

普通高等教育“十二五”规划教材

SAS

生物统计分析应用教程

周海龙 周开兵 主编

SAS

9.13 版



化学工业出版社

普通高等教育“十二五”规划教材

SAS

生物统计分析应用教程

张永华 张华 主编

SAS

应用教程

北京：科学出版社

普通高等教育“十二五”规划教材

SAS

生物统计分析应用教程

周海龙 周开兵 主编



化学工业出版社

· 北京 ·

图书在版编目 (CIP) 数据

SAS 生物统计分析应用教程/周海龙, 周开兵主编.

北京: 化学工业出版社, 2011. 6

普通高等教育“十二五”规划教材

ISBN 978-7-122-11197-5

I. S… II. ①周…②周… III. 生物统计-统计分析-应用软件, SAS-高等学校-教材 IV. Q-332

中国版本图书馆 CIP 数据核字 (2011) 第 079232 号

责任编辑: 刘 畅 赵玉清 吴 俊

装帧设计: 史利平

责任校对: 战河红

出版发行: 化学工业出版社 (北京市东城区青年湖南街 13 号 邮政编码 100011)

印 装: 北京市兴顺印刷厂

710mm×1000mm 1/16 印张 9½ 字数 175 千字 2011 年 6 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888 (传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书, 如有缺损质量问题, 本社销售中心负责调换。

定 价: 25.00 元

版权所有 违者必究

前 言

SAS 系统是全球最权威的统计分析软件，为大型集成信息分析管理系统，其集数据访问、数据管理、数据分析、数据挖掘功能于一体，为国际公认的国际标准统计分析软件。信息技术的迅猛发展使得数据管理和数据分析成为各领域发展的重要支柱，SAS 系统顺理成章地成为数据处理的首选工具。我国日益增多的 SAS 用户已充分认识到 SAS 系统强大的功能，SAS 软件应用于科研数据处理也被认为是发表高水平学术论文和提高试验结果可靠性的重要工具。

近 7 年来，笔者在海南大学为生命科学、农业科学和管理学专业研究生、本科生讲授《生物统计学》和《SAS 软件在统计学中应用》等课程，同时为校内外科研工作者提供科研数据统计分析咨询服务，从中积累了许多教学和应用 SAS 软件的经验，勤于思索而受到许多启迪。尽管一些不熟悉 SAS 软件应用的学生和科研工作者可能熟练掌握了统计学原理和基本计算步骤，但是，在数据处理分析实践中由于计算量过大，致使其只能应用较为低级的试验设计方法和对数据信息作出极其有限的分析处理，不能提高工作效率和浪费许多数据信息。还有一些在计算机理论和操作上拥有天赋的学生和科研工作者因为统计学知识掌握得不扎实，一样不能灵活和合理应用 SAS 软件。因此，为了适应生命科学、农业科学和管理学等学科的发展，必须将统计学和 SAS 软件的应用完美结合，这就要求学生和科研工作者既要了解统计分析方法的原理和主要计算步骤，又要了解 SAS 软件主要过程的功能和分析结果的判读技巧，甚至学会编写程式技能。在学习和应用要求上，应该将通过传统数学计算方法和通过应用 SAS 软件的数据统计分析方法的优势进行互补，克服各自缺点，使统计学和 SAS 软件成为科研工作的重要手段。即通过学习统计学来熟练掌握统计分析方法的数学原理；通过应用 SAS 软件而避免记忆有关公式含义、推导和证明过程，同时增强计算结果的正确性和准确性，极大程度加快计算速度和完成繁杂的计算任务。正是掌握了 SAS 软件应用技能，才促进试验设计方法的发展和提高试验设计水平，从而提高科研工作效率和学术水平。

通过学习 SAS 软件应用，可以减轻学习统计学的负担，但是，要精通 SAS 软件应用技能，需要对 SAS 软件程式和运行结果进行理解，做到程式能灵活变化和正确应用，输出结果能够抓住重点和正确判读，这就要求必须对 SAS 软件的应用知识展开系统的学习。国内外有很多 SAS 软件应用学习丛书，但是，大

