R语言在统计中的应用

薛毅 陈立萍 | 编著

Application of
R Language in Statistics

用R轻松实现**数据挖掘**

学懂分析，解决统计中的烦杂**计算问题**

从**学习统计过程**出发，全面掌握**R软件**的使用

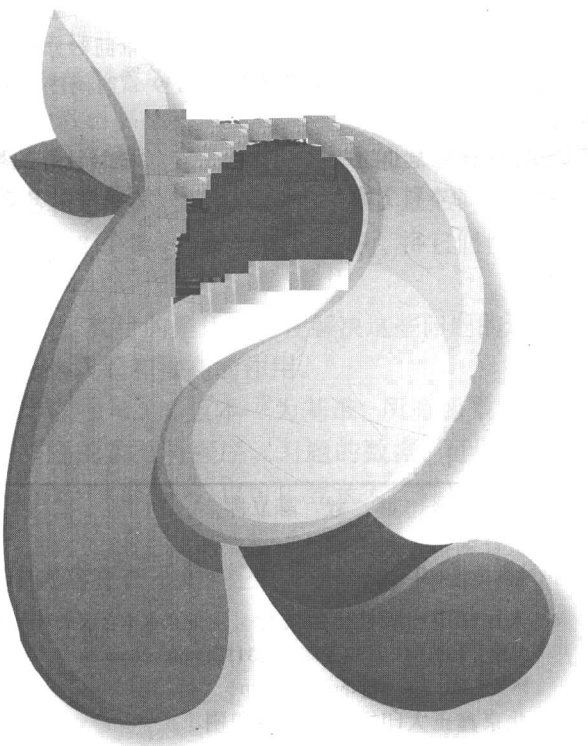
中国工信出版集团

人民邮电出版社
POSTS & TELECOM PRESS

R语言在统计中的应用

薛毅 陈立萍 | 编著

Application of
R Language in Statistics



人民邮电出版社
北京

图书在版编目（CIP）数据

R语言在统计中的应用 / 薛毅，陈立萍编著. -- 北京 : 人民邮电出版社，2017.4
ISBN 978-7-115-44395-3

Ⅰ. ①R… Ⅱ. ①薛… ②陈… Ⅲ. ①统计分析－统计程序 Ⅳ. ①C819

中国版本图书馆CIP数据核字(2016)第313688号

内 容 提 要

本书按照统计学的结构来编排，在介绍完相关的统计知识后，着重介绍如何用 R 求解统计问题。因此，本书并不是简单的 R 使用手册，而是将统计知识、统计模型及 R 的求解过程融为一体的教科书。

本书共 9 章，分别是：第 1 章绪论，介绍统计学及 R 的基本概念；第 2 章 R 语言入门，介绍 R 软件的下载与安装，以及 R 使用的基本方法；第 3 章数据的描述性分析，介绍描述数据的图形和数值方法；第 4 章概率、随机变量及其分布，介绍概率的基本知识和几个重要的分布；第 5 章参数估计与假设检验，介绍参数估计与检验的基本方法；第 6 章非参数检验，介绍秩检验、分布的检验及列联表检验；第 7 章方差分析，介绍单双因素方差分析的方法；第 8 章回归分析，介绍回归分析中参数的计算与检验、回归方程的诊断，以及回归分析的建模方法；第 9 章时间序列分析与预测，介绍时间序列最基本的建模与预测方法。

本书可作为经济管理、统计等专业的本科生学习统计学、统计计算的教材或教学参考书，也可作为理、工、农、医、生物等专业的本科生或者相关专业的技术人员学习 R 的教材或参考书，还可作为数学建模竞赛培训的辅导书。

◆ 编　　著　薛　毅　陈立萍
责任编辑　孙燕燕
责任印制　杨林杰

◆ 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：22.75　　2017 年 4 月第 1 版
字数：572 千字　　2017 年 4 月北京第 1 次印刷

定价：59.80 元

读者服务热线：(010)81055256　印装质量热线：(010)81055316
反盗版热线：(010)81055315

前　言

R 是现今最受欢迎的数据分析和可视化软件之一, 它的发展经历了 3 个阶段: 由于它的开源代码和免费而在学术界流行; 相关的研究人员将 R 带入他们的工作环境中; 在商业分析中, 非统计专业的分析人员开始广泛使用。因此, 目前越来越多的人员在使用 R 来分析数据, 使 R 成为学习统计的必备工具。

