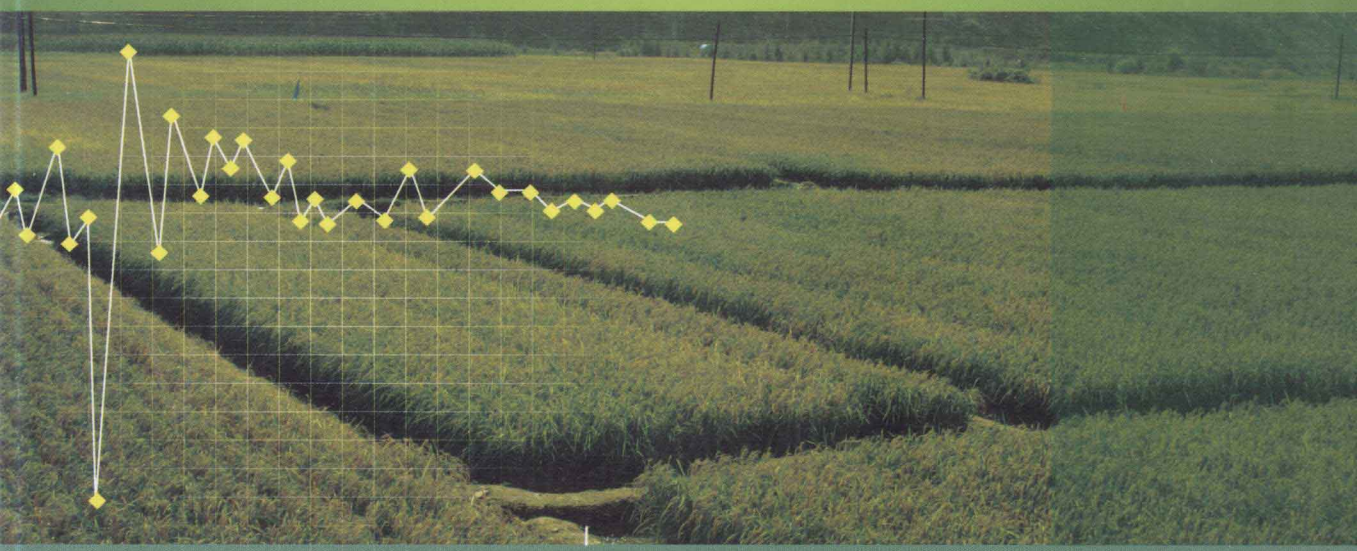
普通高等教育“十二五”规划教材

田间试验数据的计算机分析

宁海龙　主编



科学出版社

普通高等教育“十二五”规划教材

田间试验数据的计算机分析

宁海龙　主编

科学出版社
北京

内 容 简 介

本书是根据植物生产类本科专业应用型创新人才培养目标的要求及课程的学时，针对目前科学研究中常用的、基本的、重要的田间试验设计资料，介绍SAS软件和Excel软件的统计分析方法，由多所高校著名教师联合编写而成。全书共十三章，包括统计软件使用基础、试验资料的整理与描述分析、概率分布、统计假设测验、χ^2 测验、单因素试验资料的统计分析、多因素试验的方差分析、品种区域试验资料的统计分析、正交设计资料的统计分析、直线相关和回归分析、多元线性回归与相关分析、非线性回归、正交回归设计等内容。本书内容循序渐进，每种统计方法都在叙述基本原理的基础上，以实例说明SAS软件和Excel软件的分析过程，各章后都配备习题供读者练习。

本书可作为高等农林院校植物生产类专业本科生教材，也可供教师及科研人员参考使用。

图书在版编目(CIP)数据

田间试验数据的计算机分析/宁海龙主编. —北京：科学出版社，2012
（普通高等教育“十二五”规划教材）
ISBN 978-7-03-033523-4

Ⅰ.田… Ⅱ.宁… Ⅲ.①田间试验-实验数据-计算机辅助分析
Ⅳ.①S3-33

中国版本图书馆CIP数据核字(2012)第020299号

责任编辑：吴美丽 丛 楠 / 责任校对：朱光兰
责任印制：张克忠 / 封面设计：谜底书装

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

北京市文林印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2012年3月第 一 版 开本：787×1092 1/16
2012年3月第一次印刷 印张：13
字数：312 000

