

南开大学经济类系列实验教材

S TATA使用指南与 应用案例

王群勇 著

 中国财政经济出版社

南开大学经济类系列实验教材
南开大学亚洲研究中心资助项目

第1版 (2010) 目次照录封面

STATA 使用指南与应用案例

王群勇 著

中国财政经济出版社

图书在版编目 (CIP) 数据

STATA 使用指南与应用案例 / 王群勇著. —北京: 中国财政经济出版社, 2008. 2
(南开大学经济类系列实验教材)

ISBN 978-7-5095-0624-0

I. S… II. 王… III. ①统计分析-应用软件, STATA-高等学校-教材②经济
计量分析-应用软件, STATA-高等学校-教材 IV. C812 F224.0-39

中国版本图书馆 CIP 数据核字 (2008) 第 051502 号

中国财政经济出版社出版

URL: <http://www.cfeph.cn>

E-mail: cfeph@cfeph.cn

(版权所有 翻印必究)

社址: 北京市海淀区阜成路甲 28 号 邮政编码: 100036

发行处电话: 88190406 财经书店电话: 64033436

北京财经印刷厂印刷 各地新华书店经销

787×1092 毫米 16 开 16.5 印张 380 000 字

2008 年 2 月第 1 版 2008 年 2 月北京第 1 次印刷

定价: 30.00 元

ISBN 978-7-5095-0624-0/TP·0014

(图书出现印装问题, 本社负责调换)

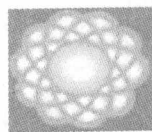
编委会

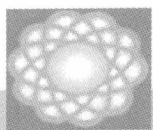
主 任：马君潞

副主任：李秀芳 何自力 沈亚平 侯文强

委 员：（按姓氏笔画排列）

马君潞	王群勇	关路祥	刘晓峰	华 钧
孙佳美	何自力	吴 浙	张骅月	张晓峒
攸 频	李冰清	李秀芳	沈亚平	邹 洋
周爱民	侯文强	赵胜民	涂宇清	秦海英
郭 玲	谢娟娟			





总 序

ZONGXU

南开大学经济学科多年来一直在探讨如何适应改革开放、如何根据理论与实践的发展进行教学改革,包括教学理念与教学内容的更新、教学方式与教学方法的创新。实验教学的内容和方法是教学改革多方面的具体体现。南开大学也为经济学科实验课程教学的开设提供了重要的物质保证,在整合相关资源的基础上投资建设的实验教学中心成为经济学科各专业本科生、硕士生、博士生实验教学和实践教学基地,是经济学科教学、科研和社会服务重要的基础支撑。

经过多年的建设与探索,南开大学经济学实验教学中心逐步建立起与学科发展和人才培养目标相适应、比较科学的实验教学体系,同时组织实验课程教师开发适合于不同专业、不同教学层次的实验课程,并在课程中广泛引入了演示法、案例法、模拟法、仿真法、棋块式沙盘规划法等教学方法。经过数年的积累,实验课程教师在教学的过程中组织学生自主研发教学软件,将科研成果注入实验教学体系,从而使科研成果与教学内容相结合,也使教学软件有了自我更新的能力。

实验课程教材内容体现了实验课程教师多年来不断研究和实践的成果,也体现了南开大学经济学科对教学改革内容的探索。目前,实验教学已经成为经济学教学的重要组成部分。当然,无论是实验课程教材体系还是实验课程教学内容,都有待于根据理论与实践的发展,以及技术手段的提升不断更新和完善。

我们期待读者与同行的意见和建议。

马君潞

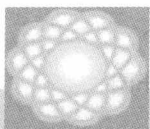
2008年1月于南开大学



南开大学数量经济研究所

王健

山东大学经济学院硕士生



前言

QIANYAN

为了推动国内应用统计学、计量经济学的发展,提高广大理论工作者和实践工作者的数据分析水平,南开大学数量经济研究所推出了 STATA 系列丛书。2007 年 9 月出版了该系列丛书之一《STATA 在统计与计量分析中的应用》。我们在国内一些高校研究生的统计学、计量经济学课程教学中采用该书作为 STATA 软件的教材,得到了广大师生的好评。同时,广大 STATA 用户对本套丛书提出了很多中肯建议,推动了这项工作的不断改进和完善。

