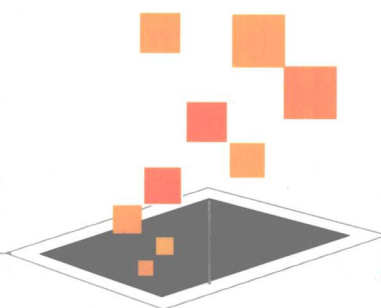




挖掘数据宝藏
助力决策支持

SAS

统计分析 从入门到精通

阮敬 编著

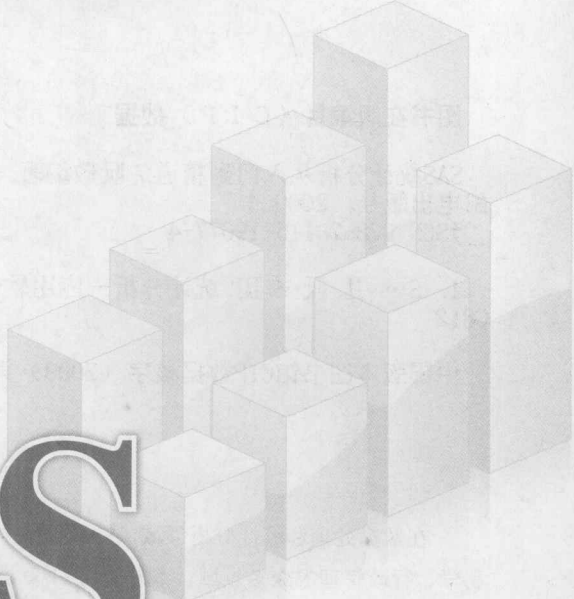


应用案例
全程视频演示

CD-ROM



人民邮电出版社
POSTS & TELECOM PRESS



SAS

统计分析 从入门到精通

○ 阮敬 编著

人民邮电出版社
北京

图书在版编目 (CIP) 数据

SAS统计分析从入门到精通 / 阮敬编著. —北京: 人民邮电出版社, 2009. 4
ISBN 978-7-115-19677-4

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

中国版本图书馆CIP数据核字 (2009) 第009701号

内 容 提 要

在数据处理和统计分析领域, SAS 软件被誉为标准软件, 在我国广泛应用于医学、农林、财经、社会科学、行政管理等众多领域。本书分为 11 章讲解 SAS 统计分析相关知识, 全面透彻地讲解统计分析与数据挖掘技术, 内容包括数据预处理、数据的描述、统计推断、相关与回归分析、因子分析、聚类分析与判别分析、列联分析与对应分析、定性数据分析和时间序列分析。

本书实用性强, 避免复杂的数学公式推导, 并且通过菜单和编程两种方式实现统计分析和得到结论的全部过程。

本书既可作为高等院校本科生和研究生的统计学教材, 又可以作为管理、金融、医学领域进行数据分析的自学教材, 同时还可以作为从事数据分析与数据管理的研究人员的参考用书。

SAS 统计分析从入门到精通

- ◆ 编 著 阮 敬
责任编辑 刘 浩
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
三河市海波印务有限公司印刷
- ◆ 开本: 787×1092 1/16
印张: 19.75
字数: 496 千字
印数: 1—3 500 册
- 2009 年 4 月第 1 版
2009 年 4 月河北第 1 次印刷

ISBN 978-7-115-19677-4/TP

定价: 39.00 元 (附光盘)

读者服务热线: (010)67132692 印装质量热线: (010)67129223

反盗版热线: (010)67171154

前言

在当今知识爆炸的年代，人们日常生活的每一个方面几乎都会产生各种各样的数据，同时也离不开数据，这些数据的类型各异，其表现形式纷繁芜杂。随着数据分析工作的作用凸显，如何对现有数据进行整理、加工、处理和分析，以期得到所谓的结论，作为我们进行决策的依据？如何利用现有数据对将来可能出现的数据结果或结论进行预测？不管是针对企事业单位的管理者或决策者还是从事具体数据分析的工作人员而言，都需要进行合理数据分析流程的规划，区分数据类型，利用适合的数据分析方法，使用方便、快捷、可靠的统计软件作为工具，对特定数据进行分析与预测，从而洞察市场动向、观测人心所在、把握商机，从而提升所在单位的竞争力。

具有深厚数学背景的数据分析方法往往会成为人们继续深入学习的“拦路虎”因此，本书就是要降低学习难度，通过笔者积累的大量真实案例和数据，以文字阐述而不是复杂公式推导的形式深入浅出地剖析统计分析方法的基本原理和步骤，重点在于理清数据分析的基本思路，得到恰当的分析结果。

