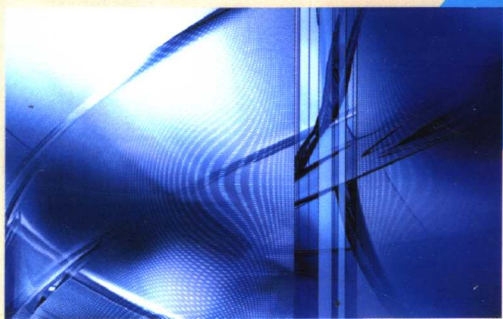


SAS 与 **统计分析**



胡希远 编著

西北农林科技大学出版社

SAS 与统计分析

胡希远 编著

西北农林科技大学出版社

内容简介

本书在介绍 SAS 软件基本操作、编程、数据管理和主要模块功能的基础上,通过大量实例详细讲解了描述性统计、假设检验、相关与回归分析、方差分析、协方差分析、非参数检验、属性数据分析、主成分分析与因子分析、聚类分析与判别分析、典型相关与对应分析等统计方法的要点与 SAS 软件实现过程,并对每例分析结果作了详细解释,便于读者融会贯通。本书凝聚了作者多年的教学与应用经验,深入浅出地介绍各统计分析方法及应用,实例丰富。既可作为高等院校本科生、研究生的教材,也可作为科研人员数据分析的参考用书。

图书在版编目(CIP)数据

SAS 与统计分析/胡希远编著. —杨凌:西北农林科技大学出版社,2007
ISBN 978-7-81092-363-7

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

中国版本图书馆 CIP 数据核字(2007)第 129045 号

SAS 与统计分析 胡希远 编著

出版发行 西北农林科技大学出版社
地 址 陕西杨凌杨武路 3 号 邮 编:712100
电 话 总编室:029-87093105 发行部:87093302
电子邮箱 press0809@163.com
印 刷 西安华新彩印有限责任公司
版 次 2007 年 8 月第 1 版
印 次 2007 年 8 月第 1 次
开 本 787mm×1092mm 1/16
印 张 15
字 数 346 千字

ISBN 978-7-81092-363-7

定价:22.00 元

本书如有印装质量问题,请与本社联系

前 言

利用计算机进行数据处理和统计分析,不仅可以提高工作效率,而且可以帮助对统计方法的理解和进行手工无法实现的复杂统计分析,大大拓展利用统计方法解决问题的深度和广度。而统计软件正是计算机数据处理和统计分析不可或缺的工具。

SAS 软件拥有完备的数据访问、数据管理、数据分析和数据呈现功能,是目前应用最为广泛的统计分析软件之一。SAS 在数据统计分析方面的功能尤为完善,包含了从简单的描述性统计到复杂的现代统计分析方法。利用 SAS 语言也可以很容易地自行编程进行特定的分析。这些特点使得 SAS 软件成为十分有力的统计分析工具。

SAS 系统是一个庞大的信息管理与统计分析系统。虽然软件的帮助系统提供了大量、详尽的说明,SAS 研究所也提供了相应的使用说明书,但对初学者而言,直接阅读这些英文说明书,既要花费大量的精力和时间,而且也不易读懂。近年来有关 SAS 的中文教程也纷纷面世,但对于非专业统计工作者,在应用中仍感有一定难度。鉴于此,编者拟结合自身的使用与教学经验,提供一本内容适中、简明易懂的参考书。设定的读者对象是有一些统计基本知识,并准备用 SAS 软件解决实际问题的研究人员和学生。

全书共分为 14 章,第 1 章到第 3 章介绍了 SAS 的基本几面操作、必要的语句、编程基础、数据管理等方面的内容,使读者对 SAS 系统有一个初步了解;第 4 章介绍了 ASSIST、INSIGHT 和 ANALYST 等 3 个不需编程就能够完成基本的数据管理、生成统计图、统计报表和常用的统计分析的视窗的使用。在它们的帮助下,使得初学者可以较快地学会使用 SAS 进行的常用操作,并在操作过程中进一步熟悉 SAS 的特点,为深入学习程序操作打下良好的基础。第 5 章到第 11 章介绍了基本统计描述和常用统计分析的 SAS 实现方法,如统计图、t 检验、方差分析、相关与回归分析、协方差分析、非参数检验和分类数据分析等。第 12 章到第 14 章为较复杂的多元统计分析 SAS 编程实现方法,包括主成分分析与因子分析、聚类分析与判别分析以及典型相关与对应分析。为求分析例题典型和便于举一反三,书中引用的例子参阅了较多的文献,并基于作者的使用经验和理解给予了详尽的解释。

