

# 实用 SAS

## 统计分析教程

SAS 9.3版本

阮敬 纪宏 编著




中国统计出版社  
China Statistics Press

# 实用 SAS

## 统计分析教程

SAS 9.3版本

阮敬 纪宏 编著

 中国统计出版社  
China Statistics Press

## 内容简介

本书通过真实案例,全面介绍 SAS 9.3 的编程方法并培养读者利用数据来分析问题、解决问题以及对统计分析结果评价的能力。全书内容包括:SAS 编程基础、数据预处理、数据描述、参数推断与非参数分析、相关与回归分析、因子分析、聚类分析与判别分析、列联分析与对应分析、定性数据分析、时间序列分析等,将读者关注的统计分析与初步数据挖掘技术进行剖析。

本书实用性强,强调对统计知识的领悟和应用。对于统计分析的基本原理,编者避免用复杂数学公式来推导,而尽量用文字来描述,并且通过 SAS 来实现统计分析和得到结论的全部过程。为方便读者全面掌握本书内容,本书配有配套案例数据(下载地址:<http://csp.stats.gov.cn>)

本书既可作为高等院校本科生、专业硕士与低年级学术型硕士研究生的统计学及统计软件教材,也可作为社会科学领域数据分析相关知识的自学教材,同时还可以作为从事数据分析与数据管理的研究人员的参考用书,也适合想用 SAS 来解决实际统计分析问题的工程技术和管理人员。

### 图书在版编目(CIP)数据

实用 SAS 统计分析教程 / 阮敬, 纪宏编著. — 北京 :  
中国统计出版社, 2013. 4

ISBN 978-7-5037-6798-2

I. ①实… II. ①阮… ②纪… III. ①统计分析—应用软件—教材 IV. ①C819

中国版本图书馆 CIP 数据核字(2013)第 059225 号

### 实用 SAS 统计分析教程

作 者/阮敬 纪宏

责任编辑/张 赏

封面设计/李雪燕

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://csp.stats.gov.cn/>

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/880×1230mm 1/16

字 数/800 千字

印 张/31.75

版 别/2013 年 4 月第 1 版

版 次/2013 年 4 月第 1 次印刷

定 价/62.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区以任何文字翻印、拷贝、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

## 前 言

数据分析是科学研究中的重要环节,随着社会经济的快速发展,其越来越受到社会和市场环境的重视,是经营管理与决策过程中必不可少的工作和基础。SAS 系统是目前世界上最为权威的统计软件之一,市场占有率极高,在高校、银行、电力、电信、航空、保险、精算、咨询、社会调查、电子商务等行业或部门应用十分广泛,可以实现从数据管理到数据分析和挖掘的完整过程,满足日常分析工作的需求。

随着数据分析作用的日益凸显,如何对现有数据进行整理、加工、处理和分析,以期得到所谓的结论,作为我们进行决策的依据?如何利用现有数据对将来可能出现的结果或结论进行判断或预测?不管是针对企事业单位的管理者或决策者还是从事具体数据分析的工作人员,都需要进行合理数据分析流程的规划,区分数据类型,利用适合的数据分析方法,使用方便、快捷、可靠的统计软件作为工具,对特定数据进行分析与预测,从而洞察市场动向、观测人心所在、把握商机,提升所在单位的竞争力。

而具有深厚数学背景的数据分析方法往往会成为相关人员继续深入学习的门槛,甚至成为枯燥乏味的代名词,无法使企业体验到数据分析成果带来的成效。本书就是要力求降低学习难度,通过编者积累的大量真实案例和数据,主要以文字阐述而不是复杂公式推导的形式深入浅出剖析统计分析方法的基本原理和步骤,重点在于厘清数据分析的基本思路,合理得到恰当的分析结果。在分析过程中,本书使用当前最新版本 SAS 9.3 系统,从编程基础入手,用大量的实例来展示统计分析每一步骤的细节,带领读者走入统计数据分析的奇妙世界。

本书的第 1 部分由前 5 章组成,主要介绍 SAS 系统 9.3 版本的编程基础和数据处理方面的内容,具体包括 SAS 语言的基本结构和语法、SAS 宏编程、SAS 输入输出控制、SAS 数据库及数据集的管理和预处理等;

第 2 部分由第 6 章和第 7 章组成,主要介绍利用 SAS 系统进行描述统计分析的基本过程和方法,涵盖了各种常用统计图形的绘制和解读以及统计量和统计表等具体内容;

第 3 部分由第 8、9 和 10 章组成,主要介绍利用 SAS 系统进行参数统计推断和非参数统计分析;

第 4 部分由第 11、12 和 13 章组成,主要介绍利用数据之间的统计关系建立

单方程模型的内容,具体涵盖了相关分析、典型相关分析、线性回归、非线性回归、多项式回归、分位数回归以及广义线性模型中的离散因变量模型等内容;

第 5 部分由第 14、15、16 和 17 章组成,主要就日常数据分析中所使用的多元统计分析方法进行介绍;

