



新世纪高等学校教材

经济管理实验教程

沈越 总主编

计量经济学实验教程

(赠光盘)

刘泽云 孙志军 编 著

Econometrics Using Stata



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

新世纪高等学校教材

经济管理实验教程

沈 越 总主编

计量经济学实验教程 (赠光盘)

Econometrics Using Stata

刘泽云 孙志军 编 著



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

图书在版编目(CIP)数据

计量经济学实验教程 / 刘泽云, 孙志军编著. —北京: 北京师范大学出版社, 2011.9

(经济管理实验教程)

ISBN 978-7-303-13235-5

I. 计·· II. ①刘··②孙· III. 计量经济学—实验—高等学校—教材 IV. ① F224.0-33

中国版本图书馆 CIP 数据核字(2011)第 154688 号

营销中心电话 010-58802181 58808006
北师大出版社高等教育分社网 <http://gaojiao.bnup.com.cn>
电子信箱 beishida168@126.com

J I L I A N G J I N G J I X U E S H I Y A N J I A O C H E N G

出版发行: 北京师范大学出版社 www.bnup.com.cn

北京新街口外大街 19 号

邮政编码: 100875

印 刷: 北京中印联印务有限公司

经 销: 全国新华书店

开 本: 170 mm × 230 mm

印 张: 12.25

字 数: 208 千字

版 次: 2011 年 9 月第 1 版

印 次: 2011 年 9 月第 1 次印刷

定 价: 22.00 元 (含光盘)

策划编辑: 马洪立 责任编辑: 高 玲

美术编辑: 毛 佳 装帧设计: 毛 佳

责任校对: 李 茵 责任印制: 李 啸

版权所有 侵权必究

反盗版、侵权举报电话: 010-58800697

北京读者服务部电话: 010-58808104

外埠邮购电话 010-58808083

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 010-58800825

前 言

当代经济学研究的一个普遍趋势是采用经验分析(empirical analysis)的方法,即建立数理模型表达理论构想,然后通过统计数据或抽样调查数据对变量之间的关系做出推断,从而证实或证伪理论构想。计量经济学(econometrics)是从事经济学经验研究的方法论基础,目前已成为高等院校经管类专业的核心课程,并越来越多地受到教师和学生的重视。

自2005年以来,我们在北京师范大学经济与工商管理学院为本科生讲授计量经济学,并为硕士研究生讲授计量分析方法与应用。目前有不少优秀的计量经济学理论教材,理论教材注重基本理论和基础知识,特别是数学推导,这有助于学生掌握计量经济学的理论框架,并有助于培养学生的逻辑思维能力。但在教学过程中,我们发现仅仅依靠理论教材会产生两个问题:其一,很多学生把计量经济学当做一门数学或数理统计课程,缺乏对计量经济学应用价值的直观感受,无法把计量经济学与实际的经济现象和经济问题联系起来;其二,过多的数学内容容易使部分学生(特别是本科学生)对计量经济学望而却步,失去学习兴趣。为此,我们努力尝试理论教学与实验教学相结合的模式,在强调理论知识的同时增加了计量分析软件应用的内容。我们发现,这样做有助于学生初步掌握计量分析的工具,理解计量分析的步骤与方法,在实际应用中增强对基本理论和基础知识的认识。同时,通过实际操作,学生感受到了将抽象的理论应用于分析实际问题的可能性,调动了学生的学习兴趣和积极性。因此,我们一直想编写一本与计量经济学理论教学相配套的应用教材,以方便教师教学和学生学习。呈现在各位读者面前的这

本教材，就是近年来我们不断探索的成果。

本教材的主要特色是：第一，目前国内的计量经济学实验教材绝大部分都是自成体系，与计量经济学理论教材缺乏联系，不利于教学工作的开展。而本教材与当前国内外最流行的计量经济学本科教材——《计量经济学导论》配套^①，紧扣教材内容和教学进度，对教师教学和学生学学习都极为方便。第二，目前国内的计量经济学实验教材大部分使用 Eviews 统计软件，而本教材使用的 Stata 统计软件在计量经济学中应用更为广泛，也更适合本科层次的初学者。本教材详细介绍 Stata 在计量分析中的初步应用，通过大量的例题帮助学生学学习 Stata 操作和深入理解计量经济学基本理论和方法，培养学生进行简单计量分析的兴趣和能力。第三，目前计量经济学教材中的例题和习题均为国外（特别是美国）的实例，缺乏与中国问题相结合的数据资源。本教材的实验大都采用中国的数据，增强了现实感。同时，附赠的光盘收录了所有实验使用的数据文件，为教学提供了便利。