本书通过菜单操作和编程两种途径，用大量的图形展示每一步操作的细节，一步一步地带领读者走入统计数据分析的美妙世界。

为与国际接轨，本书采用 SAS 9.1.3 Service Package 4 的英文版本进行讲解，希望读者边操作边学习。

为了提高学习效率，本书还附送随书案例的全部数据以及利用 SAS 系统进行统计分析的详细操作视频。

主要内容

本书全面、细致地讲解了 SAS 系统进行数据预处理和数据分析的全过程，全书分为 7 个部分共 11 章。

第 1 部分——数据预处理。该部分内容包括第 1 章，主要结合笔者实际工作经验对数据分析之前的准备工作和处理流程进行系统讲解。同时，对于 SAS 系统环境和界面的基本元素和操作方式进行了介绍，详细阐释了 SAS 编程语言的基本结构，并明确了 SAS 系统分析的对象是数据库当中的数据集。本部分内容是掌握 SAS 系统的基本知识，也是读者必须学习和掌握的内容。

第 2 部分——数据的描述。该部分内容主要为第 2 章，主要讲解如何通过 SAS 系统绘制常见的统计图形和统计表格来描述数据的概貌，并在此基础上计算反映数据集中趋势、离散程度和分布状况的统计量来进行简单的描述统计分析。该部分内容是数据分析工作的基础和前提，同时也是各种高级统计方法的数据描述基础。

第 3 部分——统计推断。该部分包含第 3~5 章，内容涵盖了简单统计推断、方差分析和非参数检验。本部分主要讲解如何利用已经搜集到的一个或多个样本数据，根据样本数据的特征，在总体分布形式已知或未知的情况下，通过特定的方法，推断总体参数或对总体参

数进行判断的基本思路和详细分析流程。

第4部分——相关与回归分析。该部分内容主要是第6章，主要研究两个或多个变量之间的相互依存或影响的统计关系，具体内容包括两个变量之间的相关分析、两组变量之间的典型相关分析以及经典回归分析。本章是进行数据统计建模的基础，也是数据分析工作的核心内容之一。

第5部分——多元统计分析。该部分内容包括第7~9章，具体内容涵盖了因子分析、主成分分析、聚类分析、判别分析、列联表分析和对应分析。本部分内容对多元统计分析中各种常用的方法和原理进行了具体阐述，是对数据分析工作的进一步深入介绍。

第6部分——微观计量分析。该部分内容主要是第10章，讲解微观计量分析中的常见离散因变量模型，具体内容包括概率线性模型、二元选择模型、多重选择模型及计数模型。这部分内容也是日常问卷数据分析的重点。

第7部分——时间序列分析。该部分内容主要为第11章，着重讲解时间序列的平稳性以及利用Box-Jenkins法进行时间序列建模的模型识别、估计及预测等问题。

本书特色

- ▶ 本书是笔者多年一线数据分析工作的经验总结和倾情奉献，所有分析案例均来源于作者第一手的调查数据。
- ▶ 使用SAS 9.1.3 Service Pack 4最新版本作为分析工具，其程序和菜单操作过程同样适用于较低版本。
- ▶ 数据分析方法齐全，涵盖了描述统计分析、统计推断、多元统计分析、微观计量分析、时间序列分析等常用方法。
- ▶ 以说理而非公式推导的方式进行原理和理论讲解。
- ▶ 全程真实数据案例引导学习，数据分析报告顺理成章、水到渠成。
- ▶ 数据文件+录像示例，可对照录像进行手把手的实践，学习效率高。

本书约定

本书的插图和运行结果可能会与读者实例环境中的操作界面或结果略有差别，这可能是由于操作系统平台、SAS版本不同而导致的，在此特别说明，一切以实际环境为准。

致谢

本书由首都经济贸易大学统计学院阮敬老师编著，成都易为科技有限责任公司审校，参与编辑、排版、校对的同志有：王斌、黄中林、张强林、王晓、夏惠军、余松、江广顺、姚新军。在本书编写过程中，还得到了首都经贸大学统计学院纪宏教授、台湾辅仁大学谢邦昌教授、清华大学朱世武教授的大力支持和帮助，我的师妹刘欢同学提供了部分章节的素材。在此要特别感谢家人对我创作的大力支持，没有他们的支持，本书不可能这么快速地和读者见面。

