

统计分析系统 SAS与SPSS



何宁 吴黎兵 滕冲 主编

本书为教师配有电子教案



机械工业出版社
China Machine Press

统计分析系统 SAS与SPSS



清华大学出版社

清华大学出版社

清华大学出版社

C819/77

2008

统计分析系统 SAS与SPSS



何 宁 吴黎兵 滕 冲 主编
余艳霞 汪同庆 王广平 谭明新 刘德喜 周 成 参编



机械工业出版社
China Machine Press

本书讲解了统计分析软件的基本用法,并以 SAS 和 SPSS 为例讲解了具体的数据分析和处理方法。主要内容包括 SAS 系统概述、SAS 编程基础及数据集建立、描述性统计分析、统计绘图、相关分析、假设检验、回归分析、主成分分析和因子分析、时间序列分析和预测、判别分析、聚类分析、SPSS 统计分析软件使用基础、SPSS 常用统计分析方法及应用。

本书适用于理工科农林类、经济类、文科类院校的学生学习统计分析使用,也可供使用计算机进行数据统计分析与课题研究的教师和科研人员参考。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

图书在版编目(CIP)数据

统计分析系统 SAS 与 SPSS/何宁,吴黎兵,滕冲主编. —北京:机械工业出版社,2008.3

ISBN 978-7-111-23662-7

I. 统… II. ①何… ②吴… ③滕… III. 统计分析—软件包, SAS、SPSS IV. C819

中国版本图书馆 CIP 数据核字(2008)第 029242 号

机械工业出版社(北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑:王 璐

北京市慧美印刷有限公司印刷 · 新华书店北京发行所发行

2008 年 3 月第 1 版第 1 次印刷

184mm × 260mm · 19 印张

标准书号: ISBN 978-7-111-23662-7

定价: 32.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换
本社购书热线:(010)68326294

计算机基础课程系列教材

编 委 会

主 任：刘 国 唐碧龙

副主任：刘春燕 汪同庆 何 宁

委 员：熊建强 黄文斌 康 卓 吴黎兵

黄 华 高建华 熊素萍

前言

面对信息时代下的海量数据,如何获取有价值的信息,从而作出科学的评估和决策?数据库类软件支持用户将日常数据存入计算机并在需要的时候进行快速访问,是信息处理的基础;统计分析软件则从原有信息中提取有用数据进行分析,帮助用户决策。不仅农林类院校、经济类院校的教师和科研人员需要使用计算机进行数据统计分析与课题研究,文科类院校的师生也面临这一需求。另外,社会上许多行业的数据分析人士也需要使用计算机进行数据统计分析。

SAS(Statistical Analysis System)是世界上著名的统计分析软件之一。它是一个集数据整理、数据分析、输出结果等功能于一体的统计分析软件包。SAS自诞生以来,经过不断发展,已由最初的统计分析系统演变为大型的集成应用软件系统。SAS灵活方便、功能齐全,在国际上被誉为数据分析的标准软件。在充斥着大量数据资料等待处理的信息时代,SAS系统在医学、心理学、生物学、经济学、体育、农林、商业、金融等领域中得到越来越广泛的应用。

SAS的版本更新很快,本书内容以SAS for Windows V9中文版为操作平台,介绍数据集的建立、基本统计分析、多元统计分析和统计绘图等内容。

本书还将介绍利用SPSS软件进行数据文件的建立与编辑,以及常用统计分析方法及其应用。

统计分析软件(SAS和SPSS)已经将统计分析算法编成了优秀的软件,使用者不需要记住统计分析公式的推导和证明,只需要弄清相关统计量的来源的概念,就可以轻松掌握统计分析方法、处理实验数据。因此,本书在每个章节中,简要介绍了基本概念和统计分析方法,没有对分析原理作详细介绍。有关统计分析原理的知识,读者可以参考相关的书籍。

本书侧重于统计功能、统计分析与工具软件操作相结合的应用,可以作为高等院校相关专业本科或硕士研究生的教材或参考书,也可以作为从事数据分析和统计工作人员的工具书。

第1章、第7章和第10章由吴黎兵编写;第2章由汪同庆编写;第3章由王广平编写;第4章和第9章由余艳霞编写;第5章由谭明新编写;第6章由滕冲编写;第8章、第11章和第12章由何宁编写;第13章由周成编写;第14章由滕冲和刘德喜编写。

