

Statistical  
Analysis  
System

**SAS**

# 统计分析及应用

黄燕 吴平 等编著

为教师配有电子教案



机械工业出版社  
China Machine Press

# Statistical Analysis System

# SAS

## 统计分析及应用

黄燕 吴平等编著



机械工业出版社  
China Machine Press

本书基于 SAS (Statistical Analysis System) 系统软件 8.02 版写成。全书以统计分析方法为主线,从实际应用角度出发,通过大量真实的试验数据,详细介绍了 SAS 系统软件中的各种常用统计分析方法的用法。每章都通过大量实例详细讲解 SAS 软件中相应的统计分析过程的使用方法,并对每个实例的统计输出结果做了详细的解释,尤其是给出了统计分析思路、具体应用方法、相关统计量的判断准则和选择条件,并指出了下一步的分析思路与解决方案。

本书既可作为高等院校研究生、本科生的统计学教材,也可作为从事数据分析与数据管理的研究人员的参考用书。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

### 图书在版编目(CIP)数据

SAS 统计分析及应用/黄燕等编著. -北京:机械工业出版社,2006.1  
ISBN 7-111-17566-2

I .S… II .黄… III .统计分析-应用软件,SAS IV .C812

中国版本图书馆 CIP 数据核字(2005)第 119021 号

机械工业出版社(北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑:李云静

北京慧美印刷有限公司印刷·新华书店北京发行所发行

2006 年 1 月第 1 版第 1 次印刷

787mm×1092mm 1/16·23 印张

印数:0 001-4 000 册

定价:33.00 元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

本社购书热线:(010) 68326294

# 前 言

SAS 系统是世界公认的权威性统计分析软件之一, 是一个大型集成信息分析管理系统。SAS 系统集数据访问、数据管理、数据分析、数据挖掘功能于一体, 是一个功能强大且被国际公认的国际标准统计分析软件。随着信息技术的发展, 数据管理与数据分析已成为各领域发展的重要支柱, SAS 系统也成为数据处理的首选工具。在中国, 越来越多的使用者感受到了 SAS 系统强大功能所带来的益处, SAS 的使用者逐年增多, 它已成为研究人员、学者发表高水平学术论文的重要工具。

近 10 年来, 笔者在农业院校为研究生讲授此课, 并为广大科研人员做了大量的统计咨询工作, 从中积累了许多经验, 也受到了很多启迪。众多非数理专业的研究生及研究人员缺乏系统的多元统计分析知识, 仅有常规的概率统计知识, 而从事的研究项目又主要是定性及定量分析相结合的问题。要对试验数据进行科学的分析, 就需要自学相关的统计知识。然而, 传统的多元统计分析教程主要从统计分析公式推导、证明的角度进行讲述, 虽然很多研究人员试图学好统计学知识, 但由于种种原因仍不能在短期内了解并掌握自己所需要使用的统计学知识, 要达到正确使用或灵活应用的目标更是有一定难度。而 SAS 系统软件已将统计分析算法编成了优秀的软件, 使用者不需记住各种统计分析公式的推导、证明过程, 完全可以将繁杂、深奥的统计计算简化为对各统计量的理论意义及实际作用的理解, 搞清相关统计量的来源及概念就可以轻松地掌握统计分析方法、处理科研试验数据, 从而得到科学的研究结论。

根据长期的教学经验, 针对非数理专业研究人员的统计知识水平, 我们从应用角度出发, 依据具体研究问题和真实的试验数据讲解各种统计方法, 编写了本书。书中通过对大量应用实例的详细剖析, 将复杂、难于理解的统计计算、统计分析方法、统计量的理论意义与实际作用进行了详细解释, 内容通俗易懂、深入浅出。尤其针对一些进行统计分析时无从下手的人员, 书中重点讲解了统计分析思路、统计量的判断准则以及相应的解决方案。这些对缺乏系统的统计知识而又从事数据分析的研究人员提供了良好的帮助。

全书共 11 章, 第 1 章是 SAS 系统软件的基础入门篇; 第 2 章介绍 SAS 系统的数据步 (DATA STEP, 也称 DATA 步) 及常用的一些相关语句; 第 3 章介绍 SAS 过程步的一些基本功能, 主要介绍常用的统计分析过程 (包括 SORT 过程、MEANS 过程、UNIVARIATE 过程、FREQ 过程以及 PRINT 过程等) 的用途及常用统计概念; 第 4 章介绍  $t$  检验和非参数检验的基本概念和检验方法; 第 5 章介绍方差分析的基本概念、原理, 以及 ANOVA 和 GLM 过程的使用方法; 第 6 章讲解相关分析与回归分析, 重点介绍 CORR 过程、PLOT 过程、REG 过程、RSREG 过程、NLIN 过程的使用方法; 第 7 章介绍聚类分析的基本概念, 讲解了 VARCLUS 过程、CLUSTER 过程、FASTCLUS 过程、ACECLUS 过程以及 TREE 过程的使用方法; 第 8 章介绍主成分分析的基本概念和数据结构, 讲解了 PRINCOMP 过程的使用方法; 第 9 章介绍典型相关分析方法的基本概念, 介绍了 CANCORR 过程的使用方法; 第 10 章讲解判别分析的基本概念, 介绍了 DISCRIM 过程、CANDISC 过程的使用方法; 第 11 章介绍因子分析的基本概念和数学模型, 讲解了 FACTOR 过程的使用方法。

本书的第 1、2、3、7、8 章由黄燕、吴平、汪安共同编写, 第 4、5、6 章由黄燕、张日俊编写, 第 9、10、11 章及各章习题由黄燕编写、整理。

本书的内容由浅入深, 可适用对统计知识要求不同层次的学科专业, 根据课程学时数及专业要求不同, 在教学中可适当进行取舍。本书可用作研究生、本科生、大专生的统计学教材, 也可作为从事数据分析和数据研究的高等院校和科研机构的研究人员的参考书。

本书得以顺利出版, 与我校研究生院领导的热情关心、帮助以及机械工业出版社华章分社的支持是分不开的, 在此一并表示衷心的感谢!

