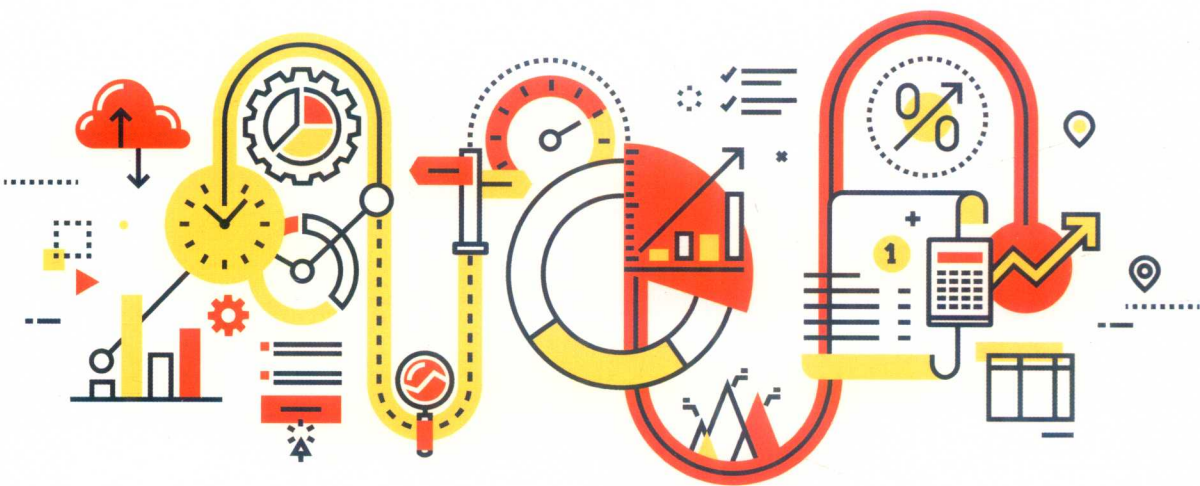




# 数据分析与量化投资

## 基于SAS的应用

林煜恩◎著

实战性强，帮你解决实际问题，快速完成工作，感受SAS数据分析的力量



中国工信出版集团



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

# 数据分析与量化投资

## 基于SAS的应用

林煜恩◎著

電子工業出版社

Publishing House of Electronics Industry

北京•BEIJING

## 内 容 简 介

本书内容基于 SAS EG 平台,采用项目管理的过程流方式,介绍数据分析及量化投资策略。书中提供了作者编写的 autoexec 命令文件,可以让 SAS 在运行时就自带实用的宏语句,读者不需要编写复杂的语句,只要按照书中的指导步骤,就可以轻松地完成数据分析,对投资策略效果进行检验,并且可以将所有的数据表格结果输出到 Word 文件中。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有,侵权必究。

## 图书在版编目(CIP)数据

数据分析与量化投资:基于 SAS 的应用 / 林煜恩著. —北京:电子工业出版社, 2019.10  
ISBN 978-7-121-37320-6

I. ①数… II. ①林… III. ①统计分析—应用软件 IV. ①C819

中国版本图书馆 CIP 数据核字(2019)第 187239 号

责任编辑:黄爱萍

印 刷:三河市良远印务有限公司

装 订:三河市良远印务有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

开 本:720×1000 1/16 印张:15.5 字数:252 千字

版 次:2019 年 10 月第 1 版

印 次:2019 年 10 月第 1 次印刷

定 价:69.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888, 88258888。

质量投诉请发邮件至 [zltts@phei.com.cn](mailto:zltts@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式:(010) 51260888-819, [faq@phei.com.cn](mailto:faq@phei.com.cn)。

# 前 言

自 2015 年入职吉林大学以来，已经过了 4 个年头，在这 4 年里我承担了研究方法的课程，在讲授研究方法内容时出现了一些难题，如果仅仅讲授研究方法的内容，那么部分学生可能无法了解这些内容是如何实现的。我想这也是一些承担研究方法课程的老师面临的难题，主要是在课程设计中，我们无法同时让学生学习程序语言和掌握研究方法材料，这启发我对 SAS 编程进行进一步的思考，究竟是让学生掌握初级的 SAS 语言，让他们由基础慢慢进步，还是直接将 SAS 编程提升到新的高度，以宏语法为主，让他们了解我所编写的宏语法，只要掌握这些宏语法，就能够完成实证。