由于时间有限，书中不足之处在所难免，恳请读者批评指正（电子函件：book_better@sina.com）。

作者

2009.1

目 录

第 1 章 数据预处理	1
1.1 SAS 环境与操作界面	1
1.2 SAS 编程基础	3
1.2.1 SAS 编程语言的基本结构	3
1.2.2 SAS 结构化编程语句	5
1.3 SAS 的数据处理对象	8
1.3.1 SAS 数据库和 SAS 数据集	8
1.3.2 SAS 系统的外部数据文件	15
1.4 数据预处理原理和基本方法	17
1.4.1 数据整理	19
1.4.2 数据分拆与合并	22
1.4.3 数据清洗	24
1.4.4 数据变换	27
1.5 本章小结	30
第 2 章 数据的描述	31
2.1 统计图	31
2.1.1 直方图	31
2.1.2 条形图	34
2.1.3 线图	36
2.1.4 散点图	37
2.1.5 饼图	39
2.1.6 盒式图	40
2.1.7 茎叶图	41
2.2 统计量	42
2.2.1 集中趋势	42
2.2.2 离散程度	45
2.2.3 分布形状	48
2.2.4 利用菜单和程序进行详细的描述统计分析	50
2.3 统计表	55
2.3.1 统计表的基本要素	56
2.3.2 用 TABULATE 过程绘制统计表	56
2.4 数据分布	58
2.4.1 总体分布	59

2.4.2 样本分布	59
2.4.3 抽样分布	59
2.5 本章小结	61
第3章 简单统计推断	62
3.1 简单统计推断的基本原理	62
3.1.1 参数估计	63
3.1.2 假设检验	64
3.2 单总体参数的估计及假设检验	68
3.2.1 单总体的参数估计	68
3.2.2 单总体参数的假设检验	71
3.3 两总体参数的估计及假设检验	82
3.3.1 独立样本的参数估计和检验	82
3.3.2 成对样本的参数估计和检验	91
3.4 本章小结	95
第4章 方差分析	96
4.1 方差分析的基本原理	96
4.2 单因素方差分析	99
4.2.1 单因素方差分析与方差同质性检验	99
4.2.2 方差分析的多重比较	102
4.2.3 方差分析模型的参数估计和预测	104
4.3 多因素方差分析	109
4.3.1 只考虑主效应的多因素方差分析	110
4.3.2 存在交互效应的多因素方差分析	116
4.4 协方差分析	118
4.5 本章小结	122
第5章 非参数检验	123
5.1 非参数检验的基本问题	123
5.2 单样本非参数检验	124
5.2.1 单样本均值的 Wilcoxon 符号秩检验	124
5.2.2 单样本的 Kolmogorov-Smirnov 检验	125
5.3 两个样本的非参数检验	128
5.3.1 两个独立样本中位数比较的 Wilcoxon 秩和检验	128
5.3.2 两个独立样本分布的 Kolmogorov-Smirnov 检验	132
5.3.3 成对样本中位数的 Wilcoxon 符号秩检验	134
5.4 多个样本的非参数检验	136
5.4.1 多个独立样本位置的 Kruskal-Wallis 检验	136
5.4.2 多个独立样本位置的 Jonckheere-Terpstra 检验	138
5.4.3 多个独立样本中位数的 Brown-Mood 检验	139
5.5 本章小结	139
第6章 相关与回归分析	141

6.1 相关分析	141
6.1.1 简单相关分析	142
6.1.2 偏相关分析	146
6.1.3 等级相关分析	147
6.2 典型相关分析	151
6.2.1 典型相关分析基本原理	151
6.2.2 典型相关系数的显著性检验	155
6.2.3 典型相关的冗余分析	156
6.3 线性回归分析	158
6.3.1 回归分析的基本原理	158
6.3.2 一元线性回归分析	161
6.3.3 多元线性回归分析	168
6.4 定性自变量回归分析	172
6.4.1 虚拟变量的设定	172
6.4.2 含有虚拟变量的回归分析	173
6.5 本章小结	174
第7章 因子分析	175
7.1 数据降维	175
7.1.1 数据降维的基本问题	175
7.1.2 数据降维的基本原理	176
7.2 主成分分析	177
7.2.1 主成分分析的基本概念与原理	177
7.2.2 主成分分析的基本步骤和过程	178
7.3 因子分析	184
7.3.1 因子分析的基本原理	185
7.3.2 因子分析的基本步骤和过程	186
7.4 本章小结	194
第8章 聚类分析与判别分析	195
8.1 聚类分析的基本原理	195
8.1.1 分类的基本原则	195
8.1.2 单一指标的系统聚类过程	197
8.1.3 多指标的系统聚类过程	198
8.2 聚类分析的步骤和过程	202
8.2.1 系统聚类	202
8.2.2 快速聚类	210
8.2.3 变量聚类	212
8.3 判别分析的基本原理	215
8.4 判别分析的步骤和过程	216
8.4.1 距离判别	217
8.4.2 Bayes 判别	217
8.4.3 Fisher 判别	224