本书的目的是将 R 语言（和 R 软件）应用到统计中, 解决统计中烦杂的计算问题, 利用 R 在统计计算上的优势, 使数据分析的过程变得简明和清晰。并且, 本书可以使研究者有更多的时间关注研究问题的本质, 或将计算结果应用到实际工作中。

为达到这一目的, 本书将按照统计学（或商务统计学）的章节或结构编排, 在介绍完相关的统计知识后, 着重介绍如何使用 R 来解决本章中的统计问题, 以及应用 R 来解决本章中的应用案例。因此, 本书并不是简单的 R 使用手册, 而是将统计问题、统计知识、统计模型及 R 的求解过程融为一体的教科书。

本书的每章以一个引例作为开始, 提出本章将要讨论的问题, 给出要点。在每章的结束, 会对讨论问题的重点加以概括和总结, 强化本章的知识点。最后, 将应用本章介绍的统计知识, 结合 R 中的相关函数, 对 2~3 个案例进行求解与分析, 将它们作为本章知识点（统计与 R）的综合应用。

本书的各章配有大量的习题, 其目的是让读者在解决这些问题的过程中, 对所学知识, 特别是 R 中相关函数的使用, 起到巩固和提高的作用。

本书所介绍的 R 函数均以 R-3.2.3 版本[①]为基准, 所有函数（包括自编函数）均通过测试, 读者如果需要书中例题的数据与程序、习题的数据, 可以发送电子邮件向编者索取, 邮件地址：xueyi@bjut.edu.cn （薛毅）; 也可以登录人民邮电出版社教育社区免费下载（www.ryjiaoyu.com）。

由于编者水平有限, 书中内容存在不足, 甚至错误之处, 欢迎读者不吝指正。

编　者

2016 年 10 月

于北京工业大学

①每隔一段时间 R 的版本会有一次更新.

目　录

第1章 绪 论

导入案例

为什么是相反的结论

张先生是一位从事实验的工作者, 为了研究动物对颜色的喜好, 他将 10 只小白鼠关在一个笼子内, 并在笼子的两侧各安装一个门, 一个门涂成红色, 另一个门染成蓝色. 当他同时打开两个门时, 他发现 10 只小白鼠中有 7 只从蓝色的门逃出, 另外 3 只从红色的门逃出. 因此, 他断定, 小白鼠更喜欢蓝色. 他的同事李先生看了他的报告后, 对他说, 这个结论不正确, 小白鼠从哪个门逃出, 可能是随机的.

张先生又接到一个测试某种药品是否有毒的试验. 他将这种药喂给 10 只小白鼠, 结果有 3 只死了, 他想说, 这种药品有毒. 但他想起李先生的话, 小白鼠的死亡可能是随机的. 他带着这个结果去问李先生, 李先生明确地告诉他, 这种药品确实有毒.

为什么同样的实验结果会得出两个完全相反的结论呢? 张先生有点糊涂了.

这个问题正是本书要回答的问题. 问题的回答应从两方面考虑: 一是如何建立合理的统计模型; 二是如何对数据进行计算与分析, 以及对计算结果做出合理的解释.

本章要点

- 统计、统计学与统计模型的介绍.
- R 语言与 R 软件.

什么是统计学? 根据《兰登书屋大学字典》(The Random House College Dictionary) 的定义, 统计学是“对用数字表示事实或数据进行收集、分类、分析以及解释的科学”. 简而言之, 统计学就是数据的科学.

什么是 R? R 是进行统计分析、绘图以及统计编程的平台, 是进行统计分析的重要工具, 是现今最受欢迎的数据分析和可视化软件. 同时, 它还是一款免费的开源软件, 从这一点来说, 它比其他软件更有意义. 目前, R 已成为学习统计学的必备工具.

§1.1 统计、统计学和统计模型

§1.1.1 什么是统计

什么是统计? 它是数字相加吗? 是图表、人们的平均收入、物价上涨率吗? 总之, 它是不是对社会和自然的数值描述?

统计是一套科学原理和技术, 用于在可能得到的信息既有限又富于变化时, 从中得出关于总体和过程的结论. 也就是说, 统计是关于从数据中学习的科学.