多数书目所讲述内容较为庞杂，与统计学教材出现很多内容重复，未注意到科研工作中常用的统计分析方法而重点不突出，给读者尤其是 SAS 软件应用的初学者带来许多麻烦，甚至于不能正确应用 SAS 软件。笔者结合多年的教学和科研实践，与当前高校主打的生物统计学教材相匹配，力求编写出通俗易懂、深入浅出、详略得当、案例经典和科研思想清晰的实用性 SAS 软件应用教程。全书共 15 章，第一章是 SAS 软件系统的基础入门篇；此后，自绘图应用开始，直至各种常用多元统计分析方法，由浅入深、层层递进。只需了解统计方法的基本原理，就可以理解程式结构精髓和正确判读输出结果，为正在高校学习的生命科学、农业科学和管理学等专业的研究生和本科生，以及广大科研工作者提供合适的教学用书和工具书。

本书在 7 年的内部讲义的基础上，不断修改完善而成，部分章节由海南大学汤华博士编写。本教材得到了“211 工程”热带作物遗传育种与生态保育项目、转基因生物新品种培育重大专项（编号：2008ZX08011-001，2009ZX08011-010B）、海南大学科研基金资助项目（项目编号：kyqd1102）的资助；同时，得益于海南大学副校长胡新文教授、海南大学农学院副院长袁潜华研究员的关心与指导，也得益于化学工业出版社的支持与帮助，在此笔者对他们谨致谢忱！

由于编写时间仓促、笔者水平所限，书中难免有误，敬请各界同行和读者批评指正！

编者
2011 年 4 月

目 录

第一章 SAS 系统简介	1
第一节 SAS 系统主要窗口	1
第二节 SAS 编程基础	2
第三节 SAS 程式常用语句	5
第二章 绘图	7
第一节 绘制散点图	7
第二节 绘连线图程式	9
第三节 散点图和连线图在同一坐标系中输出	10
第三章 描述性统计	12
第一节 Means 过程	12
第二节 Univariate 过程测验变数的正态分布适合程度	14
第四章 T 检验	18
第一节 单样本平均数为常数的测验程式	18
第二节 两个样本成组数据的比较分析程式	19
第三节 成对数据的比较分析程式	22
第五章 卡方检验——Freq 过程检验独立性	24
第六章 单因素试验的方差分析	27
第一节 完全随机试验设计	27
第二节 完全随机区组试验设计	33
第三节 拉丁方试验设计统计分析程式	36
第四节 巢式试验设计统计分析程式	38
第七章 双因素试验统计分析	41
第一节 不考虑交互效应的双因素试验	41
第二节 考虑交互效应的完全随机试验设计的双因素试验	43
第三节 双因素完全随机区组试验设计的统计分析	46
第四节 裂区试验设计统计分析	49
第五节 条区试验设计统计分析	52
第八章 三因素试验统计分析	59
第一节 三因素完全随机试验统计分析	59


第二节	三因素完全随机区组试验统计分析	63
第九章	正交设计试验统计分析	68
第一节	不考虑交互效应的正交试验的方差分析	68
第二节	考虑交互效应的正交试验方差分析	72
第十章	一元线性回归与相关分析程式	77
第一节	一元线性回归分析程式	77
第二节	一元线性相关分析	79
第三节	协方差分析	80
第十一章	多元线性回归与相关性分析	90
第一节	多元线性回归分析	90
第二节	多元逐步线性回归分析	93
第三节	多元线性相关性分析	98
第四节	两组变量的典型相关性分析	103
第十二章	一元非线性回归分析	109
第一节	一元多项式回归分析	109
第二节	Logistic 曲线方程的分析	112
第三节	其他一元曲线回归分析	114
第十三章	多元聚类与判别	117
第一节	多元聚类分析	117
第二节	成批调整法多元聚类分析	121
第三节	Bayes 判别分析	124
第四节	逐步判别程式分析	130
第十四章	主成分分析法	134
第十五章	因子分析	138
参考文献	143