8.4.4 逐步判别	227
8.5 本章小结	231
第9章 列联分析与对应分析	233
9.1 列联分析	233
9.1.1 列联表	233
9.1.2 列联表的分布	237
9.1.3 χ^2 分布与 χ^2 检验	238
9.1.4 列联表中的关联度分析	240
9.1.5 χ^2 分布的期望值准则	241
9.2 对应分析	242
9.2.1 对应分析的基本思想	242
9.2.2 对应分析的步骤和过程	243
9.3 本章小结	249
第10章 离散因变量模型	250
10.1 线性概率模型	250
10.2 二元选择模型	251
10.2.1 线性概率模型的缺陷与改进	252
10.2.2 二元选择模型的基本原理	252
10.2.3 BINARY PROBIT 模型	253
10.2.4 BINARY LOGIT 模型	263
10.3 多重选择模型	269
10.3.1 多重选择模型的基本原理	269
10.3.2 ORDINAL PROBIT 模型	271
10.3.3 ORDINAL LOGIT 模型	276
10.3.4 MULTINOMIAL LOGIT 模型	279
10.4 计数模型	280
10.4.1 POISSON 回归模型的基本原理	281
10.4.2 POISSON 回归模型的分析过程和步骤	281
10.5 本章小结	285
第11章 时间序列分析	286
11.1 时间序列的基本问题	286
11.1.1 时间序列的组成部分	286
11.1.2 时间序列的平稳性	288
11.2 ARIMA 模型的分析过程	291
11.2.1 ARIMA 模型	291
11.2.2 ARMA 模型的识别、估计与预测	292
11.2.3 利用 SAS 时间序列预测系统进行菜单操作	301
11.3 本章小结	308

第1章

数据预处理

数据预处理是统计分析过程中不可缺少的重要内容，是定量分析研究的基础，也是进行社会科学研究的的基本前提。本章将结合实际定量分析工作中的常见问题，从 SAS 系统的基本界面、操作方式及编程基础等方面对数据预处理的方法进行详细介绍，并通过 SAS 软件的实际操作和真实数据展现一个完整的数据预处理过程。

