

统计分
析
软件系列

SAS 8.2 统计 应用教程

薛富波 张文彤 田晓燕 编



兵器工业出版社



北京希望电子出版社
Beijing Hope Electronic Press
www.bhp.com.cn

统计分析
软件系列

SAS 8.2 统计 应用教程

薛富波 张文彤 田晓燕 编



兵器工业出版社



北京希望电子出版社
Beijing Hope Electronic Press
www.bhep.com.cn

内 容 简 介

本教程基于 SAS 系统的 8.2 版本写成, 全书分成四大部分, 第一部分(前 5 章)为 SAS 8.2 的入门篇, 主要介绍了 SAS 的基本界面操作、编程基础和数据管理入门等方面的内容; 第二部分(第 6 章到第 15 章)为统计应用篇, 介绍了在 SAS 中各种基本统计分析和常用高级统计分析的编程实现方法; 第三部分(第 16 章到第 19 章)为多元分析篇, 介绍了多元统计模型在 SAS 中的编程实现方法, 包括了典型相关与对应分析等近年来应用较广的方法; 第四部分(第 20 章到第 24 章)为操作进阶篇, 介绍了对专业用户非常重要的 SAS 系统高级功能, 如宏语言、SQL 等, 其中 IML 矩阵语言、ODS 功能和 DDE 的自动报表功能在国内均属首次介绍。教程中的大部分内容都是几位编者基于对 SAS 的使用经验与理解编写而成, 大都经过实际应用的检验, 因此教程内容具有较高的客观性和实用性。

本书既可作为大专院校统计专业师生的入门教材, 也可作为有关数据管理与分析处理专业人员的参考用书。

图书在版编目(CIP)数据

SAS 8.2 统计应用教程 / 薛富波, 张文彤, 田晓燕编.

— 北京: 兵器工业出版社; 北京希望电子出版社, 2004.8

(统计分析软件系列)

ISBN 7-80172-217-5

I. S… II. ①薛… ②张… ③田… III. 统计分析

—应用软件, SAS 8.2 IV. C812

中国版本图书馆 CIP 数据核字(2004)第 035269 号

出 版: 兵器工业出版社 北京希望电子出版社

邮编社址: 100089 北京市海淀区车道沟 10 号

100085 北京市海淀区上地信息产业基地 3 街 9 号

金隅科技中心 C 座 610-612

发 行: 北京希望电子出版社

电 话: (010) 82702660 (发行) (010) 62541992 (门市)

经 销: 各地新华书店 软件连锁店

印 刷: 北京媛明印刷厂

版 次: 2004 年 8 月第 1 版第 1 次印刷

封面设计: 王 翼

责任编辑: 王 琦 宋丽华 郭淑珍

责任校对: 向 云

开 本: 787×1092 1/16

印 张: 33

印 数: 1-5000

字 数: 766 千字

定 价: 48.00 元

(版权所有 翻印必究 印装有误 负责调换)

序

自从有了人类的生产活动，就有了统计。自从有了统计，就有了统计计算。统计计算工具从古时的“结绳记事”，经历了笔、纸张、算盘、对数表、乘方开方表、计算尺、手摇式机械计算机和电子计算器（Calculator）等数千年的发展过程，今天已发展到网络化计算机时代。

统计计算的历史，也折射出统计学发展的历史。独立发现正态分布的德国数学家 Gauss (1777-1855)，在其孩提时代回答过一道数字计算题： $1+2+3+\cdots+100=?$ 他巧妙地把计算过程看作 50 个 101，即 $1+100=101$ ， $2+99=101$ ， $\cdots$ ， $50+51=101$ ，快速做出答案 5050，表现出他极高的统计计算天赋。

最先提出“相关与回归”的英国人类学家 Galton，在人类学和优生学研究中萌发的统计学思想，其中一个重要思想是“在任何时候你都可以做的事——计数”（Whenever you can, count）。小样本均数 Student- t 检验的发现者、英国生物统计学家 Gosset (1876-1937) 为了解决 t 检验的理论和应用问题，通过使用卡片进行抽样试验和计算统计量得出 t 分布曲线，成为 Monte-Carlo 计算方法的先行者。现代实验设计方法和统计分析技术的奠基者、英国生物统计学家 Fisher (1890-1962) 在 20 世纪 20 年代撰写他的实验设计和统计方法专著时，因不可能计算出所有小概率对应的 t 分布和 F 分布的临界值，只好给出了 $\alpha=0.05, 0.01$ 时所对应的 t 分布和 F 分布的临界值表。令 Fisher 未曾想到的是，这种因统计计算方法限制而人为设定的两个数字，至今仍然被许多人当作小概率的惟一数量标准。因此，Fisher 的同事、 χ^2 检验“Yates 校正”公式的提出者 Yates (1902-1994) 强烈主张统计学家使用电子计算机，并且说：“优秀的理论统计学家必须学会计算，并且需要最好的计算工具”（To be a good theoretical statistician one must also compute, and must therefore have the best computing aids.）。Yates 对 20 世纪 60 年代初英国建立计算机学会发挥了重要的作用，并且还担任了两年的英国计算机学会主席。

