



SAS

林煜恩◎著

在财务研究中的应用

深入讲解SAS数据处理和统计分析，知识全面、系统，实战性强，
科学安排内容的层次架构，由浅入深，循序渐进，适合读者的学习规律。



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>

SAS

在财务研究中的应用

林煜恩 / 著

電子工業出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

本书分为两部分,第一部分为第1章到第7章,主要针对财务管理、金融、投资领域需要用到SAS语法进行介绍;第二部分为第8章到第17章,主要针对财务实践研究中会使用到的研究方法进行介绍。

本书的第二部分中提供了标准化的宏语法,在第8章中提供了stat以及correlation两个宏语法,读者仅需将文档准备好,输入文档名称,希望进位到小数点后几位以及要分析的变项名称,宏语法就能够直接将叙述统计表以及相关系数表自动生成出来,第9章提供了npar及ttest两个宏语法,可以快速地进行两群体的中位数及均值检定,第10章提供了Jagaseesh and Titman动能投资策略的语法,而第11章到第15章阐述了各个回归模型,以及相对应的宏语言,提供给读者进行实践分析,第16章则针对短期及长期事件研究法做了详细的介绍,最后在第17章则介绍了如何使用财务数据构建出Fama and French的5因子报酬率。

本书适用于高等院校财经类专业本科生以及硕士研究生,它提供了撰写学科论文相当好用的宏语法;而对于博士研究生而言,本书撰写的宏语法逻辑,也可以提供其在进行更进一步的模型分析时的撰写架构。另外,本书对其他行业从事数据分析的工作人员也有一定的参考价值。

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

SAS在财务研究中的应用 / 林煜恩著. —北京:电子工业出版社, 2017.1

ISBN 978-7-121-30225-1

I. ①S… II. ①林… III. ①统计分析—应用软件—应用—财务管理 IV. ①C819 ②F275-39

中国版本图书馆CIP数据核字(2016)第258355号

策划编辑:师纬风

责任编辑:徐津平

特约编辑:顾慧芳

印 刷:北京天宇星印刷厂

装 订:北京天宇星印刷厂

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:787×980 1/16 印张:23 字数:510千字

版 次:2017年1月第1版

印 次:2017年1月第1次印刷

印 数:3000册 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 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式:010-51260888-819 faq@phei.com.cn。

推荐序一

能得天下英才而教之是身为老师最感到快乐的事情，这也是十多年来，我的学生林煜恩给我的感受。我是他硕士班以及博士班的指导老师，直到他取得博士学位之后，我们仍持续进行学术研究。如今得知煜恩要出书了，身为老师的我真的为他感到高兴。

煜恩在硕士班的时候就崭露了对于撰写 SAS 程序的天赋与热诚。让我印象最深刻的是他曾经为了写出好的程序，连续花了六个小时的时间待在计算机屏幕前仔细看着程序运行中的 LOG 文件，并藉此观察出 SAS 速度越变越慢的原因，逐步改进，这样的努力让他从一个不会写程序的门外汉变成一个热爱写程序、并能自行写程序来解决问题的专家。现在的煜恩不仅擅长 SAS 指令的撰写，亦将其这些年来的 SAS 撰写经验编纂成书，他将 SAS 程序以简驭繁，将过去复杂的研究语法变得简洁方便，且可以让 SAS 初学者能快速上手。

事实上这本书最早是煜恩为了让学弟学妹可以更容易使用 SAS 软件，而把自己研究 SAS 的心得写成的秘籍。还记得煜恩将本书第一版的书名取为《恶魔党私传秘籍》。那时我看到这本秘籍，真的很开心，因为这是一本学长对学弟学妹的关爱，也是煜恩为师门留下的宝贵资产。这几年来，我指导的学生们（也是煜恩的学弟学妹们）在研究时遭遇到的问题，煜恩也不吝给予帮助，同时也修正并更新了他的秘籍。

