21 世纪高等院校教材

SAS软件实用教程

张 瑛 雷毅雄 主编

科学出版社
www.sciencep.com

21世纪高等院校教材

SAS软件实用教程

主 编 张 瑛 雷毅雄

科学出版社
北京

内 容 简 介

本教材以简明、实用的写作手法，从实例出发，介绍 SAS 软件统计分析的相关基础知识和应用。全教材共十章，第一、二章主要介绍 SAS 系统的操作环境，SAS 语言的语句、函数、程序结构，SAS 数据集的建立与修改；第三章介绍 SAS 常用程序在描述性统计中的应用；第四章至第十章介绍 SAS 常用程序在推断性统计的应用。教材中实例均有 SAS 程序编写和程序运行结果的说明与解释。

本教材既可作为高等院校本科生和研究生《卫生统计学》理论教材配套的 SAS 实验教材，又可作为医学工作者学习统计软件的参考书或工具书。

图书在版编目（CIP）数据

SAS 软件实用教程 / 张瑛，雷毅雄主编. —北京：科学出版社，2009
21 世纪高等院校教材
ISBN 978-7-03-024372-0

Ⅰ. S… Ⅱ. ①张… ②雷… Ⅲ. 统计分析-应用软件，SAS-高等学校-教材 Ⅳ. C812

中国版本图书馆 CIP 数据核字（2009）第 052716 号

策划编辑：李国红　周万灏 / 责任编辑：周万灏　李国红 / 责任校对：桂伟利
责任印制：刘士平 / 封面设计：黄　超

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
*
2009 年 4 月第　一　版　开本：787×1092　1/16
2009 年 4 月第一次印刷　印张：8
印数：1—4 000　　字数：178 000
定价：19.80 元
如有印装质量问题，我社负责调换

《SAS 软件实用教程》编委会

前　言

《卫生统计学》早已是我国高等医药院校本科生和研究生教育的基础课之一。随着计算机应用的普及，利用统计软件进行统计实验的高校也越来越多。统计软件不仅为统计应用者消除了大量数据处理的烦恼，同时可以促进使用者对统计理论和方法的深入理解，不断提高人们应用统计学的能力，统计软件已成为统计应用最有力且不可缺少的工具。目前，国内大部分医学统计教材都增加了软件实验的内容，一般采用最权威的国际标准统计软件——SAS 软件系统，限于篇幅，教材里只是简单地给出实例的 SAS 程序，但对程序和结果的解释较少，导致学生们在理解上存在很大的障碍，不能灵活应用。因此，我们急需要与理论教材配套的 SAS 实验教材，能够比较详细地介绍 SAS 操作，作为使用指南供读者随时翻阅参考。实际上，国内也有一些专门介绍 SAS 的书籍，因为编写的目的不同，适合医学统计配套教学的教材较少，这就鼓励我们尽全力编写出一本好的、符合目前国内教育环境的 SAS 实验教材，供读者参考。

本教材主要作为预防医学专业《卫生统计学》(108 学时)和《高级统计学》(72 学时)的配套教材，用于统计学的实验教学，实验课为 36 学时。经过多年的教学实践，我们从 DOS 版 SAS6.03 起，经过 SAS6.12、SAS8.0 和 SAS9.0 等，随 SAS 的版本一次次升级，我们编写教材的内容也不断改版完善，满足了本科生和研究生的统计实验教学的需要，且在为社会提供继续教育培训中，也取得了良好的教学效果。

本教材分为两大板块共十章。第一板块由第一、二章组成，主要介绍 SAS 软件的基础知识，包括 SAS 系统的操作环境，SAS 语言的语句、函数、程序结构和 SAS 数据集的建立与修改。第二板块由第三章至第十章组成，着重介绍 SAS 在卫生统计学上的应用，其中第三章为 SAS 常用程序在描述性统计的应用，第四章至第十章为 SAS 常用程序在推断性统计中的应用。

在本教材的编写过程中，我们力求突出简明和实用的特点。简明性表现为着重介绍常用的统计分析过程和语句的使用，不常用的语句和过程不做累赘的介绍。实用性通过两方面来体现，一是解题思路符合统计学逻辑思维的要求；二是注重于解决实际问题，如第二板块各章节以实例引出，对 SAS 程序和程序运行结果给予较为详细的说明和解释，便于学习者对程序、结果和结论进行比对学习，加深对 SAS 程序的理解，做到“知其然也知其所以然”，能较快地灵活运用 SAS 软件来解答实际问题。