本书是该系列丛书之二,与《STATA 在统计与计量分析中的应用》相比,它侧重于介绍 STATA 的界面操作方法。全书分为四部分。第一部分介绍 Stata 的数据管理功能和模型估计的基本概念。第二部分介绍了统计分析方法,包括描述统计、参数检验、非参数检验、方差分析等内容。第三部分介绍多元统计分析方法,包括多元方差分析、主成分分析、因子分析、典型相关分析、聚类分析、判别分析、对应分析和多维标度分析。第四部分介绍经济计量分析方法,包括单方程结构模型、系统方程模型、离散选择模型、计数模型、截断模型和归并模型、时间序列模型和面板数据。

本书介绍的界面都是基于 Stata 10.0 版本。书中利用案例对各种模型的估计、检验、预测等进行分析和说明。每一章都介绍了 Stata 的操作方法及案例分析,用户可以在网址 <http://202.113.23.180/teacher/showjiaoshi.asp?id=229> 下载这些数据,进行实际练习和操作。

本书的写作得到南开大学亚洲研究中心“计量经济模型的稳定性研究”项目组和南开大学经济学院的资助,在此表示感谢。南开大学数量经济研究所王健、杜勇宏副教授参与了本书多元统计部分的校对,山东大学经济学院硕士生

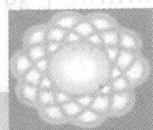
花小安、王凯传参与了本书第2章、第18章部分内容的写作，在此表示诚挚的感谢。

书中的错误或不妥之处，敬请读者批评指正。

王群勇

2008年2月于南开园

电子信箱: nkstata@yahoo.com.cn



目 录

MULU

第一篇 Stata 概述

第 1 章 Stata 简介	(1)
1.1 Stata 的版本	(1)
1.2 Stata 的特点与功能	(2)
1.3 Stata 的界面介绍	(3)
1.4 基本设定方法	(4)
第 2 章 数据管理	(6)
2.1 数据文件的内容描述	(6)
2.2 对变量和观测值的操作	(8)
2.3 标签	(11)
2.4 分类数据的处理	(13)
2.5 字符变量与数值变量的互换	(15)
2.6 数据的排序	(18)
2.7 缺失值的处理	(19)
2.8 重复观测值	(20)
2.9 长数据与宽数据的转换	(22)
第 3 章 矩阵与函数	(24)
3.1 矩阵运算	(24)
3.2 变量与矩阵的相互转换	(28)
3.3 函数	(30)
第 4 章 作图	(35)

4.1	绘制图形	(35)
4.2	修改图形	(36)
4.3	图形的其他操作	(38)
第5章	模型估计、检验与预测概述	(39)
5.1	估计模型的基本选项	(39)
5.2	估计模型的相关操作	(41)
5.3	模型的基本统计指标	(46)
5.4	参数检验	(48)
5.5	参数函数的估计	(53)
5.6	变量的边际影响或弹性	(55)
5.7	模型预测	(57)
5.8	建立参数约束	(59)

第二篇 单元统计分析

第6章	描述统计学	(60)
6.1	描述统计量	(60)
6.2	单因素分类表	(64)
6.3	两个分类变量的列联表	(65)
第7章	参数假设检验	(67)
7.1	总体均值的假设检验	(67)
7.2	总体方差的假设检验	(69)
7.3	总体比率的假设检验	(71)
第8章	非参数检验	(73)
8.1	单变量的分布检验	(73)
8.2	两组样本的同分布检验	(74)
8.3	多组样本的同分布检验	(76)
8.4	列联表的独立性检验	(77)
8.5	序列自相关的游程检验	(77)
第9章	(协) 方差分析	(79)
9.1	单因素方差分析	(79)
9.2	一般方差分析	(81)
9.3	线性约束检验	(82)
9.4	案例	(83)