由于编者水平有限, 书中难免出现差错, 如有差错, 望读者批评、指正。

黄 燕

2005 年 8 月 25 日于北京

# 目 录

前言	
第 1 章 SAS 系统入门	1
1.1 系统概述	1
1.2 SAS 系统界面操作指南	1
1.2.1 SAS 的启动与退出	1
1.2.2 SAS 系统界面简介	2
1.2.3 简单运行样例	4
1.3 SAS 系统的文件管理	7
1.3.1 SAS 文件与逻辑库	7
1.3.2 数据集文件的基本操作	9
1.4 数据集的导出/导入	12
1.4.1 导出 SAS 数据集	12
1.4.2 外部数据导入 SAS 数据集	13
1.5 数据集的编辑与整理	15
1.5.1 用 Viewtable 表编辑数据集	15
1.5.2 数据集变量名与类型的更改	17
1.5.3 数据集变量、观测记录的增加 与筛选	18
1.5.4 数据集排序	19
1.5.5 数据集的纵向连接	20
1.5.6 数据集的横向合并	21
1.5.7 数据合并的菜单操作	21
1.6 习题	24
第 2 章 SAS 编程简介	25
2.1 SAS 程序设计语言基本常识	25
2.1.1 SAS 程序书写规范和运行方法	25
2.1.2 SAS 程序中的注释	26
2.2 SAS DATA 步简介	26
2.2.1 DATA 步基本结构	26
2.2.2 用 DATA 步生成 SAS 数据集	27
2.2.3 DATA 步中的常用语句	30
2.3 综合举例	39
2.4 习题	40
第 3 章 SAS 过程步及简单统计分析	41
3.1 SAS 过程步简介	41
3.2 PRINT 过程	43
3.2.1 PRINT 过程简介	43
3.2.2 变量标签语句 LABEL	44
3.2.3 全程语句	45
3.2.4 用 TABULATE 过程制作统 计量表格	46
3.3 简单描述统计分析	49
3.3.1 简单描述统计量的基本概念	49
3.3.2 MEANS 过程	50
3.3.3 UNIVARIATE 过程	51
3.3.4 FREQ 过程	52
3.3.5 应用实例	53
3.4 菜单操作进行简单统计分析	58
3.4.1 利用 SAS/INSIGHT 进行简单 分析	58
3.4.2 利用 SAS/ASSIST 进行简单 分析	63
3.5 习题	64
第 4 章 $t$ 检验与非参数检验	67
4.1 常用统计概念简介	67
4.1.1 常用统计概念	67
4.1.2 概率与概率分布	67
4.1.3 二项分布	68
4.1.4 泊松分布	68
4.1.5 正态分布	69
4.1.6 极限定理	69
4.1.7 样本平均数的分布	70
4.1.8 $t$ 分布	71
4.1.9 $\chi^2$ 分布	71
4.1.10 $F$ 分布	72
4.2 统计检验的基本原理与方法	73
4.2.1 假设检验的基本概念	73
4.2.2 样本平均数的检验—— $u$ 检验与 $t$ 检验	75
4.2.3 双尾检验与单尾检验	79
4.2.4 非参数检验	79
4.3 TTEST 过程——比较 $t$ 检验	80
4.3.1 TTEST 过程简介	80
4.3.2 TTEST 过程举例说明	81
4.4 NPARIWAY 过程	83
4.4.1 NPARIWAY 过程简介	84
4.4.2 NPARIWAY 过程举例说明	85
4.5 综合应用	88
4.5.1 配对数据的统计分析	88
4.5.2 成组试验数据的统计分析	93

4.5.3	成组试验数据经对数转换后的 $t$ 检验	96	举例	192	
4.5.4	SAS/ASSIST 中的 $t$ 检验与非参数检验	98	6.2	相关过程与回归过程	197
4.6	习题	99	6.2.1	CORR 相关过程	197
第 5 章	方差分析	103	6.2.2	PLOT 过程	199
5.1	方差分析简介	103	6.2.3	REG 回归过程	202
5.1.1	方差分析基本概念	103	6.2.4	回归诊断方法	208
5.1.2	总体方差检验	104	6.2.5	用各种回归方法筛选变量	209
5.1.3	方差的同质性检验	104	6.3	直线回归分析	212
5.1.4	方差分析的基本假定和数据转换	105	6.3.1	概述	212
5.2	ANOVA 过程和 GLM 过程简介	106	6.3.2	可直线化的简单曲线拟合	213
5.2.1	ANOVA 过程	106	6.3.3	应用举例	213
5.2.2	GLM 过程	109	6.3.4	一般多项式曲线拟合	227
5.3	单因素多水平试验设计及方差分析	113	6.4	非线性回归过程 NLIN	230
5.3.1	完全随机设计	113	6.4.1	NLIN 过程简介	230
5.3.2	一元方差分析应用举例	114	6.4.2	NLIN 应用举例	231
5.3.3	多元方差分析应用举例	118	6.4.3	多条回归直线的方差比较	242
5.3.4	完全随机区组设计	123	6.5	多元线性回归分析	243
5.3.5	单因素多水平间的多重比较	125	6.5.1	多元线性回归分析的概述	243
5.4	配伍组设计及其统计分析	129	6.5.2	应用举例	244
5.4.1	一元应用实例	129	6.6	RSREG 过程	251
5.4.2	多因素多水平间的多重比较	132	6.6.1	RSREG 过程说明	251
5.5	拉丁方设计及其统计分析	135	6.6.2	RSREG 过程语句格式	252
5.5.1	拉丁方设计及其线性模型	135	6.6.3	应用举例	254
5.5.2	拉丁方设计应用实例	136	6.7	习题	259
5.6	析因设计及其统计分析	141	第 7 章	聚类分析	265
5.6.1	CONTRAST 语句说明	141	7.1	聚类分析方法概述	265
5.6.2	应用实例	143	7.2	TREE 过程	266
5.7	正交设计及其统计分析	164	7.3	应用举例	266
5.7.1	正交设计简介	164	7.3.1	用 VARCLUS 过程实现变量聚类分析	266
5.7.2	正交设计应用举例	164	7.3.2	用 CLUSTER 过程实现样品聚类分析	275
5.8	协方差分析	170	7.3.3	用 FASTCLUS 过程进行大样本聚类分析	280
5.8.1	协方差分析简介	170	7.3.4	用 ACECLUS 过程进行预处理	285
5.8.2	协方差分析的数学模型及基本假定	171	7.4	习题	288
5.8.3	协方差分析应用举例	171	第 8 章	主成分分析	293
5.9	习题	185	8.1	主成分的几何意义与一般数学模型	293
第 6 章	相关与回归分析	189	8.2	PRINCOMP 过程	296
6.1	相关与回归分析概述	189	8.3	应用举例	297
6.1.1	相关与回归分析的任务和种类	189	8.4	习题	303
6.1.2	直线回归与相关分析	190	第 9 章	典型相关分析	305
6.1.3	直线相关与回归分析的应用		9.1	典型相关分析概述	305