在这样的构想之下，从 2017 年开始，我开发了 `autoexec` 程序，并将多年来编写（与收集）的宏语法都打包进去，同时修正了《SAS 在财务研究中的应用》书中的一些宏语法，并且让这些宏语法可以在 SAS 9.4 M3 中运行，也使用了 SAS EG 平台进行分析，让不熟悉编程的读者可以利用类似于 SPSS 的点击功能去完成统计分析，更重要的是可以调用本书中提供的宏语法快速进行分析。

本书主要内容如下：

第 1 章介绍 SAS EG 模块的基础界面，并阐述 SAS EG 的过程流与项目管理；第 2 章介绍自动运行宏文件，以弥补 SAS 在菜单功能上的不足，并且介绍如何将程序运行结果输出为 Word 文档、在 SAS 中发送电子邮件的命令功能，以及 SAS EG 的排程设计。

第 3 章介绍本书后续章节中所要使用的数据，包括报酬率数据、财务

报表数据和公司治理数据；第 4 章介绍数据处理的宏语法，分别是变量转换宏、移动窗口转换宏、领先/滞后宏、缩尾宏、缺失值处理宏以及行业变量计算宏。这 6 种宏在进行数据分析时是最常遇到的，因此本书特别将其写成标准化的宏，让读者了解其用法，以便在日后可以随时随地调用。

第 5 章探讨关于基本投资策略的绩效；第 6 章介绍 Lewellen 投资策略，一次性考虑多个投资变量，并借此建立投资组合；第 7 章介绍能够获利的投资组合。

第 8 章介绍描述性统计宏，可以生成实证研究中的描述性统计表和相关系数表；第 9 章介绍样本分配趋势宏、两群体差异性检验宏和组合差异性检验宏；第 10 章介绍回归宏语法架构，在设置中可以运行 5 种基础回归方法，以及使用回归宏语法；第 11 章综合了第 8~10 章的内容，设计出基础论文宏、中介论文宏和调节论文宏，通过运行论文宏，可以直接生成实证论文中的基础表格，这对于研究分析者而言，可以节省很多制表的时间。

第 12 章讲述解释变量内生性的解决之道，第 13 章讲述样本选择性问题。

读者可以在 <http://www.broadview.com.cn/37320> 处下载本书提供的资源文件。

林煜恩

2019 年 6 月 30 日于吉林大学匡亚明楼

# 目 录

## 第 1 篇 基础篇

|                              |    |
|------------------------------|----|
| 第 1 章 SAS EG 基础界面 .....      | 2  |
| 1.1 SAS EG 菜单 .....          | 2  |
| 1.2 过程流与项目管理 .....           | 6  |
| 1.3 小结 .....                 | 13 |
| 第 2 章 数据分析小工具 .....          | 14 |
| 2.1 SAS 自带宏语法与本书提供的自带宏 ..... | 15 |
| 2.2 直接输出成 Word 的报表 .....     | 20 |
| 2.3 E-mail 功能 .....          | 23 |
| 2.4 每天都要做的工作——排程 .....       | 28 |
| 2.5 小结 .....                 | 32 |

## 第 2 篇 数据处理篇

|                   |    |
|-------------------|----|
| 第 3 章 数据说明 .....  | 34 |
| 3.1 股票报酬率数据 ..... | 34 |
| 3.2 财务报表数据 .....  | 37 |
| 3.3 公司治理数据 .....  | 40 |
| 3.4 小结 .....      | 41 |
| 第 4 章 数据处理宏 ..... | 42 |
| 4.1 变量转换宏 .....   | 42 |

|     |                  |    |
|-----|------------------|----|
| 4.2 | 移动窗口描述性统计宏 ..... | 47 |
| 4.3 | 领先/滞后宏 .....     | 52 |
| 4.4 | 神奇的缩尾宏 .....     | 55 |
| 4.5 | 数据缺失处理 .....     | 57 |
| 4.6 | 行业变量的计算 .....    | 61 |
| 4.7 | 小结 .....         | 64 |

