

SAS统计分析与应用



从入门到精通

● 汪海波 罗莉 汪海玲 徐俊杰 编著

源于实践 成就行录

SAS

上海软件行业协会 秘书长 杨根兴
江苏省软件行业协会 副会长 徐雷

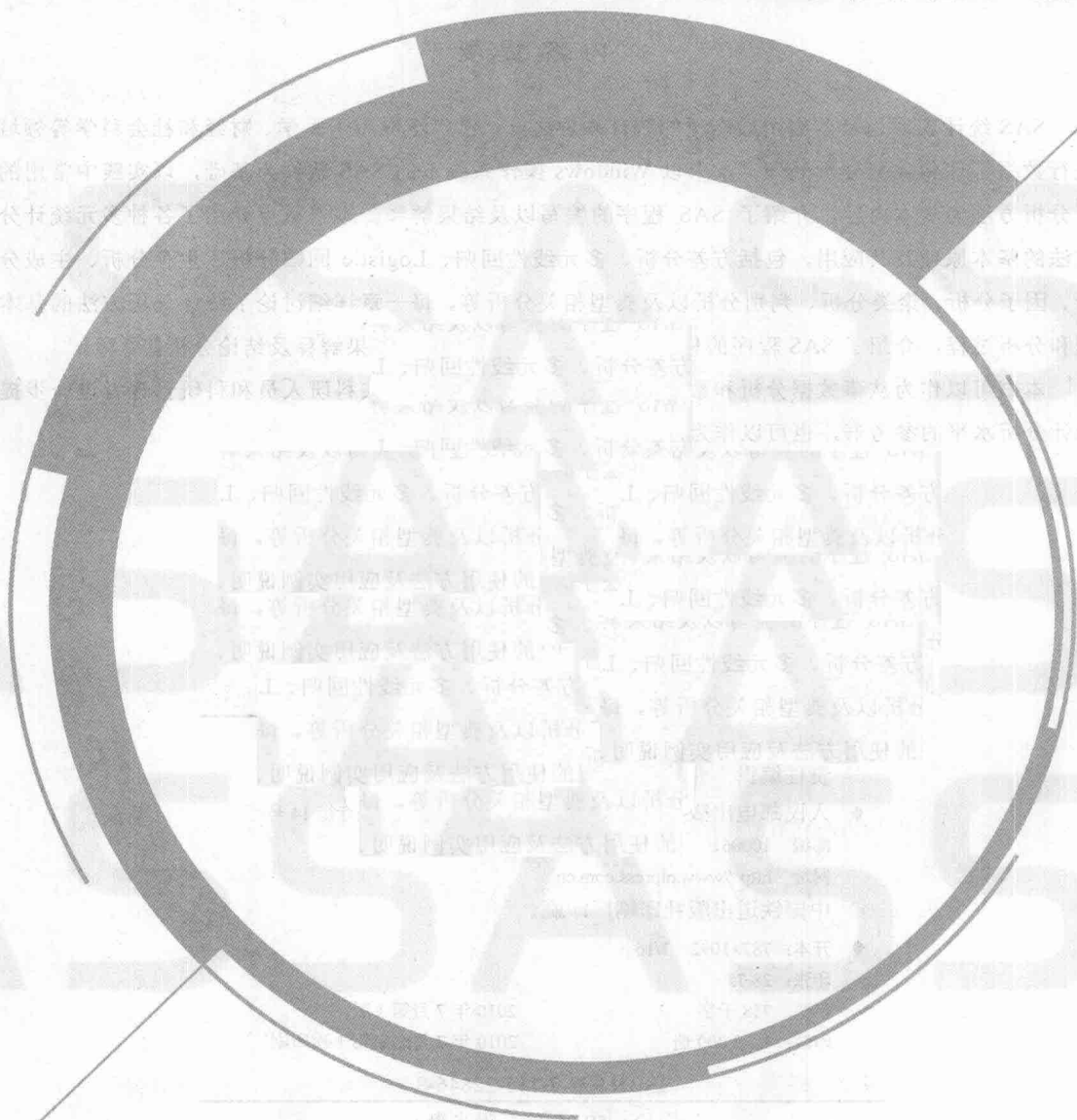
鼎力推荐

- ▶ 20个综合案例
- ▶ 143个贯穿SAS数据统计应用的实例
- ▶ 46个疑难解答和实战技巧
- ▶ 620分钟的视频讲解及案例程序（见光盘）

 人民邮电出版社
POSTS & TELECOM PRESS

SAS统计分析与应用 从入门到精通

◎ 汪海波 罗莉 汪海玲 徐俊杰 编著



人民邮电出版社
北京

图书在版编目 (CIP) 数据

SAS 统计分析与应用从入门到精通 / 汪海波等编著

— 北京 : 人民邮电出版社, 2010. 7

ISBN 978-7-115-22846-8

I. ①S… II. ①汪… III. ①统计分析—应用软件,
SAS IV. ①C812

中国版本图书馆CIP数据核字(2010)第075889号

内 容 提 要

SAS 统计软件是目前应用最广泛的统计软件之一, 已广泛应用于医学、财经和社会科学等领域中进行数据管理和数据分析处理。本书以 Windows 操作系统下的 SAS 软件为基础, 以实践中常用的统计分析方法为基本内容, 介绍了 SAS 程序的编写以及结果解释。本书重点介绍了各种多元统计分析方法的基本原理及其应用, 包括方差分析、多元线性回归、Logistic 回归分析、生存分析、主成分分析、因子分析、聚类分析、判别分析以及典型相关分析等。每一章详细讨论了统计分析方法的基本原理和分析过程, 介绍了 SAS 程序的使用方法及应用实例说明、结果解释及结论分析等。

