



新世纪高校统计学专业系列教材

S-Plus应用统计教程

S-Plus YINGYONG TONGJI JIAOCHENG

王小明 韩小亮 编著



上海财经大学出版社

新世纪高校统计学专业系列教材

S-Plus应用统计教程

王小明 编著
韩小亮



上海财经大学出版社

图书在版编目(CIP)数据

S-Plus 应用统计教程/王小明,韩小亮编著. —上海:上海财经大学出版社,2005.4

(新世纪高校统计学专业系列教材)

ISBN 7-81098-356 3/O · 010

I. S… II. ①王…②韩… III. 统计分析-应用软件, S-Plus-教材
IV. C819

中国版本图书馆 CIP 数据核字(2005)第 027826 号

☐ 责任编辑 徐超
☐ 封面设计 优典工作室

S-Plus Yingyong Tongji Jiaocheng
S-Plus 应用统计教程

王小明 编著
韩小亮

上海财经大学出版社出版发行
(上海市武东路 321 号乙 邮编 200431)

网 址: <http://www.sufep.com>

电子邮箱: webmaster@sufep.com

全国新华书店经销

上海市印刷十厂印刷

上海远大印务发展有限公司装订

2005 年 4 月第 1 版 2005 年 4 月第 1 次印刷

890mm×1240mm 1/32 6.375 印张 183 千字

印数 0 001—4 000 定价: 19.00 元

(本书含光盘一张)

前 言

现代统计技术的蓬勃兴旺是与高速计算机的发展与日俱进的。因为有了价廉物美的高速计算机,现代统计方法解决问题能力的深度和广度都有了很大的拓展。而统计软件正是我们应用统计方法不可或缺的工具。

时下市场上统计软件名目繁多,各具特色,令人有无所适从之感。然而以 S 语言环境为基础的 S-Plus 软件由于其鲜明的特色一直以来都受到了统计专业人士的青睐,是国外大学里相当标准的统计软件。可惜在中国它的普及程度并不高,也没有合适的中文教材。

S-Plus 在中国的代理人上海宏能软件有限公司委托笔者尽快编写一本 S-Plus 的初等教材,笔者感到不胜惶恐。我们使用这种软件当然已经颇有年头,笔者在牛津大学工作时与该软件的高级开发人员也有相当接触,但我们与 S-Plus 的专家水平尚有距离。我们在这里抛砖引玉的目的是:为学生们和广大打算使用 S-Plus 软件进行统计分析的研究人员提供一本急需的基础教科书。请专家和广大使用者多加批评,不吝赐教。

这不是一本纯粹的统计学教科书,也不是一本纯粹的软件教科书。我们设定的读者对象是有一些统计知识,准备应用 S-Plus 软件解决实际问题的研究人员和学生。我们不要求读者有任何软件背景知识。如果读者有一些计算机语言的知识,则对本课程的学习是大有裨益的。我们的主要参考书是 The Basics of S-Plus (Third Edition), by A. Krause & M. Olson 和 Modern Applied Statistics with S (Fourth Edition), by W. N. Venables & B. D. Ripley。

本书的编写相当匆忙,在这里我们要特别感谢我们的学生冯玲、彭

定忠、刘燕、殷书峰为我们做的准备工作。我们还要感谢上海宏能软件有限公司给予我们的帮助和支持。

上海财经大学统计系

王小明

韩小亮

2005 年 2 月

目 录

前 言	(1)
第一章 导 言	(1)
§ 1.1 S-Plus 特点简介	(1)
§ 1.2 S 和 S-Plus 的历史	(2)
§ 1.3 符号惯例	(4)
* 附:不同操作系统上的 S-Plus	(5)
第二章 S-Plus 的桌面环境	(7)
§ 2.1 主窗口	(8)
§ 2.2 图形使用界面的其他部分	(11)
§ 2.3 图形使用界面的一些基本操作	(15)
第三章 数据和运算	(23)
§ 3.1 基本知识	(23)
§ 3.2 数据类型	(32)
§ 3.3 S-Plus 数据对象的构造	(35)
§ 3.4 常用函数	(64)