"不确定的知识 + 所含不确定性量度的知识 = 可用的知识."[①]

这就是学习统计的目的.

§1.1.2 统计学

什么是统计学? 它是科学、技术、逻辑, 还是艺术? 它是一门像数学、物理、化学、生物那样有确切定义的独立的科学吗? 统计学中, 我们研究的现象是什么?

统计学是数据的科学. 它包括数据的收集、分类、概括、整理、分析及解释. 统计学通常应用于两种类型的问题: (1) 概括、描述以及探索数据, (2) 利用样本数据推断被选取样本的数据集的性质.

全国人口普查可以看成是描述统计应用的典型例子, 它涉及数据的收集与整理, 包括全国人口的状况、人口的年龄比例及社会经济特征等. 对于计算机软件的工程师来说, 管理巨大的数据库需要使用统计方法描述数据库. 类似地, 一位环境工程师要利用统计学的方法描述过去一年中每天 PM2.5 的含量等.

致力于数据集的整理、概括以及描述的统计分支称为描述性统计.

有时数据集 (称为总体) 刻画的是一种感兴趣的现象, 但这样的数据在自然状态下无法得到, 或者是代价昂贵, 或者是耗时很长才能得到. 在这种情况下, 我们可得到一个子集 (称为样本), 利用这个样本来推断它的性质.

例如, 一个灯泡厂每天大约生产 50 万只灯泡, 质量控制部门必须检验灯泡的次品率. 这个任务可以通过检验每一只灯泡来完成, 但这样做的花费巨大, 而且有时是不可能的. 另一种方法是从每天生产的 50 万只灯泡中选出 1 000 只, 然后检验这 1 000 只灯泡. 如果这 1 000 只灯泡是以正确的方式被选出的, 那么从中检验的次品率, 可被用于估计全天所有产品的次品比例.

简单地说, 你想知道一锅汤的味道如何, 是咸, 还是淡? 你不必将一锅汤全部喝掉, 品尝一勺就足够了, 当然, 品尝的方法要合理.

利用样本数据对一个很大的数据集做出推断的统计学分支称为推断统计学.

§1.1.3 统计学的基本要素

1. 总体与样本

总体是指与所研究的问题有关的全部个体的集合. 例如, 研究某城市大学生的身高状况, 则总体包括该市全体大学生; 研究一批产品的合格率, 则总体包括该批中的全部产品. 在前面的例子中, 需要研究每天生产 50 万只灯泡的次品率, 则这 50 万只灯泡就是总体.

以一定方式从总体中抽取的若干个体称为样本, 人们也将其中的单个个体称为样本. 样本中所含个体的数目称为样本量. 例如, 在灯泡质量控制中, 从 50 万只灯泡中抽取的 1 000 只灯泡就是样本, 这里的 1 000 就是样本量.

2. 参数与统计量

参数是用来描述总体特征的概括性数字度量, 它是研究者想要了解的总体的某种特征. 例如, 总体的平均值、方差、比例等. 在灯泡质量控制中, 50 万只灯泡的次品率就是研究者想要知道的参数.

①C. R. Rao 统计与真理 —— 怎样运用偶然性. 北京: 科学出版社, 2004.

统计量是用来描述样本特征的概括性数字度量, 它是根据样本数据计算出来的量. 样本是随机的, 因此, 统计量是样本的函数. 例如, 研究者可以通过样本计算出样本均值、样本方差、样本比例等. 在灯泡质量控制中, 1 000 只灯泡的次品率就是样本统计量.

统计推断的任务是从样本统计量推断出总体参数, 例如, 用 1 000 只灯泡的次品率推断出 50 万只灯泡的次品率.

3. 变量

在研究总体和样本的过程中, 会专注于总体试验中一个或多个人们感兴趣的特征或性质, 统计学称这些特征为变量. 例如, 在饮用水质量的研究中, 感兴趣的两个变量是在 100 ml 的水样本中, 氯的残留量及大肠杆菌的数量.

4. 推断的可靠性

在统计推断中, 还有一个需要关心的要素就是推断的有效程度, 即推断的可靠性. 例如, 我们用 1 000 只灯泡的次品率来估计 50 万只灯泡的次品率, 需要给出一个估计误差的界, 这个界是一个数 (如 5%), 估计误差不大可能超过它 (如估计误差不超过 5%). 可靠性度量是关于统计推断不确定程度的一个陈述, 通常是定量的.