本书可以作为从事数据分析和数据管理人员的参考用书, 有关科研人员和科研管理者进一步提高统计分析水平的参考书, 也可以作为高等院校师生统计教材。

SAS 统计分析与应用从入门到精通

- ◆ 编 著 汪海波 罗 莉 汪海玲 徐俊杰
责任编辑 张 涛
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
中国铁道出版社印刷厂印刷
 - ◆ 开本: 787×1092 1/16
印张: 28.75
字数: 758 千字
印数: 1—3 000 册
- 2010 年 7 月第 1 版
2010 年 7 月北京第 1 次印刷

ISBN 978-7-115-22846-8

定价: 59.00 元 (附光盘)

读者服务热线: (010)67132692 印装质量热线: (010)67129223
反盗版热线: (010)67171154

前 言

SAS 系统是世界领先的信息系统，是一个大型规模化的集成应用软件，具有完备的数据存取、管理、分析和显示的功能。SAS 统计分析软件使用灵活方便、功能齐全，SAS 语言编程能力强且简单易学，其数据处理和统计分析融为一体，已被广泛应用于医学、财经和社会科学等领域中。

本书内容充实、重点突出、实用性强，不仅介绍了常用的统计分析方法，也详细阐述了多元统计分析方法及其应用，例如方差分析、生存分析等，是覆盖多元统计分析方法较完整的一本参考书，适合于不同层次的人员使用。本书对每一种多元统计分析方法的基本原理、分析过程进行了详细的阐述，并介绍了 SAS 程序的使用方法、应用实例说明、结果解释、结论分析等，使读者可以将统计分析方法与 SAS 软件应用结合起来学习，全面掌握统计分析方法。统计基础知识比较薄弱者可以忽略计算公式等理论部分，把重点放在对基本原理的理解和利用计算机解决实际问题的应用上。统计基础知识较好者可以精读每一部分内容，使自己的统计分析能力达到更高的一个层次，并为今后进一步学习和研究打下扎实的实践基础。

本书在编写方法上有所创新。目前，许多统计分析教程存在理论和实践操作脱节，或者说计算例题和解决实际问题脱节，使得多元统计在研究上的应用价值不能真正体现出来。本书从讲求实效的原则出发，将统计分析原理和 SAS 软件的输出结果有机地结合在一起，将理论学习与实际操作融为一体，使得读者有机会把各种统计分析方法的理论知识和解决实际问题的能力结合起来。

本书内容

本书第 1 章介绍了 SAS 系统的特点以及结构组成、安装和启动、SAS 菜单操作和 SAS 系统的文件管理。第 2 章介绍了 SAS 编程入门，包括数据集的建立以及各种操作方法等。第 3 章~第 5 章分别介绍了 SAS 宏编程、输出传送系统和 PROC SQL。第 6 章和第 7 章概括地总结了基本统计分析方法的主要内容。从第 8 章开始逐个介绍了工作实践中常用的一些多元统计分析方法，内容包括各种多元统计分析方法的基本原理、计算方法、分析步骤、实例应用、SAS 程序和使用方法以及结果解释和结论分析等。从第 21 章开始探讨了 SAS 软件的拓展应用，包括 ROC 分析、一致性检验、概率抽样方法、样本量估计和统计图。另外，本书每一章都给出了使用 SAS 软件进行数据分析的实例，以帮助读者理解和掌握各章的内容。

本书特色

- SAS 资深专家执笔。作者深入理解了 SAS 内涵、精髓，结合自己丰富的培训经验，并结合大量的一线工程实践经验，潜心编写而成。
- 软件版本采用当前最为流行的 SAS 版本。在知识点讲解过程中穿插了新功能的讲述与应用。
- 知识全面、系统，科学安排内容的层次架构，由浅入深，循序渐进，适合读者的学习规律。
- 理论与实践应用紧密结合。基础理论知识穿插在知识点的讲述中，言简意赅、目标明确，其目的是使读者知其然，亦知其所以然，达到学以致用目的。

- 知识点+针对每个知识点的小实例+综合实例的讲述方式，可以使读者快速地学习、掌握 SAS 软件操作及应用该知识点解决实践中的问题。综合实例部分，深入细致地剖析数据统计分析应用的流程、细节、难点、技巧，可以起到融会贯通的作用。

本书附带所有实例操作的视频光盘。

本书由汪海波、罗莉、汪海玲、徐俊杰主编，参与编写的还有郝旭宁、李建鹏、赵伟茗、刘钦、于志伟、张永岗、周世宾、姚志伟、曹文平、张应迁、张洪才、邱洪钢、张青莲、陆绍强、李成。

读者对象

本书可以作为从事数据分析和数据管理人员的参考用书，有关科研人员和科研管理者进一步提高统计分析水平的参考书，也可以作为高等院校师生统计教材。

因编者水平有限，不妥之处敬请读者和同行提出宝贵意见，联系邮箱为 zhangtao@ptpress.com.cn。

编者

目 录

第1章 SAS for Windows入门 --1

1.1 SAS 简介-----	2
1.1.1 SAS 系统特点-----	2
1.1.2 SAS 系统结构组成-----	3
1.1.3 SAS 系统资料-----	4
1.1.4 SAS 软件要求-----	4
1.2 SAS for Windows 的安装和 启动-----	4
1.3 SAS 菜单操作-----	5
1.4 SAS 系统的文件管理-----	6
1.5 本章小结-----	7