将计算机用于统计计算，主要通过统计软件实现，其中最具有代表性和全球用户最多的统计软件是 SAS 软件。经过近半个世纪的发展，SAS 已经由单一的

统计软件发展为大型决策支持系统，由大型计算机的专用软件发展为家庭和办公室普及的 PC 机软件。利用 SAS 软件进行数据处理和统计分析，不仅能得到各种检验统计量对应的确切概率值(不必再查 Fisher 和 Yates 等制定的界值表)，而且许多复杂的统计计算通过轻点鼠标就可以完成。SAS 软件的发展和普及，为各个领域统计方法的运用开辟了更广阔的天地，相信本书的出版发行对提高各个领域、特别是医学研究领域研究人员 SAS 的应用水平起到积极的促进作用。

本书主要介绍了 SAS 软件的使用方法。严格按照各章介绍的操作步骤，可以很快得到数据的处理结果。但要真正“用好、用活”SAS，把 SAS 作为数据处理和统计分析的工作平台，一方面要掌握基本的统计知识，理解各种计算步骤和统计指标的含义；另一方面要理论结合实际，把电脑里资料的研究背景和设计方法搞清楚。有“用”的需求，才会有“学”的愿望。本书的编者大多是因研究工作和教学工作的需要才接触 SAS、使用 SAS 和钻研 SAS 的，所以本书的内容贴近实际、编排合理、便于检索，不仅可以作为学习 SAS 的入门书，对于已经熟悉 SAS 的读者，还可以作为手册和工具书使用。

徐勇勇

前 言

SAS 系统是大型集成的应用信息系统,拥有完备的数据访问、数据管理、数据分析和数据呈现功能。尤其是数据分析这一方面, SAS 以其强大完善的数据管理与统计分析功能被公认为国际上标准的统计分析软件,在某些领域,比如新药临床试验中 SAS 已经成为数据处理的首选工具。有关于 SAS 的种种优点,这里我们不再多说,相信随着读者对 SAS 的理解和应用的加深,每个人都会有自己的体会。

在中国, SAS 系统的使用者近些年增长很快,有关 SAS 的教程近几年来也纷纷面世,但是多数介绍的都是其 6.12 版,内容上也大多以基础操作和基本的统计分析方法为主,而每次版本的升级, SAS 都会加入许多新的功能。鉴于此,编者决定结合自身的使用与教学经验,为使用 SAS 的专业统计分析人员提供一本内容全面又简单易懂的参考书。

本教程基于 SAS 系统的 8.2 版本写成,全书分成四大部分,第一部分(前 5 章)为 SAS 8.2 的入门篇,主要介绍了 SAS 的基本界面操作、编程基础和数据管理入门等方面的内容;第二部分(第 6 章到第 15 章)为统计应用篇,介绍了在 SAS 中各种基本统计分析和常用高级统计分析的编程实现方法;第三部分(第 16 章到第 19 章)为多元分析篇,介绍了多元统计模型在 SAS 中的编程实现方法,包括了典型相关与对应分析等近年来应用较广的方法;第四部分(第 20 章到第 24 章)为操作进阶篇,介绍了对专业用户非常重要的 SAS 系统高级功能,如宏语言、SQL 等,其中 IML 矩阵语言、ODS 功能和 DDE 的自动报表功能在国内均属首次介绍。教程中的大部分内容都是几位编者基于对 SAS 的使用经验与理解所编写而成,大都经过实际应用的检验,因此教程内容具有较高的客观性和实用性。

本书既可作为大专院校统计专业师生的入门教材,也可作为有关数据管理与分析处理专业人员的参考资料。

由于在国内 SAS 的用户多数为医学统计学背景,本教程的绝大部分实例均来自于医学统计学教材,特别是徐勇勇主编的教育部面向 21 世纪课程教材——《医学统计学》和孙振球主编的研究生规划教材——《医学统计学》,有关实例的详细情况,各位读者可参考这两本教材的相应内容。但是本书对其他行业需

要使用 SAS 进行统计分析的用户仍然具有非常强的参考价值。

本教程由薛富波、张文彤、田晓燕共同主编，参编人员分别来自第四军医大学卫生统计学教研室（薛富波，刘丹红，杨鹏）、复旦大学公共卫生学院社会医学与卫生统计学教研室（张文彤）、上海日新医药发展有限公司（田晓燕，张子豹）以及第二军医大学卫生统计学教研室（陆健，曹阳，贺宪民），均为 SAS 系统的长期使用者与教学者。各编委所在单位对本书的编写给予了热情帮助，在此谨代表本教程全体编写人员对他们表示诚挚的谢意。作为 SAS 8.2 的正版用户，上海日新医药发展有限公司大力支持了本书的编写工作，编者在此特致谢忱。

本教程的主审徐勇勇教授，现任第四军医大学预防医学系卫生统计学教研室主任、博士生导师、中国卫生信息学会常务理事、中华预防医学会卫生统计学专业委员会副会长、解放军医学会和陕西省医学会卫生统计学专业委员会主任委员等职。他在百忙之中主审了本教程的全部内容并为本书作序，在此谨代表全体编写人员对徐勇勇教授表示诚挚谢意。