煜恩曾经开授过财务软件应用（SAS）的课程，他的教学态度认真积极，与学生相处融洽，学生们非常喜欢向他请教 SAS 问题。许多学生第一次跑程序，对于程序往往很彷徨，煜恩非常了解初学者常犯的错误及不懂之处，因此他所撰写的秘籍可以让初学者更简单地认识 SAS，不再害怕写程序，深获学生们的好评。

在这样的情况下，经过煜恩不断的研究、修改与更新，这本秘籍逐渐形成为各位读者眼前的这本书，字里行间所透露的，是煜恩对后进入者的叮咛与指导，也是他送给我的最佳礼物。

我在此向希望踏入学术界的初学者们强烈推荐《SAS 在财务研究中的应用》这本书。这本书由浅入深，一步一步带领着读者进入 SAS 这个领域，指引着读者在论文实证上使用更好、更有帮助的程序代码。本书清楚地写出了每个程序代码的含义以及使用上常犯的错误，并搭配着程序跑出的结果，让读者更容易看懂程序语言。若能跟着作者引导的步骤一步一步地尝试，相信你们很快就会使用 SAS，甚至可以成为 SAS 达人，使你们的研究更为顺利。在此，我也强烈建议已经踏入学术界已有写程序经验、想另外学习 SAS 的学者们将此书作为一本工具书，当你所写的程序跑得太慢或者遇到瓶颈时，相信这本书会提供许多的帮助。

池祥萱

台湾东华大学财务金融学系教授

推荐序二

撰写论文主要是在学术理论及实证研究中的新发现，现今学术领域研究的过程中，由于统计软件、绘图软件等工具可使用，使得学术领域发展更为迅速，产出更多。

撰写一篇好的论文，除了有好的想法之外，还需要搜集整理资料、实证研究及稳健性检验。研究发展刚开始时，资料搜集较为困难，且资料量较少；到现今有许多数据库提供给研究人员使用，数据收集简单许多但数据量越来越庞大，使得数据的整理及合并数据库较为困难，因此需要有好的软件协助。当数据整理完成后，紧接着实证研究的部分，需选择适当的统计模型且正确地计算出结果，此时依然需要好的软件功能才能快速且正确地完成。而 SAS 在财务数据处理的功能上，是其他统计软件无法比拟的。但当前教育课程的设计，财务专业的学生虽然具备足够的财务知识，可是对各种财务软件的特性却是一知半解，更缺乏处理数据的能力；而当前有关 SAS 软件的书籍，多是针对生物统计、市场营销领域撰写的，鲜少有专门为财务专业背景而撰写的 SAS 教材；而针对财务专业的 SAS 教材中，其程序过于偏向传统指令的介绍，对于进阶编程技巧较少涉猎，导致运行效能较差，应用价值不高。

我很高兴得知煜恩针对财务专业撰写这本教科书，它突破过去教材传统思维，提供了不同的视角，确实与应用紧密结合。这本教科书可成功指引本科生以及研究生在论文实证上使用更好且更正确的程序语言，并且搭配 SAS 所提供的功能，将软件使用得淋漓尽致。本书清楚地写出了财务上经常使用的 SAS 程序语言，并且介绍各种财务模型及投资组合策略的程序语言，这样的分析手法使程序语言更为清楚，并且提供许多统计模型及宏语法，可以快速地整理出论文所需要的表格，也有许多细节部分也说明得非常仔细，让初学者不再害怕程序语言，使资深的财务人员节省许多时间，我相信读者熟读这本书之后，可以感觉到作者的用心。

在这本《SAS 在财务研究中的应用》中，煜恩从许多不同的角度，分析程序语言中过去没有想过的方式。通过他的诠释，读者会发现许多复杂的程序语言，变得简单而且容易了解，同时也凸显了各种统计模型以及形成投资组合的差异。读者往往会通过那些投资组合 SAS 之模型，验证自己所学的理论模型。更难得的是，借助这本教科书，读者可非常简易地通过参数的修改，进行更多不同的研究或形成不同的投资组合，事半功倍。相信经过本教科书的介绍，读者可以自行发挥，通过自我学习的过程，获得更多、更广泛的专业知识。