本教材由 4 篇构成：第 1 篇介绍 Stata 的基本操作与描述性统计分析；第 2 篇介绍横截面数据的回归分析，包括简单回归分析、多元回归分析的估计与推断、多元回归分析的应用以及异方差性四章；第 3 篇介绍时间序列数据的基本回归分析；第 4 篇是计量经济学专题，涉及面板数据模型、工具变量法和两阶段最小二乘法以及限值因变量模型等三个专题。这些内容涵盖了本科层次计量经济学课程的基本要求，当然也可作为研究生学学习计量经济学应用的参考。

除了第 1 篇，每一章由三部分构成：知识要点、基本命令和实验。我们希望这样的安排有助于实验教学的开展。同时，我们建议计量经济学的实验教学与理论教学同步进行，这样有利于学生在应用中理解理论知识，并增强学生的方法论意识、应用能力和学学习兴趣。

除了第 1 篇，本实验教材宜与美国计量经济学家伍德里奇所著的《计量经济学导论》配套使用，各章的对应关系如下：

本书	《计量经济学导论》
第 3 章	第 2 章
第 4 章	第 3、4、5 章
第 5 章	第 6、7 章以及第 9 章的部分内容
第 6 章	第 8 章
第 7 章	第 10、11、12 章以及第 18 章的部分内容

^① [美]伍德里奇(J. M. Wooldridge):《计量经济学导论》，第 3 版，北京，中国人民大学出版社，2007。

第 8 章 第 13、14 章

第 9 章 第 15 章

第 10 章 第 17 章

本教材的撰写分工如下：

第 1 篇和第 4 篇：孙志军

第 2 篇和第 3 篇：刘泽云

刘泽云负责了全书的统稿工作。

本书的撰写及出版，要感谢北京师范大学经济与工商管理学院沈越教授和孙运传老师的大力支持与帮助，感谢北京师范大学出版社马洪立老师、高玲老师的辛苦劳动。书中的纰漏与不当之处，恳请读者批评指正。

计量经济学的初学者常常感到枯燥乏味，但当你通过 Stata 把抽象的计量模型转化为一个个具体的结果，而这些结果又推动着你进行更深入的理论和经验思考时，莫大的乐趣油然而生。希望读者和我们一样能够不断地享受这一过程。

作 者

2011 年 8 月

目 录

第 1 篇 Stata 基本操作与描述性统计分析

第 1 章 Stata 基本操作 /3

- 1.1 Stata 软件与相关资源 3
- 1.2 开始认识 Stata 4
- 1.3 开始使用 Stata 5
 - 1.3.1 Stata 文件 6
 - 1.3.2 命令和语句 7
- 1.4 数据管理 12
 - 1.4.1 数据浏览与变量定义 12
 - 1.4.2 数据文件的合并 13
 - 1.4.3 删除、生成或替代变量 14
 - 1.4.4 转换数据的排列方式 15
 - 1.4.5 使用函数 16
 - 1.4.6 设定条件 16
 - 1.4.7 内存管理 17
 - 1.4.8 错误提示信息 17

第 2 章 描述性统计分析 /18

- 2.1 变量与数据 18
 - 2.1.1 变量的类型 18
 - 2.1.2 数据的层次、类型与结构 19
 - 2.1.3 描述性统计分析的内容 21
 - 2.1.4 本章所用数据 21
- 2.2 单变量描述统计分析 22
 - 2.2.1 基本描述性统计分析 22
 - 2.2.2 计算收入不平等 24

2.3	定性—定性变量描述统计分析	26
2.4	定性—定量变量描述统计分析	28

第2篇 横截面数据的回归分析

第3章 简单回归分析 /31

3.1	知识要点	31
3.1.1	简单线性回归模型的定义	31
3.1.2	普通最小二乘法(OLS)	33
3.1.3	OLS估计量的性质	34
3.2	基本命令	35
3.3	实验	36
3.3.1	实验一 简单回归分析:教育对工资的影响	36
3.3.2	实验二 简单回归分析:学校投入对学生成绩的影响	43

第4章 多元回归分析:估计和推断 /45

4.1	知识要点	45
4.1.1	多元线性回归模型的OLS估计	45
4.1.2	OLS估计量的性质	46
4.1.3	推断	47
4.2	基本命令	54
4.3	实验	54
4.3.1	实验一 多元回归分析:工资方程	54
4.3.2	实验二 多元回归分析:学习努力程度对大学英语成绩的影响	57