9.2	CANCORR 过程简介 .....	306	功能 .....	327
9.2.1	CANCORR 语法格式 .....	306	10.3.2 应用举例 .....	328
9.2.2	CANCORR 语句说明 .....	306	10.4 习题 .....	336
9.3	应用举例 .....	308	第 11 章 因子分析 .....	341
9.4	习题 .....	313	11.1 因子分析简介 .....	341
第 10 章	判别分析 .....	315	11.1.1 因子分析基本概念 .....	341
10.1	判别分析概述 .....	315	11.1.2 因子分析的数学模型 .....	342
10.2	DISCRIM 判别分析过程 .....	318	11.2 PROC FACTOR 过程 .....	342
10.2.1	DISCRIM 过程语句说明 .....	318	11.3 应用举例 .....	345
10.2.2	应用举例 .....	320	11.4 习题 .....	359
10.3	CANDISC 典型判别过程 .....	327	参考文献 .....	362
10.3.1	CANDISC 过程的语句用法和			

# 第 1 章 SAS 系统入门

## 1.1 系统概述

SAS(Statistical Analysis System)系统于 1966 年开始研制, 1976 年由美国 SAS 公司实现商业化, 1985 年推出 SAS/PC 版本, 1987 年推出 6.03 版, 目前已推出 Windows 系统支持的 8.02 版和 9.0 版, 是目前国际上公认的著名的数据统计分析软件系统之一。

SAS 系统的最大特点是将数据管理和统计分析融为一体。Windows 版的 SAS 系统软件操作简单, 它提供一种交互方式运行 SAS 程序。SAS 系统的基础是 SAS 语言, SAS 语言是一种近似自然语言的非过程语言(第四代语言), 用户不用告诉 SAS“怎样做”, 只需告诉它你想要它“做什么”。目前 SAS 已经可在不同的操作系统(如 MS-DOS、UNIX、VMS 等)和不同的机器类型下运行。该系统应用范围广, 包括统计学、心理学、经济学、生物学、商业、金融、医药卫生、图形显示、数据输入、数据检索、报表生成、统计计算、市场研究、销售管理、调查分析、建立预测模型等方面。

SAS 系统是集成数据分析、生成报表于一体的一种集成软件系统。它由基本部分和外加模块组成, 其中主要部分包含的功能如下:

- 基本部分——BASE SAS 部分。
- 统计分析计算部分——SAS/STAT。
- 绘图部分——SAS/GRAPH。
- 矩阵运算部分——SAS/IML。
- 运筹学和线性规划——SAS/OR。
- 经济预测和时间序列分析——SAS/ETS。

BASE SAS 是 SAS 系统的核心平台; SAS/STAT 是 SAS 系统的精华, 是用来解决实际问题的主要部分。

随着数据挖掘技术的发展和应用, SAS 推出了 Enterprise Mine 模块, 其具有强大的数据挖掘功能。由于这部分内容很多, 通常单做介绍, 因此本书对此不做介绍。本书将以 V8.02 版本为主线, 重点介绍 BASE SAS 和 SAS/STAT 这两部分中常用的统计分析模块。

BASE SAS 提供 31 个初等 SAS 过程, 按照用途归为 4 类:

- 1) 简单统计计算——means 过程、univariate 过程、freq 过程等。
- 2) 报表生成——table 过程、chart 过程、plot 过程等。
- 3) 得分方法——standard 过程、rank 过程等。
- 4) 实用方法——sort 过程、format 过程、dbf 过程、dif 过程等。

SAS/STAT 部分包括: 回归分析、方差分析、属性数据分析、多变量分析、判别分析、聚类分析、得分方法和残存分析共 8 类 26 个过程。

SAS/GRAPH 绘图软件具有强大的绘图能力, 能完成多种绘图功能, 如等值线图、二维曲线图、条型图、饼图、区块图等多种功能。

## 1.2 SAS 系统界面操作指南

### 1.2.1 SAS 的启动与退出

SAS V8.02 安装完成以后, 在系统“开始”菜单列表中添加了一个“The SAS System”项目组。点击“The SAS System for Windows V8”, 可以启动 SAS 系统。启动操作如图 1-1 所示。除使用“开始”菜单启动外, 也可以在桌面上建立快捷方式, 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 \ V8 \ SAS.EXE 文件并双击以启动 SAS 系统。



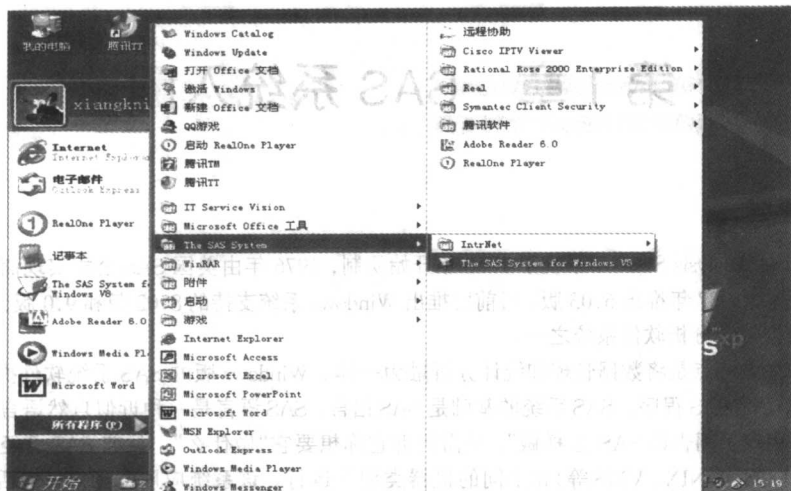



图 1-1 启动 SAS 系统

退出 SAS 系统时，只需要选择菜单 File|Exit，或者直接点击主界面右上角的控制按钮 ，SAS 系统就会弹出确认框，点击确定按钮可以退出 SAS 系统。

### 1.2.2 SAS 系统界面简介

SAS 系统简单易用，为用户所要完成的基本 SAS 任务都提供了窗口。一旦熟悉了 SAS 任务的起点，就可以迅速地完成任务。

SAS 系统启动后，见图 1-2 所示的 SAS 工作界面，称为“SAS 工作空间(SAS Application WorkSpace)”。它像其他 Windows 应用程序一样，可以在一个主窗口内，包含若干个子窗口，并有菜单栏、工具栏、状态栏等。

