

刘顺忠 编著

数学★软件★应用系列教材

管理统计学和SAS软件应用



WUHAN UNIVERSITY PRESS

武汉大学出版社

数学★软件★应用系列教材

管理统计学和SAS软件应用

刘顺忠 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

管理统计学和 SAS 软件应用/刘顺忠著. —武汉: 武汉大学出版社,
2006. 1

数学·软件·应用系列教材

ISBN 7-307-04835-3

I. 管… II. 刘… III. 统计分析—应用软件, SAS IV. C812

中国版本图书馆 CIP 数据核字(2005)第 135820 号

责任编辑: 杨 华

责任校对: 黄添生

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.com.cn)

印刷: 武汉大学出版社印刷总厂

开本: 787×980 1/16 印张: 16.5 字数: 302 千字

版次: 2006 年 1 月第 1 版 2006 年 1 月第 1 次印刷

ISBN 7-307-04835-3/C·155 定价: 28.00 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。

内 容 简 介

本书基本包括了经济管理学的研究中常用的统计分析方法,并将统计学原理和管理实践紧密结合。从实用的角度出发,本书省去了大量的理论证明,着重从经济管理学应用的角度出发对各种统计方法进行详尽的阐述。通过简单易懂的管理学数据分析例题,使读者更容易接受所学知识,更有利于读者使用 SAS 软件解决管理过程中遇到的实际问题。

本书论述详实,深入浅出,通俗易懂,同管理实践和科研结合紧密。图书中光盘附有本书所有例题数据和程序,便于读者进行 SAS 软件的实际操作。本书不但可以作为学习 SAS 软件的工具书,也可以作为经济管理专业高年级本科生和研究生学习管理统计学的教材或参考书,也是从事管理实践人员和科研工作者运用管理统计学解决实际问题的有力助手。

前 言

随着知识和信息爆炸的知识经济时代的到来,在管理过程中信息的需求量和处理量日益增多。计算机的普及和统计软件的不断完善使大规模数据处理变为可能,这反过来又促进了企业对数据处理的需求不断增加。因此,从事管理工作的人员和从事管理学研究的科研工作者不但需要掌握管理统计学的理论,更要掌握统计分析软件的操作和使用。由于从事管理工作的人员和从事管理学研究的科研工作者往往不是统计学专业出身,因此编写一本适合非统计学专业出身人员学习管理统计学理论、统计软件的使用和指导统计学方法在管理工作和科研中使用的图书就成为一项非常有意义的工作。

为能够使从事管理工作的人员和从事管理学研究的科研工作者在工作和科研实践中准确理解和认识管理统计学,首先必须对管理统计学的内涵和外延进行科学界定。我们认为,管理统计学是一门收集、整理和分析数据,并利用分析结果辅助管理决策的科学,其目的在于通过对随机性数据的分析,探索管理的方法和策略。管理科学是决策的科学,管理统计学是通过对随机性数据进行处理以辅助决策的科学,因此,任何能够对随机性数据进行处理的方法都可以纳入管理统计学领域。基于上述对管理统计学的认识,考虑到从事管理工作的人员和从事管理学研究的科研工作者的专业背景特征、管理工作的实际需求和我国目前本科专业统计学的课程设置,本书主要内容体系包括管理统计学的预备知识(概率论和线性代数基础)、随机变量及其数字特征、抽样方法和统计量、常用统计量分布和分位数、参数估计和假设检验、相关分析、方差分析、回归分析、定性数据分析、聚类分析、判别分析、主成分分析、因子分析、典型相关分析、通径分析等内容。本书可以作为管理专业高年级本科生和研究生学习管理统计学的教材,也可以作为从事管理学教学和科研的教师以及从事管理工作的人员学习管理统计学的参考书。

读者在阅读本书时应当学习完线性代数和概率论的基本知识,并且具备简单的计算机软件操作能力。我国一般正规大学的经管专业本科生在大学一二年级已经修完线性代数和概率论两门课程,因此高年级本科生、研究生、教师、研究人员和从事管理工作的人员阅读本书不存在较大障碍。为了方便读者学习,本