定价：28.00元

（如有印装质量问题，我社负责调换）

前　言

田间试验与统计方法是植物生产类专业的基础课程，也是农业科学研究的基础工具。但由于其统计方法的复杂性，该课程的学习一直很困难，并且在科研和生产实践中难以广泛应用。解决这种困难的途径是加强统计方法与计算机科学的结合，培养在理解统计方法的基础上应用统计软件分析数据的能力。目前很多院校已经开设田间试验与统计方法实验课或统计软件应用类的选修课程。应用的软件或是 SAS，或是 Excel，也有用 SPSS 的。最近几年出版的一些供本科生使用的此类教材，只在附录中介绍了部分例题的计算机分析程序，但是内容论述过于简单，涉及的统计方法不全面，对于较复杂的试验设计和统计方法的配套程序没有介绍，不能满足实际科研工作的需要。也有一些介绍统计软件的教材与著作相继出版，这些教材对统计软件的介绍全面系统，对于应用的示范例题的介绍却较少，并且这些教材侧重于医学和财经领域中应用的讲述，多不涉及农学和生物学领域的示范例题，导致非统计专业本科生和研究人员阅读困难，尤其不适合植物生产类学生和科研人员的使用。因此，编写一本面向植物生产类专业读者的田间试验数据的计算机分析教材，对于提高本课程的教学效果和推进其在农业科学研究中的应用，具有重要意义。

全书共十三章，包括统计软件使用基础、试验资料的整理与描述分析、概率分布、统计假设测验、χ^2 测验、单因素试验资料的方差分析、多因素试验的方差分析、品种区域试验资料的方差分析、正交设计资料的统计分析、直线相关和回归分析、多元线性回归与相关分析、非线性回归、正交回归设计等内容。本书内容循序渐进，第一章介绍 SAS 软件和 Excel 软件的应用基础，以后各章节在说明各种统计方法的基本原理的基础上，以实例说明 SAS 软件和 Excel 软件的分析过程，各章后都配备习题供读者练习。本书可与《田间试验与统计方法》(宁海龙，2012)配合使用，也可单独使用。

本书的第一章由东北农业大学宁海龙编写，第二章由塔里木大学张丹编写，第三章由内蒙古农业大学王树彦、于晓芳、齐冰洁编写，第四章由东北农业大学屈淑平、董玲和黑龙江八一农垦大学王霞编写，第五章由河南农业大学许海霞、董中东编写，第六章由东北农业大学李文霞编写，第七章由吉林农业大学张君编写，第八章由东北农业大学宁海龙编写，第九章由东北农业大学李文霞编写，第十章由沈阳农业大学陈志斌编写，第十一章由河南农业大学许海霞编写，第十二章由东北农业大学李文霞编写，第十三章由沈阳农业大学关欣编写。除以上编者交叉审稿外，由东北农业大学金益和河南农业大学崔党群主审，全书由宁海龙统稿。

本书可作为高等农业院校植物生产类专业及其相近专业的教材，讲授完全部内容大约需要 40 个学时，各学校可根据学生基础和教学学时数选择章节讲授。本书也可作为广大农业科技工作者、农业生产管理人员和其他专业人员的工具性参考书。

本书中引用了书后参考文献中的部分理论和图示，在此谨向有关作者表示感谢，同时也对关心与支持本书编写、出版的科学出版社编辑和编者所在院校的领导表示衷心感谢。

最后，诚挚地希望读者对书中的谬误和不足之处指正，以利于今后的修订。

编　者

2011 年 10 月

目　录

第一章　统计软件使用基础

SAS是国际著名的统计分析软件，不但具有数据管理、计算和分析的高级语言编程功能，还有基于菜单系统的SAS/Analyst(分析家)模块。Excel是电子表格软件，在管理数据的基础上，通过编辑公式、分析工具库和函数数据，也具有部分常用的数据统计功能。本章简单介绍SAS软件的编程基础和SAS/Analyst(分析家)模块操作，Excel软件公式编辑、分析工具库和函数数据操作的基础知识，为通过SAS软件和Excel软件实现试验数据的统计分析奠定基础。

