第 1 篇　SAS 软件及相关知识介绍

第 1 章　SAS 软件与 SAS 用法简介

由于 SAS 软件产品很多，不同产品用法不尽相同。本章仅围绕与统计学内容密切相关的方面，扼要介绍 SAS 软件及与使用 SAS 有关的基本概念和用法。本书中绝大部分内容都属于详细介绍如何用 SAS 软件实现各种统计分析及其结果解释，本章仅是这些详细内容的一个缩影。确切地说，本章重点是介绍全书中其他章节不做详细介绍的与统计学应用密切相关的一些“基本常识”和“小知识点”，为读者顺利学习其他各章做一点铺垫。

1.1　SAS 软件简介

1.1.1　SAS 软件结构

SAS 软件采取模块式结构，每个模块可被称为一个 SAS 产品。有些 SAS 产品中仅有具体的 SAS 过程（即编译后的 SAS 程序），如 SAS/STAT；有些 SAS 产品中不仅有 SAS 过程，还有其他内容，如 SAS 语言、SAS 窗口、SAS 宏、SAS SQL 等，见 SAS/BASE；有些 SAS 产品本身就是一个功能比较齐全的软件，可以完成一系列相关的功能，如 SAS/ASSIST 模块、SAS/ANALYST 模块和 SAS/INSIGHT 模块等；还有些 SAS 产品是其他 SAS 产品的开发工具，如 SAS/AF 等。

1.1.2　SAS 界面简介

不同版本的 SAS 软件，其界面不尽相同。比较新的 SAS 软件（SAS 9.3）的主界面如图 1-1 所示。

由图 1-1 可知，在主界面上，顶部有两行，分别为“菜单栏”和“工具栏”，左边有一个“SAS 资源管理器”窗口，与它重叠的一个窗口叫“结果”窗口；右边有两个窗口，上方的窗口叫“日志”窗口，下方的窗口叫“程序编辑器”窗口，通过顶部菜单条中的“窗口”选项，可以切换到其他窗口，如“输出”窗口。

可以说，基本的 SAS 窗口有“SAS 资源管理器”、“结果”、“程序编辑”、“日志”和“输出”窗口，但另外还有 30 多个窗口可供用户处理打印和微调 SAS 会话之类的操作。这些窗口的名称和窗口命令的详细列表从略。若用户想获得此列表，可通过下面的方法实现：先选中主界面第 2 行工具栏最后一个选项（图标为一本书），即帮助（help）。然后按照“帮助（help）→SAS 产品→Base SAS→SAS 窗口引用→SAS 窗口索引”的步骤显示全部 SAS 窗口列表。

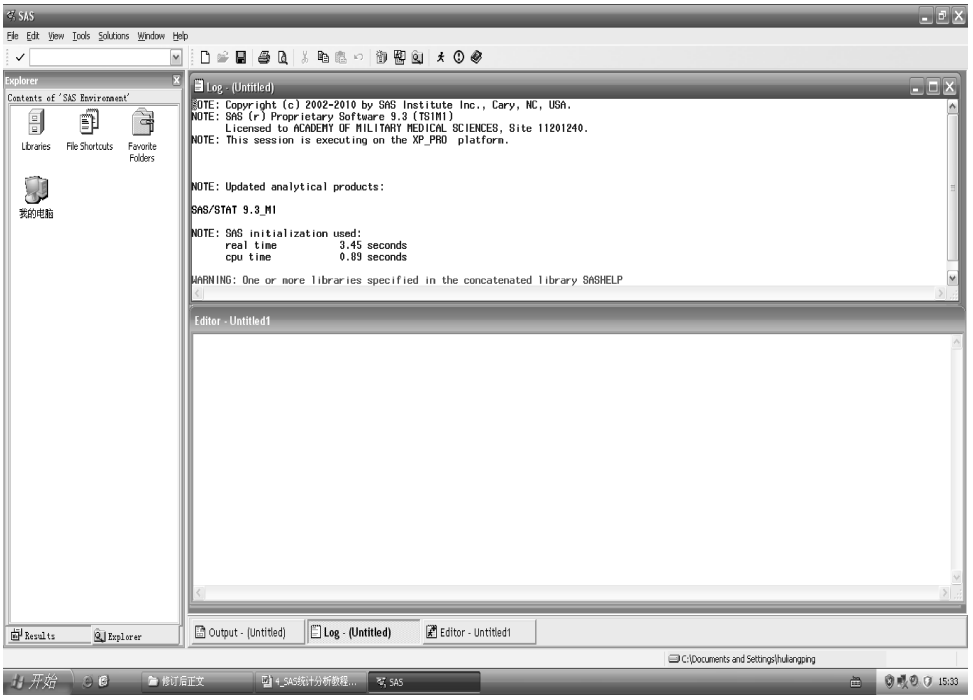

图 1-1　进入 SAS 9.3 后的主界面

1.1.3　SAS 过程与 SAS 程序

SAS 过程是 SAS 软件中经过编译后的程序，这些程序解决问题所依赖的理论和方法是被公认的，因此，可以做到标准化、程序化和系统化。然而，用户要解决的问题却是千变万化的，对于用户的数据是什么，存放在何处，都是事先无法预知的。用户在调用某个具体的 SAS 过程之前，必须将上述信息传递给 SAS 系统，这些信息必须依据 SAS 语言规则来组织，它们被称为 SAS 引导程序，简称为 SAS 程序。