第四章 S-Plus 图形功能介绍	(68)
§ 4.1 点击式作图	(68)
§ 4.2 S 语言作图功能介绍	(83)
§ 4.3 Trellis 作图	(94)
第五章 探索性数据分析	(98)
§ 5.1 描述性数据分析	(98)
§ 5.2 置信区间和假设检验	(111)
§ 5.3 缺省值和无穷值	(121)
第六章 统计模型	(124)
§ 6.1 一元线性回归模型	(125)
§ 6.2 多元线性回归模型	(134)
§ 6.3 方差分析模型	(147)
§ 6.4 广义线性回归模型	(160)
第七章 编程简介	(172)
§ 7.1 编程	(172)
§ 7.2 函数编制	(175)
§ 7.3 程序和函数调试	(178)
第八章 数据的输入输出	(183)
§ 8.1 使用命令输入输出	(183)
§ 8.2 使用菜单输入输出	(190)
参考文献	(195)

第一章 导 言

本章简单介绍 S-Plus 的特点和历史,用以澄清什么是 S 语言、什么是 S-Plus 系统以及 S-Plus 系统不同版本之间的区别。

§ 1.1 S-Plus 特点简介

S-Plus 统计软件是美国 Insightful 公司的旗舰产品,是世界最流行的统计分析软件之一,尤其为专业人士所喜爱。它主要用于统计分析、统计作图、Data Mining 及 Machine Learning 等等,为人们提供了一个弹性的、互动的可视化环境来分析和展示数据。

1. S-Plus 的最大特点在于它既可以像所有通用统计软件一样通过简易的界面操作来实现基本的输入输出、统计分析和统计作图,又可以由它所特有的 S 高级语言所支持的环境来完成各种复杂的特定任务。S 语言的扩展功能使得它可以很容易地实现一种新的统计方法,或编制和调试一个专用的程序。最新 S-Plus 6 的基础是 S 语言 version 4,它是一个获奖的面向对象的程序语言,由朗讯科技贝尔实验室(Lucent Technologies' Bell Labs)特别为数据的可视化、探索和程序化而开发的,现为 Insightful 公司所拥有。

2. S-Plus 提供了最全面的各种统计模型和分析手段,包括各种线性、非线性回归分析模型,随机效应模型,生存分析模型,方差分析,聚类分析,时间序列分析等等。Insightful 公司一直在密切关注统计学发展的最新动态,保持与全球多个领域的顶尖统计学家的紧密合作,因此,各种统计方法上的进展都会很快被扩展到 S-Plus 中。

3. S-Plus 具有很强的图形功能,拥有独一无二的可视化交互式图形显示,包括著名的 Trellis 图形分析。

- S-Plus 产生的图形可以是面向对象的,你可以修改图形中的每一个细节,来产生适合于报告用的图形。

- S-Plus 提供广泛的可供选择的 2D 及 3D 图形种类,包含广泛的图形库,你可以方便地选择你希望得到的图形形式,轻易地调整图形的属性,例如线条、颜色和字体等。

- 创新的 S-Plus Graphlets 技术可以产生互动式的图形,让使用者能够通过图形逐层下探来观察和探索数据或者链接到 Web 网页,让图形及时地与你的数据互动。

4. S-Plus 的兼容性极好,可以直接实现与 Excel、Lotus、Access、SAS、SPSS 等常用软件的数据转换,也可以方便地插入 C 语言或 Fortran 语言编制的计算机程序。

§ 1.2 S 和 S-Plus 的历史

为了更好地了解 S-Plus 系统,对该系统的发展历史做一个简单的回顾是有益的。以下我们将简要介绍 S 和 S-Plus 的历史发展概况,澄清二者之间的差别。

S 语言是 20 世纪 70 年代末 80 年代初在美国电话电报公司贝尔实验室开发出来的。瑞克·贝克、约翰·詹波斯和阿尔·威尔克斯这三位 S 语言的主要创造者,他们组成了 S 语言开发的核心小组。

1976 年或许可看作是 S 的创立之年。起初,“S 系统”仅包括程序库和一个界面。为以保证核心代码不被修改,系统被设计为封闭式的。1981 年,S 小组决定用 C 语言重新编写系统,并移植到 UNIX 操作系统中去。此时的 S 系统是开放式的,从 1981 到 1993 年,原代码可以被贝尔实验室外的有兴趣人员得到。

在接下来的一些年份里,很多人都对 S 系统的开发做过贡献。统计学家和数据分析人员应用该系统的兴趣与日俱增,贝尔实验室外的人们也兴致盎然地参与到开发中来。S 和 S-Plus 这种开发方式沿用至今并已发展到非常庞大的程度。

值得一提的是,S 发展中重要的几步都是以书为标记的,今天 S 用户仍然会谈及褐皮书时期、蓝皮书时期和白皮书时期。绿皮书则是最近的里程碑。作为 S 系统的参考书,它们对 S 系统的应用和发展起到巨大的推动作用。迄今为止,S 系统共有 4 版。第 4 版的 S 出现在 1998 年,是由朗讯科技贝尔实验室特别为数据的可视化、程序化和数据探索而开发的,这些被详细记录在詹波斯(1998)的绿皮书里。