第一节　SAS编程基础

SAS是“统计分析系统”(Statistical Analysis System)的英文缩写。该系统是由北卡罗来纳州立大学统计系的两位教授A. J. Barr和J. H. Goodnight于20世纪60年代末开发的。最初是以统计分析和线性统计模型为主，至今已开发成为功能强大的集成应用软件系统。本软件包括30多个工具模块，广泛应用于实用统计、运筹学、质量控制、大型矩阵计算等，是国际上公认的统计软件。目前最新版本为SAS 9.3。

本书以Windows XP操作系统下的SAS 9.0为主，介绍SAS统计功能的具体使用方法。本书所做的介绍仅供读者对SAS有一些肤浅的了解，能够使用SAS软件处理本书中的各种统计方法。若要对SAS统计功能有深入了解，请参考书后所引用的相关书籍。

一、SAS的界面操作

单击“开始”菜单“程序”项，光标移到“SAS”程序项处，选择“The SAS System for Windows 9.0(简体中文)”，单击即启动SAS。

SAS启动后在屏幕上出现的是显示管理系统(display manager)(图1-1)。

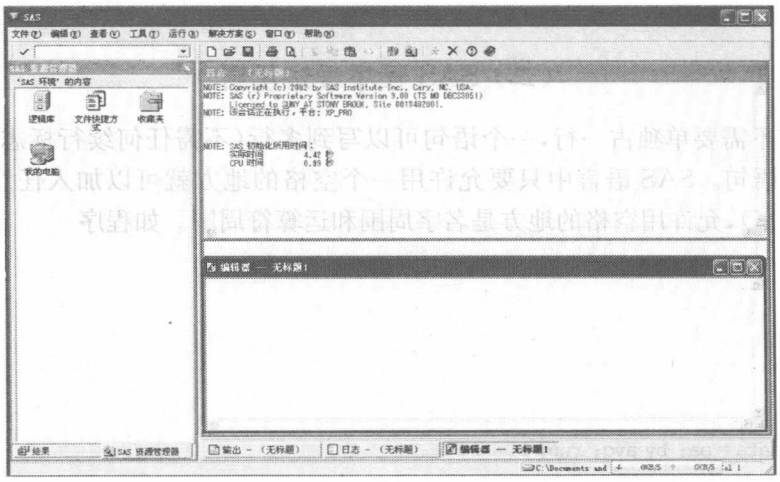


图1-1　显示管理系统窗口：编辑器窗口，日志窗口，输出窗口

显示管理系统主要有三个窗口，一个是编辑器(PROGRAM EDITOR)窗口，一个是日志(LOG)窗口，一个是输出(OUTPUT)窗口(图 1-2)。其中编辑器和日志两个窗口在启动后可直接看到，屏幕的左上角是命令框。

