

SAS

统计分析 实用大全

阮桂海 等编著 蔡福金 审校



清华大学出版社

SAS 统计分析实用大全

阮桂海 等编著

蔡福金 审校

清华大学出版社

北京

内 容 简 介

本书前 12 章包括了各种基本统计分析法。从第 13 章开始,几乎包括了所有的高级统计和专业统计,对于各个领域的 SAS 读者完全适用。特别是第 37 章的结合分析(Conjoint Analyze)等新技术,是当今市场调研的锐利武器,也是本书的独特之处。

本书是国内 SAS 分析应用的精品教程,可作为各大院校经济学、统计学、市场营销学、医学、心理学、人文地理学、社会学、管理学、运筹学学生的计算机数理统计必修课教材,也可作为非计算机专业学生自学的指南;本书也是统计师、科研人员、行政管理人员、文秘以及广大自然科学工作者进行课题研究及科研定量分析的有力工具。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

SAS 统计分析实用大全/阮桂海等编著.—北京:清华大学出版社,2003

ISBN 7-302-06416-4

I.S... II.阮... III.统计分析-应用软件,SAS IV.C812

中国版本图书馆 CIP 数据核字(2003)第 018763 号

出 版 者:清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

责任编辑:许存权

印 刷 者:北京鑫海金澳胶印有限公司

发 行 者:新华书店总店北京发行所

开 本:787×1092 1/16 印张:40.75 字数:933 千字

版 次:2003 年 6 月第 1 版 2003 年 6 月第 1 次印刷

书 号:ISBN 7-302-06416-4/TP·4835

印 数:0001~5000

定 价:53.00 元

前 言

SAS (Statistical Analysis System), 称为统计分析系统软件, 这是 20 世纪 60 年代初美国 SAS 公司推出时的说法。如今在医学、市场学、经济学乃至自然科学领域都得到了广泛的普及和应用。

SAS 集“问卷设计、数据汇总和精辟分析”于一身。SAS 在国内之所以与 SPSS 一样得到普及, 是因为 SAS 统计结果科学清晰, 图形极具立体感。因而 SAS 与 SPSS 和 BMDP 一样, 为国际上最知名的三大统计软件之一。

在国际学术界有条不成文的规定, 凡是用 SPSS 和 SAS 统计分析的结果, 在国际学术交流中可以不必说明算法, 由此可见其权威性和信誉度。国内众多人士体会到: 当今市场经济时代若不学习 SAS 和 SPSS, 则其业务很难有所发展, 在激烈的市场竞争中很难站稳脚跟。

作者从 1987 年开始, 进行 SAS 和 SPSS 统计分析软件的二次开发研究, 一边研发一边教学。在教学的基础上, 于 1992 年在国内率先著有《SPSS 与 SAS 6.02 的应用教程》。本书《SAS 统计分析实用大全》继往开来, 而且突出“新”与“全”两字, 全书共 39 章, 每一章后面有习题, 附录有习题答案。书中内容几乎囊括了社会学、医学、生物学、心理学、人口学、经济学中所有常用的初高级统计方法, 称得上实用统计大全。本书对每一种数据, 既介绍编程统计法, 又传授了对话框统计法。编程统计法有助于读者熟悉 SAS, 对话框统计法使不了解 SAS 功能的读者能轻松上手。本书适合于多层次的读者。

读书是系统学习, 听课是掌握要领, 况且限于篇幅, 有些内容未能列入; 为使读者学以致用, 每年元旦、五一节、国庆节、暑假, 北京大学将分别举办 SAS 和 SPSS 的培训, 联系电话: 010-62756149。

在西方发达国家, SAS 和 SPSS 是大学学生的必修课, 也是其毕业后谋生的资本。在我国, SAS 和 SPSS 也逐渐成为统计师和文、理科大学生的“看家本领”。

参加本书编著工作的人员有:

阮桂海、蔡建平、建瓴、蔡建琼、蔡文英、朱志海、苏红、周爽、阮开林、朱志洪、阮潮海、阮开顺、阮晨辉、杨武栓、张才彬、胡忠华、陈耀辉、张同延、刘云、吴谅谅、蔡敬堂、郑卫平、罗安娜、吴宝科、曲庆云、刘旭东、于惠芳、李欣、祝迎春等。

在编著过程中, 统计学家们给予了宝贵的指导, 北京大学社会系领导给予了大力支持, 特别是本书的责任编辑始终一丝不苟、严把质量关, 在此一并表示由衷的感谢!