书将全部例题数据和程序都存放在本书所附的光盘之中,读者只要按照书中的操作步骤即可完成相应例题的 SAS 计算,使读者有了同实际相结合的数据分析机会。

在本书编写过程中得到许多同事和专家的帮助,他们为本书的编写提出了有益的建议,增强了本书的实用性和可读性。在这里需要特别感谢刘继红博士,他为本书“参数估计和假设检验”一章的编写提出了许多建设性的建议,并对我的工作给予了极大支持。

作者在撰写本书时,试图为读者提供一本通俗易懂和实用性较强的图书,但是实际效果究竟如何,有待广大专家、学者、教师、学生和各界读者对本书提出批评和建议。限于作者的水平,书中不可避免地存在疏漏之处,希望读者和同行提出宝贵意见。

刘顺忠

2005 年 12 月于长春净月潭

目 录

前 言	1
第一章 SAS 统计分析软件概述	1
第一节 SAS 8.0 界面和菜单简介	1
第二节 SAS 编程的基础知识	2
第三节 SAS 中 Analyst 视窗的使用	4
第四节 SAS 统计分析程序模块	7
第二章 概率论和线性代数基础知识	12
第一节 概率论的基础知识	12
第二节 线性代数的基础知识	17
第三章 随机变量及其数字特征	21
第一节 随机变量分布	21
第二节 随机变量的数字特征	25
第四章 抽样方法和统计量	27
第一节 抽样方法和测量水平	27
第二节 简单随机抽样误差及其估计	31
第三节 样本容量确定	34
第四节 统计量及其计算	36
第五章 常用随机变量分布及其性质	41
第一节 常用统计量分布	41
第二节 常用分布的分位数	46
第六章 参数估计和假设检验	49
第一节 点估计和区间估计的基本原理	49

第二节	假设检验的基本原理	52
第三节	正态分布的假设检验	56
第七章	相关分析	67
第一节	简单相关分析	68
第二节	秩相关分析	75
第三节	偏相关分析	81
第八章	方差分析	87
第一节	单因素方差分析	87
第二节	无交互作用双因素方差分析	92
第三节	有交互作用双因素方差分析	99
第九章	一元回归分析	108
第一节	回归分析概述	108
第二节	一元线性回归分析的基本理论	109
第三节	一元线性回归模型计算实例	116
第四节	一元线性回归分析 SAS 计算步骤和程序	118
第十章	多元线性回归	121
第一节	多元线性回归的基本理论	121
第二节	多元回归模型计算实例	127
第三节	多元回归分析 SAS 计算步骤和程序	130
第十一章	回归模型违反假设及其处理	133
第一节	自相关问题及其解决	133
第二节	异方差问题及其解决	142
第三节	多重共线性问题及其解决	147
第四节	其他回归模型	156
第十二章	定性数据统计分析	161
第一节	列联表分析	161
第二节	Logistic 回归分析	167

第十三章 聚类分析	172
第一节 距离的测量	172
第二节 系统聚类法	173
第三节 快速聚类法	177
第十四章 判别分析	182
第一节 距离判别分析	182
第二节 Fisher 判别分析	193
第十五章 主成分分析	202
第一节 主成分分析的基本原理	202
第二节 主成分分析求解实例	206
第三节 主成分 SAS 计算步骤和程序	207
第十六章 因子分析	211
第一节 因子分析的基本思想	211
第二节 因子分析的原理	212
第三节 因子分析的计算实例	215
第四节 因子分析 SAS 计算程序	219
第十七章 典型相关分析	224
第一节 典型相关的基本原理	224
第二节 典型相关分析的求解实例	228
第三节 主成分 SAS 计算步骤和程序	231
第十八章 通径分析	235
第一节 通径分析的基本原理	235
第二节 通径分析的求解实例	239
第三节 通径分析的软件计算	242
附 录	244
参考文献	254