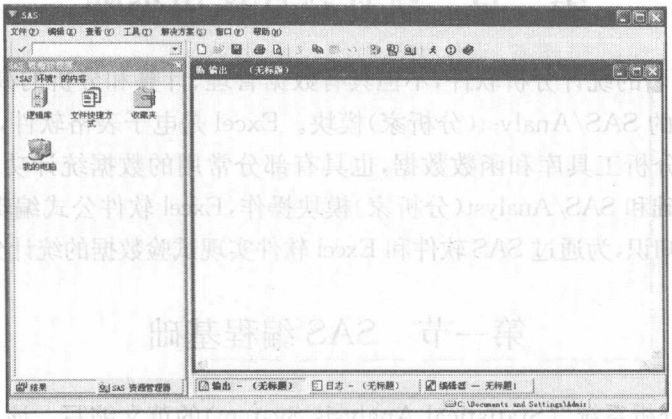


图 1-2 显示管理系统的输出窗口

上述三个窗口的主要功能如下：①编辑器窗口：输入 SAS 程序和数据；②日志窗口：显示执行程序过程中的有关信息；③输出窗口：显示程序执行的结果。

二、SAS 的基础知识

(一) SAS 语句

SAS 语言的基本单位是语句。每个 SAS 语句一般由一个关键词(如 DATA，PROC，INPUT，CARDS，BY)开头，包含 SAS 名字、特殊字符、运算符等，以分号“;”结束。SAS 关键词是用于 SAS 语句开头的特殊单词，SAS 语句除了赋值、累加、注释、空语句以外都以关键词开头。SAS 名字在 SAS 程序中标识各种 SAS 成分，如变量、数据集、数据库等。SAS 名字由 1 到 8 个字母、数字、下划线组成，第一个字符必须是字母或下划线。SAS 关键词和 SAS 名字都不分大小写。

(二) SAS 程序语法规则

SAS 语句不需要单独占一行，一个语句可以写到多行(不需任何续行标志)，也可以在一行连续写几个语句。SAS 语言中只要允许用一个空格的地方就可以加入任意多个空白(空格、制表符、回车)，允许用空格的地方是名字周围和运算符周围。如程序

```
proc print
data = aa;
by avg;
run;
```

和

```
proc print data = aa; by avg; run;
```

是等效的。

SAS 关键词和名字不分大小写，但字符型数据值要区分大小写，如“Beijing”和“BEI-

JING"被认为是不同的数据值。

在SAS程序中可以加入注释，用/*和*/在两端界定注释，这种注释可以出现在任何允许加入空格的位置，可以占多行。我们一般只把注释单独占一行或若干行，不把注释与程序代码放在同一行。注释的另一个作用是把某些代码暂时屏蔽使其不能运行。下面是一个注释的例子：

```
data split;        /* 裂区设计资料分析 */
input r a b x;
cards;
  …
;
```

（三）SAS语句的赋值

在SAS中用赋值语句计算一个值并存放到变量中。格式为"变量名＝表达式"。例如：

```
y=(x1+x2)/2;
isfem=(color='紫');
y=arsin(sqrt(x/100))*180/3.1415926;
obs=.;
```

其中，第一个赋值语句用一个公式计算平均分数；第二个生成一个取值为0或1的变量，花色为紫时为1，否则为0；第三个使用了正弦函数和乘方运算；第四个给变量赋了缺失值。

注意：想试验上述语句要把它们放入数据步中，并且等号右边的表达式中的各变量应该是存在的，否则会得到缺失值结果。

三、SAS的数据步

SAS程序包括数据步和过程步两种结构，每一个步是一段相对完整的可以单独运行的程序。SAS语言的编程计算能力主要由SAS数据步提供。数据步用来生成、整理数据和自编程计算，过程步调用SAS已编好的处理过程对数据进行处理。应用SAS编程序进行计算主要在数据步中进行。

（一）数据步的基本结构

数据步中可以使用INPUT、DATALINES(CARDS)、INFILE 、SET、MERGE等语句指定数据来源输入数据，也可以用赋值、分支、循环等编程结构直接生成数据或对输入的数据进行修改。

1. DATA语句

DATA语句以关键词DATA开头，后面给出一个数据集名，这是本数据步要生成的数据集的名字。例如：

```
data aa;
```

也可以省略数据集名，这时SAS自动生成一个临时数据集名。

2. 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＞；

在数据步中输入数据可以从原始数据输入，也可以从已有数据集输入。从原始数据输入

要使用 INPUT 语句来指定输入的变量和格式。

最简单的 INPUT 语句使用自由格式：按顺序列出每个观测的各个变量名，中间用空格分开。变量如果是字符型的需要在变量名后面加一个 $ 符号，$ 符与变量名可以直接相连也可以隔一个空格。

