

SAS

《软件与金融数据库》

¥ SAS ¥
RUANJIAN YU JINRONG SHUJUKU



肖枝洪 / 编 著



科学出版社

SAS 软件与金融数据库

肖枝洪 编著

科学出版社

北京

内 容 简 介

本书以 RESSET 金融数据库为数据来源,以国际流行的标准统计软件 SAS 软件为数据处理工具,向读者逐步介绍 SAS 软件进行数据分析的方法、金融数据的存取以及探索性分析方法。全书共分 11 章,首先对 SAS 软件系统、金融数据库进行介绍;然后介绍数据的导入与导出、全局通用语句、数据步语句的操作、过程步语句的操作、SAS 函数的使用、数据的可视化操作——图表制作等基础知识;最后进一步介绍比较高级的 SAS 编程技术,如宏编程方法、SQL 查询以及 IML 编程等内容。本书所有的例子都给出了用 SAS 语言编写的程序,基本都可以独立运行。

本书具有较强的实用性与操作性,可作为金融数学、金融工程、应用统计学和数据科学等专业的本科生和研究生教材,也可供相关专业人士参考。

图书在版编目(CIP)数据

SAS 软件与金融数据库 / 肖枝洪编著. —北京:科学出版社, 2017.10

ISBN 978-7-03-054587-9

I. ①S… II. ①肖… III. ①金融-数据库-统计分析-应用软件
IV. ①F83-39

中国版本图书馆 CIP 数据核字 (2017) 第 234849 号

责任编辑:李小锐 唐 梅 / 责任校对:韩雨舟

封面设计:墨创文化 / 责任印制:罗 科

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

成都锦瑞印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2017年10月第 一 版 开本:787×1092 1/16

2017年10月第一次印刷 印张:17.75

字数:420千字

定价:48.00元

(如有印装质量问题,我社负责调换)

前 言

在 21 世纪信息化时代，只要有人类生活的地方，就会有金融活动。如何把控好金融活动的规律，使之为人们的美好生活服务是金融工作者的责任与义务。

20 世纪初至今，金融行业积累的数据已经浩如烟海。如何从这些数据中探索出金融活动的规律，成为目前金融数学或金融工程研究的主要任务。对金融数据的分析与挖掘，什么软件比较合适呢？著名的 SAS 统计软件是最合适的。虽然现在流行着开源软件 R，但其软件包需要热心的研发人员自愿提供，其结果的可读性和丰富性还需改进。而 SAS 是一种标准软件，既有丰富的过程步可直接使用，也可以自己编写程序以解决特定的问题，还可以与其他软件有很多接口，而且对于海量数据也可以通过并行计算进行处理，速度比 R 软件要快很多。

本书也是后续课程“金融计算与建模”的基础。“金融计算与建模”是金融工程专业和金融数学专业学生的核心课程，是培养学生专业核心能力的关键课程，本门课程学习的好坏直接影响到“金融计算与建模”课程学习的好坏。

本书的特点体现在以金融数据库为载体，以实际问题驱动对概念、方法进行介绍，内容安排上循序渐进、逐步贯通、以点带面，达到综合训练，符合学习者的认知心理。本书的特色是按版块组合，而且每个版块通过层次化逐步深入，使读者容易接受；案例中数据主要是金融数据，而且数据量庞大，可以为读者开阔视野！