蕭朝興

台湾东华大学财务金融学系教授

Tel: +886-38633135

Fax: +886-8633130

Email: cschiao@mail.ndhu.edu.tw

前言

12年前，笔者于台湾东华大学攻读企业管理系硕士班，当时完全没有编程基础就必须学习动能投资策略（见本书10.3节与10.4节）的语法，所以就一股脑儿一行一行运行程序，一行一行修改，在暑假期间，每天盯着电脑的日志文件六个钟头，与同行见面讲的话就是，“你的程序跑得如何？”当时觉得研究财务金融的人好苦，为了做好财务金融研究，不得不去接触从大学到研究生前期都不会接触的统计软件，因为是研究生，所以这些东西要自己学；因为要完成论文，这些东西拼死拼活也得学好，也因此开始了我学SAS的路。

10年前开始攻读博士学位后，要引领硕士生学习SAS，也因此要接触更多的研究方法，更多的研究语言，当时很多语法都是20世纪90年代的前人留下来的语法，程序确实能够顺利运行，但是问题是运行速度慢。跟我合作的老师问，“煜恩，为什么程序运行得这么慢，却没有人要去改善呢？”，当时我没有答案，但当年我指导一名硕士生，请他把程序改好，使运行结果会比较快时，他这么回答：“学长，我没有时间去改程序，我还是会把结果跑出来，自己再动手转到Word感觉会比较快。”于是我得到答案了，大家都知道运行慢，可是没人有勇气去修改这些程序，因为不知道会花多少时间才能修改完，而就算完成以后速度变快，自己也毕业了，以后也用不上了，所以如何能期待硕士生修改这些语法呢？

于是我开始撰写一本讲义，专门为我师门“恶魔党”写的SAS攻略，最早的书名为“SAS在财务研究上的应用：恶魔党的私传秘籍”。最初的版本仅有129页，内容上以财务领域常常涉及的语言为主，之后接触到SQL语言以及SAS中的Output Delivery System（ODS）的功能，更进一步开启了笔者将实践表格输出到Excel的进程，于是在本书的第9章到第17章所呈现的SAS语法，都能将实验结果整理成可发表的论文格式，并且将其输出到Excel文档，可以节省人工输入结果的时间，我想如果这本书可以帮助读者节省进行实验的时间，不管是学生、研究者都能够把时间花在更有生产力的工作上。

虽然本书主要是针对SAS在财务研究上的编程，但是SAS在应用上不仅限于财务研究分析，SAS除了是一款优良的统计软件外，还是一家跨国公司的名称，其主要是为公司以大数据的思维逻辑构建适合的数据库，并且提供良好的分析模组，其运用层面包含商业智能分析、客户智能、数据管理、决策管理器、绩效管理、风险管理、供应链管理，而由于SAS本身最早起源于大型数据库的分析，因此在大数据背景下，还具备了云分析以及文本分析等功能，对于各行各业都起着重要的功用，其目前应用的行业包含教育、医疗保健、制造业、媒体分析、旅游与运输业、汽车行业、电信业、零售业、资本市场、银行业以及保险业等相关行业。在美国，如果取得了SAS的Base以及Advance的国际证照，一般都能取得10万美元以上的年薪。

而在大数据时代，金融行业面临着微信钱包、余额宝等手机理财侵蚀传统银行业提供的投资理财

服务，又面临着 P2P 网贷来抢占传统消费金融贷款以及企业金融贷款的业务，在此情况下，金融业将由传统以人力为主的服务转向数据挖掘的服务，在理财上，需要了解什么样的客户会希望接受人力服务的理财规划，或者是提供相关的理财材料供其选择；在借贷上，需要快速地了解该位客户未来违约的可能性，这些都需要进行数据收集规划，并且进行分析，而 SAS 公司这几年来还为大型银行提供其数据挖掘以及消费者信用分析的定制化服务，在全球 100 强银行中，有 99 家银行使用 SAS 的产品服务，应运而生的是，这 99 家银行产生了对于专精 SAS 的金融从业人员的需求，而这样的趋势现象也在我国逐渐蔓延。