3. CARDS/DATALINES 语句

用于在 SAS 系统中直接输入数据，表明所列数据的开始。数据行写在 CARDS 语句（或 DATALINES 语句）和一个只有一个顶头的分号的行之间。

DATA 语句、INPUT 语句、CARDS 语句的使用可见下例。

```
data soybean;
input name $ nods  seeds SW  yield;
cards;
Heihe33        23.1     1495     22.3     333.4
Beijiao392     34.9     2305     17.6     405.7
Huajiang1      27.0     1868     17.7     330.6
Heihe45        29.9     2015     18.3     368.7
Heihe48        35.3     2000     16.5     330.0
Beiken0412     23.4     1333     20.8     277.3
Heinong37      31.1     1971     18.2     358.7
Dongnong48     29.9     2053     17.0     349.0
Hefeng45       24.5     1682     17.8     299.4
Kenfeng12      26.8     1926     17.9     344.8
Suinong14      29.5     1841     17.6     324.0
;
run;
```

注意：这个例子的数据有 11 个观测，5 个变量，每行数据的各变量之间用空格分隔。为输入这些数据，INPUT 语句中依次列出了 5 个变量名，并在字符型变量 name 后加了 $ 符。要生成一个数据集这是最简单的写法。

（二）循 环 结 构

SAS 数据步可以使用循环结构读入数据，循环结构包括计数 DO 循环，当型循环和直到型循环，应用最多的是计数 DO 循环。计数 DO 循环的写法是：

DO　计数变量＝起始值　TO　结束值　<BY　步长>；

<，参数表>???

END；

在 DO 和 END 之间可以有多个语句。程序先把计数变量赋值为起始值，如果此值小于等于结束值则执行循环体语句，然后把计数变量加上步长，再判断它是否小于等于结束值，如果是则继续执行循环体，直到计数变量的值大于结束值为止。上述结构中“BY 步长”可以省略，这时步长为 1。如果步长取负值，则继续循环的条件是计数变量大于等于结束值。例如：

```
data;
do a = 1 to 4;
do b = 1 to 3;
```

```
input x @@;output;
end;
end;
run;
```

（三）函　　数

SAS 提供了比一般程序设计语言多几倍的标准函数可以直接用在数据步的计算中，其中包括所有语言都有的数学函数、字符串函数，还包括特有的统计分布函数、分位数函数、随机数函数、日期时间函数、财政金融函数等。

SAS 函数格式为函数名(<变量 1>,<变量 2>,…)，"函数名(OF 变量名列表)，其中变量名列表可以是任何合法的变量名列表。

1. 数学函数

ABS(x)：求 x 的绝对值。

MAX(x1,x2,…,xn)：求所有自变量中的最大一个。

MIN(x1,x2,…,xn)：求所有自变量中的最小一个。

MOD(x,y)：求 x 除以 y 的余数。

SQRT(x)：求 x 的平方根。

ROUND(x,eps)：求 x 按照 eps 指定的精度四舍五入后的结果，如 ROUND(5654.5654,0.01)结果为 5654.57，ROUND(5654.5654,10)结果为 5650。

LOG(x)：求 x 的自然对数。

LOG10(x)：求 x 的常用对数。

EXP(x)：指数函数 e^x。

SIN(x)，COS(x)，TAN(x)：求 x 的正弦、余弦、正切函数。

ARSIN(y)：计算函数 $y=\sin(x)$ 在 x 取值区间为 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 的反函数，y 取[−1,1]间值。

2. 分布密度函数、分布函数

作为一个统计计算语言，SAS 提供了多种概率分布的有关函数。分布密度、概率、累积分布函数等可以通过几种统一的格式调用，格式为

分布函数值 = CDF('分布', x <, 参数表>);

密度值 = PDF('分布', x <, 参数表>);

概率值 = PMF('分布', x <, 参数表>);

对数密度值 = LOGPDF('分布', x <, 参数表>);

对数概率值 = LOGPMF('分布', x <, 参数表>);

CDF 计算由'分布'指定分布的分布函数，PDF 计算分布密度函数，PMF 计算离散分布的分布概率，LOGPDF 为 PDF 的自然对数，LOGPMF 为 PMF 的自然对数。函数在自变量 x 处计算，<，参数表>表示可选的参数表。

分布类型取值可以为 BINOMIAL，CHISQUARED，F，NORMAL 或 GAUSSIAN，POISSON，T 等。可以只写前四个字母。

例如，PDF('NORMAL', 1.96)计算标准正态分布在 1.96 处的密度值(0.05844)，CDF('NORMAL', 1.96)计算标准正态分布在 1.96 处的分布函数值(0.975)。PMF 对连续型分