希望读者通过一节的学习进步一小点，一章的学习进步一大点，几章之后有显著的进步，整本书学完之后有较强的数据处理能力。

学习本课程还要注意，本书前面的知识可能要到后面才有所领悟，学到后面再读前面的内容又会感觉到是一种升华，是一种新的境界！

学习本门课程一定要坚持学与练紧密结合，只有上机练习才能体会到编写程序的乐趣，才能体会到一种成就感！学习理论又会使读者感觉到是一种升华！

本书为重庆理工大学“十三五规划教材”。本书主要参考了朱世武编著的《SAS 编程技术教程》以及朱世武和严玉星编著的《金融数据库》。重庆理工大学金融数学专业学生赵赞豪、熊正兵、朱云烽和陆江发等同学帮助收集了部分资料。本书的出版得到了科学出版社、重庆理工大学教务处和重庆理工大学理学院的大力支持。科学出版社唐梅编辑对书稿的认真修订和许多建设性建议使本教材在质量上有很大的提高。在此谨向以上各位表示诚挚的谢意，同时也对本书的出版给予关心和支持的同仁致以衷心的感谢！

本书不仅可以作为高等院校统计专业、金融工程专业和金融数学专业本科教材和研究

生教材，也可以作为“汇丰杯”中国高校 SAS 数据分析大赛的培训教材，还可以作为从事金融量化分析人员和数据分析人员的参考资料。

限于作者水平，书中难免存在不足之处，敬请读者批评指正！

Email: zhihongxiao@126.com

QQ: 695086564

肖枝洪

2017 年 8 月于重庆花溪

目 录

第1章 SAS 软件简介	1
1.1 SAS 的显示管理系统	1
1.1.1 工具栏	2
1.1.2 命令行窗口	3
1.1.3 SAS 资源管理窗口	3
1.1.4 增强型编辑窗口	3
1.1.5 日志窗口	3
1.1.6 输出窗口	4
1.1.7 结果窗口	4
1.2 SAS 资源管理器窗口操作	5
1.3 SAS 程序编辑窗口操作	6
1.4 SAS 功能键窗口操作	6
1.5 SAS 词段和 SAS 名称	7
1.6 SAS 语言组件	8
1.7 SAS 变量	9
1.7.1 SAS 变量类型	9
1.7.2 创建 SAS 变量	10
1.7.3 SAS 变量的输入与输出格式	11
1.7.4 SAS 系统的自动变量	12
1.8 SAS 表达式	13
1.8.1 引用 SAS 常数	13
1.8.2 SAS 算符	14
习题 1	16
第2章 金融数据库简介	18
2.1 金融数据库的起源	18
2.2 金融数据库的作用	19
2.3 金融数据库的类别	19
2.3.1 按内容和用途划分	20
2.3.2 按金融数据库的种类划分	20
2.4 中外金融数据库概况	21
2.4.1 CRSP	21

2.4.2	PACAP	21
2.4.3	聚源	22
2.4.4	锐思数据	22
2.5	金融数据库的选择标准	22
	练习 2	24
第 3 章	数据的导入导出	25
3.1	外部数据集的菜单方式导入	25
3.2	SAS 数据集的菜单方式导出	26
3.3	远程登录	27
3.4	数据步导入原始数据	27
3.4.1	简单方式输入	28
3.4.2	列方式输入	28
3.4.3	列表方式输入	29
3.4.4	格式化方式输入	31
3.4.5	命名化方式输入	32
3.5	指针控制	33
3.5.1	行列指针控制	33
3.5.2	使用行固定说明符	34
3.5.3	数据行引导语句 CARDS 与 CARDS4	38
3.6	IMPORT 过程步导入数据	38
3.7	数据步导出数据	41
3.8	过程步导出数据	44
	练习 3	45
第 4 章	数据集操作与数据管理	46
4.1	数据集语句 DATA	46
4.2	SET 语句	48
4.3	横向合并 SAS 数据集 MERGE 语句	53
4.4	BY 语句	55
4.5	删除变量语句 DROP 与保留变量语句 KEEP	56
4.5.1	DROP 语句	56
4.5.2	KEEP 语句	57
4.5.3	DROP 语句和 KEEP 语句使用规则	57
4.6	更改变量名语句 RENAME 与保留数值语句 RETAIN	58
4.7	输出外部文件语句 FILE	61
4.8	定义输入外部数据文件语句 INFILE	63
4.9	更新数据语句 UPDATE	65
4.10	修改数据语句 MODIFY	67
4.11	添加观测	72