为方便读者使用本教程，凡书中涉及的数据及示例程序，我们将全部放在笔者共同开设的医学统计之星网站（<http://www.MedStatStar.com>）的相应板块中，以便于各位读者下载使用。

由于编者水平所限，难免存在错误与疏漏，请各位读者多多提出宝贵意见。相应的意见和建议可以直接发表在“医学统计之星”网站论坛相应的版面，我们会及时回复。

作者

目 录

第1章 SAS 8.2 使用入门..... 1	
1.1 SAS 系统简介..... 1	
1.1.1 概况..... 1	
1.1.2 SAS 的模块化结构..... 1	
1.1.3 SAS 的操作方式..... 2	
1.1.4 SAS 的功能与运行速度..... 2	
1.2 界面操作入门..... 3	
1.2.1 SAS 的启动与退出..... 3	
1.2.2 SAS 的主窗口..... 4	
1.2.3 SAS 的子窗口..... 5	
1.2.4 菜单栏与工具栏..... 7	
1.3 SAS 系统的文件管理..... 10	
1.3.1 SAS 文件和 SAS 的逻辑库..... 10	
1.3.2 使用资源管理器窗口管理 SAS 文件..... 11	
第2章 SAS 8.2 编程入门..... 16	
2.1 SAS 程序的使用常识..... 16	
2.1.1 SAS 语句的基本结构..... 16	
2.1.2 数据步与过程步..... 16	
2.1.3 SAS 程序的书写规则..... 17	
2.1.4 SAS 程序的运行..... 17	
2.1.5 SAS 程序中的注释..... 18	
2.2 SAS 程序的数据步..... 18	
2.2.1 逻辑库和逻辑文件名..... 18	
2.2.2 数据步的基本结构..... 19	
2.3 SAS 程序的过程步..... 20	
2.3.1 过程步的基本结构..... 21	
2.3.2 过程步的语法说明..... 21	
2.4 程序的修改与调试..... 23	
2.5 结构化语句简介..... 24	
2.5.1 分支(条件)语句..... 24	
2.5.2 循环语句..... 26	
第3章 数据管理入门..... 28	
3.1 SAS 数据集的基本格式..... 28	
3.1.1 数据集的基本结构..... 28	
3.1.2 数据集的描述部分..... 28	
3.1.3 数据集的数据部分..... 29	
3.2 数据集的建立..... 30	
3.2.1 直接输入方式建立数据集..... 30	
3.2.2 外部文件读入方式建立数据集 31	
3.2.3 更灵活地建立数据集..... 32	
3.3 数据的导入和导出..... 33	
3.3.1 使用向导实现数据的导入 和导出..... 33	
3.3.2 编程实现数据的导入和导出... 34	
3.4 数据文件的编辑与整理..... 35	
3.4.1 SAS 数据步中的运算符..... 36	
3.4.2 数据集内容的拷贝..... 36	
3.4.3 变量的增加..... 36	
3.4.4 变量的筛选..... 37	
3.4.5 变量名称与类型的更改..... 37	
3.4.6 数据集的排序..... 38	
3.4.7 数据集的记录筛选..... 38	
3.4.8 数据集的纵向连接..... 39	
3.4.9 数据集的横向合并..... 39	
3.5 DATASETS 过程介绍..... 40	
3.5.1 DATASETS 过程的基本格式... 41	
3.5.2 语句说明及示例..... 41	
3.6 利用菜单和窗口操作实现数据管理 42	
3.6.1 数据集的建立..... 42	
3.6.2 数据集浏览格式的更改..... 44	
3.6.3 数据集的编辑整理..... 46	
第4章 SAS/ASSIST 视窗简介..... 48	
4.1 ASSIST 模块概述..... 48	
4.1.1 ASSIST 模块的启动..... 48	
4.1.2 ASSIST 模块功能介绍..... 49	
4.2 ASSIST 模块操作入门..... 51	
4.2.1 打开主对话框..... 51	
4.2.2 对话框的基本操作..... 52	
4.2.3 进一步了解 ASSIST 的界面... 54	