5. 统计学的基本要素

描述性统计问题有 4 个要素: (1) 感兴趣的总体或样本; (2) 被研究的一个或多个变量 (总体或样本中感兴趣的特征); (3) 表格、图形或数字概括工具; (4) 确定数据类型.

推断统计问题有 5 个要素: (1) 感兴趣的总体; (2) 被研究的一个或多个变量 (试验中感兴趣的特征); (3) 试验中的样本; (4) 基于包含在样本中信息对总体的推断; (5) 推断的可靠性度量.

§1.1.4 数据的分类

数据类型可分为两类: 定量数据和定性数据.

1. 定量数据

定量数据表示事物的数量或个数, 用数值标度度量. 定量数据还可以细分为计量数据和计数数据.

计量数据属于连续型变量, 它们的取值可以为某个区间内的任意一个实数, 如人的身高和体重, 产品的长度、直径和重量, 股票的价格和市盈率等. 我们对这类数据可以进行计算, 如求和、计算平均值等.

计数数据属离散型变量, 它们在整数范围内取值, 大部分还仅在非负整数范围内取值, 如企业的职工人数、成交股票的股数、单位时间内通过某个交叉路口的车辆数和每天到医院就诊的人数等. 尽管计数数据是离散的, 但我们可以对它们进行各种运算, 如计算均值, 因为每天平均有 13.5 人到医院看感冒是合理的.

2. 定性数据

定性数据没有量的解释, 它们只能是分类或顺序. 定性数据还可以细分为名义数据和有序数据.

当观察值不是数, 而是事物属性时, 也可以用数值来表示, 但这些数只起一个名义作用, 因此, 称其为名义数据. 它们之间没有大小关系, 也不能进行运算. 例如, 人的性别分为男、女, 可以用数“1”和“2”来表示, 在这里“2”和“1”不能比较大小, “1+2”也没有任何意义.

描述事物属性的顺序关系的数据称为有序定性数据, 简称有序数据. 例如, 人的文化程序由低到高可分为文盲、小学、初中、高中、大学和研究生 6 个等级, 分别用 0, 1, 2, 3, 4 和 5 表示. 又如, 对某项服务的评价分为“很满意”“基本满意”“一般”和“不满意”4 类, 可用 4, 3, 2 和 1 表示. 这些数只起到一个顺序作用, 数字之间不能进行运算. 例如, 对服务的评价, 只知道“4”要优于“3”, 但“4 – 3”没有意义.

§1.1.5 统计模型

一个量或几个量的取值受到偶然因素的影响时, 无法用确定的数量关系或函数关系描述它们, 在统计学中, 这些量称为随机量或随机序列. 在这些量之间, 或其自身前后之间往往存在着某种统计依赖关系, 也就是说, 在大量的重复观察或丰富的数据资料中, 存在着相对稳定的规律, 它被称为统计规律.

当这种规律能用某一模型方式描述, 或近似描述时, 称这种随机量或随机序列适合此模型. 这种模型可以通过相应量的实测数据的计算分析而获得估计. 所以, 这种模型称为统计模型.

统计模型的具体形式, 在少数情况下能够依靠被考查的各量的实际背景所决定, 在大多数情况下并非都能如此. 因此, 目前所使用的各种统计模型, 在绝大多数情况下都是对真实统计规律的近似描述. 另一方面来讲, 真实模型形式总是比较复杂, 而实际使用的模型又不能太复杂, 因此, 近似描述手段又是十分必要的.

§1.2 R 语言与 R 软件

§1.2.1 R 语言

R 语言是主要用于统计分析、绘图的语言和操作环境. R 最初是由来自新西兰奥克兰大学的 Ross Ihaka 和 Robert Gentleman 开发的, 因此, 称为 R. 其现在由“R 开发核心团队”负责开发和维护. R 语言是基于 S 语言的一个 GNU 项目, 所以也可以当作 S 语言的一种实现, 通常用 S 语言编写的代码都可以不做修改地在 R 环境下运行.