4.12	数据集排序	73
4.13	数据集转置	74
4.14	数据集加密	76
	练习 4	78
第 5 章	DATA 步操作语句	79
5.1	基本语句——赋值与累加	79
5.1.1	赋值语句	79
5.1.2	累加语句	80
5.2	观测的选择与输出	82
5.2.1	IF 语句	82
5.2.2	WHERE 语句	83
5.2.3	WHERE 语句与 IF 语句的比较	86
5.2.4	DELETE 语句	86
5.2.5	OUTPUT 语句	87
5.3	变量属性控制	88
5.3.1	INFORMAT 语句与 FORMAT 语句	88
5.3.2	LENGTH 语句	91
5.3.3	LABEL 语句	93
5.3.4	字符自动转换为数值	94
5.4	DO 语句及循环控制	96
5.5	选择控制语句 SELECT	99
5.6	数组 (ARRAY 语句)	100
5.6.1	显式下标数组语句	100
5.6.2	引用显式下标数组元素	102
5.6.3	隐含下标数组语句	105
5.7	GOTO 语句与语句标号	108
5.8	LINK 语句	109
5.9	STOP 语句与 ABORT 语句	111
5.9.1	STOP 语句	111
5.9.2	ABORT 语句	111
5.10	REMOVE 与 REPLACE 语句	113
5.10.1	REMOVE 语句	113
5.10.2	REPLACE 语句	114
5.11	MISSING 语句	115
5.12	其他语句	116
5.12.1	LIST 语句	116
5.12.2	CALL 语句	117
5.12.3	NULL 语句	118

5.12.4	ERROR 语句	119
5.12.5	RETURN 语句	120
5.12.6	CONTINUE 语句	121
5.12.7	LEAVE 语句	122
练习 5		123
第 6 章 常用函数		125
6.1	SAS 函数定义	125
6.2	SAS 函数分类	127
6.3	日期时间函数	128
6.4	常用算术函数与数学函数	132
6.5	常用概率分布函数与分位数函数	133
6.6	多个变量的统计函数	137
6.7	随机数函数	139
6.8	随机数子程序	141
6.9	LAG 函数和 DIF 函数	143
练习 6		145
第 7 章 通用语句		146
7.1	全局通用语句	146
7.2	过程步通用语句	153
练习 7		159
第 8 章 数据可视化过程		161
8.1	制表过程 TABULATE	161
8.2	频数过程	167
8.3	作图过程 GPLOT	170
8.4	图形存储利用	176
8.5	图表过程 GCHART	178
练习 8		182
第 9 章 宏编程技术		184
9.1	宏变量	184
9.1.1	宏变量定义及引用	184
9.1.2	宏变量范围	187
9.2	宏	189
9.3	宏参数	190
9.3.1	创建宏参数及给宏参数赋值	190
9.3.2	宏调用宏	191
9.3.3	条件表达式	192
9.4	宏表达式	193
9.4.1	宏处理器对算术表达式的处理	193

9.4.2 宏处理器对逻辑表达式的处理	194
9.5 宏程序语句和宏函数	194
练习 9	196
第 10 章 SQL 编程	197
10.1 SQL 过程概要	197
10.2 SQL 查询语句	198
10.2.1 SELECT 子句	198
10.2.2 WHERE 子句	202
10.2.3 汇总函数	203
10.2.4 GROUP BY 子句	205
10.2.5 HAVING 子句	206
10.2.6 ORDER BY 对结果排序	206
10.3 创建与更新表	208
10.3.1 创建新表	208
10.3.2 在表中插入行与删除行	210
10.3.3 列修改与更新表	212
10.4 多表查询	214
10.4.1 内部 JOIN 连接查询	216
10.4.2 JOIN 外部连接查询	218
10.4.3 JOIN 连接和 MERGE 语句比较	219
10.4.4 子查询	222
10.4.5 混合子查询	222
10.4.6 合并查询 (SET 算符)	223
10.5 在 PROC SQL 过程运用宏	227
10.6 PROC SQL 选项及其作用	230
10.7 PROC SQL 用表词典查看 SAS 系统信息	233
练习 10	235
第 11 章 IML 编程技术	236
11.1 SAS /IML 语言	236
11.1.1 定义矩阵	236
11.1.2 由矩阵标识创建矩阵	237
11.1.3 PROC IML 表达式与语句类型	239
11.2 矩阵操作	243
11.2.1 创建特殊矩阵函数	243
11.2.2 使用矩阵表达式	245
11.2.3 利用行列标展现矩阵	248
11.3 IML 编程语句	249
11.3.1 各种 IML 编程语句	249

