全面的统计研究平

台，提供了各式各样的数据分析技术，几乎任何类型的数据分析工作皆可在 R 中完成。



▲图 1-1 R 的特点

(3) 交互式平台。R 是一个可进行交互式数据分析和探索的强大平台，任意一个分析步骤的结果均可被轻松保存、操作，并作为进一步分析的输入。有各式各样的图形用户界面工具通过菜单和对话框提供了与 R 语言同等的功能。

(4) 扩展性强。编写函数和程序包十分方便，跨平台，可以胜任复杂的数据分析、绘制精美的图形。R 拥有顶尖水准的制图功能。如果希望复杂数据可视化，那么 R 拥有最全面且最强大的一系列可用功能。

(5) 适用性强。R 可运行于多种平台之上，包括 Windows、UNIX 和 Mac OS X。这基本上意味着它可以运行于你所能拥有的任何计算机上。R 可以轻松地从各种类型的数据源导入数据，包括文本文件、数据库管理系统、统计软件，乃至专门的数据仓库。它同样可以将数据输出并写入这些系统中。

(6) 应用面广。R 适用于几乎任何应用的需要。目前在 R 网站上约有 2400 个程序包，涵盖了基础统计学、社会学、经济学、生态学、空间分析、系统发育分析、生物信息学等诸多方面。R 可用来解决自然科学和社会科学领域中的各种问题，如统计学、社会学、经济学、生物、医疗卫生等，其功能包括图形显示、数据分析、数据输入、数据检查、数据管理、报表生成、统计计算、市场研究、销售管理、供应管理、调查分析和建立预测模型等。

(7) 完备的帮助系统。每个函数都有统一格式的帮助，运行实例。

1.1.2 R 支持资料

R 的官方网页拥有大量英文学习资源，还可以参考 <http://cran.r-project.org/other-docs.html> 上的中文翻译文档。统计之都 BBS 是一个不错的学习 R 的中文网站，你可以在这个论坛中找到大量学习资料或直接提出问题同大家探讨。

R 的帮助系统非常强大，可以直接使用 `?topic` 或 `help(topic)` 来获取 `topic` 的帮助信息，也可使用 `help.search("topic")` 来搜索帮助系统。如果你只知道函数的部分名称，那么可以使用 `apropos("tab")` 来搜索得到载入内存所有包含 `tab` 字段的函数。如果还没有得到需要的资料，还有 R Site Search: <http://finzi.psych.upenn.edu/search.html>，等价于在 R 平台上使用 `RSiteSearch()` 函数。

正如大家应用 R 的过程中看到的，大部分经典的 R 书籍都为英文版的，例如：

Modern Applied Statistics with S (Venables and Ripley)

The New S Language: A Programming Environment for Data Analysis and Graphics (Richard A. Becker, John M. Chambers, Allan R. Wilks)

A Handbook of Statistical Analysis Using R (Brian S. Everitt, Torsten Hothorn)

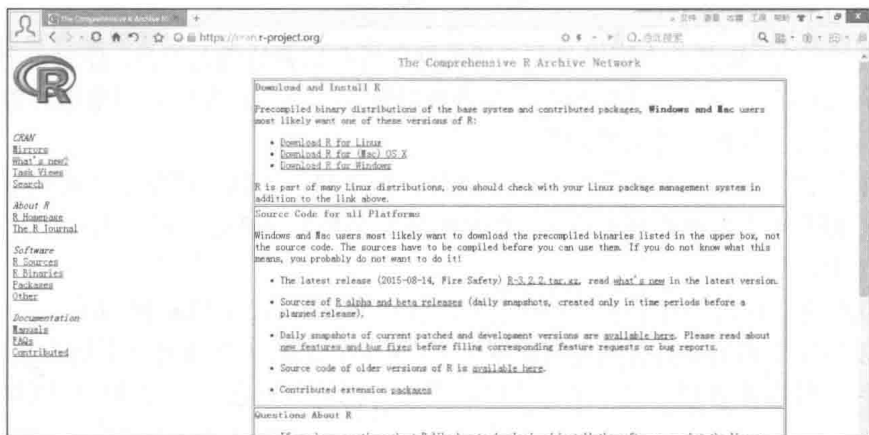
Data Analysis and Graphics using R (Maindonald and Braun)