第一章 SAS 系统简介

SAS (statistical analysis system) 于 1966 年由美国北卡罗莱州 (North Carolina) 大学开发研究, 历经十年, 于 1976 年完成该系统, 同时成立 SAS 研究所, 负责该系统的维护、继续开发、销售和 SAS 教育等工作。现在, 已广泛应用于统计分析的诸多领域, 具有较强的语言编程能力、统计方法丰富、可对数据进行连续处理、较强的报表输出能力、信息存储简单方便、强大的在线帮助、方便的视窗操作和系统设置以及强大的文本和图形处理能力等特点。

第一节 SAS 系统主要窗口

一、SAS 启动

进入 WinXP 系统后, 在桌面上将看到 “” 图标。将鼠标移到图标上双击, 则很快出现 SAS 启动界面, 等待几秒钟, 进入 SAS 环境。或通过开始程序菜单启动 SAS。

二、SAS 窗口

进入 SAS9.13 后, 其窗口主要有 Editor、Output、Log、Explorer、Results 等窗口, 点击窗口条上相应按钮即可使其成为当前活动窗口, 如图 1-1 所示。

1. **Editor 窗口** (SAS 程序编辑窗口): 该窗口主要对 SAS 程序进行打开、编辑、调试、存储、打印等。另外, 该窗口可放大缩小, 可移动, 可全屏幕编辑; 具有可见、可写、可存储、可打印。

2. **Output 窗口** (SAS 结果输出窗口): 主要显示运算结果。该窗口特性为: 出现在此窗口的所有信息只可看, 不可在本窗口内直接修改, 但可将所有内容进行打印或存储。即可看, 可打印, 但不允许编辑。

3. **Log 窗口** (SAS 日志窗口): 主要显示监控 SAS 程序的运行情况, 包括系统信息, 执行某段程序所需时间。语句运行中出现的各种错误 (用红色字表示), 并加上一定的注解。该窗口特性同 Output 窗口。