11.3.2 停止执行.....	252
11.3.3 创建和运行模块语句	253
11.4 IML 数据集操作.....	259
11.4.1 打开与编辑 SAS 数据集	259
11.4.2 由 SAS 数据集创建矩阵	261
11.4.3 由矩阵创建 SAS 数据集	262
11.5 用 IML 访问外部数据文件	264
练习 11	269
参考文献	271

第1章 SAS 软件简介

Statistical Analysis System 简称为 SAS，可用来查询数据库、分析数据和编写报告。它是美国 SAS 软件研究所的产品，在国际上被誉为标准通用软件，在我国深受医学、农林、财经、社会科学、行政管理等众多领域的专业工作者的好评；也是“汇丰杯”SAS 数据分析大赛的指定参赛软件。有关 SAS 的最新信息，可以查看 <http://www.sas.com.cn>。

SAS 采用积木式模块结构，其中的 SAS/STAT 模块是目前功能最强的多元统计分析程序集，可以做回归分析、聚类分析、判别分析、主成分分析、因子分析、典型相关分析、各种试验设计的方差分析、协方差分析以及时间序列分析。其他模块还可以进行运筹与优化分析。

建议安装 SAS 系统时同时选择简体中文和英文版安装。

1.1 SAS 的显示管理系统

启动计算机，点击 SAS 图标后，即可进入 SAS 的显示管理系统，如图 1.1 所示。

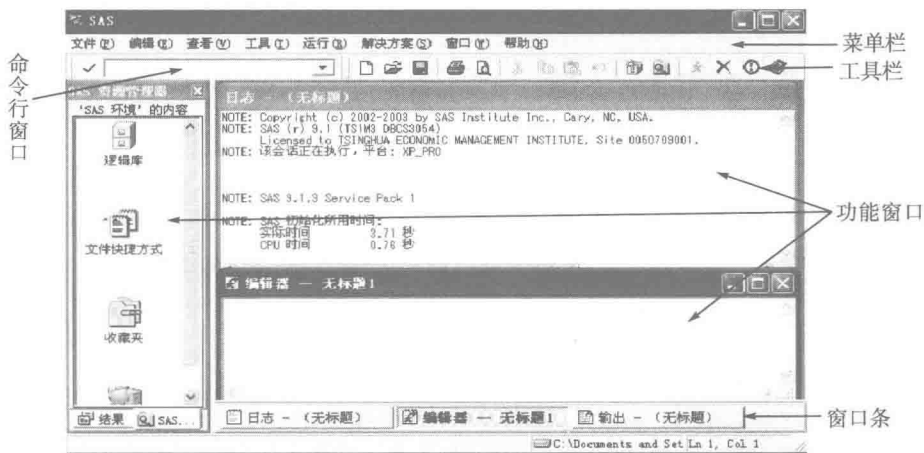


图 1.1 SAS 显示管理系统

显示管理系统的界面分为菜单栏、工具栏、命令行窗口、功能窗口(包括资源管理器窗口、日志窗口、编辑器窗口、结果窗口)4 部分。不同功能部分为 SAS 用户提供了不同功能的运行环境。

