




全国高等农林院校“十一五”规划教材

# SAS统计分析教程

唐燕琼 主编

 中国农业出版社

全国高等农林院校“十一五”规划教材

# SAS 统计分析教程

唐燕琼 主编

中国农业出版社

## 图书在版编目 (CIP) 数据

SAS 统计分析教程 / 唐燕琼主编. — 北京: 中国农业出版社, 2006. 12

全国高等农林院校“十一五”规划教材

ISBN 7-109-11482-1

I. S… II. 唐… III. 统计分析-应用软件, SAS-教材  
IV. C812

中国版本图书馆 CIP 数据核字 (2007) 第 008760 号

中国农业出版社出版

(北京市朝阳区农展馆北路 2 号)

(邮政编码 100026)

责任编辑 张 利

中国农业出版社印刷厂印刷 新华书店北京发行所发行

2006 年 12 月第 1 版 2006 年 12 月北京第 1 次印刷

开本: 720mm×960mm 1/16 印张: 23

字数: 407 千字

定价: 29.60 元

(凡本版图书出现印刷、装订错误, 请向出版社发行部调换)

主 编 唐燕琼

副 主 编 韩汉鹏

编写人员 薛文龙 王 英 唐燕琼 韩汉鹏

主 审 林德光

# 前 言

展现在读者面前的这本《SAS 统计分析教程》，是一部建立在当代科技水准上的新教材。编者站在当代统计的前沿高度，系统介绍了 SAS 技术的最精辟成就及具体操作方法。取材精湛、理论完备、循序渐进、深入浅出、叙述清晰为其写作特点。此外，各部分内容都赋予实例示范，详细阐述编程过程及输出结果的解释。既有理论意义，又有实用价值。

编者以国际上最新 SAS9.0 版本为首选软件，介绍了 SAS 系统在 Windows 操作系统下的安装、启动及众多基础运算规则。编者是假定读者接受过计算机基础训练、修完统计学基础知识为前提的。因此，对于未掌握这两方面基础知识的读者，必须尽快补上，以便顺利进入 SAS 学习。

本教材的核心是学会各种经典试验设计资料的 SAS 程序编写，正确解释结果的输出。统揽资料输出的全部信息，并用于指导科技工作。对于自己所从事的研究课题，不仅要作单因子分析，最后还必须作多因子综合分析（即多元分析）。只有这样，才能揭示事物全貌及本质联系，正确反映研究对象的联系规律。

与国内外同类型教材相比较，本书还详尽介绍了具有创新性的精彩内容，如各种经典试验设计资料的多元分析，正交试验的多元分析及套设计的多元分析等。此外，对各种多元分析的特殊功能，也作了详尽介绍（见第十一章）。这些内容，是院校历代统计专家长期勤奋耕耘的结晶与贡献。

本书包括内容极其丰富。一些较深内容初学者可暂时略去，待以后实际工作中不断加深理解。此项工作，须在任课教师指导下安排进行。编者这样做的目的，在于尽量扩大读者的知识视野，丰富本学科的知识结构。

编程上机、用计算机分析试验资料，是学好本课程最重要、最关键的学习方法。不要把学习停留在书本上，应该手指落实到键盘上。这是国内外众多 SAS 学者共同的成功学习经验。

这是一本好书，我乐于推荐给读者。祝愿同学们、同志们从本书获得 SAS 技术的宝贵知识，以便使自己所从事的科研工作达到国内外一流水平，成为一名掌握当代科技发展水准的研究工作者。

林德光

2006 年 8 月于海口

# 编者的话

SAS是目前国际上应用最广泛的标准统计分析软件之一。我国在自然科学和社会科学的各个领域，也广泛应用到SAS。我国许多高等院校的研究生、本科生教育中也相继开设有关SAS软件及其应用方面的课程。本书就是为此类课程的教学而编写的。

编者吸收了我校历代统计专家，特别是林德光教授在SAS方面的教学和科研成果，参考借鉴了国内外同类教材的相关内容，结合农林领域科研实例，编写了此书。本书可以作为高等农林院校有关专业研究生和本科生教材，也可供教师和科研人员参考。