## 第 3 篇 量化投资篇

|       |                     |     |
|-------|---------------------|-----|
| 第 5 章 | 常见的投资策略 .....       | 66  |
| 5.1   | 规模投资策略 .....        | 66  |
| 5.2   | 价值投资策略 .....        | 75  |
| 5.3   | 动能投资策略 .....        | 81  |
| 5.4   | 定投 .....            | 91  |
| 5.5   | 小结 .....            | 94  |
| 第 6 章 | Lewellen 投资策略 ..... | 95  |
| 6.1   | 二维投资策略 .....        | 95  |
| 6.2   | 化腐朽为神奇的预测报酬率 .....  | 101 |
| 6.3   | 小结 .....            | 106 |
| 第 7 章 | 定价模型和风险套利模型 .....   | 107 |
| 7.1   | 定价模型 .....          | 107 |
| 7.2   | 风险套利模型 .....        | 118 |
| 7.3   | 小结 .....            | 126 |

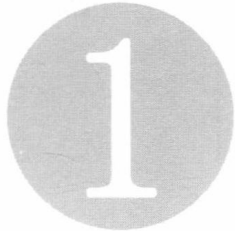
## 第 4 篇 基础实证篇

|       |                  |     |
|-------|------------------|-----|
| 第 8 章 | 描述性统计与相关分析 ..... | 129 |
| 8.1   | 描述性统计表 .....     | 129 |
| 8.2   | 相关分析 .....       | 134 |
| 8.3   | 描述性统计宏 .....     | 137 |
| 8.4   | 小结 .....         | 141 |

|                             |     |
|-----------------------------|-----|
| 第 9 章 样本分配趋势与组合差异性检验 .....  | 142 |
| 9.1 样本分配趋势 .....            | 142 |
| 9.2 两群体差异性检验 .....          | 147 |
| 9.3 组合差异性检验 .....           | 148 |
| 9.4 小结 .....                | 153 |
| 第 10 章 回归实证 .....           | 154 |
| 10.1 辨认被解释变量 .....          | 154 |
| 10.2 以股利支付进行回归 .....        | 161 |
| 10.3 交互作用模型 .....           | 173 |
| 10.4 中介回归宏 .....            | 178 |
| 10.5 小结 .....               | 185 |
| 第 11 章 论文实证 .....           | 186 |
| 11.1 基础论文宏 .....            | 186 |
| 11.2 中介论文宏 .....            | 195 |
| 11.3 调节论文宏 .....            | 202 |
| 11.4 小结 .....               | 207 |
| <br><b>第 5 篇 内生性检验篇</b><br> |     |
| 第 12 章 解释变量内生性的解决之道 .....   | 210 |
| 12.1 连续型变量的解决方案 .....       | 211 |
| 12.2 二元变量的解决方案 .....        | 218 |
| 12.3 小结 .....               | 223 |
| 第 13 章 样本选择性问题 .....        | 225 |
| 13.1 倾向得分匹配 .....           | 225 |
| 13.2 双重差分法 .....            | 231 |
| 13.3 小结 .....               | 240 |

# 第 1 篇

## 基础篇



1

在大数据时代，越来越多的工作需要进行分析，而在日常数据分析中，往往进行的都是重复性工作，因此很多人使用 Excel 或者 SPSS 通过菜单操作来进行数据分析。然而，Excel 或者 SPSS 的使用虽然方便，但是其缺少记录操作过程的功能。而在 SAS EG 模块中，所有操作过程都会被记录下来，并且可以反复运行该操作过程，因此只要进行过一次操作，就可以以此为标准来重复运行，并且也能够让新进入的工作伙伴直接使用该操作过程。鉴于此，本书就以 SAS EG 为主来进行数据分析的介绍。