由于时间仓促, 加上我们水平有限, 书中不到和不妥之处, 还请同仁拨冗赐教。E-mail: rgh114716@sina.com, 谢谢!

阮桂海

2002 年 10 月于北京大学

目 录

第 1 章 SAS 的安装与启动	1
1.1 SAS 的安装	1
1.2 SAS 系统的启动	2
1.3 SAS 的专用功能键	3
第 2 章 SAS 系统功能简介	5
2.1 SAS 自身的程序结构（即过程）	5
2.2 SAS 的基本命令（即过程行）	5
第 3 章 SAS 的数据定义与数据变换	14
3.1 SAS 的数据定义	14
3.2 定义自由格式的数据	15
3.3 定义固定格式的数据	15
3.4 一行多个观察值（自由格式的扩充）	18
3.5 数据的读取与显示	18
3.6 用 INFILE 命令调用外部数据文件	20
3.7 用 IF 进行条件转换	20
3.8 用 DELETE 作有条件的删除	22
3.9 用 GO TO 作有条件的“转移”	22
3.10 用 LABEL 命令定义变量标签	23
3.11 用 FORMAT 过程定义数值标签	23
3.12 用 RETAIN 命令预置变量值	24
3.13 创造新变量以及用 SUM 等函数求和	24
3.14 数据重新编码	27
3.15 本章知识摘要	29
第 4 章 采用 SAS 对话框进行数据输入管理与统计	32
4.1 Analysis 对话框（第 1 个视窗）	32
4.2 ASSIST 视窗	34
4.3 LAB 视窗	38
4.4 在 LAB 视窗新建 SAS 数据集的利弊	42
4.5 在 Work 工作区中直接输入 SAS 数据集	46

第 5 章 SAS 基本统计概念	48
5.1 总体与样本	48
5.2 参数与统计量	49
5.3 正态分布	49
5.4 经验规则的应用	50
5.5 正态性的假设与检验	51
5.6 显著性水平与 α (Alpha) 值	56
第 6 章 SAS 程序设计技巧及报表的汉化	58
6.1 PUT 语句	58
6.2 用 SET 定义 (复制) 数据集	60
6.3 数据集定义 (DATA 语句) 与 BY 语句	61
6.4 用 _NULL_ 快速运行程序	63
6.5 用 END=FINAL 显示观测数 (个案数)	64
6.6 注释行 PUT 语句和 PRINT 格式	65
6.7 输出结果的汉化	67
第 7 章 数据行排序	69
7.1 数据行排序的语法规则	69
7.2 采用 ASSIST 视窗中的 Sort 菜单排序	71
第 8 章 SAS 平面图形	76
8.1 曲线式散点图的制作	76
8.2 调用 CHART 过程的 VBAR 及 HBAR 命令画条形图	79
8.3 调用 CHART 过程的 BLOCK 命令绘制区域图 (即 BLOCK 图)	86
8.4 实心圆形图的制作	93
第 9 章 频数统计	98
9.1 采用 Analyst 中的对话框统计频数	98
9.2 调用 FREQ 过程中的 TABLES 命令作 sex 和 edc 变量的频数统计	102
9.3 调用 Univariate 过程进行详尽的频数统计	105
第 10 章 交叉汇总表及相关测量	111
10.1 采用 Analyst 对话框进行 双变量的交叉汇总分析	111
10.2 简单易学的编程法	117
10.3 多变量交叉汇总表及结合测量	120
10.4 分析交叉汇总表中的数据特征	123
10.5 两个地区老年人与受教育水平相互关系的比较	124

