

21世纪高等院校计算机教材系列



# SAS

## 软件与统计应用教程

●汪远征 徐雅静 主编

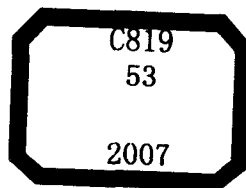


购书可获得增值回报  
提供教学用电子教案

 机械工业出版社  
CHINA MACHINE PRESS



21 世纪高等院校计算机教材系列



# SAS 软件与统计应用教程

汪远征 徐雅静 主编

机械工业出版社

SAS 是当今国际上最著名的数据分析软件系统之一。本书以 SAS 系统 8.2 版本为背景, 通过丰富的实例来介绍如何使用 SAS 中的 INSIGHT、“分析家”以及编程方法来实现常用的一些统计分析。本书涉及的 SAS 应用包括: 描述性统计、假设检验和区间估计、回归分析、方差分析、主成分分析与因子分析、典型相关分析与对应分析、聚类分析与判别分析、属性数据分析等。

本书概念清楚, 逻辑性强、层次分明、例题丰富, 书中每章均附有典型习题, 适合教师课堂教学和学生自学。

本书既可作为本、专科应用统计专业的教材, 也可作为 SAS 统计软件培训班的教材, 还可作为有关数据处理专业人员的参考用书。

### 图书在版编目 (CIP) 数据

SAS 软件与统计应用教程 / 汪远征, 徐雅静主编. —北京: 机械工业出版社, 2007.1

(21 世纪高等院校计算机教材系列)

ISBN 978-7-111-20695-8

I. S... II. ①汪...②徐... III. 统计分析—应用软件, SAS—高等学校—教材 IV. C819

中国版本图书馆 CIP 数据核字 (2007) 第 000940 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策 划: 胡毓坚

责任编辑: 李馨馨

责任印制: 李 妍

北京诚信伟业印刷有限公司印刷

2007 年 2 月第 1 版·第 1 次印刷

184mm×260mm·19.5 印张·479 千字

0001—4000 册

定价: 28.00 元

凡购本书, 如有缺页, 倒页, 脱页, 由本社发行部调换  
销售服务热线电话: (010) 68326294

购书热线电话: (010) 88379639 88379641 88379643

编辑热线电话: (010) 88379739

封面防伪标均为盗版

# 出版说明

计算机技术是一门发展迅速的现代科学技术，它在经济建设与社会发展中，发挥着非常重要的作用。近年来，我国高等院校十分注重人才的培养，大力提倡素质教育、优化知识结构，提倡大学生必须掌握计算机应用技术。为了满足教育的需求，机械工业出版社组织了这套“21世纪高等院校计算机教材系列”。

在本套系列教材的组织编写过程中，我社聘请了各高等院校相关课程的主讲老师进行了充分的调研和细致的研讨，并针对非计算机专业的课程特点，根据自身的教学经验，总结出知识点、重点和难点，一并纳入到教材中。

本套系列教材定位准确，注重理论教学和实践教学相结合，逻辑性强、层次分明、叙述准确而精炼、图文并茂、习题丰富，非常适合各类高等院校、高等职业学校及相关院校的教学，也可作为各类培训班和自学用书。

参加编写本系列教材的院校包括：清华大学、西安交通大学、上海交通大学、北京交通大学、北京邮电大学、北京化工大学、北京科技大学、山东大学、首都经贸大学、河北大学等。

机械工业出版社

# 前 言

SAS 系统 (Statistical Analysis System) 是数据处理和统计领域的国际标准软件之一, 是世界领先的数据分析和信息系统。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。

在众多的统计软件中, SAS 系统以运行稳定、功能强大而著称。近 20 年来, SAS 系统一直占据统计软件的高端市场, 它被成功应用于 120 多个国家和地区的 31000 多个机构中, 直接用户超过 3500000 人。在世界财富 500 强中, 有 90% 的公司使用 SAS 软件。而在财富 500 强的前 100 家企业中, 更是有 98% 的公司使用 SAS 软件。目前 SAS 软件正在国内迅速地普及、发展。尽管 SAS 软件在国内的普及程度还远远不及美国和欧洲一些国家, 但是相信, 随着对数据统计需求的增加, 企业对数据统计分析将会越来越重视, 会有越来越多的企业使用 SAS 系统。

在西方发达国家, SAS 软件是大学生的必修课, 也是其毕业后谋生的资本。在我国, 熟练地应用 SAS 系统也逐渐成为统计师和文、理科大学生的“谋生本领”。

本书通过丰富的实例来介绍如何使用 INSIGHT、“分析家”以及编程方法来实现常用的一些统计分析。这些实例涉及多个领域, 并有详细的分析过程, 便于不同背景的读者学习参考。

本书涉及的 SAS 软件在统计方面的应用包括: 描述性统计、假设检验和区间估计、回归分析、方差分析、主成分分析与因子分析、典型相关与对应分析、聚类分析与判别分析、属性数据分析等。作为 SAS 软件系统与应用统计的入门书, 本书不强调全面介绍 SAS 软件功能, 而是强调掌握 SAS 统计软件的基本用法, 进一步掌握统计分析的基本方法, 然后通过阅读更详细的资料以达到熟练使用统计软件完成统计应用的目的。

