

计算机公共课系列教材

统计分析系统 SAS

主 编 何 宁 吴 黎 兵



WUHAN UNIVERSITY PRESS
武汉大学出版社

计算机公共课系列教材

统计分析系统 SAS

主 编 何 宁 吴黎兵
参 编 余艳霞 宋 麟



WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

统计分析系统 SAS/何宁,吴黎兵主编. —武汉: 武汉大学出版社,
2005. 8

(计算机公共课系列教材)

ISBN 7-307-04661-X

I. 统… II. ①何… ②吴… III. 统计分析—应用软件, SAS—高等学校—教材 N. C812

中国版本图书馆 CIP 数据核字(2005)第 081856 号

责任编辑: 解云琳 黄 杨 责任校对: 王 建 版式设计: 支 笛

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: wdp4@whu.edu.cn 网址: www.wdp.whu.edu.cn)

印刷: 武汉大学出版社印刷总厂

开本: 787×980 1/16 印张: 20.5 字数: 420 千字

版次: 2005 年 8 月第 1 版 2005 年 8 月第 1 次印刷

ISBN 7-307-04661-X/C·149 定价: 29.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有缺页、倒页、脱页等质量问题, 请与当地图书销售部门联系调换。

计算机公共课系列教材

编委会

主任：杨健露

副主任：熊建强 李俊娥 殷朴

编委：（以姓氏笔画为序）

刘英 何宁 汪同庆 杨运伟

吴黎兵 罗云芳 黄文斌 康卓

执行编委：黄金文 杨华



内 容 简 介

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。本书共分为 12 章,主要介绍了数据集的建立、基本统计分析、多元统计分析和统计绘图等。本书可以作为高等院校相关专业本科或硕士研究生教材,也可以作为各行各业从事数据分析和统计工作人员的实用参考书。

前 言

随着计算机技术和信息技术的发展,信息的增长速度呈指数上升。面对海量数据,如何存储、整理和分析处理这些数据,从中发现有价值的信息或知识,是一件很重要又很复杂的工作。目前有各种各样的软件可以用来管理数据、分析数据: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快速访问它们,是信息处理的基础;统计分析软件则从原有的信息中提取有用的数据进行分析,帮助我们决策。不仅是理工科的农林类院校、经济类院校,文科类院校的教师和科研人员也需要使用计算机进行数据统计分析与课题研究,而且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,经过不断的发展,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、功能齐全的整理数据、分析数据及报告结果的软件系统,在国际上被誉为数据分析的标准软件。在充斥着大量等待处理的数据资料的信息时代,SAS正在各个领域得到广泛的应用。该软件可以应用在医学、心理学、生物学、经济学、体育、农林、商业、金融等各领域。

SAS for Windows的版本更新很快,本书内容以SAS for Windows V8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。为了便于初学者学习数理统计的有关知识,每一章节中均对涉及的统计分析方法作了简要的介绍。有关统计分析原理方面的知识,读者可以参考相关的书籍。

本书可以作为高等院校相关专业本科或硕士研究生自学、教学用书,也可以作为各行各业从事数据分析和统计的人员的重要工具书。

本书的第三章、第八章、第十章和第十二章由何宁编写;第一章、第七章、第十一章由吴黎兵编写;第二章、第四章和第九章由余艳霞编写;第五章和第六章由宋麟编写。

本书的编写得到了武汉大学教务部、计算机学院及计算中心领导的大力支持,许多老师在本书的编写过程中给予了帮助并提出了宝贵的意见,在此表示衷心的感谢。同时,感谢武汉大学出版社对本书的出版给予的支持。

由于编者水平有限,书中难免有错漏之处,欢迎大家批评、指正。

作者

2005年8月

目 录

第一章 SAS 简介	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	4
1.2 SAS 的安装	5
1.3 SAS 的启动、退出及重要的系统文件	7
1.3.1 启动 SAS	7
1.3.2 退出 SAS	8
1.3.3 SAS 的几个重要文件	9
1.4 SAS 的主要界面介绍	11
1.4.1 SAS 的菜单栏	11
1.4.2 SAS 的工具栏	15
1.4.3 Explorer 和 Results 窗口	16
1.4.4 SASforWindows 的四个基本视窗	17
1.4.5 其他	19
1.5 SAS/ASSIST 窗口介绍	20
1.5.1 启动 SAS/ASSIST	20
1.5.2 数据分析示例	22
1.5.3 简单报表图形示例	23
1.5.4 ASSIST 各菜单项介绍	27
第二章 SAS 数据集的建立	29
2.1 常量、变量、观察值与数据集	29
2.1.1 常量	29
2.1.2 观察值	29
2.1.3 变量	29
2.1.4 数据集	32
2.2 用 SAS 数据步建立数据集	33
2.2.1 临时数据集的建立	33