全书共十一章，编写与分工如下：薛文龙老师编写第三章；王英老师编写第六章；韩汉鹏老师编写第十一章；其余各章均由唐燕琼编写，并对全书统稿和校对。此外，吴紫云同学参与部分编排和校对工作，林正洪同学录入了部分资料。

全书由林德光教授主审。编写过程中，始终得到导师林教授热情鼓励和悉心指导，并为本书撰写了前言。恩师无私奉献精神、治学严谨态度、深厚学术造诣和渊博知识，关爱提携年轻后辈的高尚品德，让编者敬仰，并成为完成此书的精神动力。为了让读者了解我院校开展统计学教学科研的简况，特将林德光教授撰写的《我院校开展统计学与SAS教学科研简史》作为本书的后记。

本书的编写与出版，得到我校特色教材编写项目的资助，在此表示衷心的感谢。同时对支持和关心本书编写出版工作的领导、同事表示深深的谢意。

若读者需书中的例题及 SAS 程序可发邮件 TYQ68@126.com 向编者索取。

限于我们的知识和水平，教材中难免有疏漏和不妥之处，敬请读者不惜赐教，编者将不胜感谢。

唐燕琼

2006 年 8 月



# 目 录

前言

编者的话

第一章 SAS 概述 .....	1
1.1 Windows 系统下 SAS 简介 .....	1
1.2 SAS9.0 系统的安装和启动 .....	2
1.2.1 安装 SAS .....	2
1.2.2 启动 SAS .....	3
1.3 SAS 主要窗口介绍 .....	4
1.3.1 增强型编辑器 .....	5
1.3.2 程序编辑器窗口 .....	6
1.3.3 输出窗口 .....	7
1.3.4 日志窗口 .....	7
1.3.5 Web 输出窗口 .....	7
1.3.6 SAS 资源管理器窗口 .....	9
1.3.7 结果管理窗口 .....	10
1.4 SAS 程序的组成 .....	11
1.5 SAS 程序的编写规则 .....	11
1.6 SAS 程序的运行、修改与存储 .....	12
1.7 一些常用 SAS 语句 .....	12
1.7.1 TITLE 语句 .....	12
1.7.2 DM 语句 .....	12
1.7.3 OPTIONS 语句 .....	13
1.7.4 DATA 语句 .....	13
1.7.5 INPUT 语句 .....	13
1.7.6 PROC 语句 .....	16
1.7.7 OUTPUT 语句 .....	17
1.7.8 PRINT 语句 .....	17
1.8 一个简单的 SAS 程序例子 .....	18

<b>第二章 建立 SAS 数据集</b> .....	20
2.1 编辑程序导入数据 .....	20
2.2 将其他格式的数据文件导入数据集 .....	21
2.2.1 文本格式的数据文件 .....	21
2.2.2 微机格式的数据文件 .....	22
2.3 利用 ASSIST 创建数据集 .....	25
2.4 数据集的复制、拆分与合并 .....	30
2.4.1 数据集的复制 .....	30
2.4.2 数据集的拆分 .....	31
2.4.3 数据集的合并 .....	32
<b>第三章 报表与图形输出</b> .....	36
3.1 SORT 过程 .....	36
3.2 GPLOT 过程 .....	37
3.2.1 散点图 .....	37
3.2.2 连线图 .....	37
3.2.3 回归曲线图 .....	38
3.3 GCHART 过程 .....	40
3.4 G3D 过程与 GCONTOUR 过程 .....	42
<b>第四章 SAS 常用的函数、数组与矩阵</b> .....	44
4.1 SAS 常用的函数 .....	44
4.1.1 算术运算符 .....	44
4.1.2 比较运算符 .....	44
4.1.3 逻辑运算符 .....	44
4.1.4 数学函数 .....	45
4.1.5 截取函数 .....	45
4.1.6 三角函数 .....	45
4.1.7 概率分布函数 .....	45
4.1.8 样本统计函数 .....	46
4.2 数组 .....	46
4.2.1 数组的定义 .....	46
4.2.2 数值型数组 .....	46
4.2.3 字符型数组 .....	47