第 11 章 调用 MEANS 作描述性统计.....	126
11.1 首先采用 Analyst 中的对话框统计.....	126
11.2 采用 PROC MEANS 计算均值分布.....	130
11.3 结果分析.....	135
第 12 章 相关分析.....	136
12.1 相关分析概述.....	136
12.2 皮尔逊积差相关 (比例-比例变量).....	137
12.3 皮尔逊二分“点-距”相关 (二分变量-区间以上变量).....	141
12.4 肯氏 (Kendall) 等级相关 τ_b (“次序-次序”等级变量).....	141
12.5 计算次序-比率数据的肯氏相关系数.....	146
12.6 斯皮尔曼等级相关 (“次序-次序”等级变量).....	150
12.7 Phi (Φ) 系数和 V 系数 (“定类-定类”变量).....	153
12.8 Cronbach 的 Alpha 系数.....	154
12.9 Hoeffding 的相关系数 D.....	154
12.10 调用 PROC CORR 过程作各种相关分析 (编程法).....	155
第 13 章 采用 GLM 过程进行回归分析 和协方差分析.....	164
13.1 最小二乘法的原理.....	164
13.2 GLM 中各语句的格式.....	168
13.3 GLM 程序各语句的使用说明.....	169
13.4 调用 GLM 程序作一元线性回归.....	175
13.5 调用 GLM 程序进行多元线性回归分析.....	180
13.6 调用 GLM 程序作多项式回归.....	182
13.7 虚拟变量的用法.....	185
第 14 章 采用 REG 过程进行多元线性回归分析.....	188
14.1 采用 Analyst 等对话框进行多元线性回归分析.....	189
14.2 REG 过程的语句格式.....	194
14.3 REG 程序的进一步实例.....	199
14.4 MAXR 回归法和 RSQUARE 回归法.....	204
第 15 章 方差分析 (过程 ANOVA).....	211
15.1 概述.....	211
15.2 关于 ANOVA (方差分析) 的主对话框.....	214
15.3 One-Way ANOVA (单因素方差分析) 实例.....	217
15.4 双因素方差分析 (ANOVA).....	230
15.5 多元 (多因变量) 方差分析——MANOVA.....	237
15.6 方差反复测量 (Repeated ANOVA).....	244

第 16 章	过程 FACTOR	253
16.1	具有共性值的元素.....	253
16.2	因子分析模型	254
16.3	因子分析的步骤.....	254
16.4	析取因子 (Factor extraction)	254
16.5	前三个因子	256
16.6	对前三个因子的进一步研究.....	256
16.7	转轴方式	258
16.8	解释因子与因子得分.....	260
16.9	运行过程 FACTOR.....	261
16.10	因子得分 (FACTOR SCORE) 语句格式.....	263
第 17 章	主成分分析 (过程 PRINCOMP)	270
17.1	主成分分析实例 (犯罪率实例)	270
17.2	采用 Analyst 对话框进行主成分分析.....	271
17.3	采用编程法 (语句) 进行主成分分析.....	276
第 18 章	聚类分析	285
18.1	四种聚类方法 (四个聚类过程)	285
18.2	CLUSTER 凝聚算法.....	286
18.3	CLUSTER 过程的语句命令	286
18.4	大样本的观察值聚类法 (过程 FASTCLUS 的应用)	292
18.5	对变量的聚类	297
18.6	树形结构图	304
第 19 章	独立组样本与成对组样本的 T 检验	309
19.1	从独立组样本中概括数据.....	309
19.2	独立组样本 TTEST 的假设与检验.....	312
19.3	两组独立样本的 Wilcoxon 秩和检验	318
19.4	从成对组样本中概括数据.....	320
19.5	成对差值的 T 检验	322
19.6	成对差值的 Wilcoxon 秩和检验	324
19.7	两组比较的总结.....	325
19.8	试对图 19.15 的结果进行男女生 两组成绩的独立样本 T 检验。	327
第 20 章	计划与统计表 (过程 TABULATE)	328
20.1	表格的维	328
20.2	运行过程 TABULATE	331
20.3	统计表命令小结.....	341