第2章 SAS编程入门 -----8

2.1 SAS 语言构成-----	9
2.1.1 使用 SAS 语言-----	9
2.1.2 DATA 数据步-----	11
2.1.3 PROC 过程步-----	14
2.2 SAS 常用函数及其应用----	17
2.2.1 算术函数 (Arithmetic Functions) -----	17
2.2.2 样本统计函数 (Sample Statistic Functions) -----	17
2.2.3 截取函数 (Truncation Functions) -----	18
2.2.4 随机数函数 (Random Number Functions) -----	18
2.2.5 数学函数 (Mathematical Functions) -----	19
2.2.6 金融函数 (Financial Functions) -----	20
2.2.7 三角函数 (Trigonometric Functions) -----	20
2.2.8 字符函数 (Character Functions) -----	20
2.2.9 概率函数 (Probability Functions) -----	21
2.2.10 日期和时间函数 (Date and	

Time Functions) -----	22
2.2.11 分位数函数 (Quantile Functions) -----	23
2.2.12 数组函数 (Array Functions) -----	24
2.2.13 特殊函数 (Special Functions) -----	24
2.3 数据集-----	25
2.3.1 数据集的建立与保存-----	25
2.3.2 数据集的导入与导出-----	27
2.3.3 数据集的排序-----	28
2.3.4 数据集的连接-----	28
2.3.5 数据集的合并-----	30
2.3.6 数据集的复制-----	32
2.3.7 数据集的拆分-----	33
2.3.8 数据集的转置-----	35
2.4 SAS 语句-----	36
2.4.1 赋值语句-----	36
2.4.2 输出语句-----	36
2.4.3 条件语句-----	38
2.4.4 循环语句-----	39
2.4.5 WHERE 语句-----	41
2.4.6 Delete 语句-----	42
2.4.7 累加语句-----	43
2.4.8 By 语句-----	43
2.4.9 Update 语句-----	44
2.5 本章小结-----	45

第3章 SAS宏编程----- 46

3.1 宏变量-----	47
3.1.1 宏变量的产生-----	47
3.1.2 宏变量的引用-----	48
3.1.3 宏变量的特殊用法-----	50
3.2 宏-----	52
3.2.1 宏的定义-----	52
3.2.2 宏的调用-----	52
3.3 本章小结-----	57

第4章 SAS输出传送系统**(ODS) -----58**

- 4.1 ODS 功能 ----- 59
- 4.2 ODS 语句 ----- 59
- 4.3 ODS 程序实例 ----- 60
 - 4.3.1 缺省的 Listing 目标 ----- 60
 - 4.3.2 ODS TRACE 语句 ----- 61
 - 4.3.3 ODS SELECT 和 ODS EXCLUDE 语句 ----- 62
 - 4.3.4 ODS OUTPUT 语句 ----- 62
 - 4.3.5 ODS RTF 语句 ----- 63
 - 4.3.6 ODS HTML 语句 ----- 64
- 4.4 本章小结 ----- 64

第5章 PROC SQL简介 -----66

- 5.1 SQL 过程概述 ----- 67
- 5.2 SQL 语句格式 ----- 67
- 5.3 SQL 过程应用实例 ----- 69
- 5.4 本章小结 ----- 74

第6章 定量资料的统计描述 ---75

- 6.1 统计描述基础理论知识 --- 76
 - 6.1.1 集中趋势描述 ----- 76
 - 6.1.2 离散趋势描述 ----- 78
 - 6.1.3 正态分布 ----- 80
- 6.2 统计描述分析实例 ----- 81
 - 6.2.1 proc means 分析实例 ----- 81
 - 6.2.2 proc univariate 分析实例 ----- 83
 - 6.2.3 对数正态分布资料的统计描述 ----- 86
- 6.3 本章小结 ----- 87

第7章 t 检验 -----88

- 7.1 单样本 t 检验 ----- 89
 - 7.1.1 单样本 t 检验的基础理论 ----- 89
 - 7.1.2 单样本 t 检验分析实例——MEANS、UNIVARIATE 过程 ----- 89
 - 7.1.3 无原始数据的单样本 t 检验 SAS 程序 ----- 91

7.2 配对设计资料的 t 检验 ---- 92

- 7.2.1 配对设计资料 t 检验的基础理论 ----- 92
- 7.2.2 配对 t 检验实例——MEANS、UNIVARIATE、TTEST 过程 ----- 93
- 7.2.3 无原始数据的配对设计的 t 检验分析实例 ----- 97

7.3 两独立样本的 t 检验 ----- 98

- 7.3.1 两独立样本 t 检验的基础理论 ----- 98
- 7.3.2 独立样本 t 检验分析实例——TTEST 过程应用 ---- 99
- 7.3.3 无原始数据的两独立样本 t 检验分析实例 ----- 101