1.1.4　运行 SAS 软件的两种常用方式

在 SAS 系统中运行 SAS 软件通常有两种方式。第一种是非编程法(或称为菜单驱动法)，即用户不需要编写 SAS 程序就可以直接调用 SAS 过程实现希望达到的目的。事实上，当用户通过菜单驱动系统选择某些项时，SAS 系统内部就在进行自动编程(即自动产生 SAS 程序)，当用户的选择工作结束时，SAS 程序也就全部被产生出来，故也实现了用户的目的。但并非所有的任务都能通过此法来实现。第二种是编程法，即用户亲自在“程序编辑”窗口中写 SAS 程序(或直接调用别人事先写好的 SAS 程序)并提交给 SAS 系统执行。提交 SAS 程序的方法为按图 1-1 中上方第 2 行功能键倒数第 4 个图标(一个小人图像)。

本章仅介绍如何用编程法运行 SAS 软件。

1.1.5　SAS 程序结构

SAS 程序通常由两部分组成：一部分用于提供待分析的数据，称为 SAS 数据步；另一部分调用 SAS 系统中已编译过的能处理某个具体问题的真正程序，称为 SAS 过程步。一段 SAS 程

序可以有多个 SAS 数据步和多个 SAS 过程步，有时，也可以仅有其中的一种。若仅有 SAS 数据步，就需要利用 SAS 语言编写程序，以达到某个特定的研究目的；若仅有 SAS 过程步，必须事先提供可供调用的数据(称为 SAS 数据集)。先通过下面的一个简单实例，来直观了解这些抽象的概念。

【例 1-1】 两总体率差异性检验问题。从两个同类和规模相近的大工厂里随机地各抽取 254 个相同规格的零件，其中甲厂出现的次品数为 5 个；乙厂出现的次品数为 10 个。请问：甲、乙两工厂生产的这种零件的次品率之间的差别是否具有统计学意义？

【分析与解答】 这是两样本率比较问题，更确切地说，是关于两个总体率是否相等的假设检验问题，通常称为四格表资料的统计分析。可通过在“SAS 编程”窗口中输入下面的 SAS 程序来实现，设程序名为 SASTJFX1_1. SAS。

```
DATA rate;
    DO a =1 TO 2; DO b =1 TO 2;
           INPUT f @@ ; OUTPUT;
    END; END;
CARDS;
 5  249
10  244
;
RUN;
ODS HTML;
PROC FREQ DATA = rate;
    WEIGHT f;
    TABLES a* b / CHISQ;
RUN;
ODS HTML CLOSE;
```

程序说明与修改指导：(1)SAS 数据步。从“DATA rate;”语句开始到第一个“RUN;”语句结束的这一段 SAS 程序被称为 SAS 数据步，其功能是创建 SAS 数据集，即为 SAS 系统提供待分析的数据所需要的 SAS 程序，包括与数据有直接联系的变量名。这里，变量 a 代表工厂类别，即代表两行的总名称，a = 1 代表第一行上的甲工厂、a = 2 代表第二行上的乙工厂；变量 b 代表产品检查结果，即代表两列的总名称，b = 1 代表次品、b = 2 代表合格品；f 代表与特定工厂及特定检查结果对应的产品数，即四格表资料中每个格子上的频数，如 249 代表来自甲工厂的合格品。

(2)SAS 过程步。介于“ODS HTML;”与“ODS HTML CLOSE;”两语句(其作用是使 SAS 输出结果以网页格式呈现)之间的语句段被称为过程步，即从“PROC FREQ DATA = rate;”到最后的“RUN;”语句的这段 SAS 程序被称为 SAS 过程步，其功能是调用 SAS 软件中能达到用户要求的真正程序(即 FREQ 过程)并产生相应的输出结果。

1.1.6　简单 SAS 程序中的 SAS 语句简介

全部 SAS 语句内容十分丰富，因篇幅所限，这里仅介绍 SAS 程序 SASTJFX1_1. SAS 中所涉及的几个 SAS 语句的大致含义。

每个 SAS 语句以分号(注意：必须是英文状态下的分号)结束，一行上可写多个 SAS 语句；不区分字母的大小写。上节这段 SAS 程序的数据步涉及 8 种 SAS 语句，现扼要介绍如下。

(1)DATA 语句，创建一个名为 rate 的临时 SAS 数据集(一旦退出 SAS 系统，它就消失了)。

(2)DO 语句，循环体的起始语句。

DO 语句与 END 语句成对出现，构成一个循环体。在上节 SAS 程序中，第一个 DO 语句与第二个 END 语句构成一个外循环，第二个 DO 语句与第一个 END 语句构成一个内循环。循环体的功能是创建标识性变量，同时使循环体内部的语句被重复执行规定的次数。

两个 DO 语句中的 a 和 b 被称为循环变量，DO 语句中 TO 前面的数被称为循环变量将取的最小值(或起始值)，TO 后面的数被称为循环变量将取的最大值(或终止值)。这段 SAS 程序中外循环变量执行两次，内循环变量先后各执行两次，故对循环体内的两个语句而言，总共被执行了 4 次。

(3)INPUT 语句，创建一个名为 f 的变量，通过它去读取 CARDS 语句和空语句(即独占一行的分号)之间的数据，该语句中的“@@”为指针控制符，当 INPUT 语句中变量的个数少于一行上数据个数时，必须加上这两个符号，以确保完整地读入全部数据。