第一章 SAS 统计分析软件概述

SAS系统全称为 Statistics Analysis System,最早由北卡罗来纳大学的两位生物统计学研究生编制,并于1976年成立了SAS软件研究所,正式推出了SAS软件。自从SAS软件推出以来,应用领域已经遍及金融、医药卫生、生产、运输、通信、政府和教育科研领域,成为国际最富知名度的统计软件之一。SAS的数据处理和统计分析系统被誉为国际上的标准软件系统,在国际学术交流中不必说明算法,直接使用SAS统计分析结果,可见SAS的权威性。SAS不但可以使用对话框方式处理数据,而且可以使用编程方法处理数据,因此成为广大科研工作者和数据处理分析人员喜爱的统计分析软件。本书将以SAS 8.0版本来说明使用SAS软件编程模块和菜单进行统计数据处理和分析的方法。

第一节 SAS 8.0 界面和菜单简介

进入SAS 8.0后,出现如图1-1所示的SAS主窗口。

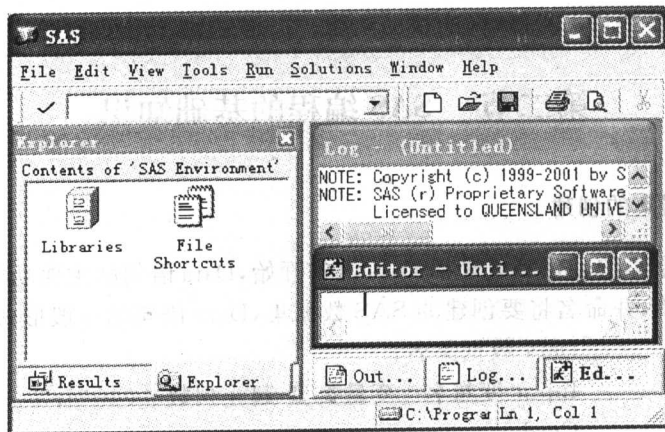


图 1-1 SAS 主窗口

在 SAS 主窗口中可以看到 SAS 的主要窗口和菜单,这些窗口和菜单的基本功能如下。

一、SAS 主要窗口

Editor 窗口,也称为增强编辑窗口,主要用来编辑 SAS 程序。使用该窗口编辑 SAS 程序时,其基本操作同 Windows 视窗操作系统提供的记事本程序编辑功能类似,而且该窗口能够使用不同颜色对程序不同部分进行区分,并对 SAS 语言进行语法检查。用户可以打开多个 Editor 窗口来编辑多个 SAS 程序。编辑完 SAS 程序后,按照 Run 菜单 → Submit 选中菜单,即可运行 SAS 程序。

Log 窗口主要用来观察程序运行状况,找出程序的错误并指明错误的位置,以便用户进行修改。

Output 窗口主要用来输出程序计算结果,当程序错误时不输出结果。

二、SAS 主要菜单

File 菜单完成文件打开、建立、保存、导入和打印功能。

Edit 菜单完成剪切、复制、查找和替换等编辑功能。

View 菜单可以用来观察和新建窗口。

Tools 菜单提供 SAS 的常用工具。

Run 菜单完成程序提交运行。

Solutions 菜单提供一个图形界面完成 SAS 功能的操作。

Window 菜单完成视窗的控制。

Help 菜单提供 SAS 软件的帮助功能。

第二节 SAS 编程的基础知识

一、数据集的创建

SAS 数据集的创建和处理以 Data 语句开始,Data 语句的主要功能是标志数据定义的开始,并命名将要创建的 SAS 数据集,Data 语句的一般形式如下:

```
Data 数据集名;  
Input 变量名 [变量类型 起止列数][@@];  
Cards;  
数据块  
;
```

Data 语句表示数据集的开始,其主要功能是使用“数据集名”为数据集命名。

Input 语句用于向系统表明如何读入每一条记录,它的主要功能是:读入由语句指定的数据列,为相应的数据域定义变量名和确定变量的读入模式。其中方括号表示括号内的内容为可选,如果不输入,系统会以默认值代替。变量名根“@@”表示可以在一行里连续读入数据,直至本行结束或到达分号为止。