布即 PDF。

除了用上述统一的格式调用外，SAS 还单独提供了常用分布的密度和分布函数。

PROBNORM(x)：标准正态分布函数。

PROBT(x,df<,nc>)：自由度为 df 的 t 分布函数。可选参数 nc 为非中心参数。

PROBCHI(x,df<,nc>)：自由度为 df 的卡方分布函数。可选参数 nc 为非中心参数。

PROBF(x,ndf,ddf<,nc>)：自由度为 ndf，ddf 的 F 分布的分布函数。可选参数 nc 为非中心参数。

PROBBNML(p,n,m)：设随机变量 Y 服从二项分布 $B(n,p)$，此函数计算 $P(Y\leqslant m)$。

POISSON((lambda,n)：参数为 lambda 的 Poisson 分布 $Y\leqslant n$ 的概率。

3. 分位数函数

分位数函数是概率分布函数的反函数。其自变量在 0 到 1 之间取值。分位数函数计算的是分布的左侧分位数。SAS 提供了六种常见连续型分布的分位数函数。

PROBIT(p)：标准正态分布左侧 p 分位数。结果在－5 到 5 之间。

TINV(p, df <,nc>)：自由度为 df 的 t 分布的左侧 p 分位数。可选参数 nc 为非中心参数。

CINV(p,df<,nc>)：自由度为 df 的卡方分布的左侧 p 分位数。可选参数 nc 为非中心参数。

FINV(p,ndf,ddf<,nc>)：$F(ndf,ddf)$ 分布的左侧 p 分位数。可选参数 nc 为非中心参数。

GAMINV(p,a)：参数为 a 的伽马分布的左侧 p 分位数。

BETAINV(p,a,b)：参数为 (a,b) 的贝塔分布的左侧 p 分位数。

4. 样本统计函数

样本统计函数把输入的自变量作为一组样本，计算样本统计量。其调用格式为“函数名(自变量 1,自变量 2,…,自变量 n)”或者“函数名(OF 变量名列表)”。比如 SUM 是求和函数，如果要求 x1,x2,x3 的和，可以用 SUM(x1,x2,x3)，也可以用 SUM(OF x1－x3)。这些样本统计函数只对自变量中的非缺失值进行计算，比如求平均时把缺失值不计入内。

各样本统计函数如下。

MEAN：均值。

MAX：最大值。

MIN：最小值。

N：非缺失数据的个数。

NMISS：缺失数值的个数。

SUM：求和。

VAR：方差。

STD：标准差。

STDERR：均值估计的标准误差，用 STD/SQRT(N)计算。

CV：变异系数。

RANGE：极差。

CSS：离差平方和。

USS：平方和。

SKEWNESS:偏度。

KURTOSIS:峰度。

注意:数据集的存储一般是每行是一个个体的观测值,每列是个体的一个属性(变量),所以统计一般应该对列进行,而不是象这里对行进行,把各变量作为一个样本的各个观测处理。这里提供的函数主要用于进行一些自编程的计算。

(四)读入外部数据

1. 文本格式的数据文件

对于小量的数据,用 CARDS 语句和分号把数据夹在中间放在数据步程序中就可以用 INPUT 语句输入数据。如果数据量很大时,一种办法是把原始数据放在一个普通的文本格式的文件中,然后用 INFILE 语句指定输入文件名。例如,可以单独生成一个文本文件"soy. txt",假设放在了"C:\SAS\"下,可以用如下程序读入文件中的数据并生成数据集:

```
data soybean;
infile 'c:\sas\soy. txt';
input name $ nods seeds 100SW  yield;
run;
proc print;
run;
```

注意:INFILE 语句要写在 INPUT 语句之前,有 INFILE 语句就不再有 CARDS 语句和分号。INFILE 关键词后面跟的是一个包含文件名的字符串,可以使用全路径名,如果只有文件名则在当前工作目录下寻找。