若 INPUT 语句中包含字符型变量，例如，若数据中性别变量 SEX 的具体取值为 SEX = MALE 代表男性、SEX = FEMALE 代表女性，则在 INPUT 语句中，变量 SEX 应写成“SEX $”。若数据中第 1 列为人的姓名，其中最长的姓名占 18 个字符，不仅要将 NAME 写成“NAME $”，还应在 INPUT 语句前加一个定义字符型变量长度的语句，即“LENGTH NAME $ 18.;”。

(4)OUTPUT 语句，即输出语句，将 INPUT 语句变量 f 读取的每一数值送到循环体外面去(即放入计算机缓冲区中)，以免后读取的数据将先读取的数据覆盖掉。

(5)END 语句，循环体结束语句。

(6)CARDS 语句，标志着其后为数据。

(7)空语句，以一个“;”独占一行的语句被称为空语句，标志着数据行的结束。千万不要将这个分号与最后一个数据写在同一行上。

(8)RUN 语句，SAS 数据步中的该语句标志着 SAS 数据步的结束。此语句可以不写，但写上显得完整。

上节这段 SAS 程序的过程步涉及 4 种 SAS 语句，现扼要介绍如下。

(1)PROC 语句，它是 SAS 过程步开始的标志，其后跟随着一个具体的 SAS 过程名，本例中为 FREQ，它是用于分析定性资料的一个 SAS 过程，其后写“DATA = rate”，它是一个选择项，意味着待分析的 SAS 数据集名为 rate。若在此之前程序中只有一个数据集，可以不加此选择项。

(2)WEIGHT 语句，指明 SAS 数据集中哪个变量代表列联表资料中各格上的频数，本例为变量 f。

(3)TABLES 语句，指明 SAS 数据集中哪个定性变量为行变量(本例为 a)、哪个为列变量(本例为 b)，“/”代表其后为选择项，CHISQ 选项要求 FREQ 过程对 SAS 数据集中的数据进行 χ^2 检验。

(4)RUN 语句，SAS 过程步中的该语句标志着 SAS 过程步的结束，此语句不可省略。

初学者只需将自己的四格表资料中的 4 个数正确地替换掉此段程序中的 4 个数，将 SAS 程序提交给 SAS 系统执行(按 SAS 窗口上方第二行菜单条倒数第 4 个按钮，即小人像)即可。

何为正确替换？就本例而言，来自一个工厂的两个数据必须写在同一行(或列)上，而同一种检查结果必须写在同一列(或行)上。还应注意：检查结果必须按“次品与合格品”频数来表达，不能是“次品数与被检查总数”。

1.1.7　SAS 语言简介

1. SAS 语言概述

在编写像 SASTJFX1_1. SAS 那样的 SAS 程序时，所涉及的全部内容可概括为 SAS 语言。以笔者之见，SAS 语言由基本 SAS 语言(如 SAS 文件、基本 SAS 语句、常用 SAS 函数、SAS 数组和 SAS 过程等)和高级 SAS 语言(如宏、SAS SQL、SAS ODS、SAS/IML 等)两部分组成。这些内容十分丰富，需要大量篇幅才能对它们做一粗略的介绍。详细内容请查阅 SAS 说明书或从 SAS 软件的帮助窗口中查询，下同，不再赘述。

2. SAS 函数概述

SAS 中提供的常用 SAS 函数近 500 个，如求 -12.3 的绝对值，其对应的 SAS 语句为："Y = ABS(-12.3);"，也可这样写："X = -12.3; Y = ABS(X);"；再比如，求 49 的平方根，其对应的 SAS 语句为："Y = SQRT(49);"，也可这样写："X = 49; Y = SQRT(X);"。有些函数使用起来就没有这么简单了，因为那些函数中可能带有多个参数，需要搞清楚这些参数的含义，才能正确调用那些函数，获得用户需要的结果。

3. SAS 过程概述

SAS 软件中有 30 多个模块，每个模块中一般都有几十个 SAS 过程。这些 SAS 过程都能完成些什么任务，如何正确调用这些 SAS 过程，也不是轻而易举能交代清楚的。像 SASTJFX1_1. SAS 中所调用的 SAS 过程名为"FREQ"，总共写了 4 个 SAS 语句即可完成，其实，仅这一个 SAS 过程的全部语句和选择项所涉及的内容就相当多，不花上大约一周时间，可能无法全部掌握它。

1.1.8　SAS 数据集简介

1. SAS 数据集的概念

无论是采用非编程法还是编程法运行 SAS，首先都必须创建 SAS 数据集。直接在 SAS 编辑窗口录入原始数据，将其存储在某个外部设备上，只能称其为一个外部数据文件，不能称其为 SAS 数据集。何为 SAS 数据集呢？数据只有经过 SAS 系统加工后并按特定方式存储才能被称为 SAS 数据集，它将变量名及其取值(即具体数据)有机地结合在一起。

2. SAS 数据集的种类

SAS 数据集分为两类：一类被称为临时 SAS 数据集(如程序 SASTJFX1_1. SAS 中的数据集 rate)，它被存储在 SAS/WORK 库(即文件夹)中，一旦退出 SAS 系统，它就消失了；另一类被称为永久 SAS 数据集，即使退出 SAS 系统，它仍被保留。这种数据集必须被保存在非 SAS/WORK 库中，可以是 SAS 系统默认的某个库，也可以是用户自己创建的某个库(即文件夹)。