在我国，SAS 自从 2013 年开始举办汇丰杯中国高校数据分析大赛以来，除了提供高额奖项外，也提供了企业招才的机会，以 2015 年的赛事为例，SAS 还邀请了毕马威、京东金融、中信银行、丰田汽车金融（中国）有限公司、交通银行太平洋信用卡中心、南方航空电子商务部、广汽汇理汽车金融有限公司、奇瑞徽银汽车金融股份有限公司、招商银行、平安银行、京东方、光大银行等企业作为大赛的颁奖者，且这些企业也提供参赛同学实习以及未来工作的机会，这说明中国金融行业已经开始运用 SAS 作为进行数据分析的主要工具，而 SAS 更是未来金融业学生必备的工作技能。

最后，在进行数据分析的过程中，我常提醒自己，“如果数据站在你这边，就引用数据；如果理论站在你这边，就引用理论；如果两者都背离你而去，你不是发现了新的现象，就是做错了，而放弃是在做错的时候才做的决定。”也希望各位读者，在进行研究分析时，共勉之。

林煜恩

2016 年于吉林大学匡亚明楼

目 录

第 1 章	SAS 入门介绍	1
1.1	SAS 的基本接口介绍	2
1.2	SAS 语法的基础架构	2
1.3	如何输入数据	4
1.4	如何输出数据	13
1.5	总结	15
第 2 章	SAS 数据的运算与函数	16
2.1	四则运算	17
2.2	统计函数	18
2.3	随机函数	21
2.4	时间函数	22
2.5	文本变量的处理	26
2.6	总结	30
第 3 章	数据与变量的产生和选取	31
3.1	利用 SAS 产生数据	32
3.2	保留、删除变量	36
3.3	保留、删除观测值	38
3.4	抽样方法	38
3.5	总结	43
第 4 章	数据的排序、分组与转置	44
4.1	数据的排序 (proc sort)	45
4.2	数据的分组 (proc rank)	49
4.3	数据的转置	53
4.4	总结	57

第 5 章 数据的合并	58
5.1 垂直合并	60
5.2 水平合并	65
5.3 总结	72
第 6 章 SAS 的数据库管理	73
6.1 文档的复制、删除与保留 (proc datasets)	74
6.2 结构化查询语言	78
6.3 总结	91
第 7 章 宏语法 (%macro)	92
7.1 基础宏语法	93
7.2 进阶宏语法	96
7.3 宏语法撰写技巧	107
7.4 总结	111
第 8 章 描述统计	112
8.1 常见的描述统计量	113
8.2 相关系数	121
8.3 个人化表格宏解析	126
8.4 趋势图基础语法介绍	129
8.5 离群值的处理 (winsorize)	140
8.6 总结	144
第 9 章 两群体差异性检定	145
9.1 均值检定	146
9.2 中位数检定	151
9.3 两群体检定宏进阶用法	154
9.4 总结	156
第 10 章 投资组合与报酬率检定	157
10.1 投资组合股票的数目与风险	158
10.2 效率前缘的绘制	164
10.3 初探动能投资策略	168
10.4 再探动能投资策略	177
10.5 Newey and West 的调整语法	185
10.6 总结	190

