



SAS

程序设计

王惠群 李韵华 编著



科学出版社



文魁资讯股份有限公司

SAS 程序设计

王惠群 李韵华 编著

科学出版社

2002

内 容 简 介

SAS 是一种适用于企业资讯发展的整合性的套装软件,它具有对不同数据的强大的读取与整合能力,主要用在数据读取、数据整理、数据分析和展示结果等工作上。

全书共 10 章,前 5 章讲解基础知识,后 5 章讲解高级应用。主要介绍了 SAS 的常用语法、函数、程序,从数据的输入、管理、分析、展示,到使用各种数据库,再扩展到 Web 的应用上。

本书内容编排合理,语言生动、流畅,适用于从事市场调查、分析工作的相关人员阅读。

本书繁体字版原书名为《SAS 程式设计》,由文魁资讯股份有限公司出版,版权属王惠群、李韵华所有。本书简体字中文版由文魁资讯股份有限公司授权科学出版社独家出版。未经本书原版出版者和本书出版者书面许可,任何单位和个人不得以任何形式或任何手段复制或传播本书的部分或全部。

版权所有,翻印必究。

图字:01-2002-0311 号

图书在版编目(CIP)数据

SAS 程序设计/王惠群,李韵华编著.——北京:科学出版社,2002

ISBN 7-03-010201-0

I.S... II.①王... ②李... III.统计分析—应用软件,SAS—程序设计 IV.C812

中国版本图书馆 CIP 数据核字(2002)第 011797 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002 年 3 月第 一 版 开本:720×1000 1/16

2002 年 3 月第一次印刷 印张:17 1/2

印数:1—4 000 字数:346 000

定价:29.00 元

(如有印装质量问题,我社负责调换〈路通〉)

序

从 1979 年作者第一次参加工作，便开始使用 SAS。不管在大型主机、UNIX 还是 PC 上，SAS 帮助作者完成了许多工作。在使用中，作者发现，市面上介绍 SAS 的图书并不多，且仅有的几种也多半是介绍 SAS 统计模组的功能，在实用方面涉及的不多，因此，想把自己多年所学整理出来，与所有对 SAS 有兴趣的人共同研究。

本书是适合于初级和中级用户使用的一本书。前 5 章讲解基础知识，后 5 章讲解高阶应用。以下为各章的内容介绍，希望能帮助用户学习使用 SAS。

第 1 章：介绍 SAS 的设置、使用，以及 SAS 操作环境中的各个窗口的功能。

第 2 章：介绍 SAS 的基本语法、常用程序及函数。根据作者多年使用 SAS 的经验，将使用过的程序语法、函数和程序运用在浅显易懂的范例中，并介绍了各种语法的使用。

第 3 章：介绍如何使用 SAS 获取各式各样的数据。数据可能存在于 SAS 数据文件、文字文件、Excel 文件和复杂的文字文件中。

第 4 章：介绍如何整理读取后的 SAS 数据文件以产生其他的数据，使用条件判断、循环处理和阵列，皆可帮助用户完成各种可能的数据整理。

第 5 章：介绍各种分析报表的产出、文字文件的建立、地址栏格式的报表，使 SAS 在报表的应用上能更多样化。

第 6 章：学习完此章，在分析时，对于数据的分类方式能简单地使用用户自定义的格式完成各种不同的分类规则，不用修改数据内容或建立其他变量来进行分类，更不必用很长的判断条件来进行分类的工作。

第 7 章：介绍在 SAS 中使用 SQL 的方法。本章将使用户能更加简洁地使用 SQL 来完成数据的读取和分析。

第 8 章：介绍 SAS MACRO 的使用。在编写 SAS 程序代码时，加入一些 MACRO 的协助，使 SAS 程序更加强而有力，在简洁中富于变化。

第 9 章：介绍作者在 SAS 的各种高级技巧上的使用状况，包括 SAS 设置、主从架构、数据库的使用、MACRO 的使用，以及在更新版本的 SAS 中如何来设置及完成以上的工作。

第 10 章：本章介绍串联 SAS 的分析报表产出到 Web 环境，用户可以很容易地将产出的分析报表完整展现在 Web 环境中。

由于作者水平有限，书中难免有错误疏漏之处，恳请专家、同行和广大读者批评指正。

作者

目 录

第 1 章 基本介绍	1
1-1 关于 SAS	1
1-2 开始执行	4
1-3 操作环境	5
习题	8
第 2 章 程序语法	9
2-1 程序架构	9
2-2 数据类型	10
2-3 数据文件夹	11
2-4 数据步骤	14
2-5 程序步骤	15
2-6 常用语法	16
2-7 常用函数	43
2-8 常用 PROC	62
习题	83
第 3 章 数据输入	84
3-1 读取卡片	84
3-2 读文字文件方式	89
3-3 读取 SAS 数据文件	95
3-4 读取 Excel 文件	98
3-5 读取 DBF 文件	104
3-6 读取复杂文字文件	105
习题	114
第 4 章 数据整理	115
4-1 变量类型的转换	115
4-2 循环处理	121
4-3 阵列处理	127
4-4 排序	130
4-5 合并	136
习题	143
第 5 章 产出报表	144
5-1 FILE	144
5-2 FORMS	145

5-3	FREQ	148
5-4	MEANS.....	151
5-5	PRINT	156
5-6	REPORT	161
5-7	TABULATE.....	163
	习题.....	172
第 6 章	格式应用	173
6-1	基本介绍.....	173
6-2	用户自定义格式.....	175
6-3	日期格式.....	178
6-4	格式在打印报表的功能.....	180
	习题.....	183
第 7 章	SQL 语法.....	184
7-1	基本介绍.....	184
7-2	数据加总.....	190
7-3	数据合并.....	194
7-4	子查询.....	198
	习题.....	199
第 8 章	MACRO 应用.....	200
8-1	基本概念.....	200
8-2	使用者自定义 MACRO 变量	202
8-3	使用者自定义 MACRO 程序	204
8-4	MACRO 程序	211
8-5	动态 MACRO 变量.....	212
	习题.....	215
第 9 章	高级应用	216
9-1	Client-Server 架构	216
9-2	如何使用 SAS/Access ODBC 读取数据	219
9-3	自动化的工作.....	222
	习题.....	236
第 10 章	Web Solution.....	237
10-1	使用 SAS Intranet 产品架构	237
10-2	产出用户的 HTML.....	238
10-3	输出转成 HTML.....	242
10-4	数据文件转换成 HTML.....	244

10-5	TABULATE 输出转换成 HTML	247
10-6	GRAPH 圆形输出转换成 GIF	250
10-7	构建企业成功指标系统	253
习题		269

第 1 章 基本介绍

1-1 关于 SAS

1-1-1 基本概念

SAS 是适用于企业信息发展的整合性软件包，在数据运用上有数据读取、数据整理、数据分析和展示结果等 4 项功能。

利用 SAS 不仅便于数据的读取，而且可以让用户在各平台间使用同一种语言，使用户的操作得心应手。

企业内部信息的传递，有可能因不同的数据库、不同的应用软件、不同等级的用户以及不同的计算机平台，造成传递过程不流畅。利用 SAS 系统提供的一系列策略及技巧可以解决这些问题。

1. 不同数据库：全方位的数据选取

企业数据库的复杂程度根据企业建立系统的多寡而定。无论是企业的数据存储方式还是存储位置，SAS 都提供全方位的数据选取策略以帮助用户进行及时的数据选取，并进一步为企业提供最恰当的决策。

SAS 系统对不同的数据库可以完全地变向沟通，如 Mainframe 上的 IMS/DB、IMS/DLI、VSAM，关系型数据库 DB2、ORACLE、SYBASE、INFORMIX、INGRES、DEC RDB、ODBC、DIF、DBF、WK_n 和 XLS 等。

SAS 系统对不同的数据库有强大的读取和整合能力，使 SAS 系统成为企业完成最佳数据库的统合者，而且是协助完成企业构建数据存储的最佳工具。

2. 不同应用软件：应用软件整合

不同应用软件的使用造成软件间无法整合，也无法有效发展整体的信息力量。SAS 系统提供了功能强大且非常实用的应用模组，来帮助企业完成数据选取、数据管理、数据分析、数据展示的整体应用系统，从而解决不同应用系统的需求。