4.2.4 临时数组 .....	47
4.2.5 数组函数 .....	47
4.2.6 数组的使用 .....	48
4.3 矩阵 .....	48
4.3.1 资料矩阵、特征根、特征向量、迹的定义 .....	48
4.3.2 SAS/IML 模块简介 .....	50
4.3.3 SAS 定义矩阵 .....	52
4.3.4 SAS 矩阵的运算 .....	54
<b>第五章 描述性统计分析 .....</b>	<b>57</b>
5.1 常用的统计分析术语 .....	57
5.2 MEANS 过程 .....	59
5.3 UNIVARIATE 过程 .....	62
<b>第六章 统计推断 .....</b>	<b>68</b>
6.1 统计推断基本思想 .....	68
6.1.1 统计假设检验 .....	68
6.1.2 参数置信区间 .....	68
6.2 正态性检验 .....	69
6.2.1 单样本正态性检验 .....	69
6.2.2 多样本正态性检验 .....	70
6.3 t 检验 .....	71
6.3.1 单个样本均数的 t 检验 .....	71
6.3.2 两个样本均数的 t 检验 .....	73
6.3.3 配对设计资料的 t 检验 .....	75
6.4 联列表独立性检验 .....	76
6.4.1 相关原理与公式 .....	76
6.4.2 SAS 分析 (FREQ 过程) .....	77
6.4.3 2×2 联列表的 $\chi^2$ 测验法 .....	78
6.4.4 2×C 联列表的 $\chi^2$ 测验法 .....	80
6.4.5 R×C 联列表的 $\chi^2$ 测验法 .....	81
<b>第七章 方差分析 .....</b>	<b>83</b>
7.1 方差分析概述 .....	83
7.1.1 相关名词解释 .....	83

7.1.2	方差分析基本思想	83
7.1.3	常用的统计代换	84
7.1.4	常用的均数多重比较方法	85
7.2	方差分析的 SAS 过程	86
7.2.1	ANOVA 过程的常用格式	87
7.2.2	GLM 过程的常用格式	87
7.3	完全随机设计资料的方差分析	89
7.3.1	单因子完全随机等重复资料分析	89
7.3.2	单因子完全随机不等重复资料分析	94
7.3.3	复因子完全随机资料分析	95
7.4	随机区组设计资料分析	100
7.4.1	单因子随机区组设计资料分析	100
7.4.2	复因子随机区组设计资料分析	105
7.5	拉丁方设计试验资料分析	115
7.6	裂区设计试验资料分析	117
7.7	正交设计试验资料	121
7.7.1	无重复正交试验资料的分析	121
7.7.2	有重复无交互正交试验	124
7.7.3	有交互作用正交试验资料分析	127
7.8	套设计资料分析	132
7.8.1	A、B 固定	133
7.8.2	A 固定、B 随机	135
7.9	混杂设计资料的方差分析	138
7.9.1	完全混杂	138
7.9.2	部分混杂	146
7.10	不平衡设计的方差分析 (GLM 过程)	149
7.10.1	单因子不平衡设计的分析	150
7.10.2	两因子不平衡设计的分析	152
7.11	多年生作物的试验	153
7.12	多点品种区域试验的方差分析	157
7.13	多年多点品种区域试验的方差分析	160
<b>第八章</b>	<b>相关分析和回归分析</b>	<b>165</b>
8.1	相关分析和回归分析概述	165
8.2	简单相关与偏相关分析	166
8.2.1	基本原理	166