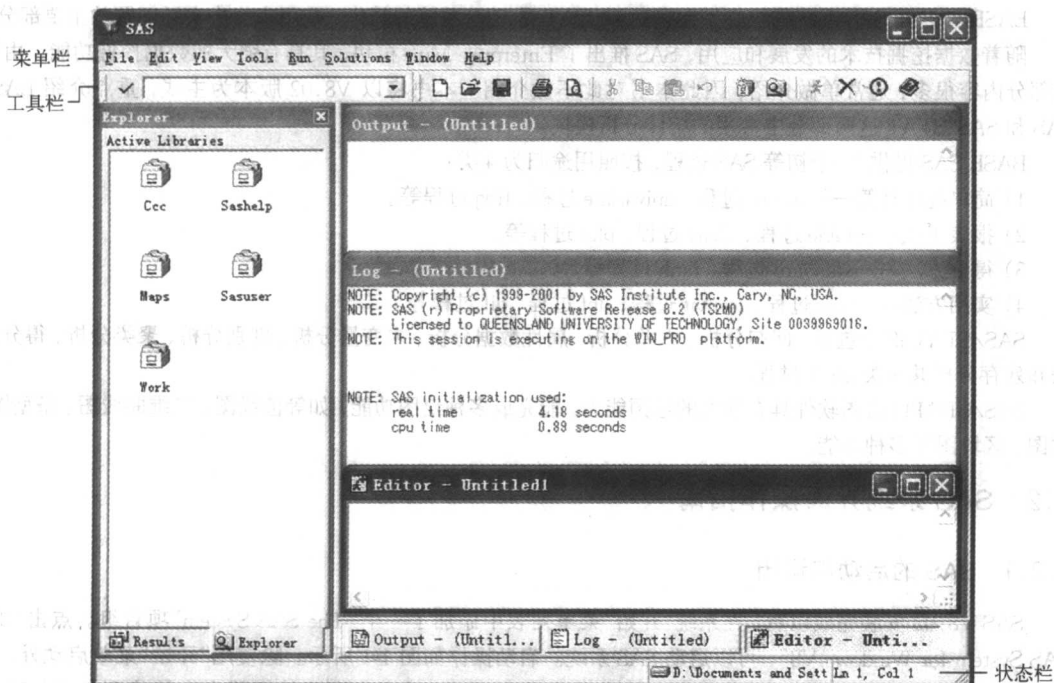


图 1-2 SAS 工作界面

SAS 三个最重要的窗口为：Editor 程序编辑窗口，Log 运行记录窗口，Output 输出记录窗口(参考图 1-2)。

Editor 程序编辑窗口用于编辑 SAS 源程序文件，编辑窗口支持全窗口编辑功能，即光标可以在整个窗口内随意移动，并支持 Windows 系统常规编辑操作的方式，如复制、剪切、粘贴等功能操作。SAS V8 提供了智能编辑功能，它可根据用户输入的 SAS 关键字、语句选项、变量名、数据、标记等不同内容显示不同的颜色。当用户输入的字串不正确时，对应的颜色不对，警告使用者有错误发生。

Log 运行记录窗口用于显示和记录 SAS 程序的运行情况，说明程序运行成功或存在错误。当程序出现错误时，Log 记录窗口中以红色字符显示错误信息，以绿色显示一些警告信息，提示使用者修改程序中的错误。

Output 输出记录窗口分页显示 SAS 程序运行所产生的文本型输出结果(图形输出通过 Graphics 窗口显示)。对于文本结果的输出，可以使用主界面的菜单 File|Save As 将结果文档保存在磁盘中，保存的文件扩展名为 .lst，但实际上是一种文本文件，可以使用文字处理软件，如 Word 或写字板、记事本等打开并进行编辑修改。对于 Graphics 窗口的图形，可以使用菜单 File|Export as Image 将图形导出保存在磁盘中，图形文件格式可以在“保存类型”下拉列表中选择。

使用 SAS 主菜单命令或功能键，可以根据需要在不同的窗口之间进行切换，利用功能键 F5 可以切换(激活)到 Editor 程序编辑窗口，利用 F6 可以进入 Log 运行记录窗口，利用 F7 可以激活 Output 输出记录窗口。

SAS V8 版系统在主界面增加了 Explorer 窗口与 Results 窗口。

Explorer 窗口用于显示 SAS 库(SAS 系统命名的库名与磁盘某文件夹的关联)及其 SAS 数据集。

Results 窗口用于显示 SAS 程序运行成功时程序输出结果的目录。通过它可方便地查找到要看的结果。Results 窗口中的输出目录与 Output 窗口的对应关系同资源管理器的左右窗口一样，使用很方便。

SAS 系统还有许多窗口，每一窗口都有其特定功能。根据需要可以关闭窗口，也可以通过主菜单 View 的下拉菜单打开或切换窗口。








在 SAS 系统工作过程中，系统主菜单是动态变化的，随当前激活的窗口不同而有不同的组织结构，提供不同的主菜单命令。









其中的主要命令及功能如下：

- File(文件)——支持 SAS 文件的调入、保存及打印的功能。
- Edit(编辑)——支持编辑窗口的基本编辑操作(例如：清空、复制、剪切、粘贴、查找、替换等)。
- View(浏览)——支持 SAS 系统用户在多窗口之间切换。
- Tools(工具)——提供对各种输出结果进行编辑的工具，如：表格、图形、报告等，并支持进行系统环境与状态的设置。
- Run(运行)——用于执行或调用程序。
- Solutions(解决方案)——支持用户进行统计分析。