4. **Results 窗口**: 主要用于管理和操作 SAS 程序运行后的输出结果, 其内容与 Output 窗口相对应, 可查看、保存、删除或打印结果。

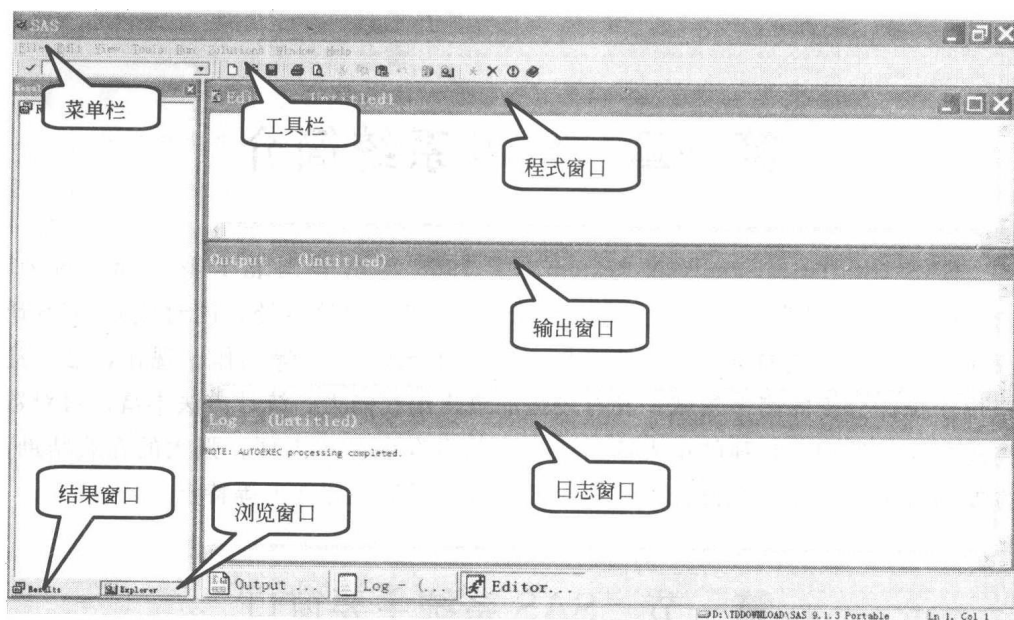


图 1-1 启动 SAS

5. **Explorer 窗口**：主要用于管理和查看 SAS 文件，类似于 WindowsXP 中的资源管理器。

其他窗口的作用，功能等详细介绍可通过按 F1 键进入 SAS 帮助系统获得。

第二节 SAS 编程基础

一、SAS 程式结构

SAS 系统中，一个完整的处理过程均包括：数据步和过程步。

1. **数据步**：将数据读入 SAS 系统，且每一个数据步均由 Data 语句开始，以 Run 语句结束。

2. **过程步**：调用各种已编好的过程来分析由数据步建立的数据集中的数据。每一个过程步均以 Proc 语句开始，Run 语句结束。且每个语句后面一定要加上“;”。例如：

一个简单完整的 SAS 程式可能只有几行，而复杂的 SAS 程式有数十至数百行。

例.

```
data zkb;                /*数据集名称为 zkb*/
  a=20;                  /*变量 a=20*/
```

```
b=100* a;          /*计算变量 b=100* 20*/  
c=200* b;  
d=350* c;  
  
run;  
proc print;        /*在 Output 窗口中输出数据*/  
run;
```

由此可见，SAS 程式具有以下 3 个特点：

- (1) 以 SAS 关键字开始，如：data, proc, cards, input, class 等；
- (2) 单个语句都是以分号“;”结尾，数据步中不能有分号“;”出现；
- (3) 以 run 结尾。


当一个程式中有多个数据步时，由于后一个 Data 或 Proc 语句具有前一步的 Run 语句作用，因此，可省略两步中间的 Run 语句，但程式最后一步必须有 Run 语句，否则程式运行结果要等到下一个程序运行时才能出来。

二、程式的输入、编辑和运行

SAS 程式只能在 Editor 窗口输入、编辑，并写在 PGM 窗口预先设置好的行号区的右边。程式语句可使用大写或小写字母或同时使用来，每个语句中单词或数据间以空格隔开。每行输入完后加上“;”再按回车。但在数据步中 Cards 语句必须独占一行，并且其下所有数据行不能加“;”，必须等到数据输入完后提行单独加“;”。在输入过程中可对错误进行修改。

SAS 语句书写格式自由，不区分大小写，可在各行的任何位置开始。一个语句可以连续数行，一行中也可写多个语句，SAS 语句中的关键字可使用空格或特殊字符分割开，每个语句后面必须用“;”隔开。

另外，若要在 Editor 窗口中回忆当前打开 SAS 后所有运行过的程式，点击 run，选中 recall last submit 并点击即可；也可以对现有程式进行修改。当一个 SAS 程式输入、编辑完毕后，有以下 3 种方式提交 SAS 程式：

1. 点击 Editor 窗口工具栏上的提交图标 。
2. 使用快捷键 F3。
3. 在 Run 下拉菜单中选择 Submit (见图 1-2)。

如果程式运行出错，则需要在 Editor 窗口对程式行修改，然后再继续提交，直至得到结果满意为止。下面是例 1 程式运行后在 Output 和 Log 窗口中的结果 (图 1-3、图 1-4)。