按功能键 F5、F6、F7 也可以进入(或叫激活)编辑、日志及输出口。同时按下 Ctrl+E 清除窗口内容。按 F9 可以查看所有功能键。

退出 SAS 有两种方法：①点击 File 菜单中的 Exit 命令；②点击窗口右上角的×。

1.1.1 工具栏

工具栏在菜单栏的右下方，包含一系列快捷按钮。这些按钮是一些常用的 SAS 命令，如保存、终止运行等。不同窗口状态下工具栏显示不同的图标。图 1.2 分别是资源管理器窗口和编辑器窗口的工具栏。注意：当前不可使用的图标呈现灰色色彩！



(a) 资源管理器窗口工具栏



(b) 编辑器窗口工具栏

图 1.2 工具栏

最为常用的两个工具是提交程序工具  与中断运行程序工具 。

如果直接将鼠标停放在工具栏按钮上，就会出现该按钮的功能说明。也可以点选菜单栏中“工具”，弹出“定制”下拉框，选中“定制”标签，就会出现工具定制窗口，如图 1.3 所示。单击每个按钮就可以查看其详细说明。也可通过这个窗口添加、删除工具栏中的按钮。

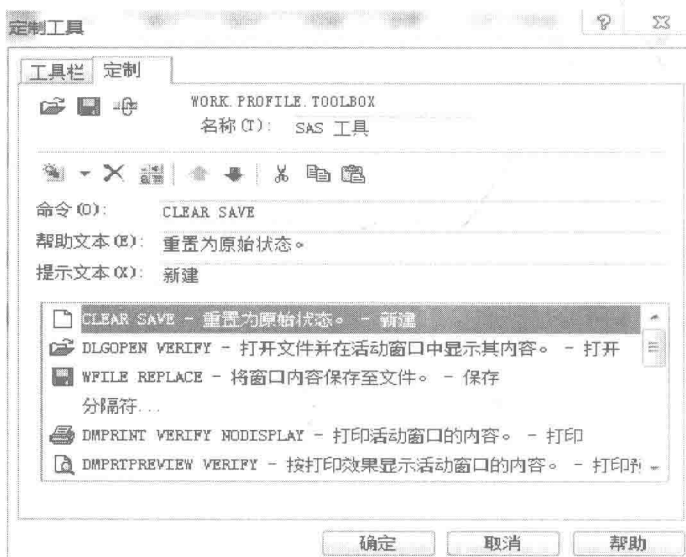


图 1.3 “定制工具”窗口

1.1.2 命令行窗口

在命令行中输入 SAS 命令可以快速完成一些 SAS 任务，如键入 Keys 可查看功能键，键入 Submit 就可提交程序等。

1.1.3 SAS 资源管理窗口

SAS 资源管理器窗口是处理数据的中心位置，数据包括目录、表(数据集)、逻辑库和主机文件数据，可以通过资源管理器窗口对 SAS 文件进行浏览。SAS 资源管理器窗口可以实现以下任务：①创建新的逻辑库和文件快捷方式；②创建新的逻辑库成员和目录条目；③打开和编辑 SAS 文件；

打开资源管理器窗口有以下 4 种方法：

- (1) 在菜单栏中选中“视图”，在弹出的下拉框中点击“SAS 资源管理器”；
- (2) 在“命令行”窗口输入“explorer”；
- (3) 在工具栏点击按钮 ；
- (4) 使用快捷方式 Ctrl+W。

1.1.4 增强型编辑窗口

在 SAS 软件中，一般使用增强型编辑窗口编写和提交程序，该窗口可以通过以下方式打开：

- (1) 选中菜单栏“视图”，再点击“增强型编辑(I)”；
- (2) 选中菜单栏“文件”，再点击“新建程序(E)”；
- (3) 单击工具栏按钮 ；
- (4) 使用快捷方式 Ctrl+N。