3. 创建 SAS 数据集的方法

创建临时 SAS 数据集很简单，只需要像程序 SASTJFX1_1. SAS 那样去做就可以了。下面介绍一种创建永久 SAS 数据集的方法。

【例 1-2】　创建永久 SAS 数据集的方法。假定在 D 盘上有一个文件夹名为 SASTJFX，试将程序 SASTJFX1_1. SAS 中的数据步改造成能创建永久 SAS 数据集，并将永久数据集存储在 D 盘 SASTJFX 文件夹中。

【分析与解答】　所需要的 SAS 数据步如下，设程序名为 SASTJFX1_2. SAS。

```
LIBNAME table 'D:\SASTJFX';
DATA table.rate;
    DO a =1 TO 2; DO b =1 TO 2;
           INPUT f @@ ; OUTPUT;
    END; END;
CARDS;
 5  249
10  244
;
RUN;
```

程序说明：LIBNAME 语句是创建文件关联名的重要语句，其后引号中的内容为路径(包括盘符和文件夹名)，table 是用户自己取的关联名(一般不要超过 8 个字符)，它就代表其后所写的路径。在 DATA 语句中，写“table. rate”，就意味着要创建一个名为 rate 的永久 SAS 数据集，它被存储在 D 盘 SASTJFX 文件夹中，存储后的实际数据集名为：rate. sas7bdat。

4. SAS 数据集名的种类

通常，SAS 数据集名有两类：一类被称为“一词名”，另一类被称为“两词名”。如上所述，临时 SAS 数据集为一词名(如 rate)，而永久 SAS 数据集为两词名(如 table. rate)。在“一词名”中又可细分为以下三种。

(1)用户自己取的一个词(如 rate)。

(2)直接将 DATA 语句写成“DATA；”的形式，用户未给数据集取名，SAS 系统会自动给数据集取名，当执行一次 SAS 数据步后，系统将所创建的 SAS 数据集取名为 DATA1，再运行一次，取名为 DATA2……

(3)使用 SAS 系统保留的特殊数据集名，如_DATA_、_NULL_和_LAST_。可以写成“DATA _DATA_;”，与写成“DATA;”是等价的，将给各次创建的数据集依次命名为DATA1，DATA2，…；若写成“DATA _NULL_;”则表明 SAS 系统将执行数据步，可用“PUT”等 SAS 语句输出中间结果，但观测值并不被写入 SAS 数据集_NULL_，这样可以节省计算机资源；用“DATA _LAST_;”时，表明在此之前(指自此次进入 SAS 系统以来)无论创建了多少个 SAS 数据集，只调用最后被创建的那个 SAS 数据集。

1.1.9　如何利用 SAS 帮助窗口

与 SAS 语言、SAS 过程和 SAS 用法对应的内容非常丰富，谁也记不全那么多内容。怎么办？只要用户知道一些基本的线索，就可以通过 SAS 软件提供的丰富帮助功能来查询。通过单击 SAS 窗口界面上菜单栏第一行或第二行最后一个按钮，就可进入 SAS 帮助窗口，也可在 SAS 窗口界面左上角的命令盒内发送命令，如 HELP FREQ，回车后，就可直接查询有关 SAS 中 FREQ 过程的全部信息。

1.2　SAS 用法简介

1.2.1　初学者学习 SAS 的快捷方式

上节中，在介绍 SAS 程序结构和常用 SAS 语句时已顺便介绍了如何用编程法使用 SAS。由此可见，使用 SAS 并不是一件非常困难的事。难就难在 SAS 语言(包括 SAS 语句、SAS 函数、SAS 过程、SAS 高级编程技术)内容很多，需要花很多时间去学习和实践。

笔者给初学者提供一种学习 SAS 快速入门的方法，即调用他人编制好的 SAS 程序，只要解决了“对号入座”问题(即每个程序是干什么的)，用自己的数据替换掉已有的 SAS 程序中的数据，将程序发送给 SAS 系统去执行，就可获得自己所需要的计算结果。每次结合他人的 SAS 程序和程序语句的讲解，一次学一点，不需用多长时间，自己就慢慢掌握了很多常用的 SAS 语言，也就是说，边学边解决实际问题，不仅不会望而生畏，而且见效很快。

当数据少时，直接将数据写在程序中即可，但是当数据量很大时，这样做就不够方便了，尤其是遇到第三方格式数据，SAS 是不能直接读取的。如何用 SAS 进行实验设计、进行资料表达等内容，虽然很简单，但却是必须了解的内容，下面就这些基本内容做一扼要介绍。

1.2.2　实际运行 SAS

什么叫实际运行 SAS？若拟采用非编程法运行 SAS，只需根据 SAS 说明书中所交代的步骤去“点菜单”就能获得所需要的结果，这是第一种实际运行 SAS 的方法。第二种实际运行 SAS 的方法：在 SAS 程序编辑器中输入一段正确的 SAS 程序，如将本章第 1.1 节介绍的 SAS 程序发送给 SAS 系统执行，就称为实际运行 SAS，会产生如下的输出结果：

FREQ 过程

频数 百分比 行百分比 列百分比	a	a * b 表 b		合计
		1	2	
	1	5	249	254
		0.98	49.02	50.00
		1.97	98.03	
		33.33	50.51	
	2	10	244	254
		1.97	48.03	50.00
		3.94	96.06	
		66.67	49.49	
	合计	15	493	508
		2.95	97.05	100.00

这是对输入的原始数据的详细描述，每个格内第一行为观察的频数；第二行为百分比，即以每个格上的频数为分子，以总频数为分母计算得到的相对数；第三行为行百分比；第四行为列百分比。

a * b 表的统计量

统计量	自由度	值	概率
卡方	1	1.7174	0.1900
似然比卡方	1	1.7497	0.1859
连续校正卡方	1	1.0991	0.2945
Mantel-Haenszel 卡方	1	1.7140	0.1905
Phi 系数		-0.0581	
列联系数		0.0580	
Cramer 的 V		-0.0581	

