


21 世纪
人口学系列教材

数据管理与模型分析： STATA 软件应用

• 杨菊华 / 著

 中国人民大学出版社

图书在版编目 (CIP) 数据

数据管理与模型分析: STATA 软件应用/杨菊华著. —北京: 中国人民大学出版社, 2012. 7
21 世纪人口学系列教材
ISBN 978-7-300-16163-1

I. ①数… II. ①杨… III. ①数据管理系统-应用软件-高等学校-教材 IV. ①TP317

中国版本图书馆 CIP 数据核字 (2012) 第 172811 号

21 世纪人口学系列教材

数据管理与模型分析: STATA 软件应用

杨菊华 著

Shuju Guanli yu Moxing Fenxi; STATA Ruanjian Yingyong

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)	010-62511398 (质管部)	
	010-82501766 (邮购部)	010-62514148 (门市部)	
	010-62515195 (发行公司)	010-62515275 (盗版举报)	
网 址	http://www.crup.com.cn http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京东君印刷有限公司		
规 格	185 mm×230 mm 16 开本	版 次	2012 年 8 月第 1 版
印 张	26.5	印 次	2012 年 8 月第 1 次印刷
字 数	486 000	定 价	45.00 元

版权所有 侵权必究 印装差错 负责调换

《21 世纪人口学系列教材》编委会

主 编 翟振武
委 员 杜 鹏 段成荣
姚 远 刘 爽
刘金塘



21世纪人口学系列教材

总序

中国是世界第一人口大国。无论是在宏观上，还是在微观上，中国社会经济的发展都与人口的状况和问题紧密相关。在中国追求现代化的历史过程中，人口问题始终是我们长期面临的重大问题。人口问题的复杂性世所公认，但中国人口问题可能更复杂。一个多世纪以来，人口问题几乎成了所有经济学家、社会学家和政治家最为关注的热门话题之一。尽管如此之多的人士研究人口问题，但是近百年来，在对中国人口问题的认识上，却经历了反复的曲折，积累了无数的教训。

对一个事物的正确认识，需要有科学的理论和科学的方法。中国人口问题如此重大，如此复杂，任何人要想深入了解、分析和研究人口问题，必须要掌握基本的人口学理论与方法。20世纪70年代中期，中国开始了具有历史意义的现代人口学学科的创建过程。人口学教材建设从一开始就被列为学科创建的最重要工作之一。到20世纪80年代中期，中国的大学里已经系统地建立了完整的人口学学士、硕士和博士的培训课程，出版了一系列与国际前沿接轨的人口学教材，如《人口统计学》、《人口经济学》、《世界人口》、《人口理论教程》、《人口思想史》、《人口社会学》、《人口规划与预测》、《人口地理学》、《现代人口分析技术》、《人口普查的分析方法》等，并培养了一大批人口学的专门人才。从那时到现在，国际上人口科学的发展日新月异，新方法、新理论、新流派层出不穷。与此同时，中国人

口的研究也迅速向深度和广度推进。收集到的人口数据堆积如山，人口数量、结构和质量，人口与经济，人口与社会，人口与环境资源等各种问题几乎全部纳入了研究视野。但从20世纪90年代初期以来，人口学教材的编写高潮却逐渐回落。反映国际最新进展并与中国人口实践紧密结合的新教材，无论是在种类上还是在水平上，都远远不能满足现实的需要。

中国人民大学人口学系（所）是国内最早开设人口学学位教育的机构，曾经培养过上千名人口学专业的本科生、硕士生和博士生，20世纪80年代初编写的《人口统计学》、《世界人口》、《人口理论教程》等著作曾在人口学界产生了重大而深远的影响。进入21世纪后，为了满足人口学教育的需要，在教育部和中国人民大学出版社的大力支持下，中国人民大学人口学系（所）决定牵头组织编写一套能够反映国际国内人口学最新进展的《21世纪人口学系列教材》。系列教材将涵盖人口学主要分支学科的内容，为各个大学和培训机构开设人口学公共课和专业课提供科学、准确的教学资料。

尽管本套教材是在长期教学实践基础上编写的，编写过程中我们也力争逻辑严谨、内容充实、形式活泼，但疏漏在所难免，望广大读者指正，以便再版时修改。

翟振武

关联课程教材推荐

书号	书名	作者	定价
978-7-300-15779-5	《社会调查教程》(精编本)	江立华 水延凯	29.80 元
978-7-300-10738-7	《社会学研究方法》(第三版)	风笑天	39.80 元
978-7-300-10912-1	《社会调查研究方法》(第二版)	郝大海	22.00 元
978-7-300-02898-5	《社会统计分析方法: SPSS 软件应用》	郭志刚	39.00 元
978-7-300-09639-1	《STATA 实用教程》	王天夫 李博柏	49.80 元
978-7-300-07564-9	《社会研究方法: 定性和定量的取向》(第五版)	[美] 劳伦斯·纽曼	89.80 元

配套教学资源支持

尊敬的老师:

衷心感谢您选择使用人大版教材!