还可以同时打开多个增强型编辑窗口，对其程序代码可以按  建立文件名保存以备下次调用。一般，SAS 文件名不区分大小写，但在文本中要区分大小写。

1.1.5 日志窗口

日志窗口用来查看程序运行信息，可以通过日志窗口的信息来查看程序是否出错。其显示的信息如下：①提交的程序语句；②系统信息和错误提示；③程序运行速度和时间。

如图 1.4 所示，日志窗口中的信息根据以下类型呈现不同颜色：①以语句标号开头的程序行显示黑色；②以“Note”开始的显示蓝色，提示注释；③以“Warning”开始的显示绿色，提示警告；④以“Error”开始的显示红色，提示错误。



图 1.4 SAS 日志窗口

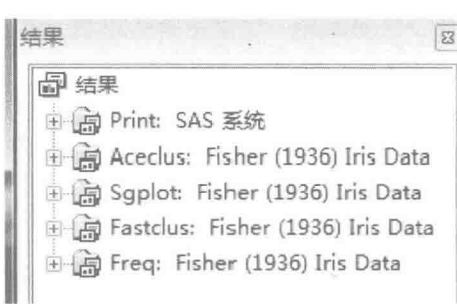


图 1.5 “结果”窗口

1.1.6 输出窗口

输出窗口用于查看 SAS 程序输出的结果。一般情况下，若有结果输出到输出窗口，则该窗口会被自动激活。若有多个输出结果则在结果窗口以树状图形式排列，单击结果表符号即可查看相应的输出。运行下列程序可以查看股票 stock000001 和 stock000002 的记录情况。

```
data qtttn;set Sasfindt.stk000001 Sasfindt.stk000002;  
proc print data=qtttn(obs=10);run;
```

1.1.7 结果窗口

可以通过结果窗口对程序的输出结果进行浏览和管理。结果窗口有三种呈现方式，SAS 系统默认为树视图。在树视图下可以通过单击“+”号节点来展开每个过程语句产生的结果文件，如图 1.5 所示。

SAS 可以用一种或多种格式(或类型)产生输出。默认类型是“列表”输出，其他输出类型包括 HTML、“输出数据集”和 PostScript。

如要设置输出类型，在菜单栏中点击“工具”下拉框，点击“选项”下拉框，点击“参数选择”下拉框，再点击“结果”标签，进入参数选择窗口进行设置，如图 1.6 所示。



图 1.6 输出结果参数选择

再次运行下列程序可以查看股票 stock000001 和 stock000002 的记录情况。

```
data qtn;set Sasfindt.stk000001 Sasfindt.stk000002;  
proc print data=qtn(obs=5);run;
```

1.2 SAS 资源管理器窗口操作

SAS 逻辑库由一组 SAS 文件组成，而 SAS 文件就是由 SAS 创建、存储、管理的文件，例如 SAS 数据集、SAS 目录册等。所有的 SAS 文件都保存在 SAS 逻辑库中。SAS 软件系统的信息组织分两层，第一层为 SAS 逻辑库，第二层是 SAS 文件。SAS 逻辑库是一个逻辑概念，其本身不是物理实体，对应的实体是操作系统下一个文件夹或几个文件夹中的一组 SAS 文件，如图 1.7 所示。

1. 新建逻辑库

如果想要在 E 盘建立逻辑库 Sasfindt，首先在 E 盘新建文件夹“Sasfindt”。然后可以采用以下两种方式进入如图 1.8 所示的“新建逻辑库”对话框。

