

APPLIED STATISTICS AND THE SAS[®]
PROGRAMMING LANGUAGE

SAS[®] 应用统计分析

第 5 版

【美】 罗纳德·科迪 杰弗里·史密斯 著
辛 涛 译



Ronald P. Cody Jeffrey K. Smith



人民邮电出版社
POSTS & TELECOM PRESS

SAS[®] 应用统计分析

(第5版)

[美] 罗纳德·科迪 杰弗里·史密斯 著

辛涛 译

人民邮电出版社

北 京

图书在版编目 (CIP) 数据

SAS 应用统计分析: 第 5 版 / (美) 科迪 (Cody P.R.), (美) 史密斯 (Smith J.K.) 著; 辛涛译.

—北京: 人民邮电出版社, 2011.7

ISBN 978-7-115-25278-4

I. S… II. ①科… ②史… ③辛… III. ①统计分析—应用软件, SAS—教材 IV. ①C812

中国版本图书馆 CIP 数据核字 (2011) 第 073300 号

Ronald P. Cody, Jeffrey K. Smith

Applied Statistics and the SAS® Programming Language, 5e

ISBN 0-13-146532-5

Posts & Telecom Press is authorized by Pearson Education to publish and distribute exclusively this reprint edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong SAR, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Copyright © 2011 by Pearson Education, Inc. and Posts & Telecom Press.

本书中文简体字版由 Pearson Education 公司授权人民邮电出版社独家出版发行。此版本仅限在中华人民共和国境内 (不包括中国香港、澳门特别行政区及中国台湾地区) 销售。未经授权的本书出口将被视为违反版权法的行为。

未经出版者预先书面许可, 不得以任何方式复制或发行本书的任何部分。

本书封底贴有人民邮电出版社和 Pearson Education 公司防伪标签, 无标签者不得销售。

北京市版权局著作权合同登记号: 01-2009-5761

版权所有, 侵权必究。

SAS® 应用统计分析 (第 5 版)

- ◆ 编 [美] 罗纳德·科迪 杰弗里·史密斯
- 译 辛 涛
- 策 划 刘 力 陆 瑜
- 责任编辑 肖 莹

- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号 A 座

邮编 100061 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

电话 (编辑部) 010-84937150 (市场部) 010-84937152

(教师服务中心) 010-84931276

三河市李旗庄少明装订厂印刷 新华书店经销

- ◆ 开本: 787 × 1092 1/16

印张: 36

字数: 590 千字 2011 年 7 月第 1 版 2011 年 7 月第 1 次印刷

ISBN 978-7-115-25278-4/F



定价: 88.00 元

本书如有印装质量问题, 请与本社联系 电话: (010) 84937153

内 容 提 要

随着计算机应用的发展,在统计学领域,先后出现了数百种各具特色的统计软件,其中 SAS (Statistics Analysis System) 是目前国际上最有影响的一种软件,功能非常强大,广泛应用于经济管理、商业金融、医学、教育、心理、生物、地理等各个领域,具有重要的应用价值。据报道,全世界 500 强企业有百分之八十在使用 SAS,它已被誉为统计分析的标准软件。

本书以帮助读者实现 SAS 入门和学会应用 SAS 软件解决实际问题为目的,逻辑结构清晰,内容系统完整,表述简单明了,用丰富的实例说明了运用 SAS 软件进行常见的统计分析的程序和步骤,以及对结果的解释与报告。全书共分为 20 章,内容如下:第 1 章介绍 SAS 基础,第 2~10 章是用 SAS 来实现常见的统计分析过程;第 11 章描述了 SAS 运用于心理测量学分析的例子,第 12 章介绍 INPUT 语句,第 13~16 章讲解如何用 SAS 来读取和处理数据,第 17~18 章介绍 SAS 函数,第 19~20 章详细描述了 SAS 程序的范例。

本书主要针对非计算机专业和非数理统计专业从事数据处理与分析的人员,力求说明统计方法在求解实际问题中的应用,为他们提供应用 SAS 软件进行数据统计分析的指导。它适用于应用领域的研究人员、SAS 程序的初学者,以及想进一步学习 SAS 程序的读者。