1.1 SAS 环境与操作界面

SAS 系统安装完毕之后，进入“开始”菜单，选择“The SAS System 9.1.3”，或者双击  图标，进入 SAS 系统的主界面，如图 1-1 所示。

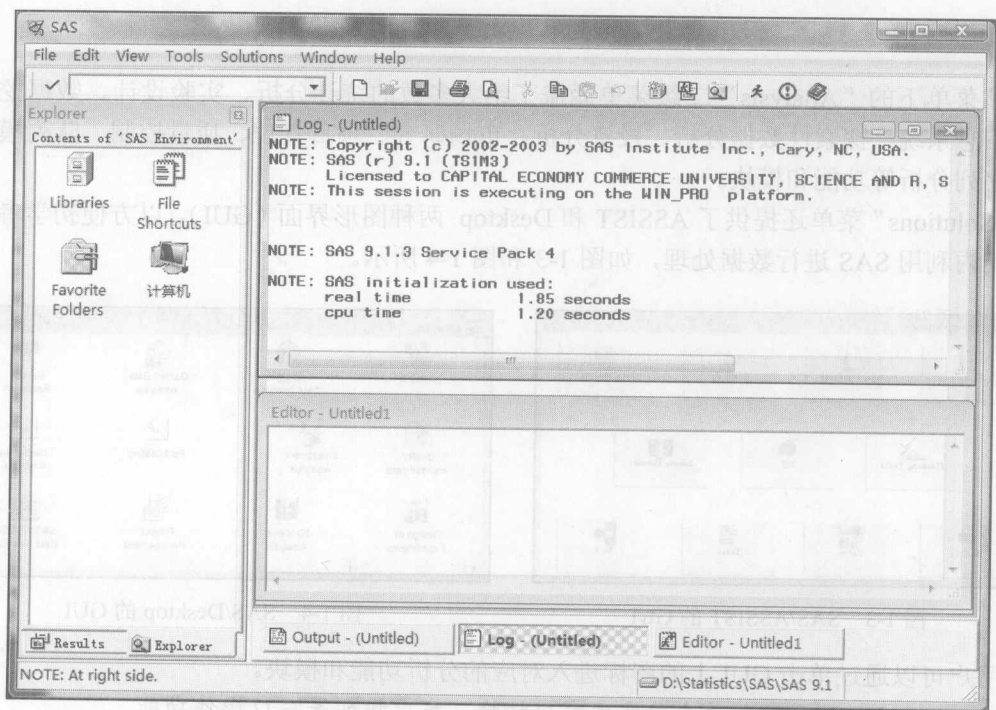


图 1-1 SAS 9.1.3 主界面

在 SAS 主界面中，处于最上面的系统菜单可以实现 SAS 文件和数据库的操作、编辑、视图、分析、作图等一系列功能。系统菜单会因所调用的 SAS 模块不同而有所不同。

- “File” 菜单主要实现 SAS 系统文件的基本操作，包括新建、打开、存储、打印文件，以及导入和导出外部数据文件、利用 E-mail 传送文件等。

- “Edit” 菜单主要实现对文本进行选择、查找、替换、复制、剪切、粘贴等编辑功能。
- “View” 菜单的功能主要是进行窗口切换或打开对应的窗口。
- “Tools” 菜单可以打开数据库查询与管理器，以及表格、图形、报告、文本编辑器，并且可以对 SAS 系统的全局参数、界面、字体、颜色等方案进行调整。
- “Solutions” 菜单是利用 SAS 进行数据分析、程序开发的最主要的菜单，如图 1-2 所示。

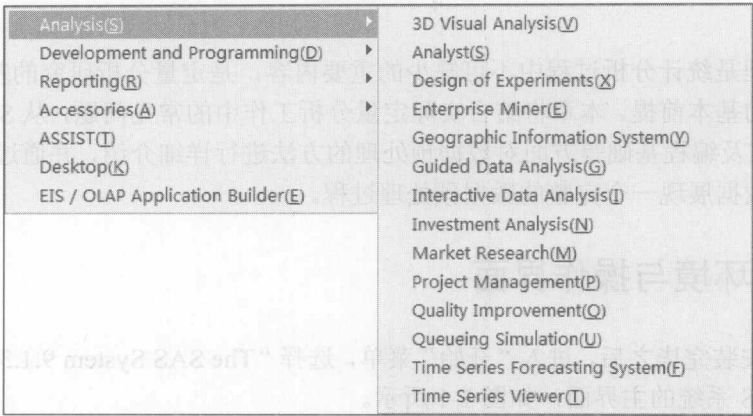


图 1-2 SAS 系统的“Solutions”菜单

该菜单下的“Analysis”二级菜单涵盖了绝大多数的统计分析、实验设计、数据挖掘、地理信息系统、探索性数据分析、投资分析、市场研究、项目管理、质量控制、队列模拟、时间序列分析等功能和模块。

“Solutions”菜单还提供了 ASSIST 和 Desktop 两种图形界面（GUI），以方便初学者在较短时间内利用 SAS 进行数据处理，如图 1-3 和图 1-4 所示。

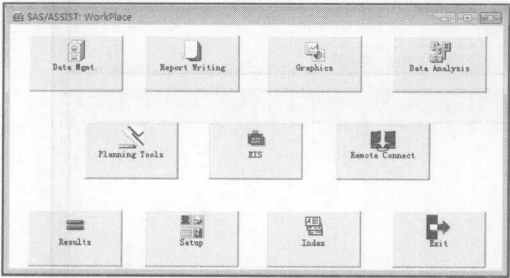


图 1-3 SAS/ASSIST 的 GUI

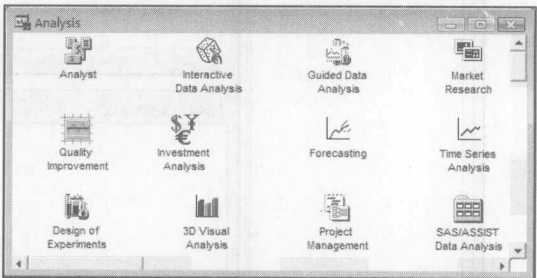


图 1-4 SAS/Desktop 的 GUI

用户可以通过单击 GUI 上的图标进入对应的分析功能和模块。

- “Window” 系统菜单主要实现窗口切换、窗口排列等窗口操作功能。
- “Help” 菜单是 SAS 系统的强大帮助系统，提供了从菜单操作到编程语言一系列的帮助功能，同时在任意对 SAS 系统进行分析的过程中，只要单击工具栏上的  图标，便可以快速地转到该过程对应的帮助信息上。