7.4 本章小结 ----- 102

第8章 方差分析 ----- 103

8.1 方差分析概述 ----- 104

8.2 完全随机设计资料的方差分析 ----- 105

- 8.2.1 单因子方差分析介绍 --- 106
- 8.2.2 单因子方差分析的 SAS 程序实例 ----- 106

8.3 随机区组设计资料的方差分析 ----- 111

- 8.3.1 随机区组方差分析介绍 ----- 111
- 8.3.2 随机区组方差分析的 SAS 程序实例 ----- 114

8.4 拉丁方设计资料的方差分析 ----- 116

- 8.4.1 拉丁方方法介绍 ----- 117
- 8.4.2 拉丁方分析的 SAS 程序实例 ----- 118

8.5 析因设计资料的方差分析 ----- 121

- 8.5.1 析因设计方法介绍 ----- 121
- 8.5.2 析因方差分析的 SAS 程序实例 ----- 124

8.6 正交试验设计资料的方差分析 ----- 127

8.6.1 正交试验设计方法 介绍-----	127	11.1.1 非条件 Logistic 回归 介绍-----	182
8.6.2 正交试验设计资料分析的 SAS 程序实例-----	129	11.1.2 非条件 Logistic 回归 模型的建立和检验----	183
8.7 重复测量资料的方差 分析-----	130	11.1.3 非条件 Logistic 回归的 SAS 程序-----	185
8.7.1 重复测量设计方法 介绍-----	131	11.2 条件 Logistic 回归-----	202
8.7.2 重复测量资料分析的 SAS 程序实例-----	132	11.2.1 条件 Logistic 回归 介绍-----	203
8.8 协方差分析-----	140	11.2.2 条件 Logistic 回归的 SAS 程序-----	204
8.8.1 协方差分析方法介绍---	140	11.3 本章小结-----	206
8.8.2 协方差分析的 SAS 程序 实例-----	141		
8.9 本章小结-----	145		
第9章 直线回归与相关----- 146		第12章 相对数----- 207	
9.1 直线相关分析-----	147	12.1 相对数简介-----	208
9.1.1 直线相关分析介绍-----	147	12.1.1 率的标准化-----	209
9.1.2 直线相关分析的 SAS 程序实例-----	149	12.1.2 率的假设检验-----	212
9.2 直线回归分析-----	152	12.2 SAS 分析实例-----	214
9.2.1 直线回归分析介绍-----	152	12.2.1 率的标准化 SAS 程序--	214
9.2.2 直线回归分析的 SAS 程序实例-----	154	12.2.2 率的 Z (U) 检验的 SAS 程序-----	214
9.3 本章小结-----	158	12.3 本章小结-----	216
第10章 多元线性回归与相关-- 159		第13章 行×列表分析----- 217	
10.1 多元线性回归与相关的 基础理论-----	160	13.1 四格表资料-----	218
10.1.1 多元线性回归-----	160	13.1.1 四格表卡方检验介绍---	218
10.1.2 复相关系数与偏 相关系数-----	172	13.1.2 四格表卡方检验的 SAS 程序-----	220
10.2 分析实例-----	173	13.2 配对计数资料的卡方 检验-----	224
10.2.1 多元线性回归方程的 建立-----	173	13.2.1 四格表配对卡方检验 介绍-----	224
10.2.2 复相关系数与偏相关系 数的 SAS 程序实例----	177	13.2.2 四格表配对卡方检验的 SAS 程序-----	225
10.3 本章小结-----	180	13.3 列变量为顺序变量的行 均分检验-----	226
第11章 Logistic回归分析 --- 181		13.3.1 行均分检验介绍-----	226
11.1 非条件 Logistic 回归-----	182	13.3.2 行均分检验的 SAS 程序-----	227
		13.4 行列均为顺序变量的相关 检验-----	230
		13.4.1 行列均为顺序变量的 相关检验介绍-----	231

13.4.2	行列均为顺序变量的 相关检验的 SAS 程序 --	231	14.5	等级分组资料的非参数 检验 -----	263
13.5	分层行列表的分析 -----	237	14.5.1	等级分组资料的非参数 检验方法概述 -----	263
13.5.1	分层行列表的分析 简介 -----	237	14.5.2	等级分组资料非参数 检验的 SAS 程序 -----	264
13.5.2	分层行列表的分析的 SAS 程序 -----	238	14.6	随机区组资料的非参数 检验 -----	266
13.6	趋势卡方检验 -----	243	14.6.1	随机区组资料的非参数 检验方法概述 -----	266
13.6.1	趋势卡方检验简介 ----	243	14.6.2	随机区组资料非参数 检验的 SAS 程序 -----	267
13.6.2	趋势卡方检验的 SAS 程序 -----	244	14.7	等级相关 (秩相关) ----	268
13.7	卡方分割与卡方合并 ----	246	14.7.1	秩相关概述 -----	268
13.7.1	卡方的分割与合并 简介 -----	246	14.7.2	spearman 秩相关的 SAS 程序 -----	269
13.7.2	卡方分割与卡方合并 分析实例 -----	246	14.8	本章小结 -----	270
13.8	本章小结 -----	248			
第14章 非参数统计 ----- 249			第15章 生存分析 ----- 271		
14.1	单样本资料与已知总体 参数的非参数检验 -----	250	15.1	生存分析简介 -----	272
14.1.1	单组资料的符号及符号 秩和检验 -----	250	15.1.1	生存数据 -----	272
14.1.2	单组资料的非参数检验 SAS 程序 -----	252	15.1.2	生存时间函数 -----	272
14.2	配对设计资料的非参数 检验 -----	253	15.1.3	均数、中位数和半数 生存期 -----	273
14.2.1	配对设计资料的符号 及符号秩和检验 -----	253	15.1.4	生存分析的基本方法 -	274
14.2.2	配对设计资料的非参数 检验 SAS 程序 -----	254	15.2	生存曲线 -----	274
14.3	两组定量资料的非参数 检验 -----	256	15.2.1	寿命表法及 SAS 分析 实例 -----	275
14.3.1	两组定量资料的非参数 检验方法概述 -----	256	15.2.2	乘积极限法 (Kaplan-Meier) 及 SAS 分析实例 -----	281
14.3.2	两组定量资料非参数 检验的 SAS 程序 -----	257	15.2.3	Cox 回归及 SAS 分析 实例 -----	285
14.4	多组定量资料的非参数 检验 -----	259	15.3	本章小结 -----	291
14.4.1	多组定量资料的非参数 检验方法概述 -----	259			
14.4.2	多组定量资料非参数 检验的 SAS 程序 -----	261	第16章 主成分分析 ----- 292		
			16.1	主成分分析简介 -----	293
			16.1.1	主成分分析的数学 模型 -----	293
			16.1.2	主成分分析的方法 步骤 -----	294
			16.1.3	主成分分析的应用 ----	296
			16.2	PRINCOMP 过程实例 --	297
			16.3	本章小结 -----	306