本书的编写得到了武汉大学计算中心、武汉大学东湖分校的领导和老师的大力支持,在此表示衷心的感谢。同时,感谢机械工业出版社对本书出版给予的支持。

由于编者水平有限,书中难免有错漏之处,欢迎大家提出宝贵意见。

编 者

2008年1月



前言

第1章 SAS 概述	1
1.1 SAS 的组成与特点	1
1.1.1 SAS 的组成	1
1.1.2 SAS 的特点	3
1.2 SAS 的安装	5
1.3 SAS 的启动退出及重要的系统文件	6
1.3.1 启动 SAS	6
1.3.2 退出 SAS	7
1.3.3 SAS 的重要文件	7
1.4 SAS 的主界面介绍	9
1.4.1 SAS 的菜单栏	9
1.4.2 SAS 的工具栏	13
1.4.3 资源管理器和输出结果窗口	13
1.4.4 SAS 的四个基本视窗	14
1.4.5 其他	16
1.5 SAS/ASSIST 窗口介绍	16
1.5.1 启动 SAS/ASSIST	16
1.5.2 ASSIST 各菜单项介绍	17
1.5.3 数据分析示例	19
1.5.4 简单报表图形示例	20
第2章 SAS 编程基础及数据集的建立	23
2.1 SAS 语言初步	23
2.1.1 常量	23
2.1.2 观测值	23
2.1.3 变量	23
2.1.4 数据集	26
2.2 用 SAS 数据步建立数据集	26
2.2.1 临时数据集的建立	27

2.2.2 永久数据集的建立	29
2.3 用 VIEWTABLE 窗口建立数据集	30
2.3.1 生成新数据集	30
2.3.2 打开已有的数据集	32
2.3.3 编辑表	32
2.3.4 数据操作	33
2.3.5 改变数据显示	35
2.3.6 从外部文件导入数据	35
2.3.7 导出数据到其他数据格式文件	37
2.4 用 SAS/INSIGHT 模块建立数据集	37
2.4.1 数据集的建立	37
2.4.2 数据集的保存	39
2.4.3 数据集的打开	39
第3章 描述性统计分析	42
3.1 描述性统计分析概述	42
3.1.1 基本统计概念	42
3.1.2 UNIVARIATE 过程	44
3.1.3 应用举例	45
3.2 用 SAS/ASSIST 进行描述性统计分析	48
3.2.1 确定工作任务	48
3.2.2 数据集的确定	50
3.2.3 变量的确定及统计参数的选择	53
3.2.4 任务的提交与统计结果	55
3.3 用 SAS/INSIGHT 进行分布检验	56
3.3.1 确定工作任务	56
3.3.2 进行分布检验	58
3.4 用分析家进行列联表分析	59
3.4.1 确定工作任务	60
3.4.2 关联分析	62
第4章 统计绘图	64
4.1 用 INSIGHT 绘制图形	64

4.1.1 生成条形图和直方图	64	6.3.1 用分析家来作 t 检验	107
4.1.2 生成盒形图和马赛克图	67	6.3.2 用 ASSIST 作 t 检验	110
4.1.3 生成散点图	71	6.4 独立样本的 t 检验	111
4.1.4 生成连线图	72	6.4.1 用分析家来作独立样本的 t 检验	111
4.2 用“分析家”绘制图形	72	6.4.2 用 ASSIST 作参数检验	114
4.2.1 条形图	73	6.4.3 用程序中 ttest 过程做配对样本的 t 检验	115
4.2.2 盒图	75		
4.2.3 饼分图	76		
4.2.4 散点(线)图	77		
4.2.5 直方图	79		
4.2.6 曲面图	79		
4.3 用编程方法绘制图形	80	第7章 方差分析	116
4.3.1 用 GCHART 过程绘图	80	7.1 单因素方差分析	116
4.3.2 用 GPLOT 过程绘图	81	7.1.1 方差分析的问题与模型	116
4.3.3 用 G3D 过程绘图	83	7.1.2 编程作单因素方差分析	117
		7.1.3 用 Insight 作单因素方差分析	122
第5章 相关分析	85	7.1.4 用 ASSIST 作单因素方差分析	124
5.1 相关分析的概念及数学模型	85	7.2 多因素方差分析	126
5.1.1 相关分析的概念	85	7.2.1 方差分析的问题与模型	127
5.1.2 相关分析的数学模型	85	7.2.2 编程作多因素方差分析	129
5.2 两个变量间的直线相关分析	86	7.2.3 用 Insight 作多因素方差分析	132
5.2.1 用 S 相关分析	86	7.2.4 用 ASSIST 作多因素方差分析	135
5.2.2 用分析家作直线相关分析	90		
5.3 偏相关分析	95	第8章 回归分析	137
5.3.1 用 ASSIST 作偏相关分析	95	8.1 回归分析概述	137
5.3.2 用分析家作偏相关分析	96	8.1.1 回归模型	137
5.4 等级相关	97	8.1.2 回归方程的假设检验与自变量的 选择	138
5.4.1 用 ASSIST 作等级相关	97	8.1.3 基本回归诊断	139
5.4.2 用分析家作等级相关	98	8.2 用编程拟合线性回归	140
5.4.3 通过编程计算相关系数	99	8.3 用分析家拟合线性回归	143
第6章 假设检验	101	8.4 用 Insight 拟合多项式回归	149
6.1 假设检验的概念及数学模型	101	8.4.1 拟合多项式回归的方法	149
6.1.1 假设检验的概念	101	8.4.2 拟合多项式回归	150
6.1.2 假设检验的数学模型	102	8.5 用 ASSIST 拟合多元线性回归	152
6.2 单样本 t 检验	103	8.6 Logistic 回归	155
6.2.1 用分析家作单样本均值检验	103	8.6.1 Logistic 回归的模型	155
6.2.2 用程序中 means 过程做单样本的 t 检验	106	8.6.2 用分析家作 Logistic 回归	157
6.3 配对样本的 t 检验	107	8.6.3 用程序的方法拟合 Logistic 回归	159
		第9章 主成分分析和因子分析	164
		9.1 主成分分析	164