第 21 章	SAS 调用 FoxBASE、DBASE 和 SPSS 的数据	343
21.1	用 PROC DBF 过程的语句调用 数据库数据进行统计	343
21.2	在 SAS 对话框中调用数据库数据进行统计	344
21.3	在 SAS 对话框中调用 SPSS 数据进行统计	347
第 22 章	SAS 数据集文件转换为 SPSS 文件	352
22.1	采用对话框将 SAS 数据集转换为 SPSS 的*.SAV 数据	352
22.2	TOSPSS 过程命令概述	358
22.3	SAS 过程的命令格式 (语句)	359
22.4	与 TOSPSS 过程连用的子命令(即选择项)	359
22.5	数据的转换	360
22.6	实际应用	361
第 23 章	SAS 高级绘图系统 GRAPH	364
23.1	前言	364
23.2	GSLIDE 过程	368
23.3	GCHART 过程	371
23.4	GPLOT 过程	377
23.5	G3D 过程	382
23.6	G3GRID 过程	387
23.7	GCONTOUR 过程	390
23.8	用对话框绘制立体图	394
第 24 章	路径分析	400
24.1	用以路径分析的数据和程序	400
24.2	图形输出	403
24.3	路径图的分析方法	406
第 25 章	非线性回归分析之一 (对数、多项式回归)	408
25.1	对数曲线回归	408
25.2	对数曲线回归分析	410
25.3	多项式回归分析 (拟合抛物线)	413
25.4	多项式回归的结果与分析	414
第 26 章	非线性回归之二 (拟合 Logistic 曲线与正负指数的回归)	422
26.1	拟合 Logistic 曲线 (人口增长方面的一个例子)	422
26.2	从 Logistic 曲线模型解出初始值	423
26.3	拟合 Logistic 曲线回归的分析	425
26.4	负指数生长曲线	429

26.5	拟合负指数生长曲线回归的分析.....	431
26.6	拟合指数曲线 $Y=Ae^{BX}$	434
第 27 章	Logistic Regression 回归分析	438
27.1	Logistic Regression 模型.....	438
27.2	做 Logistic Regression 的变量及其数据	439
27.3	采用 Analyst 对话框进行 Logistic 回归	440
27.4	采用 PROC Logistic 编程做逻辑斯谛回归	446
27.5	假设与检验	449
27.6	解释回归系数	450
27.7	概率预测	451
27.8	多分变量(Categorical Variable)的编码.....	452
第 28 章	四方表 Logistic Regression 回归分析	454
28.1	四方表 Logistic Regression 模型.....	454
28.2	做 Logistic Regression 的变量及其数据	454
28.3	采用 Analyst 对话框进行 Logistic 回归	455
28.4	结果分析	460
第 29 章	多个随机实验组的协方差分析 (GLM 的应用)	463
29.1	两组随机数据的协方差分析.....	463
29.2	多因变量协方差分析.....	467
第 30 章	趋势面回归分析 (GLM 的应用)	474
30.1	原始数据与趋势面分析程序.....	474
30.2	输出结果	475
30.3	结果分析	483
第 31 章	多因素二水平排列组合方差分析 (Anova 的应用)	484
31.1	2×2 排列组合式方差分析	484
31.2	三因素二水平排列组合式方差分析.....	487
第 32 章	随机配伍组与对照组的方差分析 (Anova 的应用)	492
32.1	三个随机配伍组设计的方差分析.....	492
32.2	多个随机实验组与一个对照组的均值比较	499
第 33 章	典型相关 (Cancorr 过程的应用)	503
33.1	计算典型相关的 SAS 数据	503
33.2	采用 CANCERR 过程计算典型相关.....	504