S 语言是由 AT&T Bell 实验室的 Rick Becker、John Chambers 和 Allan Wilks 开发的一种用来进行数据探索、统计分析、作图的解释型语言. 最初 S 语言的实现版本主要是 S-Plus. S-Plus 是一个商业软件, 它基于 S 语言, 并由 MathSoft 公司的统计科学部进一步完善.

§1.2.2 R 软件

R 软件是 R 语言的实现环境, 是一套完整的数据处理、计算和制图软件系统, 其功能包括数据存储和处理系统、数组运算工具、完整连贯的统计分析工具、优秀的统计制图功能、简便而强大的编程语言、可操纵数据的输入和输出、可实现分支和循环以及用户可自定义功能.

与其说 R 软件是一种统计软件, 还不如说 R 软件是一种数学计算环境. R 软件提供了有弹性的、互动的环境来分析、可视及展示数据. 它提供了若干统计程序包, 以及一些集成

的统计工具和各种数学计算、统计计算的函数, 用户只需根据统计模型, 指定相应的数据库及相关的参数, 便可灵活机动地进行数据分析等工作, 甚至创造出符合需要的新的统计计算方法.

使用 R 软件可以简化数据分析过程, 从数据的存取到计算结果的分享, R 软件提供了更加方便的计算工具, 能帮助用户更好地决策. 通过 R 软件的许多内嵌统计函数, 用户可以很容易地学习和掌握 R 软件的语法, 也可以编制自己的函数来扩展现有的 R 语言, 完成相关的科研工作. 你可以下载其他的扩展程序包, 帮助你完成你的工作或科研所需的计算工作 [①].

本章小结

- 统计、统计学和统计模型.
- 统计学的基本要素: 总体与样本、参数与统计量、变量, 以及推断的可靠性.
- 数据的分类: 定性数据与定量数据.
- R 语言与 R 软件.

习 题

1. 为了调查可乐的消费者是喜欢可口可乐, 还是百事可乐, 在某次的促销活动中, 随机地选择了 1 000 名可乐的消费者进行双盲品味测试. 所有消费者按照品味的喜好程度将品牌 A 和品牌 B 的可乐进行排序. (1) 描述总体与样本; (2) 描述所关注的变量; (3) 描述所需进行的推断.

2. 如果习题 1 中有 56% 的可乐消费者更喜欢可口可乐, 这是否表明在所有可乐的消费者中有 56% 的人喜欢可口可乐, 如何刻画统计推断的可靠程度?

3. 为了提高医院的医疗服务水平, 医院的管理者让康复出院的病人填写一张问卷调查表, 表中包括如下问题, 试分析调查表中数据的类型.

(1) 你住院的时间有多长?

(2) 你是住在哪个病房? 选项有: 内科、外科、心血管科、妇产科、儿科、特护中心.

(3) 你选择看病就医时, 你认为医院的地理位置重要吗? 选项有: 非常重要、很重要、不太重要、根本不重要.

(4) 你住院时, 你的病情是否严重? 选项有: 非常严重、很严重、不大严重、一般.

(5) 你认为你的主治医师的医疗水平如何? 选项有: 医术高明、水平很高、水平高、水平一般.

(6) 你认为医院的护理水平如何? 有 7 个选择: 1, 2, 3, 4, 5, 6, 7, 其中 1 为差, 7 为好.

4. 高速公路桥的检测. 美国联邦公路局 (FHWA) 定期对美国全境内所有的高速公路进行结构检测, 检测数据被录入国家桥况数据库 (NBI). 下面是从库中变量选出的几个, 请判断每个变量是定性的, 还是定量的.

(1) 桥梁的最大跨度的长度; (2) 车道的数量; (3) 是不是收费桥 (是与否); (4) 日平均交通流量; (5) 分道设施的条件 (好、一般和差); (6) 旁道或弯道的长度; (7) 线路的类型 (州际、全美、州、县或市).

①截止到 2015 年 8 月 1 日, CRAN 网站共有 6 957 个 R 包, 涵盖了不同领域的应用.

5. 有结构缺陷的高速公路桥, 参考习题 4. NBI 的数据分析结果可在网上查询, 根据 FHWA 的检测结果, 美国的 608 272 座高速公路桥被按照有结构缺陷、功能过时和安全分类. 大约有 13.5% 的桥被认定为有结构缺陷, 3.5% 被认定为功能过时.