9.1.1 使用 INSIGHT 模块作主成分分析	164	第 13 章 SPSS 统计分析软件使用基础 ...	239
9.1.2 使用“分析家”作主成分分析 ...	167	13.1 概述	239
9.1.3 使用 PRINCOMP 过程进行主成分分析	171	13.1.1 SPSS 的发展简史	239
9.2 因子分析	173	13.1.2 SPSS 的主要特点	239
9.2.1 使用 INSIGHT 模块作因子分析	173	13.1.3 SPSS 软件的安装、启动和退出	240
9.2.2 用 FACTOR 过程进行因子分析	178	13.1.4 SPSS 的帮助系统	242
第 10 章 时间序列分析和预测	184	13.2 SPSS 使用基础	244
10.1 时间序列分析	184	13.2.1 三种运行方式	244
10.1.1 建立时间序列数据集	185	13.2.2 SPSS 的窗口及其设置	244
10.1.2 季节性调整分析	186	13.3 SPSS 数据结构及其定义	245
10.1.3 回归分析	193	13.4 SPSS 数据的录入和编辑	249
10.2 时间序列预测	197	13.4.1 SPSS 数据的内容	250
10.2.1 设置数据集及相关信息	197	13.4.2 SPSS 数据的录入	250
10.2.2 选择预测模型	197	13.4.3 SPSS 数据的编辑	251
10.2.3 生成和查看预测结果	199	13.5 数据文件的操作	252
10.2.4 管理预测项目	200	13.5.1 数据文件的保存	252
第 11 章 判别分析	202	13.5.2 获取外部数据的方法	254
11.1 多类判别分析	202	13.6 小结	260
11.1.1 分析原理	202	第 14 章 SPSS 常用统计分析及应用	261
11.1.2 判别分析的假设检验	204	14.1 基本描述统计量	261
11.1.3 鉴别错误分类的比率	205	14.1.1 描述集中趋势的统计量	261
11.1.4 DISCRIM 过程	205	14.1.2 描述离散趋势的统计量	262
11.2 逐步判别过程分析过程 STEPDISC ...	212	14.1.3 描述分布形态的统计量	263
11.2.1 逐步判别分析方法	212	14.2 正态分布的重要特征和经验法则 ...	263
11.2.2 STEPDISC 过程	213	14.3 基本统计分析	264
第 12 章 聚类分析	221	14.3.1 采用 Frequencies 过程进行频数分析	265
12.1 聚类分析概述	221	14.3.2 采用 Descriptives 过程描述分析	269
12.2 SAS 聚类分析过程	223	14.3.3 采用 Explore 过程进行探索分析	271
12.2.1 CLUSTER 聚类分析过程	223	14.4 假设检验和 t 检验	275
12.2.2 FASTCLUS 快速聚类过程	224	14.4.1 假设检验概述	275
12.2.3 VARCLUS 聚类过程	225	14.4.2 单样本的 t 检验	276
12.2.4 TREE 聚类树型输出过程	226	14.4.3 两独立样本的 t 检验	278
12.3 应用举例	227	14.4.4 两配对样本的 t 检验	280

14.5 方差分析	281
14.5.1 方差分析概述	282
14.5.2 单因素方差分析的主要思想和步骤	282
14.5.3 SPSS 中单因素方差分析的基本操作	283
14.5.4 多因素方差分析	289
14.6 非参数检验	290

14.6.1 卡方检验	291
14.6.2 二项分布检验	291
14.6.3 游程检验	292
14.6.4 单样本 K-S 检验	292
14.6.5 两独立样本的非参数检验	292
14.7 小结	293
参考文献	294