译者序

随着计算机的应用发展,在统计学领域,先后出现了数百种各具特色的统计软件,其中 SAS (Statistics Analysis System, 统计分析软件) 是目前国际上最有影响的一款软件,广泛应用于经济管理、商业、金融、医学、教育、心理、生物、地理等各个领域。

1966 年,美国北卡罗来纳大学开发了 SAS 统计分析软件,并于 1976 年成立了 SAS 软件研究所和 SAS 软件公司,开始进行 SAS 系统的维护、开发、销售和培训工作。到目前为止, SAS 公司已经出版了此软件的多个版本,并不断地完善和发展。作为目前国际最权威的数据统计分析软件之一, SAS 功能非常强大,具有重要的应用价值。据报道,全世界 500 强企业中有 80% 在使用 SAS, SAS 已被誉为统计分析的标准软件。

简单来说, SAS 是一个模块化、集成化的大型应用软件系统。它由数十个专用模块构成,其功能包括数据访问、数据储存及管理、应用开发、图形处理、数据分析、报告编制、运筹学方法、计量经济学与预测等。SAS 系统基本上可以分为四大部分: SAS 数据库部分、SAS 分析核心、SAS 开发呈现工具、SAS 对分布处理模式的支持及数据仓库设计。SAS 系统主要完成以数据为中心的四大任务: 数据访问、数据管理、数据呈现和数据分析。当前(2010 年)软件最高版本为 SAS 9.2。虽然 SAS 是用于决策支持的大型集成信息系统,但该软件系统最早的功能仅限于统计分析,至今,统计分析功能仍是它的重要组成部分和核心功能。

《SAS 应用统计分析》(第 5 版)这本书主要针对从事数据处理与分析的非计算机专业和非数理统计专业人员,力求说明统计方法在解决实际问题中的应用,指导 SAS 用户如何用 SAS 软件进行数据统计分析。本书表述简单明了,逻辑结构清晰,内容系统完整,并以帮助读者实现 SAS 入门和学会应用 SAS 软件解决实际问题为目的,用丰富的实例说明了运用 SAS 软件进行常见的统计分析的程序和步骤以及对结

果的解释与报告。除此之外，在每一章的后面，还附有大量的习题，帮助读者进一步熟悉程序，锻炼解决实际问题的能力。作为从事心理和教育统计工作的研究人员，我相信这本书对于从事社会科学研究的专业技术人员，对于高校相关学科的本科生和研究生掌握统计方法是大有裨益的。

本书与其他统计软件类书籍的最大不同在于，它不是简单枯燥地介绍 SAS 的操作方法，而是从社会科学，特别是教育学和心理学的研究方法切入，力图使读者将研究方法、统计手段与 SAS 软件的操作结合起来，层层递进，吸引读者基于研究方法而有效地掌握 SAS 的操作。我曾将这本书作为心理学研究生高级统计课程的教材，在学习的过程中，大家对这本教材称赞有加。正是因为这些原因，在新曲线出版咨询有限公司的支持下，我们将其译为中文。

英文翻译是一个枯燥而又费神的过程，支持我们翻译完这本书的主要动力来源于我和研究生同学之间的良好互动，是他们的热心、期望让我深切体会到教学相长的乐趣。在这里我要特别感谢北京师范大学心理学院 07 级研究生黄慧静同学、08 级研究生高慧健和李珍同学，以及数学科学学院 07 级研究生王奋际同学，感谢她们为翻译本书所做的出色工作。此外，本书的翻译出版有赖于新曲线公司的鼎力支持，我还要特别感谢责任编辑肖莹小姐为本书中译版的出版所付出的辛苦劳动。尽管已经尽了最大努力，但是个人之力绵薄，其中定有不当之处，望各位专家、同行及广大读者批评指正。

辛涛

2010 年 11 月于北京师范大学

前言（第 5 版）

关于第 5 版的出版,我们需要澄清几个事实:第一,SAS 程序本身一直在不断改进;第二,我们的编程技术也有所提高;第三,我们觉得是时候用 SAS Graph™ 的输出结果来替换所有基于文本的图形了。