本书既可作为大专院校学生的教材使用,也可作为科研人员数据分析的参考资料。

由于编者水平所限,不妥和错误之处在所难免,恳请各位读者和广大同行多提宝贵的批评和建议,使本书不断得到完善。另外,对于所引用例题的有关参考书籍,列在书末的参考文献中,在此特向其作者表示衷心的感谢。

目 录

第 1 章 SAS 系统入门	(1)
1.1 SAS 系统概述	(1)
1.2 SAS 系统界面操作	(2)
1.3 SAS 基本概念	(5)
1.4 SAS 系统的文件管理	(8)
1.5 SAS 系统文件库操作	(11)
1.6 数据集的导入和导出	(14)
第 2 章 SAS 编程与数据步	(17)
2.1 SAS 语言	(17)
2.2 SAS 程序规则	(23)
2.3 SAS 数据步	(24)
2.4 数据步赋值、输出与结构语句	(28)
2.5 数据集的编辑整理	(34)
2.6 用 PROC SQL 管理数据	(38)
第 3 章 SAS 过程步	(42)
3.1 SAS 过程步用法及常用语句	(42)
3.2 PRINT 过程应用	(45)
3.3 用 TABULATE 过程做统计量表格	(50)
3.4 用 SORT 过程排序	(53)
第 4 章 无需编程模块简介	(54)
4.1 SAS ASSIST 简介	(54)
4.2 INSIGHT 模块	(60)
4.3 分析家 Analyst 简介	(62)
第 5 章 数据的统计描述	(65)
5.1 描述性统计的基本概念	(65)
5.2 用 SAS 计算统计量	(67)
5.3 统计图形简介	(74)
5.4 用 SAS 绘制统计图形	(75)
第 6 章 正态性检验和 t 检验	(85)
6.1 样本数据总体的正态性检验	(85)
6.2 <i>t</i> 检验	(86)
第 7 章 方差分析	(91)
7.1 方差分析概述	(91)
7.2 方差分析的 SAS 主要过程	(92)

7.3	试验资料的方差齐性检验	(95)
7.4	完全随机设计试验资料的方差分析	(97)
7.5	系统分组设计试验资料的方差分析	(101)
7.6	随机区组设计试验分析	(103)
7.7	拉丁方设计试验资料的方差分析	(107)
7.8	裂区设计试验分析	(110)
7.9	正交设计试验结果的方差分析	(112)
第 8 章	相关与回归分析	(115)
8.1	线性相关	(115)
8.2	线性回归分析	(119)
8.3	非线性回归分析	(128)
第 9 章	协方差分析	(137)
9.1	单向分组资料的协方差分析	(137)
9.2	两向分组资料的协方差分析	(141)
第 10 章	非参数检验	(145)
10.1	非参数检验概述	(145)
10.2	非参数分析过程 NPAR1WAY	(145)
10.3	分析实例	(147)
10.4	等级(秩次)相关	(153)
第 11 章	分类数据分析	(155)
11.1	分类数据与列联表	(155)
11.2	分类数据分析方法	(156)
11.3	分类数据分析 SAS 过程	(157)
11.4	分析实例	(159)
第 12 章	主成分分析与因子分析	(168)
12.1	主成分分析	(168)
12.2	因子分析	(172)
第 13 章	聚类分析与判别分析	(183)
13.1	系统聚类法简介	(183)
13.2	样品系统聚类	(185)
13.3	变量聚类	(191)
13.4	判别分析	(197)
第 14 章	典型相关与对应分析	(215)
14.1	典型相关分析	(215)
14.2	对应分析	(224)
参考文献	(234)

第 1 章 SAS 系统入门

1.1 SAS 系统概述

1. SAS 与统计分析

在众多的统计软件中,SAS 以运行稳定、功能强大而著称。近 20 多年来,SAS 一直占据着统计软件的高端市场,已被全世界 120 多个国家和地区的近三万家机构所采用,直接用户则超过三百万人,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,堪称统计软件界的巨无霸。在以苛刻严格著称于世的美国 FDA 新药审批程序中,新药试验结果的统计分析规定只能用 SAS 进行,其他软件的计算结果一律无效!哪怕只是简单的均数和标准差也不行!由此可见 SAS 的权威地位。在国际学术界有条不成文的规定,凡是用 SAS 统计分析的结果,在国际学术交流中可以不必说明算法,由此可见其权威性和信誉度。

2. SAS 版本介绍

SAS 系统全称为 Statistics Analysis System,最早是在 20 世纪 60 年代末期由两位北卡州立大学统计系的教授开发。第一版的 SAS 只含一般线性模型的分析法,而且只适用于 IBM 的主机;1976 年成立 SAS 公司负责软件的发展、维护并提供相关服务。PC 版本的 SAS 于 1987 年推出 V6.02 版本,1989 年推出 SAS/PC(V6.04)版本;1997 年下半年推出适用于多种操作系统的 V6.12 版本(Windows 版);2000 年 2 月又推出 SAS 系统 V8 版本,2001 年推出 SAS 系统 V8.2 版本;目前 SAS 最新版本为 V9.01。