2.2.2	永久数据集的建立	36
2.3	用 VIEWTABLE 窗口建立数据集	38
2.3.1	生成新数据集	38
2.3.2	打开已有的数据集	40
2.3.3	编辑表	41
2.3.4	数据操作	43
2.3.5	改变数据显示	45
2.3.6	从外部文件导入数据	45
2.3.7	导出数据到其他数据格式文件	47
2.4	用 SAS/INSIGHT 模块建立数据集	48
2.4.1	数据集的建立	48
2.4.2	数据集的保存	51
2.4.3	数据集的打开	51
2.4.4	SAS/INSIGHT 功能概述	55
第三章 描述性统计分析		56
3.1	描述性统计分析概述	56
3.1.1	基本统计概念	56
3.1.2	UNIVARIATE 过程	58
3.1.3	应用举例	60
3.2	用 SAS/ASSIST 进行描述性统计分析	64
3.2.1	确定工作任务	64
3.2.2	数据集的确定	66
3.2.3	变量的确定及统计参数的选择	70
3.2.4	任务的提交与统计结果	73
3.3	用 SAS/INSIGHT 进行分布检验	74
3.3.1	确定工作任务	74
3.3.2	进行分布检验	77
3.4	用分析员应用进行列联表分析	79
3.4.1	确定工作任务	79
3.4.2	关联分析	81
第四章 统计绘图		86
4.1	用 INSIGHT 绘制图形	86
4.1.1	用 Histogram/BarChart(Y) 命令生成条形图	86
4.1.2	用 BoxPlot/MosaicPlot(Y) 命令生成盒形图和 Mosaic 图	90



4.1.3	用 ScatterPlot(YX)命令生成散点图	94
4.1.4	用 LinePlot(YX)命令生成连线图	97
4.2	用分析员应用模块绘制图形	98
4.2.1	条形图	98
4.2.2	盒形图	103
4.2.3	饼分图	104
4.2.4	散点图	105
4.2.5	直方图	107
4.2.6	曲面图	108
4.3	用编程方法绘制图形	109
4.3.1	用 GCHART 过程绘图	110
4.3.2	用 GPLOT 过程绘图	113
4.3.3	用 G3D 过程绘图	117
第五章	相关分析	120
5.1	相关分析的概念及数学模型	120
5.1.1	相关分析的概念	120
5.1.2	相关分析的数学模型	121
5.2	两个变量间的直线相关分析	122
5.2.1	用 ASSIST 作相关分析	122
5.2.2	用分析员应用作直线相关分析	126
5.3	偏相关分析	131
5.3.1	用 ASSIST 作偏相关分析	131
5.3.2	用分析员应用作偏相关分析	133
5.4	等级相关	134
5.4.1	用 ASSIST 作等级相关	134
5.4.2	用分析员应用作等级相关	136
5.4.3	用编程方法计算相关系数	137
第六章	假设检验	140
6.1	假设检验的概念及数学模型	140
6.1.1	假设检验的概念	140
6.1.2	假设检验的数学模型	142
6.2	单样本 t 检验	143
6.2.1	用分析员应用作单样本均值检验	143
6.2.2	用程序中 means 过程作单样本的 t 检验	147