本书可以作为大专院校应用统计专业的教材, 也可作为 SAS 统计软件培训班的教材, 还可以作为专业统计人员使用 SAS 软件的参考书。

本书由汪远征、徐雅静主编, 参加编写的还有曲双红、李湘露、吕延华、张文卿、张愿章、孔杰、仝允战、胡尧、朱跃峰、罗楚新、刘庆胜、李溢佳、姜胜利、李慧、刘大明。由于编者水平有限, 难免存在错误与疏漏, 请读者批评指正。

编 者

# 目 录

出版说明

前言

<b>第 1 章 SAS 基础</b> .....	1
1.1 初识 SAS .....	1
1.1.1 SAS 简介 .....	1
1.1.2 使用 SAS .....	2
1.1.3 SAS 对数据文件的管理 .....	5
1.2 初识 INSIGHT 模块 .....	9
1.2.1 调用 INSIGHT 模块 .....	9
1.2.2 INSIGHT 的功能概述 .....	10
1.2.3 数据的预处理操作 .....	11
1.3 初识“分析家” .....	13
1.3.1 调用“分析家” .....	13
1.3.2 功能概述 .....	14
1.3.3 “分析家”的有关操作 .....	14
1.4 SAS 编程初步 .....	19
1.4.1 SAS 程序概述 .....	19
1.4.2 指定逻辑库 .....	20
1.4.3 数据步简介 .....	21
1.4.4 过程步简介 .....	25
1.5 习题 .....	31
<b>第 2 章 SAS 的描述统计功能</b> .....	32
2.1 描述性统计的基本概念 .....	32
2.1.1 统计学的基本概念 .....	32
2.1.2 表示数据位置的统计量 .....	32
2.1.3 表示数据分散程度的统计量 .....	33
2.1.4 表示数据分布形状的统计量 .....	34
2.1.5 其他统计量 .....	34
2.2 在 SAS 中计算统计量 .....	35
2.2.1 用 INSIGHT 计算统计量 .....	35
2.2.2 用“分析家”计算统计量 .....	37
2.2.3 编程实现描述性统计 .....	41
2.3 统计图形 .....	48
2.3.1 定量变量的图形表示 .....	48
2.3.2 分类变量的图形表示 .....	50

2.4	用 SAS 绘制统计图形 .....	51
2.4.1	用 INSIGHT 绘制统计图形 .....	51
2.4.2	用“分析家”绘制统计图形 .....	56
2.4.3	编程绘制统计图形 .....	57
2.4.4	图形的调整与输出 .....	65
2.5	习题 .....	66
<b>第 3 章</b>	<b>区间估计与假设检验 .....</b>	<b>68</b>
3.1	区间估计与假设检验的基本概念 .....	68
3.1.1	区间估计 .....	68
3.1.2	假设检验 .....	69
3.2	总体均值的区间估计与假设检验的 SAS 实现 .....	71
3.2.1	使用 INSIGHT 模块 .....	71
3.2.2	使用“分析家” .....	75
3.2.3	使用 TTEST 过程 .....	78
3.3	总体比例的区间估计与假设检验的 SAS 实现 .....	83
3.3.1	总体比例的置信区间 .....	83
3.3.2	单样本总体比例的假设检验 .....	85
3.3.3	两总体比例的比较 .....	86
3.4	总体方差的区间估计与假设检验的 SAS 实现 .....	87
3.4.1	总体方差的置信区间 .....	88
3.4.2	单样本总体方差的假设检验 .....	88
3.4.3	两样本总体方差的比较 .....	90
3.5	分布检验 .....	91
3.5.1	数据的分布研究 .....	91
3.5.2	在 INSIGHT 模块中研究分布 .....	93
3.5.3	在“分析家”中研究分布 .....	96
3.5.4	使用 UNIVARIATE 过程 .....	98
3.6	习题 .....	98
<b>第 4 章</b>	<b>相关分析与回归分析 .....</b>	<b>101</b>
4.1	简单相关分析 .....	101
4.1.1	相关分析的基本概念 .....	101
4.1.2	用 INSIGHT 模块作相关分析 .....	103
4.1.3	用“分析家”作相关分析 .....	107
4.2	回归分析 .....	110
4.2.1	回归分析的基本概念 .....	110
4.2.2	用 INSIGHT 模块作回归分析 .....	117
4.2.3	用“分析家”作回归分析 .....	124
4.2.4	使用 REG 过程作回归分析 .....	129
4.3	非线性回归 .....	134