2. Excel 文件

SAS 可以用 Import 过程读入 Excel 文件,其语法格式如下:

PROC IMPORT

DATAFILE="数据的地址及名称"

OUT=SAS 数据集

DBMS=Excel2000 REPLACE;

GETNAMES=yes|no;

其中,DATAFILE=读入数据的地址及名称,OUT=给出要输出 SAS 数据集的名称,REPLACE 要求替换已有文件,GETNAMES=指出是否读入外部文件标题,yes 为读入,no 为不读入。例如:

```
proc import out = aa
datafile = "E:\yuandata\md16. xls"
dbms = excel2000 replace;
getnames = yes;
run;
```

四、过　程　步

(一)过程步的通用语句

能够出现在 PROC 步的 SAS 语句包括过程信息语句、变量属性语句、可用在任何地方的

全局语句。这里列出语句是 PROC 过程中最常用的一些通用语句，还有很多其他语句对不同的过程是专用的。

1. VAR 语句(变量语句)

VAR 语句被用来给出要分析的变量。该语句的格式为：

VAR 变量列表；

变量列表给出过程将要分析的数据集中的一些变量。变量列表的任意有效形式都是可以使用的。通常 VAR 语句是放在过程的开始处。VAR 语句中的变量顺序，也是将来输出结果时的变量顺序。

2. MODEL 语句(变模型语句)

MODEL 语句被用来规定分析的模型。该语句的格式为：

MODEL 因变量列表＝自变量列表 </选项>；

3. WEIGHT 语句(权数语句)

WEIGHT 语句用来规定一个变量，它的值是这些观测相应的权数。该语句的格式为：

WEIGHT 变量；

WEIGHT 语句常常用在这样一些分析中，如与每个观测有联系的方差不等时，那么可引入一个权数变量，其值和方差的倒数成比例。

4. FREQ 语句(频数语句)

FREQ 语句用来规定一个变量，它的值表示这个观测出现的频数。该语句的格式为：

FREQ 变量；

如果在某个观测中，FREQ 变量的值小于 1，这个观测在分析中不使用；如果 FREQ 变量的值不是整数，仅取整数部分使用。

5. ID 语句

ID 语句用来规定一个或几个变量，它们的值在打印输出或这个过程产生的 SAS 数据集中用来识别观测。该语句的格式为：

ID 变量列表；

使用了 ID 语句后，最左边的 OBS 列被取消了，且 ID 语句所指定的变量被排列在输出结果报告的最左边。例如，当一个 ID 语句同 PRINT 过程一起使用时，输出的观测用 ID 变量的值来识别，而观测本来的序号没有被打印输出。

6. CLASS 语句

CLASS 语句用来指定一些分类变量，SAS 过程按分类变量的不同值分别进行分析处理。该语句的格式为：

CLASS 变量列表；

7. BY 语句

当用户要求 SAS 系统对数据集进行分组处理时，可在 PROC 步中使用 BY 语句。但处理过程要求数据集事先已经按 BY 变量排序好了。该语句的一般格式为：

BY<DESCENDING>变量 1<…变量 2><NOTSORTED>；

DESCENDING 选项表示它后面的一个变量按降序排列。要特别注意 BY 后面的变量排列的先后次序，表示分组的先后次序。例如，有一个关于大豆品种的数据集，我们要育种的生态区(Loca)降序排列，同一生态区中按叶型(Leaf)的升序排列。BY 语句的使用格式为：

BY DESCENDING Loca Leaf；

NOTSORTED选项并不是说数据不要求排序，而是要求数据按组整理，并且这些组不必按字母顺序或数值顺序排序。

但如果要处理的数据集事先没有按BY变量的升序排序，可在SORT过程中用相同的BY语句对观测进行排序。

例如，我们有一个没有按任何变量排序过的soybean数据集，现在要想按熟期(Sq)分组显示观测的Name和Yield变量的内容。程序如下：

```
libname Study "d:\sasdata\mydir";
proc sort data = soybean;
by Sq ;
proc print data =  soybean;
by Sq;
var Name Yield;
id Leaf;
run;
```

