

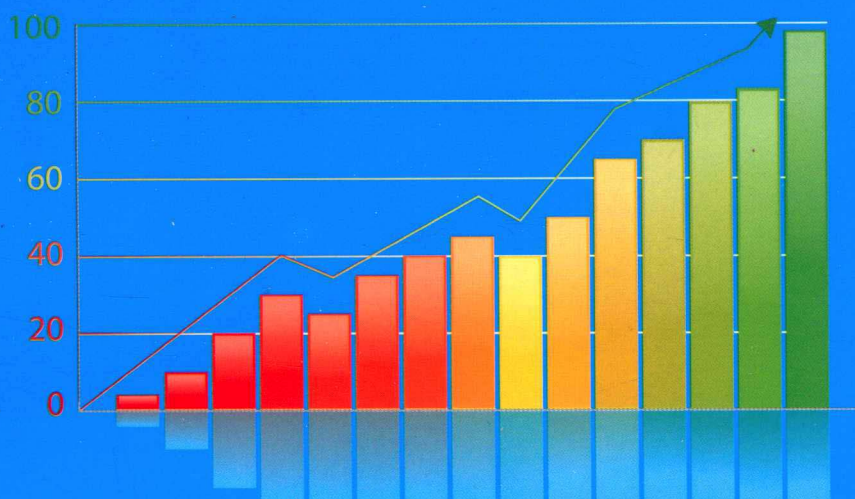
160个实例、6个案例，详细介绍18种常用的统计分析方法
以贴近实际的例子引导读者学习，亲切自然，生动有趣

SAS

统计分析实用宝典

(17.3小时多媒体教学视频)

姚鑫锋 王薇 等编著



超值DVD光盘内容

- ◎ 本书配套教学视频
- ◎ 本书源文件及数据文件



清华大学出版社

SAS

统计分析实用宝典

姚鑫锋 王薇 等编著



清华大学出版社

北 京

内 容 简 介

本书用通俗易懂的语言阐述了 SAS 统计分析软件的使用和各种统计分析方法的基本原理。内容由浅入深、循序渐进,书中的每个知识点都有相应的实例演示,每章都附有练习题,帮助读者从实际角度体验统计方法的 SAS 实现过程。本书配 1 张光盘,内容为本书配套教学视频和涉及的源数据。

本书共 19 章,分 4 篇。第 1 篇介绍了 SAS 的基础知识,重点讲述入门操作,包括程序软件的安装、配置等。第 2 篇介绍了常用统计方法的 SAS 实现,包括描述性统计分析、列联表分析、统计假设测验、非参数统计分析、方差分析、相关与回归分析、聚类分析、判别分析、主成分分析、因子分析、典型相关分析、时间序列分析和生存分析等。第 3 篇介绍了常用统计图形的绘制与 SAS 宏编程。第 4 篇介绍了 SAS 在社会调查、企业销售、医学、电力、证券和农业等领域的具体应用。

本书面向零基础、编程基础差或不了解 SAS 统计软件的初学者,尤其适合有一定的数学或统计学背景、需要使用 SAS 软件完成各种统计分析操作的读者。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

SAS 统计分析实用宝典/姚鑫锋等编著. —北京:清华大学出版社,2013.2
ISBN 978-7-302-30736-5

I. ①S… II. ①姚… III. ①统计分析—应用软件 IV. ①C819

中国版本图书馆 CIP 数据核字(2012)第 284031 号

责任编辑:夏兆彦

封面设计:欧振旭

责任校对:徐俊伟

责任印制:何 芊

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座

邮 编:100084

社总机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京鑫丰华彩印有限公司

装 订 者:三河市李旗庄少明印装厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:23.25

字 数:584 千字

附光盘 1 张

版 次:2013 年 2 月第 1 版

印 次:2013 年 2 月第 1 次印刷

印 数:1~4000

定 价:49.80 元

产品编号:049843-01

前言

SAS 全称是 Statistics Analysis System，是目前国际上权威的统计分析软件之一，具有强大的数据管理与分析功能，广泛地应用于教育、经济、医学等各个领域。SAS 软件具有强大的统计分析功能，为了更好地使用这些功能，往往需要依赖于程序设计。SAS 系统全面涵盖了各项统计分析功能，具有操作简单、功能强大、结果专业等特点，为用户提供了一种专门用于 SAS 软件的编程语言，其语言类似于一般的编程语言，用户可以方便地在程序编辑窗口完成程序的设计工作。

SAS 9.2 版本支持多种语言，提供了中文的操作环境，便于国内读者的使用，但目前以 SAS 9.2 为基础的 SAS 教程较少。本书为了帮助国内读者更好地掌握 SAS 语言，特意以较新的 SAS 9.2 中文版为例，向读者讲述 SAS 语言的操作，包括 SAS 安装、配置到各种统计学操作在 SAS 统计学软件中的实现方式。

本书全面介绍了 SAS 常用统计分析功能，涉及 SAS 编程技术与 SAS 界面操作，以满足不同读者群体的需要。因此，本书不但适合不擅长编程的初学者，也适合需要进一步提高 SAS 软件实际操作能力的读者。本书致力于用户对 SAS 软件的从零开始的学习，并配备大量的实例操作和视频教程，以帮助读者更好地掌握 SAS 语言，提高解决统计分析问题的能力。

本书特色

1. 配教学视频、实例文件、习题，教学资源丰富

本书每章都配备一定量的习题，帮助读者练习，巩固所学。同时，本书中的所有实例都提供源程序和数据文件，方便读者自行演练。另外，本书还配备了长时间的教学视频，以便于读者更快地掌握各统计操作的实现过程。

2. 内容全面

本书全面涵盖了 SAS 统计软件常用的各种统计分析功能，并对各知识点都做了详细介绍，让读者能全面掌握各种操作在 SAS 中的具体实现。

3. 较新的软件版本

本书以目前较新的 SAS 9.2 版本为例，向读者介绍 SAS 软件的操作。该版本的 SAS 软件支持中文的操作环境，因而更适合国内读者的学习和使用。

4. 实例丰富，实用性强

本书各知识点的讲解都配备相应的实例，向读者具体演示各知识点的操作，力求培养

读者解决实际问题的能力。

5. 适用面广，适合各层次读者

本书由浅入深地介绍了 SAS 的各项统计分析功能，并详细介绍了编程和界面操作两种 SAS 统计实现方法，可以满足不同层次读者的需要。

本书主要内容

本书共分为 4 篇，各部分对应的章节和具体内容如下。

第 1 篇为基本统计操作入门，包括第 1、2 章。主要介绍了 SAS 的基础知识，包括软件的背景介绍、特点、功能、安装、启动、操作界面、语言规范等。

第 2 篇为常用的统计分析，包括第 3~16 章。主要介绍了描述性统计分析、非参数检验、方差分析、相关分析、回归分析、聚类分析、判别分析、因子分析、主成分分析、生存分析等内容。

第 3 篇为图形绘制与 SAS 宏编程，包括第 17、18 章。主要介绍了常用的统计图形在 SAS 系统内的绘制方式，包括散点图、直方图、饼图、线图等；另外介绍了 SAS 宏编程的基础知识。

第 4 篇为综合实例分析，包括第 19 章。在社会调查、企业销售、医学、电力、证券和农业等领域的具体应用。

本书读者对象

本书内容全面，可读性强，适应面广，适合阅读的人员有：

- ☐ SAS 入门新手；
- ☐ 统计学专业的学生和老师；
- ☐ 统计分析从业人员；
- ☐ 各个专业领域需要使用 SAS 解决问题的用户；
- ☐ SAS 中高级用户。

本书作者

本书主要由姚鑫锋和王薇主笔编写。其他参与编写的人员有安静、崔莎、韩静、彭维、滕川、张帆、张玉媛、周曦、陈世琼、陈欣、陈智敏、董加强、范礼、郭秋滢、郝红英、蒋春蕾、黎华、刘建准、刘霄、刘亚军、刘仲义、柳刚、罗永峰。

本书在写作过程中参考和借鉴了相关图书，在此对原作者致谢！另外特别感谢我们的师长、家人和朋友及所有帮助过我们的人。

编著者

目 录

第 1 篇 基本统计操作入门

第 1 章 SAS 入门 (教学视频: 75 分钟)	2
1.1 SAS 系统概况	2
1.1.1 SAS 的历史背景	2
1.1.2 SAS 的系统特点	2
1.1.3 SAS 的功能模块	3
1.2 SAS 的安装与启动	3
1.2.1 SAS 的安装	3
1.2.2 SAS 的启动	6
1.3 SAS 的操作界面	7
1.3.1 主界面	7
1.3.2 程序编辑窗口	10
1.3.3 运行日志窗口	11
1.3.4 输出窗口	11
1.3.5 结果窗口	11
1.3.6 资源管理器窗口	11
1.3.7 帮助窗口	13
1.4 SAS 的界面操作简介	14
1.4.1 INSIGHT 界面操作	14
1.4.2 ANALYST 界面操作	17
1.4.3 ASSIST 界面操作	20
1.5 SAS 系统的文件管理	21
1.5.1 SAS 数据集	21
1.5.2 SAS 逻辑库	23
1.6 本章小结	25
1.7 习题	25
第 2 章 SAS 程序设计入门 (教学视频: 109 分钟)	26
2.1 SAS 语言基础	26
2.1.1 SAS 程序的组成	26
2.1.2 结构化语句	27

2.2	SAS 程序结构	31
2.2.1	数据步介绍	31
2.2.2	过程步介绍	33
2.3	SAS 中数据步常用语句	35
2.3.1	赋值语句	35
2.3.2	INPUT 语句	36
2.3.3	INFILE 语句	38
2.3.4	PUT 和 FILE 语句	38
2.3.5	LENGTH 语句	40
2.3.6	LABEL 语句	41
2.3.7	KEEP 语句	42
2.3.8	DROP 语句	42
2.3.9	WHERE 语句	42
2.4	SAS 中过程步常用语句	43
2.4.1	输出语句	43
2.4.2	打印语句	43
2.4.3	标注语句	44
2.4.4	MODEL 语句	45
2.4.5	VAR 语句	45
2.4.6	ID 语句	45
2.4.7	BY 语句	46
2.4.8	CLASS 语句	47
2.5	常用函数	48
2.5.1	数学函数	48
2.5.2	统计函数	49
2.5.3	数组函数	50
2.5.4	字符串函数	50
2.5.5	时间函数	51
2.5.6	概率分布函数	52
2.5.7	分位数函数	52
2.5.8	随机数函数	53
2.6	程序的常见错误	54
2.6.1	词法错误	54
2.6.2	语法错误	56
2.7	本章小结	58
2.8	习题	58

第 2 篇 常用的统计分析

第 3 章	SAS 数据管理 (教学视频: 76 分钟)	60
3.1	数据的导入和导出	60

3.1.1	数据的导入	60
3.1.2	数据的导出	63
3.2	数据集的基本操作	65
3.2.1	数据集的复制	65
3.2.2	数据集的新增	67
3.2.3	数据集的拆分	68
3.2.4	数据集的连接	69
3.2.5	数据集的更新	72
3.2.6	数据集的排序	73
3.2.7	数据集的转置	74
3.2.8	数据集的删除	75
3.3	利用 INSIGHT 模块实现数据管理	76
3.4	利用 ANALYST 模块实现数据管理	80
3.5	本章小结	84
3.6	习题	84
第 4 章	描述性统计分析 (教学视频: 51 分钟)	86
4.1	描述性统计分析概述	86
4.1.1	数据位置描述	86
4.1.2	离散程度描述	87
4.1.3	分布形状描述	88
4.2	描述性统计分析过程	88
4.2.1	MEANS 均值过程	88
4.2.2	UNIVARIATE 单变量过程	92
4.3	利用界面操作进行简单描述性统计	94
4.3.1	INSIGHT 模块操作	94
4.3.2	ANALYST 模块操作	96
4.4	本章小结	100
4.5	习题	100
第 5 章	列联表分析 (教学视频: 28 分钟)	102
5.1	列联表分析的编程实现	102
5.1.1	FREQ 频数过程	102
5.1.2	单变量列联表分析	103
5.1.3	2×2 列联表分析	105
5.1.4	n 维列联表分析	107
5.2	利用界面操作进行列联表分析	109
5.3	本章小结	111
5.4	习题	111
第 6 章	统计推断 (教学视频: 49 分钟)	113
6.1	统计推断的基本原理与方法	113

6.1.1 参数估计	113
6.1.2 假设测验	113
6.2 统计推断的 TTEST 过程语句	114
6.3 统计假设测验的界面实现	115
6.3.1 INSIGHT 模块操作	115
6.3.2 ANALYST 模块操作	118
6.4 统计假设测验的实例	120
6.4.1 单个样本平均数的假设测验	121
6.4.2 成对样本的平均数比较的假设测验	122
6.4.3 两个样本的平均数比较的假设测验	123
6.5 本章小结	124
6.6 习题	124
第 7 章 非参数统计分析 (教学视频: 28 分钟)	125
7.1 非参数统计分析概述	125
7.2 非参数统计分析的 NPAR1WAY 语句	127
7.3 非参数统计分析实例	128
7.3.1 单个样本的非参数检验	128
7.3.2 两个样本的非参数检验	129
7.3.3 多个样本的非参数检验	131
7.4 利用界面操作实现非参数统计分析	132
7.5 本章小结	133
7.6 习题	133
第 8 章 方差分析 (教学视频: 45 分钟)	134
8.1 方差分析概述	134
8.1.1 方差分析基本概念	134
8.1.2 单因素方差分析	134
8.1.3 多因素方差分析	135
8.1.4 多重比较	135
8.2 方差分析的编程实现	135
8.2.1 ANOVA 过程	135
8.2.2 GLM 过程	140
8.3 方差分析的界面操作	144
8.3.1 利用 INSIGHT 模块实现方差分析	144
8.3.2 利用 ANALYST 模块实现方差分析	145
8.4 本章小结	148
8.5 习题	149
第 9 章 相关与回归分析 (教学视频: 90 分钟)	150
9.1 相关分析	150
9.1.1 相关分析概述	150

9.1.2	相关分析的 CORR 过程	150
9.1.3	相关分析的界面操作	152
9.2	一元线性回归分析	155
9.2.1	一元线性回归的基本概述	155
9.2.2	一元线性回归模型的 REG 过程	156
9.2.3	一元线性回归的界面操作	161
9.3	多元线性回归分析	165
9.3.1	多元线性回归模型概述	165
9.3.2	多元线性回归模型的建立	166
9.4	逐步回归分析	173
9.4.1	逐步回归模型概述	173
9.4.2	逐步回归模型的建立	173
9.5	非线性回归分析	177
9.5.1	可化为线性回归的非线性回归	177
9.5.2	非线性回归 NLIN 过程	178
9.6	Logistic 回归	182
9.6.1	Logistic 回归模型概述	182
9.6.2	Logistic 回归过程	183
9.7	本章小结	186
9.8	习题	186
第 10 章	聚类分析 (教学视频: 47 分钟)	189
10.1	聚类分析方法概述	189
10.2	系统聚类	190
10.2.1	系统聚类方法概述	190
10.2.2	CLUSTER 过程	190
10.2.3	TREE 过程	191
10.2.4	系统聚类实例演示	192
10.3	变量聚类	195
10.3.1	变量聚类方法概述	195
10.3.2	VARCLUS 过程	195
10.3.3	变量聚类实例演示	196
10.4	快速聚类	200
10.4.1	快速聚类方法概述	200
10.4.2	FASTCLUS 过程	201
10.4.3	快速聚类实例演示	202
10.5	本章小结	205
10.6	习题	205
第 11 章	判别分析 (教学视频: 49 分钟)	208
11.1	判别分析方法概述	208

11.1.1	判别分析的基本思想	208
11.1.2	判别分析的基本步骤	208
11.2	一般判别分析	209
11.2.1	一般判别分析方法概述	209
11.2.2	DISCRIM 一般判别分析过程	209
11.2.3	一般判别分析实例	211
11.3	典型判别分析	214
11.3.1	典型判别分析方法概述	214
11.3.2	CANDISC 典型判别分析过程	214
11.3.3	典型判别分析实例	215
11.4	逐步判别分析	217
11.4.1	逐步判别分析方法概述	217
11.4.2	STEPDISC 逐步判别分析过程	218
11.4.3	逐步判别分析实例	218
11.5	本章小结	221
11.6	习题	221
第 12 章	主成分分析 (教学视频: 35 分钟)	224
12.1	主成分分析方法概述	224
12.1.1	主成分分析的数学模型	224
12.1.2	主成分分析的基本步骤	225
12.2	主成分分析的 SAS 过程	225
12.3	主成分分析的界面操作	226
12.3.1	ANALYST 模块实现主成分分析	227
12.3.2	INSIGHT 模块实现主成分分析	232
12.4	主成分分析实例	236
12.5	本章小结	238
12.6	习题	238
第 13 章	因子分析 (教学视频: 25 分钟)	241
13.1	因子分析方法概述	241
13.1.1	因子分析的基本原理	241
13.1.2	因子分析的基本步骤	241
13.2	因子分析的 SAS 过程	242
13.3	因子分析的界面操作	247
13.4	本章小结	252
13.5	习题	252
第 14 章	典型相关分析 (教学视频: 27 分钟)	255
14.1	典型相关分析方法概述	255
14.1.1	典型相关分析的基本思想	255
14.1.2	典型相关分析的基本步骤	255

14.2	典型相关分析的 SAS 过程	256
14.3	典型相关分析的界面操作	260
14.3.1	INSIGHT 模块实现典型相关分析	260
14.3.2	ANALYST 模块实现典型相关分析	263
14.4	本章小结	264
14.5	习题	265
第 15 章	时间序列分析 (教学视频: 36 分钟)	266
15.1	时间序列分析方法概述	266
15.1.1	时间序列分析的数学模型	266
15.1.2	时间序列分析的基本步骤	267
15.2	时间序列分析的 SAS 过程	267
15.3	时间序列分析实例	269
15.4	本章小结	274
15.5	习题	274
第 16 章	生存分析 (教学视频: 55 分钟)	276
16.1	生存分析概述	276
16.1.1	生存分析的基本概况	276
16.1.2	生存分析的基本方法	277
16.2	生存分析的相关过程	277
16.2.1	LIFEREG 回归过程	277
16.2.2	LIFETEST 检验过程	280
16.2.3	PHREG 回归过程	284
16.3	利用界面操作实现生存分析	287
16.4	本章小结	291
16.5	习题	291

第 3 篇 图形绘制与 SAS 宏编程

第 17 章	图形绘制 (教学视频: 96 分钟)	294
17.1	GPlot 过程绘图	294
17.1.1	GPlot 过程	294
17.1.2	散点图	295
17.1.3	线图	297
17.2	GChart 过程绘图	298
17.2.1	GChart 过程	298
17.2.2	直方图	299
17.2.3	饼图	300
17.3	GContour 过程绘图	302
17.4	其他图形绘制	303

17.4.1	盒形图	303
17.4.2	茎叶图	305
17.4.3	Q-Q 图	306
17.5	图形编辑	307
17.6	利用图形界面操作绘制图形	309
17.5.1	利用 INSIGHT 模块绘制统计图形	309
17.5.2	利用 ANALYST 模块绘制统计图形	314
17.7	本章小结	318
17.8	习题	318
第 18 章	SAS 宏编程 (教学视频: 62 分钟)	322
18.1	宏变量	322
18.1.1	宏变量的定义	322
18.1.2	宏变量的引用	326
18.2	宏过程	327
18.2.1	宏过程的定义	327
18.2.2	宏过程的调用	328
18.2.3	宏过程的参数	329
18.3	宏语句	329
18.4	宏函数	332
18.5	本章小结	335
18.6	习题	335

第 4 篇 综合实例分析

第 19 章	SAS 应用实例 (教学视频: 54 分钟)	338
19.1	在社会调查统计中的应用	338
19.1.1	实例问题描述	338
19.1.2	采用的统计分析方法及其 SAS 实现	339
19.1.3	主要分析结果	341
19.2	在企业销售数据中的应用	342
19.2.1	实例问题描述	342
19.2.2	采用的统计分析方法及其 SAS 实现	342
19.2.3	主要分析结果	343
19.3	在医学中的应用	343
19.3.1	实例问题描述	343
19.3.2	采用的统计分析方法及其 SAS 实现	344
19.3.3	主要分析结果	345
19.4	在电力行业的应用	346
19.4.1	实例问题描述	346

19.4.2	采用的统计分析方法及其 SAS 实现	347
19.4.3	主要分析结果	348
19.5	在证券行业的应用	349
19.5.1	实例问题描述	349
19.5.2	采用的统计分析方法及其 SAS 实现	350
19.5.3	主要分析结果	351
19.6	在农业领域的应用	352
19.6.1	实例问题描述	352
19.6.2	采用的统计分析方法及其 SAS 实现	353
19.6.3	主要分析结果	353
19.7	本章小结	354
19.8	习题	354

第 1 篇 基本统计操作入门

» 第 1 章 SAS 入门

» 第 2 章 SAS 程序设计入门

第 1 章 SAS 入门

本章将向读者介绍 SAS 的入门知识。首先,将通过 SAS 系统的概述,帮助读者对 SAS 系统有初步的了解,了解其历史背景、系统特点和主要功能模块;接着,将向入门读者简单介绍 SAS 的安装与启动;最后,对 SAS 主要工作界面的介绍将使读者对软件的操作环境有所熟悉。

1.1 SAS 系统概况

SAS (Statistics Analysis System) 是目前国际上权威的统计分析软件之一,具有强大的数据管理与分析功能,广泛地应用于教育、经济、医学等各个领域。为使读者对 SAS 软件基本情况有所了解,本节将从 SAS 的历史背景、系统特点和功能模块三个方面向读者介绍 SAS 概况。

1.1.1 SAS 的历史背景

SAS 最初是由美国北卡罗纳大学的研究生编写的,随后,成立了专门研究 SAS 的公司,正式推出了 SAS 软件。SAS 研究公司为 SAS 软件的发展作出了很大的贡献,使 SAS 从最初的仅具有线性模型分析功能的软件发展到现在的集数据分析、数据管理、数据挖掘、数据可视化等功能于一体的统计分析系统。在国际上,SAS 是公认的最权威的统计分析软件,在科研、金融、医药、交通、通信等重要领域,SAS 都是广大用户首选的统计分析软件。

早期的 SAS 软件主要是依赖于编程解决问题,面向专业人士。目前,SAS 的版本在不断更新,提供了方便的图形界面操作,以满足更多用户的需求。本书将以较新的 SAS 9.2 版本为基础向用户介绍 SAS。SAS 9.2 版本支持多种语言,提供了中文操作环境,便于国内用户的使用。而其他版本的 SAS 用户,通过学习本书,也可以触类旁通。

1.1.2 SAS 的系统特点

SAS 系统全面涵盖各项统计分析功能,具有操作简单、功能强大、结果专业等特点。下面来一一了解。

1. 操作简单

SAS 软件操作简单,用户无须详细地了解各种统计分析具体的计算步骤,通过现成的 SAS 语句,即可方便地实现各种统计分析功能。同时,对于没有编程基础的用户,SAS 还

提供了方便的图形界面操作方式，通过鼠标操作即可完成常用的统计分析功能。

2. 功能强大

SAS 系统涵盖了常用的数据统计分析功能，包括假设测验、参数估计、描述性统计分析、方差分析、判别分析、聚类分析、时间序列分析等。同时，与一般的统计软件相比，SAS 又具有强大的数据、文件管理能力。

3. 结果专业

作为目前国际上公认的最为权威的统计软件，SAS 为用户提供了详实、专业的分析结果。用户执行相应的统计分析程序后，完整的统计分析结果将在结果输出窗口以文本的形式输出。同时，结果也可以以专业图表的形式展现。

1.1.3 SAS 的功能模块

SAS 通过模块化的方式实现各种功能，包括 BASE（基础模块）、STAT（统计模块）、ACCESS（数据库模块）、GRAPH（绘图模块）、EM（数据挖掘模块）等 30 余个模块。其中：

- 基础模块（BASE），是 SAS 系统功能实现的核心，利用该模块，用户可以实现简单的统计分析、报表制作、对数据的排序等常用的操作。
- 统计模块（STAT），是 SAS 系统中实现统计分析功能的主要模块，可以实现的功能包括方差分析、回归分析、判别分析、聚类分析、主成分分析、因子分析等。
- 绘图模块（GRAPH），是 SAS 系统中为用户提供的专门用于绘制各类统计图形的模块，可以绘制的图形包括条形图、饼图、曲线图等。
- 预测模块（ETS），可用于经济学常用的时间序列分析。
- 数据库模块（ACCESS），为用户提供了方便的主流数据库文件与 SAS 程序的接口，通过数据库模块，用户可以实现数据的导入和导出。
- 矩阵计算模块（IML），为广大科学工作者提供了矩阵计算的交互式的编程环境，通过矩阵计算模块，用户可以方便地调用其中的数学函数，完成基于矩阵计算的程序。
- 质量控制统计模块（QC）：SAS 为控制领域的质量统计分析提供了完善的统计分析模块，利用该模块，用户可以方便地进行试验设计，完成控制领域的相关图形绘制。

1.2 SAS 的安装与启动

本节将以 SAS 9.2 为例，向用户演示 SAS 软件详细的安装和启动过程。

1.2.1 SAS 的安装

SAS 9.2 的安装步骤如下：