第5章 多元回归分析:应用 /60

5.1	知识要点	60
5.1.1	标准化系数	60
5.1.2	多项式回归模型	61
5.1.3	含交互项的回归模型	61
5.1.4	预测	62
5.1.5	虚拟变量在回归分析中的应用	63
5.1.6	代理变量	66
5.1.7	测量误差	67
5.2	基本命令	68
5.3	实验	69

5.3.1	实验一 虚拟变量和多项式回归：工资方程	69
5.3.2	实验二 预测和标准化系数：学习努力程度对大学英语成绩的影响	76
5.3.3	实验三 代理变量的应用：教育收益率	79

第 6 章 异方差性 /83

6.1	知识要点	83
6.1.1	异方差性对 OLS 估计的影响	83
6.1.2	稳健性检验	83
6.1.3	对异方差性的检验	84
6.1.4	加权最小二乘估计	86
6.2	基本命令	88
6.3	实验	89

第 3 篇 时间序列数据的基本回归分析

第 7 章 时间序列数据的基本回归分析 /99

7.1	知识要点	99
7.1.1	有限样本条件下时间序列数据模型的 OLS 估计	99
7.1.2	平稳性和弱相关	100
7.1.3	大样本条件下时间序列模型的 OLS 性质	101
7.1.4	平稳性检验和非平稳时间序列的 OLS 估计	102
7.1.5	序列相关的检验和校正	104
7.1.6	时间序列模型中的异方差性	107
7.2	基本命令	108
7.3	实验	110
7.3.1	实验一 平稳性检验和非平稳时间序列的回归分析： 生育方程	110
7.3.2	实验二 序列相关的检验与校正：波多黎各的就业率 ..	118
7.3.3	实验三 时间序列模型中的异方差性：股票收益率 ..	123

第 4 篇 计量经济学专题

第 8 章 面板数据模型 /131

8.1	知识要点	131
8.1.1	面板数据的排列结构	131
8.1.2	基本模型	133

8.1.3	固定效应模型	134
8.1.4	差分模型	135
8.1.5	随机效应模型	135
8.1.6	FE、RE、FD 的比较与选择	136
8.2	基本命令	137
8.3	实验	137

第9章 工具变量估计与两阶段最小二乘法 /142

9.1	知识要点	143
9.1.1	工具变量(IV)与两阶段最小二乘法(2SLS)	143
9.1.2	检验工具变量的有效性	144
9.1.3	工具变量的构建	147
9.2	基本命令	148
9.3	实验	148

第10章 限值因变量模型 /153

10.1	知识要点	154
10.1.1	Probit 与 Logit 模型	154
10.1.2	Tobit 模型	160
10.1.3	截取回归模型	161
10.1.4	样本选择纠正	162
10.2	基本命令	164
10.3	实验	165
10.3.1	实验一 二元 Probit 模型：影响农村居民外出打工的因素研究	165
10.3.2	实验二 多元 Logit 模型：教育对职业选择的影响	171
10.3.3	实验三 有序 Probit 模型：工作满意度的影响因素	174
10.3.4	实验四 Tobit 模型：劳动供给时间	175
10.3.5	实验五 Censored 模型：儿童在校持续期的分析	176
10.3.6	实验六 样本选择偏差纠正：教育收益率估计	178

附录 数据文件索引 /181

参考文献 /182

北师大版经管类高校教材书目(赠教学课件) /183

教师用免费教材样本申请表 /185

第 1 篇 Stata 基本操作与 描述性统计分析

第1章 Stata 基本操作

1.1 Stata 软件与相关资源

Stata(Data Analysis and Statistical Software)是由美国计算机资源中心(Computer Resource Center)研制,一个用于分析和管理数据的功能强大又小巧玲珑的实用统计分析软件。从1985年至今20多年间,已连续推出了十几个版本,最新的是Stata 11。本书所使用的版本是Stata 10。随着版本的不断升级,其功能也越来越多。

目前,Stata 软件尤其受到从事经济学经验研究或计量分析的研究者的喜爱。这是因为,几乎所有的计量经济学中所提到的方法,在该软件中都可以找到相应的命令,而且新的计量方法一旦被验证较为成熟,就有人将该方法写成程序,并对命令附有详细的说明。在大部分命令的帮助文件中,甚至列出了该命令源自哪位经济学者的哪篇文献中提出的计量方法。由此可见,该软件简直就是为经济学研究的人士定做的。

对于一名Stata 的使用者来说,首先要消除对它的陌生感。你可以花很短的时间——可能是两三个小时,甚至是几十分钟——就可以初步掌握该软件。你会发现,就如同会操作Windows、Word、Excel或其他计算机软件那样,操作Stata 软件竟然与操作这些软件同样简单。但是,这需要一个前提条件:你必须首先掌握数据描述统计分析和计量经济学知识,Stata 软件仅仅是一个帮助你完成无法用手工计算完成的复杂的数据分析的工具而已。