系统菜单下面是 SAS 工具栏。与大多数软件一样，该工具栏提供了能够实现系统菜单中的常用功能的快捷方式，并且以 Windows 操作系统中的常用图标表示，如图 1-5 所示。

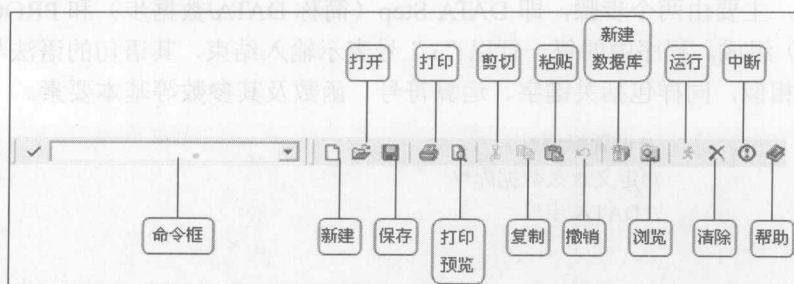


图 1-5 SAS 系统的工具栏

SAS 工具栏的特殊之处在于，工具栏的左边有一个文本输入框（称为文本框），用户可以在该框中输入命令或者调用 SAS 系统的各种模块，故称之为命令框。单击命令框前的✓按钮即可向 SAS 系统提交命令，其功能等同于按键盘上的 Enter（回车）键；单击命令框后的▾按钮，则可从中选择以前使用 SAS 系统时输入的历史命令。

处于主界面最左边的是“Results”（结果）窗口和“Explorer”（浏览）窗口，这两个窗口处于同一界面中，可以通过单击最下边的对应按钮进行切换。

- “Explorer”窗口类似于 Windows 系统下的资源管理器，可以实现对 SAS 数据库、数据文件、程序及其他文件的查看、打开、新建、删除等基本操作，甚至可以直接运行 SAS 系统之外的其他软件。

- “Results”窗口可以以目录或条目的形式提供分析结果的浏览功能。

处于主界面右边的分别是“Output”、“Log”和“Editor”窗口，同样可通过单击最下边的对应按钮进行切换。

- “Output”窗口主要呈现利用 SAS 系统进行分析的结果，按 F7 键可切换至该窗口。
- “Log”窗口是日志窗口，按 F6 键可以切换至该窗口。用户对 SAS 的每一步操作都会以日志的形式列示在“Log”窗口当中（如每次打开 SAS 系统时，该窗口中默认显示一些授权文件的信息）。同时，在进行数据分析的过程中，一些出错信息和程序执行的反馈信息也会出现在该窗口中。
- “Editor”窗口是一个文本编辑器，按 F5 键可以切换至该窗口。用户可以在该窗口中进行编程或输入文字信息（在“Log”和“Output”窗口中，用户不能自行写入信息）。对于这 3 个窗口中的信息，用户可以通过系统菜单或工具栏上的按钮进行复制、新建、存储、打印等操作。

可以根据用户的需要用两种方式关闭和打开这些默认的窗口。一种方式是在命令框中输入对应窗口的名字，另一种方式是在“View”系统菜单中单击对应的窗口名字即可。

1.2 SAS 编程基础

SAS 系统不仅可以使使用菜单方式进行数据操作和统计分析，而且还具备强大的编程语言功能，供用户灵活调用各种模块以及调整各种分析参数。

1.2.1 SAS 编程语言的基本结构

在 SAS 系统中，可以利用“Editor”或“Program Editor”窗口书写程序。SAS 编程语言

结构比较简单，主要由两个步骤，即 DATA Step（简称 DATA/数据步）和 PROC Step（简称 PROC/过程步）组成。程序中的每一行以“;”号表示输入结束，其语句的语法与常见的高级语言语法大体相似，同样包括关键字、运算符、函数及其参数等基本要素。

```
title "...";           /*设置标题*/
libname ...;          /*定义永久数据库*/
data ...;             /*DATA 步*/
    ...;
    ...;
run;
proc ...;             /*PROC 步*/
run;
```

在 SAS 语言中，通常利用“/*.....*/”表示对程序的注释，DATA 步和 PROC 步之间用“run;”或者直接用“;”隔开。通常，一些全局变量的设置语句可以放在 DATA 步之前，如为输入表格或图形定义表头或标题、建立 SAS 数据库等。