第17章 因子分析-----307	
17.1 因子分析简介 -----308	
17.2 FACTOR 过程实例 -----316	
17.3 本章小结 -----323	
第18章 聚类分析-----325	
18.1 聚类分析简介 -----327	
18.2 聚类分析 SAS 实例 -----332	
18.2.1 VARCLUS 过程 实例 ----- 332	
18.2.2 CLUSTER 过程 实例 ----- 341	
18.2.3 FASTCLUS 过程 实例 ----- 346	
18.2.4 ACECLUS 过程 实例 ----- 350	
18.3 本章小结 -----357	
第19章 判别分析-----358	
19.1 判别分析简介 -----359	
19.2 判别分析 SAS 实例 -----364	
19.2.1 DISCRIM 过程实例 --- 364	
19.2.2 CANDISC 过程 实例 ----- 371	
19.2.3 STEPDISC 过程 实例 ----- 376	
19.3 本章小结 -----378	
第20章 典型相关分析 -----380	
20.1 典型相关简介 -----381	
20.1.1 典型相关分析的理论 架构及基本假设 ----- 382	
20.1.2 冗余分析 ----- 383	
20.1.3 典型相关系数的假设 检验 ----- 384	
20.2 CANCELL 过程实例 ---384	
20.3 本章小结 -----394	
第21章 诊断试验的ROC分析 -395	
21.1 诊断试验简介 -----396	
21.1.1 诊断试验介绍 ----- 396	
21.1.2 诊断试验评价指标 --- 396	
21.1.3 ROC 分析资料收集与 整理 ----- 399	
21.1.4 ROC 曲线构建 ----- 399	
21.2 ROC 分析及 SAS 分析 实例 ----- 400	
21.3 本章小结 ----- 403	
第22章 一致性检验Kappa --- 404	
22.1 二分类资料一致性分析 --- 405	
22.1.1 Kappa 值的计算及 检验 ----- 405	
22.1.2 一致性检验的 SAS 分析 实例 ----- 406	
22.2 有序分类资料一致性 分析 ----- 407	
22.2.1 有序分类资料一致性 分析简介 ----- 407	
22.2.2 有序分类资料一致性 分析的 SAS 实例 ----- 408	
22.3 本章小结 ----- 410	
第23章 概率抽样方法 ----- 411	
23.1 简单随机抽样及 SAS 程序 ----- 413	
23.2 系统（机械、等距）抽样 及 SAS 程序 ----- 417	
23.3 分层抽样及 SAS 程序 --- 418	
23.4 整群抽样及 SAS 程序 --- 425	
23.5 多阶抽样及 SAS 程序 --- 426	
23.6 本章小结 ----- 429	
第24章 样本量估计 ----- 430	
24.1 样本量估算的影响因素 --- 431	
24.2 分类资料的样本量 估计 ----- 431	
24.2.1 抽样调查样本量的估计 及 SAS 程序 ----- 431	
24.2.2 单样本与已知总体检验 时样本量的估计及 SAS 程序 ----- 432	
24.2.3 两样本率比较样本量的 估计及 SAS 程序 ----- 433	

24.2.4	配对设计总体率比较 样本量的估计及 SAS 程序	433
24.3	定量资料的样本量 估计	435
24.3.1	抽样调查总体参数估计 时样本量的估计及 SAS 程序	435
24.3.2	单样本与已知总体检验 时样本量的估计及 SAS 程序	435
24.3.3	两总体均数比较样 本量的估计及 SAS 程序	436
24.3.4	配对设计两样本均数 比较样本量的估计及 SAS 程序	436

24.4	本章小结	437
------	------	-----

第25章 统计图 ----- 438

25.1	直条图	439
25.2	圆图	441
25.3	线图	442
25.4	半对数线图	443
25.5	箱线图	444
25.6	散点图	447
25.7	直方图	448
25.8	统计地图	449
25.9	本章小结	449

参考文献 ----- 450

第 1 章

SAS for Windows入门

学习目标

- 了解SAS系统的特点以及结构组成。
- 熟悉SAS for Windows的安装和启动。
- 掌握SAS菜单操作。
- 掌握SAS系统的文件管理。

内容概要

SAS系统是世界领先的信息系统，是一个大型规模化的集成应用软件系统，具有完备的数据存取、管理、分析和显示等功能。本章对SAS软件以及SAS系统的特点进行了较详细的介绍，简要地概述了SAS系统结构组成（分别简要介绍各组成模块的作用），描述了SAS软件的要求和环境支持。希望大家对SAS系统能够有较全面的了解。