3. SAS 的功能模块

SAS 系统是一个组合软件系统,它由多个功能模块组合而成,其基本部分是 SAS/BASE 模块。SAS/BASE 模块是 SAS 系统的核心,承担着主要的管理任务,并管理用户使用环境,进行用户语言的处理,调用其他 SAS 模块和产品。也就是说,SAS 系统的运行,首先必须启动 SAS/BASE 模块,它除了本身所具有数据管理、程序设计及描述统计计算功能以外,还是 SAS 系统的中央调度室。它除可单独存在外,也可与其他产品或模块共同构成一个完整的系统。各模块的安装及更新都可通过其安装程序非常方便地进行。SAS 系统具有灵活的功能扩展接口和强大的功能模块,在 SAS/BASE 的基础上,还可以增加如下不同的模块而增加不同的功能: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。SAS 有一个智能型绘图系统,不仅能绘各种统计图,还能绘出地图。SAS 提供多个统计过程,每个过程均含有极丰富的任选项。用户还可以通过对数据集的一连串加工,实现更为复杂

的统计分析。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。

在使用 SAS 系统前,应首先熟悉 SAS 软件的集成环境、各个功能窗口和窗口切换,掌握基本菜单命令、工具条按钮、文件库、浏览器和帮助菜单的用法。掌握 SAS 程序的结构,熟悉过程、语句、选项等编程概念。掌握利用计算机整理试验数据的方法和技能。

1.2 SAS 系统界面操作

1.2.1 SAS for Windows 系统的启动与退出

1. SAS 的启动

若已正确安装了 SAS 系统,开机后在 windows 桌面操作环境下,从开始菜单的程序文件夹中找到 SAS 系统文件夹,用鼠标单击相应的菜单,或直接双击桌面上 SAS 系统的快捷键图标,即可进入 SAS 系统。SAS 系统启动后,自动显示出其主界面(见图 1.1)。

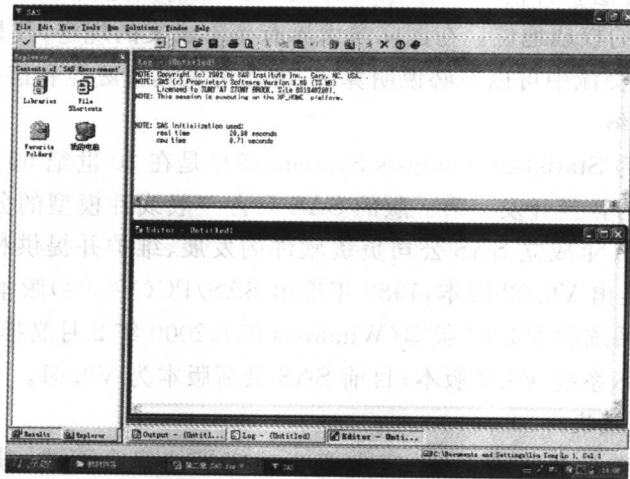


图 1.1 SAS 主界面

2. SAS 的退出

当需要退出时,可以单击菜单 File,选择其中的 Exit 命令,或者单击 (关闭)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 按钮,就可以彻底退出 SAS 系统;如果需要继续使用 SAS 系统,单击 按钮。

1.2.2 SAS for Windows 系统的基本窗口

SAS 启动后,出现如图 1.1 的 SAS 运行界面,术语称为“SAS 工作空间(SAS Application Work Space)”。它象其他 Windows 应用程序一样,在一个主窗口内,包含若干个子窗口,并有菜单条、工具栏、状态栏等。

SAS 在启动时,默认打开五个窗口: SAS 增强程序编辑(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。其中前三个窗口较为常用,后两个相当于索引或者目录树窗口,起辅助的作用,所以这两个窗口默认地固定在工作空间的左侧。SAS 子窗口的排列方式也可以自行设置。在 SAS 系统进行操作的过程中,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。用鼠标点击某个窗口就可以激活该窗口。由于不同窗口有不同的功能,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,在应用时需注意这一点。

1. 增强程序编辑窗口(Enhanced Editor)

该窗口的使用类似于 Windows 中的记事本程序,可以在其中编辑文本文件,主要是编辑 SAS 程序。程序可以直接在窗口中键入,然后可用复制、剪切、粘贴等编辑命令或菜单进行编辑,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,并同时语法检查。
 - ※ 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;
 - ※ 浅黄色表示数据块;
 - ※ 红色表示可能的错误。
- ▶ 程序段的展开和收缩功能,单击每个程序段左侧的加号或减号即可实现,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。
- ▶ 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,实际上就是扩展名为“.sas”的纯文本文件。