1. DATA 步（数据步）

DATA 步是 SAS 进行数据管理和操作的基本步骤，其主要功能包括：建立 SAS 数据集，导入外部程序数据文件，分割、合并、修改、更新现有的 SAS 数据集，分析、呈现和管理数据，利用数据集中已有的数据计算或生成新变量等。DATA 步中常用的 SAS 编程语句如下。

- **infile 语句**：从外部文件获取数据。如要使用该语句，必须把其放在其他 data 语句之前，其主要语法如下所示。

```
infile “外部数据路径及文件名” <选项>;
```

如从 D:\student.txt 文件中获取数据，可利用以下 infile 语句。

```
infile “D:\student.txt”;
```

- **input 语句**：指定读入数据的格式以及为读入的数据指定变量名及格式，其语法如下所示。

```
input <变量名 1 变量名 2 ... 变量名 n> <选项>;
```

如从上例读入的外部数据中读入两个变量的数据，并分别命名为 height、weight。

```
input height weight;
```

- **cards 语句**：用于在 SAS 系统中直接输入数据，表明所列示数据的开始。

对于 DATA 步编程的详细内容，将在下节详细讲解。

2. PROC 步（过程步）

SAS 系统的过程步可引用现有的程序或过程进行相应的数据处理和分析活动。其主要语法如下所示。

```
proc 过程名 <data=数据库名.数据集名> <选项>;
    <var <变量名 1 变量名 2 ... 变量名 n>>;
    <where <条件或表达式>>;
    <by <变量名 1 变量名 2 ... 变量名 n>>;
```

run;

其中的语句功能如下。

- data 语句表示该 PROC 步所处理的数据集。
- var 语句表示处理数据集中的特定变量。对于没有列示的变量，系统则不予处理。
- where 语句表示指定系统处理符合一定条件或表达式的样本。
- by 语句表示指定系统按照所列示的变量进行分组处理。但是要注意的是，使用该语句时，必须先对该语句中指定的分类变量进行排序。

表 1-1 列出了最常用的几种 PROC 步过程。

表 1-1 常见 PROC 步过程的作用及功能

过 程 名	作 用	输 出 结 果
PRINT	显示数据集的变量名及变量值	变量和变量值
SORT	对指定变量进行排序	可对指定变量进行升序、降序排列
MEANS	对数值型变量进行描述统计分析	均值、标准差、极值等统计量
UNIVARIATE	对数值型变量进行描述统计分析	常见统计量、t 检验、分位数、极端值等
FREQ	对定序变量进行描述统计分析	频数、频率、累计频数、累计频率等
CHART	对指定变量绘制文本形式的图形	饼图、横向/纵向直方图、星形图
GCHART	在“Graph”窗口中对指定变量绘制图形	饼图、横向/纵向直方图、星形图

以上输出结果均可在“Output”窗口中显示，并且在“Results”窗口中列示对应结果的标签。双击标签，便会在“Output”窗口中显示对应的具体内容。

PROC 步所能调用的程序或模块非常多。本书出于应用的目的，在后续各章节中就对应的过程或模块进行详细讲解。

3. SAS 编程语言的表达式

SAS 的常用表达式主要有运算表达式和逻辑表达式。SAS 的运算符主要有+(加)、-(减)、*(乘)、/(除)、** (乘方)等；SAS 的逻辑符号主要有=(等于)、<(小于)、>(大于)、<=(小于等于)、>=(大于等于)、<>(不等于)、and (和)、or (或)、xor (异或)，这些符号也可以用英文字母等价表示，如表 1-2 所示。

表 1-2 SAS 逻辑运算符的表达方式

符 号	=	<	>	<=	>=	<>
英 文	eq	lt	gt	le	gt	ne

1.2.2 SAS 结构化编程语句

SAS 的结构化编程语句主要有顺序语句、条件语句和循环语句。这 3 种基本形式的语句均可在 DATA 步和 PROC 步中使用。顺序语句是最为常见的语句形式，系统按照语句自身顺序进行解释形式的执行。下面简要介绍条件语句和循环语句的基本使用方法。

1. 条件语句

条件语句能够使程序按照一定的表达式或条件实现不同的操作或执行顺序跳转的功能，其语法如下所示。

```
if 条件或表达式 then
    ...;                                /*当条件或表达式满足时执行的程序*/
else
    ...;                                /*当条件或表达式不满足时执行的程序*/
```

此外，条件语句还可以进行嵌套。



例 1-1

比较 x 和 y 两个变量的大小。如果 $x > y$ ，则输出 “ $x > y$ ”；如果 $x < y$ ，则输出 “ $x < y$ ”；如果 $x = y$ ，则输出 “ $x = y$ ”。