SAS 程式提交后，Log 窗口中的信息非常有用，可帮助调试 SAS 程式，主要有以下几种内容：

- (1) 程序行用黑色表示：以顺序行号 1 2 3...开始，记录了 SAS 程式运行的每一条语句。

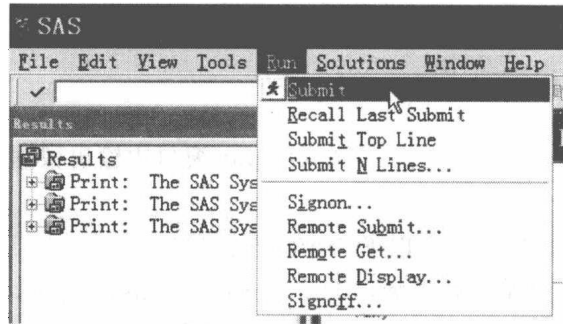


图 1-2 提交 SAS 程式

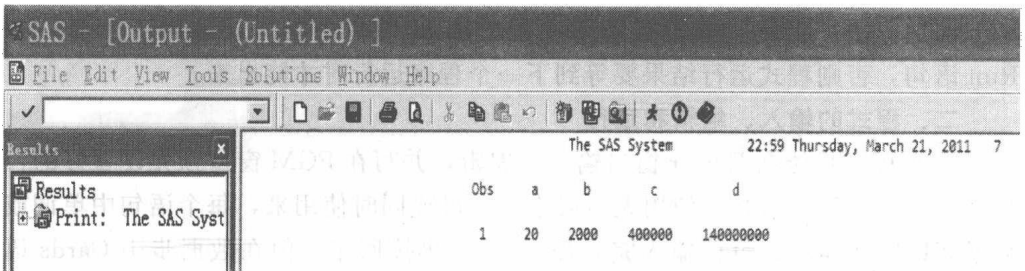


图 1-3 SAS 运行结果

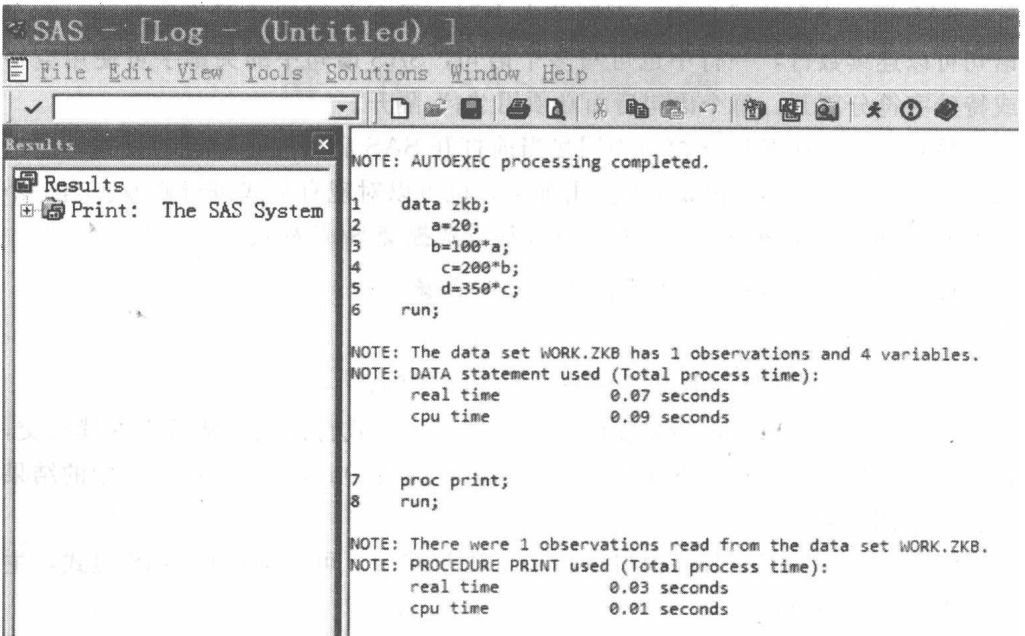


图 1-4 SAS Log 窗口

(2) NOTE 用蓝色表示：提供程序运行的一些常规信息，是一些有用的提示。

(3) WARNING 用绿色表示：一般在程序中含有系统可以自动更正的小错误时出现，此时会提供错误序列号。出现警告时应仔细阅读，如程式有误应立即修改。

(4) ERROR 用红色表示：说明程式代码有误，SAS 不能处理数据步的数据集，终止对数据的分析和处理。这时需要查询程式错误，并在 Editor 中修改。

本例中 Log 窗口中有 4 条提示信息，第一条是数据中的观测值个数和变量数，第二条则标出了程式运行所需时间。

第三节 SAS 程式常用语句

SAS 系统管理和分析数据时，用户需要先利用 SAS 语言编写相应的指令，称为 SAS 程式。而 SAS 语句是指要求 SAS 系统执行某种操作的命令。

一个 SAS 语句通常由 SAS 关键字数据步准备好数据后，SAS 过程步对数据集进行分析处理。而过程步中语句多而复杂，不同过程有其选项不同，下面先简单介绍一些常用语句。

一、Proc 语句

该语句是 SAS 程序过程步的开始语句。

1. 格式：Proc <SAS 过程名> [选择项]；

2. 功能：指定要调用的过程和该过程的一个或多个选项。

3. Proc 语句的选择项有三种类型：

(1) 关键字 规定反映本过程特征的关键字。

(2) 关键字=值 规定初值，该值可是数值或字符串。