4.3.1	多项式回归 .....	134
4.3.2	可化为线性回归的一元非线性回归 .....	142
4.4	习题 .....	146
<b>第 5 章</b>	<b>方差分析 .....</b>	<b>150</b>
5.1	方差分析中的有关概念 .....	150
5.1.1	单因素方差分析问题与模型 .....	150
5.1.2	双因素方差分析问题与模型 .....	152
5.1.3	方差分析中的基本假定 .....	154
5.2	单因素方差分析 .....	155
5.2.1	用 INSIGHT 作单因素方差分析 .....	155
5.2.2	用“分析家”作单因素方差分析 .....	158
5.2.3	用过程进行单因素方差分析 .....	160
5.3	双因素方差分析 .....	162
5.3.1	用 INSIGHT 作双因素方差分析 .....	162
5.3.2	用“分析家”作双因素方差分析 .....	167
5.3.3	用 GLM 过程进行双因素方差分析 .....	171
5.4	均值估计与多重比较 .....	172
5.4.1	概述 .....	172
5.4.2	用“分析家”进行均值比较 .....	173
5.5	习题 .....	177
<b>第 6 章</b>	<b>主成分分析与因子分析 .....</b>	<b>180</b>
6.1	主成分分析 .....	180
6.1.1	主成分分析的概念与步骤 .....	180
6.1.2	使用 INSIGHT 模块作主成分分析 .....	183
6.1.3	使用“分析家”作主成分分析 .....	186
6.1.4	使用 PRINCOMP 过程进行主成分分析 .....	190
6.2	因子分析 .....	194
6.2.1	因子分析的概念与步骤 .....	194
6.2.2	使用 INSIGHT 模块作因子分析 .....	198
6.2.3	使用 FACTOR 过程进行因子分析 .....	202
6.3	习题 .....	210
<b>第 7 章</b>	<b>典型相关分析与对应分析 .....</b>	<b>213</b>
7.1	典型相关分析 .....	213
7.1.1	典型相关分析的概念与步骤 .....	213
7.1.2	用 INSIGHT 模块实现典型相关分析 .....	215
7.1.3	用“分析家”实现典型相关分析 .....	219
7.1.4	用 CANCELL 过程实现典型相关分析 .....	224
7.2	对应分析 .....	230
7.2.1	对应分析的基本概念 .....	230



7.2.2	使用 CORRESP 过程实现对应分析	231
7.3	习题	239
<b>第 8 章</b>	<b>聚类分析与判别分析</b>	<b>244</b>
8.1	聚类分析	244
8.1.1	聚类分析的基本概念	244
8.1.2	系统聚类法的基本思想和步骤	245
8.1.3	用 CLUSTER 过程和 TREE 过程进行系统聚类	248
8.1.4	用 VARCLUS 过程进行变量聚类	252
8.2	判别分析	259
8.2.1	判别分析的基本概念	259
8.2.2	SAS 中作判别分析的过程	263
8.2.3	分析实例	267
8.3	习题	281
<b>第 9 章</b>	<b>属性(分类)数据分析</b>	<b>287</b>
9.1	属性数据及其分析	287
9.1.1	属性数据分析与列联表	287
9.1.2	属性变量关联性分析	288
9.1.3	属性变量关联度计算	290
9.1.4	有序变量关联性分析	290
9.2	SAS 中的属性数据分析	291
9.2.1	$2 \times 2$ 表的分析	291
9.2.2	$r \times c$ 表的分析	293
9.2.3	分层列联表分析	295
9.2.4	有序变量的关联性分析	298
9.3	习题	300

# 第 1 章 SAS 基础

SAS (Statistical Analysis System) 是由美国北卡罗来纳州的 SAS institute 公司开发的一种统计软件, 它被广泛应用在商业、科研领域。SAS 不仅具有强大的统计分析功能, 而且具有一般数据库软件的数据管理功能, 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。

本章介绍统计软件 SAS 的基础知识, 包括 SAS 系统工作环境的使用、SAS 系统对数据文件的管理以及 SAS 系统对数据的简单处理。

## 1.1 初识 SAS

SAS 是一个大型的应用软件系统, 具有完备的数据访问、管理、分析、呈现, 以及应用开发功能。

### 1.1.1 SAS 简介

#### 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 年来, SAS 一直占据着统计软件的高端市场, 用户遍及金融、医药卫生、生产、运输、通信、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。在美英等国, 能熟练使用 SAS 系统进行统计分析是许多公司和科研机构选才的条件之一。在国内 SAS 软件也正迅速普及、发展, 尽管普及程度还远远不及美国和欧洲一些国家, 但是随着对数据统计需求的增加, 企业对数据统计分析将会越来越重视, 会有越来越多的企业使用 SAS 系统。

#### 2. SAS 的版本介绍

SAS 是在 20 世纪 60 年代末期由两位北卡州立大学 (North Carolina State University) 统计系的教授开发的。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; 1972 年 SAS 软件及其使用手册首次对外公开租赁; 1976 年 SAS 公司成立, 主要负责软件的发展、维护并提供相关服务。由于 SAS 公司认为软件多元化是未来发展的必然趋势, 在 1984 年决定以 C 语言改写 SAS 系统的同时, 以多元架构 (Multi-Vendor Architecture, MVA) 为设计重点, 使得 SAS 在不同的操作系统平台下几乎不须修改即可运行, 提高了使用性。至于 PC 版本的 SAS 则于 1987 年推出 (V6.02), 1989 年推出 SAS/PC (V6.04) 版本; 1997 年下半年推出适用于 Windows 的 V6.12 版本; 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。