---

8.2.2 SAS分析 .....	167
8.3 直线回归 .....	173
8.3.1 直线回归概述 .....	173
8.3.2 直线回归 SAS分析 .....	174
8.4 多元线性回归 .....	178
8.4.1 多元线性回归方程的建立 .....	178
8.4.2 多元线性回归方程的显著性检验 .....	179
8.4.3 SAS分析 .....	179
8.5 逐步回归 .....	183
8.6 非线性回归 .....	187
8.6.1 曲线直线化 .....	187
8.6.2 非线性回归的 NILIN 过程 .....	191
8.6.3 多项式回归分析 .....	193
<b>第九章 协方差分析 .....</b>	<b>199</b>
9.1 协方差分析概述 .....	199
9.2 含有一个协变量的协方差分析 .....	199
9.2.1 完全随机设计协方差分析 .....	199
9.2.2 随机区组的协方差分析 .....	203
9.2.3 正交试验的协方差分析 .....	205
9.3 多元协方差分析 .....	208
<b>第十章 多元方差分析 (MANOVA) .....</b>	<b>211</b>
10.1 MANOVA 语句 .....	211
10.2 单因子试验的多元分析 .....	212
10.2.1 数学模型与数据矩阵的分解 .....	212
10.2.2 Wilks 的 $\Lambda$ 统计量及其分布 .....	213
10.2.3 处理效应的联合置信区间 .....	214
10.2.4 Bonferroni 检验法 .....	215
10.2.5 单因子多元方差分析实例 .....	216
10.3 二因子试验多元方差分析 .....	223
10.3.1 数学模型 .....	223
10.3.2 $\Lambda$ 统计量的构造 .....	223
10.3.3 实例分析 .....	224
10.4 正交试验的多元分析 .....	231
10.5 多元套设计的 SAS 程序 .....	235

10.5.1 二阶套设计多元分析方法 .....	235
10.5.2 Wilks 的 $\Lambda$ 统计量的构造 (假定 A、B 固定) .....	236
10.5.3 多元套设计的 SAS 程序 .....	236
10.6 裂区设计的多元方差分析 .....	242
10.6.1 引言 .....	242
10.6.2 数学模型及平方和、乘积和矩阵的分解 .....	242
10.6.3 Wilks 的 $\Lambda$ 统计量的构造 (假定 A、B 固定) .....	243
10.6.4 多元裂区设计的 SAS 分析 .....	243
<b>第十一章 SAS 多元统计分析 .....</b>	<b>249</b>
11.1 主成分分析 .....	249
11.1.1 主成分分析的基本原理 .....	249
11.1.2 PRINCOMP 过程 .....	251
11.1.3 实例分析 .....	252
11.2 因子分析 .....	260
11.2.1 因子分析的基本思想 .....	260
11.2.2 因子分析模型 .....	261
11.2.3 SAS/STAT 中的 FACTOR 过程 .....	264
11.2.4 实例分析 .....	266
11.3 对应分析 .....	271
11.3.1 对应分析的基本思想 .....	271
11.3.2 对应分析的 SAS 书写格式与实例分析 .....	272
11.4 聚类分析 .....	277
11.4.1 聚类分析的方法 .....	277
11.4.2 距离与相似系数 .....	278
11.4.3 系统聚类过程 .....	279
11.4.4 SAS 格式与 Cluster 过程 .....	284
11.4.5 实例分析 .....	285
11.5 判别分析 .....	301
11.5.1 距离判别 .....	301
11.5.2 两组 (两个总体) 距离判别 .....	302
11.5.3 多组 (总体) 距离判别 .....	309
11.6 典型相关分析 .....	314
11.6.1 典型相关分析的基本思想 .....	314
11.6.2 典型相关分析的 SAS 书写格式与实例 .....	315

目 录

---

附录 英汉对照表 .....	326
参考文献 .....	346
后记 .....	347

# 第一章 SAS 概述

## 1.1 Windows 系统下 SAS 简介

SAS (Statistical Analysis System) 是当代最著名的统计分析软件之一。SAS 是用于数据分析和决策支持的大型集成式、模块化系统。SAS 由 30 多个专用模块组成, 其中 Base SAS 模块是 SAS 系统的核心。SAS 分析模块主要有 STAT、QC、IML 等模块, 而 STAT 模块几乎覆盖了所有的实用数理统计分析方法。SAS 功能包括: 客户机/服务器计算、数据访问、数据存储及管理、应用开发、图形处理、数据分析、报告编制、质量控制、项目管理、计算机性能评估、运筹学方法、计量经济学与预测等。各个模块之间既相互独立又相互交融与补充, 可以根据具体应用建立相应模块的信息分析与应用系统。