现假定 x 赋值为 10， y 赋值为 20，在 Editor 窗口中输入以下程序。

```
data;
x=10;y=20;
if x>y then
    put "x>y";        /*在“Log”窗口中显示引号内的字符*/
else
    if x<y then
        put "x<y";
    else
        put "x=y";
run;
```

单击工具栏上的 （运行）按钮，则在“Log”窗口中输出结果 “ $x < y$ ”。

2. 循环语句

循环语句可以使 SAS 系统循环执行一定的程序，主要有计数（DO）循环、当（WHILE）循环、直到循环（UNTIL）3 种形式。

（1）计数循环。

计数循环的主要表达方式如下所示：

```
do 计数变量=初始值 to 终止值 by 步长;
    ...;
end;
```

如果 by 步长省略，则表示计数变量按照默认步长 1 计数。步长也可以是负数，此时计数变量初始值需大于终止值。



例 1-2

计算 1~100 之内的所有奇数自然数之和。具体程序如下所示。

方法一：

```
data;
    y=0;
do x=1 to 99 by 2;
```

```

y=y+x;
end;
put "y=" y;
run;

```

方法二:

```

data;
y=0;
do x=99 to 1 by -2;
y=y+x;
end;
put "y=" y;
run;

```

单击工具栏上的  按钮, 则上述两种方法均会在“Log”窗口中显示结果“y=2 500”。

(2) 当循环。

当循环的主要表达式如下所示:

```

do while (继续循环条件表达式);
...;
end;

```

该语句执行时会首先判断条件表达式是否成立。如果成立, 则系统执行 DO WHILE 中的语句, 遇到 END 时返回条件表达式的判断。如此重复, 直到条件表达式不能够满足为止。

如在例 1-2 中, 同样也可以利用当循环进行计算, 程序如下所示:

```

data;
x=1;
y=0;
do while (x<100);
y=y+x;
x=x+2;
end;
put "y=" y;
run;

```

程序运行结果同样显示“y=2 500”。

(3) 直到循环。

直到循环的主要表达式如下所示:

```

do until (退出循环条件表达式);
...;
end;

```

该语句会首先执行循环语句内部的程序, 然后判断条件是否成立。如果成立, 则退出循环过程, 否则继续执行循环语句内部的程序。

仍旧以例 1-2 为例, 采用直到循环的程序如下所示:

```

data;
x=1; y=0;
do until (x>100);

```

```

y=y+x; x=x+2;
end;
put "y=" y;
run;

```

循环语句均可以使用条件语句跳出循环过程。仍以例 1-2 为例，现计算 50 以内的奇数自然数之和。除了修改以上各种循环语句的条件之外，还可以在不修改循环条件的前提下进行计算，具体程序如下所示：

```

data;
x=1;
y=0;
do until (x>100);
y=y+x;
x=x+2;
if x>50 then leave;
end;
put "y=" y;
run;

```

1.3 SAS 的数据处理对象

要进行数据预处理，并进一步进行统计分析，必须先弄清利用 SAS 软件包进行统计分析的对象是什么。本节将从两个方面对该问题进行阐述。

1.3.1 SAS 数据库和 SAS 数据集

SAS 的数据对象是存在于 SAS 数据库中的数据集。具体而言，数据集是 SAS 的数据处理对象，同时也是数据分析的基础。

1. SAS 数据库

SAS 数据库具体是指存放 SAS 数据文件（即数据集）的文件夹，它与计算机存储器中的某一个具体文件夹相对应。

（1）SAS 数据库的分类。

为了让 SAS 系统识别文件夹所对应的数据库，要为每一个数据库指定一个库标记（库名）以识别该库。库标记是逻辑存在的，只存在于 SAS 系统中，同一个文件夹可以对应不同的逻辑库标记，就像一个人可以有很多个称呼或绰号一样。根据数据处理的时效性不同，SAS 数据库又可以分为临时库和永久库。

- 临时库：只有一个，名为 WORK。在每次启动 SAS 时由系统自动生成。关闭 SAS 时，该数据库中的所有数据文件自动被清除。
- 永久库：可有多个。用户可以自己指定永久库的库标记，库中的所有数据文件永久保留。但库标记是临时的，每次启动 SAS 系统都要重新指定。

在每次启动 SAS 时，系统都会根据用户安装 SAS 时授权文件的模块自动指定若干个库标记，其中有 3 个库标记是不可缺少的。