本书介绍目前最为流行的 SAS V8.2, 所有程序和操作都在 SAS V8.2 中调试通过。

#### 3. SAS 的功能模块

SAS 系统由 30 几个模块组成, 主要包括: BASE、GRAPH、ETS、FSP、AF、OR、IML、

SHARE、QC、STAT、INSIGHT、ANALYST、ASSIST、CONNECT、CPE、LAB、EIS、WAREHOUSE、PC File Formats、GIS、SPECTRAVIEW、SHARE、R/3、Online Tutor: SAS Programming、MDDDB Server、IT Service Vision Client、IntrNet Compute Services、Enterprise Reporter、MDDDB Server common products、Enterprise Miner、AppDev Studio、Integration Technologies 等。

SAS 的分析功能散布在几乎所有的模块之中，较为集中的具有统计分析功能的是 SAS/BASE、SAS/STAT、SAS/QC、SAS/INSIGHT、SAS/ETS 等模块，通过编程可以调用各种分析功能。

随着图形界面、用户友好等程序思想的发展，SAS 也陆续提供了一些不需要学习 SAS 编程就能进行数据管理、分析、报表、绘图的功能，其中做得比较出色的有 INSIGHT 模块和 Analyst（分析家）。INSIGHT 是在基本的 SAS 系统基础上添加的一个模块，提供了数据交互输入、数据探索、分布研究、相关分析，以及各种图形功能。“分析家”则为常见的数据管理和统计功能提供了一个简单易用的图形界面，还能够把用图形界面进行的操作以普通 SAS 程序的形式记录下来供学习参考。

对于常用的一些统计分析方法，SAS 系统中的如下三种方法可以达到同样的目的：

- INSIGHT（交互式数据分析）模块
- Analyst（分析家）模块
- 直接编程方式

一般来说，INSIGHT 模块在数据探索方面比较有特色，最为直观，便于步步深入；“分析家”可提供自动形成的程序，而且在属性数据分析和功效函数计算方面较 INSIGHT 强；编程方式是功能最强的，尤其是一些特殊或深入的分析功能只能用编程实现，但相对来说，编程较难熟练掌握。

本书将分别介绍使用 INSIGHT 模块、“分析家”和直接编程方式来完成常用统计分析功能并说明 SAS 输出结果的含义。

## 1.1.2 使用 SAS

### 1. SAS 的启动

在 Windows 环境中，用如下方法可以进入 SAS 系统的窗口运行环境：

- 用鼠标选择系统菜单项：“开始”→“程序”→The SAS System→The SAS System for Windows V8。
- 如果 SAS 安装系统在桌面上建立有快捷方式图标，双击 SAS.EXE 图标启动 SAS。
- 生成 SAS.EXE 的快捷方式（把 SAS.EXE 用鼠标右键拖到桌面），双击 SAS.EXE 图标启动 SAS。

### 2. SAS 的操作界面

SAS 系统启动后，进入 SAS 操作界面。SAS 操作界面的标准名称为 SAS Application WorkSpace（SAS 应用工作空间），简称 SAS AWS。

与 Windows 应用程序一样，SAS AWS 是一个多窗口界面：在一个主窗口内包含若干个窗口，并有菜单栏、工具栏、状态栏等，如图 1-1 所示。

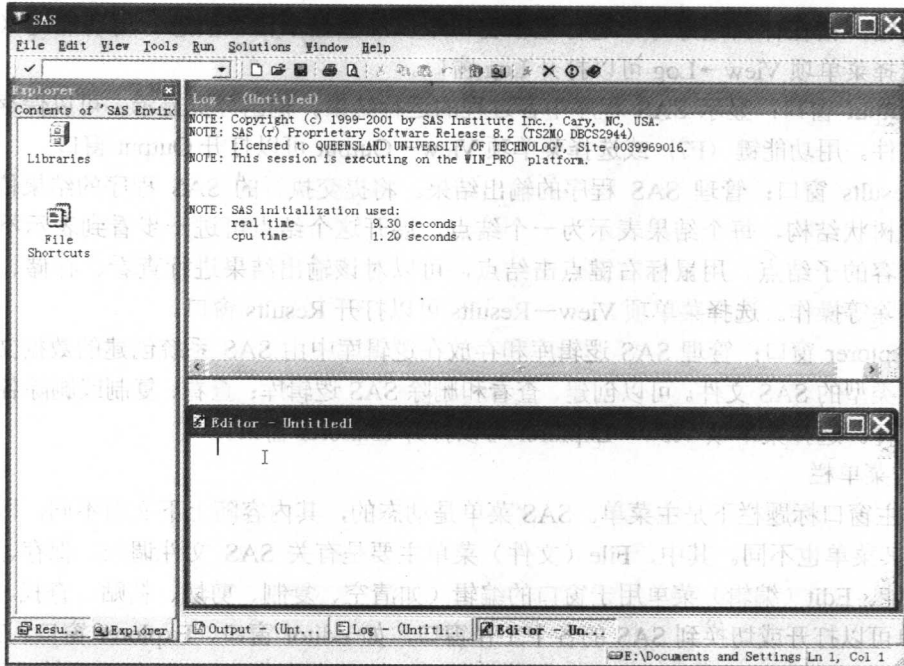



图 1-1 SAS AWS

### (1) 子窗口

SAS 是一个典型的多文档界面程序，因此在其窗体内可以有多个子窗口。但当前子窗口只有一个，即标题栏深色显示的那个。我们所做的操作均是针对当前子窗口的，如果要对其他子窗口进行操作，则需切换为当前子窗口。