主菜单下是一个命令输入栏和图标工具栏。命令栏用于提供系统对 SAS 早期版本使用方式的兼容操作，在这里可以键入 SAS 的显示管理命令。工具栏图标提供了常见任务的快捷操作方式，比如保存、打印、帮助等。

工具栏图标的主要功能解释如下：

-  Submit——提交编辑窗口中的程序运行，或 ASSIST 功能的运行。
-  New——清空编辑窗口。
-  Open——在编辑窗口中打开文件。注意：打开后编辑窗口自动保持当前打开文件与实际存储的物理文件的连接关系。
-  Save——存盘，保存编辑窗口中的内容。注意：如果此时编辑窗口的内容已经与一个物理文件相连接，此功能将覆盖文件的原物理文件的内容。
-  Print——打印当前编辑窗口内容。
-  Print preview——打印预览。
-  Cut——剪切选定的文本。

-  Copy——复制选定的文本。
-  Paste——粘贴。注意这些操作是对 Windows 剪贴板进行的，所以它不仅支持 SAS 编辑窗口内的复制和剪切操作，还可以用来与其他 Windows 应用程序交换文本、数据。
-  Undo——撤销刚才进行的编辑操作。
-  DOS prompt——临时进入 DOS 状态。
-  Browse——打开 WWW 浏览器并进入 SAS 公司的主页 www.sas.com。
-  Directories——进入 Directory(目录)窗口，还可以浏览各 SAS 数据库的内容，还可以浏览数据库中的数据、SAS 目录的内容。
-  SAS/ASSIST——启动 SAS 的菜单驱动模块 SAS/ASSIST。
-  Help——启动帮助系统进入 SAS 的帮助。

### 1.2.3 简单运行样例

例 1.1 学生管理数据库中的数据表如下，要求建立一个 SAS 数据集，并打印输出。

姓 名	出生日期	年 龄	入学日期	学 号	数 学	英 语
王 红	1973/6/2	22	1995/9/10	9510012	89	73
李 明	1974/3/23	21	1995/9/9	9510004	91	68
徐凯歌	1973/11/14	21	1995/9/10	9510034	87	78
吴青云	1974/4/12	21	1995/9/9	9510023	92	84
李清华	1973/10/24	21	1995/9/10	9510024	92	78
张萍江	1973/12/23	21	1995/9/9	9510005	87	78

#### SAS 程序 data1\_1.sas

```
LIBNAME L3 'D:\mysas';
/* 建立一个库名为 L3 的逻辑库，该库指向 D 盘的 mysas 文件夹 */
DATA L3.Stu01 ;          /* 在库 L3 中建立一个永久 SAS 数据集 */
    INPUT NAME $ 1-9 BIRTH Yymmdd10. +1 AGE +1 INDAY Yymmdd10. +1
           SNO MATH ENGLISH ;
CARDS ;
王 红 1973/6/2 22 1995/9/10 9510012 89 73
李 明 1974/3/23 21 1995/9/9 9510004 91 68
徐凯歌 1973/11/14 21 1995/9/10 9510034 87 78
吴青云 1974/4/12 21 1995/9/9 9510023 92 84
李清华 1973/10/24 21 1995/9/10 9510024 92 78
张萍江 1973/12/23 21 1995/9/9 9510005 87 78
;
PROC PRINT;
    FORMAT BIRTH INDAY MmDdYy8. ;
RUN ;
```


以下是建立的 SAS 数据集 Stu01，通过 Print 过程显示输出：

Obs	NAME	BIRTH	AGE	INDAY	SNO	MATH	ENGLISH
1	王 红	06/02/73	22	09/10/95	9510012	89	73
2	李 明	03/23/74	21	09/09/95	9510004	91	68
3	徐凯歌	11/14/73	21	09/10/95	9510034	87	78
4	吴青云	04/12/74	21	09/09/95	9510023	92	84
5	李清华	10/24/73	21	09/10/95	9510024	92	78
6	张萍江	12/23/73	21	09/09/95	9510005	87	78

例 1.2 若学生数学成绩满分为 120, 英语成绩满分为 100, 要求计算学生的平均分数(按百分制)并按降序排名, 可用下述程序完成。

SAS 程序 data1\_2.sas

```
Title '02 级 1 班学生成绩排名';
LIBNAME L3 'D:\mysas'; /* 如果前面已建立了这个逻辑库, 此语句应省略 */
data L3.Stu02;
  input name $1-10 sex $math english;
  avg = math/120 * 100 * 0.5 + english * 0.5;
  cards;
王 红   F 89 73
李 明   M 91 68
徐凯歌  M 87 78
吴青云  F 92 84
李清华  F 92 78
张萍江  M 87 78
;
proc sort data = L3.Stu02;
  by descending avg; /* 按平均成绩降序排序 */
run;
proc print; run;
```

在 Editor 窗口写好上述程序后, 只要用鼠标点击工具栏的 Submit(提交)图标, 或用菜单 Run|Submit 命令即可。程序运行后, Log 窗口中记录如下内容:

```
50 Title '02 级 1 班学生成绩排名';
51 LIBNAME L3 'D:\mysas';
NOTE: Libref L3 was successfully assigned as follows:
      Engine:          V8
      Physical Name: D:\mysas
51 ! /* 如果前面已建立了这个逻辑库, 此语句应省略 */
52 data L3.Stu02;
53   input name $1-10 sex $math english;
54   avg = math/120 * 100 * 0.5 + english * 0.5;
55   cards;
NOTE: The data set L3.Stu02 has 6 observations and 5 variables.
NOTE: DATA statement used:
      real time          0.01 seconds
      cpu time           0.01 seconds
62 ;
63 proc sort data = L3.Stu02;
64   by descending avg; /* 按平均成绩降序排序 */
65 run;
NOTE: There were 6 observations read from the data set L3.Stu02.
NOTE: The data set L3.Stu02 has 6 observations and 5 variables.
NOTE: PROCEDURE SORT used:
      real time          0.00 seconds
      cpu time           0.00 seconds
66 proc print; run;
NOTE: There were 6 observations read from the data set L3.Stu02.
NOTE: PROCEDURE PRINT used:
      real time          0.00 seconds
      cpu time           0.00 seconds
```