CLASS语句与BY语句是有所区别的。CLASS语句使用时，不要求数据集事先按CLASS指定的变量排序，按指定变量的不同值进行分类计算和分析后，输出的分类结果列在一张报表里。而BY语句在使用时，要求数据集事先按BY指定的变量排序，且输出的结果也按分组列出许多报表。

SAS的过程步就是已经编好了的用于实现各种统计分析功能的计算机程序，只要按照规定好的格式调用就可以了。过程步总是用一个PROC语句开始，后面紧跟着过程步名，用以区分不同的过程步。表1-1是本书涉及的过程步的名称及功能。

表1-1　PROC过程步的名称及功能

过程名	功能
PRINT	显示数据集的变量名及变量值
SORT	对指定变量进行升序、降序排列
MEANS	对数值型变量进行描述分析
UNIVARIATE	对数值型变量进行描述分析和正态分布适合性测验
FREQ	对次数资料进行卡平方测验
GCHART	在“Graph”窗口中对指定变量绘制图形
TTEST	进行单个样本和两个样本的t测验
ANOVA	进行方差分析
GLM	进行方差分析、协方差分析
CORR	进行变量间的相关分析
REG	进行线性回归分析
RSREG	进行正交回归分析

(二) FREQ过程

FREQ过程语法如下。

PROC FREQ 选项表：

TABLES 请求式</选项表>;

WEIGHT variable </选项表>;

BY <DESCENDING> 变量 1 …<DESCENDING>变量 n <NOTSORTED>;

OUTPUT statistic-keywords <OUT=SAS-data-set>;

PROC FREQ 语句为必须语句,其他语句为可选语句,且该过程只能使用一个 OUTPUT 语句。

1) PROC FREQ

常用的选择项有:

DATA=SAS-dataset(SAS 数据集)

PAGE 要求 FREQ 每页只输出一张表。否则按每页行数允许的空间输出几张表。

2) TABLES

TABLES 请求式:请求式由一个或多个由"*"号连接起来的变量组成。一维表由一个变量产生;二维表由"*"隔开两个变量,任何数量的变量能被"*"连起来得到多维的表格。

二维频数表:在 TABELS 语句中用星号"*"连接两个变量。第一个变量的值构成表的行,而第二个变量的值构成表的列。例如:

```
PROC FREQ;
TABLES  A * B;
```

产生一个列联表,A 的值构成表的行,B 的值构成表的列。

(1) TABLES 语句的选择项列表如下。

MISSING——像分析非缺项值那样分析缺项值,且在百分数计算和其他统计计算时包括缺项值。若 TABLES 语句中没有规定该选择项,则 FREQ 过程产生的列联表中每一变量的缺项值从表中删除,但缺项的总频数在每个表下面输出。

LIST——不是用列联表而是用列表格式打印二维或多维表格。但当需要统计检验和联合测量时,不能使用 LIST 选择项。

(2) 请求统计分析的选择项如下。

CHISQ——请求卡方(χ^2)检验和基于卡方的有关测量。检验包括 Pearson 卡方、似然比卡方和 Mantel-Haenszel 卡方。测量值包括斐(phi)系数、列联系数和克莱姆系数(Cramer'v)。对于 2 * 2 表也包括费雪尔(Fisher)精确检验。

FISHER——要求对大于 2 * 2 的表进行 Fisher 精确检验。

其他:此外还有 CMH、CMH1、CMH2、ALL、MEASURES、ALPHA= 等选择项。

(3) 请求增加表格信息的选择项如下。

EXPECTED——请求打印在独立(或齐性)假设下的期望格频数。

DEVIATION——请求打印出各格的格频数和期望值的偏差。

CELLCHI2——请求打印出每一格对总χ^2 统计的贡献。

CUMCOL——请求在格中打印累计列百分数。

MISSPRINT——要求打印缺项值频数。

SPARSE——使过程打印出在请求表中各个变量水平的所有可能组合的信息。即使某些水平的组合不在数据中,此选择项影响在 LIST 选择项下的打印输出和输出的数据集。

(4) 禁止打印选择项如下。

NOFREQ—— 禁止打印列联表中的格频数。