这是用四种方法分析四格表资料所得到的结果。第一种为一般 χ^2 检验，$\chi^2 = 1.7174$，$P = 0.1900$；第二种为似然比 χ^2 检验，$\chi^2 = 1.7497$，$P = 0.1859$；其他从略。

最后三行为度量列联表中行变量与列变量间关联性强弱的系数，其绝对值越接近于 1，表明关联越密切。因未对其进行假设检验，故无太大参考价值。

Fisher 精确检验

单元格(1,1)频数(F)	5
左侧 Pr <= F	0.1472
右侧 Pr >= F	0.9435
表概率(P)	0.0907
双侧 Pr <= P	0.2944

样本大小 = 508

以上是采用 Fisher 精确法分析四格表资料所得到的结果。

统计和专业结论：因 $\chi^2 = 1.7174$，$P = 0.1900$，说明两个工厂该产品的次品率之间的差别无统计学意义。虽然乙厂的次品数是甲厂次品数的 2 倍，但 2 个次品率(即 $P_{甲} = 5/254 = 1.97\%$ 与 $P_{乙} = 10/254 = 3.94\%$)之间的差别无统计学意义。说明 2 个工厂此种零件的次品率接近相等，即质量水平相当。

1.2.3　从实验设计角度谈 SAS 用法

与实验设计有关的内容可大致分为三类：其一是进行随机化(SAS 中有 PLAN 过程等)；其二是估计样本含量和检验效能(SAS 中有 POWER 过程和 GLMPOWER 过程等)；其三是给出实验设计方案(具体地说，就是与特定设计类型对应的可用于安排实验的设计表格，SAS 中有多个过程可用于此目的)。以上内容都可以通过非编程法和编程法来实现，具体做法参见本书有关章节。

1.2.4　从资料录入角度谈 SAS 用法

1. 按数据库格式录入统计资料

人们收集的科研资料往往错综复杂，但绝大部分统计资料都可表达成如表 1-1 所示的形式，它常被称为“数据库格式”的复合型统计资料。这种呈现资料的方式把每个变量在每个个体身上的具体取值都清楚地展示出来了，可以说是比较准确的原始资料。

【特例】 如何用数据库格式呈现多因素多指标资料。有人对 103 例冠心病患者(G = 1)和 100 例正常对照者(G = 2)进行了多项指标的观测，资料见表 1-1。请问：如何在 SAS 系统中录入这些资料，以便于采用 SAS 软件对数据进行各种统计表达与描述或进行各种统计分析？

表 1-1　冠心病人与正常人多项指标的观测结果

编号	组别	性别	年龄	高血压史	吸烟史	胆固醇	甘油三酯	低密度脂蛋白	高密度脂蛋白	脂蛋白 α	载脂蛋白 As	载脂蛋白 B	基因型 xbal	基因型 EcoRI	服药情况
N	G	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	1	男	60	无	无	223	205	122	30	106	0.92	0.74	-/-	-/-	未服
2	1	女	46	无	无	166	51	84	57	56	1.14	0.54	-/-	+/-	β 阻滞剂
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
203	2	男	69	有	无	224	110	58	49	132	1.10	0.96	-/-	+/+	未服

若这些数据是写在纸上的，则只能在 SAS 软件的编辑窗口内一行一行地输入数据。每行代表一位受试者的全部信息（在 SAS 中，称其为一个观测），通常，同一行上的数据之间用空格符隔开；每一列代表一个变量在不同受试者身上的具体取值。变量的含义很广，它可以代表第 1 列的序号，可以代表第 2 列的组别，…，可以代表最后一列的合并用药情况。变量代表的内容可以是辅助信息，可以是分组标志或影响因素，可以是定量或定性的观测结果。

①若在 SAS 编辑窗口内输入的仅仅是数据（包括字符型数据），将其以文本格式存入外部媒介（硬盘、优盘或软盘）上，就被称为“数据文件”。这种格式的数据文件可以在 SAS 编辑窗口中编写 SAS 程序实现调用，调用的关键 SAS 语句是 INFILE 语句和 INPUT 语句。

例如，在 SAS 程序编辑窗口中输入例 1-1 的数据，形式如下：

1　1　　5
1　2　249
2　1　　10
2　2　244

说明：第一列代表工厂编号，“1”代表甲厂、“2”代表乙厂；第二列代表产品检查结果，“1”代表次品、“2”代表正品；第三列代表各条件下的样品数。按下面的方法可将此数据存储在 D 盘 SASTJFX 文件夹内，形成数据文件，假定数据文件名为 PRODUCT. DAT：

在 SAS 程序编辑窗口左上角：文件（FILE）→Save As（另存为）→在弹出的“另存为”窗口的左上角寻找并确定路径：D/SASTJFX→在此窗口下方倒数第二行的文件名命令盒内输入数据文件名：PRODUCT→在此窗口倒数第一行的文件类型命令盒中选择“. DAT”作为文件的扩展名→单击“保存（或“确定”）”按钮。

【例 1-3】 如何将文本格式的数据读入 SAS 编程窗口。如何将数据文件读入 SAS 编程窗口，进行四格表资料的各种统计分析呢？在 SASTJFX1_1. SAS 程序中，只需将数据步修改成如下形式，而不需改动过程步。