第10章 相关分析	(85)
10.1 Pearson 简单相关系数	(85)
10.2 偏相关系数	(86)
10.3 分类变量的相关系数	(87)
第三篇 多元统计分析	
第11章 多元方差分析	(89)
11.1 多元方差分析	(89)
11.2 轮廓分析	(92)
11.3 霍特林均值向量检验	(94)
第12章 主成分分析	(96)
12.1 主成分估计	(97)
12.2 主成分分析的几种工具	(99)
12.3 作图	(102)
12.4 旋转	(104)
12.5 预测	(105)
12.6 案例	(106)
第13章 因子分析	(109)
13.1 因子估计	(109)
13.2 模型诊断与检验	(110)
13.3 因子旋转与作图	(112)
13.4 预测	(113)
13.5 案例	(113)
第14章 典型相关分析	(114)
14.1 典型相关估计	(114)
14.2 预测	(115)
14.3 典型载荷	(116)
14.4 案例	(117)
第15章 聚类分析	(120)
15.1 系统聚类法	(120)
15.2 非系统聚类法	(123)
15.3 聚类停止法则	(124)
15.4 分层树	(125)



15.5	生成分类变量	(126)
15.6	其他命令	(126)
15.7	案例	(128)
第16章	判别分析	(131)
16.1	判别分析	(131)
16.2	模型诊断与检验	(133)
16.3	预测	(138)
16.4	案例	(139)
第17章	对应分析	(142)
17.1	简单对应分析	(142)
17.2	多重、联合对应分析	(154)
17.3	双重信息图	(157)

第四篇 计量经济分析

第18章	单方程结构模型	(159)
18.1	单方程线性模型的 OLS 估计	(159)
18.2	异方差的检验与修正	(163)
18.3	自相关的检验与修正	(165)
18.4	多重共线性的检验与修正	(169)
18.5	工具变量法	(170)
18.6	非线性最小二乘法	(174)
第19章	系统方程模型	(177)
19.1	多元回归	(177)
19.2	似不相关回归	(178)
19.3	联立方程模型的估计	(180)
19.4	预测	(182)
第20章	离散选择模型	(183)
20.1	二项选择模型	(183)
20.2	排序选择模型	(188)
第21章	计数模型	(191)
21.1	泊松回归和负二项回归	(191)
21.2	广义负二项回归	(194)
21.3	零膨胀计数模型	(195)

21.4 零截断计数模型	(196)
第 22 章 受限因变量模型	(197)
22.1 截断模型	(197)
22.2 归并模型	(200)
22.3 Tobit 模型	(202)
第 23 章 时间序列模型	(205)
23.1 基本命令介绍	(205)
23.2 相关图	(208)
23.3 ARIMA 模型与 SARIMA 模型	(211)
23.4 单位根检验	(214)
23.5 向量自回归模型	(216)
23.6 结构向量自回归模型	(226)
23.7 向量误差修正模型	(229)
23.8 ARCH 族模型	(232)
第 24 章 面板数据	(236)
24.1 面板数的定义与描述统计	(236)
24.2 跨方程相关结构模型的估计	(237)
24.3 固定效应、随机效应模型	(240)
24.4 工具变量估计	(243)
24.5 动态面板数据	(246)
参考文献	(249)

第一篇

Stata 概述

第1章

DIYIZHANG

Stata 简介

Stata 是一个用于数据管理和分析的统计软件，由 Stata 公司（StataCorp）研制开发。从 1985 至今，先后开发了 1.1、1.2 等多个版本。目前 Stata 的最新版本为 2007 年 6 月份推出的 Stata 10.0。

1.1

Stata 的版本

Stata 10 包括四种版本：Small（小型版）、IC（标准版）、SE（特殊版）和 MP（多处理器版），其中属 MP 版本最为强大。MP 版与 SE 版的功能完全相同，但 MP 版的运算速度比 SE 版的要快很多。