但是对数据进行分析并不是我们的真正工作，我们的真正工作是能够在 Word 文档中呈现表格，并且即时发送给所需要的人。因此，在基础篇中除介绍宏语法外，还会介绍将 SAS 的结果输出为 Word 文档，并且利用 SAS 编程将结果直接发送给所需要的人。

本篇的第 1 章介绍 SAS EG 模块的基础界面，并阐述 SAS EG 的流程与项目式管理；第 2 章介绍自动运行宏文件，以弥补 SAS 在菜单功能上的不足，并且介绍如何将程序运行结果输出为 Word 文档、在 SAS 中发送电子邮件的命令功能，以及 SAS EG 的排程设计。

# 1

## 第 1 章

# SAS EG 基础界面

首先介绍 SAS EG 模块的主要菜单功能,以便让用户能够顺利操作 SAS EG。然后介绍过程流与项目式管理,协助用户重复运用项目进行日常分析工作。

本书使用的是 SAS 9.4 中的 EG 7.1,有些功能仅在 EG 7.1 版本以上才能使用,若读者的 SAS 版本过低,则请更新。

### 1.1 SAS EG菜单

SAS EG 是 SAS Enterprise Guide 的简称,即 SAS 企业向导模块,其强大的功能对企业和科研人员都有很大的帮助作用。下面介绍 SAS EG 模块的启动界面,如图 1-1 所示。

从图 1-1 中可以看到 SAS EG 有文件、编辑、视图、任务、程序、工具以及帮助菜单。利用 SAS EG 的“文件”菜单中的功能选项,我们可以进行文件的编辑修改和数据的输入,即它纳入了 Office 中的 Word 和 Excel 的功能。