33.3	采用 Cancorr 对话框计算典型相关	505
33.4	输出结果与分析	506
第 34 章	判别分析	513
34.1	判别分析的原理	513
34.2	三种主要的判别分析法	513
34.3	一般判别分析法实例	514
34.4	采用逐步判别分析法	528
34.5	采用典型判别分析法	528
第 35 章	时间序列分析	530
35.1	季节性调整	530
35.2	从回归分析中建立回归预测模型	538
第 36 章	时序（调整的）自相关回归分析	541
36.1	时序自相关的数据	541
36.2	数据文件的编辑	542
36.3	采用 ASSIST 对话框进行时序自相关回归分析	542
36.4	结果分析	546
36.5	最小二乘法回归与调整的自相关回归的相对误差比较	550
第 37 章	用 Conjoint Analysis（结合分析）研究品牌	553
37.1	结合分析的基本原理与思路	553
37.2	通过正交设计建立卡片	553
37.3	精品购物的问卷调查例子	555
37.4	如何做好结合分析	555
37.5	结合分析的输出与分析	556
37.6	市场预测与决策	559
第 38 章	SAS 函数	560
38.1	算术函数	560
38.2	截取函数（Truncation function）	563
38.3	数学函数	564
38.4	三角及双曲函数	566
38.5	概率函数	568
38.6	样本统计函数	573
38.7	随机数函数	577
38.8	字符串函数（同一行的字处理）	578
38.9	时间与日期函数	580

38.10 特殊函数（只用于非微机上）	582
第 39 章 SAS 图形的画法比较	584
39.1 绘图所用的 SAS 数据集	584
39.2 绘图对话框之一：Graphs 对话框	585
39.3 SAS 图形的画法比较	587
39.4 用 ASSIST 对话框绘制中国地图（轮廓）	595
39.5 用 Interactive Data Analysis 对话框绘图	598
附录 1 习题及答案	601
附录 2 如何快速入门并学好 SAS 和 SPSS	633
参考书目	635

第 1 章 SAS 的安装与启动

SAS (Statistical Analysis System) 是 20 世纪 60 年代推出的一种科学的统计分析技术, 如今已广泛普及和应用于医学、社会学、市场学、经济学和自然科学各个领域的信息处理、定量研究和科研分析中。

SAS 和 SPSS、BMDP 并称为国际上最富知名度的三大统计软件。在国际学术界有条不成文的规定: 凡是用 SPSS 和 SAS 统计分析的结果, 在国际学术交流中可以不必说明算法, 由此可见其权威性和信誉度。通过我们数十年的二次开发和应用, 深感 8.x 版本的 SAS, 人机对话的界面越来越友好, 既可以通过编程进行数据的统计分析, 也可以在对话框中选择命令进行信息处理和科研分析, 大大方便了不熟悉命令的 SAS 用户。

1.1 SAS 的安装

SAS 安装要注意软件与硬件条件。

1.1.1 SAS 对硬件的要求

SAS 统计分析软件的版本不断更新。本书着重介绍的是 8.1 版本。但是, 无论是哪一种版本的 SAS, 软件的功能和操作运用都是向上兼容, 因此, 学习本书之后, 对于所有的 SAS 软件版本, 必将触类旁通, 应用自如。

SAS 都能安装在微机 (IBM 系列微机、国产 0520 系列微机、COMPAQ 机、AST 机等所有微型计算机) 上; 硬盘容量要求 10GB 以上。

SAS 软件的安装与运行还有时间上的限制, 必须将微机上的时间下调到 SAS 出品时的日期, 方能运行程序。

1.1.2 SAS 的安装

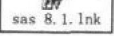
所有版本的 SAS 软件, 在 Windows 系统中的安装步骤如下:

- (1) 从桌面“我的电脑”中直接单击 SAS 光盘中的 SETUP 图标。
- (2) 然后根据系统提示, 单击相应的按钮, 而后单击 SAS 图标启动 SAS 系统。

1.2 SAS 系统的启动

操作步骤:

(1) 启动 SAS 之前将时间调到 SAS 出品日期之前 (8.0 以上版本不必调时间), 见图 1.1。

(2) 然后单击桌面上的 SAS 图标  (如图 1.2 所示)。

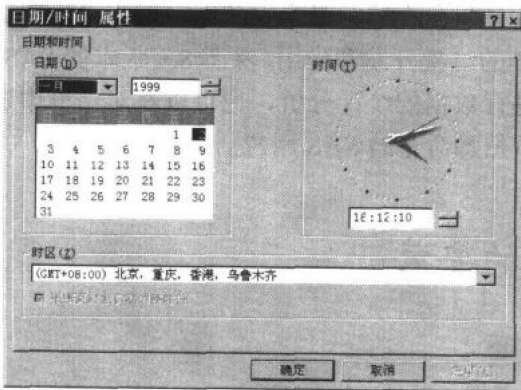


图 1.1 调整后的电脑时钟



图 1.2 桌面上的 SAS 8.1 图标

(3) 自动启动 SAS 系统后, 主画面便出现如图 1.3 所示的 SAS (系统) 标志。

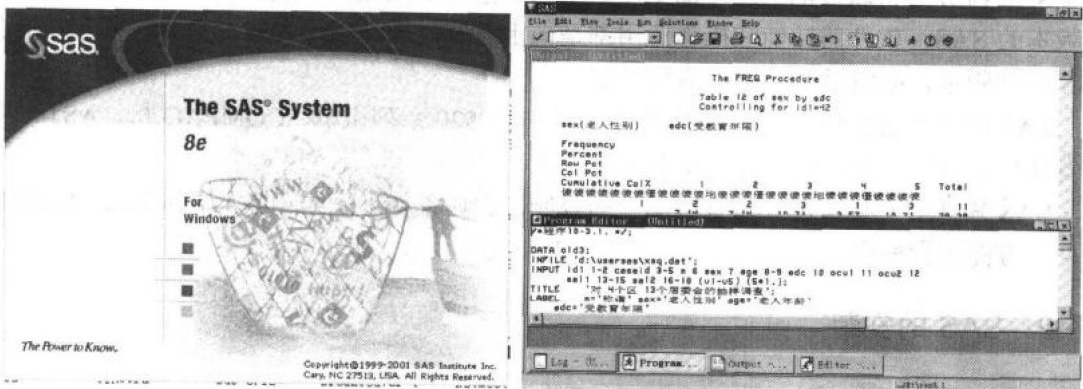


图 1.3 SAS (系统) 标志

图 1.3 (右) 任务栏说明: 3 个窗口图标可用鼠标单击, 自由选择一个。

① Program Editor 窗口: 程序编辑区。可打开“文件”菜单, 从磁盘调入一个 SAS 用户程序, 也可在 Program Editor 窗口编辑新的程序语句或删、改程序。此窗口称为 Program Editor 窗口。