SAS V8.2 的子窗口有十几个，但初始状态下能见到的是最常用的五个子窗口：程序编辑窗口 (Editor)、运行记录窗口 (Log)、输出窗口 (Output)、结果窗口 (Results) 和浏览器窗口 (Explorer)，通过点击工作区下方的标签可以切换当前子窗口。

通常，在 Editor 窗口输入 SAS 程序，选择菜单项 Run → Submit，或者直接点击工具栏上的  按钮就可以运行程序；在 Output 窗口查看结果，如果有图形输出，还会出现一个 Graph 窗口；如果进行了多次输出，Output 便保存了多次的结果，从 Results 窗口查看结果可以很容易分清楚；Log 窗口给出了程序运行的信息，如果程序有错误，则给出相应的提示，可以根据提示修改错误。如果 Output 或 Log 窗口的信息过多，造成混乱，也可以在主菜单 Edit 中选择 Clear 清除当前窗口的信息。

下面是五个常用子窗口的功能和特点：

- Editor 窗口：增强的程序编辑器。可以用不同颜色显示 SAS 程序的不同部分，并对 SAS 命令的语法进行检查；可以自动缩进排列程序文本，可以折叠一段程序。其操作类似于 Windows 中的记事本，可以对所编辑的文本（主要是 SAS 程序）进行选中、复制、剪切、粘贴等操作。可以同时打开多个 Editor 窗口。用功能键 <F5> 或选择菜单项 View → Enhanced Editor 可以打开 Editor 窗口。
- Log 窗口：记录程序的运行情况。显示运行是成功还是出错，运行所用时间，如果

出错，错在什么地方等。运行记录窗口中以红色显示错误信息。用功能键〈F6〉或选择菜单项 View→Log 可以打开 Log 窗口。

- Output 窗口：显示 SAS 程序中各过程的运行结果。输出分页显示，可以保存结果为文件。用功能键〈F7〉或选择菜单项 View→Output 可以打开 Output 窗口。
- Results 窗口：管理 SAS 程序的输出结果。将提交执行的 SAS 程序的结果依次排列为树状结构，每个结果表示为一个结点，展开这个结点可进一步看到表示不同输出内容的子结点。用鼠标右键点击结点，可以对该输出结果进行查看、存储、打印和删除等操作。选择菜单项 View→Results 可以打开 Results 窗口。
- Explorer 窗口：管理 SAS 逻辑库和存放在逻辑库中由 SAS 系统创建的数据文件和其他类型的 SAS 文件。可以创建、查看和删除 SAS 逻辑库；查看、复制或删除各种 SAS 文件。选择菜单项 View→Explorer 可以打开 Explorer 窗口。











### (2) 菜单栏

SAS 主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、转换及打印等功能。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）。View 菜单可以打开或切换到 SAS 的各个工作窗口，如 Editor 窗口、Explorer 窗口。Tools 菜单可以打开一些 SAS 提供的小工具如图像编辑器，还包含修改 SAS 运行选项等功能。Run 菜单用于程序执行、远程调用等，仅当 Editor 窗口为当前窗口时有效。Solutions 菜单是 SAS 图形操作界面模块的入口，比如 INSIGHT、Analyst（分析家）、ASSIST 等模块的调用都可以通过 Solutions 菜单。






### (3) 工具栏

主菜单下是一个命令条和工具栏。命令条主要用于与 SAS 较早版本的兼容，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。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。工具栏也是动态的，当光标在编辑窗口时工具栏图标的解释如表 1-1 所示。

表 1-1 光标在编辑窗口时工具栏图标说明

图标	名称	功能
	New	建立新的编辑窗口
	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，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贴，其他应用程序放到剪贴板的内容也可以粘贴到 SAS 的编辑窗口中
	Undo	撤销刚才的编辑操作
	New Library	建立新的 SAS 逻辑库

(续)

图标	名称	功能
	SAS Explorer	打开 SAS 管理器窗口查看、管理 SAS 的各个库和库中的文件
	Submit	运行 (提交) 编辑窗口中的程序
	Clear All	清空当前窗口内容
	Break	中断正在运行的 SAS 程序
	Help	进入 SAS 的帮助界面。因为 SAS 是一个十分庞大的系统, 我们不可能完全了解 SAS 的每一个细节, 所以善于利用 SAS 的帮助系统对于学好用好 SAS 是极为重要的。也可以从菜单进入 SAS 帮助