同时简要地介绍了SAS的安装和启动，对SAS菜单操作也进行了详细的讲解，尤其是那些SAS操作中经常应用的菜单。SAS虽然是编程软件，但是菜单操作仍有非常重要的作用，熟练地掌握SAS菜单操作是学好SAS的基础。还概述了SAS的文件管理，重点介绍了临时和永久数据库以及它们之间的区别。掌握SAS数据库的作用以及意义，有助于对SAS数据集和其他文件的管理。

1.1 SAS 简介

SAS 系统于 1966 年由美国北卡罗来纳州州立大学开始研制, 1976 年成立美国 SAS 软件研究公司, 并对 SAS 系统进行维护、开发、销售和培训工作。SAS 软件研究所自创建以来, 就以 SAS 系统的卓越功能和可靠技术支持著称于世, 逐步发展成为全球最大的独立软件开发商之一, 在世界 6 大洲的 60 多个国家设有分公司及分支机构。

1.1.1 SAS 系统特点

SAS 系统是世界领先的信息系统, 是一个大型规模化的集成应用软件系统, 具有完备的数据存取、管理、分析和显示等功能。所谓软件系统, 就是一组在一起作业的计算机程序。SAS 系统的用户都可以根据自己的需要, 随时作出调整。在数据处理和统计分析领域, SAS 系统被誉为国际上的标准分析软件系统。不论是复杂的还是简单的工作, SAS 都可以满足你的需要。SAS 系统在世界范围内被广泛应用于政府、科研、教育和生产等不同的领域, 发挥着积极的作用。SAS 系统的最大特点就是把数据管理和数据分析融为一体。SAS 系统特点如图 1-1 所示。



图 1-1 SAS 系统特点

具体来说, 它有以下几方面的特点。

(1) 使用灵活方便、功能齐全。用户把要解决的问题用 SAS 语言表达出来, 提交给 SAS 系统就可以解决问题。并且 SAS 系统提供有很多语句和选项, 供用户灵活地选择使用。

(2) SAS 语言编程能力强且简单易学。SAS 语言是一种近乎自然英语的非过程语言。

(3) 数据处理和统计分析融为一体。SAS 软件克服了通常软件偏重于数据管理功能或统计分析的弊病, 把数据管理功能和统计分析功能有机地结合在一起, 它不仅具有一整套从数据输入、加工处理直到文件操作等完备灵活的数据管理功能, 而且还能对所存储的数据进行统计分析。SAS 系统提供有完备的 SAS 语句和函数用于数据加工处理, 并具有对数据进行编辑、整理、连接和更新等功能。

(4) 扩展性强。SAS 系统的模块式结构可以让用户灵活地按自己的需要去选择所需功能的模块。

(5) 适用性强。SAS 系统可以在不同的操作系统下运行 (PC-DOS/MS-DOS、UNIX、VES、Windows 等), 因为它有 DOS 和 Windows 的不同版本; SAS 系统适用于任何类型的数据, 可以访问各种数据资源; SAS 系统适用于任何人员 (初学者或有经验的用户)。

(6) 应用面广。SAS 系统适用于几乎任何应用的需要。SAS 系统可用来解决自然科学和社会科学领域中的各种问题, 如统计学、社会学、经济学、生物、医疗卫生等, 其功能包括图形显示、数据分析、数据输入、数据检查、数据管理、报表生成、统计计算、市场研究、销售管理、供应管理、调查分析和建立预测模型等。

1.1.2 SAS 系统结构组成

SAS 系统是由众多产品组成的模块化的大型集成系统, 其中 Base SAS 软件是 SAS 系统的基础核心。

下面对 SAS 系统中的一些模型进行简要的介绍。

(1) Base SAS。是 SAS 系统的基础, 提供 SAS 数据库管理功能, 所有其他的模块必须与之结合起来使用。模块中的一些基本过程和 SAS 函数能够完成基础的统计计算, 其中包括以下几个方面。

- 对数据进行标准化、求秩及有关统计。
- 生成并分析列联表。
- 计算概率分布函数、百分位数、样本统计量及产生随机函数。
- 输出与读入数据时可使用任何格式的报表。
- 绘制直方图、饼图、星形图, 还可以画散点图、曲线图、时间序列图等。

(2) SAS/STAT。这是一个完整可靠的统计分析模块。包括回归分析、方差分析、属性数据分析、多变量分析、判别分析、聚类分析、残存分析、心理测验分析和非参数分析等 9 类方法共 40 多个过程。

(3) SAS/ETS。是用于计量经济与时间序列分析的专用模块。利用该模块可建立各种统计模型, 进行所关心系统的模拟和预测。如主要分析领先经济指标的综合指数和道琼斯股票指数的趋势, 以及预测每月总用电需求量、健康保险索赔额等。还提供有许多处理时间序列数据的实用程序, 如时间频率转换和季节调整等。

(4) SAS/OR。是用于运筹学和工程管理的专用模块。该模块可以帮助人们实现对人力、时间以及其他各种资源的最佳利用。包括通用的线性规划、正数规划以及混合整数规划和非线性规划等方法。

(5) SAS/QC。是用于质量控制的模块。它提供从发现和明确问题所在及进行试验设计到过程控制图和进行过程的能力分析。还提供了一套全屏幕菜单系统, 以引导用户进行标准的统计过程控制和过程的能力分析。

(6) SAS/IML。用于矩阵运算的模块。它提供了功能强大的矩阵运算的编程语言, 用户可根据自己的需要编写各种矩阵运算的程序。

(7) SAS/GRAPH。这是一个强有力的绘图模块。它拥有多种绘图功能, 如直线图、二维和三维图、直方图、饼图、星形图、地理图和各种映像图等。除此之外还具有对图进行全屏幕编辑和修改等功能。

