

信息与计算科学丛书

# SAS统计分析

S A S T o n g j i f e n x i

沈其君 主编

东南大学出版社

信息与计算科学丛书

高等医药院校教材

# SAS 统计分析

沈其君 主 编

东南大学出版社  
(南京)

## 内 容 提 要

SAS 统计软件是目前国际上应用最广泛的统计软件之一,国内已广泛用于医学、理学、财经、社会科学等一切从事数据管理和数据分析处理的领域中。自 20 世纪 90 年代初,我国高等医药院校研究生和预防医学等专业的本科生教学中已陆续开设“SAS 统计软件及其应用”课程。本书以 Windows 操作系统下的 6.12 版 SAS 软件为基础,以目前医学院校使用的《医学统计学》、《卫生统计学》和《多元统计分析》教材中的统计方法为基本内容,介绍 SAS 程序的编写及结果解释。本书的特点是以统计方法为主线介绍 SAS 软件的程序编写,并将统计方法和 SAS 软件有机结合。

本书可作为高等医药院校研究生和本科生的教材,也可供有关科研人员参考。

### 图书在版编目 (CIP) 数据

SAS 统计分析/沈其君主编. —南京:东南大学出版社, 2001. 4  
(信息与计算科学丛书)  
ISBN 7-81050-674-9

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

中国版本图书馆 CIP 数据核字 (2001) 第 21153 号

东南大学出版社出版发行  
(南京四牌楼 2 号 邮编 210096)

出版人: 宋增民

江苏省新华书店经销 南航飞达印刷厂印刷

开本: 787mm×1092mm 1/16 印张: 14.5 字数: 361 千字

2001 年 4 月第 1 版 2001 年 4 月第 1 次印刷

定价: 23.00 元

(凡因印装质量问题,可直接向发行科调换。电话: 025-3792327)

## 《SAS 统计分析》编委会名单

**主 编** 沈其君 (东南大学医学院)

**副主编** 闵 捷 (东南大学医学院)

于 浩 (南京医科大学)

**主 审** 陈启光 (东南大学医学院)

### 编写组成员 (按章节顺序)

恽振先 (东南大学医学院)

张 宏 (东南大学医学院)

于 浩 (南京医科大学)

陈启光 (东南大学医学院)

彭猛业 (上海医科大学)

闵 捷 (东南大学医学院)

樊美梅 (东南大学医学院)

余小金 (东南大学医学院)

孙晓武 (University of Rhode Island)

沈其君 (东南大学医学院)

## 前 言

SAS 统计软件是目前世界上应用最广泛的统计软件之一，我国高等医学院校研究生和预防医学等专业的本科生教学中已普遍开设“SAS 统计软件及其应用”的课程。

多年来，在该课程的教学过程中，我们体会到编写一本将统计方法和 SAS 软件有机结合，并适用于不同层次学生需要的教材很有必要。

本书编写力求通俗易懂，并结合实例分析，使读者能方便地应用于实际工作。全书以目前医学院校使用的《医学统计学》和《卫生统计学》教材的基本内容在 SAS 软件中的实现为主要内容，并在此基础上增加了多元统计分析方法与 SAS 程序部分。其特点是以统计方法为主线介绍 SAS 程序编写及结果解释。对于本科教学中，学生已学过的统计方法可采取复习方式，提纲挈领地进行回顾，重点叙述其在 SAS 软件中的实现；对本科教学中未讲授的内容如方差分析(二)、Logistic 回归、多元统计分析方法和生存分析等，在统计方法和 SAS 程序两方面都作了详细叙述。这样，便于教师根据教学需要选择相应的章节进行教学；也便于有一定统计学知识基础的学生进行自学和提高。

本书除适用于医学院校研究生和不同专业、不同层次的学生选用外，还可作为高等学校其他专业人员学习 SAS 统计软件课程的教材和科研参考书。

本书编写分工如下：恽振先、张宏编写第一、二章；于浩、恽振先编写第三、四章；陈启光、彭猛业编写第五章和第九章；闵捷、樊美梅编写第六、七章；孙晓武、余小金编写第八章；沈其君、余小金编写第十、十一章。主编沈其君和主审陈启光对全书各章进行了必要的修改和补充。

由于我们的知识和经验有限，不妥之处敬请读者和同行提出宝贵意见。

**沈其君**

于东南大学

2001 年 1 月



# SAS统计分析

S A S T o n g j i f e n x i

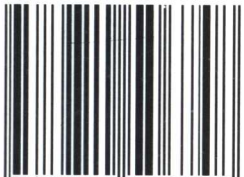
信息与计算科学丛书

责任编辑 朱经邦

责任印制 张文礼

封面设计 顾晓阳

ISBN 7-81050-674-9



9 787810 506748 >

ISBN 7-81050-674-9  
O·34 定价: 23.00元  
(丛书总定价: 180元)

# 目 录

第一章 SAS 系统简介 .....	(1)
1.1 SAS 系统的特点 .....	(1)
1.2 SAS6.12 的安装、启动和退出 .....	(2)
1.3 SAS 显示管理系统 .....	(4)
1.4 一个简单的 SAS 程序 .....	(6)
第二章 数据集的建立和编辑 .....	(9)
2.1 SAS 数据集的建立 .....	(9)
2.2 DATA 步 .....	(12)
2.3 SAS 数据集的整理和浏览 .....	(14)
2.4 SAS 数据集的排序、连接与合并 .....	(18)
2.5 常用的 SAS 运算符和 SAS 函数 .....	(21)
第三章 数值资料的统计描述 .....	(23)
3.1 数值资料常用统计指标 .....	(23)
3.2 MEANS 过程 .....	(23)
3.3 UNIVARIATE 过程 .....	(27)
第四章 数值资料的统计推断 (一) —— $t$ 检验和 $u$ 检验 .....	(33)
4.1 总体均数的区间估计 .....	(33)
4.2 总体均数的假设检验 —— $t$ 检验与 $u$ 检验 .....	(34)
第五章 数值资料的统计推断 (二) —— 方差分析 .....	(41)
5.1 方差分析的意义、应用条件及常见设计类型 .....	(41)
5.2 多样本的正态性检验和方差齐性检验 .....	(41)
5.3 完全随机设计资料的方差分析 (含多重比较) .....	(44)
5.4 随机区组设计资料方差分析 .....	(48)
5.5 拉丁方设计资料的分析 .....	(50)
5.6 析因试验设计的方差分析 .....	(52)
5.7 正交试验设计资料的方差分析 .....	(54)
5.8 交叉试验设计资料的方差分析 .....	(58)
5.9 不平衡设计的方差分析 (GLM 过程) .....	(60)
5.10 系统分组 (嵌套) 设计资料的方差分析 .....	(63)
5.11 裂区试验设计 .....	(66)
第六章 分类资料的统计方法 .....	(71)
6.1 分类资料统计描述与统计推断简介 .....	(71)
6.2 $\chi^2$ 检验 .....	(71)
6.3 二项分布与 Poisson 分布 .....	(85)

<b>第七章 秩和检验</b> .....	(88)
7.1 非参数统计概述 .....	(88)
7.2 配对符号秩和检验 .....	(88)
7.3 成组设计两样本比较的秩和检验 .....	(90)
7.4 多个样本比较的秩和检验 (Kruskal - Wallis 法) .....	(91)
7.5 频数表资料的秩和检验 .....	(93)
<b>第八章 相关分析与回归分析</b> .....	(95)
8.1 相关分析 .....	(95)
8.2 回归分析 .....	(99)
8.3 直线回归 .....	(101)
8.4 多元线性回归 .....	(104)
8.5 分类变量的处理 .....	(116)
8.6 非线性回归 .....	(119)
<b>第九章 Logistic 回归</b> .....	(124)
9.1 概述 .....	(124)
9.2 Logit 变换及 Logistic 回归方程 .....	(124)
9.3 反应变量为两分类时不同类型自变量下的 Logistic 回归 .....	(126)
9.4 反应变量为多分类有序变量的 Logistic 回归 .....	(138)
9.5 不同设计类型研究资料的 Logistic 回归 .....	(140)
<b>第十章 多元统计分析与 SAS 程序</b> .....	(151)
10.1 判别分析 .....	(151)
10.2 主成分分析与 PRINCOMP 过程 .....	(166)
10.3 因子分析与 FACTOR 过程 .....	(169)
10.4 聚类分析与 VARCLUS、CLUSTER 过程 .....	(175)
10.5 典型相关分析与 CANCORR 过程 .....	(187)
<b>第十一章 生存分析与 SAS 程序</b> .....	(194)
11.1 生存分析的基本概念 .....	(194)
11.2 生存率估计与组间比较 .....	(196)
11.3 生存率估计与组间比较的 LIFETEST 过程 .....	(201)
11.4 Cox 回归模型与 PHREG 过程 .....	(205)
<b>练习题</b> .....	(211)
<b>参考文献</b> .....	(221)



# 第一章 SAS 系统简介

SAS (Statistical Analysis System)系统意为统计分析系统,是一个模块化、集成化的应用软件系统,具有完备的数据存取、管理、分析和显示功能。该系统于1964年由美国北卡罗莱纳州立大学开始研制,1976年成立了SAS公司(SAS Institute Inc.),1985年推出SAS/PC(6.02)版本,1989年推出SAS/PC(6.04)版本,目前最新版本为Windows操作系统下的8.1版。SAS软件在世界范围内被广泛应用于政府、科研、教学、工业和商业等不同领域的的数据管理与处理,是当前国际上最流行、并具有权威性的统计分析软件。本书主要介绍Windows操作系统下的6.12版SAS软件。

## 1.1 SAS 系统的特点

### 1.1.1 SAS 系统组成

SAS 系统是由三十多个专用程序模块组成的大型集成应用软件系统,由四大部分构成:

1. Base 模块。是 SAS 系统的基础,它提供 SAS 数据库管理功能,所有其他模块必须与之结合起来使用。
2. SAS 分析模块。是 SAS 系统的灵魂,具有对数据进行分析与决策的功能,这部分主要模块有 STAT、QC、ETS、IML 等十大模块。
3. SAS 开发及显示模块。提供便捷的面向用户的开发工具,支持客户/服务器的应用开发,主要模块有 AF、EIS 与 GRAPH 模块。
4. 其他。有 ACCESS、CONNECT、SHARE 等。

### 1.1.2 SAS 系统的基本功能

SAS 系统主要完成以数据为中心的四大功能。

1. 数据管理功能。SAS 系统可手工输入数据值,然后把数据组成 SAS 数据集;它亦能够方便地将流行的数据格式,如 Excel、Lotus、DBF、TXT 等转为 SAS 数据集。SAS 数据可用临时数据集或永久数据集两种形式储存。SAS 系统提供完备的 SAS 语句和函数用于数据加工处理,并具有对数据进行编辑、整理、连接和更新等功能。
2. 基础统计计算功能。计算简单的描述统计量,多达 39 项;对数据进行标准化、编秩及计算有关统计量;生成并分析列联表;计算概率分布函数、分位数及产生随机数。
3. 报表生成和图形显示功能。产生固定格式报表、用户自定义报表;绘制二维与三维的基本的统计图,如直方图(水平或垂直)、圆形图、散点图、曲线图、时间序列图等。
4. 统计分析功能。是 SAS 系统的核心功能。提供包括方差分析、相关与回归分析、属性数据分析、判别与聚类分析、非参数分析、生存分析、时间序列分析、质量控制、运筹学等统计分析过程。