1987 年,为了推广 S,道格拉斯·马丁在西雅图的华盛顿大学创立了 StatSci 公司。1994 年,StatSci 公司成了 MathSoft 公司的数据分析产品部。

公司为 S 增加了更多新的功能,并移植到其他硬件平台上,而且对技术上和统计上的问题研究提供了必要的支持。S 的这个提高版本有了新的名字:S-Plus。

S-Plus 使得 S 在非技术人员中流行起来。StatSci 将 S-Plus 移植到非 UNIX 平台上;1989 年将 S-Plus 发布到 DOS 系统上,1993 年又将其发布到 Windows 系统上。直到版本 3,S-Plus 对 DOS/Windows 和 UNIX 提供的功能实质上是一样的。

第 4 版本的 S-Plus 只向 Windows 平台发布。1998 年秋,发布了面向 Windows 的第 4.5 版本 S-Plus,该版本被分成包括所有功能的“专业版”和只提供菜单功能的“标准版”。国内最早从国外购买的 S-Plus 系统正是第 4.5 版本的 S-Plus。

第 5 版本的 S-Plus 在 1998 年末和 1999 年只面向 UNIX 系统发布。它基于詹波斯 1998 年描述的第 4 版 S,并且第一次支持安装在英特尔个人电脑上的 Linux 系统。

基于 Windows 系统的 S-Plus2000 在 1999 年中期发布,而且依然基于第 3 版 S。

2000 年,Mathsoft 公司的数据分析产品部(StatSci 的前身)从该公司中分离出来。新公司被命名为 Insightful 公司。直到这时,S-Plus 才正式获准在中国大陆推广和发行。

Insightful 分别在 2000 年和 2001 年发布了面向 UNIX 和 Win-

dows 的第 6 版 S-Plus。并且第一次实现了基于同一代码基础开发出两个系统的版本。UNIX 系统版本有基于 Java 语言的图形化用户界面,并且支持一些特殊的视窗组件,如读取 Excel 数据文件、把图形写为图元文件格式等。S 依然是系统的核心,并且 S 核心开发组继续在 S 系统上工作。S 所有的功能都已包含在 S-Plus 中并得到提高,今天 S 系统已不再公开可见了。

在本书的后续部分,我们将以 Windows 版本下的 S-Plus 6 作为标准来加以介绍。

§ 1.3 符号惯例

在开始之前,我们先介绍一下本书中运用的一些符号惯例。在 Commands Window 视窗,你将看到一个大于号“>”提示你输入新的命令,这是一种交互式的命令输入方式。如果一个命令在输入中超过了一行,S-Plus 将会确定该命令是不完整的并且要求你完成。提示符将变为加号“+”,并且你可以输入命令的余下部分,如:

```
>2*3+5*(7-3  
+)
```

为使“+”不会被意外地当作加号运算符键入,在本书出现的程序中,我们忽略连续提示符。对于更长的命令语句,我们也忽略提示符,因为我们并不采用交互式输入命令而是采用批处理方式输入。

在书写命令时,我们有时需要包括一些描述性文本,或者对命令进行必要的注释。在符号“#”之后命令行结束前的所有语句都被看作是注释,而不被看作是命令。

这些惯例汇总在表 1-1 里。

表 1-1 符号惯例

惯 例	解 释
>	S-Plus 提示符
+	命令延续到下一行(被 S-Plus 显示但本书省略)
Commands	打字机字体中显示的 S-Plus 命令
#	注释符,表示注释的开始
Ctrl-C	按下键 Ctrl 和字符键 C
Menu	进入菜单并且提到的按钮也是这种字体

由于 S-Plus 是一个非常人性化的软件系统,有许多符号和惯例是不言自明的,即使与本书在细节上存在差异,也不会影响到对本书内容的理解。

* 附:不同操作系统上的 S-Plus

目前, S-Plus 支持两类主要的操作系统: UNIX(以及其大多数变体)和 Windows(以及其大多数变体)。S-Plus 对内存和硬盘大小都有最小的要求说明。这些要求依赖于具体的版本和操作系统。对于完全安装, 300M 硬盘空间就足够了(包括 PDF 参考手册等)。为了合理运行, 对同时运行的每个用户, 128M 内存是最低要求。大型数据集或是高强度运算将需要更多的内存。顺便提一下, S-Plus 在一段时间内将根据需要动态地消耗或释放内存。因此, 它不会溢出内存, 直到不再有内存和交换空间可利用为止。如果 S-Plus 耗尽内存, 操作系统将指定虚拟内存(如硬盘空间)作为替代。机器应该配备合理数量的内存, 这将大大降低运行时间。为了改善运行, 内存是第一加速器。如果你不满意运行状况, 可以在执行命令过程中查看永久内存访问, 或者利用检测工具(如 UNIX 的 top 或 Windows 的任务管理器)来跟踪交换行动。