(8) SAS/FSP。这是一个用来进行数据处理的交互式菜单系统。可用来进行全屏幕的数据输入、编辑、查询及数据文件的创建等。

(9) SAS/AF。这是一个应用开发工具, 利用它的屏幕设计能力及 SCL 语言的处理能力, 可以快速开发各种功能强大的应用系统。

(10) SAS/ASSIST。该模块为 SAS 系统提供了面向任务的菜单驱动界面。它可免去用户学习 SAS 语言的负担, 同时生成的 SAS 程序既可辅助有经验的用户快速编写 SAS 程序, 又可帮助用户学习 SAS 语言。

(11) SAS/EIS。该模块是 SAS 系统中采用新兴的面向对象的编程模式的又一个开发工具。

(12) SAS/ACCESS。该模块是对目前许多流行病数据的接口集，它提供的与外部数据库的接口是透明和动态的。

(13) SAS/ENGLISH。该模块为用户提供利用自然英语获取信息的能力。

(14) SAS/CALC。是一个功能完善的多维电子表格模块，具有财务分析、数据建模、数据组合及管理等功能。

(15) SAS/INSIGHT。该模块为可视化数据探索工具。

(16) SAS/LAB。这是菜单驱动面向任务的解释引导式数据分析模块。

(17) SAS/CPE。这是计算机及网络系统运行状况分析评估模块。

(18) SAS/GIS。这是集地理信息系统功能与空间数据的显示分析为一体的模块。

(19) SAS/SHARE。这是 SAS 系统中进行数据库的并发性控制的专门模块。

(20) SAS/CONNECT。在标准网络环境下，通过该模块可使各平台的 SAS 系统建立内在联系，实现分布处理，从而有效地利用各平台数据和机器资源。

(21) SAS/NVISION。这是用于高标准影像制品的设计和制作的模块。

(22) SAS/SPECTRAVIEW。这是进行多维数据观察、分析、研究的一个交互立体可视化工具。

(23) SAS/TOOLKIT。这是 SAS 功能扩充工具，用户可使用各种高级语言编写 SAS 过程、函数和输入输出格式等，然后通过该模型转换为 SAS 的组成部分。

1.1.3 SAS 系统资料

SAS 系统提供有非常完整的资料，主要有以下一些指南：SAS 入门、SAS 语言、SAS 过程、SAS/STAT、SAS/GRAPH、SAS/IML、SAS/AF、SAS/FSP、SAS 宏、SAS 表等。

1.1.4 SAS 软件要求

- 操作系统：安装者权限。为了确保 SAS9.1.2 全部安全的安装，安装者必须具有安装各种更新的管理权限。
- Windows XP。SAS9.1.2 支持使用 Windows XP 专业版。
- Windows 2000 系列。SAS9.1.2 支持 Windows 2000 专业版和 Windows 2000 服务器版本。
- 硬件要求：英特尔或英特尔奔腾 II 及以上处理器。
- 内存要求：Windows、Windows NT、Windows 2000、Windows XP 处理器的内存要求为最低 128M 或更大，交换文件的最小空间为 128M。
- 支持监视器：SVGA (800 × 600 或更高分辨率)。

1.2 SAS for Windows 的安装和启动

SAS 的安装盘一般包括 SAS Setup Disk、Software Disk 1、Software Disk 2、Software Disk 3、Software Disk 4、Software Disk 5、SAS Shared Components，附带一张高密度的授权软盘。安装方法与其他 Windows 下的软件安装一样。双击 SETUP 安装盘中的 Setup.exe，启动安装程序，选择安装过程中使用的语言，如“中文（简体）”等；采用“终端用户安装步骤”，选择“检验系统需求”：系统检查到需要安装的组件；返回到主界面，选择“安装 SAS 软件”；接下来需要指定“SAS 安装数据”文件，简称 SID 文件，也就是常说的 License 文件；然后按照对话框的提示选择安装目录，根据提示插入正确的安装盘即可。

SAS 的启动常用方法有两种：一是快捷方式，即按创建 Windows 系统中应用程序的快捷方式的步骤，把 SAS 目录中的 SAS.EXE 拖放至桌面，形成快捷方式，用鼠标双击快捷方式即可启动 SAS。二是菜单方式，从开始菜单的程序文件夹中找到 SAS 系统文件夹，从中启动 SAS 系统。启动后，会出现图 1-2 所示的 SAS 运行界面，术语称为“SAS 工作空间 (SAS Application Workspace)”。它和其他的 Windows 应用程序一样，在一个主窗口内包含若干个子窗口，并有菜单条、工具栏、状态栏等。

SAS 有 3 个最重要的子窗口：程序窗口 (Program Editor)、日志窗口 (LOG)、结果输出口 (Output)。程序窗口的使用类似于 Windows 中的记事本程序，可以在其中编辑文本文件，主要是编辑 SAS 程序。日志窗口记录程序的运行情况、运行是成功还是出错，如果出错，则显示出错的地方。日志窗口中以红色显示的是错误信息。