本教材采用“案例”叙述法，案例既有来自不同版本的《卫生统计学》、《医学统计学》，也有来自编写组的科研数据，在本教材中不一一注明出处，都列入参考文献。

由于我们水平有限，编写时间仓促，难免有错漏之处，敬请读者批评指正，以便我们不断改进。

张　瑛　雷毅雄

2008 年 10 月

目　录

第一章　SAS 系统概述

第一节　SAS　简　介

一、SAS 的创立和发展

SAS 系统是 20 世纪 70 年代早期由美国 North Carolina 州立大学研制并逐渐发展起来的。最初它主要用于农业领域试验的数据管理和分析，所以 SAS 字母的原意是统计分析系统(Statistical Analysis System，SAS)。但从推出之日至今，经过近 40 多年的不断发展和完善，SAS 已由最初的统计分析软件，逐渐成为一个用来管理、分析数据和编写报告的大型集成应用软件系统，具有完备的数据访问、管理、分析、呈现及应用开发等功能，完全超出了单纯统计应用的功能。因此，目前 SAS 已不再表示任何含义的首字母缩写。在数据处理和统计分析领域，SAS 系统被誉为国际上标准软件系统，属于世界领先，使用最为广泛的统计软件之一。

二、SAS 系统的组成部分

SAS 系统是一个模块化的组合软件系统，它提供了约 20 多个模块，各个模块之间既相互独立又相互交融补充。本课程用得最多的是 Base SAS 模块和 SAS/STAT 模块。

Base SAS 是 SAS 系统的基础。它既可以单独使用，也可以与其他模块组成一个用户化的 SAS 系统，但是其他模块必须与之结合起来才能使用。Base SAS 主要承担着数据及用户使用环境的管理、SAS 语言程序的处理，并具有基本的数据分析和报告等统计功能。

SAS/STAT 提供了当今流行的主要统计分析方法，是国际上统计分析领域的标准权威软件。它具有回归分析、方差分析、属性数据分析、多元分析、聚类分析、判别分析、非参数分析、生存分析和心理测量分析等统计功能。该软件时常更新，充分反映了统计方法的新进展。

SAS 系统中其他常见的模块包括：

SAS/GRAPH 是可以完成多种绘图功能的图形软件包，包括二维及三维的曲线图、直方图、圆饼图、区块图、星形图、地理图，以及各种映像图。这些图形非常形象、直观地表达各变量之间的关系及数据的分布状态，对解决各种实际问题起着重要的辅助作用。

SAS/IML 主要用于进行矩阵运算，它提供功能强大的面向矩阵运算的编程语言。它处理的基本数据是一个矩阵，用户可直接用矩阵代数的记号来组成 IML 的程序语句，用于研究新算法或作为解决 SAS 系统中没有现成方法的工具。

SAS/ETS 主要用于经济预测和时间序列分析。它提供了经济分析、时间序列分析、时间序列预测、建立计量经济和财务模型、周期性调整、金融分析和报表、经济和金融数据集接口，以及时间序列资料管理等过程。此外，SAS/ETS 也包括了一个进行交互式时间序列预测的菜单驱动系统。

SAS/INSIGHT 是可视化的数据探索工具，是进行数据挖掘的有力工具。它采用交互式数据分析，制作各种统计图形以及用于方差分析、线性拟合、Logistic 回归和 Poisson 回归等。

SAS/ASSIST 为 SAS 系统提供了面向任务的、菜单驱动用户的友好界面。它可免去用户学习 SAS 语言的负担，同时 SAS/ASSIST 生成的 SAS 程序既可帮助用户学习 SAS 语言，又可辅助有经验的用户快速编写 SAS 程序。

SAS/ACCESS 是对目前许多流行数据库的接口组成的接口集，这种接口是透明和动态的。

SAS/OR 主要用于运筹学和线性规划，是运筹学和工程管理的专用软件，该软件不仅包含通用的线性规划、混合整数规划和非线性规划的方法，还包含用于项目管理、时间安排和资源分配等问题的一整套方法。

SAS/QC 提供了根据产品观测数据进行产品质量管理的各种分析工具。它是质量管理的专用软件，可提供完整的实验设计和质量管理的菜单系统，用于实验设计、质量管理和过程控制。

SAS/FSP 提供全屏幕交互式数据输入、编辑、查询功能，还可设计数据输入屏幕，并具有书信撰写方面的功能。

SAS/AF 是一个应用灵活的开发工具，利用 SAS/AF 的屏幕设计能力及 SCL 语言的处理能力可快速开发各种功能强大的应用系统，通过 SAS/AF 的 OOP(面向对象编程)的技术，可建立用户化菜单系统，连接各种应用程序。

此外 SAS8.0 以上版本还包括 **SAS/GIS**，**SAS/IntrNet**、**SAS/LAB**、**SAS/MDDB Server**、**SAS/Spectraview**、**OLE DB Providers** 和 **IT Service Vision** 等模块。

第二节　SAS 窗口工作环境

一、启　动　SAS

双击桌面上的快捷方式图标  ，或从开始→程序菜单中找到 The SAS System for Windows，点击启动(图 1-1)。

图 1-1　启动 SAS

二、SAS 窗口环境

启动 SAS 后，可看到图 1-2 的界面。本部分有五个主要的 SAS 窗口，分别是：**Editor** 窗口、**Log** 窗口、**Output** 窗口、**Explorer** 窗口和 **Results** 窗口。这些窗口可以帮助我们轻松完成很多最基本的 SAS 任务。点击窗口条上相应的按钮可将某窗口移至前台，成为当前活动窗口。

下面我们分别介绍五个窗口的主要功能。

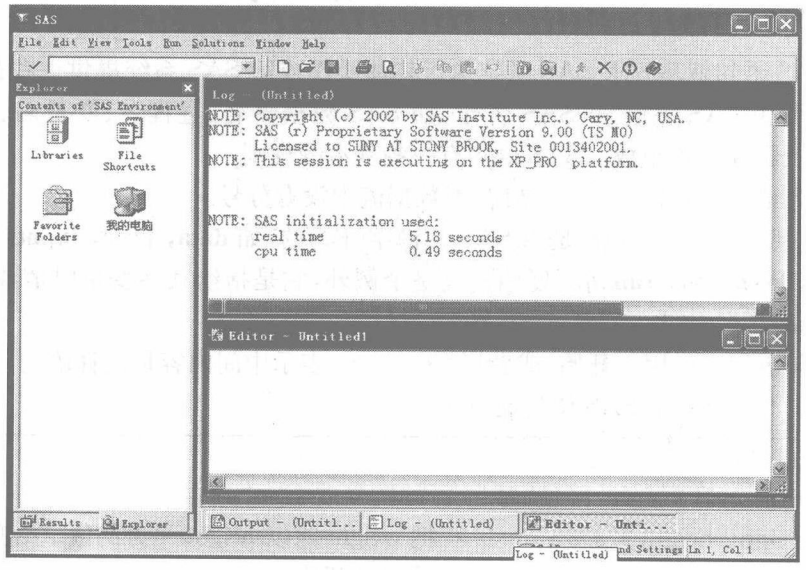

图 1-2　SAS 的窗口工作界面

Editor 窗口：主要用于打开 SAS 程序文件（SAS 程序文件扩展名为 *.sas）、编辑和修改 SAS 程序、并提交全部或部分 SAS 程序。根据程序中编码的性质可以显示不同的颜色，并且对 SAS 语言进行语法检查。在 SAS 中可同时打开多个 **Editor** 窗口进行操作。

Log 窗口：该窗口显示有关的 SAS 会话和提交 SAS 程序的信息，包括程序的出错信息等（log 文件的扩展名为 *.log）。

Output 窗口：该窗口主要用于显示提交 SAS 程序后的运行结果（output 文件的扩展名为 *.lst）。缺省时，该窗口位于 **Editor** 窗口和 **Log** 窗口的后面，如果运行程序有结果输出时，该窗口自动移至前台。

Explorer 窗口：这个窗口主要用于查看和管理所有 SAS 文件，而且可以对非 SAS 文件创建快捷方式。它类似 Windows 操作系统中的资源管理器，在这里可以创建新的库（Libaries）和 SAS 文件（SAS files），并且对文件进行移动、复制、粘贴、重命名、删除等操作。

Results 窗口：该窗口主要用于操作和管理提交 SAS 程序后的输出结果。它的内容与 **Output** 窗口的内容一一对应，可以看做是 **Output** 窗口内容的名称，可以用它来查看、删除、保存和打印部分或全部结果。缺省时，它位于 **Explorer** 窗口的后面，点击窗口条上的**Results** 按钮可将它移至前台。

第三节　SAS 语言的语句和程序

运用 SAS 系统管理和分析数据，用户通常需要先利用 SAS 语言编写一系列指令，称为 SAS 程序。通过 SAS 程序，用户可以告诉系统如何定义数据以及要求系统对数据做哪些统计分析。学习用 SAS 语言编写 SAS 程序，可以帮助应用者更灵活地应用 SAS。

一、SAS 语　句

一个 SAS 语句就是要求 SAS 系统执行某种操作或给 SAS 系统提供一些信息的命令。SAS 语句通常由 SAS 关键字、SAS 名称（如 SAS 数据集名、过程名、变量名、输出格式名等）、运算符及特殊字符组成。SAS 语句有一些基本的规则：

(1) SAS 语句都是以分号“;”结尾；但数据流中没有分号。

(2) 几乎所有 SAS 语句都是以 SAS 关键字开始的：如 **data**，**proc**，**input**，**cards**，**model**，**class**，**if**，**keep**，**set**，**run** 等。赋值语句是个例外，它是指给某些变量赋值或把某些变量的值赋给其他变量，如 $x=2$；$y=3$；$z=x+y$ 等。

(3) 注释语句可以用 * 开始，或者用/ * … * /表示中间内容是注释语句。

例 1.1　下面是几个 SAS 语句的例子：

```
data sas1_1;
input id name$ height weight;
bmi=weight/height**2;          /* 把 weight 除以 height 平方的值赋给变量 bmi */
cards;                         /* 数据流开始 */
1 Judy 1.56 45
2 Lucy 1.67 53
;                              /* 数据流结束 */
proc print data=sas1_1;
run;
```

其中 **data**，**input**，**cards**，**proc**，**run** 等是 SAS 关键字；sas1_1 是数据集名；**id**，**name**，**height**，**weight** 和 **bmi** 是变量名，**name** 是个用 $ 定义的字符变量，其余是数值变量；“=”、“/”和“**”属于 SAS 运算符；**print** 是 SAS 过程名。

1. SAS 关键字

几乎所有的 SAS 语句都是由 SAS 关键字开始的，说明 SAS 语句的类型。如例 1.1 中的语句可以分别称为：**data** 语句，**input** 语句，**cards** 语句，**proc** 语句，**run** 语句等。

2. SAS 数据集名和 SAS 变量名

SAS 数据集名和 SAS 变量名也有一些基本规则：

(1) SAS 名长度不能超过 32 个字符。

(2) 第一个字符必须是字母 A、B、…、Z 或下划线“_”；从第二个字符开始，可以为字母

A、B、…、Z,阿拉伯数字0、1、…、9或下划线“_”等。

(3) 所有SAS名称中的英文字母不区分大小写。

(4) 空格和特殊字符(如◎＃￥％＄等)不允许在SAS名中使用。

例1.2 一些SAS名称:

```
xt5_1;_num3;year2008; _n_;age;sex;name
```

3. SAS运算符

SAS运算符是用于比较运算、算术运算或逻辑运算的符号。常用的SAS运算符包括:算术运算、比较运算、逻辑运算或布尔运算符,最大或最小连接等运算符(表1-1)。SAS表达式的运算次序与通常的算术运算规则相同,如括号内运算优先和较高级运算符优先等。

表1-1 SAS运算符

运算符	说明	例子
算术运算符		
+	加	$x+y$
−	减	$x-y$
*	乘	$x*y$
/	除	x/y
* *	乘方	$x**y$
比较运算符		
=	等于	$x=y$
^=	不等于	$x^=y$
>	大于	$a>b$
>=	大于等于	$a>=b$
<	小于	$a<b$
<=	小于等于	$a<=b$
逻辑运算符		
And/&	逻辑与	$x>2$ and $y>3$;
Or/|	逻辑或	$x>2$ or $y>3$;
Not/^	逻辑非	
其他		
<>	最大	3<>5:结果为5
><	最小	3><5:结果为3
||	连接	A=“my name is”,B=“SAS”,C=A||B,那么C=“my name is SAS”

4. SAS函数

SAS函数是一个独立的子程序,它对0个或多个自变量进行计算后返回一个值,每个函

数都有一个关键字名，为了调用一个函数，需要先写出它的函数名，再用括号将 0 个或多个自变量括起来，跟在函数名后面，表示这个函数对这些自变量执行某种运算。函数一般形式为：函数名(自变量，自变量，…)。

SAS 函数有多种，这里介绍部分常用函数。

(1) SAS 常用概率和密度函数：

1) 标准正态分布函数：PROBNORM(x)。

该函数计算服从标准正态分布的随机变量 u 小于给定 x 的概率。即 $p(u<X)$。如 y=probnorm(−1.96)，结果为 0.025。

2) t 分布概率函数：PROBT(x,df,nc)。

计算自由度为 df，非中心参数为 nc 的 t 分布随机变量小于给定值 x 事件的概率，当 nc=0或不规定这项时，该分布为中心分布。如 y=probt(0.95,100)，结果为 0.8278。

3) F 分布概率函数：PROBF(x,$df1$,$df2$,nc)。

计算服从分子自由度为 $df1$，分母自由度为 $df2$ 的 F 分布的随机变量小于给定值 x 事件的概率，当分布为中心分布时，nc=0 或不规定该项。

4) χ^2分布概率函数：PROBCHI(x,df,nc)。

计算服从自由度为 df，非中心参数为 nc 的 χ^2分布的随机变量小于给定值 x 事件概率，如 nc 没有规定或取为 0，那么就是中心 χ^2分布。

5) 二项分布概率函数：PROBBNML(p,n,m)　　其中 $0\leqslant p\leqslant 1$，$n\geqslant 1$，$0\leqslant m\leqslant n$。

计算率为 p，样本例数为 n 的二项分布，随机变量 $x\leqslant m$ 的概率。

如求 $p(x=k)$的值，可计算 probbnml(p,n,k)−probbnml(p,n,$k-1$)。

6) 泊松分布概率函数：POISSON(m,n)　　其中 $m\geqslant 0$，$n\geqslant 0$。

计算参数为 m 的泊松分布的随机变量 $x\leqslant n$ 的概率。

如计算 $P(x=k)$的值，可用 Poisson(m,k)−Poisson(m,$k-1$)。

(2) SAS 常用分位数函数：

1) 正态分布分位数函数：PROBIT(p)($0\leqslant p\leqslant 1$)。

计算标准正态分布的分位数，是概率函数 PROBNORM 的逆函数。

如 probit(0.025)，结果为−1.96。

2) t 分布的分位数函数：TINV(p,df,nc)。

计算自由度为 df，非中心参数为 nc 的 t 分布的 p 分位数，如 nc 没有规定或取 nc=0，就计算中心 t 分布的 p 分位数。

3) F 分布的分位数函数：FINV(p,$df1$,$df2$,nc)。

计算分子自由度为 $df1$，分母自由度为 $df2$，非中心参数为 nc 的 F 分布的 p 分位数，如 nc 没有规定或取 nc=0，就计算中心 F 分布的 p 分位数。

4) χ^2分布的分位数函数：CINV(p,df,nc)。

计算自由度为 df，非中心参数为 nc 的 χ^2分布的 p 分位数，如没有规定 nc 或取 nc=0，就计算中心 χ^2分布的 p 分位数。

(3) 其他 SAS 常用函数见表 1-2。

表 1-2 其他 SAS 常用函数

函 数	说 明
算术函数	
ABS(X)	取 X 的绝对值。
SQRT(X)	计算 X 的平方根。
MAX(X_1,…,X_n)	求 X_1,…,X_n 中的最大值。
MIN(X_1,…,X_n)	求 X_1,…,X_n 中的最小值。
MOD(X,Y)	求 X/Y 的余数。
	如 MOD(10,3)=1,MOD(6,2)=0。
SIGN(X)	当 $X<0$ 时其值为−1,当 $X>0$ 时其值为 1,当 $X=0$ 时其值为 0。
	如 SIGN(3.5)=1,SIGN(−5.4)=−1,SIGN(0)=0。
EXP(X)	计算 e 的 X 次幂。EXP(X)=e^X。
LOG(X)	对自变量 X 求以 e 为底的自然对数。
LOG2(X)	对自变量 X 求以 2 为底的对数。
LOG10(X)	对自变量 X 求以 10 为底的对数。
截取函数	
CEIL(X)	取≥自变量 X 的最小整数。
	如 CEIL(5.7)=6,CEIL(−2.3)=−2。
FLOOR(X)	取≤自变量 X 的最大整数。
	如 FLOOR(6.9)=6,FLOOR(−7.2)=−8。
INT(X)	取 X 的整数部分。
	如 INT(5.6)=5,INT(−3.7)=−3。
ROUND(X,n)	X 按 n 指定的精度取舍入值。
	如 ROUND(73.58,0.1)=73.6。
随机数函数	
UNIFORM(seed)或 RANUNI(seed)	产生服从均匀分布 UNI(0,1)的随机数。
NORMAL(seed)或 RANNOR(seed)	产生服从标准正态分布 $N(0,1)$的随机数。经如下变换:$M+s*$ NORMAL(seed),可得到服从正态分布 $N(M,s^2)$的随机数。
RANEXP(seed)	产生一个参数 $\lambda=1$ 的指数分布的随机数。
	如果 Y=RANEXP(seed)/λ,产生参数为 λ 的指数分布随机数。
RANBIN(seed,n,p)	产生服从均值 np,方差为 $np(1-p)$的二项分布的随机数。
RANPOI(seed,λ)	产生服从均数为 λ 的泊松分布的随机数。

注:表中所列统计符号、函数等均为 SAS 程序自动列出,故为真实反映软件的运行情况,我们未做任何改动。全书余同。

二、SAS 程 序

将一系列 SAS 语句按逻辑顺序排列起来,构成 SAS 程序。通常 SAS 程序包含数据步和过程步两部分。数据步以 data 语句开头,以 run 语句结束,其主要作用是建立 SAS 数据

集。过程步以 proc 开头，以 run 语句结束，其主要作用是激活 SAS 过程对数据进行处理和分析。通常，在数据处理过程中，可有多个数据步和多个过程步或多个数据步和多个过程步混合使用。后一个 data 或 proc 语句起到前一步 run 语句的作用，故两步中间的 run 语句常省略，但最后一步的后面必须要有 run 语句，否则最后一步将不能运行。Endsas 语句可以终止交互式的 SAS 工作。

SAS 语句书写格式自由，可在各行的任意位置开始输入程序，一条 SAS 语句可以连续写在几行中，也可以一行写几个语句，每个语句的后面一定要用“;”结束。

例 1.3 SAS 程序的例子。

```
data sas1_3;                          /* 创建名为 sas1_3 的 SAS 数据集 */
input id name $ height weight;
bmi=weight/height * * 2;              /* 在 cards 语句之前创建分析用的所有变量 */
cards;
1 Judy 1.56 45                        /* 数据流中每个数据值之间最少有 1 个空格 */
2 Lucy 1.67 53
;
run;                                  /* 数据步结束 */
proc print data=sas1_3;               /* 打印输出 sas1_3 数据集到 output 窗口 */
run;                                  /* 过程步结束 */
```

SAS 程序编辑完成后，可选择以下任一方式提交 SAS 程序：

(1) 点击程序编辑窗口工具栏上的提交图标。

(2) 在 run 下拉菜单中选择 submit。

(3) 使用 F3 键。

注意：用户可以用光标选定部分程序进行提交。

SAS 程序提交后，SAS 会在 Log 窗口写入一些信息(一定要去读这些信息!)，这些信息是非常有用的，可以帮助你调试 SAS 程序。Log 窗口中通常包括三类信息，分别用蓝色，绿色和红色表示。

NOTE：用蓝色表示，主要是 SAS 程序运行的一般情况，一些有用的提示等。

WARNING：用绿色表示，不算错误，但 SAS 会通知你程序可能有一些问题。SAS 处理过程不会终止，SAS 仍然会创建数据集。

ERROR：用红色表示，反映程序代码有错误，SAS 将不能处理数据步，并终止对数据的处理和分析。如果你正在运行数据步来代替一个旧的数据集，那么旧数据集将不会被代替。

当运行一个 SAS 程序后，如果有输出结果，SAS 会将它写在 Output 窗口中，同时加到 Results 窗口中。用户可以通过点击工具栏上的，来清除该窗口中的内容。