第 5 章 SAS/Analyst 模块操作入门.....	57
5.1 分析员模块概述.....	57
5.1.1 分析员模块的操作界面.....	57
5.1.2 分析员模块的菜单分类.....	58
5.1.3 分析项目的管理.....	59
5.2 使用分析员进行数据管理.....	60
5.2.1 新建数据集.....	60
5.2.2 数据集的打开和保存.....	61
5.2.3 变量列的冻结、隐藏和移动... ..	61
5.2.4 对数据集排序.....	62
5.2.5 计算新变量.....	62
5.3 使用分析员进行统计分析.....	64
5.3.1 主对话框.....	65
5.3.2 子对话框.....	66
第 6 章 数据的统计描述.....	69
6.1 执行描述性统计功能的 SAS 过程... ..	69
6.1.1 MEANS 过程.....	70
6.1.2 SUMMARY 过程.....	78
6.1.3 UNIVARIATE 过程.....	79
6.1.4 TABULATE 过程.....	96
6.1.5 GCHART 过程.....	103
6.1.6 Gplot 过程.....	104
6.2 分析实例.....	105
6.2.1 计量资料描述性统计指标的 计算.....	105
6.2.2 统计图的绘制.....	109
6.2.3 计数资料的统计描述.....	111
第 7 章 单组或两组资料均数的比较.....	114
7.1 均数差别比较的 SAS 过程 ——TTEST 过程.....	114
7.1.1 TTEST 过程的一般格式.....	114
7.1.2 TTEST 过程各条语句的用法 和功能.....	115
7.2 分析实例.....	117
7.2.1 单组样本均数和已知总体均数 比较的 t 检验.....	117
7.2.2 配对设计资料的 t 检验.....	118
7.2.3 成组设计资料两样本均数比较的 t 检验.....	119

7.2.4 两组均数差别比较的 u 检验	120
第 8 章 多组资料均数比较的方差分析.....	123
8.1 有关方差分析的 SAS 过程.....	123
8.1.1 ANOVA 过程.....	123
8.1.2 GLM 过程.....	130
8.2 分析实例.....	138
8.2.1 完全随机设计资料的方差分析	138
8.2.2 随机区组设计资料的方差分析	140
8.2.3 析因设计资料的方差分析.....	142
8.2.4 重复测量设计资料的方差分析	143
第 9 章 分类资料的统计推断.....	147
9.1 处理分类资料的 SAS 过程 ——FREQ 过程.....	147
9.1.1 FREQ 过程的一般格式.....	148
9.1.2 FREQ 过程各条语句的用法 和功能.....	148
9.2 分析实例.....	158
9.2.1 两个率比较 (四格表数据) 的 χ^2 检验.....	158
9.2.2 $R \times C$ 表资料的 χ^2 检验.....	160
第 10 章 非参数统计分析方法.....	162
10.1 非参数统计分析 SAS 过程 ——NPAR1WAY 过程.....	162
10.1.1 NPAR1WAY 过程的一般格式	163
10.1.2 NPAR1WAY 过程各条语句的 用法和功能.....	163
10.2 分析实例.....	166
10.2.1 两独立样本差别比较的秩 和检验.....	166
10.2.2 配对设计资料的秩检验.....	170
10.2.3 完全随机设计多样本差别 比较的秩和检验.....	171
第 11 章 相关分析与线性回归分析.....	174
11.1 相关分析.....	174

11.1.1 执行相关分析的 SAS 过程 ——CORR 过程.....	174	15.2 分析实例.....	278
11.1.2 相关分析实例.....	177	15.2.1 参数法生存分析.....	279
11.2 线性回归分析.....	180	15.2.2 非参数法生存分析.....	282
11.2.1 执行回归分析的 SAS 过程 ——REG 过程.....	180	15.2.3 半参数法生存分析.....	286
11.2.2 线性回归分析实例.....	196	第 16 章 聚类分析	293
第 12 章 非线性回归	202	16.1 有关聚类分析的 SAS 过程.....	293
12.1 NLIN 过程.....	202	16.1.1 CLUSTER 过程.....	293
12.1.1 NLIN 过程的语句格式.....	202	16.1.2 FASTCLUS 过程.....	296
12.1.2 NLIN 过程中各语句的语法 和功能.....	203	16.1.3 ACECLUS 过程.....	298
12.2 非线性回归实例.....	206	16.1.4 VARCLUS 过程.....	300
12.2.1 NLIN 过程分析实例.....	207	16.1.5 TREE 过程.....	302
12.2.2 使用多个初始参数值.....	209	16.2 分析实例.....	305
第 13 章 LOGISTIC 回归	212	16.2.1 样品聚类.....	305
13.1 与 LOGISTIC 回归有关的 SAS 过程.....	212	16.2.2 指标聚类.....	308
13.1.1 LOGISTIC 过程的语句格式	213	第 17 章 判别分析	316
13.1.2 LOGISTIC 过程中各条语句 的用法和功能.....	213	17.1 有关判别分析的 SAS 过程.....	316
13.2 LOGISTIC 回归实例.....	220	17.1.1 DISCRIM 过程.....	316
第 14 章 对数线性模型	228	17.1.2 CANDISC 过程.....	321
14.1 有关对数线性模型的 SAS 过程.....	230	17.1.3 STEPDISC 过程.....	323
14.1.1 CATMOD 过程提供的统计 分析类型.....	230	17.2 分析实例.....	326
14.1.2 CATMOD 过程的语法.....	232	第 18 章 主成分分析与因子分析	341
14.1.3 对数线性模型分析.....	242	18.1 主成分分析.....	341
14.2 对数线性模型实例.....	244	18.1.1 有关主成分分析的 SAS 过程 ——PRINCOMP 过程.....	342
14.2.1 两分类变量实例.....	244	18.1.2 主成分分析实例.....	343
14.2.2 多个分类变量关联性的分析	246	18.2 因子分析.....	346
14.2.3 含有零频数的数据实例.....	250	18.2.1 有关因子分析的 SAS 过程 ——FACTOR 过程.....	347
第 15 章 生存分析	252	18.2.2 因子分析实例.....	353
15.1 有关生存分析的 SAS 过程.....	253	第 19 章 典型相关分析和对应分析	358
15.1.1 LIFEREG 过程.....	253	19.1 典型相关分析.....	358
15.1.2 LIFETEST 过程.....	263	19.1.1 有关典型相关分析的 SAS 过程 ——CANCORR 过程.....	358
15.1.3 PHREG 过程.....	269	19.1.2 分析实例.....	361
		19.2 对应分析.....	366
		19.2.1 有关对应分析的 SAS 过程 ——CORRESP 过程.....	367
		19.2.2 分析实例.....	372
		第 20 章 SAS SQL 简介	381

20.1	SQL 过程的语句格式	381
20.1.1	PROC SQL 语句	382
20.1.2	ALTER TABLE 语句	383
20.1.3	CREATE INDEX 语句	384
20.1.4	CREATE TABLE 语句	385
20.1.5	CREATE VIEW 语句	386
20.1.6	DELETE 语句	387
20.1.7	DESCRIBE 语句	387
20.1.8	DROP 语句	389
20.1.9	INSERT 语句	389
20.1.10	RESET 语句	391
20.1.11	SELECT 语句	391
20.1.12	UPDATE 语句	395
20.1.13	VALIDATE 语句	395
20.2	SQL 过程应用实例	396
20.2.1	统计输出功能	396
20.2.2	数据表的合并	399
第 21 章	SAS 矩阵运算语言 (IML)	404
21.1	IML 中矩阵运算的基础知识	404
21.1.1	IML 内容概述	404
21.1.2	矩阵的性质	404
21.1.3	矩阵的命名	405
21.1.4	矩阵的元素	405
21.1.5	矩阵的创建	406
21.1.6	矩阵的运算	411
21.1.7	矩阵下标的运用	414
21.1.8	矩阵的标题	418
21.1.9	矩阵中缺失值的处理	419
21.2	IML 程序设计语言	420
21.2.1	IF-THEN/ELSE 语句	420
21.2.2	模块的定义与实施	421
21.2.3	DO...END 语句	422
21.2.4	跳转语句	423
21.2.5	停止语句	423
21.3	IML 中矩阵与 SAS 数据集的 相互转换	425
21.3.1	打开 SAS 数据集	425
21.3.2	显示数据集的属性信息	426
21.3.3	创建永久数据集	426

21.3.4	读取并转换 SAS 数据集	427
21.3.5	将矩阵转换为数据集	429
21.4	在 IML 环境中对 SAS 数据集 的操作	430
21.4.1	在 IML 中对 SAS 数据集排序	430
21.4.2	在 IML 中对 SAS 数据集的 统计运算	431
21.4.3	小结	433
21.5	IML 矩阵运算与 DATA 步运算 的联系与区别	434
第 22 章	SAS 的宏功能	437
22.1	宏变量	437
22.1.1	简介	437
22.1.2	SAS 系统定义的宏变量	438
22.1.3	用户定义的宏变量	439
22.2	宏	446
22.2.1	简介	446
22.2.2	宏语言中的元素	446
22.2.3	包含 SAS 程序的宏	449
22.2.4	宏参数	450
22.2.5	宏表达式	451
22.3	宏应用实例	454
22.3.1	用宏变量替代文本串	454
22.3.2	用宏产生 SAS 代码	454
22.3.3	用宏参数在宏中传递信息	455
22.3.4	根据条件产生 SAS 代码	455
22.3.5	用循环语句 %DO 产生重复的 文本	456
22.3.6	为宏变量引用产生后缀	456
22.4	如何提高宏的运行效率	456
22.4.1	根据需要使用宏	457
22.4.2	使用宏名调用	457
22.4.3	避免使用嵌套的宏定义	457
22.4.4	将宏函数的值赋给一个宏变量	458
第 23 章	SAS 输出传送系统	460
23.1	ODS 概述	460
23.1.1	简介	460

23.1.2 示例	460	24.1.3 从 SAS 发送命令控制 EXCEL 的运行	496
23.1.3 ODS 基本概念和功能	463	24.1.4 从 SAS 到 WORD 简介	497
23.1.4 ODS 语句	465	24.2 使用 DDE 自动产生统计报告	499
23.2 常用 ODS 输出目标介绍	466	24.2.1 统计报表的自动化生成	499
23.2.1 SAS 格式目标	466	24.2.2 报表自动化中 DDE 技术和 ODS 技术的联合应用	502
23.2.2 第三方格式目标	484	24.2.3 统计图的自动绘制	503
第 24 章 SAS/DDE 技术的应用	493	附录 1 SAS 常用选项设置一览	505
24.1 DDE 入门	493	附录 2 SAS 常用函数介绍	508
24.1.1 将 SAS 数据集写入 EXCEL	494	附录 3 SAS 常用模块一览	513
24.1.2 读取 EXCEL 文件到 SAS 数据集	495	参考文献	517