【分析与解答】 所需的 SAS 数据步如下，设程序名为 SASTJFX1_3. SAS。

```
DATA rate;
    INFILE 'D:\SASTJFX\product.dat';
    INPUT a  b  f;
ODS HTML;
PROC FREQ DATA = rate;
    WEIGHT f;
    TABLES a* b / CHISQ;
RUN;
ODS HTML CLOSE;
```

将这段 SAS 程序发送给 SAS 系统执行，可达到同样的效果。这种使用 SAS 的方法适合数据量很大的场合。

②若在 SAS 编辑窗口内输入数据，第一行输入了变量名，从第二行开始是变量的具体取值，也将其存成数据文件，这是错误的数据文件，不能用前述的方法被 SAS 系统直接调用，只能在日后给用户提个醒，每列数据的变量名是什么。若变量名写得比较科学，其含义一看便知，则对理解这些数据能起到“备忘录”的作用，否则，没有任何价值！

③若从 SAS 窗口通过“工具→表编辑器”方式进入表编辑器窗口，在此窗口内直接输入数据，窗口第一行带有变量名（用户可修改变量名），然后，将数据存入外部设备或计算机缓存区，就成了能被 SAS 系统直接调用的 SAS 数据集了。存入 SAS 系统自动创建的逻辑库

WORK 中的数据集被称为临时 SAS 数据集，存入其他位置的 SAS 数据集被称为永久 SAS 数据集。

2. 按实验设计类型录入统计资料

人们收集完数据后，有时，习惯将它们分类整理成一张统计表，统计表的分组标志通常是定性变量，而结果变量通常是定量的。这样的数据，当属于单因素设计定量资料时，常需进行 t 检验、单因素设计定量资料的方差分析或秩和检验；当属于某种多因素设计定量资料时，常需进行相应设计定量资料的方差分析。请看下面的例子。

【例 1-4】 如何使用多个 DO – END 循环语句读取多因素析因设计一元定量数据。某实验同时涉及 A、B、C 三个地位平等的实验因素，A 分为 2 个水平、B 分为 3 个水平、C 分为 4 个水平，观测指标为 OD 值，受试对象为样品，不同实验条件下均独立地重复做了 2 个样品，资料见表 1-2。请在 SAS 编辑窗口中输入此表中的主要变量和相应的数据，以便能进行相应设计定量资料的方差分析。

表 1-2　3 个实验因素作用下 OD 值的测定结果

因素		OD 值							
A 与 B		因素 C：	C_1		C_2		C_3		C_4	
A_1	B_1		0.39	0.41	0.37	0.39	0.42	0.38	0.44	0.41
	B_2		0.37	0.36	0.43	0.45	0.41	0.37	0.42	0.39
	B_3		0.45	0.43	0.46	0.39	0.38	0.35	0.39	0.37
A_2	B_1		0.36	0.41	0.45	0.36	0.41	0.45	0.41	0.46
	B_2		0.42	0.37	0.38	0.41	0.38	0.36	0.43	0.38
	B_3		0.37	0.43	0.36	0.39	0.43	0.42	0.35	0.37

【分析与解答】 如果采取上例的方法输入数据，需要在每个定量数据前输入 4 个变量的水平代码。例如第一个数据 0.39，应当输入如下信息：1 1 1 1 0.39，这 4 个 1 分别代表因素 A、B、C 和重复实验次序都取 1 水平；同理，第二个数据 0.41，应当输入如下信息：1 1 1 2 0.41，…；最后一个数据 0.37，应当输入如下信息：2 3 4 2 0.37。这 4 个数分别代表因素 A 取 2 水平、因素 B 取 3 水平、因素 C 取 4 水平，而重复实验次序为第 2 次。显然，这样做太麻烦了，而且，很容易出错。简便的做法是，用 SAS 语言中的 DO – END 循环语句来自动产生因素 A、B、C 和重复实验次序的水平，其数据步如下，设 SAS 程序名为 SASTJFX1_4. SAS。

```
data doxunhuan;
    do A =1 to 2; do B =1 to 3;
    do C =1 to 4; do cixu =1 to 2;
           input OD @@ ; output;
    end; end; end; end;
cards;
0.39 0.41 0.37 0.39 0.42 0.38 0.44 0.41
0.37 0.36 0.43 0.45 0.41 0.37 0.42 0.39
0.45 0.43 0.46 0.39 0.38 0.35 0.39 0.37
0.36 0.41 0.45 0.36 0.41 0.45 0.41 0.46
0.42 0.37 0.38 0.41 0.38 0.36 0.43 0.38
0.37 0.43 0.36 0.39 0.43 0.42 0.35 0.37
;
run;
```

程序说明：最外层的 DO – END 循环控制表中横向上水平变化最慢的变量（因素 A），第二

层 DO – END 循环控制表中横向上水平变化较快的变量(因素 B)，这两个变量已将全部 6 行打上 A 与 B 的水平标记，即前 3 行 A 均标记为 1、后 3 行 A 均标记为 2；而 B 的标记从上到下分别为 1、2、3、1、2、3。每行有 8 个数据，先按因素 C 分为 4 组，其标记分别为 1、2、3、4，每组内再按次序(cixu)分为标记 1、2。其效果是所形成的 SAS 数据集中的排列顺序与前面用“笨方法”产生的结果一致，即(因篇幅太大，中间部分省略了)：