(邹艳晖)

第二章　建立 SAS 数据集

第一节　SAS 数据集概述

一、SAS 数据集基本格式

SAS 创建并且处理 SAS 数据集，SAS 数据集有描述数据集的信息，如变量数、变量名、文件更新的时间、数据的长度和格式等(SAS 数据集的扩展名为 *.sas7bdat)。我们常说的 SAS 数据集多指形如图 2-1 所示的数据表。

在图 2-1 中的矩形数据表中，每一行数据值称为一个观测或称为记录，而每一列数据值称为一个变量，所以在 SAS 数据集中，每一个观测由各个变量的数据值组成。

变量

	id	name	height	weight
观测	1	Judy	156	45
	2	Lucy	167	53
	3	Susan	165	55
	4	Tony	178	62

图 2-1　SAS 数据集(数据表)的观测和变量

SAS 数据集中的变量有两种类型：数值型变量和字符型变量。数值型变量的取值只能是数值，前面可以直接加正号(+)或负号(-)表示正负值，前面加小数点(.)表示小数，前面加 E 表示科学计数法。字符型变量的取值可以是字母、特殊符号或数字。

在 SAS 中所有程序处理缺失值的形式都是相同的。如果观测中的某一个或多个变量值为缺失值，SAS 处理时会将该观测全部删去。因此，如果你的数据集中含有缺失值，应该经常性的查看 Log 窗口的信息，并在 Output 窗口打印数据列表，查看观测的数目。在 SAS 语句中，可指令数据流中“999”所代表的缺失值在 Output 窗口中以“.”表示。

例 2.1　缺失值的例子。

【程序】

```
data sas2_1;
input id name $ height weight;
if name='999' then name=.;                    /*字符型变量值缺失*/
if weight=999 then weight=.;                  /*数值型变量值缺失*/
cards;
1 Judy 156 999
2 Lucy 167 53
3 999 165 55
4 Tony 178 62
;
proc print data= sas2_1;run;
```

【结果】

Obs	id	name	height	weight
1	1	Judy	156	.
2	2	Lucy	167	53
3	3	.	165	55
4	4	Tony	178	62

二、临时 SAS 数据集和永久 SAS 数据集

在 SAS 会话中，可以使用两类 SAS 数据集：临时数据集和永久数据集。

临时数据集只能在 SAS 会话过程中创建使用，一旦退出 SAS，数据集就不存在了，因而临时数据集并不占用硬盘空间。在 **Explorer** 窗口中可以看到一个名为 **work** 的 SAS 数据库，主要存放 SAS 临时数据集。创建临时数据集时，可以用两水平的命名方式，如 **work. 数据集名**命名，也可以直接用一水平的**数据集名**命名（相当于数据库名 **work** 被省略了）。

永久性数据集储存在硬盘里，因此，在以后的每一次 SAS 会话中都可以再打开。在创建和使用 SAS 永久数据集之前，需要先建立一个 SAS 数据库来指定永久数据集存放的路径。一个 SAS 数据库相当于硬盘上的一个文件夹，可以将 SAS 数据集写入或读出。使用永久性数据集时，为避免在 SAS 程序中反复指定路径，可以给该路径指定一个名字，即 SAS 数据库名，在以后的 SAS 语句中，只要用这个数据库名就可以表示永久性数据集存放的路径。在数据步和过程步中都可以使用数据库名。

命名永久性数据集时必须用两水平的命名方式，即**数据库名. 数据集名**，数据库名实际上是一个 SAS 数据库的逻辑名。永久性数据会占用硬盘空间。

创建 SAS 数据库有两种方式：

(1) 在 SAS 程序中用 **libname** 语句，如 libname libref 'drive:\directory'。

例 2.2 创建和使用 SAS 数据库的例子。

【程序】

```
libname tj 'd:\tongji';          /* 创建名为 tj 的 SAS 数据库 */
data tj. sas2_2;                 /* 创建永久性数据集 sas2_2,保存在'd:\tongji'下 */
input id name $ height weight;
cards;
1 Judy 156 45
2 Lucy 167 53
;
run;
proc print data=tj. sas2_2;      /* 将保存在'd:\tongji'下的数据集 sas2_2 打印
run;                                输出到 Output 窗口 */
```