其中记录了每段程序的运行情况、所用时间、生成数据保存情况，每行前的行号由系统自动生成。如果有错误，还会用红色指示错误。例如，若程序最后 proc print 语句的分号丢失，记录窗口中将会显示如下错误：

```
190 proc print run;
-----
181
ERROR 181-322: Procedure name misspelled.
```

该错误指出“过程名错”，但实际上是由于语句丢失了分号，导致系统将 print 和 run 作为一个词处理而产生了错误。我们可以在程序窗口用 Run|Recall last Submit 菜单或按 F4 功能键调回程序进行修改。正确运行后在输出窗口(Output)产生如下输出结果：

02 级 1 班学生成绩排名					
Obs	name	sex	math	english	avg
1	吴青云	F	92	84	80.3333
2	李清华	F	92	78	77.3333
3	徐凯歌	M	87	78	75.2500
4	张萍江	M	87	78	75.2500
5	王 红	F	89	73	73.5833
6	李 明	M	91	68	71.9167

从上述实例可以了解 SAS 程序的一些语法规定：

- SAS 程序由**语句**组成。
- 每个语句用西文状态下的分号“;”作为结束符(注：不能使用中文的分号“；”)。
- 程序命令使用的英文单词一般不区分大小写字母(在作为数据的字符串中系统区分大小写字母)。SAS V8 版中提供了智能编辑窗口，SAS 程序中的关键字、语句名、数据都由不同颜色显示，当你的输入错误时颜色显示不正确，这样我们可以方便地查找及改正错误。
- SAS 程序由两类“操作步”构成，一种叫**数据步**(data step)，一种叫**过程步**(proc step)；数据步以 DATA 语句开始，过程步以 PROC 语句开始。
- 每个“操作步”由若干个语句构成，一条语句可以占多个程序行，多条语句也可以写在一个程序行内，在语法上只要保证每个语句以分号作为结束符即可。通常我们应该在程序中保持每个语句单独分行的风格，以使得程序具有更好的可读性。
- 每个 SAS 过程一般以 Run 语句结束。有些过程需要使用 Quit 语句结束过程的运行。

另外，由于目前 SAS 系统对中文输入的处理还不够完善，在需要输入带有中文的程序内容时，大家可以利用其他编辑工具，如 Windows 的记事本(在 Windows 98/2000/XP 中用“开始”菜单中的“程序”|“附件”|“记事本”启动)。我们可以在编辑器中录入程序，接着选中全部录入内容，之后选择“复制”命令(也可以应用 Ctrl+C 快捷键)，进入 SAS 系统 Editor 程序编辑窗口，使用粘贴命令(或用 Edit 菜单的 Paste 或工具栏上的粘贴图标)，在程序编辑窗口保存程序文件。当然，也可以在录入编辑器中直接把编辑好的程序存盘，然后在 SAS 程序窗口用 File|Open 命令打开保存好的程序文件以实现程序的运行。

提示：可以通过 View(视图)(见图 1-3)中的菜单定制 SAS 工作环境的各窗口，也可以通过 Tools 菜单定制 SAS 系统的许多功能项，例如：工具栏、弹出菜单、图标等。选择 Tools(工具)菜单可查看到可用的定制选项。

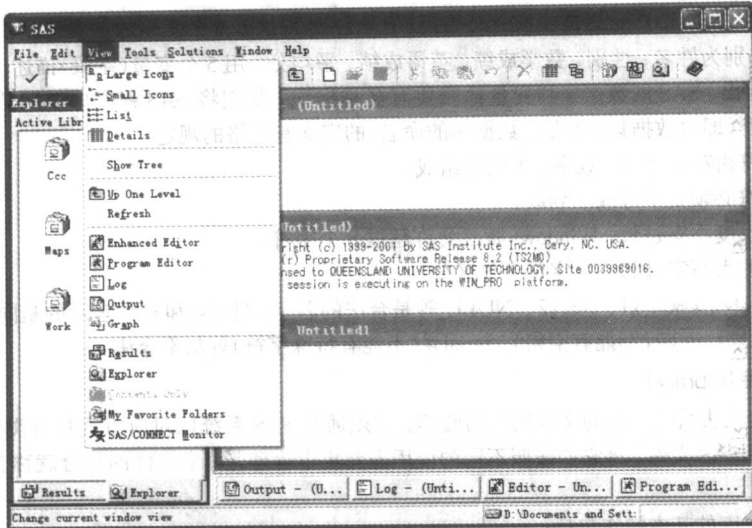


图 1-3 SAS 系统的 View 菜单

## 1.3 SAS 系统的文件管理

### 1.3.1 SAS 文件与逻辑库

通过 SAS 系统的功能建立、维护和管理的文件统称为 SAS 文件，其中含多种不同类型的文件，主要包括数据集(Datasets)文件、索引文件以及用于面向对象开发任务的目录树(catalog)类型文件等。数据集文件是 SAS 各种格式文件中最重要的一种，是 SAS 使用和分析计算的原始数据来源。SAS 目录(Catalog)用来管理各种不能表示成行列表结构的数据，例如记录系统设置、图像、声音等。

在利用 SAS 进行数据分析时，首先必须将要分析的数据正确合理地生成 SAS 数据集。SAS 系统的文件以及对于文件的管理涉及数据集、逻辑库两个基本概念，它们构成了 SAS 应用的基本数据平台。下面我们分别对这两个概念以及相关管理与操作方法进行详细的介绍。

#### 1. SAS 数据集

SAS 数据集(SAS Datasets)可以看做由若干行和若干列组成的表格，各列可以取不同的类型，如整数型、浮点型、时间型、字符型、货币型等。SAS 数据集以特殊格式的二进制文件保存。我们可以通过一个在 SAS 中定义的逻辑名来使用 SAS 数据集，而不必时时关心数据集存储在磁盘的什么位置上。