以下是学习和熟练使用Stata 软件的一些经验和常用资源:

①养成查看Stata 软件本身自带的资源——帮助(help)文件的习惯。帮助文件会介绍命令的使用方法,并且还有案例供你参考。记住,没有人能够完全将Stata 的各种使用方法和命令记在脑中,帮助文件就如同一本字典,需要的时候可以随时翻阅它。

②查看Stata 官方信息,可以访问它们的网址:<http://www.Stata.com>,或Stata 出版社的网站:<http://www.Stata-press.com>。

③国内也有一些网站有Stata 的许多资源,比如人大经济论坛(<http://www.pinggu.org>)、中国经济学教育科研网(<http://www.cenet.org>)等。这些网站上有众多关于Stata 最新信息、Stata 使用方法的讨论。

④当一个新的Stata 命令出来后,现有的Stata 版本可能不含有该命令,

这时你可以到网络上搜寻并更新到软件中。比如在衡量收入不平等时，有一个指标是 Theil 系数，当你有了一些收入数据，想计算出该系数时，你可以在 google 里搜索“theil index Stata module”，就会列出一个命令“inequal7”，以帮助你完成计算。该命令如何更新到 Stata 软件中呢？可以在 Stata 命令窗口键入如下命令：`ssc install inequal7`，如果你的计算机网络可以访问国外网站，那么 Stata 就会自动安装上该命令。

⑤当你遇到某个计量方法在 Stata 中应该用什么样的命令的问题时，除去上面借助网络资源，还可以多与别人交流，并询问，然后去看该命令的帮助文件。

⑥最后一点是，学习使用 Stata 软件应养成一个良好的习惯，那就是积累。建立一个专门的文件夹，按照计量方法将新了解到又容易忘记的 Stata 的命令整理并归类，形成自己的“Stata 小词典”。

建议你初步学会使用 Stata 软件之后，再回过头来阅读一下上面的内容，说不定你也能总结出几条学习经验供大家分享。

1.2 开始认识 Stata

在 Stata10 文件夹中，找到下面这个图标，然后用鼠标双击它。

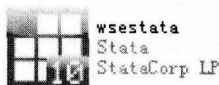


图 1-1 Stata 图标

计算机屏幕上就会出现 Stata 的界面，如图 1-2 所示。你的大部分工作都将在这个界面上完成。

Stata 的界面窗口由这几个部分构成：

- ①Command(命令，右下部分)窗口：用于向 Stata 输入命令；
- ②Stata results(结果，右上部分)窗口：显示运行结果；
- ③Variables(变量，左下部分)窗口：显示数据文件的变量名；
- ④Review(左上部分)窗口：显示已执行过的命令；
- ⑤窗口上方是工具栏，列出了常用的管理文件和数据的命令，如打开文件、存储文件、打印文件、数据管理窗口等；
- ⑥工具栏上方是菜单栏。

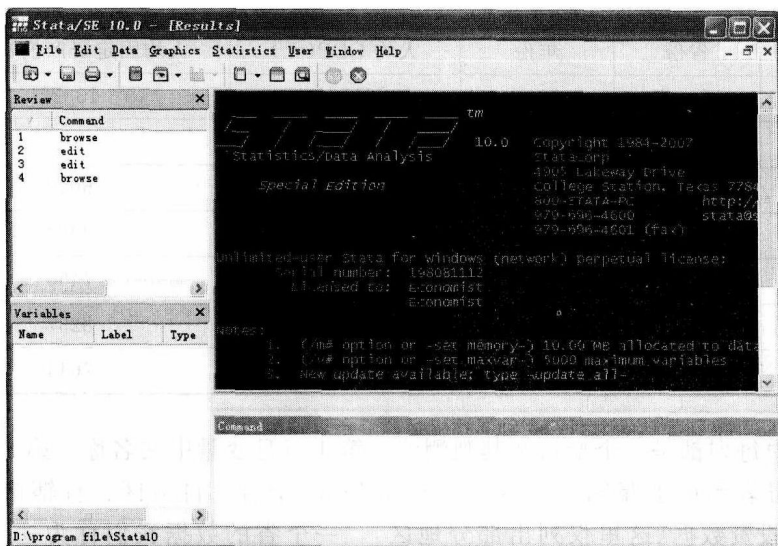


图 1-2 Stata 界面

1.3 开始使用 Stata

开始使用 Stata，首先需要了解的是文件、数据和命令。我们用一个具体的例子来说明如何使用它。

假设你打算做一项有关中国投资水平与人均 GDP 之间关系的研究，你收集到 2006 年全国各省人均 GDP 和人均固定资产投资(代表投资水平)的数据，该数据存储在 Excel 文件“中国经济与投资省级数据 2006—07.xls”中。其中 2006 年的数据如下所示。