Cards 语句用于直接输入数据,标志着数据块的开始。

数据块是中间以空格划分的数据,当遇到“;”时,数据块结束。

我们通过下面两个例子来说明数据集的创建。

Data example1;	将要建立的数据集命名为 example1
Input var1 var2;	要输入的变量为 var1, var2
Cards;	直接输入数据,数据块开始
18 32	} 数据块
47 22	
46 34	
57 43	
;	数据集结束符号

上例中数据块写成两列,使得程序过长,为书写简洁,可以采用下面数据集定义方式:

Data example1;	将要建立的数据集命名为 example1
Input var1 var2@@;	要输入的变量为 var1, var2, 并定义连续输入
Cards;	直接输入数据,数据块开始
18 32 47 22 46 34 57 43	数据块
;	数据集结束符号

Input 语句中的“@@”表示数据可以在一行里连续读入。SAS 默认按列来分隔变量,若有两个变量,则数据需按两列输入。而当 SAS 遇到变量名后有“@@”时,在按变量名依次读取一组数据后,不是跳到下一行,而是继续在该行读取数据,直至本行结束或到达分号为止。

二、SAS 程序调用的基本模式

SAS 系统提供有各种用于数据整理和统计的 SAS 程序模块,在实际工作中,可以根据数据处理要求,调用相应的模块。SAS 程序总是由一个 Proc 语句开始,后面紧跟着程序名,程序名表示将要调用的统计分析模块。SAS 的许多功能就是通过程序模块来实现的。SAS 统计分析程序模块的基本调用方式如下:

```
Proc 过程名 [Data = 数据集名] [选项];  
    该过程的专用语句描述;  
    [Var 变量序列;]  
    [Where 条件表达式...;]  
    [By 变量序列;]  
Run;
```

其中,方括号里的语句均可以省略,在省略的情况下该过程按最平常的情况来处理,即处理最新建立的 SAS 数据集并且处理所有的变量。

第四节我们将详细介绍本书使用的 SAS 统计分析程序模块的调用方式以及各种选项的意义。

第三节 SAS 中 Analyst 视窗的使用

除了可以使用编程方法调用 SAS 统计分析功能外,SAS 还提供了窗口和菜单操作方式可方便非统计专业人员对数据进行统计分析。窗口和菜单方式虽然操作方便,但是不能进行复杂的统计分析,其统计效果远不如编程方法周全。为此,我们对本书所有例题都提供了用编程方法计算的程序。如果能够使用 Analyst 视窗的相应统计功能,我们还对绝大多数例题提供了 Analyst 视窗进行统计分析的操作流程。

一、Analyst 视窗的调用

启动 SAS 主窗口,如图 1-1 所示。

按照 Solution 菜单 → Analysis → Analyst 顺序点击各级菜单,显示出 Analyst 视窗,如图 1-2 所示。

除了可以在 Analyst 视窗中直接输入数据外,也可以将外部数据导入 Analyst 视窗。本书中的例题数据全部以 Excel 文件格式存放,因此下面介绍将 Excel 数据导入 Analyst 视窗的方法。