不同版本对样本容量、变量个数、矩阵阶数、宏的字符长度等有着不同的限制。以 SE 版为例，其最大变量个数为 32 767，最大字符长度为 244 字节，最大矩阵阶数为 11 000（即 $11\,000 \times 11\,000$ ）。Stata 默认值为：最大变量个数为 5 000，最大矩阵阶数为 400，最大内存为 10 兆。

1.2

Stata 的特点与功能

在众多的统计软件中, Stata 凭借其独特的优势得到了迅速的发展。与其他统计计量软件相比, 其主要特点包括:

1. Stata 将统计功能与计量分析较完整地结合起来。这一点是 Stata 受到众多研究机构青睐的重要原因之一。Stata 不仅可以实现诸多的统计分析方法, 比如描述统计、假设检验、方差分析 (ANOVA)、多元方差分析 (ANOVA)、主成分、因子分析、典型相关、聚类分析、对应分析、多维标度等, 而且还包括了许多计量模型, 比如经典单方程回归模型、系统方程模型、特殊因变量模型 (离散选择模型、计数模型、截断模型、归并模型等)、时间序列模型 (ARIMA、GARCH、VAR、VEC)、面板数据分析等。

2. Stata 除了强大的统计和计量分析功能之外, 还可以进行强大而方便的矩阵运算。Mata 是 Stata 的矩阵运算语言。可以通过 Mata 实现大量的矩阵运算功能, 除了基本的矩阵运算功能之外, 还包括很多重要的功能, 如面板数据的处理等。

3. 精致的作图功能。在 Stata 10 中, 作图功能得到了进一步强化。用户可以在图形上面直接对其进行修改, 比如添加标题、添加箭头、修改图例等。

4. 丰富的网络资源。任何一个软件都很难将这一领域的所有方法都囊括进来, 因此与其他用户进行学术资源的交流是学术发展的重要推动力。Stata 在这一点上做得非常突出。在 Stata 网站上, 用户可以找到学习统计学和计量经济学以及 Stata 软件的诸多资源。用户可以在 Stata 的几个官方网站及其链接上寻找自己所需要的资源。

<http://www.stata.com>

<http://www.stata-press.com>

<http://www.stata-journal.com>

5. 及时便捷地更新。用户可以通过网络连接直接进行升级, 或者通过 Stata 官方网站 <http://www.stata.com> 进行下载升级。

除了上述官方升级之外, 用户也可以下载更新其他用户编写的 Stata 模块。通过 Help → SJ and User - Written Programs 下载更新。

6. 严谨、简练、灵活的程序语言。Stata 提供了严谨、简练而灵活的程序语句, 用户可以编写自己的命令和函数, 也可以制作自己的操作界面。

1.3

Stata 的界面介绍

图 1-1 是 Stata 10 运行的界面。

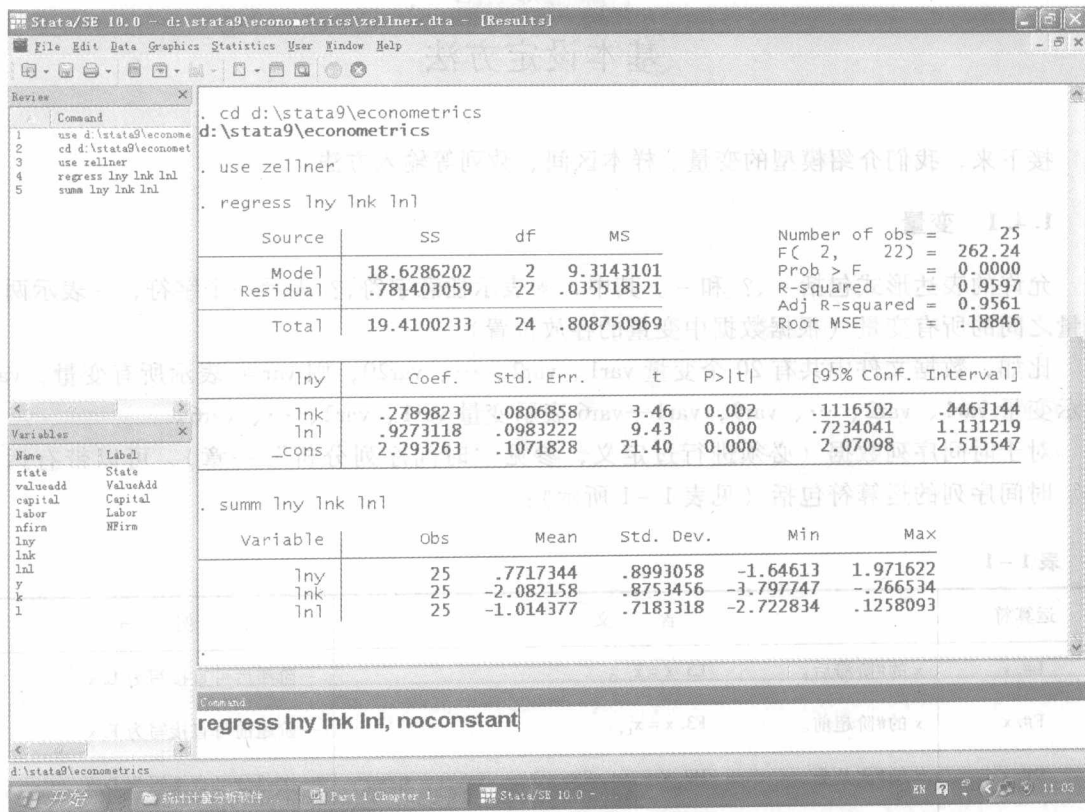


图 1-1 Stata 10 运行界面

主菜单中包括 File、Edit、Data、Graphics、Statistics、User、Window、Help。File 的下拉菜单中包括打开、保存文件以及导入、导出数据等功能。Edit 主要用于数据的管理和设置 Stata 的界面特征，比如文件合并、增加、删除变量、矩阵操作等。Graphics 用于作图，Statistics 用于各种统计和计量分析，User 用于构建用户自己的菜单，Window 用于对显示界面的操作，Help 为帮助文件。在主菜单的下方是常用功能的快捷图标，如打开文件、保存、数据编辑器等。

中间的主体部分由以下部分构成，左侧是运行过的命令（Review）和内存中的变量（Variables）。Review 包括命令的序号、内容（Command）和允许是否成功的标记结果（_rc）。Variables 则包括变量的名称（Name）、标签（Label）、存储类型（Type）和显示格

式 (Format)。

右侧则是输入命令栏 (Command) 及命令运行的结果 (Results)。最左下角显示的是当前的目录 (D: \ stata10 \ data)。

上述界面是 Stata 的默认设置特征。用户可以通过 Edit→Preference 更改界面布局, 比如输出结果的背景颜色、字体大小。

1.4

基本设定方法

接下来, 我们介绍模型的变量、样本区间、数列等输入方法。

1.4.1 变量

允许的表达式包括 *、? 和 -。其中, * 表示任意字符, ? 表示一个字符, - 表示两个变量之间的所有变量 (根据数据中变量的存放位置)。

比如, 数据文件中共有 20 个变量 var1、var2、…、var20, 则 var* 表示所有变量, var? 表示变量 var1、var2、…、var9, var1—var6 表示变量 var1、var2、…、var6。

对于时间序列数据 (必须进行过定义, 参见“时间序列分析”一章), 可以带有运算符。时间序列的运算符包括 (见表 1-1 所示):

表 1-1

运算符	含 义	例 子
L#. x	x 的 # 阶滞后。 $L3. x = x_{t-3}$	一阶滞后可直接写为 L. x
F#. x	x 的 # 阶超前。 $F3. x = x_{t+3}$	一阶超前可直接写为 F. x
D#. x	x 的 # 次差分。 $D2. x = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$	一次差分可直接写为 D. x
S#. x	x 的 # 阶季节差分。 $S12. x = (x_t - x_{t-12})$	一阶季节差分可直接写为 S. x