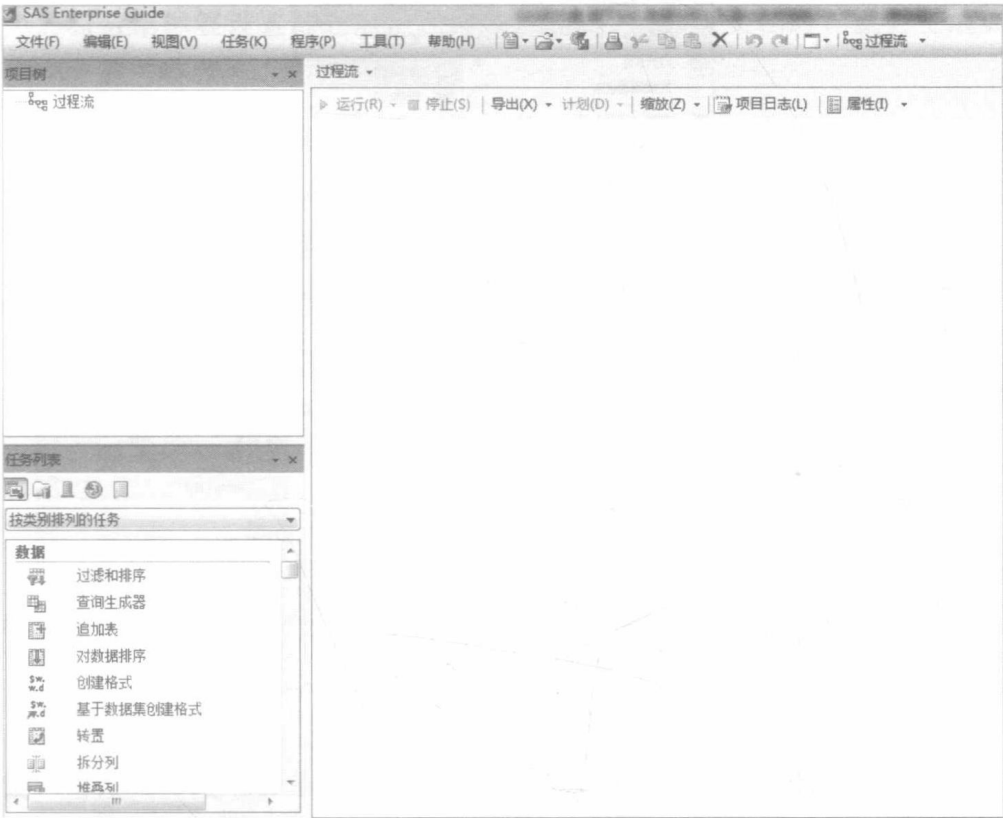


图 1-1 SAS EG 模块的启动界面

如图 1-2 所示，可以使用 SAS EG 模块的报表功能来撰写文本内容。SAS EG 模块可以将数据分析和出具报告的文本撰写功能结合在一起。

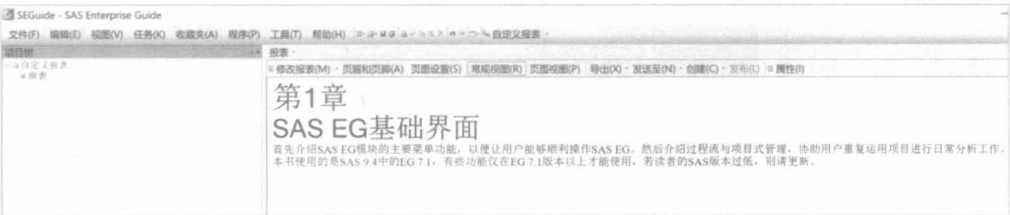


图 1-2 使用 SAS EG 的报表功能

在“文件”菜单中还有“数据”功能选项，选择该选项可以进行数据的输入，如图 1-3 所示。



图 1-3 手动输入数据的生成样式

在“新建数据”对话框中，我们可以以类似于 Excel 的方式来输入数据。需要注意的是，数据名称应为英文，不能输入中文名称，但标签可以输入中文名称，如图 1-4 所示。

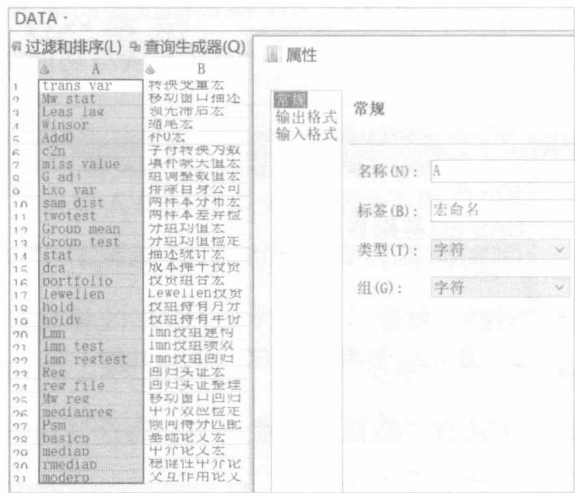


图 1-4 数据标签设置

将数据标签设置好之后，在将数据输出为 Word 文档时，就会生成如图 1-5 所示的形式。

如图 1-5 所示的信息也是本书后面要介绍的宏文件的各项说明。但由于目前大家的使用习惯仍然以 Excel 输入为主，因此本书对于数据的输入并不多加介绍，而是着重于介绍 SAS 格式数据，以及如何将外部记事本和 Excel 中的内容导入 SAS 中并进行分析。

接下来，我们介绍 SAS EG 的“任务”菜单功能。打开“任务”菜单，显示如图 1-6 所示。

| 宏命名        | 宏说明    |
|------------|--------|
| trans_var  | 转换变量宏  |
| Mw_stat    | 移动窗口描述 |
| Leas_lag   | 领先滞后宏  |
| Winsor     | 缩尾宏    |
| Add0       | 补0宏    |
| c2n        | 字符转换为数 |
| miss_value | 填补缺失值宏 |
| G_adj      | 组调整数值宏 |
| Exo_var    | 排除自身公司 |
| sam_dist   | 两样本分布宏 |
| twotest    | 两样本差异检 |
| Group_mean | 分组均值宏  |
| Group_test | 分组均值检定 |