第14章 SPSS 常用统计方法

14.1 基本统计量	291
14.1.1 描述统计量	291
14.1.2 描述统计量的计算	292
14.1.3 描述统计量的应用	293
14.2 正态分布的重要特征和检验	294
14.3 非参数统计	294
14.3.1 非参数统计的应用	295
14.3.2 非参数统计的检验	295
14.3.3 非参数统计的 SPSS 操作	295
14.4 聚类分析	296
14.4.1 聚类分析的概念	296
14.4.2 聚类分析的方法	296
14.4.3 聚类分析的应用	296
14.4.4 聚类分析的 SPSS 操作	296

14.5 判别分析	297
14.5.1 判别分析的概念	297
14.5.2 判别分析的方法	297
14.5.3 判别分析的应用	297
14.5.4 判别分析的 SPSS 操作	297
14.6 主成分分析	298
14.6.1 主成分分析的概念	298
14.6.2 主成分分析的方法	298
14.6.3 主成分分析的应用	298
14.6.4 主成分分析的 SPSS 操作	298
14.7 因子分析	299
14.7.1 因子分析的概念	299
14.7.2 因子分析的方法	299
14.7.3 因子分析的应用	299
14.7.4 因子分析的 SPSS 操作	299

第1章 SAS 概述

统计分析系统(Statistical Analysis System, SAS)是当今国际上流行的数据统计分析软件系统。SAS 最早由美国北卡罗纳州州立大学开发。1976 年, SAS 软件研究所(SAS Institute Inc.)成立, 开始进行 SAS 的开发、维护、销售和培训工作。早期的 SAS 只能在大型机上运行, 现在 SAS 可以在各种大型机、小型机、工作站和微机上运行, 其 6.06 以后的版本支持 Windows 系列 32 位操作系统。在 Windows 环境下运行的 SAS 充分利用了 Windows 操作系统良好的人机界面, 大大方便了程序编辑、数据操作, 并提供了多个不需编程而直接通过菜单调用的模块。SAS 自诞生以来, 经过不断发展, 已由最初的统计分析系统演变为大型的集成应用软件系统。SAS 是一个灵活方便, 且功能齐全的整理数据、分析数据及报告数据分析结果的软件系统, 在国际上被誉为数据分析的标准软件。在充斥着大量数据资料等待处理的信息时代, SAS 正在各个领域得到广泛应用。

1.1 SAS 的组成与特点

1.1.1 SAS 的组成

SAS 是由众多模块组成的系统。其中, Base SAS 模块是 SAS 的核心。其他各模块均在 Base SAS 提供的环境中运行。用户可选择需要的模块与 Base SAS 一起构成一个用户化的 SAS。下面介绍 SAS 的功能模块。

• Base SAS

Base SAS 是 SAS 的核心, 负责数据管理、交互应用环境管理、进行用户语言处理以及调用其他 SAS 模块。Base SAS 还可进行基本的描述性统计, 基相关系数的计算, 以及进行正态分布检验等。Base SAS 为 SAS 的数据库提供了丰富的数据管理功能, 并支持用标准的 SQL 语言对数据进行操作。Base SAS 不仅能够制作简单列表, 也可制作比较复杂的统计报表。

• SAS/GHARH

SAS/GHARH 可将数据及其包含着的深层信息以多种图形生动地呈现出来, 如直方图、圆饼图、星形图、散点相关图、曲线图、三维曲面图、等高线图及地理图等。SAS/GHARH 提供一个全屏幕编辑器, 提供多种设备程序, 支持各种类型的图形输出设备以及标准的图形交换文件。

• SAS/ASSIST

SAS/ASSIST 为 SAS 提供了面向任务的菜单界面, 借助它可以通过菜单系统使用 SAS 的其他产品。它自动生成的 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。

• SAS/AF

SAS/AF 是一个应用开发工具。用户使用 SAS/AF 可将包含众多功能的 SAS 软件作为方法库,

利用 SAS/AF 的屏幕设计能力以及屏幕控制 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。SAS/AF 由于采用了 OOP(面向对象编程)技术,可使用户方便快速地开发各类具有图形用户界面(GUI)的应用系统。

- SAS/EIS

SAS/EIS 是决策工具,也是一个快速应用开发工具。SAS/EIS 完全采用新兴的面向对象的编程模式(OOP)。EIS 以生动直观的方式(图或表)将关键性或总结性信息呈现给使用者。

- SAS/ACCESS

为了对不同格式的数据进行查询、访问和分析,SAS/ACCESS 提供了与许多流行数据库软件的接口,利用 SAS/ACCESS 可建立一个访问其他外部数据库的统一的公共数据界面。SAS/ACCESS 提供的接口是透明的和动态的,用户一般不必将外部数据文件转换为 SAS 数据集,而只需在 SAS 中建立对外部数据的描述文件。对一些经常使用的外部数据,可以利用 SAS/ACCESS 提取进入 SAS 数据库。SAS/ACCESS 提供的接口是双向的:既可将数据读入 SAS,也可在 SAS 中更新外部数据或将 SAS 数据加载到外部数据库中。目前,SAS/ACCESS 支持的数据库主要有 IML-DL/I、SQL/DS、DB2、ADABAS、Rdb、Oracle、Sybase、INGRES、Informix、DBF/DIF、ODBC 等。