不同版本间的差异

S-Plus 在 UNIX 和 Windows 两个版本间的执行有些差异。最显著的差异就是用户界面。Windows 版本的 S-Plus 有一个典型的 Windows 界面。很多功能可以通过鼠标点击来实现。产生的图形是可编辑的,利用鼠标点击就可达到图形编辑的目的。通过复制和粘贴,系统就可实现与 Windows 应用程序(如 Excel 和 PowerPoint)间数据和图形的转移。

UNIX 版本的 S-Plus 也有一个图形化用户界面,不过它在外观和感觉上稍微不同于 Windows。因为它在 Java 环境下执行,它没有工具条,而且生成的图形不可被编辑。

在 Windows 版本下的 S-Plus,其标准的图形装置被称为图表。每个图表可以容纳多个页面。如果某些 S-Plus 代码产生了多幅图片,它们将被显示在不同的页面里,用户可以通过鼠标点击在它们之间转换。UNIX 版本的 S-Plus 没有图表装置。如果一个新的图形产生在活动装置中,那么原有的页面将被清除,原来的图形也会丢失。不过最新引入的 Java 装置也拥有了容纳多幅图片的能力。

UNIX 版本的 S-Plus 运行时可以通过一个连接到系统的共享对象文件的编译,允许简单集成 C 和 Fortran 语言代码。Windows 版本的 S-Plus 则允许使用动态链接库。

Windows 版本的 S-Plus 有两个版本,专业版和标准版,其中标准版不提供命令行接口。

UNIX 版本的 S-Plus 有一个特殊的版本称为“应用服务器”。这一新工具允许通过 Java 界面进入 S-Plus(包括命令行也可以通过类似于原始命令行的 Java 界面进入)。应用服务器允许用户通过网页浏览器进入 S-Plus。

还有其他一些差异,但多数并不为用户直接观察到。除了极少数情况外,在 Windows 版本的 S-Plus 下运行的 S-Plus 程序代码可以在 UNIX 版本的 S-Plus 下运行,反之亦然。

第二章 S-Plus 的桌面环境

本章中我们简要介绍一下 S-Plus 的操作界面,它的功能和一些基本操作如输入输出数据、画图和统计分析等等。这些运作的详细内容将在后面的有关章节中叙述。

S-Plus 成功启动之后,首先呈现在用户面前的就是工作桌面,它由标题栏、菜单栏、常用工具栏、主窗口和状态栏五部分组成,如图 2-1 所示。这五部分就构成了 S-Plus 的图形使用界面(Graphical User Interface,简称 GUI),使用者只需要点击鼠标就可以实现许多功能。