第 11 章 基础回归语法	191
11.1 回归语法介绍	192
11.2 格式化回归模型输出	201
11.3 Fama-MacBeth 回归模型	212
11.4 总结	222
第 12 章 回归语法的应用	223
12.1 移动窗口 (moving window)	224
12.2 共同基金绩效评估: 移动窗口的应用	229
12.3 滚动法 (Rolling)	234
12.4 Where 语法有妙招	239
12.5 结构性改变	241
12.6 分段回归 (piecewise regression)	246
12.7 总结	254
第 13 章 panel data (proc panel、proc tscsreg)	255
13.1 固定效应与随机效应的估计方法	256
13.2 panel data 的实证流程	261
13.3 格式化 panel data 模型输出	264
13.4 总结	267
第 14 章 罗吉斯特模型	268
14.1 logit model (logistic regression)	269
14.2 conditional logistic regression	276
14.3 multinomial logistic regression	280
14.4 分类与概率转换	287
14.5 总结	291
第 15 章 tobit 模型 (proc lifereg)	292
15.1 受限数据 (censored data) 与截断数据 (truncated data)	293
15.2 格式化 tobit 模型输出	300
15.3 总结	302
第 16 章 事件研究法	303
16.1 短期事件研究法	304
16.2 日历期间投资组合法	312
16.3 买进持有异常报酬率: 配对投资组合法	318

16.4	买进持有异常报酬率：配对样本法.....	326
16.5	总结.....	335
第 17 章	特殊议题.....	336
17.1	均值抽样分配.....	337
17.2	拔靴法（Bootstrap method）.....	339
17.3	构建 5 因子与动能因子报酬率.....	347
17.4	总结.....	356

第 1 章

SAS 入门介绍

对于大多数人而言，SAS 似乎是个很难入门的软件，因为它需要自行撰写程序。但是这也意味着 SAS 的自由度相当高，用户可以自由地撰写程序。本书采用的是 SAS 9.2 版本，书中的部分语法只兼容于 SAS 9.2 以上的系统；而 Proc Panel 语法是从 SAS 9.2 以后开始提供的，若使用 SAS 9.2 以前的版本，可能无法使用。

首先简单介绍一下如何使用 SAS。SAS 由三大部分组成：editor、log 和 output，如表 1-1 所示。

表 1-1

editor	SAS 的程序编辑器，所有的程序皆在此输入
log	记录执行 SAS 过程中遇到的问题。如果 SAS 发生问题，常会出现绿色的警告 (warning) 和红色的错误 (error)，初学者往往会因此以为程序写错。事实上有些警告和错误是在撰写通用程序时产生的，这些问题不会影响 SAS 结果的正确性。不过真正严重的问题是 SAS 中止程序运行，这时就需要进行调试 (debug)
output	SAS 执行程序后的输出结果

接下来将逐一介绍以下内容：SAS 的相关基本接口；SAS 语法的基础架构；如何读取外部文件，以应用在其后的数据处理和一些程序上，文件分为记事本数据、Excel 文件数据及外部的 SAS table 文件数据；如何将已经整理好的 table 文件输出到外部文件，以便用户使用不同的统计软件进行数据的相关验证。

1.1 SAS 的基本接口介绍

初始进行程序的编写与执行时，需要善用工具栏中的相关工具，如图 1-1 所示，相关操作说明如表 1-2 所示。



图 1-1

表 1-2

	这三个图标是常见的文件管理图示，分别为打开新文件、打开旧文件和存储文件
	程序的打印图标，便于用户教学调试等。一开始不熟悉程序时，可以将程序打印出来，避免长期观看计算机屏幕
	在编辑程序时，有时候会用到相同的语法，这时就可以使用剪切、复制、粘贴操作来处理数据。如果有错也可以撤销
	手动输入新的逻辑库和检验 SAS 逻辑库，后者较常用
	程序的执行（submit）、清除所有程序、结束程序和 SAS 的 help 选项（查询指令）。对初学者最重要的是小黑人图标和惊叹号图标。进阶使用者要常常使用 help 选项来进行查询

在进行 SAS 程序的撰写时，有时候还要使用 Edit (E) 中的 Find 和 Replace。显而易见，SAS 程序的编辑方法与 Word 文档编辑文字的操作非常相似。在 SAS 程序编辑中单击鼠标的右键，也可以看到与 Word 界面相似的复制、粘贴等相关编辑功能，这使得我们在进行 SAS 的程序编辑时，可以很快地适应相关操作。接下来逐步介绍如何编辑 SAS 程序。

1.2 SAS 语法的基础架构

在 SAS 语法中总共有两大体系，第一体系为 data step (数据步)，第二体系为 proc step (程序步)。

在 data step 中，主要是进行数据与变量的整理。财务领域内公司的数据类型非常多，有股票交易数据、公司财务报表数据，甚至还有多年期的数据，因此需要对不同数据源的数据进行整理，这些整理步骤往往是在 data step 中进行整理的。除了整理数据源外，变量的整理也是财务数据整理中的一大重头戏，该部分可以分为两类：一类为变量的横向处理，即所有的数据都是同一公司、同一个年度的数据，这部分通过四则运算便可完成；另一类是变量的纵向处理，在财务研究中，往往要使用公司前一年度的数据，即会使用滞后的变量，如计算营收增长率。在数据库中，所有公司的数据是按顺序排列的，编程时就需要注意滞后的变量是否会取到另外一家公司的值。

在 proc step 中，就是使用 SAS 设计好的规范程序。笔者根据个人的财务实证经验，将 SAS 的程序步划分为三类：变量类、模型类和数据库类。变量类，顾名思义，是针对变量进行排序、分组、运算等功能的程序，在撰写过程中通常带有 var (variable; SAS 通常是采用英文单词的前三位)，其结尾都以 “run;” (运行) 结束。模型类是进行实证时的语法，撰写过程中都带有 model，语法结尾多缀以 “quit;” (停止) 结束。数据库类的程序通常不以单一数据中的变量作为整理对象，其运行多以整个或多个数据文件为对象。在撰写程序过程中常常会出现过渡数据，这不是我们最终要保留的数据，此时可以利用数据库类的程序步将其删除。以下为笔者进行实证分析时常用到的程序步¹，如表 1-3 所示。

表 1-3

变量类	
程序步	程序说明
proc sort	可将数据依照指定的变量进行排序，是非常常用的程序步
proc rank	可将数据依照指定的变量进行分组，是投资组合分析中常用的程序步
proc transpose	可将数据依照指定的变量进行分组
proc means	可针对指定的变量进行描述统计，与 proc univariate 相似
proc univariate	可针对指定的变量进行描述统计，与 proc means 相似
proc corr	可针对指定的变量进行相关分析
模型类	
程序步	程序说明
proc reg	进行普通最小二乘法 (OLS) 的回归程序步
proc model	可撰写各种模式，本书藉以计算 Newey West 的估计
proc panel	进行面板数据的程序步
proc logistic	进行罗吉斯特的程序步
proc probit	进行波比特的程序步，编程语句逻辑和 proc logistic 类似
proc lifereg	本书主要用来进行 Tobit model 的程序
proc mixed	混合线性模式，可运行随机系数模式
proc arima	进行时间序列中的自回归以及移动平均的模式
数据库类	
程序步	程序说明
proc contents	可检验数据的各项内容，确认数据的各项特征、长度
proc datasets	针对数据数据进行处理
proc copy	可将目标的文件复制到逻辑库

1 这种分类方法仅是笔者个人进行实证分析的心得分类，并非规定的分类。

续表

proc delete	可将目标的文件删除
proc append	可将新文件附加到新文件，特别注意变量的长度尽量要一致
proc sql	结构化查询语言，是运用性相当重要的数据库语言
proc compare	比较两个文件的程序步
proc import	将外部数据导入 SAS，通常是导入 Excel 以及 CSV 文件
proc export	将 SAS 数据导出到外部文件，通常是导出 Excel 以及 CSV 文件
proc surveyselect	随机抽样语法，针对文件抽取数据，进行拔靴复制法的重要程序步

笔者根据整理财务数据的经验，认为数据处理大多由四部分构成：第一部分是数据的水平处理，第二部分是数据的垂直处理，第三部分是数据的转置处理，最后一部分是数据的配对处理。前两个部分使用数据步进行处理的，后两个部分采用程序步处理，其中转置处理使用 proc transpose，配对处理则采用 proc sql。只要灵活掌握这四部分的处理逻辑，就可以轻松地将所有数据整理完毕。

1.3 如何输入数据

SAS 本身可以读取外部文件，如记事本数据和 Excel (CSV) 文件的数据。一般而言，Excel 可以采用 SAS 本身的 import 操作输入数据，所以很容易上手，下面介绍具体操作。

1. 在程序中输入数据

输入数据的简单程序示例：

```
data a;
input id a $ b c;
datalines;
1101 a 2 3
1102 b 3 4
1103 d 2 4
;
run;
```

该语法会生成一个 SAS 的 table，命名为 a，并输入 id、a、b、c 四个变量，datalines 的作用是告知 SAS 下面要输入数据，接着为四种数据，结束时输入 run，SAS 就会将数据存到表 a 中，如图 1-2 所示。

	id	a	b	c
1	1101	a	2	3
2	1102	b	3	4
3	1103	d	2	4

图 1-2

在上例中，语法结束时，都会以分号“;”来区分。这里要注意一点，如果是在程序中输入数据，分号不能与数据在同一行。

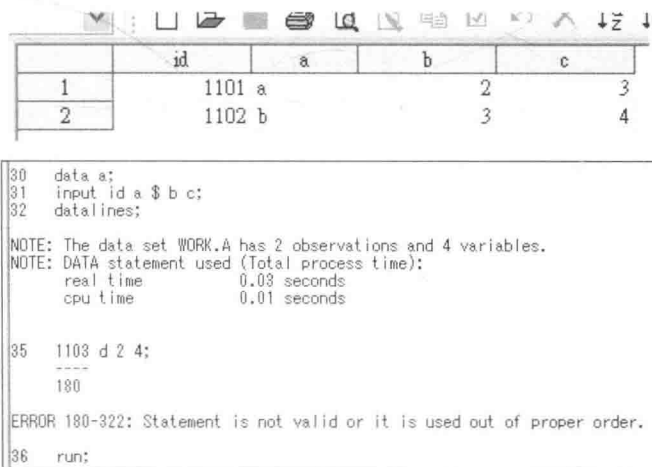
```
data a;
input id a $ b c;
datalines;
1101 a 2 3
1102 b 3 4
1103 d 2 4;
run;
```

如图 1-3 所示为出现程序警告的数据输入实例，第三笔数据无法读进去，检验 SAS 的 log 文件，发现对该语法提出了警告。ERROR 中的信息显示该语法无效或者顺序不当。

在此提醒一点，如果要在程序中输入数据，分号必须在数据列的下一行独立存在。图 1-3 显示的例子中数据都是以空格隔开的，所以导致第三笔数据无法读进去。如果数据本身是链接在一起的，可以按照下列方法来读取。

```
data a;
input id 1-4 a $ 5 b 6 c 7;
datalines;
1101a23
1102b34
1103d24
;
run;
```

首先，id 1-4 指的是读取以下数据 1 到 4 栏为 id 变量，a 为第 5 栏的数据，前面有美元符号 (\$) 表示其为文本类变量，第 6 栏读取为 b，第 7 栏读取为 c。这样的读取方式在读取记事本数据 (txt 文件) 时也可以适用。



	id	a	b	c
1	1101	a	2	3
2	1102	b	3	4

```
30 data a;
31 input id a $ b c;
32 datalines;

NOTE: The data set WORK.A has 2 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.01 seconds

35 1103 d 2 4;
----
180

ERROR 180-322: Statement is not valid or it is used out of proper order.

36 run;
```

图 1-3

有时候数据形式如时间 1996 年 1 月,要想将其拆成年、月的显示格式(笔者个人偏好的读取方式),就需要读取 1-4 这样的字段数据。但如果数据本身使用空格键进行区隔,例如 1 月之后紧接着空白数据,那么此时只要输入 m、id、a 等,SAS 就会自动区分读取数据了。具体程序如下,生成的对应表格如图 1-4 所示。

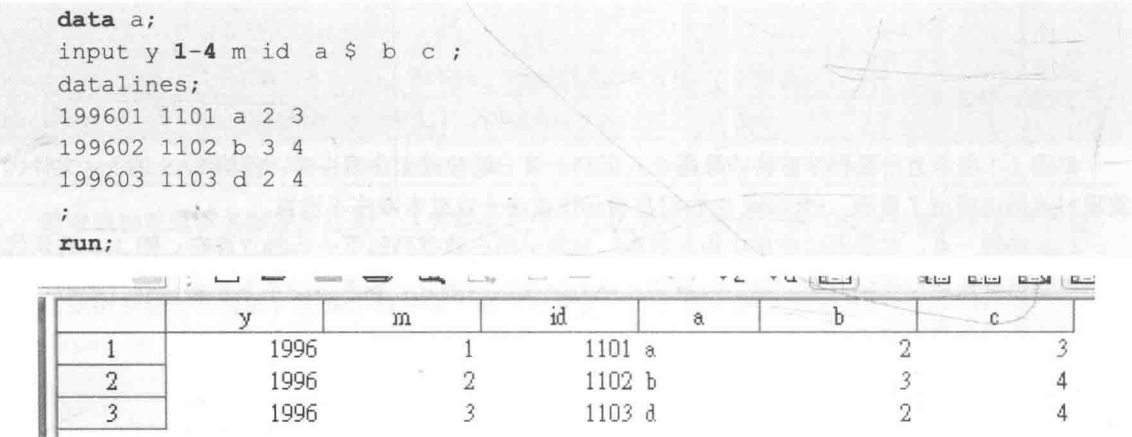


图 1-4

在撰写程序时,可能会觉得将月数据分拆成两个变量不适当,于是加入 month 1-6 的命令,希望 SAS 重复读取 month。以数据 199601 为例,我们可以读出三个变量 1996、1 和 199601。这里需要注意 SAS 有一个特性,当指定字段后,下一个变量若不指定字段,数据就会从上一个指定字段的地方重新读取。该部分的程序如下所示,生成的对应表格如图 1-5 所示。

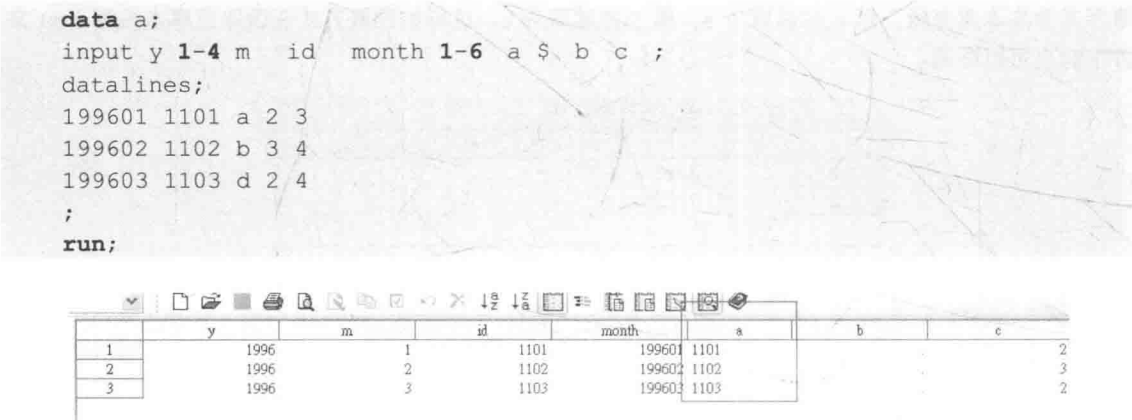


图 1-5

由图 1-5 可以了解到 SAS 在读取数据 199601 之后, a 变量又重新读取 1101 这笔数据。因此如果要重复读取数据,只有下列两种读法。

第一种方法是在读取完所有数据之后,另外要求 SAS 读取第 1~6 栏的数据。具体程序如下,生