- SAS/STAT

SAS/STAT 覆盖了所有实用的数理统计分析方法,是国际统计分析领域的标准软件。SAS/STAT 提供了十多个过程,可进行各种不同模型和不同特点数据的回归分析,如正交回归/面回归、响应面回归、logistic 回归、非线性回归等,且具有多种模型选择方法。它可处理实型数据、有序数据和属性数据,并能产生各种有用的统计量和诊断信息。在方差分析方面,SAS/STAT 为多种试验设计模型提供了方差分析工具。另外,它还有处理一般线性模型和广义线性模型的专用过程。在多变量统计方面,SAS/STAT 为主成分分析、典型相关分析、判别分析和因子分析提供了许多专用过程。SAS/STAT 还包含多种聚类准则的聚类分析方法。

- SAS/QC

SAS/QC 为全面质量管理提供了一系列工具。它也提供一套全屏幕菜单系统引导用户进行标准的统计过程和试验设计。SAS/QC 提供了多种不同类型控制图的制作与分析功能。Pareto 图(排列图)可用于发现需优先考虑的因素,Ishikawa 图可用于直观地进行因果分析。

- SAS/ETS

SAS/ETS 提供丰富的计量经济学和时间序列分析方法,是研究复杂系统和进行预测的有力工具。它提供方便的模型设定手段、多样的参数估计方法。

- SAS/OR

SAS/OR 提供全面的运筹学方法,是一种功能强大的决策支持工具。它帮助人们实现人力、时间以及其他各种资源的最佳利用。SAS/OR 包含通用的线性规划、混合整数规划和非线性规划的求解,也为专门的规划问题提供更为直接的解决办法,如网络流问题、运输问题、分配问题等。

- SAS/IML

SAS/IML 提供功能强大的面向矩阵运算的编程语言,能帮助用户研究新算法、解决 SAS 中没有现成算法的专门问题。SAS/IML 中的基本数据元素是矩阵,它包含大量的数学运算符、函数和例行程序,用户用很少的语句便可执行很复杂的计算过程。

● SAS/WA

SAS/WA (Warehouse Administrator) 是建立数据仓库的集成工具, 它在其他 SAS 软件的基础上提供了一个建立数据仓库的管理层, 包括定义数据仓库和主题, 数据转换和汇总, 更新汇总数据, Metadata 的建立、管理和查询, Data marts 和 Info marts 的实现。

● SAS/MDDB Server

SAS/MDDB Server 是 SAS 的多维数据库产品, 主要用于在线分析处理 (OLAP), 可将从数据仓库或其他数据源导入的数据以立体阵列的方式存储, 以便用多维数据浏览器等工具快速和方便地访问。

● SAS/IntrNet

SAS/IntrNet 为 SAS Web 应用提供了数据服务和计算服务, 包括 htmSQL。它是 UNIX Web 服务器的 CGI 程序, 实现了通过支持 Web 浏览器动态查询 SAS 数据或外部的关系型数据库; SAS ODBC Driver 实现了通过支持 ODBC 的 Windows Web 服务器来访问 SAS 数据; SAS Driver for JDBC 实现了通过 Java applet 来查询 SAS 数据; SAS/IntrNet Application Dispatcher 实现了通过 Web 浏览器动态地递交 SAS 程序到 SAS 应用服务器执行, 并将结果返回浏览器。

● SAS/GIS

SAS/GIS 是集地理信息系统与空间数据显示分析于一体的软件。它提供层次化的地理信息, 每一层可以是某些地理元素, 也可与用户定义的主题 (例如, 人口、产值等) 相关联。用户可交互式地缩小或放大地图, 设定各层次显示与否, 并利用各种交互式工具进行数据显示与分析。

● SAS/ITSV

ITSV (IT Service Vision) 是用于企业的全面 IT 服务的性能评估和管理的软件, 这些 IT 服务包括计算机系统、网络系统、Web 服务器和电话系统等。ITSV 将不同来源的数据进行整理和组织, 存放于性能数据仓库中, 用 GUI 或批处理的方式生成的报告。系统程序员及网络工程师能据此识别、研究并解决有关问题, 业务分析人员能据此制定资源管理的总体策略, CIO 和数据中心经理能据此定期地得到所需的 IT 运作的汇总和分析报告。