6.3	配对样本的 t 检验	147
6.3.1	用分析员应用来作 t 检验	148
6.3.2	用 ASSIST 作 t 检验	154
6.3.3	用程序中 means 过程作配对样本的 t 检验	154
6.4	独立样本的 t 检验	155
6.4.1	用分析员应用来作独立样本的参数检验	156
6.4.2	用 ASSIST 作参数检验	161
6.4.3	用程序中的 ttest 过程作配对样本的 t 检验	161
第七章	方差分析	164
7.1	单因素方差分析	164
7.1.1	方差分析的问题与模型	164
7.1.2	编程作单因素方差分析	165
7.1.3	用 Insight 作单因素方差分析	171
7.1.4	用 ASSIST 作单因素方差分析	174
7.2	多因素方差分析	177
7.2.1	方差分析的问题与模型	177
7.2.2	编程作多因素方差分析	180
7.2.3	用 Insight 作多因素方差分析	184
7.2.4	用 ASSIST 作多因素方差分析	188
第八章	回归分析	191
8.1	回归分析概述	191
8.1.1	回归模型	191
8.1.2	回归方程的假设检验与自变量的选择	192
8.1.3	基本回归诊断	193
8.2	用编程方法拟合线性回归	195
8.2.1	REG 过程的基本语句	195
8.2.2	REG 过程的基本选项	196
8.2.3	应用举例	197
8.3	用分析员应用拟合线性回归	198
8.4	用 Insight 拟合多项式回归	207
8.4.1	拟合多项式回归的方法	207
8.4.2	拟合多项式回归	208
8.5	用 ASSIST 拟合多元线性回归	211
8.6	Logistic 回归	215

8.6.1	Logistic 回归的模型	215
8.6.2	用分析员应用作 Logistic 回归	217
8.6.3	用程序的方法拟合 Logistic 回归	221
第九章	主成分分析和因子分析	227
9.1	主成分分析	227
9.1.1	原理与数学模型	227
9.1.2	用 PRINCOMP 过程进行主成分分析	228
9.1.3	用菜单命令进行主成分分析	233
9.2	因子分析	234
9.2.1	因子分析模型	234
9.2.2	用 FACTOR 过程进行因子分析	236
第十章	判别分析	248
10.1	多类判别分析	248
10.1.1	分析原理	248
10.1.2	判别分析的假设检验	250
10.1.3	鉴别错误分类的比率	251
10.1.4	DISCRIM 过程	252
10.2	逐步判别分析过程 STEPDISC	261
10.2.1	逐步判别分析方法	261
10.2.2	STEPDISC 过程	262
第十一章	时间序列分析和预测	272
11.1	时间序列分析	273
11.1.1	建立时间序列数据集	273
11.1.2	季节性调整分析	275
11.1.3	回归分析	280
11.2	时间序列预测	287
11.2.1	设置数据集及相关信息	287
11.2.2	选择预测模型	287
11.2.3	生成和查看预测结果	289
11.2.4	管理预测项目	291

第十二章 聚类分析	293
12.1 聚类分析概述	293
12.2 SAS 聚类分析过程	295
12.2.1 CLUSTER 聚类分析过程	296
12.2.2 FASTCLUS 聚类过程	297
12.2.3 VARCLUS 聚类过程	298
12.2.4 TREE 过程	299
12.3 应用举例	300
参考文献	315



第一章 SAS 简介

SAS 是英文“Statistical Analysis System”的缩写,直译为“统计分析系统”,是当今国际上非常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 SAS 只能运行在大型机上,现在 SAS 可以在各种大型机、小型机、工作站和微机上运行,6.06 以后的版本都支持 Windows 系列 32 位操作系统。在 Windows 环境下运行的 SAS 充分利用了 Windows 操作系统友好的人机界面,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,经过不断的发展,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,功能齐全的数据整理、分析数据及报告结果的软件系统,在国际上被誉为数据分析的标准软件。在大量数据资料等待处理的信息时代,SAS 正在各个领域中得到越来越广泛的应用。

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。

1. Base SAS

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

2. SAS/GRAPH

SAS/GRAPH 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、饼分图、星形图、散点图、曲线图、三维曲面图、等高线图及地理图等。SAS/GRAPH 提供一个全屏幕编辑器,提供多种设备程序,支持非常广泛的图形输出设备



以及标准的图形交换文件。

3. 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来使用 SAS 的其他产品。它自动生成的 SAS 程序既可以辅助有经验的用户快速编写 SAS 程序,又可以帮助用户学习 SAS。

4. 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 SCL 语言的处理能力来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。