其中提交程序按钮 \* 最为常用, 点击该按钮即可运行 Editor 窗口中的 SAS 程序。

#### (4) 状态栏

SAS 界面的状态栏中显示当前工作目录, 这是文件打开、保存的默认目录。双击此处可以更改当前工作目录。

### 3. SAS 的退出

在 SAS 应用工作空间中用鼠标单击关闭按钮, 或者选择菜单 File→Exit, 打开 Exit 对话框, 如图 1-2 所示, 单击“确定”按钮, 即可退出 SAS。

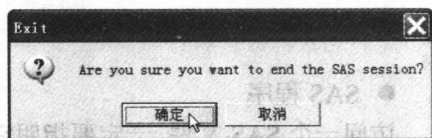


图 1-2 “Exit”对话框

### 1.1.3 SAS 对数据文件的管理


在 SAS 系统中建立的众多 SAS 文件, 可按不同需要将其归入若干个 SAS 逻辑库, 以便对 SAS 文件进行访问和管理。Explorer 窗口可以管理逻辑库和各种 SAS 文件, 一般来说, 对于 SAS 文件的复制、查看和删除等操作, 在 Explorer 窗口下进行比在 Windows 下更为方便。

#### 1. 逻辑库和 SAS 文件

SAS 逻辑库是一个逻辑概念, 在 Windows 环境下, 一个逻辑库就是存放在同一文件夹中的一组 SAS 文件。在 SAS 系统的信息组织中, 总共只有两个层次: SAS 逻辑库是高一级的层次, 低一级的层次就是 SAS 文件本身。

SAS 的逻辑库分为临时库和永久库两种。临时库只有一个, 名为 Work, 其他的库均为永久库。SAS 每次启动时会自动指定 4 个库标记: Work、Sasuser、Sashelp、Sasmap。存放在 Work 中的 SAS 文件叫临时文件, 当退出 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, Sashelp 库保存与 SAS 帮助系统、例子有关的文件。

由此看出, SAS 文件分为临时文件和永久文件: 临时文件存放在临时库中, 在退出 SAS 系统时自动被删除; 永久文件保存在永久库中, 在退出 SAS 系统时不会自动被删除。所以, 通常把作为中间结果或练习使用的数据集保存为临时数据集, 而需要以后再用的数据集则可以保存为永久数据集。如果需要备份, 最好在退出前把临时文件复制到其他的库中。

在 Explorer 窗口中, 单击工具按钮 , 或是在右键菜单中选择 New, 打开 New Library 对话框, 可以建立新库 (如把一个库名 Mylib 和一个实际的文件夹 G:\张三\Myfile 联系起来), 见图 1-3。如果选中 Enable at startup 复选框, 则每次打开 SAS 时所建逻辑库都有效。

SAS 文件是指储存在 SAS 逻辑库中的成员, SAS 的用户文件的主要类型有:

- SAS 数据集

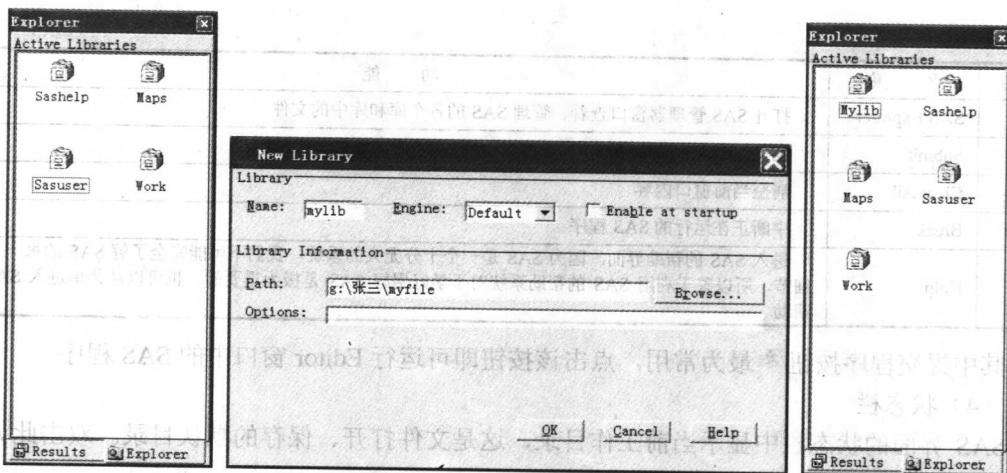


图 1-3 建立新库

### ● SAS 程序

访问一个 SAS 文件，一定要指明它所在的逻辑库，格式为：逻辑库名.文件名。如在 Sasuser 库下建立一个名为 Mydata 的数据集，访问的时候，要用 Sasuser.mydata。

如果不指明逻辑库名，则表示存储在临时逻辑库 WORK 中的 SAS 文件。

### 2. SAS 的名字

SAS 逻辑库和数据集都有名字，这些名字的命名应该尽量使其能反映数据集中储存的信息的内容。对所有的 SAS 的名字（数据集名、变量名、逻辑库名等等）有以下命名规则：

- 1) 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) 不区分大、小写字母。

另外，SAS 逻辑库名最多用 8 个字符；数据集和变量的名字最多用 32 个字符。

比如 name, abc, aBC, x1, year12, \_NULL\_ 是合法的名字，且 abc 和 aBC 是同一个名字，而 class-1（有减号）、a bit（有空格）、serial#（有特殊字符）、12abc（首字符为数字）等不是合法的名字。

### 3. SAS 数据集

SAS 的核心是 SAS 数据集。SAS 数据集是 SAS 文件的一种，一般由以下两部分组成：

#### (1) 描述部分

包含该数据集的一般信息，包括：数据集的名字及其成员类型、数据集建立的日期和时间、观测的数目、变量的数目以及数据集中每一个变量的特征信息，包括：Name（变量名）、Type（类型）、Length（长度）、Format（输出格式）、Informat（输入格式）、Label（标签）。

#### (2) 数据部分

包含该数据集中收集的数据的值，可以看作是一个矩形的表格。图 1-4 所示的数据表就是 SAS V8 自带的数据集 Sasuser.Class<sup>⊖</sup>，其中包含了 19 个学生的有关信息。

<sup>⊖</sup> 如果没有找到这个数据集，按如下步骤可以生成此数据集及其他示例数据集：选择菜单 Solutions→Analysis→Analyst，进入 Analyst，然后选择其中的 Tools-Sample Data，在出现的对话框中选中所有数据集就可以生成示例数据集。

	name	sex	age	height	weight
1	Alice	F	13	56.5	84
2	Becka	F	13	65.3	98
3	Calli	F	14	64.3	90
4	Karen	F	12	56.3	77
5	Kathy	F	12	59.8	84.5
6	Mary	F	15	66.5	112
7	Sandy	F	11	51.3	50.5
8	Sharon	F	15	62.5	112.5
9	Tammy	F	14	62.8	102.5
10	Alfred	M	14	69	112.5
11	Duke	M	14	63.5	102.5
12	Guido	M	15	67	133
13	James	M	12	57.3	83
14	Jeffrey	M	13	62.5	84
15	John	M	12	59	99.5
16	Phillip	M	16	72	150
17	Robert	M	12	64.8	128
18	Thomas	M	11	57.5	85
19	William	M	15	66.5	112

a)

	First name	Gender	Age in years	Height in inches	Weight in pounds
1	Alice	F	13	56.5	84
2	Becka	F	13	65.3	98
3	Calli	F	14	64.3	90
4	Karen	F	12	56.3	77
5	Kathy	F	12	59.8	84.5
6	Mary	F	15	66.5	112
7	Sandy	F	11	51.3	50.5
8	Sharon	F	15	62.5	112.5
9	Tammy	F	14	62.8	102.5
10	Alfred	M	14	69	112.5
11	Duke	M	14	63.5	102.5
12	Guido	M	15	67	133
13	James	M	12	57.3	83
14	Jeffrey	M	13	62.5	84
15	John	M	12	59	99.5
16	Phillip	M	16	72	150
17	Robert	M	12	64.8	128
18	Thomas	M	11	57.5	85
19	William	M	15	66.5	112

b)

图 1-4 数据集

表格的列 (columns) 称为变量 (variables), 对应于原始数据文件或其他一些外部数据库所称的字段 (fields)。如图 1-4 中的 name、sex 等。

表格的行 (rows) 称为观测 (observations), 对应于原始数据文件或其他一些外部数据库所称的记录 (records) 或数据行 (data lines)。

图 1-4a 的第一行显示的是变量名, 右图第一行显示的是相应的标签名。虽然 SAS 的变量名不支持中文字符, 但是标签可以使用中文名称。

#### 4. 数据集的导入

建立数据集的方法很多, 但是, 无论是在 Explorer 窗口中使用 VIEWTABLE 程序, 或是下面将要介绍的 INSIGHT 模块和“分析家”, 还是使用编程操作中专门的数据读入方法来建立数据集, 都需要将数据现场录入, 费时费力。较为简便的方法是, 利用 Excel 录入数据, 并作简单处理, 然后将 Excel 数据表导入到 SAS 数据集中。

导入 Excel 数据表的步骤如下 (以图 1-5 所示的 Excel 数据表为例<sup>①</sup>):

1) 在 SAS 应用工作空间中, 选择菜单 File→Import Data..., 打开导入向导 Import Wizard 第一步: 选择导入类型 (Select import type), 默认的类型为 Excel 数据表, 单击 Next 按钮, 如图 1-6a 所示。

2) 在第二步的 Select file 对话框中, 单击 Browse 按钮, 在“打开”对话框中选择所需要的 Excel 文件, 返回。然后, 单击 Option 按钮, 选择所需的工作表, 单击 OK 按钮返回, 如图 1-6b 所示。单击 Next 按钮, 进入下一步。

3) 在第三步的 Select library and member 对话框中, 选择导入数据集所存放的逻辑库以及数据集的名称, 单击 Next 按钮 (如图 1-7a 所示), 进入下一步。