● SAS/CFO Vision

SAS/CFO Vision 用于财务整合和报告, 它包含了会计知识, 为日常财务工作提供了现成的程序, 并提供了访问所有主要数据源的接口。它主要用于访问财务和非财务的有关信息整合财务数据, 通过一个财务信息仓库来管理业务结构, 通过财务报告和分析帮助理解财务的结果, 并在组织内交流关键的业务结果信息。

1.1.2 SAS 的特点

1. 信息存储简便灵活

SAS 能和绝大多数数据库系统交换数据信息, 具有很强的数据共享能力。将外部数据导入到 SAS 后, 用户就可以利用数据管理功能对数据进行更新、重整、组合编辑或求数据子集, 从而得心应手地分析数据。

2. 语言编程能力强

SAS 语言功能强大, 内含丰富的算术逻辑运算符和 100 多种函数, 用户可用赋值语句、条件语句、循环语句和数组等对变量进行各种操作。

SAS 语言有两类, DATA 语句和 PROC 语句。SAS 程序以 DATA 语句开始, 输入或处理有关的数据集(SAS 的数据文件), 让系统读有关数据。

3. 丰富的统计分析方法

统计分析是 SAS 的主要功能, 它几乎囊括了所有的统计方法, 从基本统计到多元分析, 从预测预报到运筹决策, 从时序分析到经济计量等。每一个统计分析方法在 SAS 中是一个独立的过程。各个算法过程既相互独立, 又相互联系, 一次统计分析过程可以完成多个算法, 前一过程的输出结果可以作为后一过程的输入信息, 中间不需要中断系统的运行, 因而可以从不同的角度对数据进行分析比较, 逐层次揭示出数据的蕴涵关系。

SAS 具有第四代计算机语言的特点, 只需通过菜单的选择、对话框的操作就能告诉系统做什么, 无需告之怎样做。只要初步了解统计分析原理, 无需通晓统计分析的各种算法, 即可得到统计分析结果。

4. 较强的统计报表与绘图功能

SAS 支持绘制一维到三维表格, 它既能以各种格式输入数据, 也能以任何方式输出数据。除了固定的报表格式外, SAS 还支持用户自行设计任何格式的报表。

SAS 可以绘制各种统计图形。它可以在图形中进行各种方式的标识, 并支持用彩色显示器和绘图仪显示图形。

5. 友好的用户界面

SAS 提供了一系列灵活的菜单驱动、面向任务的图形界面, 初学者和熟练的 SAS 用户都可以全面地进行以数据为中心的基本处理。SAS 采用多窗口的操作环境和显示方式, 运行一个 SAS 程序时, 用户可以从不同的窗口了解程序的运行情况、出错信息和输出结果。数据和文件也有相关的管理和显示窗口。通过窗口还可以了解到 SAS 的组成、功能键的作用等。

用户可以利用菜单或在命令行上键入命令来显示窗口, 改变窗口的形状、大小及在屏幕中的位置。SAS 还提供了一些有效的工具, 如图形/图像编辑器、文本编辑器、报表编辑器、数据表编辑器等。利用这些工具向导, 用户可以按照提示快速地创建或生成数据集、图形、报表等。

6. 宏功能

SAS 有较强的宏代换功能。如果用户需要多次做类似的工作, 而仅是参数不同, 则可以使用宏功能定义宏体, 在宏体中可以使用宏变量。随后用户就可以使用不同的参数载入宏体, 从而大大简化程序的编写。

7. 支持分布式处理

SAS 支持分布式处理和多台计算机协同工作, 从而有效减少对大数据量进行统计分析所需的时间, 并能充分利用网络环境中的软硬件资源。

8. 采用输出分发系统

从 8.0 版以后, SAS 的数据分析模块功能有较大提高。一个显著的特点是所有的输出都由输出分发系统处理。该系统允许用户定制显示哪些结果, 如显示中间的数据集、数据集的属性和变量名、统计分析结果等。输出分发系统可以生成多种格式的结果, 包括 HTML 文件。

9. 功能强大的系统阅读器

SAS 阅读器是一个浏览各类 SAS 文件的应用软件。它提供了一个快速、便利的查看打印数据集、目录、传输文件、SAS 程序、工作日志、输出结果等文件内容的通道。使用 SAS 阅读器时,

用户不必运行 SAS 甚至不需要在计算机中安装 SAS。

1.2 SAS 的安装

下面我们以 SAS 9.0 简体中文版为例,介绍 SAS 在 Windows XP 中的安装过程。先将 SAS 9.0 的光盘放入光驱中,系统会自动运行 SAS 的安装程序向导。如果不能自动播放,用户可双击光盘根目录下的 setup.exe 运行安装向导。安装程序首先提示用户选择安装语言,如图 1-1 所示。选择“中文(简体)”后单击“确定”按钮,打开 SAS 9.0 安装主界面,如图 1-2 所示。