第 1 章 SAS 8.2 使用入门

1.1 SAS 系统简介

1.1.1 概况

SAS 系统全称为 Statistics Analysis System, 最早由北卡罗来纳大学的两位生物统计学研究生编制, 并于 1976 年成立了 SAS 软件研究所, 正式推出了 SAS 软件。SAS 是用于决策支持的大型集成信息系统, 但该软件系统最早的功能限于统计分析, 至今, 统计分析功能也仍是它的重要组成部分和核心功能, 同时系统的功能也逐渐扩展到线性与非线性规划、时间序列分析、运筹决策支持、数据仓库和数据挖掘等领域。

经过多年的发展, SAS 如今已被全世界 120 多个国家和地区的近 3 万家机构所采用, 直接用户则超过 300 万人, 遍及金融、医药卫生、生产、运输、通讯、政府和教育科研等领域。在数据处理和统计分析领域, SAS 系统被誉为国际上的标准软件系统, 堪称统计软件界的巨无霸。在英美等国, 能熟练使用 SAS 进行统计分析是许多公司和科研机构选才的条件之一。

SAS 现在的最新版本为 9.0 版, 根据不同的安装方式, 所占硬盘空间大约为 1~2G。国内用户使用的主要版本为 6.12 版和 8.2 版, 本书就是以 8.2 版为准讲述 SAS 系统的操作, 但相应的绝大部分功能的实现在个版本中都是相同的。

1.1.2 SAS 的模块化结构

SAS 系统是一个组合软件系统, 连同正在开发中的模块, 它一共由 50 个左右功能模块组合而成。其基本部分是 SAS/BASE 模块, 该模块是 SAS 系统的核心, 承担着主要的数据管理任务, 并管理 SAS 的用户使用环境, 进行用户语言的处理, 调用其他 SAS 模块和产品。在 SAS/BASE 的基础上, 用户还可以增加各种模块而增加不同的功能, 如 SAS/STAT (统计分析模块)、SAS/GRAPH (绘图模块)、SAS/QC (质量控制模块)、SAS/ETS (经济计量学和时间序列分析模块)、SAS/OR (运筹学模块)、SAS/IML (交互式矩阵程序设计语言模块)、SAS/FSP (快速数据处理的交互式菜单系统模块)、SAS/AF (交互式全屏幕软件应用系统模块) 等等, 表 1.1 列出的是医学统计中常用的几个模块。更详细的模块介绍请参见本书附录。

表 1.1 医学统计中常用的几个 SAS 模块

模块名称	功能简介
Base	是 SAS 系统的核心, 它负责数据管理, 交互应用环境管理, 进行用户语言处理, 调用其他 SAS 产品。Base SAS 对 SAS 系统的数据库提供了丰富的数据管理功能, 并具有强化了 Web 功能, 但不支持中文变量名
SAS/STAT	覆盖了所有的实用数理统计分析方法, 是国际统计分析领域的标准软件

(续表)

模块名称	功能简介
SAS/ADX	专用的试验设计模块，可进行成组、配伍、拉丁方、正交等各种复杂的试验设计，并计算样本量、估计检验效能等
SAS/GRAPH	绘图模块，可将数据及其包含着的深层信息以多种图形生动地呈现出来。SAS/GRAPH 提供一个全屏幕编辑器；提供多种设备驱动程序，支持多种格式的图形交换文件
SAS/PH-Clinical	结合美国食品和药物管理局（FDA）标准，为制药、生物工艺学提供的专用临床数据管理系统

1.1.3 SAS 的操作方式

SAS 最初是由大型机系统发展而来，其核心操作方式就是程序驱动。经过多年的发展，现在已成为一套完整的第四代计算机语言，使用程序方式，用户可以完成所有工作，包括统计分析、预测、建模和模拟抽样等。此外，SAS 还提供了各类概率分析函数、分位数函数、样本统计函数和随机数生成函数，使用户能方便地实现特殊统计要求。但是，这使得初学者在使用 SAS 时必须学习 SAS 语言，入门比较困难。

SAS 的用户界面也充分体现了以程序方式为核心的这一特点：它采用 MDI(多文档界面)，用户在 PGM 视窗中输入程序，分析结果以文本形式在 OUTPUT 视窗中输出。在使用程序方式的同时，SAS 的 Windows 版本也根据不同的用户群开发了几种图形操作界面，见表 1.2)。这些图形操作界面各有特点，使用时非常方便。但是由于国内介绍它们的文献不多，并且也不是 SAS 推广的重点，因此还不为绝大多数人所了解。