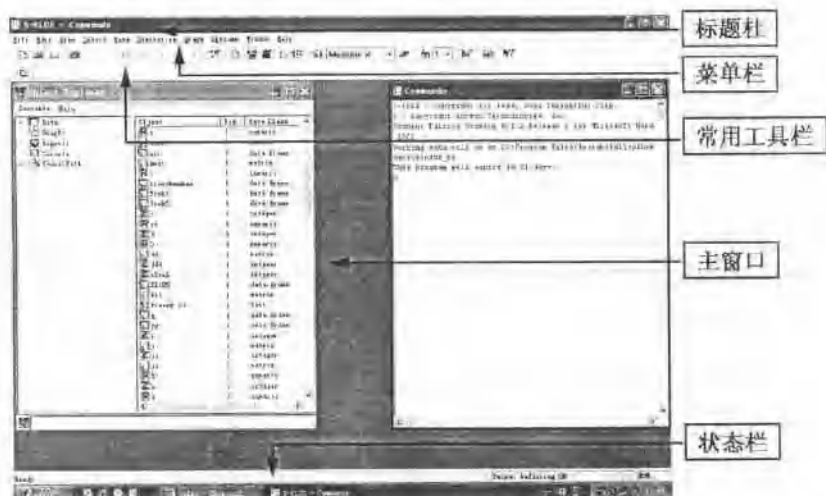


图 2-1 S-Plus 工作桌面

§ 2.1 主窗口

主窗口有四种: Command Window, Object Explorer, Script Window 和 History Log。

§ 2.1.1 Command Window

Command Window 通常出现在 S-Plus 的初始界面上(图 2-1 中右面的主窗口),它也可以通过用鼠标点击常用工具栏(toolbar)中的标志来打开或者从菜单栏的 Window 栏目下打开。Command Window 事实上是 S-Plus 系统的核心。每一个运用菜单和鼠标完成的操作都可以在 Command Window 中通过命令来实现。此外,很多函数只能在 Command Window 中运行。我们在后面的章节中将系统地介绍 Command Window 中运行的各种命令和函数。

§ 2.1.2 Object Explorer

Object Explorer 也通常出现在 S-Plus 的初始界面上(图 2-1 中左面的主窗口),它同样可以通过用鼠标点击常用工具栏中的标志来打开。在 Object Explorer 窗口可以看到当前系统中所有数据、函数、图像等等的总览。在 Object Explorer 窗口中还可以对数据进行编辑。用鼠标双击 Object Explorer 窗口中数据的名称即可打开该数据。

§ 2.1.3 Script Window

Script Window 是 S-Plus 系统中非常有用的工具。就功能而言,Script Window 中运行的命令和函数与 Command Window 中运行的命令和函数并无二致。惟一的区别只在于:Script Window 中的命令是“批量”执行的,而 Command Window 中的命令是逐个执行的。然而由于在 Script Window 中命令可以轻而易举地进行修改和运行,整个程序又可以作为 Script file 文件储存起来,还可以同时打开数个

Script Window, 它的灵活性和机动性就大大增加了。

建立一个 Script Window 的方式是:

- (1) 从菜单栏中选择 File ▸ New。
- (2) 选择 Script File。
- (3) 单击 OK。

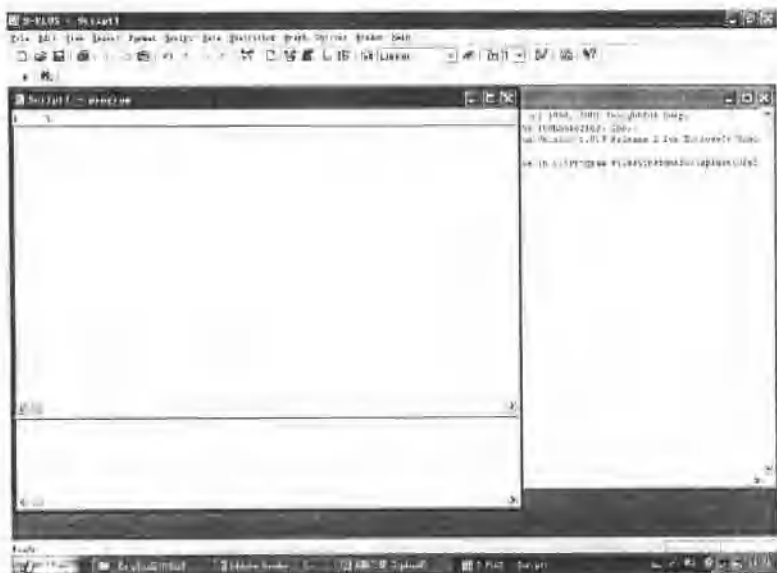


图 2-2 Script Window

S-Plus 的工作界面上就会出现一个如图 2-2 所示的 Script Window。Script Window 分成两个部分:上半部分用来书写命令,下半部分是命令执行的结果输出。倘若要执行 Script Window 中的所有命令,用鼠标点击左上角的“运行”键(用一个黑三角表示)。如只需要运行部分命令,则先用鼠标勾描出所需执行的命令,再点击左上角的“运行”键即可。(键盘上的 F10 键与运行键等价,此外还可以在菜单栏的 script 栏目中点击运行键)。图 2-3 所表示的是执行了 Script Win-

dow 中的后两个命令。



图 2-3 运行 Script Window 的命令

存储 Script Window 的命令到一个文件的方式是：

- (1) 点击这个 Script Window 的上半部分以使它“active”。
- (2) 从菜单栏中选择 File ▸ Save As。
- (3) 指定这个文件的名称和地址。
- (4) 单击 Save。

默认的文件格式是 *.ssc (S-Plus script file)。

§ 2.1.4 History Log

用鼠标点击常用工具栏中的有关标志可以打开 History Log 窗口 (见图 2-4)。在 History Log 窗口中可以看到程序运行的历史信息。