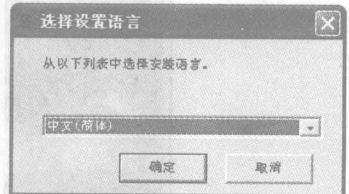


图 1-1 选择安装语言

在图 1-2 所示对话框中选择“安装 SAS 软件”开始安装 SAS。系统提供了三种安装方式:完全安装(Complete)、定制安装(Custom)、客户安装(Client)。完全安装 SAS 需要 622M 左右的硬盘空间。如果选择前两种安装方式,向导程序会询问用户是否想把 SAS 安装成一个 SAS 文件的镜像服务器,供以后采用客户安装的机器共享。客户安装方式可以不安装任何运行部分而只安装用户设置文件,运行 SAS 时从光盘或网络服务器上调用程序。定制安装可以让用户按自己的需要选择要安装的模块。在安装过程中要选择 SAS 安装的目标文件夹(默认为 C:\Program Files\SAS\SAS System\9.0),数据文件文件夹(默认为 C:\My Documents\My SAS Files\9.0)以及临时文件文件夹(默认为系统临时目录\SAS Temporary Files)。SAS 安装成功后,会在“开始/程序”菜单下创建“The SAS System”程序组,该程序组中包含的快捷方式如图 1-3 所示。

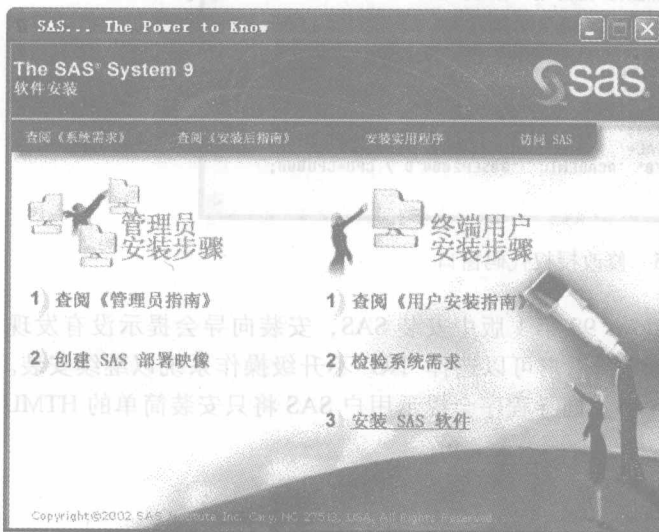


图 1-2 SAS 安装主界面

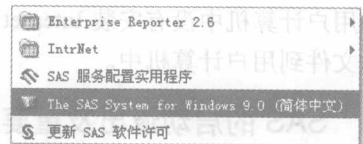


图 1-3 安装成功后生成的菜单项

安装结束后,用户可以选择图 1-3 中的“更新 SAS 软件许可”菜单项更新授权信息。在如图 1-4 所示的界面中,选择新的 SAS 安装数据文件,然后单击“下一步”更新 SAS 软件许可信息。

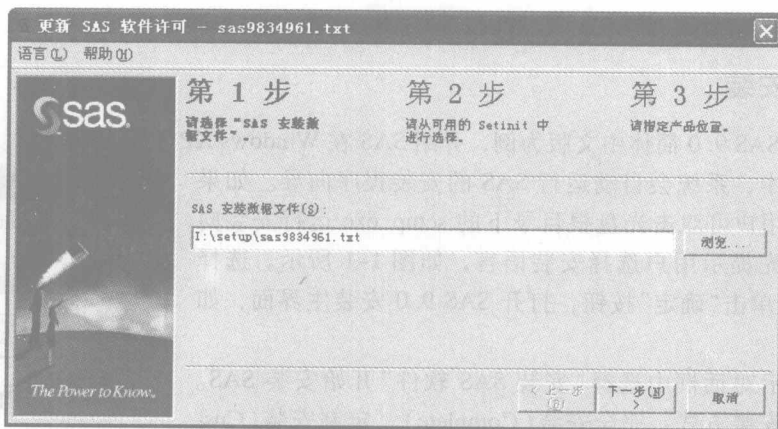


图 1-4 更新 SAS 软件许可

在安装过程中，用户可以选择手动更正授权信息，安装向导会给出如图 1-5 所示记事本编辑窗口供用户修改相关数据，修改完成后可直接关闭此窗口，向导程序执行 setinit.sas 开始更新软件许可。