第 6 部分由第 18 章组成,介绍 SAS 系统中使用 ARIMA 建模进行时间序列分析的基本方法和思路。

除此之外,编者还在相应章节中使用 SAS 宏编制了一些 SAS 系统中没有提供的数据预处理和统计分析功能供读者在分析实际问题时调用,如计算单元格数据行百分比、自动按指定变量分拆数据集、进行游程检验等。全书采用 SAS 9.3 的英文版本作为软件环境,希望读者参考本书的内容边做边学习。为了提高学习效率,本书还附送随书案例的全部数据(下载地址:<http://csp.stats.gov.cn>)。

本书由阮敬和纪宏共同在原书《SAS 统计分析从入门到精通》(人民邮电出版社 2009 年版)基础上亲自编著完成,同时也得到了台湾辅仁大学谢邦昌教授、美国中佛罗里达大学王中庆教授、David Nickerson 教授和张宁教授、美国 Merck 公司刘荣博士、清华大学朱世武教授、中国统计出版社严建辉社长和教材编辑部主任陈悟朝博士、史丹索特(北京)信息技术有限公司总经理陈庆焕先生和本书编辑张赏等的大力支持和帮助,在此一并表示衷心感谢。

尽管作者已经投入了大量时间和精力来编写此书,但由于水平有限,如有不足之处,敬请专家与同行批评指正。同时也欢迎广大读者与作者积极联系,共同探讨数据分析方面的心得与体会。Email:ruanjing@msn.com。

# 目 录

<b>第 1 章 SAS 系统环境与编程基础</b> .....	1
1.1 SAS 系统环境与界面 .....	1
1.2 SAS 数据处理的对象 .....	4
1.2.1 SAS 数据库 .....	4
1.2.2 SAS 数据集 .....	6
1.3 SAS 编程基础 .....	11
1.3.1 SAS 编程语言的基本结构 .....	11
1.3.2 SAS 编程语言的运算符和函数 .....	15
1.4 SAS 结构化编程语句 .....	17
1.4.1 条件语句 .....	17
1.4.2 循环语句 .....	20
1.5 SAS 运行方式 .....	26
1.5.1 SAS 系统内部菜单和外部加载项 .....	26
1.5.2 SAS 程序执行方式 .....	28
1.5.3 SAS Enterprise Miner .....	29
1.6 本章小结 .....	30
<b>第 2 章 SAS 数据准备</b> .....	31
2.1 直接输入 SAS 数据集 .....	31
2.1.1 基本数据输入方法 .....	31
2.1.2 变量数组输入方法 .....	33
2.1.3 数据输入指针控制 .....	34
2.1.4 数据输入格式设置 .....	41
2.1.5 数据输入修饰符的设置 .....	44
2.1.6 日期和时间型数据的输入 .....	50
2.1.7 使用循环语句自动读入数据 .....	53
2.1.8 自定义数据输入格式 .....	54
2.2 根据已有数据文件建立 SAS 数据集 .....	58
2.2.1 根据已有 SAS 数据集建立新 SAS 数据集 .....	58
2.2.2 在 SAS 系统中直接使用 EXCEL/ACCESS 的工作表数据 .....	60
2.2.3 导入 SAS 系统的外部数据文件 .....	62
2.3 输出 SAS 数据集 .....	67
2.3.1 输出 SAS 数据集的属性信息 .....	67
2.3.2 输出 SAS 数据集的数据信息 .....	69
2.3.3 自定义 SAS 数据集的数据输出格式 .....	71
2.3.4 导出 SAS 数据集到外部数据文件 .....	73
2.4 删除和更改数据库和数据集 .....	75
2.4.1 删除永久数据库 .....	75

2.4.2	删除数据集 .....	76
2.4.3	更改数据集名 .....	76
2.5	本章小结 .....	77
<b>第3章</b>	<b>SAS 数据整理 .....</b>	<b>78</b>
3.1	SAS 数据集的变量整理 .....	78
3.1.1	更改变量名称 .....	78
3.1.2	更改变量的顺序 .....	80
3.1.3	增加/删除变量 .....	80
3.1.4	变量类型的转换 .....	83
3.2	SAS 数据集的数据内容整理 .....	86
3.2.1	数据修正与更新 .....	86
3.2.2	删除观测值 .....	88
3.2.3	插入观测值 .....	89
3.2.4	调整数据位置 .....	90
3.2.5	数据查重 .....	93
3.2.6	数据变换 .....	94
3.2.7	自动插补缺失值 .....	97
3.3	SAS 数据集的整理 .....	98
3.3.1	提取子集 .....	98
3.3.2	数据合并 .....	99
3.3.3	数据追加 .....	108
3.3.4	数据集拆分 .....	110
3.4	SAS 数据整理的一个例子 .....	112
3.5	本章小结 .....	122
<b>第4章</b>	<b>SAS 宏与 CALL 子程序 .....</b>	<b>123</b>
4.1	SAS 宏变量与宏函数 .....	123
4.1.1	宏变量的定义与引用 .....	123
4.1.2	宏变量与数据变量值的交换 .....	127
4.1.3	宏函数 .....	130
4.2	SAS 宏与宏语句 .....	133
4.2.1	SAS 宏的定义与执行 .....	133
4.2.2	SAS 宏语句 .....	135
4.2.3	综合应用宏及宏语句的一些实例 .....	137
4.3	SAS 宏的调试、存储与调用 .....	143
4.3.1	SAS 宏的调试 .....	143
4.3.2	SAS 数据库中的宏存储与调用 .....	144
4.3.3	SAS 程序文件中的宏存储和调用 .....	146
4.4	自定义函数与 CALL 子程序 .....	146
4.4.1	创建和使用自定义函数 .....	147
4.4.2	创建和使用子程序 .....	149
4.4.3	自定义函数和子程序的主要区别 .....	150