(1) 哪个变量是研究者感兴趣的? (2) 变量是定量的, 还是定性的? (3) 数据分析的是总体, 还是样本? 请解释. (4) NBI 是如何获得数据的?

6. 请登录 R 的主页 (`https://www.r-project.org/`), 了解 R 的最新动态.

7. 请登录 R 的 CRAN 社区 (`http://cran.r-project.org/`), 下载最新版本的 R 软件.

第2章　R 语言入门

导入案例

为什么要使用 R

从我首次接触 R 算起来, 已经有 10 年的光景. 那时我还是 DoubleClick 公司一名年轻的产品研发经理. 我们公司出售管理网络广告销售的软件, 而我当时主要负责库存预测, 根据给定的搜索词、网页或者人口特征来估计广告的点击次数. 我想自己独立地分析数据, 但我们买不起 SAS 或者 MATLAB 这样昂贵的软件. 我尝试着寻找一个开源的软件包, 很快 R 进入了我的视野. 相比现在, 那时的 R 还有些稚嫩, 很多的功能 (如统计函数、绚丽的绘图) 都还不具备. 但是, 它很直观, 易用, 我入迷了. 从那时起, 我一直利用 R 来处理各种各样的问题: 估计信贷风险, 分析棒球比赛统计数据, 或者寻找互联网安全威胁的来源. 从数据中我学习到了很多, 并慢慢成长为一名经验丰富的数据分析师.

资料来源: (美) Josepb Adler. R 语言核心技术手册. 刘思, 等译. 北京: 电子工业出版社, 2014.

本章要点

- R 软件的下载、安装, 以及 R 语言的简单入门.
- R 语言的数据表示方法, 以及读写数据文件.
- R 语言的控制流和程序设计.

R 语言是一种为统计计算和图形显示而设计的语言环境, 是贝尔实验室开发的 S 语言的一种实现. R 是一种针对统计分析和数据科学且功能全面的开源软件, 目前在商业、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.

§2.1　R 软件的下载与安装

对于 R 的初学者来说, 首先要下载 R 软件. R 是免费的, 可在网站

`http://cran.r-project.org/`

下载, 图 2.1 显示的是 R 的 CRAN 社区网页. 对于 Windows 用户, 单击 `Download R for Windows` 进入下一个窗口. 然后单击`base`进入下载窗口[①]. 单击`Download R 3.2.3 for Windows (62 megabytes, 32/64 bit)` 下载 Windows 系统下的 R 软件[②].

R 软件安装非常容易, 运行刚才下载的程序 (如 `R-3.2.3-win`), 然后按照 Windows 的提示, 安装即可.

①`http://cran.r-project.org/bin/windows/base/` 直接进入下载窗口.
②R 软件每隔一段时间会更新一次, 本书使用的版本是 R 3.2.3.