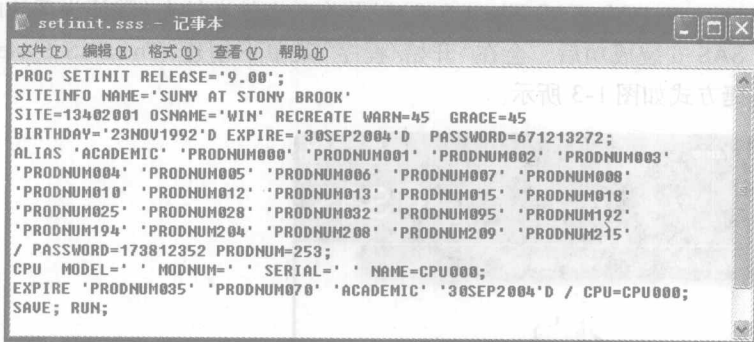


图 1-5 修改授权代码窗口

需要补充说明的是，如果用户在 Windows 98 第 1 版中安装 SAS，安装向导会提示没有发现 Windows 98 第 2 版或 Windows Service Pack 1，用户可以选择“NO”不升级操作系统以继续安装。如果用户计算机中没有安装 Internet Explorer 5，向导程序会提示用户 SAS 将只安装简单的 HTML 帮助文件到用户计算机中。

1.3 SAS 的启动退出及重要的系统文件

1.3.1 启动 SAS

在 Windows 系统中，单击图 1-3 中的“The SAS System for Windows 9.0(简体中文)”选项，即可启动 SAS。也可在“开始”菜单的“运行”对话框中输入命令“< SAS 安装路径 > \ sas.exe”运行 SAS。或在资源管理器中双击 SAS 安装目录中  sas.exe 文件图标启动 SAS。SAS 启动后，其主

窗口如图 1-6 所示。

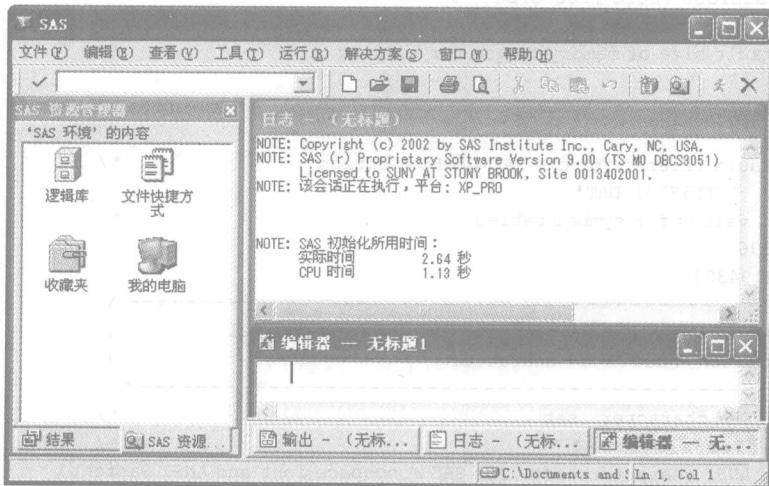


图 1-6 SAS 的主窗口

SAS 也可以用批处理的方式运行，方法是在“开始”菜单的“运行”对话框中输入下面的命令行：

```
<路径> \ sas.exe -SYSIN 文件名
```

其中，<SAS 安装路径>是 SAS 安装的位置，文件名是要批处理执行的 SAS 程序文件名。当运行 SAS 批处理时，有一个状态窗口同这个 SAS 作业相联系。这个窗口告诉用户 SAS 作业正在运行，并说明中间的记录和过程输出放在什么地方。这个窗口会一直保留到 SAS 作业完成为止。

1.3.2 退出 SAS

用户可用两种方法关闭 SAS 9.0：

- 单击“文件”菜单中的“退出”菜单项。
- 单击窗口右上角的关闭按钮。

执行以上操作后，系统会弹出如图 1-7 所示对话框，用户可单击“确定”按钮退出 SAS，也可以单击“取消”按钮返回 SAS 9.0。

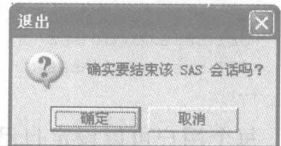


图 1-7 退出确认对话框

1.3.3 SAS 的重要文件

当 SAS 运行时，它可能要调用很多文件，不过其中有三个文件对用户比较重要，它们是 SAS 配置文件 (Sasv9. cfg)、SAS 自动执行文件 (Autoexec. sas) 和用户属性目录 (Profiles. sct)。

1. SAS 配置文件

在 SAS 9.0 中，SAS 的配置文件更名为 Sasv9. cfg，此文件用来设置 SAS 运行的选项，如 SAS 工作目录、根目录、用户属性目录的位置，用于排序的内存大小、制表线所用的字符等。其前半部分是用户可以修改的，后半部分是 SAS 安装程序设定的，用户最好不要修改。前半部分的示例如下：