2. 日志(Log)窗口

- ▶ Log 窗口用于输出程序在运行时的各种有关信息。主要有以下几种内容:
 - 程序行,黑色,记录执行过的每一条语句。
 - 提示,蓝色,以 NOTE 开始,提供系统或程序运行的一些常规信息,大多数时候对此可以视而不见。
 - 警告,绿色,以 WARNING 开始,一般在程序中含有系统可以自动更正的小错误时出现,此时会提供错误序列号。
 - 错误,红色,以 ERROR 开始,当出现该信息说明程序有错误,执行结果必然是不正确的。
 - ▶ Log 窗口中的内容在保存时应当被存为扩展名为“.log”的纯文本文件。
 - ▶ Log 窗口不能编辑更改。
- ### 3. 结果输出(Output)窗口
- ▶ 结果输出窗口在 SAS 启动后即被打开,但被其他窗口所覆盖,只有在运行了某个程序后才会自动移到显示的前端,从而被显示出来。在该窗口中,可以对 SAS 程序的输出结果进行阅读。
 - ▶ 结果输出窗口中的内容是分页显示的,每一页最上方均显示相应的页标题,结果是

生成时的日期和时间。

- ▶ 当结果输出非常长时,为了能够方便地查阅某一部分结果,可以利用结果(Results)窗口中的目录树进行快速定位。
- ▶ Output 窗口不能编辑更改。

4. 结果(Results)窗口

- ▶ 结果窗口帮助用户浏览和管理所提交 SAS 程序的输出结果,实际上称为结果索引窗口可能更为妥当。在该窗口中将 SAS 系统的所有输出结果依次按照目录树的结构加以排列,每一个过程步的结果被表示为一个点,展开该结点就可以进一步看到表示不同输出内容的子结点,使用鼠标右键单击每个结点,就可对输出结果的各部分进行查看、输出、打印或删除等操作。
- ▶ 在缺省情况下,结果窗口被隐藏在资源管理器窗口的背后,当有新的结果输出生成后,它会和结果输出窗口一起自动移到前台被显示出来,单击窗口下方 Results 窗口的标签也可激活该窗口。
- ▶ Results 窗口中的内容在保存时应当被存为扩展名为“.lst”的纯文本文件。

5. 资源管理器(Explorer)窗口

- ▶ 资源管理器窗口的作用类似于 Windows 操作系统的资源管理器,用于浏览和管理 SAS 系统中的各种文件。用户可以查看和管理 SAS 文件库,以及存放在 SAS 文件库中由 SAS 系统建立的数据文件和其他类型的 SAS 文件,用这一窗口还可以建立新的文件库和 SAS 文件。

6. 其他子窗口

- ▶ 除了增强型编辑器外,SAS 还提供了普通的程序编辑(program editor)窗口。它拥有与其他文本编辑器类似的文本编辑功能,可以作为读入、编辑和提交 SAS 程序的窗口。
- ▶ 当使用 SAS 做图时,相应的统计图会在专门的 Graph 窗口中输出。

1.2.3 菜单栏与工具栏

SAS 窗口标题栏下是主菜单。SAS 菜单是动态的,其内容随上下文而不同,即光标在不同窗口其菜单也不同。其中,File(文件)菜单主要是有关 SAS 文件调入、保存及打印的功能。Edit(编辑)菜单用于窗口的编辑(如清空、复制、剪切、粘贴、查找、替换)。Locals(局部)菜单与当前正在进行的操作有关,如果你正在程序窗口中编辑程序,则 Locals 菜单有提交运行、调回修改等项,如果在运行记录窗口或输出窗口则 Locals 菜单项根本不出现。Globals 菜单内容比较复杂,它可以打开被关闭的程序窗口、运行记录窗口、输出窗口、图形窗口,可以进入 SAS 提供的各个独立模块。主菜单下是一个命令条和工具栏菜单。命令条主要是用于与 SAS 较早版本的兼容性,可以在这里键入 SAS 的显示管理命令。工具栏图标提供了常见任务的快捷方式,比如保存、打印、帮助等等。鼠标光标在某一工具栏图标上停留几秒可以显示一个说明。工具栏图标的解释如下:

■ Submit — 提交编辑窗口中的程序。

□ New — 打开一新的编辑窗口。

- Open — 打开文件到编辑窗口。用户指定一个文件调入到编辑窗口内。这个文件从此与编辑窗口相关联,以后的存盘操作将自动存入这个文件。
- Save — 存盘,保存编辑窗口内容,注意如果此窗口已经与一个文件相联系的话此功能将覆盖文件的原有内容而不提示。
- Print — 打印当前窗口内容。
- Print preview — 打印预览。
- Cut — 剪切选定文本。
- Copy — 复制选定文本。
- Paste — 粘贴。注意这些操作是对 Windows 剪贴板进行的,可以用来与其他 Windows 应用程序交换文本、数据等。剪切或复制到剪贴板的内容可以被其他应用程序粘帖,其他应用程序放到剪贴板的内容也可以粘帖到 SAS 的编辑窗口中。
- Undo — 撤销刚才的编辑操作。
- DOS prompt — 临时进入 DOS。
- 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,可以浏览各 SAS 数据库的内容,可以浏览数据库中的数据集、SAS 目录的内容。
- SAS/ASSIST — 启动 SAS 的菜单驱动界面 SAS/ASSIST。
- Help — 启动 Windows 的帮助系统进入 SAS 的帮助。

1.3 SAS 基本概念


1. 数据步与过程步

为说明数据步与过程步,先看一个程序的例子。假设有一个班学生的数学成绩和语文成绩,数学满分为 100,语文满分为 120,希望计算学生的平均分数(按百分制)并按此排名,可以在程序窗口输入此程序:

```
title '95 级 1 班学生成绩排名';
data c9501;
  input name $ 1-10 sex $ math chinese;
  avg = math * 0.5 + chinese/120 * 100 * 0.5;
  cards;
  李 明      男      92      98
  张红艺    女      89     106
  王思明    男      86      90
  张  聪    男      98     109
  刘  颖    女      80     110
;
run;
proc print; run;
```

```
proc sort data = c9501;
  by descending avg;
run;
proc print;run;
```

实际上,输入这样包含中文的程序最好办法不是在 SAS 程序窗口直接输入,因为 SAS 目前对中文输入的处理还不够完善,好的办法是打开一个其他的编辑程序如 Windows 中的记事本(在 Win95 中用开始菜单中的“程序 | 附件 | 记事本”启动),在记事本中复制输入的程序,然后到 SAS 系统程序窗口中使用粘贴命令(用 Edit 菜单的 Paste 或工具栏上的粘贴图标),把程序复制到 SAS 中。也可以在记事本中把编好的程序存盘,然后在 SAS 程序窗口用 File 菜单的 Open 命令打开保存好的程序文件。

要运行此程序,只要用鼠标单击工具栏的提交图标,或用 Locals 菜单的 Submit 命令。运行后,运行记录窗口出现如下内容:

```
50 title '95 级 1 班学生成绩排名';
51 data c9501;
52 input name $ 1-10 sex $ math chinese;
53 avg = math * 0.5 + chinese/120 * 100 * 0.5;
54 cards;
```

NOTE: The data set WORK.C9501 has 5 observations and 5 variables. NOTE: The DATA statement used 0.11 seconds.

```
60 ;
61 run;
62 proc print;run;
```

NOTE: The PROCEDURE PRINT used 0.0 seconds.

```
63 proc sort data=c9501;
64 by descending avg;
65 run;
```

NOTE: The data set WORK.C9501 has 5 observations and 5 variables.

NOTE: The PROCEDURE SORT used 0.05 seconds.

```
66 proc print;run;
```

NOTE: The PROCEDURE PRINT used 0.0 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。如果有错误还会用红色指示错误。比如,最后的 proc print 后面的分号如果丢失,记录窗口显示如下错误:

```
67 proc print;run;
```

181

ERROR 181-322: Procedure name misspelled.

错误说明为过程名错拼,但实际上是丢了分号导致 print 和 run 连成了一个词。回到编辑窗口可以修改程序。正确运行后输出窗口出现如下结果:

95 级 1 班学生成绩排名

OBS	NAME	SEX	MATH	CHINESE	AVG
1	李 明	男	92	98	86.8333
2	张红艺	女	89	106	88.6667
3	王思明	男	86	90	80.5000
4	张 聪	男	98	109	94.4167
5	刘 颖	女	80	110	85.8333

95 级 1 班学生成绩排名

OBS	NAME	SEX	MATH	CHINESE	AVG
1	张 聪	男	98	109	94.4167
2	张红艺	女	89	106	88.6667
3	李 明	男	92	98	86.8333
4	刘 颖	女	80	110	85.8333
5	王思明	男	86	90	80.5000

这里有两页输出,第一页是输入数据后用 PROC PRINT 显示的数据集,第二页为按平均分排名后的结果。

从上面的例子程序可以看出 SAS 程序的一些特点。SAS 程序由语句组成,语句用分号结束。SAS 程序中大小写一般不区分(字符串中要区分大小写)。SAS 程序中的空格、空行一般可以任意放置,这样人们可以安排适当的缩进格式使得源程序结构清楚易读。SAS 程序由两种“步”构成,一种叫数据步(data step),以 DATA 语句开始,用于待分析数据的输入和加工,另一种叫过程步(proc step),以 PROC 语句开始,用于对输入数据的分析处理。数据步和过程步由若干个语句组成。