(3) 关键字=SAS 数据集规定输入或输出的数据集。常用的是 DATA=数据集，指出本过程所要处理的数据集名，如果缺省则处理最新建立的数据集。

而需要掌握和了解的主要过程有 Means/Plot/Gplot/Univariate/Anova/Glm/Reg/Corr/ Cluster/Discrim/Princomp/Factor/Cancorr 等，在后续章节中将分别介绍。

例如，我们在前面的例子中已多次使用的语句：

Proc Print；就是一个过程调用语句，调用的是 SAS 系统的 Print 过程。

二、Var 语句

该语句指定过程要分析的数据集中哪些变量，如果该语句缺省，则过程对数据集中的全部变量均进行统计分析。

1. 格式：VAR 变量名表；

2. 功能：定义过程分析用变量。

三、By 语句

By 语句常和 Sort 过程一起使用，对数据集进行排序。另外，By 语句用在其他过程中时，能分别处理每一个分组（By）观测值。

1. 格式：By [Descending] 变量名…… [Notsorted];

2. 功能：以指定的变量值来进行分组处理数据集。

选择项 Descending 要求紧接着的变量按降序排列数据集中的各观测值，缺省时为升序排列。选择项 Notsorted 要求进行分组时，各组不需要按数字或字母顺序进行排序。

四、Class 语句

分类变量指少数几个离散值或不连续值。可以是字符型也可是数值型。例如，性别、年龄组、年度、地区、试验小区、试验组等。

1. 格式：Class 变量名表；

2. 功能：定义分类变量。

在 Class 语句中按指定变量值进行分组。Class 语句功能与 By 语句类似，二者间的区别仅在于打印输出的格式不同，By 语句要求先排序，而 Class 语句不要求任何排序处理。

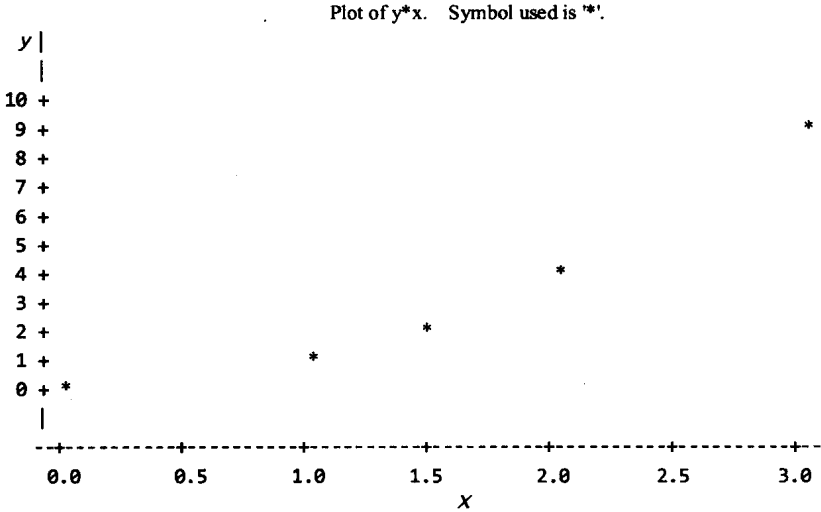
五、Output 语句

指明关键字在新的数据集中想要的统计项，而且包括给这些统计命名。

1. 格式：Output Out = 数据集名 [统计关键字 = 变量名];