在各类会议上,我们遇到了本书早期版本的很多读者,很高兴听到他们谈及此书的妙处并提出建设性的意见。我们用心领会,试着对原有材料加以改进,并增加新的相关主题。第 5 版就是基于对读者们意见的整理而产生的。

大多数研究者自然会对问题的本质更感兴趣,而不那么关注统计方法和计算机程序。但是,做这些分析是研究链中的一个环节,通常也是研究的薄弱环节。尤其在做应用研究的人需要运行统计分析的计算机程序,却找不到可参考的书或手册时,这种困境就特别凸显了。《SAS 应用统计分析》这本书就是要让从事应用研究的人能够用 SAS 进行统计分析,同时免遭大量技术文件的折磨。

读者可以看到运行 SAS 程序进行统计分析时所用到的命令语句、对计算机输出内容的解释、对结果的解释,以及如何在写报告或发表文章时制表和呈现结果。我们给出的例子涉及商业、医学、教育、心理学等多个学科。

这里要提醒一下,在使用 SAS 进行统计分析时,你可以用 Enterprise Guide™ 来管理你的数据,因为这种程序是 SAS 的交互式“前端”界面。如果你使用的是 SAS 学习版(SAS Learning Edition™,第 9 版或以上),或者你有学习版的独立许可证,建议你尝试一下。我们中的一些人对于使用像 Enterprise Guide 这样的交互式界面存有偏见,更青睐于做自己的编程。但是,Enterprise Guide 是相当强大的,你甚至可以看到它所生成的 SAS 编码。尤其在直接进行统计分析时,Enterprise Guide 非常有用,还可以更简便地启动 SAS Graph™。

我们愿意将 SAS 描述为一个统计软件包、一个数据库管理系统和一种高级编程语言的巧妙结合。正如 SPSS、BMDP、Systat 以及其他统计软件包一样,SAS 可以

用来描述一组数据并进行各种统计分析。但是，SAS 绝不仅仅是一个统计软件包。很多公司和教育机构把 SAS 当做一个高级的数据管理系统或编程语言，用它来管理和转换数据，并生成各式报告。同样，SAS 系统还可以用于交互式数据的录入或作为一种提供数据录入和检索的在线系统，这就依赖于您的电脑上安装了 SAS 系统的哪个部分（也取决于你运行的是哪种类型的计算机系统）。

SAS 系统是多个产品的集合，由位于美国北卡罗来纳州凯里镇的 SAS 研究中心负责。SAS 研究中心提供的主要产品如下：

Base SAS[®] 作为 SAS 主要的模块，提供了一些数据管理、程序性能以及基本的描述统计功能。

SAS/Stat[®] 是基础包的补充，包含了除基本程序之外的其他程序。

SAS/Graph[®] 是制作精美图形的软件包。注意“线性图”（用正态特征图形来做的图表）是在基础包和 SAS/STAT 软件包里面的。SAS/GRAPH 增加了高清照相制版的图表绘制功能。

SAS/FSP[®] 是全屏产品（Full Screen Product）的简称。这个软件包可用于直接从一个 SAS 数据文件中搜索、修改或删除记录。它也提供数据录入中复杂数据的检查功能。FSP 包含的过程有 FSBROWSE、FSEDSIT、FSPRINT、FSLIST 以及 FSLETTER。

SAS/AF[®] 表示的是 SAS 应用工具（SAS Applications Facility），职业的数据处理员用这个产品来为他们的用户创建“转换键”或菜单系统。它也可以用于创建与 SAS 系统相关的说明模块。

SAS/ETS[®] 是计量经济学与时间序列软件包（Econometric and Time Series package）。这个软件包是分析时间序列和计量经济学数据的特殊程序。

SAS/OR[®] 是运筹学程序系列。

SAS/QC[®] 是质量控制程序系列。

SAS/IML[®] 是互动矩阵语言模块（Interactive Matrix Language module）。在第 5 版发行之时，IML 的装备都包含在 PROC MATRIX 中。这个特殊的软件包便于高级统计中的矩阵操作。

无论是在美国还是其他国家，SAS、SAS/STAT、SAS/GRAPH、SAS/FSP、SAS/AF、SAS/ETS、SAS/OR、SAS/QC 以及 SAS/IML 都是以美国 SAS 公司的商标注册的。