秉承“出教材学术精品, 育人文社科英才”的出版理念, 我社为教材打造配套教学资源, 帮助老师拓展教学思路, 革新教学方式。相关的配套教学资源, 请到人文分社网站 (www.crup.cn/rw) 下载, 或随时与我们联系, 我们将向您免费提供。联系人信息:

地址: 北京海淀区中关村大街 31 号 211 室 龚洪训 收 邮编: 100080

电子邮件: gonghx@crup.com.cn 电话: 010-62515637 QQ: 6130616

欢迎您随时反馈教材使用过程中的疑问、修订建议等, 让我们与教材共成长。

建议一经采纳, 即有好书奉送。

如有相关教材的选题计划, 也欢迎您与我们联系, 我们将竭诚为您服务!

选题联系人:

电子邮件:

电话:

潘宇

pany1234@gmail.com

010-62515634

宋义平

songyp@crup.com.cn

010-62511866

俯仰天地 心系人文

www.crup.cn/rw

中国人民大学出版社 人文分社网站

欢迎登录浏览, 了解图书信息, 下载教学资源



目录

21世纪人口学系列教材

导 言 1

第一部分 数据质量控制与管理

第一章 数据的基本结构 11

- 第一节 数据结构的基本类型 12
- 第二节 数据的测量水平 27
- 第三节 数据的测量与模型的选择 32

第二章 数据的缺失及其处理 37

- 第一节 数据缺失的机制及检测 38
- 第二节 缺失数据的处理方法 46
- 第三节 几种填补方法的实证演示 49

第三章 数据的极值及其处理 57

- 第一节 极值概述 58
- 第二节 极值的判断方法 63

第三节	极值的应对方法	70
第四章	数据的简化与因子分析	91
第一节	因子分析的主要目的及基本概念	92
第二节	因子分析的类型及原理	100
第三节	公因子提取方法	102
第四节	因子旋转	110
第五节	因子结构及因子分析的基本假定	113
第六节	因子分析应用举例	116

第二部分 常规回归分析模型

第五章	连续数据与 OLS 模型	141
第一节	相关分析	142
第二节	线性回归的基本原理	154
第三节	利用 Stata 进行线性回归分析	177
第六章	分类数据与 logistic 模型	223
第一节	logistic 回归概述	224
第二节	二元 logistic 回归的应用	243
第三节	序次 logistic 回归	259
第四节	多项式 logistic 回归	274

第三部分 前沿回归分析模型

第七章	纵向数据与固定效果模型	295
第一节	固定效果模型简介	296
第二节	固定效果模型的基本原理	301
第三节	线性固定效果模型的演示	305
第四节	线性固定效果模型的应用	312
第五节	非线性固定效果模型的应用	326

第八章 多层结构数据与多层模型	332
第一节 多层模型简介	334
第二节 多层线性模型的基本原理	339
第三节 多层线性模型的应用	350
第四节 多层非线性模型的应用	364
第九章 计数数据与相关模型	373
第一节 泊松分布及泊松模型	374
第二节 负二项式模型	386
第三节 零膨胀模型	394
第四节 零截断模型	401
第五节 四类模型比较	405



导 言

社会现象极其复杂，事物之间彼此联系。这些联系有其内在机制，且以多样的形式而存在。社会科学研究的目的是要发现这些联系，解释影响的机制，找出复杂现象背后的一般性规律。合适的研究理论、科学的研究方法、系统的研究步骤有助于更好地发现事物之间关联的显性和潜在规律。西方社会科学研究积累了大量的理论和方法，而我国的社会科学研究，由于量化水平的偏低，仍旧处于规范研究阶段，侧重于理论思辨和逻辑推理；以事实为依据的实证研究才刚刚起步。