4.5	本章小结	151
<b>第5章 SAS 输出管理</b> ..... 152		
5.1	在自定义窗口中输出和获取特定内容	152
5.1.1	在 DATA 步中自定义和显示窗口	152
5.1.2	从 DATA 步自定义窗口中获取变量的观测值	155
5.1.3	自定义和显示宏窗口	156
5.1.4	从自定义宏窗口中获取宏变量值	156
5.2	把 LOG 窗口和 SAS 过程结果输出至指定文件	157
5.2.1	指定存储 SAS 反馈信息和输出结果的文件	158
5.2.2	关闭 LOG 窗口和 SAS 过程的输出结果	161
5.3	SAS 输出传送系统(ODS)	161
5.3.1	选择输出内容	163
5.3.2	指定输出位置	165
5.3.3	指定输出格式	169
5.3.4	ODS 图形系统	170
5.4	本章小结	172
<b>第6章 统计量与统计表</b> ..... 173		
6.1	统计量	173
6.1.1	集中趋势	173
6.1.2	离散程度	177
6.1.3	分布形状	180
6.1.4	利用 SAS/INSIGHT 进行描述统计分析	181
6.1.5	利用 SAS 过程进行描述统计分析	182
6.2	统计表	190
6.2.1	统计表的基本要素	190
6.2.2	利用 FREQ 过程编制频数分布表	191
6.2.3	利用 TABULATE 过程编制统计表	193
6.2.4	利用 REPORT 过程编制统计表	196
6.3	SAS 常用过程能够计算的简单统计量	203
6.4	本章小结	204
<b>第7章 统计图</b> ..... 205		
7.1	SAS 系统中的统计图形绘制工具	205
7.1.1	SAS 菜单绘制统计图	205
7.1.2	SAS 图形绘制的全局语句	207
7.1.3	SAS 绘制统计图的常用过程	210
7.2	常见统计图形	213
7.2.1	直方图	213
7.2.2	条形图	216
7.2.3	折线图	220
7.2.4	饼图	223



7.2.5	盒须图	225
7.2.6	散点图	227
7.2.7	气泡图	233
7.2.8	瀑布图	234
7.2.9	星图	235
7.2.10	雷达图	236
7.2.11	日历图	236
7.2.12	向量图	237
7.2.13	阶梯图	238
7.2.14	针状图	238
7.2.15	点图	239
7.2.16	拼砖图	240
7.2.17	龙卷风图	241
7.2.18	函数作图	243
7.2.19	等高线图	244
7.2.20	数据地图	245
7.2.21	关键绩效指标图	246
7.2.22	不同类型的图形	248
7.3	本章小结	252
<b>第 8 章 简单统计推断</b>		253
8.1	简单统计推断的基本原理	253
8.1.1	数据分布	253
8.1.2	参数估计	256
8.1.3	假设检验	257
8.2	单总体参数的估计及假设检验	260
8.2.1	单总体的参数估计	260
8.2.2	单总体参数的假设检验	265
8.3	两总体参数的估计及假设检验	272
8.3.1	独立样本的参数估计和检验	272
8.3.2	成对样本的参数估计和检验	278
8.4	本章小结	280
<b>第 9 章 方差分析</b>		281
9.1	方差分析的基本原理	281
9.2	一元单因素方差分析	284
9.2.1	一元单因素方差分析的一般过程	284
9.2.2	方差分析模型的参数估计和预测	288
9.3	一元多因素方差分析	291
9.3.1	只考虑主效应的多因素方差分析	291
9.3.2	存在交互效应的多因素方差分析	297
9.4	协方差分析	299
9.5	多元方差分析	302