二、Analyst 视窗的数据导入

下面以本书附带光盘中的“第一章导入演示数据”文件说明 Excel 数据导入方法。该文件的内容如图 1-3 所示。

在 SAS 中进入如图 1-2 所示的 Analyst 视窗,按照 File 菜单 → Open 顺序选

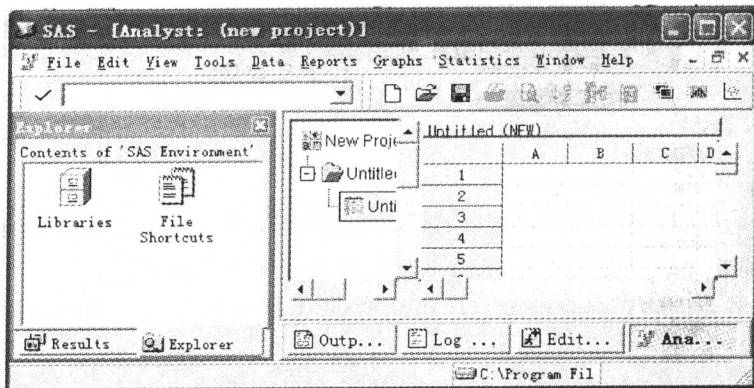


图 1-2 Analyst 视窗

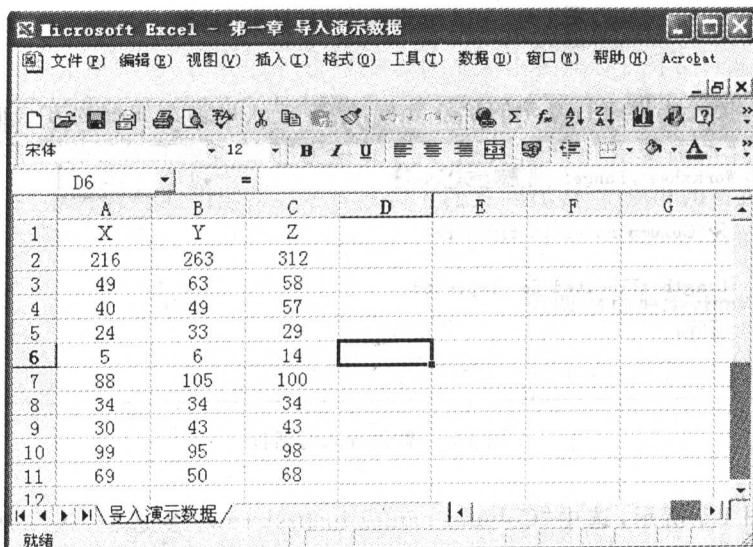


图 1-3 “第一章 导入演示数据”文件格式

中打开文件对话框,文件打开对话框如图 1-4 所示。

在图 1-4 所示打开文件对话框中,选择文件类型为 Microsoft Excel Spreadsheet,即表示打开 Excel 文件。选中光盘中的“第一章 导入演示数据”Excel 文件,单击打开按钮,进入如图 1-5 所示的数据导入对话框。

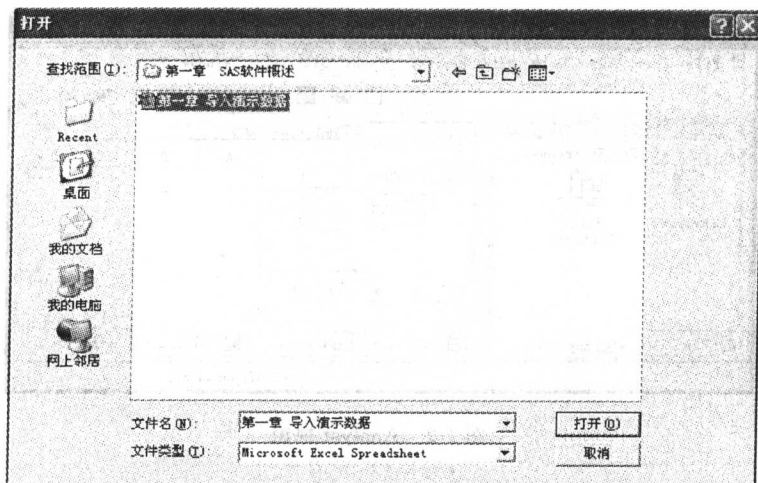


图 1-4 Analyst 视窗的文件打开对话框

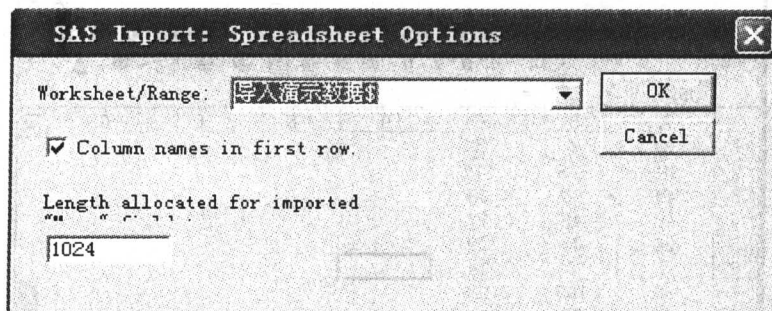


图 1-5 数据导入对话框

如图 1-5 所示,选中“Column names in first row”复选框,表示从打开的 Excel 文件第一行读取变量名。在 Worksheet/Range 下拉框中选中“第一章 导入演示数据”Excel 文件的导入演示数据页面,单击“OK”按钮,数据导入到 SAS 的 Analyst 视窗,如图 1-6 所示。