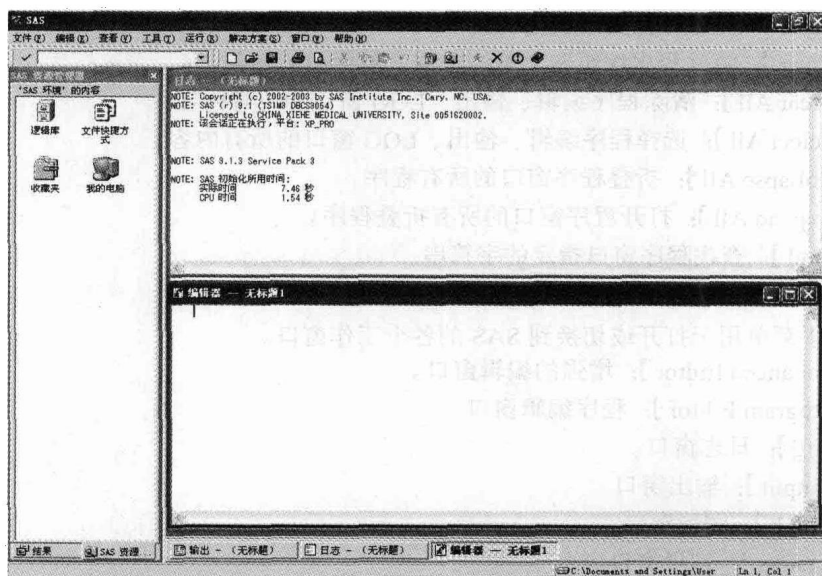


图 1-2 SAS 启动画面

1.3 SAS 菜单操作

SAS 主窗口标题栏下是主菜单。SAS 菜单是动态的，其内容随上下文而不同，即光标在不同的窗口，其菜单也不同。

- 【File】菜单主要是有关 SAS 文件调入、保存、转换及打印等功能命令。
 - 【New】：清除当前窗口（光标所在窗口）中的内容。
 - 【Open】：调入 SAS 程序（只限于程序编辑窗口）。
 - 【Import Data】：导入其他软件的数据接口。
 - 【Save】：按指定的路径和给定的文件名将文件存入外部设备（通常所存的内容为 3 个窗口的内容）。
 - 【Save As】：将同一个文件存到另一个路径下，或用另一个文件名存到相同路径下或不同路径下。
 - 【Export Data】：导出数据接口，将 SAS 数据导成其他软件的数据。
 - 【Page Setup】：纸张的设置，选择不同种类的纸张，横向或纵向打印。
 - 【Print Setup】：打印设置窗口。

- 【 Print Preview 】: 打印预览, 预先浏览排版的效果。
- 【 Print 】: 打印当前窗口的内容。
- 【 Send Email 】: 发送电子邮件。
- 【 Exit 】: 结束 SAS 工作, 退出 SAS 系统。
- 【 Edit 】菜单为文本编辑“菜单”, 可以对文本进行清空、复制、剪切、粘贴、查找字符串和检查单词的拼写错误等操作。
 - 【 Undo 】: 撤销前一次的编辑操作。
 - 【 Deselect 】: 取消所做的选择。
 - 【 Cut 】: 剪切所标记的文本。
 - 【 Copy 】: 复制被标记后存在剪贴板上的文本。
 - 【 Paste 】: 粘贴存在剪贴板上的文本。
 - 【 Clear 】: 清除程序编辑和文本编辑窗口被标记的部分。
 - 【 Clear All 】: 清除程序编辑、输出、LOG 窗口的所有内容。
 - 【 Select All 】: 选择程序编辑、输出、LOG 窗口的所有内容。
 - 【 Collapse All 】: 折叠程序窗口的所有程序。
 - 【 Expand All 】: 打开程序窗口的所有折叠程序。
 - 【 Find 】: 查找程序窗口指定的字符串。
 - 【 Replace 】: 替换指定的字符串。
- 【 View 】菜单用于打开或切换到 SAS 的各个工作窗口。
 - 【 Enhanced Editor 】: 增强的编辑窗口。
 - 【 Program Editor 】: 程序编辑窗口。
 - 【 Log 】: 日志窗口。
 - 【 Output 】: 输出窗口。
 - 【 Graph 】: 图形窗口。
 - 【 Results 】: 结果窗口。
 - 【 Explorer 】: 资源管理窗口。
 - 【 Contents Only 】: 单独的结构窗口。

还有, 【 Tools 】菜单用于打开一些 SAS 提供的小工具, 如图像编辑器, 还包含修改 SAS 运行选项的功能。【 Run 】菜单用于程序的执行、远程调用等, 只有在当前窗口是程序窗口时有效。【 Solutions 】菜单是 SAS 的一些图形界面操作的模块入口, 比如 SAS/INSIGHT、ASSIST 等。

1.4 SAS 系统的文件管理

所有的 SAS 文件都是一个 SAS 数据库的成员, 每个 SAS 数据库都有库标记, SAS 文件库标记就是同存储这些文件的目录联系的名字。在一个目录里, 所有的 SAS 文件属于相同的 SAS 数据库, 因此, 一个目录里的所有 SAS 文件能够用一个库标记来引用。

SAS 中的数据库分为永久性和临时性两种。顾名思义, 存在于永久库中的数据是永久存在的 (只要不去删除它), 临时库中的数据在退出 SAS 后会自动被删除。在 Windows 操作系统中, SAS 数据库是以路径为基础的方式组织。为了使用 SAS 数据库, 需要为每个 SAS 数据库指定一个库标记以便识别。这一标记又称库逻辑名, 是某个文件所在路径的别名。临时数据库的库标记为 WORK, 在 SAS 启动后自动生成, 结束 SAS 后, 库中的所有文件都会被删除。用 LIBNAME 定义的库逻辑名则为永久型数据库。它的一般形式如下: