

经济学与管理学实验教学系列教材

截面、面板数据分析 与STATA应用

■ 兰 草 编著



WUHAN UNIVERSITY PRESS
武汉大学出版社

经济学与管理学实验教学系列教材

截面、面板数据分析 与STATA应用

■ 兰 草 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

截面、面板数据分析与 STATA 应用/兰草编著. —武汉: 武汉大学出版社, 2012. 6

经济学与管理学实验教学系列教材

ISBN 978-7-307-09727-8

I. 截… II. 兰… III. 经济统计—统计数据—经济计量分析—高等学校—教材 IV. F224.0

中国版本图书馆 CIP 数据核字(2012)第 073038 号

责任编辑:陈 红 责任校对:黄添生 版式设计:马 佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷:湖北鄂东印务有限公司

开本: 720 × 1000 1/16 印张: 21 字数: 373 千字 插页: 1

版次: 2012 年 6 月第 1 版 2012 年 6 月第 1 次印刷

ISBN 978-7-307-09727-8/F · 1664 定价: 36.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

总 序

现代实验教学是将专业理论知识、信息技术和操作技能等融于一体的教学，是经济学与管理学新发展的重要内容，其既对已有的经济学和管理学教学体系、教学内容、教学方法、教学手段、教学条件、教学管理等提出了新的要求，也对从事经济学和管理学教育的教师和管理者提出了新的要求。

限于条件、学科发展、硬件与软件设施以及人们对经济学与管理学教学内容和方式认识的局限，过去经济学与管理学的教学长期偏重于理论教学，实验教学基本上处于可有可无的境地。随着经济与管理实践的不断发展和经济与管理实践的不断变化，随着经济学和管理学学科的进一步发展，随着实验教学硬件与软件设施的不断改善，更由于人们对经济学与管理学教学内容和方式认识的转变，以计算机为物质基础、以现代信息技术和经济学与管理学专业软件为支撑、以现代模拟教学为主要方式的经济学与管理学实验教学已经被认为是经济学与管理学的重要教学内容和方式。推进经济学与管理学的实验教学，达到理论教学与实验教学并重已经成为经济学与管理学专业教育的重要方向。实验教学已经成为培养宽口径、厚基础、高素质、强能力的具有创新、创造和创业精神和能力的高层次复合型的经济学和管理学专业人才的必不可少的重要环节。为了进一步促进经济学与管理学实验教学，我们决定推出经济学与管理学实验教学系列教材。

编写本系列教材的宗旨是要推动经济学与管理学的实验教学，提高经济学与管理学的实验教学水平，完善经济学与管理学的实验教学体系，培养和提高学生的实践能力，促进经济学与管理学学科专业的发展。

我们将以本科各专业课程教学计划为基本依据，根据实验教学需要和人才培养需要，以实验教学系列教材的形式分批推出实验教学教材，最终形成体系比较完整的适应实验教学和人才培养需要的具有学院学科专业特色和优势的系列教材。

实验教学教材的编写还没有现成的模式可以借鉴，编写实验教学教材对我们而言是一个尝试，一方面，我们希望本系列教材对国内经济学与管理学实验

教学教材的编写起到抛砖引玉的作用，尤其希望得到从事经济学与管理学教育的人士的指教；另一方面，也希望通过本系列教材的编写积累经验，探索出实验教学教材编写的可行模式。

陈继勇

2008 年春于珞珈山

前 言

Stata (Data Analysis and Statistical Software), 作为一款集数据分析、管理、作图、模拟仿真以及编程等诸多功能于一体的统计分析软件包, 最初由美国计算机资源中心 (Computer Resource Center) 研发, 1985 年成为美国 Stata 公司 (Statacorp) 的产品。目前 Stata12.0 版本已正式发布, 其在不断更新升级的过程中愈加成熟, 适用于多种操作系统 (Windows, Mac OS X, Unix, Linux) 并凭借其强大的功能、逻辑一致的命令、简洁而灵活的语句以及友好的用户操作界面受到了众多专业人士的追捧, 被广泛应用于世界各类商业和科研机构, 与 SAS、SPSS 一起被称为新的三大权威统计软件。

计量经济学, 作为以经济理论和经验事实为依据, 运用数学和统计学方法研究具有随机性特征的经济变量关系和经济活动规律的计量方法及其应用的一门经济学分支学科, 其具有庞大内容体系的理论方法已被大量应用于现代经济学和社会科学等各领域, 在现代实证研究中占据了相当重要的位置。在正确设定计量经济学理论模型的前提下, 选取恰当的样本数据并整理是完成计量经济学实证分析的基石。常见的样本数据分为截面数据 (cross-section data)、时间序列数据 (time-series data) 和面板数据 (panel data) 三类: 截面数据是指一批发生在同一时间截面上的调查数据; 而时间序列数据则是一组按照时间顺序排列的数据序列; 面板数据又称综列数据或平行数据, 它是一部分个体、企业、家庭、国家等样本在一段时间内的观测值所构成的集合, 近年来面板数据分析研究方法获得了飞跃发展并被越来越多的实证分析所应用。同时, 按照研究对象划分, 计量经济学被分为微观计量经济学 (Microeconometrics) 和宏观计量经济学 (Macroeconometrics): 微观计量经济学集中于“对个人和家庭的经济行为进行经验分析, 其原材料是微观数据, 而微观数据表现为截面数据和面板数据”; 而宏观计量经济学则主要“利用计量经济学理论方法, 建立宏观经济模型, 对宏观经济进行分析、评价和预测, 而近年来随着动态理论的发展, 动态计量经济学已成为宏观计量经济学的主要研究方向”。

集经济学、数学和统计学于一身的计量经济学, 分析中需大量运用各类代

数、矩阵运算以及各类统计量，它们随着理论和方法的发展而更加复杂，传统的人工运算已远远不能满足需求。伴随着计算机应用软件的发展，各类运用于计量分析的软件开始出现，目前主要使用的有 Stata、EViews、SAS、SPSS 等。这些软件虽然都具有基本的计量经济学分析功能，但它们的侧重点各有不同。《截面、面板数据分析与 STATA 应用》一书主要着眼于截面和面板数据这两类数据类型，在简要介绍相关计量经济学理论和方法的基础上，演示在 Stata11.0 内实现数据分析、参数估计和检验等的相关方法。

本书主要面向已具有一定计量经济学理论基础的大中院校学生，有助于提升他们的实际操作和应用能力。本书共分为十二章，前三章分别介绍 Stata 软件的基础知识、数据管理和基本统计分析应用、作图等基本功能；第四章则对经典计量经济学线性单方程模型基础理论、最小二乘估计方法等进行介绍并通过具体案例演示在 Stata11.0 内进行参数估计、统计检验以及预测；第五章介绍了蒙特卡罗随机模拟方法；第六章到第九章则一一介绍了模型违背基本假设的情况，由于本书主要针对截面数据和面板数据，因此着重于异方差性、内生解释变量以及多重共线性的检验和修正；第十章和第十一章分析了面板数据模型和动态面板数据模型；本书最后一章则简要介绍了二元离散选择模型及其估计方法等。

由于本书主要针对已有一定理论基础的软件初学者，因此书中主要介绍了 Stata 最常用命令的应用。Stata 所包含的命令以及其所涵盖的功能远远超出本书的介绍，对有兴趣深入了解的用户，本书各章节均给出了更多相关的参考链接。强烈建议用户翻阅由 Stata Press 出版的包括 *User's Guide* 在内的一系列用户手册，这些手册对各种命令、语句进行了细致的说明；用户也可以登录 Stata 公司网站 <http://www.stata.com> 查看软件最新版本以及相关参考资料；有关 Stata 的更多资源，第一章结尾部分给出了具体介绍。至于计量经济学理论和方法，更加详细的阐述可参见古扎拉蒂的《计量经济学基础》、伍德里奇的《横截面与面板数据的经济计量分析》以及李子奈的《计量经济学》等书籍。本书用以演示的数据大多来源于这些书籍的课后习题。

书中纰漏之处敬请读者提出并斧正。感谢武汉大学经济与管理实验教学改革创新研究与建设项目的大力支持，感谢所有为本书提出意见和提供帮助的老师和同事。

目 录

第一章 Stata 简介	1
1.1 Stata 界面	1
1.2 Stata 命令语句	8
1.3 do 文件	10
1.4 日志 (log) 文件	14
1.5 共用宏和局部宏 (Global and Local Macros)	15
1.6 Stata 资源	16
1.7 小结	17
第二章 数据管理与概要统计分析	18
2.1 数据形式	19
2.2 数据录入	21
2.3 数据管理	28
2.4 数据集操作	42
2.5 概要统计分析	50
2.6 小结	58
第三章 作图与概要统计分析	60
3.1 二维标绘图 (tway graph)	61
3.2 条形图和点图	90
3.3 饼图	97
3.4 直方图	101
3.5 箱线图与散点矩阵图	103
3.6 图形管理与合并图形	107

3.7	小结	110
第四章	经典单方程线性模型	111
4.1	经典线性回归模型	111
4.2	线性回归模型的参数估计	113
4.3	线性回归模型统计检验	115
4.4	基本回归分析实例：Stata 下的参数估计	118
4.5	预测	127
4.6	使用 Stata 菜单进行线性回归	130
4.7	小结	138
第五章	随机模拟	139
5.1	大样本（渐近）理论 [Large-Sample Theory (Asymptotics)]	140
5.2	伪随机数 (Pseudorandom-number) 的生成	142
5.3	随机模拟	151
5.4	小结	160
第六章	随机误差项假设诊断和广义最小二乘法	161
6.1	异方差性	161
6.2	序列相关性	165
6.3	广义最小二乘估计量 (Generalized Least Square Estimator)	167
6.4	回归诊断和 GLS 在 Stata 中的实现：案例分析	169
6.5	调查数据：加权、聚类和分层	179
6.6	使用 Stata 菜单进行检验	185
6.7	小结	190
第七章	内生解释变量	191
7.1	内生解释变量的来源及后果	191
7.2	工具变量法	197
7.3	两阶段最小二乘法 (Two Stage Least Square Method)	201
7.4	内生性的豪斯曼检验	204
7.5	Stata 下的工具变量估计量	205
7.6	小结	216

第八章 广义矩方法	217
8.1 弱工具变量 (Weak Instruments)	217
8.2 弱工具变量的检验: 实例分析	219
8.3 广义矩方法 (Generalized Method of Moments, GMM)	224
8.4 Stata 下的 GMM 估计量	229
8.5 小结	232
第九章 多重共线性	234
9.1 多重共线性的分类与后果	234
9.2 多重共线性的检验	235
9.3 多重共线性的修正	237
9.4 Stata 下的多重共线性	238
9.5 小结	243
第十章 线性面板数据模型	244
10.1 面板数据模型形式	245
10.2 面板数据模型参数估计方法	247
10.3 固定效应与随机效应	252
10.4 线性面板数据模型分析在 Stata 中的实现	252
10.5 小结	274
第十一章 动态面板数据模型	275
11.1 动态面板数据模型	275
11.2 动态面板数据模型估计法	276
11.3 带有序列相关性的动态面板数据模型	279
11.4 Stata 下的动态面板数据分析	282
11.5 案例: 香烟的动态需求	290
11.6 小结	295
第十二章 二元离散选择模型	296
12.1 二元选择模型	297
12.2 二元选择模型回归方法	298
12.3 Stata 下的 Probit 与 Logit 模型估计	302

12.4	拟合优度和预测	308
12.5	Stata 下的二元选择模型检验	315
12.6	边际效应 (Marginal Effects)	318
12.7	小结	321
参考文献		323

第一章

Stata 简介

作为一款功能强大又极具特色的统计分析软件，Stata 备受众多科研机构、高校追捧，广泛应用于各个学科领域。目前，Stata 公司已经发布了 Stata12.0 版本，进一步提升了软件各项功能并添加了更多分析内容。Stata 软件一直以来都极其重视其命令语句界面，在 Stata8.0 版本以前，所有分析和操作都只能通过输入命令语句完成；从 Stata8.0 版本开始，Stata 便开始加入了用户友好操作界面，通过软件上方菜单、下拉框以及对话框完成分析操作，在给出分析结果的同时给出相对应的操作命令，让不熟悉命令语句的用户能够了解掌握相应命令；Stata 一直不断更新其菜单操作界面并扩展其操作功能，使其更加便捷化、大众化和人性化。

本章将对 Stata 界面，图形用户界面 (graphic user interface, GUI) 进行说明和简单的操作演示；并介绍部分 Stata 常用命令。

1.1 Stata 界面

本书使用 Stata11.0 版本，它适用于 Windows、Unix 以及 Mac 等各类电脑操作系统。图 1.1 即为 Stata11.0 启动后的界面，从图中可以看出除了菜单栏外，软件界面主要由四个窗口组成。分别为命令回顾窗口 (Review)、变量名窗口 (Variables)、结果窗口、命令窗口 (Command)。命令回顾窗口位于软件界面的左上角，软件运行中所执行过的命令将会依次在此列出，错误的命令将以红色字体显示。变量名窗口位于命令回顾窗口之下，即软件界面的左下方，当数据被导入软件后，将在此窗口列出所使用的数据集中所有变量的名称、含义、属性等信息。结果窗口位于软件界面右上方，当未执行任何命令时，结果窗口中将显示类似图 1.1 中的相关内容；而软件运行后所产生的各类结果，如执行命令后的结果和出错信息等均会在结果窗口中列出，用户可以直接将结果拷贝到其他文件中；同命令回顾窗口一样，错误信息以红色字体显示。而最为

重要的命令窗口则位于结果窗口之下，即界面右下方，在此键入需要执行的命令后回车便可得到相对的结果，当所输入的为错误命令或命令不可能执行时，便会以红色字体给出相对简单的出错解释。值得注意的是，Stata 为非汉化版本，因此其不能识别中文，所有命令操作只能通过英文字符或者可识别的命令符号等进行。

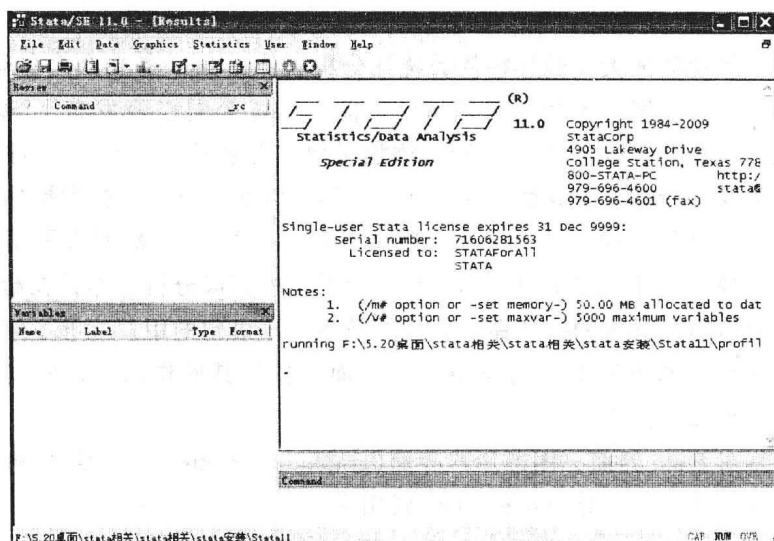


图 1.1

除了四个打开即显示出的窗口外，菜单栏内还包含文件窗口 (File)、编辑窗口 (Edit)、数据管理窗口 (Data)、绘图窗口 (Graphics)、统计分析窗口 (Statistics)、帮助窗口 (Help) 等。点击 File，得到如图 1.2 所示的下拉菜单，通过点击下拉菜单内相应选项，可以执行导入或导出、查看计算机内相应的数据集、存储运行结果、改变结果存储路径、退出软件等功能；在点击下拉菜单内相应选项后 (比如点击 Open)，将会得到如图 1.3 所示的对话框，在打开的对话框中选择查找范围、文件名和文件类型后，可以将符合软件格式的数据文件或者图片文件打开导入。对于其他诸如查看数据、图片的 View 等选项，点击它们后仍然可以得到相应对话框进行操作。在 File 菜单下有 Import 和 Export 两个选项，这两个选项提供了非 Stata 默认格式数据的直接导入和导出方法。对选项下属的几类数据类型第二章将进行详细讲述和说明，这里需要说明的是，通过点击菜单下的相关选项，可以在对话框内直接选择计算机内所存储的

相应数据类型，将其导入 Stata 后便可通过 Save 选项直接将其转换为 Stata 默认的 .dta 数据格式类型。

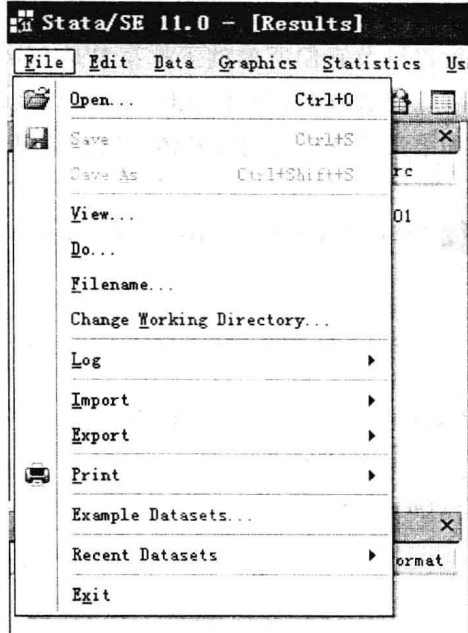


图 1.2

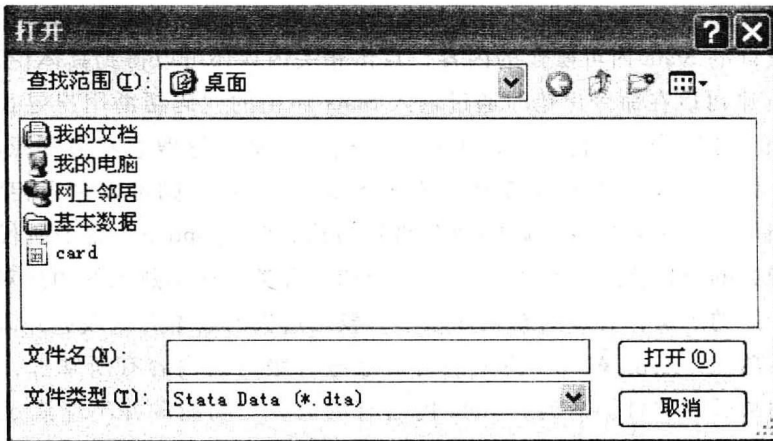


图 1.3

在文件窗口之后便是编辑窗口，点击 Edit 得到如图 1.4 所示的下拉菜单，通过其中的如 Copy 等选项，可以图表、HTML 或图片等格式复制分析所得到的结果从而将其复制粘贴到其他文件类型中；通过 Preference 选项可以对复制出的格式文件、图片的颜色、边框等进行设置。紧随编辑窗口后的数据管理、绘图、统计分析三大窗口以及它们所属的下拉菜单是 Stata 菜单操作中最常用的，通过这三大窗口可以完成数据的分析管理、作图以及统计回归分析等各项操作，由于在第二章和第三章将对它们做具体介绍，这里便不再详细说明。

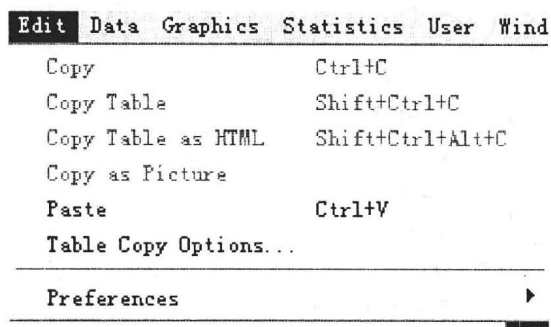


图 1.4

最后值得一提的便是 Stata 中的帮助 (Help) 窗口，其下拉菜单中的选项如图 1.5 所示，对于初学者或者使用新升级版本的用户来说，这是一个非常有用的窗口：点击 Contents 可得到如图 1.6 所示的窗口，这一窗口中列出了帮助窗口下可查询的 Stata 内可操作的内容，点击相关内容还可以得到具体说明链接；而 Search 则可以在命令出错时通过输入 Stata 给出的代码帮助用户寻找出错的具体原因并加以修正。此外，还可以通过输入关键词等搜索网络资源和 FAQs 等内容，以进一步了解 Stata 功能；点击 Stata Command 则可以输入相应命令，确认后便可得到主要命令所属的命令语句格式、各种 option、命令所代表的含义和所起到的分析作用，从而使得对命令和其各类选项不熟悉的用户对命令语句有更全面的了解；升级的新软件版本，较之过去的版本有所改进并添加了新的内容或功能，点击 What's New 便可对这些升级后的内容有所掌握，有利于熟练使用新版本软件。有时，Stata 内已有的命令、功能等并不能满足更高级的分析需要，作为一个互动的软件，可以通过编程等加入用户自写命令从而满足进一步的分析需要。经过长期的使用和发展，Stata 已包含许多用户自写命令，而 Help 下的 SJ and User-written Programs 通过给出相关链接便可查找相应

自写命令的安装地址等，从而在用户的软件内加入这些自写命令；Help 下的其他选项如 Official Update 等，则给出了与软件和其可用资源相关的进一步链接。

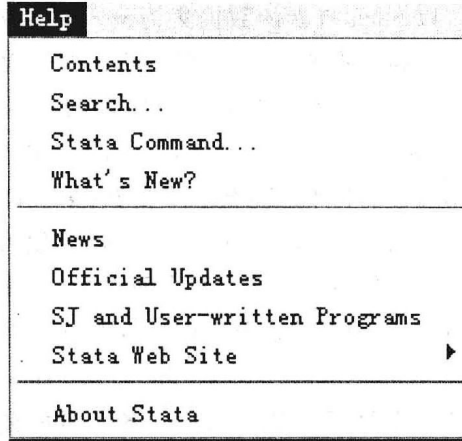


图 1.5

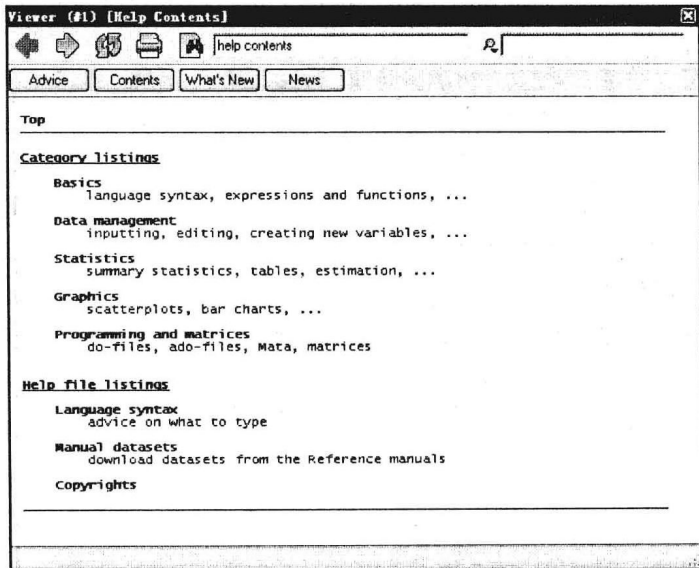


图 1.6

从对窗口菜单的简要说明中可以得知，无论是对于初学者还是对于已经初步掌握软件操作的用户来说，Stata 逐步发展起来的用户友好界面为学习、熟悉其命令、了解软件、进一步开发和分析、掌握相关资源提供了直观而便捷的途径。通过菜单操作便可产生的相应命令，对不了解命令语句的初学者来说更是提供了极大的便利。以 Stata 自带的数据集为例，简要演示如何利用菜单完成简要的统计回归分析：在双击 Stata 应用程序后，打开软件界面，点击 File> Example Datasets...(意味着打开 Stata 自带的数据集)，得到如图 1.7 所示的窗口；点击窗口中的 Example datasets installed with Stata，得到如图 1.8 所示的窗口，它列出了软件自身所带的样本数据集。点击数据集后的 use 便可将对应的数据集导入软件，这里我们选择 auto. dta，点击 use 后便可以发现，Stata 界面的变量名窗口中出现了数据集中的变量，命令回顾窗口和结果窗口相应出现命令 `sysuse auto. dta`，这便是所进行的菜单操作对应的命令语句。在导入数据后，在菜单栏选择 Statistics>summaries, tables, and tests>summary and descriptive statistics>summary statistics，便可得到数据集中各变量的概要统计量，结果如图 1.9 所示。从这个结果可以看出，得到概要统计量的命令为 `summarize`。采用同样的方式，还可以作出变量间的图形、进行参数回归等。然而，Stata 中所有的操作实际上都可以通过在命令栏内输入相应命令完成，并且命令输入方

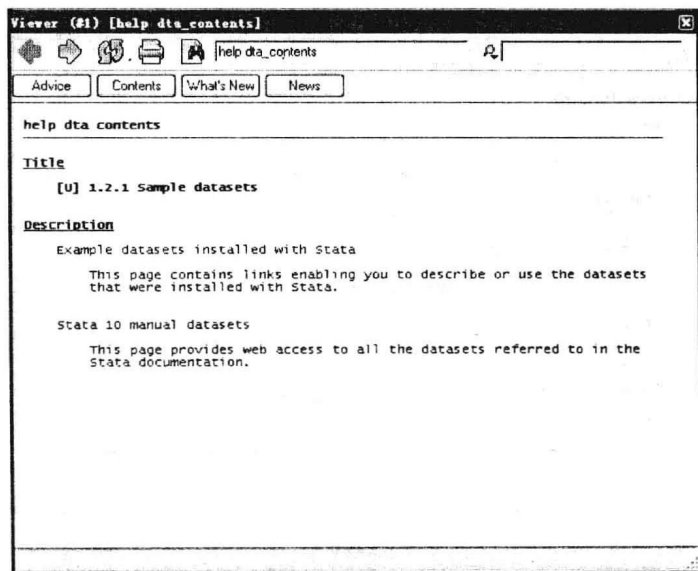


图 1.7