### 1.1.3 SAS 系统的模块组成

SAS6.12 版容量为 260+ MB,常用模块及功能为

1. BASE 模块。 14.35MB,进行数据管理和数据加工、处理。
2. STAT 模块。 9.51MB,完整的统计分析模块,可对数据进行单因素、多因素分析。
3. GRAPH 模块。 14.95MB,优异的图形软件包,能完成多种绘图功能。
4. ETS 模块。 9.55MB,用于计量经济与时间序列分析的专用软件,包括时间序列、时域分析和谱域分析。如逐步自回归、指数平滑、序列相关校正分析、分布滞后回归、ARIMA 模型、谱分析等。
5. OR 模块。 17.56MB,用于运筹学和工程管理的专用软件。提供全面的运筹方法,是一个优异的决策支持工具。包括线性规划、整数规划、混合整数规划和非线性规划。
6. QC 模块。 8.60MB,用于质量控制的专用软件。提供全面质量管理工具,进行试验设计、流程控制图及执行流程的能力分析等。
7. IML 模块。 2.26MB,提供功能强大的矩阵运算,用户可根据需要编写程序,完成特定的矩阵运算。

## 1.2 SAS6.12 的安装、启动和退出

SAS6.12 版为 Windows 操作系统下的纯 32 位软件,运行环境为

软件要求 Windows95/98

硬件要求 CPU: Intel586 以上及兼容处理器,具备协处理器

内存:至少 16MB

硬盘:具有 350MB 以上未用空间

其他:SVGA 显示器、光驱、鼠标

### 1.2.1 SAS6.12 的安装

SAS6.12 载体为一张光盘,附带一张高密度的授权软盘。与其他 Windows 下的软件安装一样,可以有多种途径进行安装。

常用方法为:在资源管理器中,用鼠标左键双击光盘根目录下的 SETUP 图标,然后按提示选择典型安装(推荐)或自定义安装。

另一种方法为:或单击“开始”—“运行”,在“打开”框中输入“E:\SETUP.EXE”(E 为光驱盘符),然后按提示进行安装。

程序安装完成后,在“开始”—“程序”菜单中会建立“The SAS System”程序组,包括“The SAS System for Windows v6.12”和“Update SAS License Information”两个程序项,在第一次启动 SAS 之前,需要更新 SAS 的授权信息,即用鼠标双击“Update”项,按提示将授权软盘中包含新的授权信息的“SETINIT.SAS”运行一次,完成授权信息的更新。如未能更新授权信息,则 SAS 只能在一定的时间范围内才能运行。

### 1.2.2 SAS6.12 的启动

SAS 的启动常用两种方法:一是快捷方式,即按创建 Windows 系统中应用程序的快捷方式的步骤,通过资源管理器,把 SAS 目录中的 SAS.EXE 拖放至桌面,形成快捷方式,再用鼠标左键双击这一快捷方式图标,即可启动 SAS。二是菜单方式,在“开始”——“程序”菜单中的“*The SAS System*”程序组中,选定其中的“*The SAS System for Windows v6.12*”程序项,即可将 SAS 启动。启动画面如图 1.1。



图 1.1 SAS6.12 启动画面

然后,系统将自动进入 SAS 的显示管理系统(*Display Manager System*),见图 1.2。

可以看出这是一个典型的 Windows 平台下的程序窗口,顶部依次为标题栏、菜单栏、命令与工具栏。程序窗的主体出现 LOG 和 PROGRAM EDITOR 两大窗口,在 LOG 窗口中显示有版权与授权信息。

### 1.2.3 SAS6.12 的退出

1. 临时退出 SAS 作业,返回到 Windows 状态。在使用 SAS 过程中,需要暂时调用 Windows 下的其他软件时,只要点击 SAS 窗口右上角的最小化按钮即可临时退出 SAS 作业。通过点击任务栏中的 SAS 图标或“Alt”+“Tab”组合键即可回到 SAS 系统。

2. 临时退出 SAS 系统,进入 DOS 状态。在使用 SAS 过程中,需要暂时调用 DOS 下的命令或程序,可通过点击 SAS 窗口中工具栏中的 DOS 图标,或在 SAS 窗口顶部左侧的命令栏中输入“X”后回车来实现。若需返回 SAS 系统则在 DOS 提示符后输入“EXIT”后回车。

3. 退出 SAS 系统。完成 SAS 操作后退出 SAS 系统,可通过点击 SAS 窗口右上角的关闭图标,或在 SAS 窗口顶部左侧的命令栏中输入“BYE”后回车。

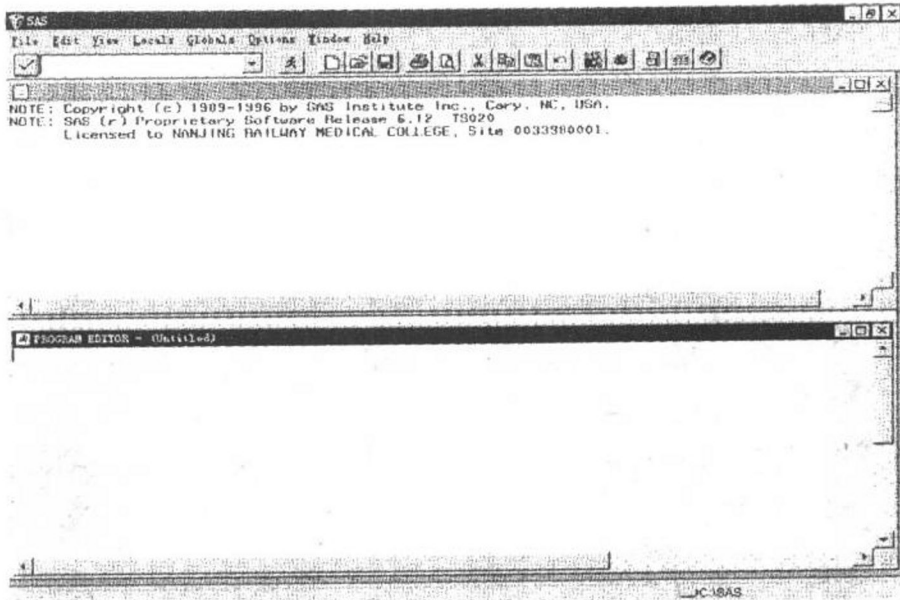


图 1.2 SAS6.12 显示管理系统(DMS)

## 1.3 SAS 显示管理系统

SAS 显示管理系统是 SAS 系统运行的主要界面,它由一系列窗口组成,通过这些窗口,用户发布执行各种不同任务的命令。SAS 显示管理系统提供众多 SAS 命令,有些命令可以在任一窗口里发布,另一些可在多个窗口里工作,还有的只能在特定的窗口里工作。下面就一些实际操作中经常使用的窗口与命令进行介绍。

### 1.3.1 显示管理系统的窗口及其功能

#### 1. 三个基本窗口及其功能

(1) PROGRAM EDITOR(PGM)窗口。具有全屏幕文本编辑功能,可以输入和编辑文本,包括用 SAS 语言编写 SAS 程序、编辑 SAS 运行结果等;提交 SAS 程序;拷贝一个外部文件到 PGM 窗口,或拷贝 PGM 窗口内容到外部文件上;拷贝一个目录条目到 PGM 窗口,或拷贝 PGM 窗口内容到一个目录条目。

(2) LOG 窗口。包含 SAS 日志,是当前提交运行的 SAS 程序执行过程的记录。一个典型的记录显示有:在当前提交执行的 SAS 语句;执行 SAS 语句过程有关的注释、警告和出错信息;DATA 步执行的一些结果和一些 SAS 过程的结果。

(3) OUTPUT 窗口。是 SAS 运行结果的显示窗口,将 SAS 过程执行后,相继产生的结果追加到该窗口的内容中。

#### 2. 其他显示管理窗口及其功能

除了三个基本窗口外,显示管理系统还包括许多其他窗口,共计 20 个。较常用的有:

KEYS 窗口	查看及修改功能键的设置
HELP 窗口	显示 SAS 系统的帮助信息
LIBNAME 窗口	显示指定的库标记和它的路径
DIR 窗口	显示已建立的所有临时 SAS 数据集名称
VAR 窗口	显示当前 SAS 数据集中的变量及它们属性的信息
OPTIONS 窗口	显示和改变 SAS 系统设置
.....	

### 3. 窗口的激活与关闭

在众多窗口中每次仅有一个窗口是当前被激活可供使用的,激活窗口的方法是在命令栏中输入相应窗口名后回车,或按指定功能键激活对应的窗口,并实现窗口间的切换。各功能键对应的窗口可以在 KEYS 窗口进行定义与修改。若要关闭当前窗口,可以通过点击该窗口右上角的关闭图标来实现。

#### 1.3.2 功能键

为了用户操作的方便,可以将常用的 SAS 命令定义成快捷功能键,在使用过程中来代替相应的命令。利用 KEYS 窗口来完成这一操作。在命令栏输入“KEYS”或按 F9 键即可进入 KEYS 窗口。在命令栏输入“END”或点击 KEYS 窗口右上角的关闭窗口图标即可退出。在 KEYS 窗口中各功能键右侧的定义栏中输入相应的 SAS 命令或修改原有的 SAS 命令,即完成了功能键的重新定义;退出 KEYS 窗口后,SAS 系统会自动保存当前各功能键的定义。若要取消当前定义,可在命令栏中输入“CANCEL”来退出 KEYS 窗口。

#### 1.3.3 常用全局命令

显示管理系统除了众多窗口外,还包括许多命令,用户通过在命令栏中发布这些命令来完成各种任务或直接在下拉菜单中选择相应的命令。按功能分类的常用命令如下:

1. 窗口调用命令。显示一个窗口并把光标移动到该窗口。共有 26 个窗口调用命令,其中最常用的 5 个为

PGM 或 Globals 菜单中的 Program Editor	进入 PROGRAM EDITOR 窗口
LOG 或 Globals 菜单中的 Log	进入 LOG 窗口
OUTPUT 或 Globals 菜单中的 Output	进入 OUTPUT 窗口
KEYS 或 Help 菜单中的 Keys	进入 KEYS 窗口
HELP 或 Help 菜单中的 Extended Help	进入 HELP 窗口

2. 窗口管理命令。帮助用户更有效地使用窗口。共有 24 个窗口管理命令,其中常用的 7 个为

BYE 或点击窗口关闭按钮	结束 SAS 作业,返回到 WINDOWS 下
CANCEL	取消在窗口中的改动,并把该窗口从显示屏幕中移走
CLEAR 或 Edit 菜单中的 Clear Text	删除激活窗口中的内容或设置的显示
END 或点击该窗口关闭按钮	关闭当前激活的窗口
ENDSAS	结束 SAS 作业,返回到 WINDOWS

X 或点击工具栏中的 DOS 提示按钮 进入到 DOS 系统

ZOOM 或点击该窗口最大化按钮 使激活的窗口最大化

3. 文件管理命令。控制文件的存贮、删除与打印等。共有 11 个命令,其中最常用的 3 个为

FILE‘文件名’或 File 菜单中的 Save 把当前窗口中的内容输出到指定的文件中

INCLUDE/INC‘文件名’或 File 菜单中的 Open 把指定文件调入到激活的 PGM 窗口中

FILE‘PRN’或 File 菜单中的 Print 打印当前窗口中的内容。

### 1.3.4 PGM 窗口全屏幕编辑

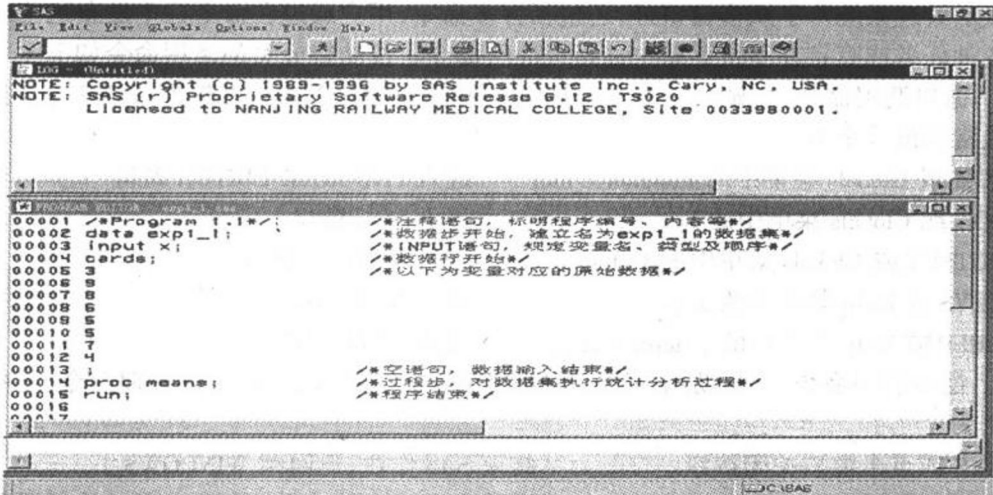
1. 编辑。PGM 窗口具有字处理的编辑功能,可以像 WORD 一样来自由编辑窗口中的文档,操作也与 WORD 相似,这里不赘述。

2. 字符串查找与替换。利用 Edit 菜单中的 Find 与 Replace 命令,可以实现对字符串的查找及替换。

3. 程序执行与重新调用命令。当 SAS 程序已经输入完毕后,可以用 F8 键或 Locals 菜单中的 Submit 命令提交给 SAS 系统执行,同时可以方便地用 F4 键或 Locals 菜单中的 Recall text 命令将已经提交给 SAS 系统的程序调回 PGM 窗口供修改与保存。

## 1.4 一个简单的 SAS 程序

下面以一个简单的 SAS 程序为例,说明 SAS 程序的组成、语句的编写规则以及程序的运行和结果。



```
NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software Release 6.12 TS020
Licensed to NANJING RAILWAY MEDICAL COLLEGE, Site 0033980001.

00001 /*Program 1.1*/
00002 data expl_1;
00003 input x;
00004 cards;
00005 3
00006 8
00007 8
00008 6
00009 5
00010 5
00011 7
00012 4
00013 ;
00014 proc means;
00015 run;
00016
00017
```

图 1.3 例 1.1 的 SAS 程序及说明

### 1.4.1 实例

**例 1.1** 8名健康成年男子血液第一小时末红细胞沉降率(mm)分别为 3,9,8,6,5,5,7 和 4,试计算其均数和标准差。将其输入 SAS 系统,如图 1.3 所示。

#### 1. SAS 程序结构

一个完整的 SAS 程序一般由数据步(DATA 步)和过程步(PROC 步)两部分组成。DATA 步通过指定数据集名称、定义数据集的结构(变量名、变量类型等)和读入原始数据来建立 SAS 数据集。在图 1.3 中第 2 行~第 13 行语句为数据步。PROC 步调用不同的分析模块来完成对数据集的统计分析。图 1.3 中的第 14 行语句即为过程步。SAS 程序每一数据步以 DATA 语句开始,而过程步以 PROC 开始,整个程序以 RUN 语句结束。

#### 2. SAS 程序的输入、修改与存储

(1) 程序输入规则。输入字母可用大、小写或混用,语句中每个字(WORD)及数字间至少以一个空格分隔,每个语句以分号“;”结束。每行可输入多个语句,但为修改方便每行最好只输入一个语句。CARDS 语句后的数据项要另起一行,数据输完后以空语句“;”结束。

(2) SAS 程序的修改。若程序中存在错误,则会导致无法得到相应的结果,而且在 LOG 窗口中会出现红色的错误提示,根据这些提示,将 SAS 程序重新调入 PGM 窗口后(F4 或 Locals 菜单中的 Recall text),即可进行修改。

(3) SAS 程序的运行。编辑好的 SAS 程序可以通过点击 SAS 窗口工具栏中的运行图标“?”或 F8 功能键运行,也可以在命令栏中输入 SUBMIT 命令提交运行。

(4) SAS 程序的储存。在命令栏中输入“file ‘文件名’”或在 SAS 窗口中的下拉菜单“FILE”中选择“SAVE”项,可将 SAS 程序以文件的形式储存起来,供以后调用。如图 1.4 中所示。

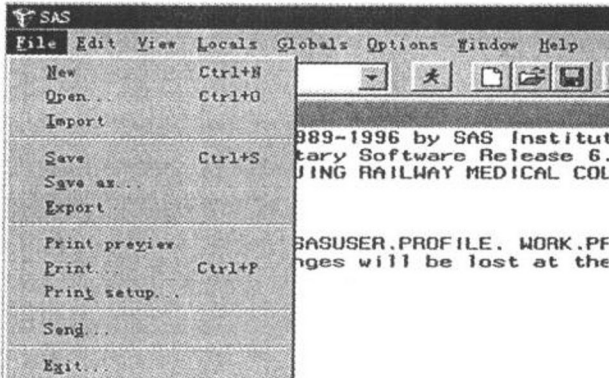


图 1.4 SAS 程序的储存

#### 3. SAS 程序的特点

通过上面的简介,可以看出以下几点:

(1) SAS 系统操作简单。SAS 系统为用户提供了交互式管理系统,使用户看到程序的输入、程序执行的过程及程序执行的结果,界面亲切、易用,为非计算机专业人员提供了一个易学而又完整的软件系统。

(2) 近似英语的非过程语言(第四代语言)。SAS 系统把所要解决的问题用 SAS 语言组成程序,只要告诉需做什么,而不必告诉如何去做。

(3) 数据处理与统计分析的有机结合。SAS 提供了几乎所有的统计分析方法,编制了相应的过程模块,使用户能够按实际需要进行各类统计分析。



## 第二章 数据集的建立和编辑

SAS 系统只能分析 SAS 数据集,因此如何建立、调用及把其他数据转换为 SAS 数据集是应用 SAS 系统的基础。本章将介绍建立 SAS 数据集的几种方法、对 SAS 数据集操作的常用语句以及各种数据转换为 SAS 数据集的方法。

### 2.1 SAS 数据集的建立

#### 2.1.1 SAS 数据集与数据库的概念

1. SAS 数据集的结构。SAS 数据集是关系型结构,分两个部分:描述部分和数据部分。描述部分包含了一些关于数据属性的信息即变量名称、类型及次序等信息;数据部分包括数据集值。SAS 数据集值被安排在一个矩阵式的表状结构中,表的列称为变量(相当于域或字段);表的行称为观测(相当于记录)。

2. SAS 数据库和库标记。SAS 数据集储存在被称为 SAS 数据库(library)的文件集中。SAS 数据集是最常用的一种 SAS 文件类型,但 SAS 数据库还可以包含其他类型的 SAS 文件,这些数据库使得 SAS 系统能够在 SAS 程序中指明并查阅文件。

SAS 数据库在不同的主操作系统下有不同的表示和命名规定。对 Windows 操作系统, SAS 数据库以路径为基础的组织方式。为了使用 SAS 数据库,需要为每个 SAS 数据库指定一个库标记来识别。这一标记又称库逻辑名,是某文件所在路径的别名。使用 LIBNAME 语句可以指定 SAS 库标记。它的一般形式如下:

```
LIBNAME 库逻辑名'路径';
```

例如指定路径'c:\sas\'的库标记(逻辑名)为 AA,可在 PGM 窗口中运行。

```
LIBNAME AA'c:\sas\';
```

SAS 数据库可分为永久型和临时型两类。临时型数据库的库标记为 WORK,在 SAS 启动后自动生成,结束 SAS 后,库中的所有文件将被删除。用 LIBNAME 定义的库逻辑名则为永久型数据库。

3. SAS 数据集命名。为了调用及识别 SAS 数据集,需对创建的每个 SAS 数据集命名,每个 SAS 数据集有一个两级名称,第一级为库标记(逻辑)名,默认 WORK;第二级为数据集名,两者间用“.”分隔。临时数据集只要有数据集名即可。

#### 2.1.2 用 DATA 步创建 SAS 数据集

1. 在 PGM 窗口创建 SAS 数据集。在 DATA 步中,用 INPUT 语句指定要读取数据的变量名及输入格式,在 CARDS 语句中依次输入原始数据。该法多用于要读取的数据较少时,实例 1.1 即用此法创建数据集。

2. 调用 ASCII 码文件创建 SAS 数据集。当原始数据已被编辑为 ASCII 纯文本文件时,