表 1-1 中国经济与投资省级数据(2006 年)

省代码	省份	年份	人均 GDP(元)	人均固定资产投资(元)
code	prov	year	rjgdp	rjcap
11	北京	2006	50467	20850
12	天津	2006	41163	16935
13	河北	2006	16962	7930
14	山西	2006	14123	6684
15	内蒙古	2006	20053	14031

续表

省代码	省份	年份	人均 GDP(元)	人均固定资产投资(元)
21	辽 宁	2006	21788	13322
...
61	陕 西	2006	12138	6642
62	甘 肃	2006	8757	3924
63	青 海	2006	11762	7455
64	宁 夏	2006	11847	8257
65	新 疆	2006	15000	7644

表中每列都是一个变量及其观测值。第 1 行是变量中文名称，第 2 行是用英文字母表示的变量名，第 3 行及以下是 31 个省、自治区、直辖市的人均 GDP 与投资数据(这里仅列出部分地区)。一个省的数据我们称为一个观测，这样，以上数据集共有 31 个观测(或样本量为 31)。该数据集共有 4 个变量：省代码、省份、年份、人均 GDP、人均固定资产投资。省代码变量用以区别每个省。

上面的数据是分析用的原始数据，用 Stata 完成你的分析，要做的工作主要包括：

- ①把原始数据转换成 Stata 格式的数据文件；
- ②根据你的任务写出 Stata 命令；
- ③Stata 命令运行后看懂分析结果，并把分析结果保存下来，或者写入你的文本文件(如 Word 文件)。

上述工作涉及 Stata 的文件、数据与命令。

1.3.1 Stata 文件

Stata 文件主要有数据文件、do-file 文件、log 文件和程序文件。

1. 数据文件

即用来保存数据的文件，是 Stata 处理和分析的对象，以扩展名“.dta”表示。数据的结构包括变量(每一列)和观测(每一行)。

2. Do-file 文件

是用来保存你的命令的文件，以扩展名“.do”表示。该类文件的好处是把你执行过的命令保存下来，以便以后你再重新执行这些命令。

3. Log 文件

是用来直接保存分析结果的文件，以扩展名“.log”表示。这类文件可以直接用文本编辑程序(如记事本、Word 等)直接打开。

4. 程序文件

是 Stata 的命令的程序，以扩展名“.ado”表示。除非你自己要编写某个计量方法的 Stata 程序，一般来说，这类文件对于仅限于应用 Stata 做数据分析的人很少用到。

1.3.2 命令和语句

命令和语句是用 Stata 做数据分析的指令，它十分简单和灵活。要想做一个计量分析，首先要知道它在 Stata 中的基本命令名称是什么。如要描述某个变量的基本统计信息(平均值、最大值、最小值等)，可以用的基本命令就是“summarize”。然后将你的任务写成一个语句。你对数据分析的所有要求都由命令语句来完成。命令语句的基本形式如下：

```
summarize varlist [if exp] [in exp] [weight] [, options]
```

summarize 是基本命令，大部分基本命令都有简写形式，如该命令的简写就是 su；varlist 是变量，实际命令语句里指数据文件中具体的变量名；[if exp]和[in exp] 是逻辑条件，你可以用它来限定你分析的观测是哪个范围内的；[weight]表示加权，加权的方式有多种，可以查看 help 文件关于加权的说明，在实际命令语句中必须加上[]；[, options]是选项，可以用来对你想要的结果进行定义。

了解了 Stata 的一些基本内容之后，下面我们以完成本节提出的研究任务为例，给出一个具体的 Stata 基本操作实例。目的是研究中国人均 GDP 与投资的关系。

要对数据进行分析，首先需要形成 Stata 数据文件。

形成 Stata 数据文件的方式有多种，最简单的是在 Stata 窗口中直接录入数据，这种方式虽然简单，但并不是特别实用。经常用的其他方式是先在其他操作软件中将数据编排好，然后导入到 Stata 中。假设上述数据已经被整理成 Excel 形式的文件，这时你只需将数据连同变量一并选取复制，然后打开 Stata 数据编辑窗口(Data Editor)粘贴就可以。具体操作如下：

①复制 Excel 窗口中的数据和变量。