图 1-5 输出数据形式



图 1-6 SAS EG 的“任务”菜单

我们发现，SAS EG 的“任务”菜单的功能与 SPSS 的“分析”菜单的功能类似，如图 1-7 所示。

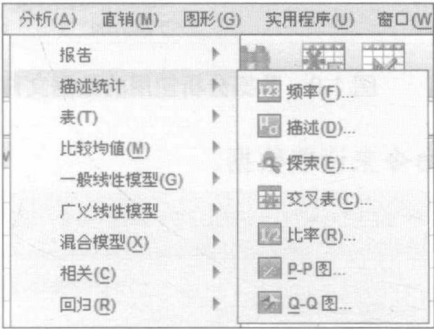
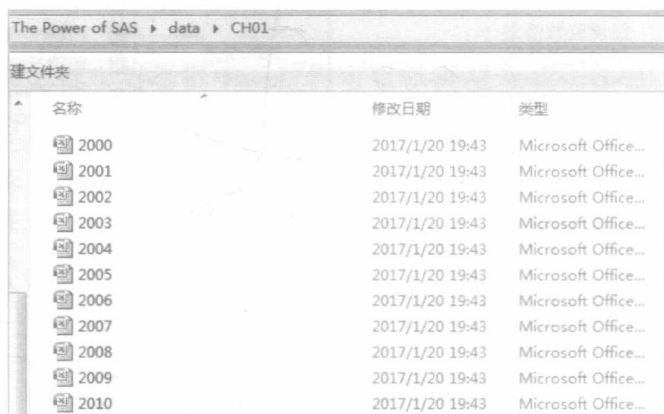


图 1-7 SPSS 的“分析”菜单

因此，本书最主要的内容不是教用户如何利用 SAS EG 来进行通过 SPSS 可以完成的操作，而是主要介绍在 SAS EG 中，结合项目管理的过程流以及笔者所撰写的宏语法形成的“标准化”数据分析，这是 SAS EG 与 SPSS 最大的差异点。SAS EG 能够记录操作过程的过程流，这些过程流只要运用得当，就能够被用户反复利用，并且可以让不同的人一起使用，以提高日常工作的效率。

## 1.2 过程流与项目管理

本节主要介绍标准化的 SAS EG 管理。此处我们使用“01-数据分析的项目管理.egp”数据，数据文件位于“D:\The Power of SAS\data\CH01”中（说明：读者需要根据自己文件的实际存储位置来修改路径），其下共有 11 个文件，如图 1-8 所示。



| 名称   | 修改日期            | 类型                  |
|------|-----------------|---------------------|
| 2000 | 2017/1/20 19:43 | Microsoft Office... |
| 2001 | 2017/1/20 19:43 | Microsoft Office... |
| 2002 | 2017/1/20 19:43 | Microsoft Office... |
| 2003 | 2017/1/20 19:43 | Microsoft Office... |
| 2004 | 2017/1/20 19:43 | Microsoft Office... |
| 2005 | 2017/1/20 19:43 | Microsoft Office... |
| 2006 | 2017/1/20 19:43 | Microsoft Office... |
| 2007 | 2017/1/20 19:43 | Microsoft Office... |
| 2008 | 2017/1/20 19:43 | Microsoft Office... |
| 2009 | 2017/1/20 19:43 | Microsoft Office... |
| 2010 | 2017/1/20 19:43 | Microsoft Office... |

图 1-8 数据分析使用的数据文件

我们采用下列命令来读取数据。

```
proc import datafile="D:\The Power of SAS\data\CH01\2000"
    dbms=xlsx
    out=anadata
    replace;

quit;
```