编程的技能很难被教会。也许你会发现我们的例子简单易懂，逻辑清晰，但是

要学着写你自己的程序就是另一回事了。只有通过大量的实践练习，你的编程技术才会有所提高。所以，每当看完一章以后，请尽量多花些时间来练习解决问题。祝您编程愉快！

罗纳德·科迪
杰弗里·史密斯

致 谢

我们要特别感谢 Mike Zdeb 和 Paul Grant，他们对本书进行了认真的审读。这是一项艰苦的任务，我们对此表示诚挚的感谢。

另外，本书从概念设想到产品上架，都有赖于一支专业团队的倾情奉献。这支团队来自于 Prentice Hall，其成员有：组稿编辑 Petra Recter、整体观念协调员 Lynda Castillo、助理编辑 Joanne Wendelken、项目经理 Jacquelyn Riotto Zupic 和 Michael Bell，以及负责生产印制环节的 Heather Meledin。

目 录

第 1 章	SAS 基础	1
1.	简介	1
2.	运行 SAS : 举个例子	2
3.	扩展此程序	6
4.	SAS 过程	9
5.	DATA 步骤	12
6.	SAS 过程的句法	12
7.	注释语句	14
8.	参考资料	17
	练习题	18
第 2 章	描述统计	23
1.	简介	23
2.	描述统计	23
3.	描述统计 (扩展)	28
4.	直方图、分位数图和概率图	35
5.	分组数据的描述统计	38
6.	频数分布	40
7.	条形图	41
8.	数据散点图	54
	练习题	60

第 3 章	分析分类数据	65
1.	简介	65
2.	问卷设计和分析	66
3.	给变量添加标签	71
4.	为变量值添加标签 (输出格式)	73
5.	重编码数据	77
6.	使用输出格式重编码变量	80
7.	双向列联表	83
8.	生成多元表格的捷径	86
9.	频数表中的卡方检验	87
10.	多个表格卡方检验的程序	88
11.	用于频数表中卡方检验的宏	89
12.	配对样本的 McNemar 检验	90
13.	Kappa 统计量 (一致性系数)	93
14.	优势比	95
15.	相对风险	98
16.	卡方趋势检验	101
17.	分层表和元分析中的 Mantel-Haenszel 卡方检验	103
18.	“允许多选” 的问题	105
	练习题	109
 第 4 章	 处理日期数据和追踪数据	 118
1.	简介	118
2.	处理日期变量	118
3.	处理两位数的年份数值 (Y2K 问题)	124
4.	追踪数据	125
5.	选择每个患者的第一次或最近一次来访	129
6.	计算追踪数据中观测值之差	131
7.	计算每个被试第一次和最近一次来访的差异	134
8.	计算追踪数据的频数	136
9.	用 PROC MEANS 或 PROC SUMMARY 创建汇总数据集	137
10.	输出其他统计量	147

练习题	149
第 5 章 相关和简单回归分析	153
1. 简介	153
2. 相关	154
3. 相关系数的显著性	156
4. 如何解释相关系数	158
5. 偏相关	159
6. 线性回归	160
7. 分解平方和	162
8. 生成散点图和回归线	163
9. 在回归方程中加入二次项	167
10. 转换数据	168
练习题	172
第 6 章 T 检验和非参数比较	177
1. 简介	177
2. T 检验：两个平均数间的差异检验	177
3. 被试的随机分配	180
4. 两组独立样本：自由分布 T 检验	183
5. 单侧检验和双侧检验	187
6. 配对 T 检验（相关样本）	187
练习题	190
第 7 章 方差分析	193
1. 简介	193
2. 单因素方差分析	193
3. 对比检验	202
4. 两个自变量的方差分析	203
5. 解释显著的交互作用	208
6. 多因素设计	216
7. 非平衡设计：PROC GLM	216