例如，例 1.2 中生成了一个名为 Stu02 的数据集，其逻辑形式如下：

02 级 1 班学生成绩排名

Obs	name	sex	math	english	avg
1	吴青云	F	92	84	80.3333
2	李清华	F	92	78	77.3333
3	徐凯歌	M	87	78	75.2500
4	张萍江	M	87	78	75.2500
5	王红	F	89	73	73.5833
6	李明	M	91	68	71.9167

数据集的一行称为一个观测(Observation)，一列叫做一个变量(Variable)，一个 SAS 数据集等价于关系型数据库系统中的一个表。利用数据库应用中熟悉的术语，一个观测常常被称做一个记录，一个变量被称

为域或者字段。在上述 Stu02 数据集中有 6 个观测(记录), 分别代表 6 个学生的基本情况, 每个学生的情况包括 5 个数据, 分别为姓名、性别、数学成绩、英语成绩、平均分, 用 5 个变量(字段)描述。

从上述实例看出, SAS 的数据集和变量都需要有名称标识, 我们将 SAS 程序中使用的名称称为“标识符”, SAS 对标识符(即对数据集、变量、数据库的命名)的定义有严格的规定:

- 1) SAS 标识符由英文字母、数字、下划线组成。
- 2) 第一个字符必须是字母或下划线。
- 3) 标识符的长度不宜过长, Windows 版已允许超过 8 个字符。
- 4) 标识符中的大写字母和小写字母不作区分。

比如, name、abc、aBC、x1、year12、\_NULL\_ 是合法的名字; 且 abc 和 aBC 表示的是同一个标识符; 而 class-1(不能有减号)、a bit(不能有空格)、serial#(不能有特殊字符)等却不合法。

## 2. SAS 逻辑库(Library)

SAS 文件通常无法通过一般的文本编辑器阅读, 必须通过 SAS 系统以对应方式打开浏览。SAS 系统为了便于管理和使用 SAS 文件, 将它们按照不同的应用需要集中管理, 每个文件的集合就称为一个 SAS 逻辑库(Library)。

建立 SAS 逻辑库的基本操作包括:

1) **逻辑库的命名** SAS 逻辑库的命名同样需要遵循上述 SAS 命名规则。在 MS DOS 和 Windows 版本中, 一个 SAS 逻辑库实际就是指向磁盘中一个子目录的逻辑指针(特殊情况下一个逻辑库可以由几个子目录组成, 也可以几个逻辑库指向同一磁盘目录)。

2) **通过程序命令建立逻辑库** 在程序中我们可以通过使用 Libname 命令建立逻辑库名和子目录的连接。命令格式为:

```
Libname 逻辑库标识符 "文件夹路径";
```

例如, 在 d:\SAS2003 子目录中保存了若干 SAS 数据集文件, 在程序中可以用下述语句实现逻辑库名 mysas 与子目录 d:\SAS2003 的连接:

```
libname mysas "d:\SAS2003";
```

3) **通过 SAS 命令菜单方式建立逻辑库名** 操作过程为: 进入 Explorer 窗口, 右击 (Library) 图标, 选择“new”命令, 之后在弹出的窗口(见图 1-4)的 Name 栏中输入库名, 在 Path 栏中输入目录路径或点击右侧的“Browse”按钮选择磁盘中的文件夹, 选中右侧的“Enable at startup”使其永久有效, 然后点击“OK”按钮完成新库名的定义。

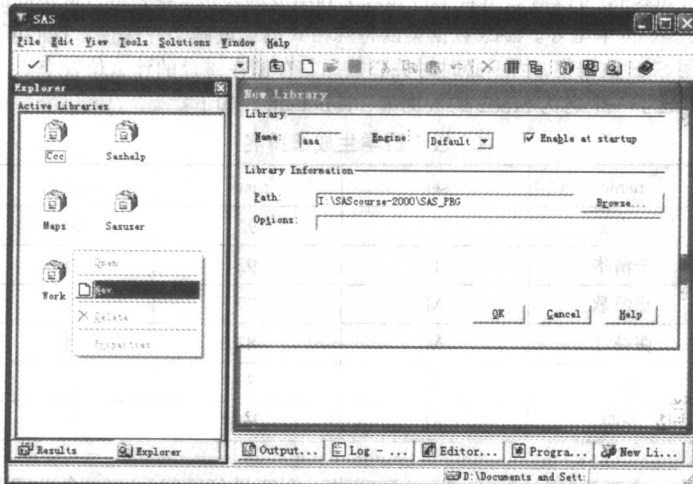


图 1-4 建立新逻辑库

4) **查看逻辑库的属性** 要查看已建好的逻辑库指向哪个文件夹等属性,可以先选中该逻辑库,右击鼠标,选择“Properties”,查看其属性。

另外,除了用户可以建立自己定义的逻辑库之外,SAS系统提供了三个预定义的SAS数据库:Work、Sasuser、Sashelp。其中,Work数据库称为**临时工作库**,用于存放SAS系统工作过程中的临时文件,当用户退出SAS系统时,这些临时文件会被自动删除。Sasuser库保存与用户个人设置有关的文件;它是永久性的,即使用户退出了SAS系统,Sasuser中的文件也不会被删除。Sashelp帮助信息库保存与SAS帮助系统、应用实例有关的文件,它也是永久性的逻辑库。

综上所述,SAS文件本质上可以分为**临时文件**和**永久文件**两类:临时文件在退出SAS系统时自动被删除,永久文件则会长期保留在磁盘中。所以,在应用过程中,我们通常把中间结果数据集或练习数据集作为临时数据集文件保存,而将以后需要使用的数据集保存为永久数据集。

从数据集名称的表达形式上来看,临时数据集和永久数据集是不同的:临时数据集可以用“单水平名”标识,即只有数据集名,比如Stu,因此,这种名字的数据集文件一定保存在Work库中,所以不需要指定文件的路径就完全可以确定所需要的数据集。而永久数据集名称由两部分组成(双水平标识),前一部分是它的库名,后一部分为数据集名,两部分中间用小数点连接,比如放在mysas库(即“d:\SAS2003”子目录)中的数据集Stu02需要用mysas.Stu02表示。这样指定的数据集名在生成时自动存放到由库名指定的子目录中,在读取时可以到指定的子目录读取,并且不会被自动删除。

临时数据集除用单水平名外,在语法上也可以用带有Work库名的“双水平名”标识,例如对于临时数据集Stu01,标识为Work.Stu01的效果与单水平名标识完全相同。