上述运算符不区分大小写, 可以重复, 也可以进行各种组合。比如, L3. x 也可以等价地表述为 LLL. x、LL2. x、L2L. x。

可以为单个数, 也可以为数列 (numlist, 参见 1.43 节的介绍), 表示对同一变量的多次运算。比如, L (1/3). GDP 表示 L. GDP、L2. GDP、L3. GDP。x 可以为单个变量, 也可以为多个变量, 表示对不同变量的同一运算。比如 L2. (GDP CS) 表示 L2. GDP、L2. CS。而 L (1/3). (GDP CS) 表示 L. GDP、L2. GDP、L3. GDP、L. CS、L2. CS、L3. CS。

1.4.2 区间

设定观测值区间的表达式见表 1-2 所示 (如果 # 为负数, 表示倒数第 # 个观测值)。

表 1-2

表达式	含 义	例 子
#	第#个观测值。	第 10 个观测值: in 10
#1/#2	第#1 到第#2 的观测值。	前 10 个观测值: in 1/10 或 f/10
#/1	第#个观测值到最后一个观测值 (last)。	后 10 个观测值: in -10/-1 或 -10/1
f/#	第 1 个观测值 (first) 到第#个观测值。	从第 21 到第 50 个观测值: in 21/50

1.4.3 数列

数列 (numlist) 即由一组数构成的序列, 经常被用于设定区间 (range), 基本的设定形式如表 1-3 所示。

表 1-3

表达式	含 义	例 子	注 释
#1 #2 #3 ...	数#1、#2、...。	2 0.5 1.8 2.9	2 0.5 1.8 2.9
#1/#2	从#1 到#2, 每步增加数为 1。	2/5	2, 3, 4, 5
#1 (#d) #2	从#1 到#2, 每步增加数为#d。	9 (-2) 1	9, 7, 5, 3, 1
#1 [#d] #2	等价于#1 (#d) #2。	9 [-2] 1	9, 7, 5, 3, 1
#1 #t to #2	从#1 到#2, 每步增加数为#t - #1。	1 4 to 13	1, 4, 7, 10, 13
#1 #t:#2	等价于#1 #t to #2。	1 4:13	1, 4, 7, 10, 13

上述形式可以进行各种组合, 比如, 1 2 3/5 8 (2) 12, 表示 1, 2, 3, 4, 5, 8, 10, 12。

1.4.4 日期数列

日期序列 (datelist) 即由一组日期构成的序列, 主要用于作图。基本的设定形式如下:

date1 date2 date3 ... <表示日期 date1、date2、...。>

date1 (#) date2 <从 date1 到 date2, 每次增加#日期单位。>

#1 (#d) #2 或者 #1 [#d] #2 <表示从#1 到#2, 每步增加数为#d。>

#1 #t to #2 或者 #1 #t:#2 <表示从#1 到#2, 每步增加数为#t - #1。>

比如下面的各个日期序列。

12jan1973 18jan1973 24jan1973

12jan1973 (6) 24jan1973 <表示 12jan1973, 18jan1973, 24jan1973。>

1998q1 (1) 2005q1 <表示 1998q1 ~ 2005q1 的所有季度。>

1998q1 (4) 2005q1 <表示 1998q1 ~ 2005q1 的第 1 季度。>

上述形式可以进行组合。比如,

5jan1973 12jan1973 (6) 24jan1973 <表示 5jan1973, 12jan1973, 18jan1973, 24jan1973。>

如果日期带有空格、斜杠或逗号, 则必须用括号括起来。比如,

(5 jan, 1973) (jan 9/1973) 12jan1973 (6) 24jan1973 <表示 5jan1973, 9jan1973, 12jan1973, 18jan1973, 24jan1973。>