A	B	C	cixu	OD
1	1	1	1	0.39
1	1	1	2	0.41
…	…	…	…	…
2	3	4	1	0.35
2	3	4	2	0.37

3. 按列联表类型录入统计资料

与前面的定量资料类似，人们在表达多因素影响下的定性资料时，习惯上将数据整理成列联表的形式，特别是高维列联表资料，很少用“数据库”的形式呈现资料，见下面的例子。

【例 1-5】 如何使用多个 DO – END 循环语句读取多因素设计一元定性数据。某临床医生收集到如表 1-3 所示的资料，请在 SAS 编辑窗口中输入此表中的主要变量和相应的数据，以便能进行相应设计定性资料的统计分析。

表 1-3　甲、乙两种治疗方法对不同病程和不同病情某病患者的治疗效果

治疗方法	病程	病情	患者例数				
			疗效：	治愈	显效	好转	无效	合计
甲	短	轻		50	46	37	12	145
		重		42	35	32	23	132
	长	轻		37	30	28	14	109
		重		31	24	25	38	118
乙	短	轻		45	49	44	16	154
		重		38	43	39	22	142
	长	轻		29	38	34	19	120
		重		22	33	30	28	113

注：这个例子是假设的。

【分析与解答】 与前例相同，输入数据的方法也有两种。第一种是在每个频数前需要提供 4 个变量的标记，它们分别是治疗方法(treatment)、病程(time)、病情(degree)、疗效(effect)。这是很麻烦的事！用 DO – END 循环语句就可方便地实现上述目标，其 SAS 数据步如下，设 SAS 程序名为 SASTJFX1_5. SAS。

```
data doxunhuan;
   do treatment = 'JIA','YI'; do time = 'short','long';
        do degree = 'light','weight';
             do effect = 'zhiyu','xianxiao','haozhuan','wuxiao';
                input number @@ ; output;
   end; end; end; end;
cards;
50 46 37 12
42 35 32 23
37 30 28 14
31 24 25 38
```

```
45 49 44 16
38 43 39 22
29 38 34 19
22 33 30 28
;
run;
```

程序说明：最外层的 DO－END 循环控制表中横向上水平变化最慢的变量(治疗方法 treatment),第二层 DO－END 循环控制表中横向上水平变化较快的变量(病程 time),第三层 DO－END 循环控制表中横向上水平变化最快的变量(病情 degree)。这三个变量已将全部 8 行打上 treatment(治疗方法)、time(病程)、degree(病情)的水平标记，即前 4 行 treatment 均标记为 JIA(甲),后 4 行 treatment 均标记为 YI(乙)；time 的标记从上到下分别为 short(短)、short(短)、long(长)、long(长)、short(短)、short(短)、long(长)、long(长)；而 degree 的标记从上到下依次是 light(轻)、weight(重)交替出现；每行上的 4 列是 effect(疗效)的水平标记，依次是 zhiyu(治愈)、xianxiao(显效)、haozhuan(好转)、wuxiao(无效)，INPUT 语句中的 number 读取各行上的频数。形成的 SAS 数据集的样式与前例相似，此处从略。

1.2.5　从不同格式数据转换角度谈 SAS 用法

1. 由非统计软件创建的数据文件与 SAS 数据集之间的互相转换

若待分析的数据已采用某些第三方非统计软件(如 Excel 软件等)创建了不同格式的数据文件，其中有些可用 SAS 系统提供的导入数据接口方便地转换为 SAS 数据集(利用导出数据接口可以实现相反的操作)。请看下面的例子。

【例 1-6】　如何将一个 Excel 数据文件转变成 SAS 数据集。设有一个用 Excel 软件创建的数据文件 zhanghongleidata1. xls，该文件中有两列数据，第一列变量名为 A，第二列变量为 B，A、B 的具体取值分别是计算机导航辅助方法与 CT 方法测定每一位骨病患者置入颈椎椎弓根螺钉的相对角度的数据。共有 140 对数据，假定它们测自 140 位患者，试将其转换为 SAS 数据集。

【分析与解答】　假定用 Excel 软件创建的数据文件 zhanghongleidata1. xls 存放在 D:\SASTJFX 内，则通过如下步骤，可将其转换为临时 SAS 数据集。

①进入 SAS 系统→文件→导入数据→在窗口右边弹出一个含有命令盒的窗口。

在命令盒中显示可导入的数据文件类型为：97、2000 或 2002 年版的 Excel 软件产生的数据文件，用户可通过此命令盒最右边的三角调整拟导入的数据文件的格式。

②单击窗口下边的“Next”按钮→弹出一个小窗口，要求通过浏览方式确定拟导入的 Excel 文件的路径和文件名→选中 D:\SASTJFX\zhanghongleidata1. xls→单击“OK”按钮。

③系统询问要导入的 Excel 文件是表几(自动显示表 1，即 sheet1 $，若不是表 1，可重新选择)→单击“Next”按钮。

④弹出一个新窗口，有两个命令盒，上行为逻辑库名(自动显示临时库名 WORK，也可改变)，下行为拟创建的数据集名→输入 dao_hang_and_CT_data→单击“Finish”按钮。

⑤在窗口左边逻辑库中 WORK 库内就有刚创建的 SAS 数据集。

⑥用鼠标左键双击此数据集，可显示此数据集的内容。

将 Excel 文件显示的界面与已转换后得到的 SAS 数据集做一个直观比较。

不难发现：变量名被 SAS 系统修改了，A、B 分别被改为 F1 和 F2；Excel 文件中的第一行

数据被 SAS 系统“吃掉了”！造成这种不良后果的主要原因是在将 Excel 文件转换成 SAS 数据集的最后一步未取消系统中一个隐含的“设置”，在上面导入数据的第二步之后，会弹出一个窗口。其内的命令盒内有“Sheet1 $”。在“Sheet1 $”之下有一个“Options”按钮，单击此按钮会弹出另一个窗口。