	A	B	C	D	E
1	姓名	年龄	性别	身高	体重
2	KATE	12	女0	145	43.1
3	LOUISE	12	女	149	55.8
4	JANE	12	女	135	33.6
5	JACLYN	12	女	162	65.8
6	LILLIE	12	女	127	29.1
7	TIM	12	男	147	38.1
8	JAMES	12	男	149	58.1
9	ROBERT	12	男	125	35.9
10	BARBARA	13	女	147	50.8
11	ALICE	13	女	149	48.6
12	SUSAN	13	女	137	30.4
13	JOHN	13	男	159	44.5
14	JOE	13	男	154	47.7
15	MICHAEL	13	男	142	43.1
16	DAVID	13	男	145	35.9
17	JUDY	14	女	149	36.8
18	ELIZABET	14	女	152	41.3

图 1-5 Excel 数据表 Bclass

① 性别一栏的第一个数据原为“女”, 现多输入一个字符“0”, 是为了让 SAS 能将该变量长度识别为 2。导入数据后, 将 0 删除再存盘。



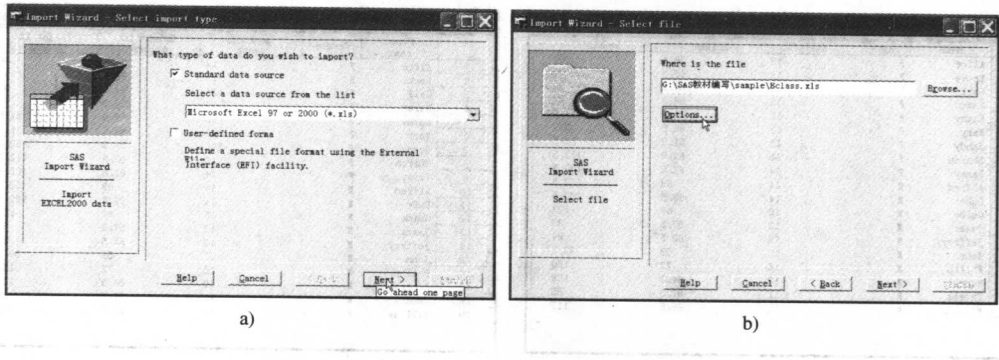


图 1-6 导入向导一、二步

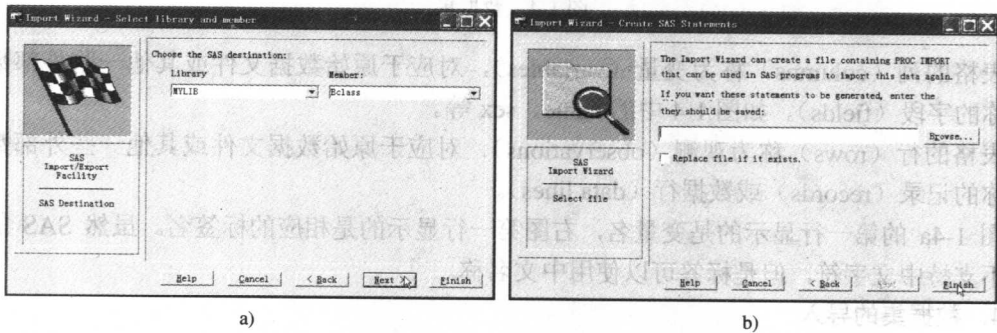


图 1-7 导入向导三、四步

4) 在第四步的 Create SAS Statements 对话框中，可以选择将系统生成的程序代码存放的位置，本例不做选择，直接单击 Finish 按钮（如图 1-7b 所示），完成导入过程。

在 Explorer 窗口中，双击逻辑库 Mylib 中的文件 Bclass，即可看到新导入的数据集，如图 1-8b 所示。

姓名	年龄	性别	身高	体重
1 KATE	12	女	145	43.1
2 LOUISE	12	女	149	55.8
3 JANE	12	女	135	33.6
4 JACLYN	12	女	162	65.8
5 LILLIE	12	女	127	29.1
6 TIM	12	男	147	38.1
7 JAMES	12	男	149	58.1
8 ROBERT	12	男	125	35.9
9 BARBARA	13	女	147	50.8
10 ALICE	13	女	149	48.6
11 SUSAN	13	女	137	30.4
12 JOHN	13	男	159	44.5
13 JOE	13	男	154	47.7
14 MICHAEL	13	男	142	43.1
15 DAVID	13	男	145	35.9
16 JUDY	14	女	149	36.8
17 ELIZABET	14	女	152	41.3
18 LESLIE	14	女	159	64.5
19 CAROL	14	女	154	38.1
20 PATTY	14	女	152	38.6

图 1-8 数据集 Bclass

VIEWTABLE 的第一行显示变量的标签，选择主菜单项 View→Column Names，显示变量名称如图 1-8b 所示。系统对不能识别的变量名依次赋予默认的名称：\_COL0、\_COL1、\_COL2 等。可以在后面介绍的 INSIGHT 模块或“分析家”中修改变量的名称。