8. 协方差分析	220
练习题	223
第 8 章 重复测量设计	229
1. 简介	229
2. 单因素实验设计	230
3. 使用 PROC ANOVA 过程的 REPEATED 语句	236
4. 使用 PROC MIXED 过程计算混合（随机效应）模型	238
5. 重复测量一个因素的双因素实验设计	240
6. 完全重复测量的双因素实验设计	250
7. 重复测量最后一个因素的三因素实验设计	254
8. 重复测量两个因素的三因素实验设计	261
练习题	269
第 9 章 多元回归	274
1. 简介	274
2. 实验设计回归	275
3. 非实验设计回归	280
4. 逐步回归和其他变量选择的方法	282
5. 创建虚拟变量	290
6. 通过方差膨胀因子查看多重共线性	291
7. Logistic 回归	292
8. 采用 PROC LOGISTIC 自动创建虚拟变量	304
练习题	306
第 10 章 因子分析	311
1. 简介	311
2. 因子分析的类型	311
3. 主成分分析	312
4. 斜交旋转	320
5. 共同度	322
6. 如何反转题目分数	324

练习题	325
第 11 章 心理测量学	327
1. 简介	327
2. 测验计分	327
3. 扩展的测验计分	330
4. 用 PROC TABULATE 美化表格	333
5. 一个完整的测验计分和题目分析程序	336
6. 测验信度	339
7. 评分者信度	340
练习题	341
第 12 章 INPUT 语句	344
1. 简介	344
2. 列表输入法：数据值用空格隔开	344
3. 读取逗号隔开的的数据	345
4. 在列表输入中使用输入格式	346
5. 栏位输入法	347
6. 格式化输入	348
7. 每个被试读取多行数据	349
8. 改变顺序和多次读取栏位	350
9. 输入格式列表	351
10. “保持此行”——@ 和 @@	352
11. 隐藏无效数据的错误信息	353
12. 读取“不规则”数据	354
练习题	361
第 13 章 外部文件：读写原始文件和系统文件	368
1. 简介	368
2. 程序自带的的数据	369
3. 从外部文本文件（ASCII 或 EBCDIC）读取数据	370
4. INFILE 选项	372

5. 从多个文件读取数据（使用通配符）	376
6. 把 ASCII 或原始数据写入外部文件	377
7. 写入 CSV（逗号分隔变量）文件	378
8. 创建永久 SAS 数据集	379
9. 读取永久 SAS 数据集	380
10. 如何查看 SAS 数据集的内容	381
11. 带有输出格式的永久 SAS 数据集	383
12. 处理大型数据集	384
练习题	391
 第 14 章 数据集处理：建立数据子集，连接与合并，更新	 395
1. 简介	395
2. 建立数据子集	395
3. 合并相似数据	397
4. 合并不同数据	398
5. “查找表”	401
6. 更新数据集	404
练习题	405
 第 15 章 数 组	 409
1. 简介	409
2. 更改变量的值	409
3. 将 999 转换为数字缺失值	411
4. 将 N/A 转换为字符缺失值	412
5. 将身高和体重从英制单位转换为公制单位	413
6. 临时数组	415
7. 使用临时数组对测验计分	416
8. 指定数组范围	418
9. 临时数组和数组范围	419
10. 内隐脚标的数组	419
练习题	421

第 16 章	用数组重组数据集	425
1.	简介	425
2.	新建数据集：从单一观测值到多个观测值	425
3.	示例二：从单一观测值到多个观测值	428
4.	多维数组：从单一观测值到多个观测值	429
5.	新建数据集：从多个观测值到单一观测值	431
6.	多维数组：从多个观测值到单一观测值	433
	练习题	434
第 17 章	SAS 函数综述	437
	第一部分：非字符函数	437
1.	简介	437
2.	算术和数学函数	437
3.	随机数函数	440
3.	时间和日期函数	441
5.	INPUT 和 PUT 函数：数字和字符变量的转换	443
6.	LAG 和 DIF 函数	446
	练习题	447
第 18 章	SAS 函数综述	451
	第二部分：字符函数	451
1.	简介	452
2.	DATA 步骤中字符变量的长度是如何设定的	452
3.	处理空格	454
4.	删除字符串中的字符	455
5.	字符数据核查	456
6.	提取子字符串的例子	457
7.	在等号左侧使用 SUBSTR 函数	458
8.	以上例子的另一种解决方法	459
9.	解压缩字符串	460
10.	分解字符串	461
11.	在一个字符串的内部定位另一个字符串	462