3. 不同等级的用户：多样的用户接口

在企业中从一般职员到高级管理者，不同等级的用户有不同的需求，对计算

机技术的要求也不尽相同。针对本身不同的工作需求而选择不同的工具，会造成企业软件资源、训练资源等方面成本的增加。SAS 系统提供了多样的用户界面供用户针对需求进行最佳的整合，如主管信息系统、用户对象导向式的用户界面和程序设计的整合界面等。

4. 不同的计算机平台：跨不同平台的技巧

企业在不同的发展阶段，因不同的需求，对不同硬件平台的需求也不尽相同。为了完成企业间信息传递的通行无阻，SAS 的主从架构技巧可协助企业在不同的平台发展 SAS 应用系统。

1-1-2 模组化

SAS 为模组化的产品，产品相关介绍请参考 SAS 产品介绍网站 <http://www.sas.com/products/index.html>。

SAS 的主要核心产品为 SAS/BASE，针对用户的不同需求，可以搭配所需的产品。例如，要开发 PC 跨平台的 SAS 应用系统，在 PC 端应有 BASE、AF 和 CONNECT，主机端应有 BASE 和 CONNECT。

要使用其他产品功能(如 Graph)，必须有该产品(SAS/Graph)的使用权限。作者将在本书中介绍 BASE 的使用方法。

1-1-3 跨平台

SAS 可安装在不同的平台，其不同平台的语法皆相同，可在不同的平台开发或在主从架构下完成系统的运作。

SAS 在各平台有不同的执行方式。总体来说，SAS 可有批处理模式和交互模式两种执行方式。

1. 批处理模式(Batch Mode)

将编写的 SAS 程序代码存放在各硬件环境中，在不同的操作环境下可将所有程序一起批处理完成，说明如下：

UNIX 环境

SAS 安装完成时应已建立执行 SAS 的批处理执行 UNIX 脚本，作者在 UNIX 上所使用的脚本名称为 `sasbatch`。Sasbatch 的内容如下：

```
$ more /usr/bin/sasbatch  
TERMINFO=/opt/sas/sas612/terminfo; export TERMINFO
```

```
/opt/sas/sas612/sas -config /opt/sas/sas612/config.batch $1  
$2 $3 $4 $5
```

在 UNIX 的批处理模式执行方式是:

```
$sasbatch main.sas
```

MVS 环境

以下是作者使用 JCL 执行 SAS 的程序代码:

```
//SAS001 JOB T001-SA-1-1111,CLASS=A,TIME=1440,  
// MSGCLASS=N,MSGLEVEL=(1,0),REGION=4M  
//*  
//STEP1 EXEC SAS609,REGION=8M  
//WORK DD UNIT=SYSDA,SPACE=(CYL,(10,5))  
//FOUT DD DSN=ISMRCHJEFF.FOUT1,DISP=OLD  
//SYSIN DD *  
OPTIONS OBS = MAX LS =90 ;  
FILENAME INP 'COMMON.ATM.PBDDTXM.KEEP' DISP=OLD ;  
FILENAME INPATM 'COMMON.YBVKLOG.SD' DISP=SHR ;  
DATA FOUT.AA_ATM ;  
INFILE INPATM ;  
INPUT @72 ACCTW $16. ;  
ACCT = INPUT (ACCTW ,16.) ;  
PROC PRINT DATA = FOUT.AA_ATM (OBS = 200);  
RUN;  
/*  
//
```

其中, DDNAME WORK 可取代 SAS609 Procedure 中所定的 WORK 空间。

SAS609 Procedure 中有以下重要的 DDNAME:

- CONFIG: 执行 SAS 时默认值设置在该文件中, 该文件是所有用户的默认值。若针对某开发系统或自行使用的环境, 可在 JCL 中指定使用自定义的文件。
- WORK: 存放临时 SAS 数据设置, 以及使用 SAS Proc 须用到的空间(如 PROC SORT)。
- SASLOG: 执行 SAS 后所有的 Log 皆放在该 Output 中。
- SASLIST: SAS Proc Output。

2. 交互模式(Interactive Mode)