有关 Analyst 视窗统计功能的介绍,将在具体章节中结合实际例子给出,本章就不详细介绍了。

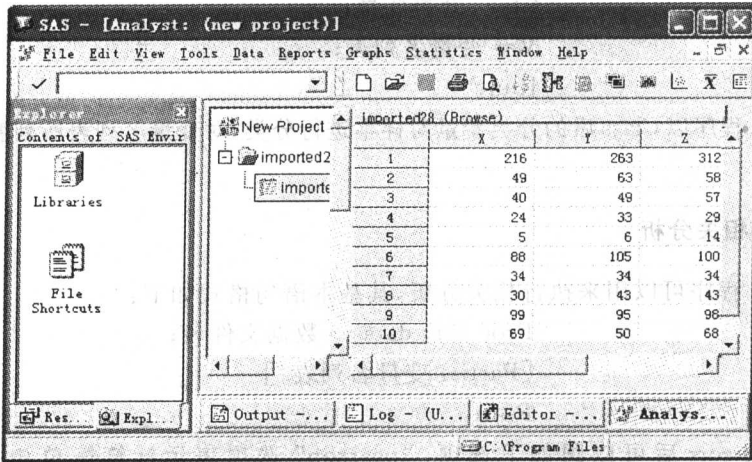


图 1-6 导入数据后的 Analyst 视窗

第四节 SAS 统计分析程序模块

SAS 为进行统计分析编程提供了各种统计分析程序模块,本节将简要介绍与本书有关的统计分析模块的基本功能,具体使用将在本书后继章节结合具体实例进行详细介绍。

一、描述统计量计算(means)

means 程序可以用来执行描述统计量计算,其基本语句格式如下:

```
Proc means data = 数据文件名;
Run;
```

Proc means 后可以跟如下选项:“maxdec”表示结果保留小数的位数,“n”表示输出样本量,“mean”、“std”、“min”、“max”、“range”、“var”和“median”分别表示输出均值、标准差、最小值、最大值、极差、方差和中位数,“t”表示对样本均值是否为零进行 t 检验,“prt”表示输出均值检验的显著水平。

二、独立样本 t 检验

ttest 程序可以用来执行独立样本 t 检验,其基本语句格式如下:


```
Proc ttest data = 数据文件名;  
Class 分类变量;  
Var 变量名列表;  
Run;
```

ttest 程序以 Class 后的分类变量对样本进行分类,对变量名列表所列变量进行 t 检验。

三、相关分析

corr 程序可以用来执行相关分析,其基本语句格式如下:

```
Proc corr data = 数据文件名;  
[Partial 变量名列表;]  
Run;
```

Proc corr 后可以跟如下选项:“pearson”选项表示计算简单相关系数,“spearman”选项表示计算 Spearman 相关系数,“kendall”选项表示计算 Kendall 相关系数。Partial 变量名列表表示对所列变量进行偏相关分析。

四、方差分析

anova 程序可以用来执行方差分析,其基本语句格式如下:

```
Proc anova data = 数据文件名;  
Class 因素变量列表;  
Model 指标 = 因素列表;  
Run;
```

其中“因素列表”可以是“因素变量列表”所列因素,当表示两个因素之间交互作用时,将两个表示因素的变量同时列出,中间加 * 号即可。例如要分析因素 f_1 、 f_2 以及 f_1 和 f_2 交互作用,这时“因素列表”为 $f_1 f_2 f_1 * f_2$ 。

五、回归分析

reg 程序可以用来执行回归分析,其基本语句格式如下:

```
Proc reg data = 数据文件名;  
Model 因变量 = 解释变量列表;  
Run;
```

Proc reg 后可以跟如下选项:“vif”选项表示输出“解释变量列表”中变量的方差膨胀因子,“dw”选项表示输出 DW 统计量,“selection”选项表示选择回归方法。