在此窗口上的第一行是处于被选中的状态，即使用 Excel 文件中的第一行作为 SAS 数据集中的变量名。若用户在 Excel 文件中第一行输入的是数据而不是变量名，则应将第一行中复选框内的“√”去掉。去掉后单击右边的“OK”按钮，接下来的操作步骤与上面的第④步到第⑥步相同，可获得正确的转换结果。

本例还可以通过在 SAS 编程窗口中运行下面的一段 SAS 程序，实现将 Excel 文件转成 SAS 数据集的目的。设程序名为 SASTJFX1_6. SAS。

```
PROC IMPORT OUT = WORK.ZHANGHONGLEI DBMS = EXCEL REPLACE
            DATAFILE = "D: \SASTJFX \zhanghongleidata1.xls";
            SHEET = "Sheet1 $";
            GETNAMES = YES;
RUN;
```

值得注意的是，若 Excel 文件中第一行不是变量名而是数据，则上面的程序倒数第二句应改为“GETNAMES = NO”；若要转换的数据在 Excel 文件的第三张表单中，则上面的程序倒数第三句应改为“SHEET = " Sheet3 $ " ”。若转换成功，则新产生的临时 SAS 数据集 ZHANGHONGLEI 存放在 SAS/WORK 库中，可通过 SAS 资源管理器找到此库，双击此库中的 SAS 数据集名，便可将其打开，也可在编程窗口中直接调用这个临时 SAS 数据集。

2. 用 SAS 系统读入其他版本或分析软件创建的数据集

若待分析的数据已采用第三方统计软件（如 SPSS、BMDP 等统计软件包）创建了不同格式的数据集，则需要通过使用 libname 语句和在 SAS 中内置的转换程序（称为读取特定格式数据的库引擎）将特定的数据文件转换为 SAS 数据集。这种方式使用起来不很方便，下面介绍如何利用 SPSS 软件提供的文件存储功能将 SPSS 数据集转换成 SAS 数据集。

如果用户正在使用的计算机上正确地安装了 SPSS 软件，直接用鼠标左键双击 SPSS 数据文件进入 SPSS 系统并打开该文件，选择“另存为”，在弹出的存储文件的窗口内选择合适的“保存类型”并输入拟创建的 SAS 数据集名，确定后，就得到转换后的 SAS 数据集。

1.2.6　从资料表达角度谈 SAS 用法

SAS 中有些过程给出的结果并不太理想，如用 FREQ 过程生成定量资料的频数分布表，只要两个数据不完全相等，就形成两个分组标志，这样形成的频数分布表很长，显得过细，没有实用价值。实际上，可以在使用 FREQ 过程的基础上配合使用 FORMAT 过程，生成比较有实用价值的频数分布表（设程序名为 SASTJFX1_6. SAS）；还可按用户的需要去选定第一组的组段下限、组距、组数等要求，利用丰富的 SAS 语言编程，产生用户自己量身定制的频数分布表。

1.2.7　从统计分析角度谈 SAS 用法

若用户需要进行的统计分析在 SAS 软件中已有相应的过程（如单因素 2 水平设计定量资料 t 检验，有 TTEST 过程；两因素析因设计定量资料一元方差分析，有 ANOVA 过程和 GLM 过

程等；对列联表资料的分析，有 FREQ 过程和 CATMOD 过程等），此时，直接调用 SAS 过程实现统计分析就比较简单了。然而，对于某些计算问题，SAS 中尚无现成的 SAS 过程，此时，可以利用 SAS 语言并按已知的计算公式或算法编写 SAS 程序，从而实现统计分析。

总之，要想利用 SAS 解决自己的各种问题，除了要会调用 SAS 中的全部过程外，还应该全面掌握 SAS 语言。运用 SAS 语言，可以编写出当前已有但 SAS 过程解决得不满意的问题或尚解决不了的一些问题。如果用户使用某种其他编程语言（如 C 语言、C ++ 语言或 Java 语言），需要写大量的代码来设置环境、构造窗口等，而用 SAS 语言，只需把精力放在要解决的具体问题上，编程工作量就小多了。此时，用户不难体会到，SAS 的用途的确是很广的。

1.3　本 章 小 结

本章用很短的篇幅介绍了 SAS 软件和 SAS 用法。显然，此举旨在起到抛砖引玉之目的。因为 SAS 软件内容十分丰富，用 SAS 能解决的问题更是不胜枚举。本章从 SAS 软件结构、SAS 界面简介、SAS 过程与 SAS 程序、运行 SAS 软件的两种常用方式、SAS 程序结构、简单 SAS 程序中的 SAS 语句简介、SAS 语言简介、SAS 数据集简介和如何利用 SAS 帮助窗口这 9 个方面宏观地扼要介绍了 SAS 软件；接着，又从初学者学习 SAS 的快捷方式、实际运行 SAS、从实验设计角度谈 SAS 用法、从资料录入角度谈 SAS 用法、从不同格式数据转换角度谈 SAS 用法、从资料表达角度谈 SAS 用法和从统计分析角度谈 SAS 用法这 7 个方面概述了 SAS 用法。

（胡良平　李子建　刘惠刚）