直接在 SAS 交互模式开发环境下编写、执行和修改程序代码，再存储完成的程序代码，以便将来可将其读回再次执行或由批处理模式执行。

SAS 在以上环境下进入 SAS 系统时，便已设置好 SAS 系统会使用的空间或路径：

- **Sashelp:** 该数据文件夹中的文件是 SAS 系统在运作时会使用到的系统文件，对环境的数据设置都会存放在此。用户不要修改该数据文件夹中的文件数据，这些数据文件是 SAS 系统环境下必须存在该数据文件夹中的。
- **Sasuser:** SAS 系统会设置 Sasuser 数据文件夹路径(MVS 环境是一个文件)。若用户没有设置其他路径，则 Sasuser 指定的路径或文件每次都相同，该路径下的数据文件是永久存在的。如果有 SAS 数据文件要存储时，用户无需在自己的程序代码中设置任何的数据文件夹名称，便可以使用 Sasuser 数据文件夹来存储 SAS 数据文件。
- **Work:** 进入 SAS 系统时会设置 Work 数据文件夹路径(MVS 环境是 Work 空间)，设置的路径每次都不同。当 SAS 系统需要工作空间时便会在该路径下建立 SAS 系统所使用的文件，当用户产生暂时性的 SAS 数据文件时，也会产生在该路径下；当离开 SAS 系统时，SAS 系统便会将该路径清除，不会将数据存储下来。若 SAS 系统并非正常结束，则必须以手动的方式由用户自行删除该路径下的数据。

1-2 开始执行

SAS 的每一版本安装方式都不尽相同，故只对安装时的注意事项做一说明。

SAS 每一产品皆有其使用权限控制。使用权限应为一年的，每年签完新合同后，SAS 公司将给用户一设置权限程序，PC 机由用户将程序调入后执行，即可完成新使用权限的设置，UNIX 或大型主机皆由系统管理者完成权限设置的操作。

权限设置内容如下：

```
PROC SETINIT RELEASE='6.12';  
SITEINFO NAME='SAS INSTITUTE TAIWAN LTD. - TRIAL INSTALLATION'  
SITE=19909028 OSNAME='xxxxx RECREATE WARN=15 GRACE=0  
EXPIRE='31OCT1998'D PASSWORD=142034954;  
ALIAS 'INTRNET' 'NEWPD002' / PASSWORD=254300292 PRODNUM=245;  
CPU MODEL=' ' MODNUM=' ' SERIAL=' '  
EXPIRE 'BASE' 'GRAPH' 'ETS' 'FSP' 'AF' 'OR' 'IML' 'SHARE' 'QC' 'STAT'
```

```
'INSIGHT' 'TOOLKIT' 'ORACLE' 'ASSIST' 'CALC' 'CONNECT' 'INGRES'  
'CFE' 'CBT101' 'CBT102' 'CBT103' 'CBT104' 'CBT105' 'PH-CLINICAL'  
'SYBASE-SQL Server' 'LAB' 'ENGLISH' 'EIS' 'WAREHOUSE' 'CBT106'  
'PC File Formats' 'GIS' 'INFORMIX' 'SPECTRAVIEW' 'SHARE*NET'  
'MDDB Server' 'IT Service Vision Server' 'INTRNET' '31OCT1998'D;  
SAVE;  
SEC PASSWORD=XXXXXXXX;  
SAVE; RUN;
```