9.6	本章小结 .....	305
<b>第 10 章</b>	<b>非参数检验 .....</b>	<b>306</b>
10.1	非参数检验的基本问题 .....	306
10.2	单样本非参数检验 .....	307
10.2.1	均值的 Wilcoxon 符号秩检验 .....	307
10.2.2	样本分布的 Kolmogorov—Smirnov 检验 .....	308
10.2.3	随机性的游程检验 .....	310
10.3	两个样本的非参数检验 .....	312
10.3.1	独立样本中位数比较的 Wilcoxon 秩和检验 .....	312
10.3.2	独立样本分布的 Kolmogorov—Smirnov 检验 .....	316
10.3.3	成对样本中位数的 Wilcoxon 符号秩检验 .....	317
10.4	多个样本的非参数检验 .....	318
10.4.1	独立样本位置的 Kruskal—Wallis 检验 .....	318
10.4.2	独立样本位置的 Jonckheere—Terpstra 检验 .....	319
10.4.3	独立样本中位数的 Brown—Mood 检验 .....	320
10.5	本章小结 .....	321
<b>第 11 章</b>	<b>相关分析 .....</b>	<b>322</b>
11.1	两变量之间的相关分析 .....	322
11.1.1	简单相关分析 .....	322
11.1.2	偏相关分析 .....	326
11.1.3	非参数相关分析 .....	327
11.2	两组变量之间的相关分析——典型相关分析 .....	329
11.2.1	典型相关分析基本原理 .....	329
11.2.2	典型相关系数的显著性检验 .....	333
11.2.3	典型相关的冗余分析 .....	335
11.3	本章小结 .....	336
<b>第 12 章</b>	<b>回归分析 .....</b>	<b>337</b>
12.1	线性回归分析 .....	337
12.1.1	回归分析的基本原理 .....	337
12.1.2	一元线性回归分析 .....	340
12.1.3	多元线性回归分析 .....	344
12.2	定性自变量回归分析 .....	348
12.2.1	虚拟变量的设定 .....	348
12.2.2	含有虚拟变量的回归分析 .....	349
12.3	非线性回归分析 .....	351
12.3.1	可线性化的非线性分析 .....	351
12.3.2	非线性回归模型 .....	353
12.4	多项式回归分析 .....	357
12.5	分位数回归分析 .....	362
12.6	本章小结 .....	366

第 13 章 离散因变量模型 .....	367
13.1 线性概率模型 .....	367
13.2 二元选择模型 .....	368
13.2.1 线性概率模型的缺陷与改进 .....	368
13.2.2 二元选择模型的基本原理 .....	369
13.2.3 二元 PROBIT 模型 .....	370
13.2.4 二元 LOGIT 模型 .....	378
13.3 多重选择模型 .....	383
13.3.1 多重选择模型的基本原理 .....	383
13.3.2 顺序 PROBIT 模型 .....	384
13.3.3 顺序 LOGIT 模型 .....	388
13.3.4 无序 LOGIT 模型 .....	391
13.4 计数模型 .....	393
13.4.1 泊松回归模型的基本原理 .....	393
13.4.2 泊松回归模型的分析过程和步骤 .....	393
13.5 本章小结 .....	397
第 14 章 因子分析 .....	399
14.1 数据降维 .....	399
14.1.1 数据降维的基本问题 .....	399
14.1.2 数据降维的基本原理 .....	400
14.2 主成分分析 .....	400
14.2.1 主成分分析的基本概念与原理 .....	400
14.2.2 主成分分析的基本步骤和过程 .....	401
14.3 因子分析 .....	406
14.3.1 因子分析的基本原理 .....	406
14.3.2 因子分析的基本步骤和过程 .....	408
14.4 本章小结 .....	414
第 15 章 列联分析与对应分析 .....	415
15.1 列联分析 .....	415
15.1.1 列联表 .....	415
15.1.2 列联表的分布 .....	418
15.1.3 $\chi^2$ 分布与 $\chi^2$ 检验 .....	418
15.1.4 列联表中的关联度分析 .....	419
15.1.5 $\chi^2$ 分布的期望值准则 .....	422
15.2 对应分析 .....	423
15.2.1 对应分析的基本思想 .....	423
15.2.2 对应分析的步骤和过程 .....	423
15.3 本章小结 .....	431
第 16 章 聚类分析 .....	432

16.1	聚类分析的基本原理 .....	432
16.1.1	聚类分析的基本原则 .....	432
16.1.2	单一指标的系统聚类过程 .....	433
16.1.3	多指标的系统聚类过程 .....	435
16.2	聚类分析的步骤和过程 .....	439
16.2.1	系统聚类 .....	439
16.2.2	快速聚类 .....	446
16.2.3	变量聚类 .....	448
16.3	本章小结 .....	451
<b>第 17 章</b>	<b>判别分析 .....</b>	<b>452</b>
17.1	判别分析的基本思想 .....	452
17.2	判别分析的步骤和过程 .....	452
17.2.1	距离判别 .....	453
17.2.2	BAYES 判别 .....	453
17.2.3	非参数判别 .....	460
17.2.4	FISHER 判别 .....	461
17.2.5	逐步判别 .....	466
17.3	本章小结 .....	470
<b>第 18 章</b>	<b>时间序列分析 .....</b>	<b>471</b>
18.1	时间序列的基本问题 .....	471
18.1.1	时间序列的组成部分 .....	471
18.1.2	时间序列的平稳性 .....	472
18.2	ARIMA 模型的分析过程 .....	476
18.2.1	ARIMA 模型 .....	476
18.2.2	ARMA 模型的识别、估计与预测 .....	477
18.2.3	利用 SAS 时间序列预测系统进行 ARIMA 建模 .....	485
18.3	本章小结 .....	492
<b>附录 I</b>	<b>本书所介绍的全局语句、宏语句、DATA 步语句和 PROC 过程通用语句索引 .....</b>	<b>493</b>
<b>附录 II</b>	<b>本书所介绍的主要 SAS 过程索引 .....</b>	<b>494</b>

“Linux - Super” 或者 “English with DUCS” 的语言版本, 或者选择 “Chinese (Simplified)” 或 “Chinese (Traditional)” 的语言版本。

在 SAS 主界面中, 位于最上面一行的系统菜单可以实现 SAS 文件和数据库操作、编辑、视图、建模分析、绘图等一系列的功能; 系统菜单视当前调用的不同 SAS 模块会有所不同。

- File: 主要实现 SAS 系统文件的基本操作, 包括新建、打开、合并程序、存储、打印文件、导入和导出外部数据文件、利用 Email 传送文件等。
- Edit: 主要实现对文本进行选择、查找、复制、粘贴、剪切、粘贴等编辑功能。
- View: 主要进行窗口切换或打开对应的窗口。
- Tools: 可以打开数据库查询与管理器、表格、图形、报告、文本编辑器, 并且对 SAS 系统的全局参数、界面、字体、颜色等方案进行调整。
- Run: 可以运行程序以及控制程序运行的方式。

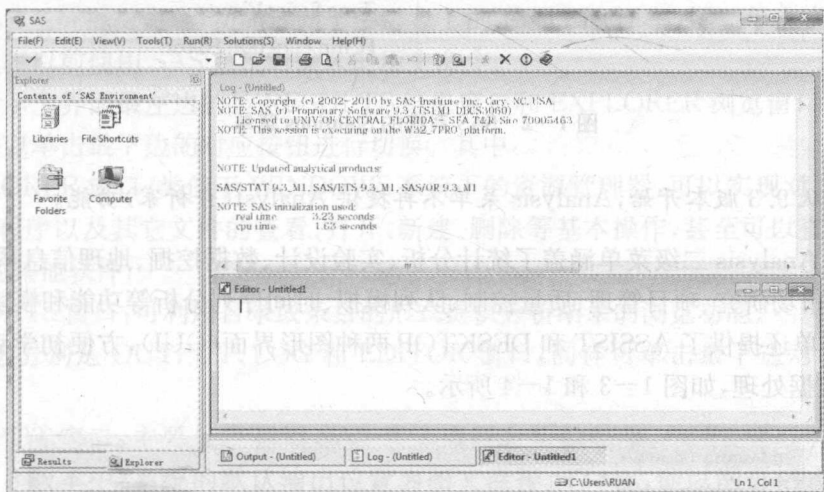
# 第 1 章

## SAS 系统环境与编程基础

SAS 系统强大的数据分析功能主要体现在编程上, SAS 所提供的语句和过程涵盖了数据搜集、数据整理整合、数据分析与建模、数据发布等数据分析的全过程。因此, 掌握 SAS 编程的基本思路和技能是利用其进行数据分析的基本要求。本章在介绍 SAS 系统环境与基本界面的基础上对编程基础进行梳理, 为后续章节打下基础。

### 1.1 SAS 系统环境与界面

选择合适的界面语言进入 SAS 系统的主界面, 如图 1-1 所示。




本书感谢美国中佛罗里达大学统计系和数据挖掘中心在编者访学和本书编写期间授权使用 SAS9.3 系统

图 1-1 SAS 9.3 主界面

**注意** SAS 系统提供有不同的语言。如果在英文界面下所处理数据中含有中文则应当选择“Unicode Support”或者“English with DBCS”的语言版本, 或者选择“Chinese (Simplified)”或“Chinese(Traditional)”的语言版本。

在 SAS 主界面中, 处于最上面一行的系统菜单可以实现 SAS 文件和数据库操作、编辑、视图、建模分析、绘图等一系列的功能, 系统菜单根据所调用的不同 SAS 模块会有所不同。

- File: 主要实现 SAS 系统文件的基本操作, 包括新建、打开、合并程序, 存储、打印文件, 导入和导出外部数据文件, 利用 Email 传送文件等;
- Edit: 主要实现对文本进行选择、查找、替换、复制、剪切、粘贴等编辑功能;
- View: 主要进行窗口切换或打开对应的窗口;
- Tools: 可以打开数据库查询与管理器, 表格、图形、报告、文本编辑器, 并且对 SAS 系统的全局参数、界面、字体、颜色等方案进行调整;
- Run: 可以运行程序以及控制程序运行的方式;

- Window: 主要实现窗口切换、窗口排列等窗口操作功能;
- Help: 是 SAS 系统的强大帮助系统, 提供了从菜单操作到编程语言一系列的帮助功能, 比市面上任何一本 SAS 书籍内容都准确和全面; 同时在使用 SAS 系统进行分析的任意过程中, 只要点击工具栏上的  图标, 便可以快速的转到该过程对应的帮助信息上;
- Solutions: 利用 SAS 进行数据分析、程序开发的最主要菜单, 如图 1-2 所示。

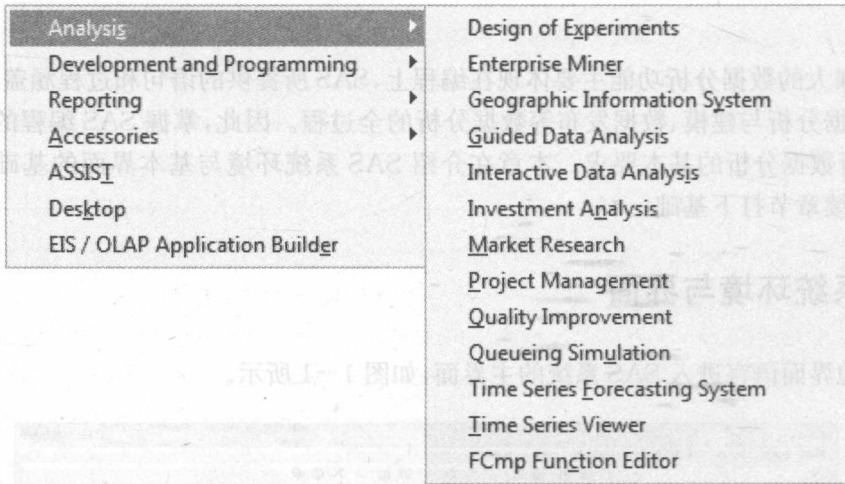


图 1-2 SAS 系统的 Solutions 菜单

**注意** SAS 从 9.3 版本开始, Analysis 菜单不再提供 Analyst(分析家)功能。

该菜单下的 Analysis 二级菜单涵盖了统计分析、实验设计、数据挖掘、地理信息系统、探索性数据分析、投资分析、市场研究、项目管理、质量控制、队列模拟、时间序列分析等功能和模块。

Solutions 菜单还提供了 ASSIST 和 DESKTOP 两种图形界面(GUI), 方便初学者在较短时间内利用 SAS 进行数据处理, 如图 1-3 和 1-4 所示。

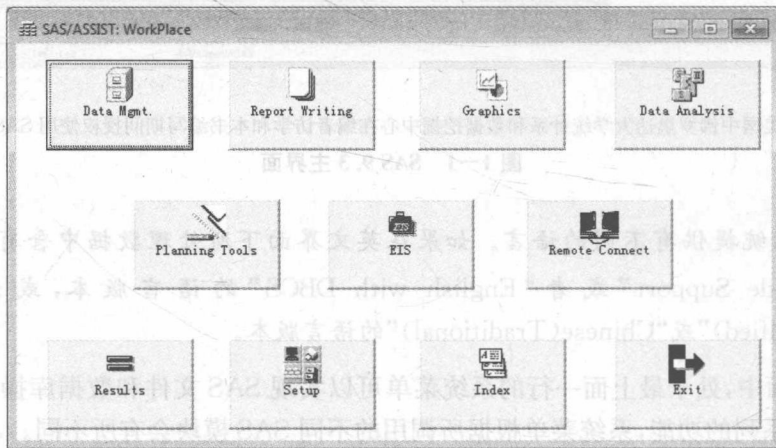



图 1-3 SAS/ASSIST 的 GUI

用户可以点击 GUI 上的图标进入对应的分析功能和模块。

在系统菜单之下是 SAS 工具栏, 与大多数软件一样, 该工具栏提供了能够实现系统菜单中常用功能的快捷方式, 并且以 MICROSOFT WINDOWS 操作系统中常用的图标来表示, 如图 1-5 所示。

SAS 工具栏的特殊之处在于工具栏的左边有一个文本输入框, 称之为命令框, 可以按 F11 键快速切换至该处, 用户可以在框中输入命令进行软件操作或者调用 SAS 系统的各种模块。命令框前的 

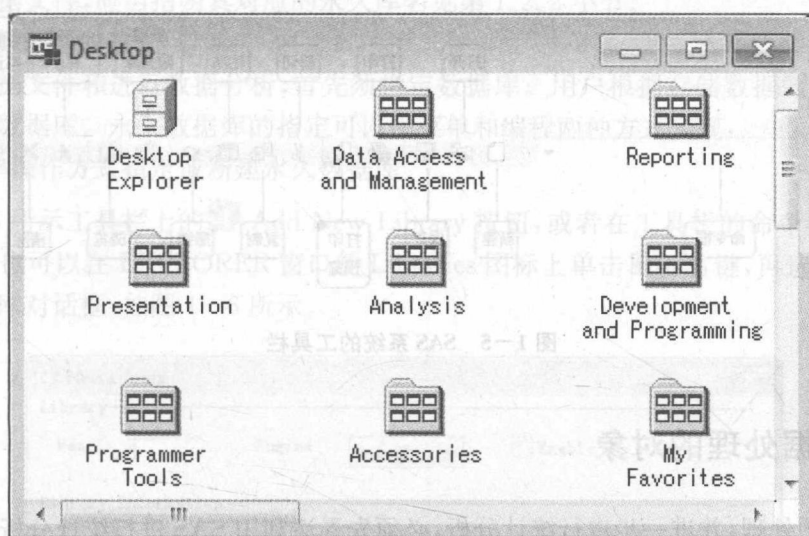


图 1-4 SAS/Desktop 的 GUI


按钮表示向 SAS 系统提交命令,功能等同于按键盘上的 ENTER(回车)键。单击命令框之后的  按钮,可以从中选择以前使用 SAS 系统时输入的历史命令。

图 1-1 所示主界面最左边的是 RESULTS 结果窗口和 EXPLORER 浏览窗口,这 2 个窗口处于同一界面下,可以单击最下边的对应按钮进行切换。其中:

- EXPLORER 窗口:类似于 WINDOWS 系统下的资源管理器,可以实现对 SAS 数据库、数据文件、程序以及其它文件的查看、打开、新建、删除等基本操作,甚至可以直接运行 SAS 系统之外的其他软件;
- RESULTS 窗口:可利用目录或条目的形式提供分析结果的浏览功能。

主界面右边分别是 OUTPUT、LOG 和 EDITOR 窗口,同样可单击最下边对应按钮进行切换。其中:

- OUTPUT 窗口:主要呈现利用 SAS 系统进行分析的结果,按 F7 键可切换至该窗口。在 SAS9.3 版本中,系统的默认输出位置为图文混排 HTML,即以网页格式默认输出至 RESULTS VIEWER 窗口,所有的统计表格和图形都可以呈现在一个 HTML 文件中。如果想要把输出结果恢复为 9.3 版本之前在 OUTPUT 窗口的输出,可以运行如下语句:

```
ods listing;
```

- LOG 窗口:即 SAS 系统的日志窗口,按 F6 键可以切换至该窗口。用户对 SAS 的每一步骤操作,都会以日志的形式列示在 LOG 窗口当中(如每次打开 SAS 系统的时候,在该窗口中默认显示一些授权文件的信息)。在进行数据分析的过程中,一些出错信息和程序执行的反馈信息也会出现在该窗口中。同时,用户也可使用 SAS 语句通过运行程序的方式在 LOG 窗口中写入文本信息;
- EDITOR 窗口:为一个文本编辑器,按 F5 键可以切换至该窗口。用户可以在该窗口中进行编程或输入文字信息(LOG 和 OUTPUT 窗口用户不能直接写入信息,只能通过程序运行的方式写入信息)。

这三个窗口的信息,用户可以通过系统菜单或者工具栏上的按钮进行复制、新建、存储、打印等文件操作。

这些默认的窗口可以根据用户的需要使用两种方式关闭和打开:一种方式是在图 1-5 所示的命令框中输入对应窗口的名字;另一种方式是利用 SAS 的 View 或 Window 系统菜单,单击对应的窗口名字即可。

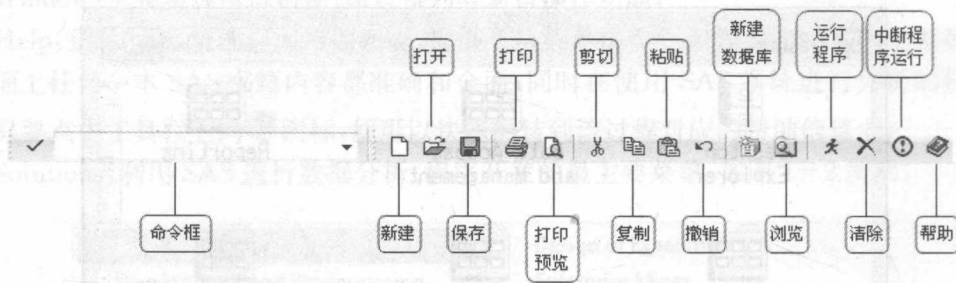


图 1-5 SAS 系统的工具栏

## 1.2 SAS 数据处理的对象

要进行数据预处理,并进一步进行统计分析,必须先弄清利用 SAS 进行统计分析的对象。

SAS 数据处理的对象是存在于 SAS 数据库中的数据集。具体而言,数据集是 SAS 的数据处理对象,同时也是统计分析的基础。但是 SAS 数据集应当存在于 SAS 数据库中,离开了数据库,就没有所谓的 SAS 数据集了。因此,SAS 数据库与 SAS 数据集之间的关系尤为重要。

### 1.2.1 SAS 数据库

SAS 数据库具体是指存放 SAS 数据文件的文件夹,它与计算机存储器中某一个具体的物理文件夹相对应。

#### 1. SAS 数据库的分类

为了让 SAS 系统识别物理文件夹所对应的数据库,要为每一个数据库指定一个库标记(即数据库的名字)来识别该库,库标记是逻辑存在的,只存在于 SAS 系统中,同一个文件夹可以对应不同的逻辑库标记,就像一个人可以有很多个称呼或绰号一样。根据数据处理的时效性不同,SAS 数据库又可以分为临时库和永久库。

- 临时库:只有一个,名为 WORK。在每次启动 SAS 时系统自动生成,关闭 SAS 时该数据库中的所有数据文件自动被清除;
- 永久库:可有多个。用户可以自己指定永久库的库标记,库中所有数据文件永久保留,但库标记是临时的,每次启动 SAS 系统都要重新指定。