Introductory Statistics with R (Dalgaard)

1.2 R 的获取、安装和启动

1.2.1 R 的获取

R 可以在 CRAN (Comprehensive R Archive Network) 网站 <http://cran.r-project.org> 中免费下载, 如图 1-2 所示。CRAN 由世界上的几十个镜像网站组成网络, 提供下载安装程序和相应软件包, 各镜像更新频率一般为 1~2 天。Linux、Mac OS X 和 Windows 都有相应编译好的二进制版本。根据所选择平台的安装说明进行安装即可。



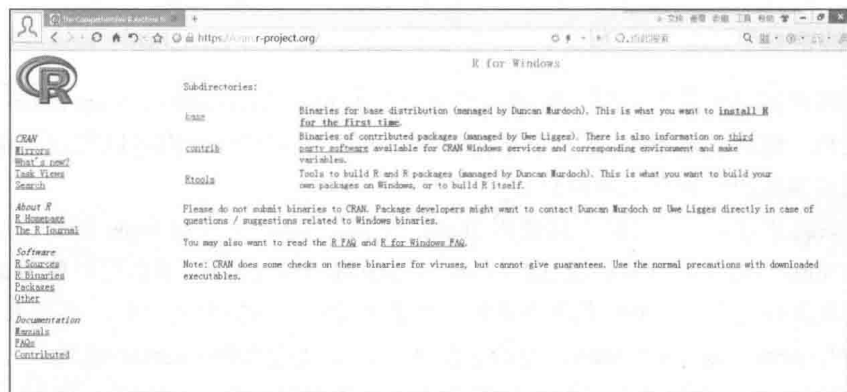
▲图 1-2 R 的下载界面 (1)

推荐镜像如下:

中国的镜像——数学所, <http://ftp.ctex.org/mirrors/CRAN>。

即时更新的 CRAN 源, <http://cran.r-project.org>。

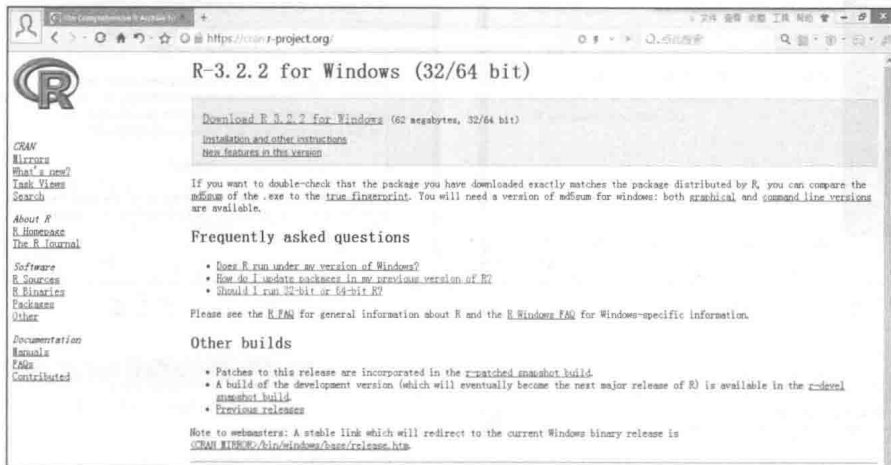
由于大部分人都是用 Windows 系统, 以“R for Windows”为例说明如何下载和安装 R。单击“Download R for Windows”进入下载界面, 如图 1-3 所示。



▲图 1-3 R 的下载界面 (2)

继续选择“install R for the first time”。

选择“Download R 3.2.2 for Windows (62 megabytes, 32/64 bit)”，即可下载所需要的 R 软件了，如图 1-4 所示。



▲图 1-4 R 的下载界面 (3)

1.2.2 R 的安装

安装方法与其他 Windows 下的软件安装一样。下面介绍 R 的安装方法。

(1) 双击下载的 R 程序，启动安装程序，如图 1-5 所示。



▲图 1-5 R 的安装 (1)

(2) 选择安装时使用的语言。为了保持较好的运行效果，最好选择英文，如图 1-6 所示。

(3) R 的安装界面如图 1-7 和图 1-8 所示。

(4) 按照对话框的提示选择安装目录，如图 1-9 所示。

(5) 如有必要，可单击“Browse”更改安装目录，如图 1-10 所示。



▲图 1-6 R 的安装 (2)



▲图 1-7 R 的安装 (3)



▲图 1-8 R 的安装 (4)



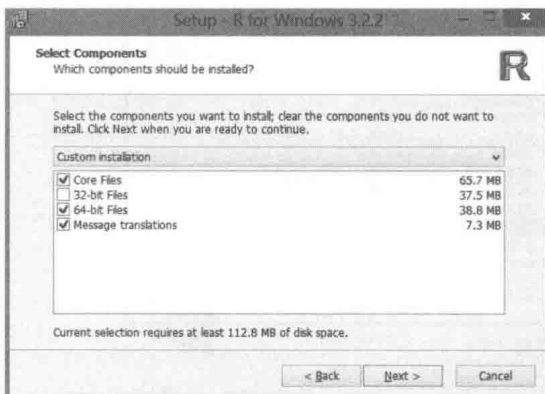
▲图 1-9 R 的安装 (5)



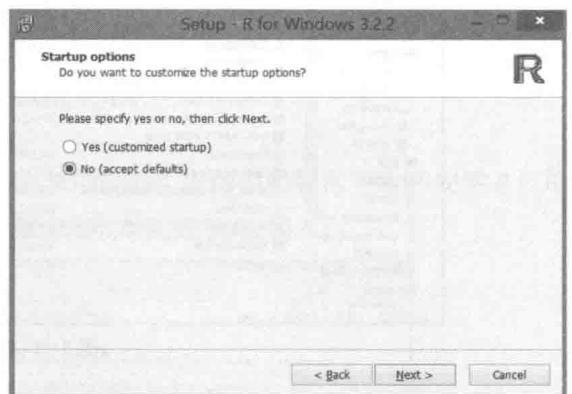
▲图 1-10 R 的安装 (6)

(6) 选择需要安装的 R 组件，“Core files”和“Message translations”为默认选择，根据操作系统选择 32 位或 64 位文件之一即可，如图 1-11 所示。

(7) 指定安装选项，可以自行设定或默认安装，本例选择默认安装，如图 1-12 所示。



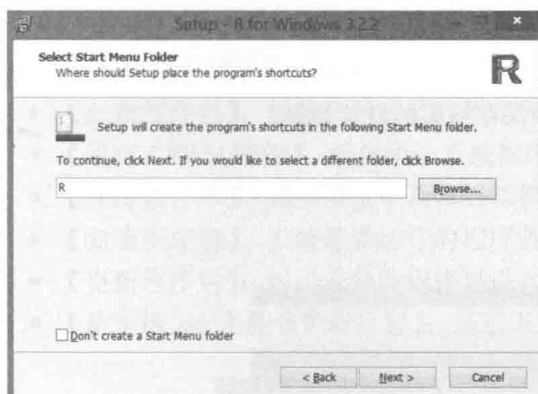
▲图 1-11 R 的安装 (7)



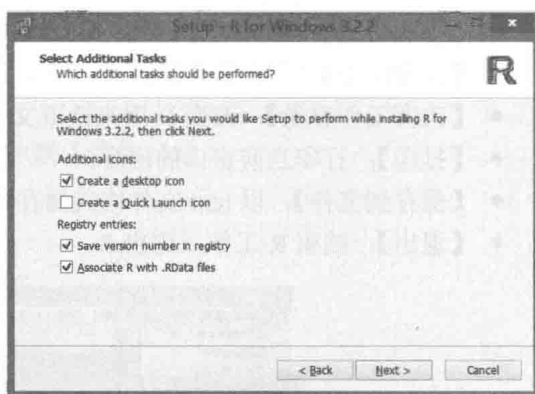
▲图 1-12 R 的安装 (8)

(8) 选择是否在“开始”菜单创建快捷方式，如图 1-13 所示。

(9) 选择附加安装选项，如图 1-14 所示。



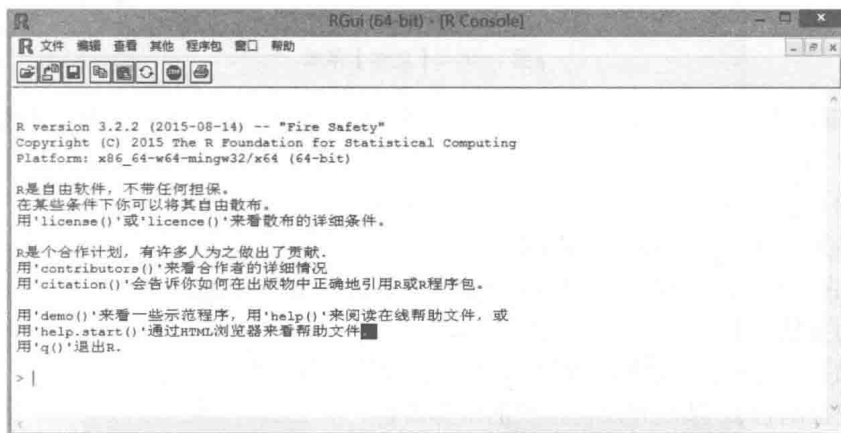
▲图 1-13 R 的安装 (9)



▲图 1-14 R 的安装 (10)

1.2.3 R 的启动

R 的启动的常用方法有两种：一是快捷方式，即按创建 Windows 系统中应用程序的快捷方式的步骤，把 R 目录中的 R.EXE 拖放至桌面，形成快捷方式，用鼠标双击快捷方式即可启动 R；二是菜单方式，从“开始”菜单的程序文件夹中找到 R 系统文件夹，启动 R 系统。启动后会出现如图 1-15 所示的 R 的运行界面。



▲图 1-15 R 的运行界面

1.3 R 菜单操作

R 主窗口标题栏下是主菜单。

(1) 【文件】菜单主要是有关 R 文件调入、保存、转换及打印等功能命令，如图 1-16 所示。

- 【运行 R 脚本文件】：调入已保存的 R 程序并输出结果。
- 【新建程序脚本】：建立新的 R 程序。
- 【打开程序脚本】：打开已保存的 R 程序。
- 【显示文件内容】：显示前次保存文件路径下的文件内容。
- 【加载工作空间】：调入已保存的工作空间，包括所有用户定义的对象（向量、矩阵、函数、数据框、列表）。
- 【保存工作空间】：保存运行中的工作空间，包括所有用户定义的对象（向量、矩阵、函数、数据框、列表）。

- **【加载历史】**: 调入运行记录。
- **【保存历史】**: 保存运行记录。
- **【改变工作目录】**: 改变 R 用来读取文件和保存结果的默认目录。
- **【打印】**: 打印当前窗口的内容。
- **【保存到文件】**: 以 text 文件格式储存记录。
- **【退出】**: 结束 R 工作, 退出 R。



▲图 1-16 【文件】菜单

(2) **【编辑】** 菜单为文本编辑“菜单”，可以对文本进行清空、复制、粘贴和数据编辑等操作，如图 1-17 所示。

- **【复制】**: 复制被标记后存在剪贴板上的文本。
- **【粘贴】**: 粘贴存在剪贴板上的文本。
- **【仅粘贴命令行】**: 仅粘贴剪贴板上文本中的命令运行内容。
- **【复制并粘贴】**: 复制的同时粘贴文本内容。
- **【全选】**: 选择程序编辑、输出、LOG 窗口的所有内容。
- **【清空控制台】**: 清空程序窗口中的所有内容。
- **【数据编辑器】**: 可以对当前工作空间的数据对象进行编辑。
- **【GUI 选项】**: 对图形用户界面进行设置。



▲图 1-17 【编辑】菜单