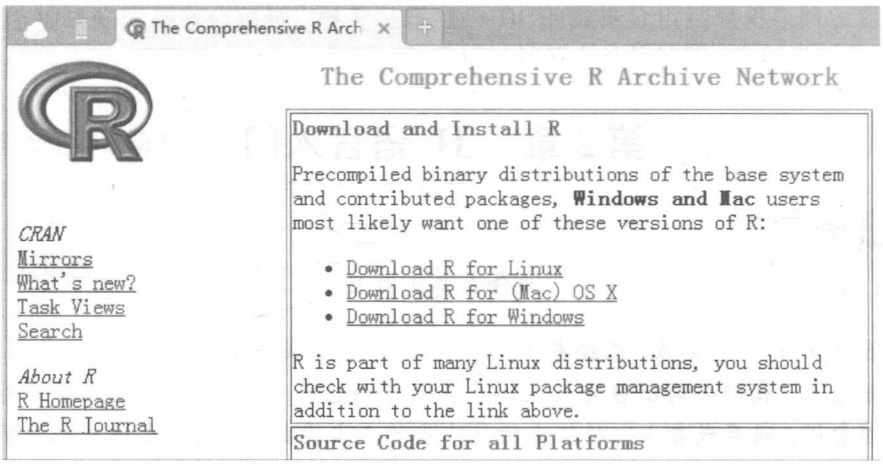


图 2.1 R 的 CRAN 社区

在开始安装后, 选择安装提示的语言, 如中文 (简体), 如图 2.2 所示, 单击“确定”按钮进入安装向导窗口. 单击“下一步”按钮进入“信息”窗口, 你可浏览相关信息, 然后再单击“下一步”按钮进入“选择目标位置”窗口. 你可单击“浏览 (R)”选择安装目录 (默认目录为 C:\Program Files\R\R-3.2.3), 接着单击“下一步”按钮进入“选择组件”窗口 (见图 2.3), 并根据所要安装计算机的性能选择相应的组件. 如果在 Message translations 前面打钩, 则使用中文系统说明. 选择后, 单击“下一步”按钮进入“启动选项”窗口.

在“启动选项”窗口中 (见图 2.4) 选择“ Yes (自定义启动)”或“ No (接受默认选项)”.

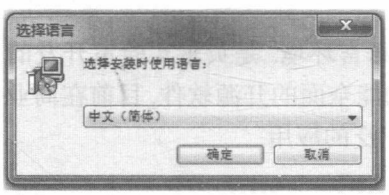


图 2.2 选择安装语言窗口

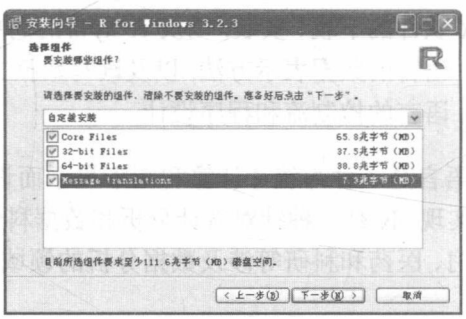


图 2.3 选择组件窗口

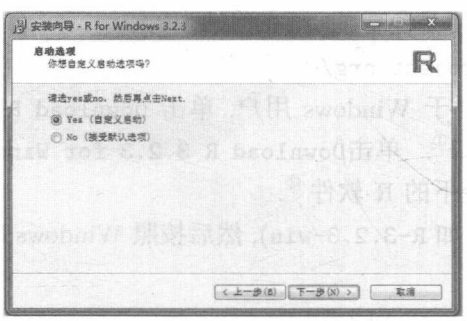


图 2.4 启动选项窗口

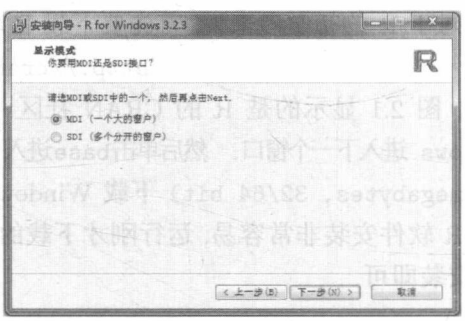


图 2.5 显示模式窗口

如果选择默认选项, 以后的帮助文件将由网页提供. 你可以选择 Yes, 进入“显示模式”界面 (见图 2.5). 在这个窗口中选择“ MDI (一个大的窗口)”或“ SDI (多个分开的窗口)”, 单击“下一步”按钮进入“帮助风格”窗口. 在这个窗口中, 选择“纯文本”, 以后的帮助文件由本地的纯文本形式提供.

单击“下一步”按钮进入“互联网接入”窗口, 选择“标准”, 接着单击“下一步”按钮进入“安装”窗口, 再单击“下一步”按钮进入安装状态. 稍候片刻, R 软件就安装成功了.

§2.2 R 软件的界面

安装完成后, 程序会创建 R 软件程序组, 并在桌面上创建 R 主程序的快捷方式 (也可以在安装过程中选择不要创建). 通过快捷方式或“开始 → 所有程序 →R→R i386 3.2.3”启动 R, 进入工作状态, 如图 2.6 所示 ①.

R 软件的界面与 Windows 的其他编程软件类似, 由下拉式菜单、快捷按钮控件和操作窗口组成, 快捷按钮控件的图形及功能如图 2.7 所示.