安装完成后即可进入如图 1-1 所示的 SAS 交互模式。

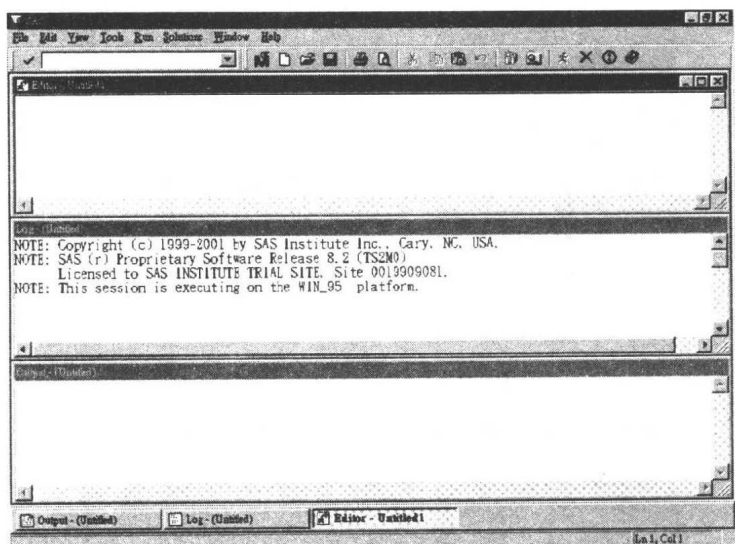


图 1-1 SAS 交互模式

1-3 操作环境

如图 1-1 所示，进入 SAS 交互模式后即可得到 3 个 SAS 主要窗口。

- **Editor 窗口：**是 SAS 第 8 版新增的窗口，可以进行编写、编辑和执行 SAS 程序。进入 SAS 后，即可在该窗口中编写和执行程序。要修改错误，可再执行测试，直到程序无误为止，存储最后结果以便下次执行。
- **Log 窗口：**程序执行时即可看到该窗口记录执行时的过程，其中包括程序错误信息、注意事项、每一程序使用的时间和 SAS 数据文件的数目。根据作者的经验，SAS 数据文件的数目常是判断的逻辑错误，或是 SAS 在某些 PROC 与自己的想法不同的重要依据。每次检查 SAS 数据文件的数

目都是必要的检查工作。

- **Output 窗口：**是程序执行后输出报表显示的窗口。

下面是几个常用窗口：

- **Libraries 窗口：**进入 SAS Explorer(图 1-2)窗口后，在该窗口左侧即可看到所建立的 Libraries 文件夹，单击该文件夹即可看到在该文件夹中的 SAS 数据文件的个数(图 1-3)。

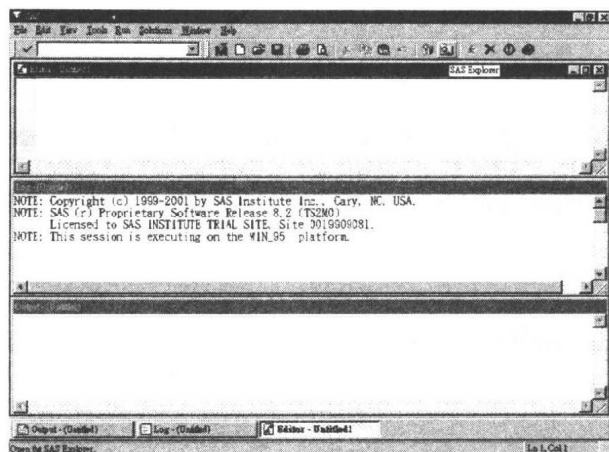


图 1-2 SAS Explorer 窗口

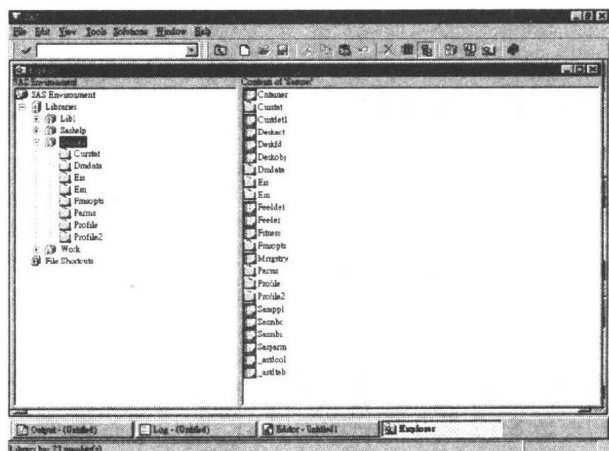


图 1-3 Libraries 窗口

在所选取的 SAS 数据文件上单击鼠标右键，在弹出的快捷菜单中选择 View Columns 命令，如图 1-4 所示。

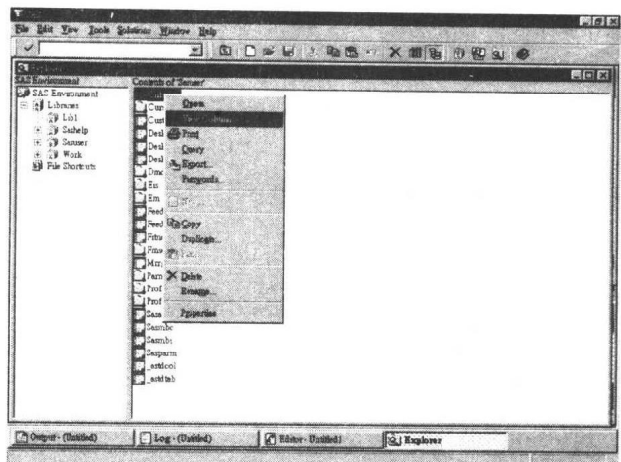


图 1-4 选择 View Column 命令

选择该命令后可看到该 SAS 数据文件变量的属性，包括该数据文件的变量名称、长度、格式、输入格式和标记。

可在本窗口中修改变量名称、格式、输入格式和标记，但无法修改长度，如图 1-5 所示。

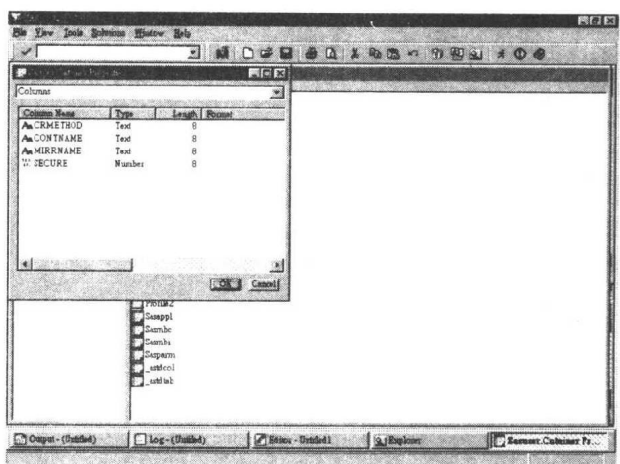


图 1-5 属性窗口

- **Keys 窗口：**在 SAS 命令下拉列表框输入“KEYS”或按 F9 键(图 1-6)，即可看到 KEYS 窗口（功能键窗口）(图 1-7)。在该窗口中可以自行建立常用的功能键，以方便程序的开发。作者喜欢将 F12 键设为“clear log;clear output”，这样可快速将其他两个窗口内容清除，以便继续执行程序，执行时的 Log 及 Output 窗口中的内容也不会混淆。

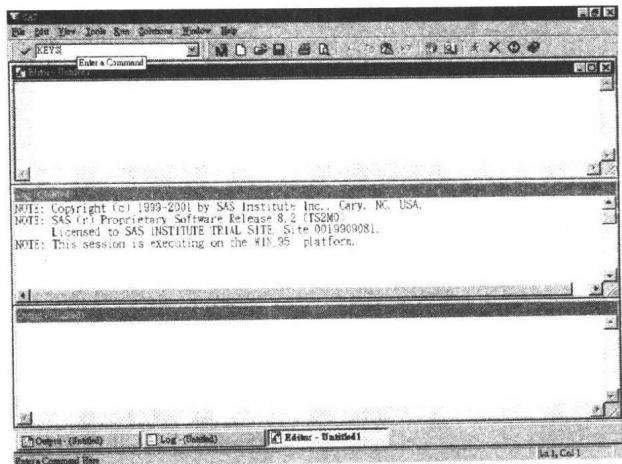


图 1-6 输入“KEYS”

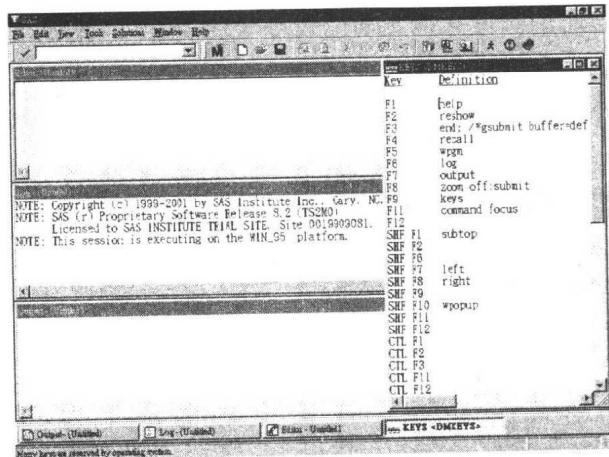


图 1-7 建立常用的功能键

习 题

1. 说明执行 SAS 程序时可用的两个模式。
2. 根据目前 SAS 使用的环境，建立批处理执行的程序样本。
3. 说明 SAS 常用的窗口有哪 3 个窗口及其功能。

第 2 章 程序语法

2-1 程序架构

现在我们开始编写一些 SAS 程序代码，以便读者能渐渐了解 SAS 的语法，并编出想要的程序。

SAS 程序代码分成 Data Step（数据步骤）和 Procedure Step（程序步骤）两种类型。

SAS 程序代码是由一至多个数据步骤或一至多个程序步骤互相交叉组合而形成整个 SAS 程序的。根据各个数据步骤的属性或程序步骤的功能，来建立 SAS 数据文件、外部数据文件、统计分析表和统计分析图等。各种程序代码的组合可以帮助用户完成产生的数据文件存储与否、分析报表的结果是否要再利用等各种需求。SAS/BASE 可以产生各种不同的分析报表和各种数据属性。有关数据步骤及程序步骤的语法及使用方法将在后面的章节中陆续介绍。

SAS 系统的编译方式是以数据步骤或程序步骤为一个区块来编译的，语法编译无误便开始执行该数据步骤或程序步骤，执行完整个数据步骤或程序步骤区块后便继续处理下一个数据步骤或程序步骤。一个区块接一个区块地将所有程序代码编译及执行完后，便完成了该程序的执行。在 SAS 系统下直接将 SAS 程序代码提交出去，便已将程序交由 SAS 系统去完成。用户必须确定 SAS 系统在 Log 窗口中是否有显示错误的程序代码，分析数据的属性是否正确。

SAS 的程序代码是自由格式的程序代码，每一程序代码语法结束皆以“;”作为一个程序语法的结束，一行可以含多段程序语法，一段程序语法也可以分成多行，但必须以空格或程序代码关键字为分隔。例如：

```
01 DATA OUT1;  
02 A = B + C ;  
03 D =  
04 A * 20 ;  
05 RUN ;
```

行 1 是一段程序语法。

行 2 是一段程序语法。

行 3 及行 4 是同一段程序语法。

行 5 是该段数据步骤区块的结束。

数据步骤或程序步骤区块的范围是以 RUN 关键字为区块的结束, 若没有 RUN 关键字, 则以下一段数据步骤或程序步骤的开始关键字为该区块的结束。数据步骤的开始关键字是 DATA, 程序步骤的开始关键字是 PROC。在交互模式时, 必须注意每次执行时可以执行某一数据步骤区块或某一程序步骤区块, 但必须有 RUN 关键字, SAS 系统才会将最后的区块进入编译执行的程序。

SAS 程序的注释有两种方式: *..... ; 以及/**/, 作者建议使用后者为注释方式, 这样不会造成程序属性的误解。

2-2 数据类型

SAS 数据类型分为字符串类型和数字类型两种。

字符串类型及数字类型的初始长度皆为 8Byte, 字符串类型长度介于 1~32767Byte 之间, 数字类型长度最大值为 8Byte。可以使用 Length 的语法来设置使用变量的长度。

字符串类型无法成为统计分析的统计值, 但如果属性为数字, 在数据步骤中是可以拿来计算的。

```
01  /* Program_2-1-1.sas */
02  Data Out1 ;
03    A = '1234' ;
04    B = '2121' ;
05    C = A + B ;
06  Run ;
```

行 3 及行 4: 指定字符串类型变量。

行 5: 使用字符串类型变量进行数字类型的运算产生数字类型变量。

日期数据在 SAS 中是以数值类型存储的。当需要显示日期时是通过日期格式的转换进行的(请参考 6-3 节)。

字符串类型的遗漏值(Missing Value)是 NULL, 数字类型的遗漏值是“.”。如果计算时有一变量的值为遗漏值, 相加后的值也是遗漏值, 可用 SUM 函数修正这个问题。

```
01  /* Program_2-2-1.sas */
02  DATA out1 ;
03    var_a = 1000 ;
04    var_b = . ;
```