SAS 系统于 1966 年由美国北卡罗来纳州 (North Carolina) 州立大学开始研制, 1976 年成立美国 SAS 软件研究所, 并开始对 SAS 系统进行维护、开发、销售和培训等工作。1985 年推出了 SAS/PC (6.02) 版本。自 SAS 系统推出以来, 它的版本更新很快, 功能也不断增加。如自 1989 年推出 SAS/PC (6.04) 版本后; 1997 年下半年推出适合于多种操作系统的 6.12 版本; 2000 年 2 月推出新的版本 8; 2004 年 3 月 31 日北京 SAS® 9 全球同步上市中国媒体见面会在北京嘉里中心召开。SAS 软件发布其新的 SAS® 9 软件家族, SAS® 9 是集成了全新的平台、增强的分析功能和精制的用户界面的软件, 旨在提供新的洞察力来帮助用户解决商业问题和赢得竞争优势。

目前最新版本已经升级到 9.1.3, SAS9.1.3 构建在可扩展性、互操作性、可管理性和可用性四大设计基石之一, 充分体现了 2005 年 SAS 提出的超越传统商业智能的企业级智能平台 (EIP, Enterprise Intelligence Platform) 概念, 开发出了丰富的商业解决方案, 支持企业级智能、客户智能、金融智能、供应链智能等, 并支持各类垂直市场。目前 SAS 在全球已有近 400 个分支机构, 拥有 4 万多家客户。其中包括世界 500 强中的前 100 家有 97 家企业, 都在使用 SAS 提供的商业智能解决方案。我国的宝钢集团、工商银行等大公司就在使用 SAS 提供的商业智能解决方案。

Windows 版本 9.0SAS 系统软件运行环境:

Windows XP Professional、Windows 2000 Professional 和所有的 Win-



dows 2000 Server 版本, Windows 2000 必须打上 SP1 补丁 (Service Pack 1)、Windows NT 则需要 Windows NT Version 4.0 更新到 Service Pack 5 或更高 (包括 Service Pack 6、Service Pack 6a 或更高)。

Windows 版本 9.0 SAS 系统最小硬件要求:

CPU: Pentium II 或更高

内存: 128M 或更高 (Windows Professional); 256M 或更高 (Windows Server)

硬盘: 128M 或更高 (Windows Professional); 256M 或更高 (Windows Server)

显示器: SVGA (分辨率 800×600 或更高)

## 1.2 SAS9.0 系统的安装和启动

### 1.2.1 安装 SAS

首先, 以管理员身份登录 Windows 系统, 将标有 “Setup Disk” 的光盘插入光驱, 在光盘根目录下面点击 setup.exe, 然后选择要安装的语言 (图 1.1):

点击 “确定”, 接着出现如下画面 (图 1.2):

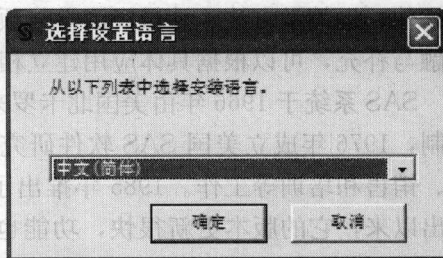


图 1.1 语言选择

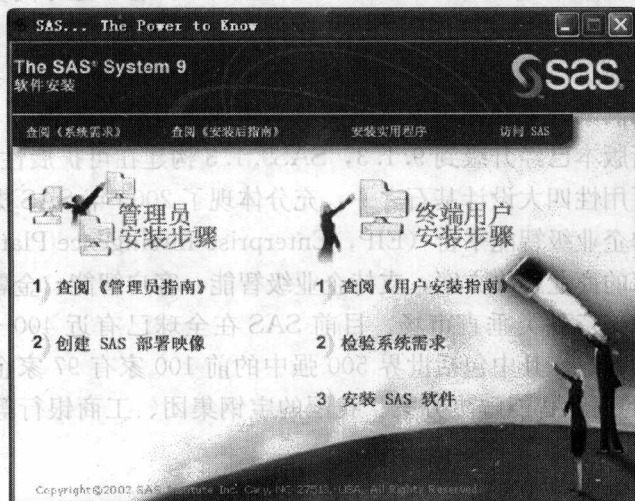


图 1.2 自动安装部署界面