表 1.2 SAS 的几种常用图形操作界面

界面名称	特 点
分析员 (Analyst)	类似资源管理器的图形操作界面，和 SPSS 的操作方式较为类似，同时可用于帮助学习 SAS 语句
SAS/INSIGHT	可视化的数据探索工具。采用交互式图形操作界面，将统计方法与交互式图形显示融合在一起，为用户提供一种全新的进行统计分析的环境。适用于对数据进行探索性分析，在数据诊断和曲线拟合上非常方便
SAS/ASSIST	全面、综合、面向任务的图形操作界面，包括了 SAS 在数据管理、统计分析、绘图等方面的常用功能
SAS/LAB	半自动化的图形操作界面，包括了常用的统计分析和绘图功能，自动生成常用的各种分析结果，可帮助初学者快速掌握如何使用 SAS 系统

1.1.4 SAS 的功能与运行速度

SAS 在数据管理上的功能非常强大，其 Base 模块本身就提供了快速而全面的数据库操作命令，并内嵌了 SQL 语言，可以满足绝大多数用户的需要。同时，针对大量多维数据的管理

和挖掘, SAS 还提供了 SAS/Warehouse Administrator、SAS/Enterprise Miner 和 SAS/MDDDB Server 等模块, 可全方位地满足各种用户的需要。

作为世界上最权威的统计软件, SAS 的统计功能极为强大, 在纳入统计新方法上也较为积极。近年来出现的一些新方法, 如非参数回归模型、左截断 Cox 模型、多水平统计模型等均已纳入该软件, 此外, 用户还可以通过对数据集的变换, 以及利用各类分析函数实现更为复杂的统计建模和分析。在 SAS 公司的产品线中, 几乎所有的新产品都是作为 SAS 系统的一个模块而出现的, 这也使得 SAS 成为了一个功能全面而强大的分析平台。

SAS 在设计上就考虑了分析大型数据集的问题, 对硬件资源的利用非常充分, 同时还针对复杂运算作了专门的算法优化, 因此它的运算速度非常快, 常用的统计分析功能一般都可以在数秒内完成, 即使在配置较低的机器上运行速度也较快。由于目前计算机硬件的发展极为迅速, 大部分微机的配置均在 PIII 以上, 完全可以满足 SAS 系统的正常运行需求。因此, 本书不再列出 SAS 相应的硬件配置要求。

1.2 界面操作入门

1.2.1 SAS 的启动与退出

1. SAS 的启动

在 Windows 的桌面操作环境下, 若已正确安装了 SAS 8.2, 则安装程序会在系统开始菜单的列表中添加一个 “The SAS System” 项目组, 这时可按照下列步骤启动 SAS 系统:

开始按钮 → 所有程序 → The SAS System → The SAS System for Windows V8

以上的操作表示用鼠标单击桌面左下方的 “开始” 按钮, 并在弹出菜单中选择 “所有程序”, 最终单击相应的菜单项, 数秒后 SAS 系统就会启动完毕, 出现主界面 (见图 1.1)。

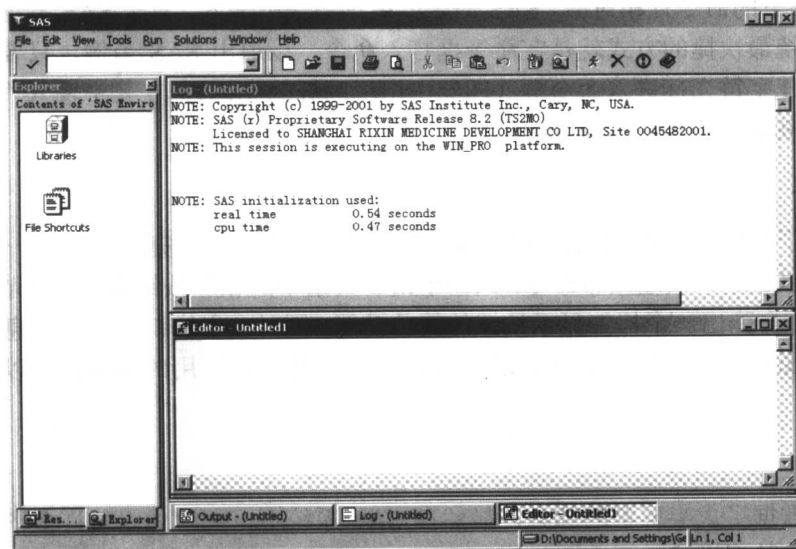


图 1.1 SAS 的主界面

除使用开始菜单启动外,对 Windows 系统熟悉的朋友也可以建立桌面快捷方式,或使用开始菜单中的“运行”,或者在资源管理器中直接找到 SAS.EXE 双击启动 SAS 系统。

2. SAS 的退出

SAS 的退出方式和绝大多数 Windows 软件相同,只需要选择菜单 File→Exit,或者直接单击主界面右上角的控制按钮“”,SAS 系统就会弹出确认框,对退出操作加以确认,此时单击确定按钮,就可以彻底退出 SAS 系统。

1.2.2 SAS 的主窗口

启动 SAS 8.2 后即展示如下的 SAS 主窗口,它是交互方式下运行 SAS 的工作界面,也称为应用工作空间(AWS: Application Work Space),又被称为 SAS 的显示管理界面(DMS: Display Management System)。现对其主要部分分别描述如下。