当一个 SAS 程序被提交时,每一个 DATA、PROC 或 RUN 语句会导致前面的程序步被执行。所以,一个 SAS 程序应以 RUN 语句结束,否则,最后一个程序步不会被执行。每一个程序步被分别执行。执行时,SAS 系统会产生一个日志(log 窗口),记录执行过程中的信息和可能发生的任何错误。执行的结果根据 SAS 程序的不同而有所不同。

SAS 程序通常包括一些过程步,它们会创建报表形式的输出(Output 窗口),有些 SAS 程序执行诸如排序或数据处理过程,除了在日志中的记录之外,不产生可见的输出结果。

2. SAS 数据集与数据文件

SAS 数据集(SAS Datasets)可以看作由若干行和若干列组成的二维表格或一个矩阵,但各列可以取不同的类型值。SAS 数据集存放在以特殊格式存放的二进制文件中,用户只用一个 SAS 中的逻辑名来使用 SAS 数据集而不需关心它到底如何存储在磁盘上。比如,上个例子生成了一个名为 C9501 的数据集,它的逻辑形式如下表:

NAME	SEX	MATH	CHINESE	AVG
李 明	男	92	98	86.8333
张红艺	女	89	106	88.6667
王思明	男	86	90	80.5000
张 聪	男	98	109	94.4167
刘 颖	女	80	110	85.8333

数据集的每一行叫做一个观测(Observation),每列叫做一个变量(Variable)。SAS数据集等价于关系数据库系统中的一个表,实际上一个SAS数据集有时也称作一张表。在数据库术语中一个观测称作一个记录,一个变量称作一个域。在C9501数据集中有5个观测,分别代表5个学生的情况,而每个学生有5个数据,分别为姓名、性别、数学成绩、语文成绩、平均分,所以此数据集有5个变量。

从上面看出,数据集要有名字,变量要有名字,SAS中对名字(数据集名、变量名、文件库名等)有约定:SAS名字由英文字母、数字、下划线组成,第一个字符必须是字母或下划线,名字最多用8个字符,大写字母和小写字母不区分。比如,name,abc,aBC,x1,year12,_NULL_等是合法的名字,且abc和aBC是同一个名字,而class-1(不能有减号)、a bit(不能有空格)、serial#(不能有特殊字符)、Documents(超长)等不是合法的名字。只有以SAS数据集格式存在的数据文件才能用SAS过程步处理。一个SAS数据集实际上就是一个SAS数据文件。

3. SAS 文件库

SAS数据集是各种特殊格式的SAS文件中最重要的一种。多个SAS文件可以放在一起,称为一个SAS文件库(Library),有人称为数据库。文件库有一个库名(Libname),其命名遵循上述SAS名字命名原则。在MS DOS/Windows环境中,一个SAS文件库实际是磁盘上的一个子目录(特殊情况下一个文件库可以由几个子目录组成)。为了把库名和子目录联系起来,使用LIBNAME语句。比如,我们在C:\Y1995子目录中保存了几个SAS数据集,可以用如下语句把库名MYLIB与子目录C:\Y1995联系起来:

```
libname mylib "c:\y1995";
```

SAS系统有三个预定义的文件库:WORK、SASUSER、SASHELP。其中,WORK文件库叫做临时库,存放在其中的SAS文件叫临时文件,这些临时文件当退出SAS系统时会被自动删除。SASUSER库保存与用户个人设置有关的文件,它是永久的,即退出SAS时文件不会被删除。SASHELP库保存与SAS帮助系统、例子有关的文件,是永久的。

1.4 SAS 系统的文件管理

1. 永久数据集的建立

从上面看出,SAS文件分为临时文件和永久文件:临时文件在退出SAS系统时自动被删除,永久文件在退出SAS系统时不自动被删除。所以,把作为中间结果使用的数据集或练习用的数据集作为临时数据集保存,而需要以后再用的数据集则可以保存为永久数据集。临时数据集和永久数据集的区别是:临时数据集可以用单水平名,即只有数据集名,比如C9501,而永久数据集名由两部分组成,前一部分是它的库名,后一部分才是数据集名,两部分中间用小数点连接,比如放在MYLIB库(即"C:\Y1995"子目录)中的数据集TEACH必须用MYLIB.TEACH表示。这样指定的数据集名在生成时可以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,也可以用库名为WORK的两水平名,如WORK.C9501和C9501是一样的。

要生成永久数据集,只要在指定要生成的数据集名时使用两水平名且库名已有定义,比如,要把上面的 C9501 数据集在生成时就放到“C:\Y1995”子目录中,可以用如下语句:

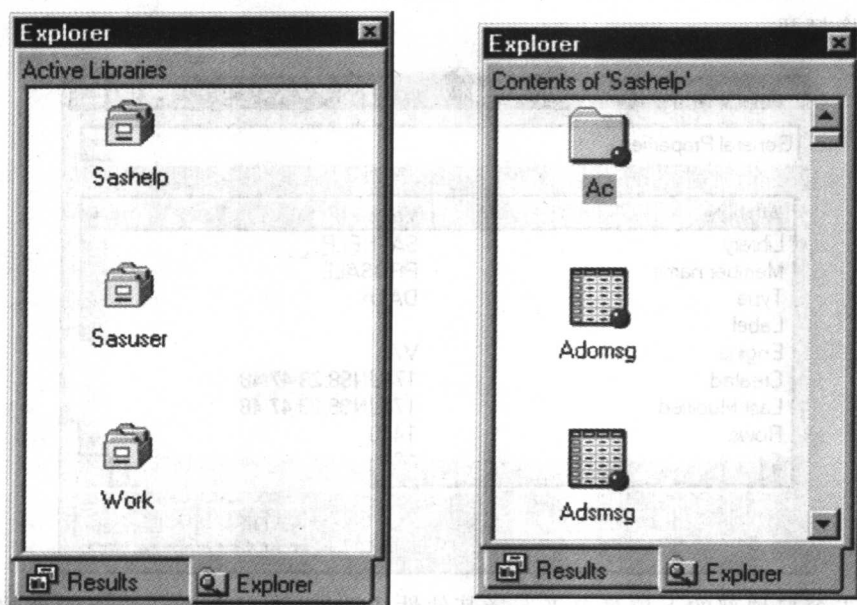
```
libname mylib“c:\y1995”;  
data mylib. c9501;  
.....  
proc sort data=mylib. c9501;  
.....
```

这个程序和前面的例子相比,只是增加了一个定义库名的 LIBNAME 语句,然后在所有用到数据集名 C9501 的地方换成了两水平名 MYLIB. C9501。要注意生成的数据集是 MYLIB. C9501,后面在用到它的时候(在 PROC SORT 中)也必须使用两水平名 MYLIB. C9501 而不能使用单水平名 C9501,这两个名字指向的不是同一个 SAS 文件。

2. 浏览 SAS 文件库中的文件

Explorer 窗口提供了一个简单的界面来管理 SAS 文件。在这里,双击一个文件夹即可打开并查看它的内容。

(1)在 Explorer 窗口中,双击 Libraries 文件夹,当前文件库被列出。双击 Sashelp 文件库,所有内容被列出。



(2)单击工具栏中的(🏠)(向上一级)两次回到 Explorer 窗口中的顶部。

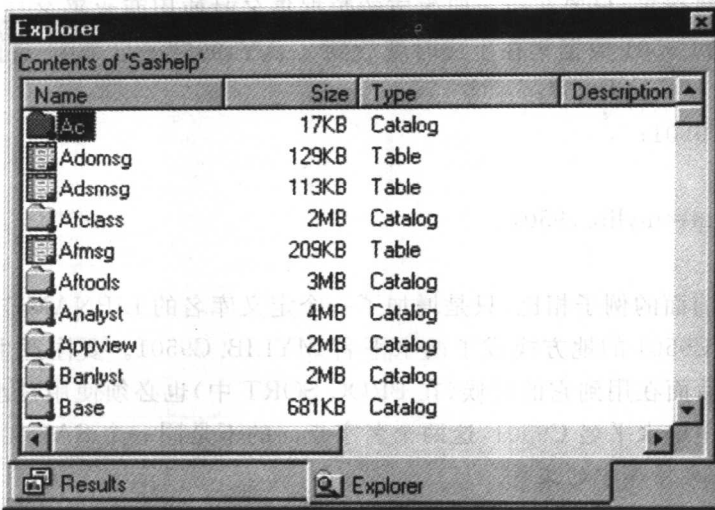
在 Explorer 窗口中,还可以通过选择“View→Up One Level”或者按 Backspace 键来返回已经打开的文件夹。

3. 查看文件属性

可以以大图标、小图标、列表或者详细资料的形式来查看一个 SAS 文件库的内容。

(1)打开 Sashelp 数据库;

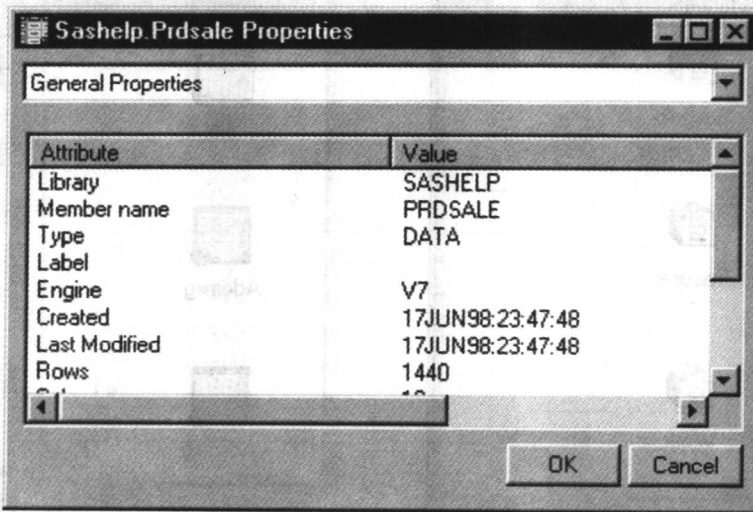
(2)单击工具栏中的(📄)图标或选择“View→Details”,有关文件的信息将被显示;



4. 查看文件性质

在 Explorer 窗口中,可以查看所有 SAS 文件的一般信息以及它的“引擎/主机信息”。针对不同的文件类型,还可以查看不同的附加性质。

(1)在 Sashelp 文件库中,右击 Prdsale 数据表并从弹出菜单中选择 Properties, Properties 窗口将打开。



(2)点击窗口顶部的下拉菜单来选择其他性质种类,选 Columns 来查看列的性质。

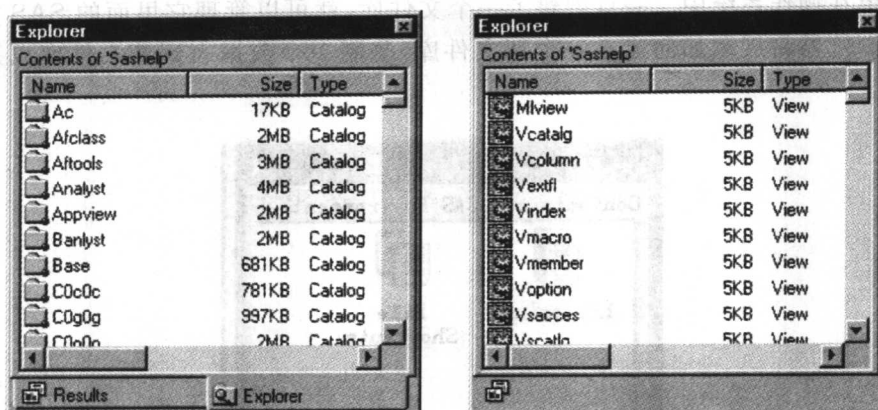
(3)点击 OK 来关闭 Properties 窗口。

对于可修改的性质,具有一个修改弹出菜单。不能修改一个文件的所有性质以及不能修改在 Sashelp 数据库中的性质。

5. 排列文件

默认情况下,Explorer 窗口中的文件是按文件名的字母顺序排列的。可以以升序或降序对任何列进行排序。

(1)点击 Type 列,则文件按类型排列;



- (2) 再点击一次 Type 列, 则按相反的方向排列;
- (3) 选择“View→Refresh”, 回到文件的原始排列。

6. 打开一个文件

可以直接在 Explorer 窗口中查看 SAS 文件的内容, 在这个例子中, 查看一个 SAS 数据表。

- (1) 双击 Prdsale 数据表打开它, 数据表在 VIEWTABLE 窗口中以浏览模式打开;

	Actual Sales	Predicted Sales	Country
1	\$925.00	\$850.00	CANADA
2	\$999.00	\$297.00	CANADA
3	\$608.00	\$846.00	CANADA
4	\$642.00	\$533.00	CANADA
5	\$656.00	\$646.00	CANADA
6	\$948.00	\$486.00	CANADA
7	\$612.00	\$717.00	CANADA
8	\$114.00	\$564.00	CANADA
9	\$685.00	\$230.00	CANADA

- (2) 当浏览完数据表中的数据后, 选择“File→Close”来关闭 VIEWTABLE 窗口;
- (3) 回到 Explorer 窗口的顶部。

1.5 SAS 系统文件库操作

在任务浏览文件中, 已经看到像数据表和 catalogs 这些 SAS 文件都是存储在文件库中的。默认情况下, SAS 系统已经定义好了三个文件库。此外, 用户也可以自定义新库。当定义一个新文件库的时候, 要为 SAS 系统指出保存 SAS 文件的位置。当删除一个 SAS 文件库的时候, 这个指向也被删除掉, SAS 系统将不再能访问这个目录, 但是, 文件库的内