在每次启动 SAS 的时候,系统都会根据用户安装 SAS 时授权文件的模块,自动指定若干个库标记,其中有 3 个库标记是不可缺少的:

- SASUSER:为永久库。该库中的数据文件可以被永久保存起来,以便下次系统启动时使用,并存储了 SAS 用户信息。其具体对应于安装 SAS 系统时指定的路径,读者可以在 EXPLORER 窗口中,双击 Libraries 图标,在名为 SASUSER 的数据库上单击鼠标右键选择 Properties 即可在弹出的对话框中看到该数据库所对应的物理路径。

**注意** 本书所用例子的数据集文件如无特殊说明均存储在 SASUSER 数据库中,读者可把本书中的数据复制到自己计算机中该数据库对应的文件夹。

- SASHELP:为永久库。该数据库包含了控制各种 SAS 运算进程的信息,具体对应于安装 SAS 系统时的根目录。
- WORK:为临时库。该库中的数据文件可以被暂时保存起来,关闭 SAS 系统时用户保存在该库内的数据文件将被清除。

在 SAS 中引用临时库中的数据文件可以省略库名,即临时库被认为是缺省的数据库。如果要引




用永久库中的数据文件,应当指明其对应的永久库名见第 1.2.2 小节。

## 2. SAS 永久数据库的建立

为了存储数据文件和进行数据分析,首先须指定数据库。用户根据存储数据的文件夹所指定的数据库都是永久数据库。永久数据库的指定可以用菜单和编程两种方式实现:

### (1) 通过菜单操作方式指定或新建永久数据库

单击图 1-5 所示工具栏上的  Add New Library 按钮,或者在工具栏的命令框中输入 DMLIBASSIGN 命令,也可以在 EXPLORER 窗口的 Libraries 图标上单击鼠标右键,再选择 New 菜单,均会弹出新建数据库对话框,如图 1-6 所示。

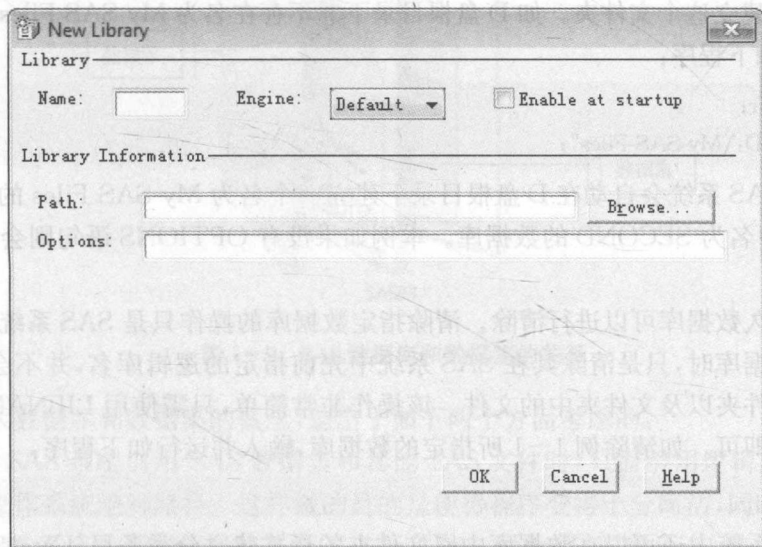



图 1-6 新建数据库的对话框

该对话框可以实现指定数据库属性的全部功能,具体如下:

Library 分栏下可以通过 Name 文本输入框输入数据库的库标记(库名),Engine 下拉列表用于指定包括 SPSS、Excel、Access、Sybase 等常用软件的 30 多种数据引擎版本,日常使用选择默认的 Default 选项就可以了。此外,如果把 Enable at startup 复选框钩上,就可以让 SAS 系统在每次启动的时候,自动建立 Name 文本输入框所指定的数据库。

Library Information 分栏主要用于指定数据库与文件夹路径的对应关系。在 Path 文本输入框中,可以输入将要建立的 SAS 数据库对应的文件夹所在的绝对路径,同时也可以利用该文本输入框右边的 Browse 按钮,调用系统资源管理器找到文件夹所处的位置。

 **例 1-1** 建立一个名为 FIRST 的永久数据库,该数据库对应的文件夹绝对路径为 D:\Statistics\SAS\SAS 93。

根据上述的菜单操作方法,打开如图 1-6 所示的新建数据库对话框,在 Name 文本输入框中输入 FIRST 作为该永久数据库的库标记,在 Path 文本输入框中输入其对应文件夹路径 D:\Statistics\SAS\SAS 93 或者单击 Browse 按钮找到该文件夹对应的位置,然后单击 OK 即可,建立的数据库会在 EXPLORER 窗口的 Libraries 内容中显示出来。

### (2) 编程方式指定或新建永久数据库

利用程序来建立永久数据库要用到 LIBNAME 全局语句,其基本语法如下:

```
LIBNAME 库标记<,路径<,数据引擎<,数据引擎的选项>>>;
```

其中库标记即数据库的名字长度不能超过 8 个字符,其命名规范与一般标准数据库系统或软件