1. 菜单条

菜单条又叫下拉菜单,是 Windows 程序的重要组成部分,利用它我们可以完成各种功能,从而轻松地驾驭 SAS 系统。以“Edit”菜单为例,用鼠标点一下 Edit 项,该菜单项就会弹出其相应的下拉菜单,请注意第三项“Cut”为灰色,表示该命令现在不可用。有的菜单项右侧有一个小黑三角(如 Solution 菜单内的许多菜单项),这表示该菜单项有自己的次级菜单,将鼠标移到该处,则次级下拉菜单自动弹出,如此反复,直至找到所需命令为止,然后将鼠标指向项目所需菜单项,单击左键,则该命令开始运行。

2. 命令框

命令框位于菜单栏的左下方,和工具栏并排,形如“”。可以将它和 DOS 状态下的命令行相对应。SAS 命令就由该命令框输入。对于 SAS 系统的熟练使用者来说,使用系统最快捷的方式就是在命令框打入命令。以退出 SAS 为例,只要在命令框中单击一下鼠标,使字符光标进入命令框(即其中出现闪烁的光标),在其中输入“bye”,然后回车,SAS 系统就和我们彻底 Bye-bye 了。

3. 工具栏

工具栏位于菜单栏下方。在工具栏中排列了若干个按钮,每个按钮相当于一个常用的命令,使用工具栏可以让用户快速发布命令,只需单击工具条中的相应按钮即可。当光标移至各个按钮时,按钮边上会弹出一个标签显示按钮的简单功能,同时,在主窗口下方的信息条上显示该按钮功能的详细说明。工具栏所包含的按钮个数、它的功能以及采用的图标都可以按用户的需要自行设定。

4. 窗体

窗体就是占据了大半屏幕的部分,显然在 SAS 系统中,主窗体内被许多子窗体所填充,对这些子窗体将随后详述。

5. 窗口栏

窗口栏位于 SAS 操作界面中窗体的下方。由于 SAS 是多文档界面，为了方便各子窗口间的切换，SAS 提供了和 windows 的任务栏功能十分相似的窗口栏，它包含了所有打开的子窗口的标签。对非激活的窗口，单击该窗口的标签就使它激活并置于前台；对激活的窗口，点击这个窗口就可以使它最小化。从窗口栏可以看到，SAS 系统在启动后默认会打开 5 个子窗口，并且默认的当前子窗口为增强型程序编辑器（Enhanced Editor）窗口。

6. 状态栏

状态栏位于窗口栏下方，又可以被分为左侧的消息栏和右侧的当前文件夹栏两部分，对于系统在运行中发生的任何情况，在消息栏上都会显示出相应的信息，它可以帮助我们及时了解系统所处的状态。而当前文件夹栏则表明系统当前的所在位置。我们所做的读、写文件等操作均默认在该文件夹中进行，双击该文件夹栏会弹出对话框，供用户更改当前文件夹。

1.2.3 SAS 的子窗口

SAS 8.2 在启动时，默认会打开以下五个窗口：增强型程序编辑器（Enhanced Editor）窗口、日志（Log）窗口、结果输出（Output）窗口、资源管理器（Explorer）窗口、结果（Results）窗口。其中前三个较为常用，后两个相当于索引或者目录树窗口，起辅助的作用，所以这两个窗口缺省地固定在工作空间的左侧(Docked)。SAS 子窗口的排列方式也可以自行设置。在 SAS 系统进行交互操作的过程中，任何时候都只有一个子窗口是被激活的，或称它为当前窗口，激活窗口上部的标题栏为蓝色，非激活窗口则为灰色。用鼠标点击某个子窗口就可激活该窗口。由于不同的窗口有不同的功能，系统为不同窗口提供了不同的命令，所以对于不同的子窗口，菜单项和快捷工具栏有可能是完全不同的，这一点在使用时需要注意。

下面依次介绍主要的子窗口功能。

1. 增强型程序编辑器（Enhanced Editor）窗口

在 Windows 操作环境下，增强型程序编辑器是 SAS 8.2 系统默认提供的程序编辑窗口（见图 1.2）。除了提供一般的文本输入和编辑功能（如复制、粘贴、剪切等）外，它还提供了许多有用的编辑功能，包括：

- 以不同的颜色显示出 SAS 程序中不同的部分，并同时进行语法检查，如用深蓝色表示数据步/程序步开始；蓝色表示关键字；棕色表示字符串，浅黄底色表示数据块，红色表示可能的错误等。这样可以大大方便程序的阅读和查错。
- 程序段的展开和收缩功能，单击每个程序段左侧的加号或者减号即可实现，便于从宏观上了解程序的结构。
- 程序输入时自动支持缩进格式。
- 支持将输入的字符串自定义为宏。
- 支持键盘快捷方式(将 Alt 键或 Shift 键与其他键同时使用)。
- 增强型程序编辑器窗口中的内容在保存时应当被存为 SAS 程序格式（SAS files），实际上就是扩展名为“.sas”的纯文本文件。虽然也可以存储为其他扩展名，但可能会引起混乱，不推荐这样做。