生成永久数据集的实例如下:把数据集放入指定的子目录“d:\SAS2003”中,在程序中可以用如下语句建立逻辑库,除Work逻辑库以外的其他逻辑库中的数据集,都可以作为永久数据集保存:

```
libname mysas "d: \ SAS2003";
data mysas.SAS2003;
.....
proc sort data = mysas.SAS2003;
.....
```

对于已经事先创立的永久SAS数据集,如果已建立了逻辑库,当需要在某个过程步中调用它时,我们不必重复定义逻辑库,而需要在该过程步中加上选项:data=mysas.SAS2003,以指明要使用的数据集。

例如要对已经创建完成的逻辑库mysas中的数据集SAS2003进行排序处理,可直接在程序编辑窗口使用语句:

```
proc sort data = mysas.SAS2003;
run;
```

并点击 Submit  就可完成对指定数据集的排序处理过程。

### 1.3.2 数据集文件的基本操作

#### 1. 文件复制

不同逻辑库之间的数据文件可以进行复制备份,也可将Work临时库中的数据集复制到一个逻辑库中永久保存到磁盘中。

操作过程为:双击逻辑库名,选择需要复制的数据文件,之后右击文件名称,在弹出的快捷菜单中选择“Copy”命令(见图1-5),选定并打开目标逻辑库名称,再右击鼠标,在弹出的快捷菜单中选择“Paste”命令实现粘贴操作,则完成了将一个逻辑库中的数据文件复制到另一个逻辑库的操作,本质上也就是将该文件从一个目录复制到另一个目录中。

#### 2. 文件浏览

在SAS系统中,你可以使用窗口来浏览和管理SAS文件和其他文件。



- ▶ 在 Explorer 窗口中, 可以查看和管理保存在逻辑库中的 SAS 文件, 逻辑库名是保存文件物理位置(路径)的指针名。你可以把库标记想象为一个目录的昵称或快捷方式。
- ▶ 选择主菜单命令 View|Show Tree, 显示 Explorer 窗口(见图 1-6)。双击左侧的图标, 则右侧窗口会显示该库中的 SAS 文件, 就像使用 Windows 中的资源管理器一样。它可以实现对逻辑库中数据文件的复制、删除、改名等操作。

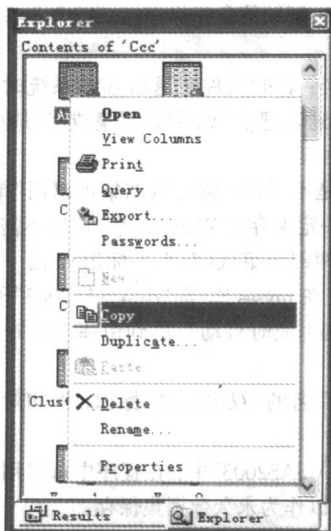


图 1-5 复制数据文件

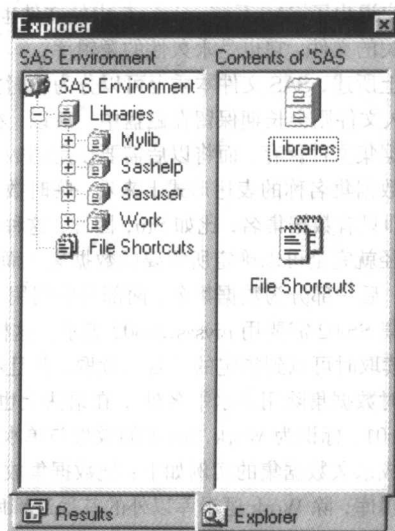


图 1-6 浏览和管理 SAS 文件

- ▶ 使用收藏夹窗口, 可以查看和管理操作系统中的所有文件。选择菜单 View|My Favorite Folders, 会弹出窗口, 如图 1-7 所示。这个窗口可以与 Explorer 类似的方式支持对非 SAS 文件的浏览。

### 3. SAS 数据文件浏览

使用 Explorer 窗口能轻松地管理 SAS 数据集, 例如查看数据文件的详细资料, 在数据库中排序文件, 查看文件属性以及打开一个文件。

操作步骤为: 在 Explorer 窗口中双击 Li-

braries, 找到数据集所在的库名(本例中库名为 CCC), 双击 CCC 库, 在 Explorer 窗口中显示该库中的所有数据集名, 找到要打开的数据集名, 双击该文件名, 此例选择数据集 Reg\_meat1, 之后 SAS 系统打开 Viewtable 窗口并显示数据集 Reg\_meat1 中的数据。图 1-8 打开了 CCC 库中的数据 Reg\_meat1, 它包含了关于人均消费肉类产品的数据。

### 4. 为数据文件建立快捷方式

利用 Explorer 窗口的功能, 可以为所有的外部文件创建一个快捷方式。例如: 将 Explorer 窗口选为当前窗口, 选择 View|Show Tree 菜单, Explorer 窗口如图 1-9 所示, 右击“File Shortcuts”文件夹, 选择“New File Shortcut”菜单, 弹出“File Shortcut Assignment”窗口。在 Name 中输入一个简单的名字, 在 Method 栏中选定 DISK, 在 File 中通过右侧的“Browse”按钮从磁盘中选择一个经常使用的文件名, 点击确定按钮, 则在 Explorer 窗口右侧中建立起这个文件的快捷方式。如图 1-9 中所示的 Name 为“NEW0”, 以后使用该文件时, 只要双击此快捷图标, 该文件就可打开。在应用中可以为频繁使用的 SAS 程序建立快捷方式。这些快捷方式存储在 Explorer 窗口的 File Shortcuts 文件夹中。

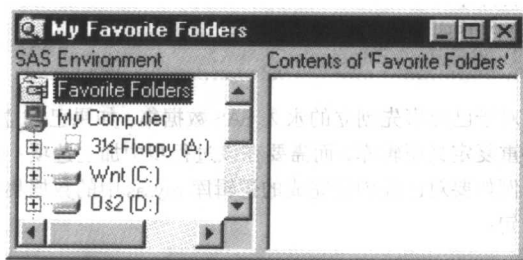


图 1-7 My Favorite Folders 窗口