将“2000”这个文件的数据导入 SAS EG 中，并命名为“anadata”。导入数据后，在 SAS EG 中将呈现如图 1-9 所示的数据形态。

|    | Stkcd  | year | month | value        | ret          | turnover     |
|----|--------|------|-------|--------------|--------------|--------------|
| 1  | 000001 | 2001 | 1     | 1012004197   | 0.0316792562 | 0.0484931603 |
| 2  | 000001 | 2001 | 2     | 518925892.01 | -0.059411345 | 0.026436523  |
| 3  | 000001 | 2001 | 3     | 5903546059.9 | 0.1511729626 | 0.2612595015 |
| 4  | 000001 | 2001 | 4     | 2778050117.7 | -0.042540727 | 0.1284040164 |
| 5  | 000001 | 2001 | 5     | 1689923878.1 | 0.0431419685 | 0.0748793586 |
| 6  | 000001 | 2001 | 6     | 1382545628.5 | -0.054935084 | 0.064820792  |
| 7  | 000001 | 2001 | 7     | 1081832104.3 | -0.096669098 | 0.0561497301 |
| 8  | 000001 | 2001 | 8     | 736130115.07 | -0.080965136 | 0.041573733  |
| 9  | 000001 | 2001 | 9     | 608094869.51 | 0.0031475216 | 0.0342350639 |
| 10 | 000001 | 2001 | 10    | 1880120562.7 | 0.0831360589 | 0.0965982978 |
| 11 | 000001 | 2001 | 11    | 755405019.23 | -0.028239853 | 0.0399397005 |
| 12 | 000001 | 2001 | 12    | 662191325.79 | -0.087181562 | 0.0395136872 |

图 1-9 导入 SAS EG 后的数据形态

从图 1-9 中可以看到导入数据后的变量名称，分别为股票代码(Stkcd)、年份(year)、月份(month)、成交值(value)、报酬率(ret)以及换手率(turnover)。

我们先将成交值取自然对数，这需要使用查询生成器的功能。单击查询生成器，显示界面如图 1-10 所示。



图 1-10 查询生成器界面

选择所有相关变量，然后单击“计算列”，打开“新建计算列”对话框，选择类型，如图 1-11 所示。

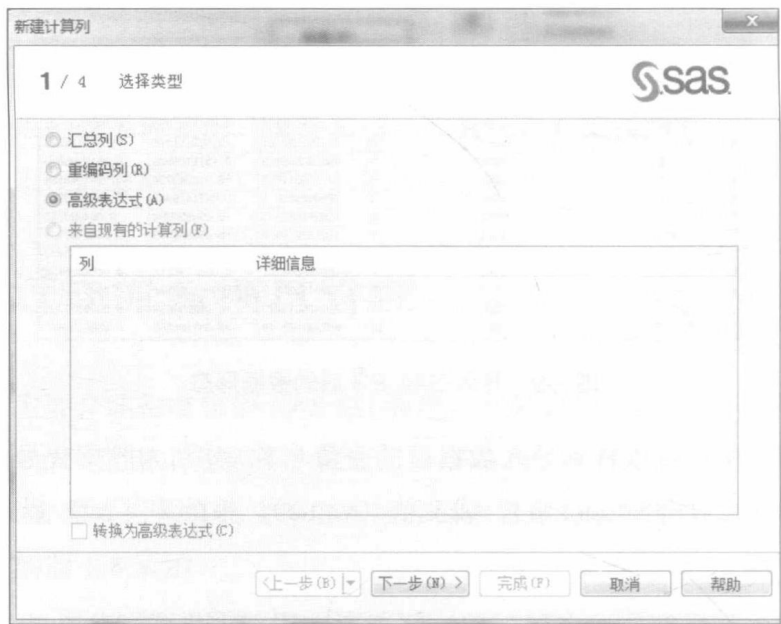


图 1-11 选择类型

由于我们的目标是针对变量做计算，所以这里选择“高级表达式”。单击“下一步”按钮，生成高级表达式，如图 1-12 所示。



图 1-12 生成高级表达式

在“输入表达式”框中，输入 `log(value)` 命令，在函数的类别中可以搜索到 SAS EG 中的所有相关函数。单击“下一步”按钮，修改其他选项，如图 1-13 所示。



图 1-13 修改其他选项

输入计算之后的变量名称和标签，单击“下一步”按钮。属性汇总如图 1-14 所示。



图 1-14 属性汇总