② Log 窗口: 程序运行时的观察窗口, 可看到程序语句正确与否。若语句有误, 则用

↑、↓ 键把光标移到编辑区的 Error 行，进行删、改。

③ Output (输出) 窗口：若程序正确无误，Output (输出) 窗口则显示出运行结果。

1.3 SAS 的专用功能键

SAS 8.x 在微机上均定义了专用的功能键，见图 1.4。

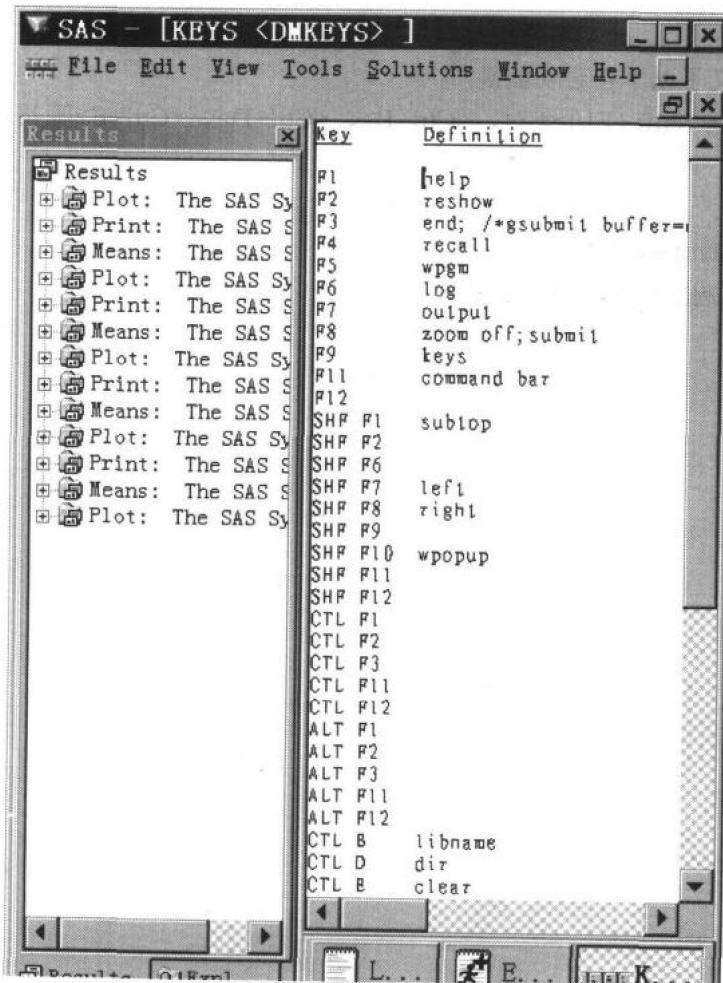


图 1.4 SAS 的功能键

说明：

F1: 功能键。按 F1 键，可出现 SAS 的所有帮助信息。

F2: REshow 键，无用。

F3: 类似 F8 键的功能 (执行源程序)。但程序被执行后程序行即隐去，再按 F4 键可重现“源程序”，有用。

- F4: 改错用, 相当于选择 Run 和 Recall Last Submit 命令, 有用。
- F5: 使屏幕从其他处返回到编辑区 (Program Editor 窗口)。
- F6: 显示 Log 窗口中的信息, 使屏幕从其他处返回到 Log 窗口。
- F7: 显示 Output 窗口中的结果。
- F8: 运行程序, 相当于选择了 Run → Submit 命令。程序执行后各窗口复原 (不放大)。
- F9: 显示出键的功能介绍。
- F10: 空着。
- F11: Command 命令键。

此外, 还定义了以下各对功能键:

<Shift +F7>~<Shift +F10>4 对, <Ctrl+B>、<Ctrl+D>、<Ctrl+E>等共 17 对 (略)。

习 题

- 1.1 SAS 的主画面有哪几个窗口?
- 1.2 运行 SAS 程序可用哪一个功能键?

第 2 章 SAS 系统功能简介

2.1 SAS 自身的程序结构 (即过程)

过程, 英文称为 Procedure。SAS 和 SPSS 等著名统计软件一样, 均有以下的主要过程:

- (1) 频数统计;
- (2) 交叉汇总表;
- (3) 数据的排序分类;
- (4) 方差分析;
- (5) 相关与回归分析;
- (6) 各种高级统计分析过程 (例如, Conjoint 结合分析、FACTOR 因子分析、Cluster 聚类分析等);
- (7) 制图和制表等。

2.2 SAS 的基本命令 (即过程行)

下面介绍 SAS 的基本命令 (也称为过程 PROC) 的用法。

2.2.1 一个简单的 SAS 程序例子

1. 在 Program Editor 窗口编程

启动 SAS 并出现类似第 1 章图 1.3 所示的屏幕窗口时, 可在 Program Editor 窗口输入以下程序行 (每行命令以 “;” 号结束), 见图 2.1。

若存储源程序, 可打开图 2.2 中的 File 菜单, 接着单击图 2.2 所示 Save As, 然后在图 2.3 中输入文件名 S2-1, 再选择 “保存”, 则把源程序 S2-1.sas 存入当前盘中。

2. 在 Program Editor 窗口运行源程序

图 2.1 中的程序行 (共 16 行) 输入完毕, 按 F3 键执行程序。也可选择图 2.4 的 Run → Submit 运行源程序。

3. 在 Output 窗口阅读结果

执行上述程序后, 在 Output 窗口将显示统计结果 (可按 F7 键浏览放大后的输出结果);