5. 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。

6. SAS/ACCESS

为了对众多不同格式的数据进行查询、访问和分析,SAS/ACCESS 提供了针对目前许多流行数据库软件的接口,利用 SAS/ACCESS 可建立访问外部其他数据库的一个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 数据集后再使用,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 将数据真正提取进入 SAS 数据库。SAS/ACCESS 提供的接口是双向的,既可以将数据读入 SAS,也可以在 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、IN-GRES、Informix、DBF/DIF、ODBC 等。

7. SAS/STAT

SAS/STAT 覆盖了所有的实用数理统计分析方法,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或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。可处理的数据有实型数据、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。

8. 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以及试验设计。SAS/QC 提供了多种不同类型控制图的



制作与分析,如 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。

9. 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。它提供了方便的模型设定手段和多样的参数估计方法。

10. SAS/OR

SAS/OR 提供全面的运筹学方法,是一种强有力的决策支持工具。它辅助人们实现对人力、时间以及其他各种资源的最佳配置。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优化办法,如网络流问题、运输问题、分配问题等。

11. SAS/IML

SAS/IML 提供功能强大的面向矩阵运算的编程语言,帮助用户研究新算法或解决 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、函数和例程序,用户使用很少的语句便可以执行复杂的计算过程。

12. 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,汇总数据的更新,Metadata 的建立、管理和查询,Data marts 和 Info marts 的实现。

13. 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),可将数据仓库中的数据或其他来源的数据以立体阵列的方式存储,以便于用多维数据浏览器等工具快速和方便地访问。

14. 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,能支持 Web 浏览器动态查询 SAS 数据或外部的关系型数据库;SAS ODBC Driver 能通过支持 ODBC 的 Windows Web 服务器来访问 SAS 数据;SAS Driver for JDBC 能通过 Java Applet 来查询 SAS 数据;SAS/IntrNet Application Dispatcher 能通过 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。

15. 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,也可以与用户定义的主题(如人口、产值等)相关联。用户可以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。



16. SAS/ITSV

IT Service Vision(ITSV)是企业全面IT服务的性能评估和管理软件,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对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或批处理的方式产生组织任意层面的报告。系统程序员及网络工程师能借此识别、研究并解决有关问题;业务分析人员能借此制定资源管理的总体策略;CIO和数据中心经理能借此定期地得到所需的IT运作的汇总和分析报告。

17. SAS/CFO Vision

SAS/CFO Vision用于财务整合和报告,内部包含了会计知识,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中后,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,从而可以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100多种函数和丰富的算术逻辑运算符,用户可用赋值语句、条件语句、数组和循环语句等对变量进行各种操作。

SAS语言有两类,即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。DATA语句的部分叫数据步。PROC语句用于指出对哪种数据进行分析。PROC语句的部分叫过程步,利用系统提供的过程(SAS提供的数据分析子程序)进行分析。

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,因此可以从不同的角度对数据进行分析比较,逐层次揭示出数据蕴涵的信息。

SAS具有第四代语言的特点,只需通过菜单的选择、对话框的操作告诉系统要做什么,而无须告之怎样做。只要初步了解统计分析原理,无须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1~3维表格,它既可以各种格式输入数据,也可以任何方式输出



数据。除了固定的报表格式外,SAS 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志,并支持彩色显示器和绘图仪。

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中了解程序的运行情况、出错信息和输出结果。对于数据和文件,也各有窗口管理和显示。通过窗口,还可以了解 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输入命令来显示窗口,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,其中仅是参数不同,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,从而大大简化了程序的编写。

7. 支持分布式处理

SAS 支持分布式处理,多台计算机协同工作,有效减少了大数据量的统计分析所需的时间,并能充分享用网络环境中的软件及硬件资源。

8. 采用输出分发系统

SAS 8.0 以后版本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,它允许用户自己定制显示哪些结果,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 SAS 各类文件的应用软件。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、输出结果等内容的通道,用户不必运行 SAS,甚至不需要在计算机中安装 SAS,即可浏览 SAS 的各类文件。

1.2 SAS 的安装

下面以 SAS 8.1 为例,介绍 SAS 在 Windows 98 第 2 版中的安装过程。将 SAS 8.1 的光盘放入光驱中,会自动运行 SAS 的安装程序向导,如图 1-1 所示。如果不能自动运行,用户可双击光盘根目录下的 setup.exe 运行安装向导。

在图 1-1 所示的对话框中,选择“SAS System Setup”即可开始安装 SAS。