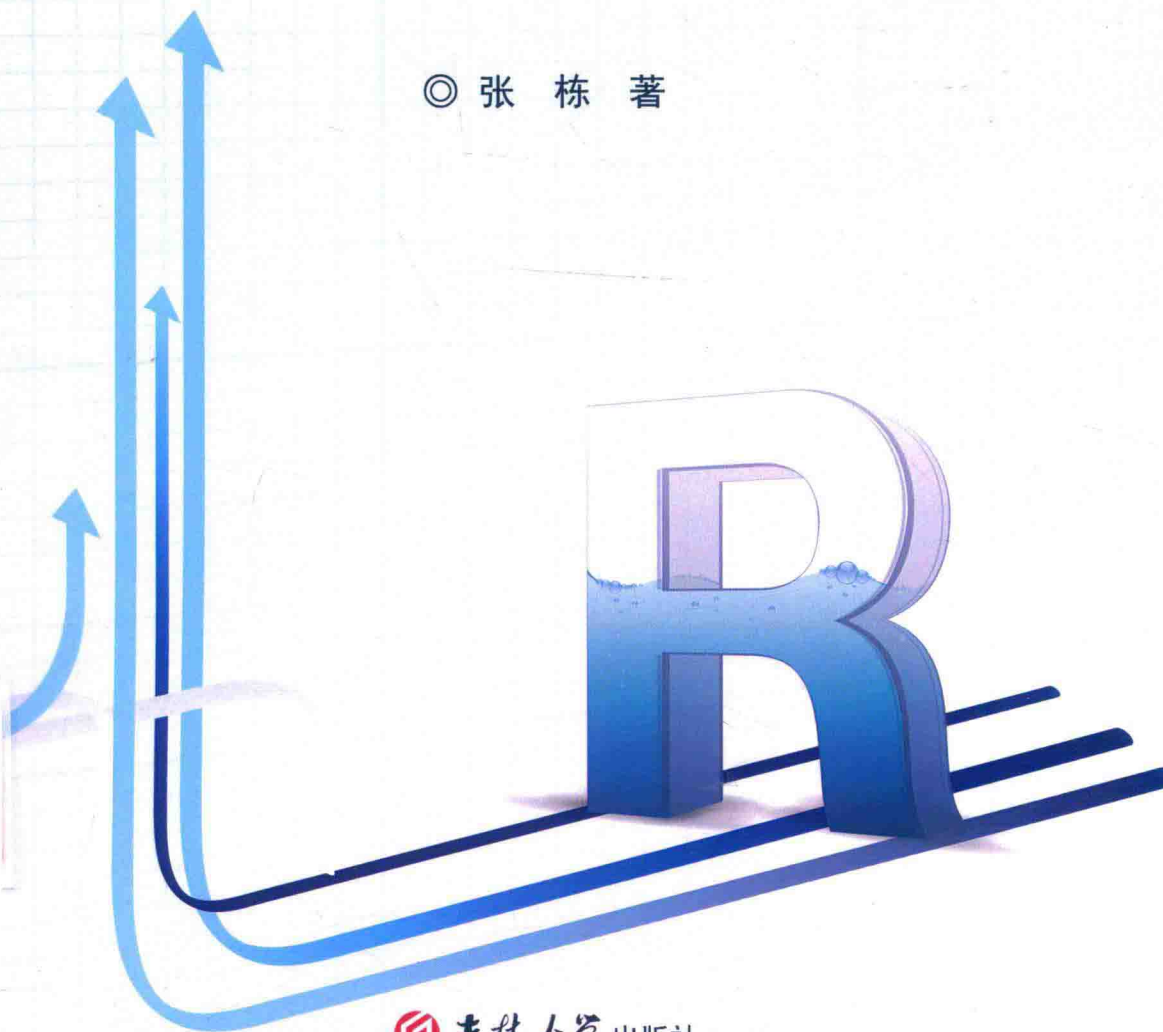


河北省科技计划项目
沧州市科技计划项目

基于R语言的多元统计分析

© 张 栋 著



 吉林大学出版社

本书系河北省科技计划项目“河北省新型城镇化与生态环境发展现状及互动关系研究”（课题编号：18456114）和沧州市科技计划项目“沧州市县域经济差异实证分析”（课题编号：183000005）的研究成果

基于R语言的多元统计分析

◎ 张 栋 著



吉林大学出版社

图书在版编目 (CIP) 数据

基于 R 语言的多元统计分析 / 张栋著. —长春: 吉林大学出版社, 2019. 2

ISBN 978-7-5692-4634-6

I. ①基… II. ①张… III. ①多元分析—统计分析—研究 IV. ①O212.4

中国版本图书馆 CIP 数据核字 (2019) 第 072667 号

书 名 基于 R 语言的多元统计分析

作 者 张栋 著

策划编辑 魏丹丹

责任编辑 魏丹丹

责任校对 王洋

装帧设计 凯祥文化

出版发行 吉林大学出版社

社 址 长春市人民大街 4059 号

邮政编码 130021

发行电话 0431-89580028/29/21

网 址 <http://www.jlup.com.cn>

电子邮箱 jdcbs@jlu.edu.cn

印 刷 河北纪元数字印刷有限公司

开 本 787mm×1092mm 1/16

印 张 10.5

字 数 161 千字

版 次 2019 年 2 月 第 1 版

印 次 2019 年 2 月 第 1 次

书 号 ISBN 978-7-5692-4634-6

定 价 45.00 元

版权所有 翻印必究

前 言

多元统计分析是研究多个随机变量之间相互依赖关系和内在统计规律的一个统计学分支，是一种综合性的分析方法，它在自然科学、经济学、管理学和社会科学等领域有着广泛的应用。因为多元统计分析的原理较为抽象，对学生数学基础的要求也较高，在教学中存在着大量的数学公式、数学符号、矩阵运算和统计计算，所以必须借助于现代化的计算工具。本书正是基于目前广泛使用的统计分析软件 R 语言完成的。

R 语言是一种集统计分析 with 图形显示于一体的统计分析软件，具有免费、源代码开放、统计分析功能完善、作图功能强大、可移植性强和使用灵活等特点，是一种进行多元统计分析的最恰当的工具。

本书在不失严谨的前提下，突出实际案例的应用和统计思想的渗透，并结合 R 语言分析出较为全面系统地研究多元分析的理论基础和使用方法。笔者在研究多元分析理论和方法的同时，尽量结合社会、经济、自然科学等领域的研究实例，将多元分析的方法与实际应用相结合，注重定性分析与定量分析的紧密结合，努力把我们在实践中应用多元分析的经验 and 体会融入其中。

本书共包括八章。第一章至第三章主要研究了多元统计分析和 R 语言的基础理论；第四章至第八章基于 R 语言的基础理论分别从多元相关分析、多元方差分析、线性回归分析、聚类分析和因子分析五个角度，以北京市空气质量为案例进行分析，并提供了相关的 R 语言程序代码。

由于笔者水平所限，书中难免存在一些疏漏甚至错误，恳请广大专家和读者批评指正。

张 栋

2018 年 11 月

目 录

第一章 多元统计分析与 R 语言概述	1
第一节 多元统计分析概述	1
第二节 R 语言概述	8
第二章 多元数据的表示及可视化	13
第一节 多元数据的矩阵表示	14
第二节 多元数据的展示及可视化	17
第三章 R 语言的基本分析和统计图形	41
第一节 R 语言的绘图基础	41
第二节 分类型单变量的基本分析	49
第三节 数值型单变量的基本分析	50
第四节 北京市空气质量监测数据的基本分析	54
第四章 基于 R 语言的多元相关分析	59
第一节 多元线性相关分析的基本思想	59
第二节 变量相关性的统计图形	68
第三节 北京市空气质量监测数据的相关性分析	70
第五章 基于 R 语言的多元方差分析	74
第一节 单因素方差分析原理和 R 实现	75
第二节 多因素方差原理和 R 实现	82
第三节 北京市空气质量监测数据的均值研究	90

第六章 基于 R 语言的线性回归分析	95
第一节 线性回归分析的一般步骤	95
第二节 建立线性回归方程	97
第三节 线性回归方程的检验	101
第四节 线性回归方程的应用	107
第五节 线性回归方程的验证	109
第六节 变量选择准则	112
第七节 北京市空气质量监测数据的线性回归分析	115
第七章 基于 R 语言的聚类分析	125
第一节 聚类分析的概述	125
第二节 系统聚类法	130
第三节 k 均值聚类法	135
第四节 北京市空气质量检测数据的聚类分析	138
第八章 基于 R 语言的因子分析	143
第一节 因子分析的基本思想与意义	143
第二节 因子载荷的估计及解释	149
第三节 因子得分计算	153
第四节 因子旋转	155
第五节 利用因子分析实现北京市空气质量的 区域综合评价	157
参考文献	161

第一节 多元统计分析概述

一、多元统计分析的历史

在统计学中,只考虑一个因素或几个因素对一个观测指标(变量)的影响大小的问题,称为一元统计分析或单因素分析。若考虑一个因素或几个因素对两个或两个以上观测指标(变量)的影响大小,或者多个观测指标(变量)间的相互关系的问题,即为多元统计分析。多元统计分析是研究客观事物中多个指标(变量)间相互依赖关系及其统计规律的数理统计学分支之一。^①

在现实生活中,随机变量受多种指标共同作用和影响的现象大量存在。当变量较多时,变量之间便不可避免地存在着相关性,分开处理不但会丢失很多信息,而且也不容易得出研究结论。有两种方法可同时对多个随机变量的观测数据进行有效的分析研究:一种方法是把多个随机变量分开分析,每次处理一个,逐次分析研究;另一种方法是对多个随机变量同时进行分析研究,即用多元统计分析的方法来解决,通过对多个随机变量观测数据的分析,来研究变量之间的相互关系并揭示变量的

^① 张曦.多元分析两个问题的理论研究和应用 [D].南京:南京信息工程大学,2008.

内在规律。所以说，多元统计分析就是研究多个随机变量之间相互依赖关系及其内在统计规律的一门学科。

统计方法是进行科学研究非常重要的辅助工具，长期以来被广泛地应用在社会学和自然科学中。1672年，英国古典政治经济学之父 William Petty（威廉·配第）发表了《政治算术》，从研究的基本方法来看，这属于一本经典的统计学著作。之后，比利时统计学家凯特勒把概率论引入了统计学，充分运用正态曲线、误差法则，增加了数理统计方法，让人们在进行世界描述的过程中，更加接近现实世界。

19世纪，经济学界出现了二维正态总体分布的方法。英国科学家戈尔登（1822年—1911年）运用生物学的统计方法，测量了近万人的身高、体重、呼吸力、拉力以及听力等数据，然后通过分析这些数据内在的联系，提出了数据相关的思想，然后运用大量数据进行描述和比较，结合自己遗传理论的基础，在进行实际统计过程中，引入了中位数、四分位数以及回归等重要的统计学方法和理论。

20世纪初期，研究者开始系统地处理多维概率分布总体的统计分析问题，提升了问题研究的系统性。1928年，Wishart（维希特）发表了论文《多元正态总体样本协差阵的精确分布》，标志着多元分析开始被作为一个独立的学科进行研究。20世纪30年代，费希尔（R. A. Fisher）与罗伊（S. N. Roy）等人对统计分析法进行了有益分析和探索。费希尔提出了变异数分析理论：在实际的实验设计过程中，坚持随机化的原则，可以进行多个参数的检测，并最大限度地减少样本出现的偏差。我国著名数学家许宝騄（1910年—1970年）在参数估计理论、多元分析、极限理论等方面都取得了卓越的成就，是我国多元统计分析学科的开拓者之一，为丰富多元统计分析方法的理论做出了重要贡献。20世纪40年代，多元统计分析方法开始应用到其他学科，比如心理学、教育学、生物学等。由于多元统计分析方法需要大量的运算，对研究者的计算能力要求很高，再加上二战的影响，多元统计分析方法的发展出现了停滞现象。

20世纪50年代，随着科技的发展，计算机的出现极大地提升了计



算能力，可以在短期内处理大量的数据，由此推动了多元统计分析方法的发展，多元统计分析方法被广泛地应用于地质、气象、医学等领域。20世纪60年代，随着社会经济的发展，经济形势出现了新的变化和特点，多元统计分析被更广泛地应用到社会实践中，形成了新的理论和方法，更加彰显了多元统计分析方法的优越性。20世纪70年代，我国学术界对多元统计分析方法进行了深入的研究，有的研究理论甚至已经达到了国际一流水平，并且我国成功建立了一支专业的研究队伍。

21世纪初期，计算机信息技术迅速发展，人们获得信息和数据的速度不断加快，人们利用计算机可以建立起超大型的数据库，这又进一步发挥了多元统计学的重要作用。如今计算机被广泛地应用在销售行业、银行产业、物理化学以及医院等多个统计领域，实现了多元统计分析方法与人工智能、数据库技术的结合，推动了人类社会的进步。随着多元正态总体参数估计和假设检验以及主成分分析方法、因子分析方法、判别分析方法等统计方法的应用，极大地推动了多元统计分析方法的进步。

二、多元统计分析的应用

采用多元统计分析技术进行数据处理、建立宏观或微观系统模型可以解决以下四个方面的问题。

(1) 变量之间的相依性分析，即分析多个或多组变量之间的相依关系，是一切科学研究尤其是经济管理研究的主要内容，简单相关分析、偏相关分析、复相关分析和典型相关分析为进行这类研究提供了必要的方法。^①

(2) 构造预测模型，进行预报控制。在自然和社会科学领域的科研与生产过程中，探索多元系统运行的客观规律及其与外部环境的关系，进行预测预报，以实现对系统的最优控制，是应用多元统计分析技术的主要目的。在多元统计分析中，用于预报控制的模型有两大类：一类是

^① 余锦华，杨维权. 多元统计分析与应用 [M]. 广州：中山大学出版社，2005.

预测预报模型，通常采用多元回归分析、双重筛选逐步回归分析、非线性回归分析和判别分析等建模技术；另一类是描述性模型，通常采用聚类分析的建模技术。

(3) 进行数值分类，构造分类模式。过去许多研究是按单因素进行定性处理，以致处理结果反映不出系统的总体特征。在多元系统分析中，往往需要将系统性质相似的事物或现象归为一类，以便找出它们之间的联系和内在规律。进行数值分类，构造分类模式一般采用聚类分析和判别分析技术。

(4) 简化系统结构，探讨系统内核。多元系统分析可采用主成分分析、因子分析、对应分析等方法，在众多因素中找出各个变量中最佳的子集合，然后根据子集合所包含的信息描述多元系统的结果及各个因子对系统的影响。这一系列过程中要注意，抓住主要矛盾，把握主要矛盾的主要方面，舍弃次要因素，以简化系统的结构，认识系统的内核。

如何选择适当的方法来解决实际问题，需要综合运用多种统计方法来对问题进行分析与考虑。例如，建立一个预报模型，可先根据有关经济学、管理学原理，确定理论模型和设计方案；再根据观察或试验结果收集相应资料，并对资料进行初步提炼；然后应用统计分析方法（如相关分析、逐步回归分析、主成分分析等）研究各个变量之间的相关性，选择最佳的变量子集合；在此基础上构造预报模型；最后对模型进行诊断和优化处理，并应用于实际生产和生活中。

三、多元统计分析的内容

多元统计分析的主要内容有：多元数据的数学表达与直观表示、多元相关分析、回归分析、判别分析、聚类分析、主成分分析、因子分析、对应分析及典型相关分析等。

1. 多元数据的数学表达

多元数据也称“多元样本值”，是对多个变量联合观测所取得的数据，即多元样本的具体值。如果将每个变量看作一个随机向量，多个变量形成的数据集将是一个随机矩阵，所以多元数据的基本表现形式就是



一个矩阵。对这些数据矩阵进行数学表示是我们的首要任务。也就是说，多元数据的基本运算是矩阵运算，而R语言是一种优秀的矩阵运算语言，这也是我们之所以选择运用它的原因之一。

2. 多元数据的直观表示

直观表示即图示法，是进行数据分析的重要辅助手段。例如，通过两个变量的散点图可以考察异常的观察值对样本相关系数的影响，通过矩阵散点图可以考察多元之间的关系，通过多元箱线图可以比较几个变量中基本统计量的大小差别。

3. 多元相关分析

相关分析就是通过对大量数字资料的分析，消除偶然因素的影响，探求现象之间相关关系的密切程度和表现形式的一种统计方法。在经济系统中，各个经济变量之间常常存在内在的关系，如经济增长与财政收入、人均收入与消费支出等。在这些关系中，有一些是严格的函数关系，这类关系可以用数学表达式表示；还有一些是非确定的关系，即一个变量产生变动会影响其他变量产生变化，且这种变化具有随机性，但是仍然遵循一定的规律。函数关系很容易解决，而那些非确定的关系，即相关关系，才是我们所要面对和解决的问题。

4. 多元回归分析

回归分析主要研究的是客观事物变量间的统计关系。它是建立在对客观事物进行大量实验和观察的基础上，用来寻找隐藏在看似不确定的现象中的统计规律的统计方法。回归分析不仅可以揭示自变量对因变量的影响程度，还可以用回归方程来进行预测和控制。回归分析的主要研究范围包括以下内容。

(1) 线性回归模型：包括简单回归模型和多元线性回归模型。

(2) 回归模型的诊断：包括判定回归模型基本假设的合理性，判定回归方程的拟合效果，以及选择回归函数的形式。

(3) 广义线性模型：含定性变量的回归，如自变量含定性变量、因变量含定性变量。

(4) 非线性回归模型：一元非线性回归，多元非线性回归。

在实际研究过程中，经常遇到一个随机变量随一个或多个非随机变量的变化而变化的情况，而这种变化呈现出明显的线性关系。怎样用一个较好的模型来表示，然后对其进行估计与预测，并对其线性进行检验就成为重要的问题。在经济预测中，经常用多元线性回归模型反映预测量与各因素之间的依赖关系，其中，线性回归分析有着广泛的应用。

5. 判别分析

判别分析是多元统计分析中用于判别样本所属类型的一种统计分析方法。所谓判别分析法，是指在已知的分类之下，有新的样品出现时，可以利用此法选定一个判别标准，以判定应将该新样品放置于哪个类别中的方法。判别分析的目的是对已知分类的数据建立由数值指标构成的分类规则，然后把这种规则应用到未知分类的样品中去分类比较。例如，我们获得了患胃炎的病人和身体健康的人的一些化验指标，就可以从这些化验指标中发现两类人的区别。把这种区别表示为一个判别公式，然后对那些被怀疑患胃炎的人就可以根据其化验指标用判别公式来进行辅助性诊断。

6. 聚类分析

聚类分析是研究“物以类聚”的一种现代统计分析方法，可以广泛应用在社会、人口、经济、管理、气象、地质及考古等众多的研究领域。例如，聚类分析可以用于不同地区城镇居民收入和消费状况的分类研究；区域经济与社会发展水平的分析及全国区域经济的综合评判；用于儿童生长发育研究时，把以形态学为主的指标归为一类，以机能为主的指标归于另一类。过去人们主要靠经验和专业知识来做定性分类处理，很少用到数学方法，致使许多分类带有主观性和任意性，不能很好地揭示客观事物内在的本质差别和联系，特别是对于多因素、多指标的分类问题，定性分类更是难以实现准确的分类。为了避免定性分类的不足，多元统计分析逐渐被引入到数值分类学中，并形成了聚类分析这个分支。聚类分析与回归分析、判别分析一起被称为多元分析的三个主要方法。聚类分析是一种分类技术，与多元分析的其他方法相比，该方法



较为粗糙，理论上还不完善，但在实际应用方面取得了很大的成功。

7. 主成分分析

在现实生活中，我们经常遇到研究多个变量的问题，然而在多数情况下，不同变量之间有一定的相关性，这必然会增加分析问题的复杂性。主成分分析就是一种通过降维技术把多个指标简化为少数几个综合指标的统计分析方法。例如，在经济管理中，可以用主成分分析的方法将一些复杂的数据综合成几个商业指数形式，如物价指数、生活费用指数、商业活动指数等。又如，在对我国各省、市、自治区经济发展做综合评价时，需要选取很多指标，为了将这些具有错综复杂关系的指标综合成几个较少的综合变量，使之既有利于分析和解释问题，又便于抓住主要矛盾做出科学的评价，此时便可以采用主成分分析的方法。

8. 因子分析

因子分析是主成分分析的推广，它也是一种把多个变量转化为少数几个综合变量的多元分析方法，但其目的是用有限个不可观测的隐变量来解释原变量之间的相关关系。在多元分析中，变量间往往存在相关性，是什么原因使变量间有关联呢？是否存在不能直接观测到但能够影响可观测变量变化的公共因子呢？因子分析就是寻找这些公共因子的统计分析方法，它是在主成分的基础上构筑若干意义较为明确的公因子，以它们为框架分解原变量，以此来考察原变量间的联系与区别。例如，在研究糕点行业的物价变动时，糕点行业品种繁多，有几百种甚至上千种，但无论哪种样式的糕点，不外乎面粉、食用油、糖等主要原料。那么，面粉、食用油、糖就是众多糕点的公共因子，各种糕点的物价变动与面粉、食用油、糖的物价变动密切相关，要了解或控制糕点行业的物价变动，只要抓住面粉、食用油和糖的价格变动规律即可。

9. 对应分析

对应分析又称为相应分析，由法国统计学家 J.P.Beozecri (J.P.比泽科里) 于 1970 年提出。对应分析是在因子分析的基础上发展起来的一种多元统计方法，是 Q 型和 R 型因子分析的联合应用。在经济管理数据的统计分析中，经常要处理三种关系，即样品之间的关系 (Q 型关

系)、变量间的关系(R型关系)以及样品与变量之间的关系(对应型关系)。例如,对某一行业的所属企业进行经济效益评价时,不仅要研究经济效益指标间的关系,还要将企业按经济效益的好坏进行分类,研究哪些企业与哪些经济效益指标的关系更为密切一些,为决策部门正确指导企业的生产经营活动提供更多有用的信息。这就需要有一种统计方法,将企业(样品)和指标(变量)放在一起进行分析、分类、作图,以便做出经济意义上的解释。而解决这类问题就可以采用对应分析的方法。

10. 典型相关分析

在相关分析中,当考察的一组变量仅有两个时,可用简单相关系数来衡量它们;当考察的一组变量有多个时,可用复相关系数来衡量它们。大量的实际问题需要我们把指标之间的联系扩展到两组变量,即分析两组随机变量之间的相互依赖关系。典型相关分析就是用来解决此类问题的一种分析方法,它实际上是利用主成分的思想来讨论两组随机变量的相关性问题,把对两组变量间的相关性研究转化为对少数几对变量之间的相关性研究,而且这少数几对变量之间又是不相关的,以此来达到简化复杂相关关系的目的。典型相关分析在经济管理实证研究中有着广泛的应用,因为许多经济现象之间都存在着多个变量对多个变量的关系。例如,在研究通货膨胀的成因时,可把几个物价指数作为一组变量,把对应的影响物价变动的因素作为另一组变量,通过典型相关分析找出几对主要综合变量,再结合典型相关系数,对物价上涨及通货膨胀的成因得出较深刻的分析结果。

第二节 R语言概述

R是一种用于统计分析和图形显示的语言和操作环境,是贝尔实验室(AT&T Bell Laboratories)的Rick Becker(里克·贝克)、Jhon Chambers(约翰·钱伯斯)和Allan Wilks(阿兰·威尔克斯)共同开



发的 S 语言的一种实现，拥有一套体系完整的统计和图形显示工具。^①

R 语言具有丰富的统计方法，大多数人使用 R 语言是因为其强大的统计功能。不过，对 R 语言比较准确的认识是，它是一个内部包含许多统计技术的操作环境。其中，部分的统计功能整合在 R 环境的底层，大多数统计功能则以包的形式提供。大约有 25 个包和 R 同时发布，这些包也被称为标准包，如果想得到更多的其他包，可以在 R 语言的网站上下载，该网站还提供了一些关于 R 语言使用的其他资料。大多数经典的统计方法和最新的技术都可以在 R 语言的网站中直接得到。

一、R 语言的特点

R 语言作为一款优秀的统计分析软件系统，具有以下几个特点。

(1) 免费和开放。R 语言是一款由志愿者维护的完全免费的统计分析软件，它的安装文件和程序包都可以从 CRAN (Comprehensive R Archive Network, R 综合归档网的简称) 社区下载，作为教学使用非常方便，国外很多大学的统计教学都在使用这款软件。而且 R 的源代码是公开的，这样方便使用者了解 R 程序的计算方法，并且可以对程序进行修改和扩展。

(2) 统计分析功能完善。R 语言内嵌了许多统计分析函数，可以直接调用进行统计分析，R 的部分统计功能整合在 R 语言的底层，但大多数功能是以各种程序包的形式提供的，大约有 25 个“标准”程序包和 R 同时发布，但更多的程序包可以通过 CRAN 社区下载、安装，而且程序包的更新比商业软件更及时，使用非常方便。

(3) 作图功能强大。R 内嵌的作图函数能在图形窗口输出漂亮美观的图形，这些图形可以保存为各种形式的文件，如 JPG、BMP、EPS、PDF、PNG、EMF、Xfig 等，使用方便。

(4) 可移植性强。R 程序可以很容易地移植到 S-PLUS (一种基于

^① 汤银才. R 语言与统计分析 [M]. 北京: 高等教育出版社, 2008.

S 语言的统计学软件) 中, 同时, S-PLUS 的程序也可以方便地移植到 R 中使用。R 可以读入很多分析软件, 如 SAS, SPSS, Excel, Stata 等的数据库文件, 而 R 的数据库文件可以保存为文本格式供其他统计软件使用, 这样 R 与其他统计软件就建立了一个良好的联系机制。

(5) 使用灵活。R 语言可以运行于 UNIX, Linux, Windows 和 Mac 等操作系统中, R 的分析结果都存放在一个对象里, 不仅用户可以选择性地显示感兴趣的结果, 而且这些结果也可以直接用于进一步的分析。

但是, R 语言的灵活性也是一把“双刃剑”, 也需要我们通过编程方式来进行统计分析。到目前为止, R 语言还缺少一个像 S-PLUS 和 SPSS 那样的菜单界面, 这对那些没有编程经验和统计方法掌握得不是很好的使用者来说是极大的挑战, 也是其在一般人群中推广的一大障碍。

二、R 语言中的基本概念

利用 R 语言进行统计分析的核心内容是: 在 R 的工作空间中创建和管理 R 对象, 调用已被加载的 R 包中的系统函数或用户自定义函数, 以逐步完成数据的管理和分析。这里涉及几个基本概念: 包、函数、工作空间、对象。掌握这些基本概念的含义, 对 R 语言快速入门将有很好的帮助。

1. 包 (Packages)

R 语言是一种面向统计分析的共享性和开源性的操作环境, 是一种面向统计分析的计算机高级语言。具体地讲, R 是一个关于包的集合。包是关于函数、数据集、编译器等集合。包是 R 的核心, 可划分为基础包 (Base) 和共享包 (Contrib) 两大类。

(1) 基础包。基础包, 顾名思义, 为 R 的基本核心系统, 是默认下载和安装的包, 由 R 核心研发团队 (Development Core Team, 简称 R Core) 管理和维护。基础包支持各类基本统计分析和基本绘图等功能, 并包含一些共享数据集可供用户使用。



(2) 共享包。共享包是由 R 的全球性研究型社区和第三方提供的各种包的集合。迄今为止，共享包中的“小包”已多达 12 000 多个，内容涵盖了各类现代统计和数据挖掘的方法，涉及诸如地理数据分析、蛋白质质谱处理、心理学测试分析等众多的应用领域。使用者可根据自身的研究目的，有选择地自行下载、安装和加载。

2. 函数 (Function)

R 函数是存在于 R 包中的实现某个计算或某种分析的程序段，每个函数都有一个函数名。使用时，用户可通过函数调用的方式，直接调用已有函数解决分析中的各类计算问题。函数名是函数调用的唯一标识。调用函数可通过以下两种格式来实现。

格式一：函数名 (形式参数列表)

函数名后括号中附带形式参数的函数调用，称为有形式参数的函数调用。函数调用时，用户须在函数名后的括号中，依顺序给出一个或多个参数值，各参数值间以英文逗号隔开。R 将这些参数值“喂给”函数，并根据参数值进行计算。

格式二：函数名 ()

函数名后括号中无任何内容的函数调用，称为无形式参数的函数调用。这类函数无须用户指定参数值，不必“喂给”函数任何具体值即可进行既定的计算。

3. 工作空间 (Workspace)

工作空间，简单来讲就是 R 的运行环境，也称工作内存。R 的所有计算都是基于工作内存进行的，即首先需要将外存中的 R 包、数据等，加载到工作内存中，然后才能够进行后续的计算。基于这种工作机制，R 成功启动后必须首先自动将基础包加载到工作空间中，旨在提供最基本的程序运行环境。

说明：

① R 的基础包由很多“小包”组成。R 启动后仅自动将其中的部分“小包”加载到工作内存，还有些“小包”是需要用户根据分析需要适时地、选择性地手工加载的。