2. 功能：将过程结果输出到一个新的 SAS 数据集。

2. 运行结果



3. 程序解释与结果说明

① 在 proc plot 后, hpercent 和 vpercent 分别控制图形输出的宽度和高度, 图形占页面的宽度和高度百分比在等于号后给定, 空格后的数字说明页面中图形个数, 0 表示输出一个图。在 plot 表达式后, vaxis/haxis=数值, 规定纵横轴刻度, to 前数值为起始刻度值, to 后数值为最大刻度值, by 后为刻度单位数值; Vtoh=数值, 规定纵横轴比例。

② 数据步中, 表达式 $y=x*x$ 可用所考察问题的表达式替代, 数据表的自变量数据用所考察问题的数据加以替代; 过程步中表示点的符号可以自定义; 修改后则可输出所考察问题的散点图。

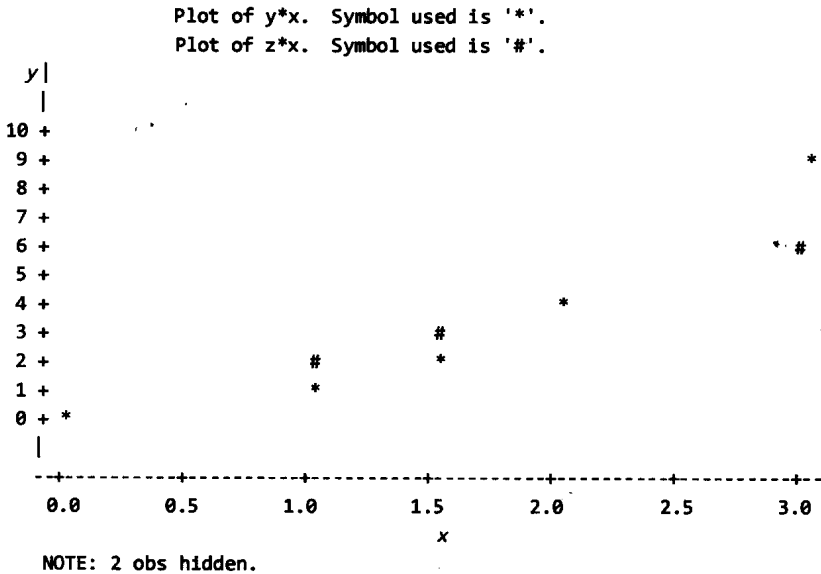
③ 程式语句的基本格式不宜随意变动, 编写程式时务必注意。

三、在同一坐标系中绘制 $y=x^2$ 和 $y=2x$ 的散点图

1. SAS 程式

```
data ex; do a=1 to 5; input x @@;
y=x*x; z=2*x; output; end;
cards;
0 1 1.5 2 3
;
proc plot hpercent=80 0 vpercent=50 0;
plot y* x='*' z* x='#'/overlay vaxis=0 to 10 by 1;
run;
```


2. 运行结果



3. 程序解释与结果说明

说明：在 plot 表达式后，“/Overlay” 说明图形叠放；其他同上述。

四、练习题

在同一坐标系中绘制散点图： $x=0\sim 360$ ， $y=\sin(3.14159x/180)$ 和 $y=(2x/180)-2$ 的散点图。

第二节 绘连线图程式

1. SAS 程式

```
data ex; do a=1 to 5; input x @@;
y=x* x; output; end;
cards;
0 1 1.5 2 3
;
symbol v=star i=spline;
proc gplot;
plot y* x;
run;
```