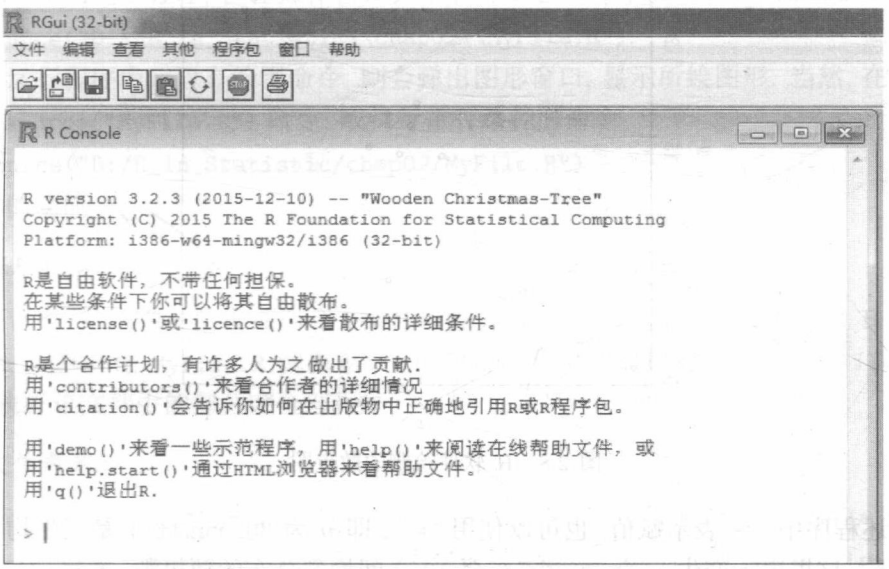


图 2.6 R 软件主界面

图 2.7 快捷按钮控件及相应的功能

①本书只显示中文系统下 R 的运行模式.

§2.2.1 主窗口

主窗口也称为控制台, 或命令窗口, 在提示符 > 下可以直接输入命令得到计算结果. 如:

```
> 2 + 2
[1] 4
> log(2)
[1] 0.6931472
```

显示的 [1] 表示第 1 个数据. 还可以绘图, 例如, 输入一段程序:

```
> n <- 30
> x <- runif(n, 0, 10)
> y <- 5 + 2*x + rnorm(n)
> plot(x, y)
```

这时, 弹出图形窗口 (R Graphics: Device2(ACTIVE)), 给出所绘的图形 (见图 2.8).

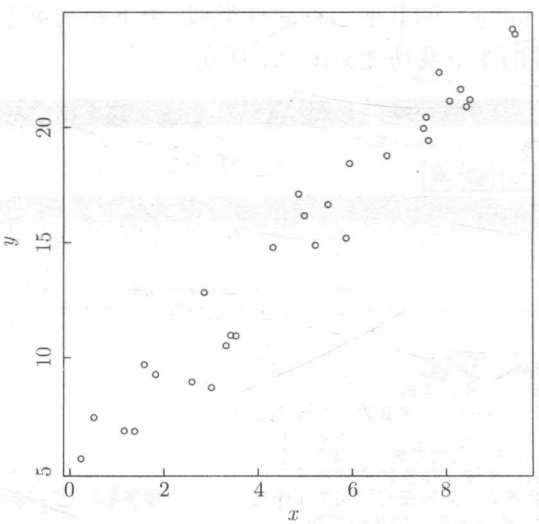


图 2.8 R 软件绘制的散点图

在上述程序中, <- 表示赋值, 也可以使用“=”, 即 n 为 30. runif() 是产生均匀分布随机数的函数, 这里表示产生 n 个 (30 个) 0 至 10 之间均匀分布的随机数. rnorm() 是产生正态分布的随机数, 这里表示产生 n 个 (30 个) 标准正态分布的随机数. 这里产生的 x 和 y 是长度为 10 的向量, plot() 函数绘出自变量为 x、因变量为 y 的散点图.

这些内容也许不能马上理解, 在后面的内容中将会逐步介绍.

在主窗口上面有 7 个下拉式菜单, 分别是“文件”“编辑”“查看”“其他”“程序包”“窗口”和“帮助”, 下面将有选择地介绍部分菜单及菜单中的部分内容.

§2.2.2 文件菜单

单击主界面中的“文件”, 弹出下拉式菜单, 分别是: “新建程序脚本”“运行R脚本文件...”“打开程序脚本...”“显示文件内容...”“加载工作空间...”“保存工作空间...”“加载历史...”“保存历史...”“改变工作目录...”“打印...”“保存到文件...”和“退出”.