- (1) 在逻辑库级别空白处右击，然后选择“新建”；
- (2) 在菜单栏选择“工具”，点击“新建逻辑库”。



图 1.7 SAS 逻辑库

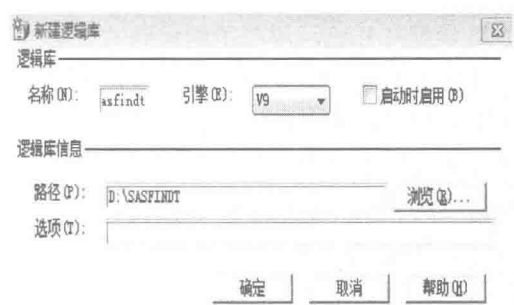


图 1.8 “新建逻辑库”对话框

在“名称”中输入新的逻辑库名称，例如“Sasfindt”。逻辑库名称一般不超过 8 个字符。在引擎中根据数据来源选择不同的引擎，如果只是想建立本地机地址上的一个普通 SAS 数据库，可以选择默认。然后选中“启动时启用”复选框，在逻辑库信息中，单击路径后面的“浏览”按钮，选择与这个逻辑库相对应的物理地址。“选项窗口”可以不选，单击“确定”产生一个新的逻辑库，例如 Sasfindt。这个逻辑库在 SAS 资源管理器中可以查看得到。也可以用 Libname 语句创建 SAS 逻辑库。其语句格式为：

```
LIBNAME libref <engine> 'SAS-DATA-Library';
```

其中，libref 为逻辑库名；SAS-DATA-Library 为对应的物理地址。
例如创建 SASFINDT 逻辑库。在增强型编辑窗口输入以下代码并提交运行：
Libname SASFINDT 'e:\sasfindt';
还可以多个文件夹创建同一个逻辑库。例如：

```
Libname sasfindt ("d:\sasfindt1",'e:\sasfindt2');
```

2. 复制粘贴数据表

如果想要将一个逻辑库的数据表或文件复制到另一个逻辑库中,只需在资源管理器中进入源逻辑库选择对应的数据表或文件,按 **Ctrl+C** 组合键,然后进入目标逻辑库中,再按 **Ctrl+V** 组合键就可以达到目的。

3. 临时库和永久库

临时库是指它的内容在 SAS 启动后才存在,退出时其内容完全被删除的逻辑库。SAS 系统默认的临时逻辑库为 **WORK**。永久逻辑库是指其内容在 SAS 系统关闭后仍保留的逻辑库。SAS 系统自动指定的逻辑库 **SASHELP**、**SASUSER** 和 **MAPS** 都是永久逻辑库。

4. 引用 SAS 文件

引用临时库 **WORK** 中文件时可以直接使用文件名,例如在增强型编辑窗口中输入 “**data=class**” 和 “**data=work.class**” 具有相同的作用。引用非临时库的 SAS 文件时必须使用两级命名方式,例如在增强型编辑窗口中输入 “**data=sashelp.class**”。

1.3 SAS 程序编辑窗口操作

1. 提交程序

如果要想打开已经保存了的代码,可以在菜单栏“文件”中选中“打开程序”,在弹出的物理地址中选中所要的程序文件,在增强型编辑窗口就会出现程序代码。也可以直接在增强型编辑窗口输入代码,或者用其他文本编写的代码复制到此编辑器中。注意标点符号需要为英文标点符号!

程序编写完后,按以下方式提交程序,程序就会被执行:

- (1) 按 **F3** 或 **F8**;
- (2) 单击工具栏中的提交按钮 ;
- (3) 在命令行窗口输入 “**submit**” 后,回车;
- (4) 在菜单栏中选中“运行(R)”,点击“提交 S”。

如果只想运行某几行,可以选中该行,按照上述任何一个方式提交即可运行。程序运行产生的信息展示在日志窗口;部分会展示在输出窗口和结果窗口中。

2. 保存程序

程序可以以其他格式的文本保存,也可以直接以编辑器的方式保存,但打开时必须在装有 SAS 软件的电脑上才能打开。

1.4 SAS 功能键窗口操作

在 SAS 软件中,一些键或组合键已经被赋予特定功能,这就是功能键。功能键可以快速执行一些常用的 SAS 命令,如程序的提交和查看 SAS 帮助文档等。