不过，近20年来，随着西方定量研究方法以及各种统计软件的传入，定量研究得到了越来越多研究者的重视，国内也掀起了一股定量研究的热潮（特别是在年轻学子之间）。但是，不少研究者受制于自身的统计学背景，或受制于数据管理能力和统计分析水平，或受制于学科理论和实践知识的积淀，或同时受上述诸多方面的制约，尽管在过去数年中，我国学者的定量研究能力有了很大的改善和提高，但依旧停留在较低的水平阶段，在实际教学和研究中，主要存在以下几个问题。

其一，研究者具有深厚的统计学理论知识，但实际数据处理和挖掘能力却欠佳。这样的研究者往往十分重视，且熟知统计理论、数学推导，然而，一旦遇到

实际的数据分析问题（特别是规模庞大、内容复杂的大型数据），他们往往或心生恐惧、望而却步，或束手无策、望“数”兴叹，出现理论和实践相脱节的现象。

其二，研究者具有一定的数据处理技术，却由于统计理论的欠缺而难以产生更高质量的研究成果。这是与第一种情形恰恰相反的另一理论与现实相脱节的现象。的确，随着大规模数据的越来越多，随着定量研究的日益流行，越来越多的人或多或少地掌握了一些常见统计软件（如 SPSS、SAS、Stata）的应用。然而，从其研究中不难发现，许多人对一些统计程序的基本原理、适用范围、优势与不足等问题都不甚了了，应用错误屡见不鲜。

其三，研究者具有深厚的理论知识，却不知如何利用数据检验自己的理论。这是第三种理论与现实相脱节的现象，但这里的理论指的是某个研究领域的具体理论，而不是统计理论。研究者具有很好的研究理论、研究假设，却不知选择何种合适的方法处理、挖掘、分析数据，并检验理论。目前，社会科学领域的绝大多数研究属于这一类型。

其四，研究者的理论（包括统计学理论和具体研究领域的理论）水平和数据处理能力都比较有限。

上述局限在很大程度上制约着我国学者的研究水平，使我们的研究往往停留在低水平的描述性分析或简单的模型分析上，并因此极大地阻碍了国内学者与国际同行接轨的步伐和层次。如何应对这些挑战、提高实际研究水平是学者们十分关心、关注的问题。显然，要避免上述局限、产出更高质量的研究成果，研究者不仅要掌握一定的统计学原理，拥有相应的研究理论，而且要具备实际的数据挖掘、管理、分析能力；同时，只有研究者具备了整合研究理论与软件应用的能力，才能最终达到其研究目的。

一、本书的写作目的和读者对象

本书是一本整合数据挖掘、管理和模型分析及相应的 Stata 统计软件应用于一体的教科书。随着国内社会科学领域定量研究的建设和发展，Stata 用户迅速增加，但是配套的教材却很稀缺，仅有的几本也是为了 Stata 的推广与应用，数据管理的基础性较弱，模型分析主要注重常规性模型，一些重要、实用、前沿的数据处理方法和挖掘技术涉及不多。针对目前国内 Stata 高级数据管理与模型分析方法教材的缺失现象，笔者编写了这本《数据管理与模型分析：Stata 软件应用》。

1. 本书的写作目的

针对上述局限，本书强调数据的质量控制、数据的处理与管理、统计理论与数据挖掘的结合以及研究理论与研究方法的整合。基于这些思路，本书的主要写

作目的如下：

- 其一，架起统计学理论和软件应用之间的桥梁；
- 其二，架起研究思路与数据管理和挖掘之间的桥梁；
- 其三，架起研究理论和数据分析方法之间的桥梁。

其最终的目的在于：有效地帮助读者在掌握基本统计原理的基础上，改善数据挖掘、管理、分析的能力，提高读者的整体研究水平，从而为产出高质量的研究成果奠定坚实的基础。

2. 本书的读者对象

本书不区分读者的学科背景。由于本书所涉及的数据管理、统计理论与技术、使用的 Stata 统计软件都不是给特定用户的，故本书对读者的学科背景并无特殊界定，任何在实际工作中需要使用数据处理与统计分析的学生、学者、研究者以及其他相关人员都可以通过本书进入数据管理与挖掘领域中。

但是，本书区分读者的统计知识和研究背景层次：它是一本适于研究生层次（或相当水平）的教材，旨在针对具有较好统计学基础的博士生、硕士生、高年级本科生以及其他相同水平的研究人员。

当然，这并不是说低年级的学生就不可以或不能参考本书了。数据管理知识和常规的线性回归模型是所有层次的研究者都必须掌握的，如果低年级的学生希望自己的研究更为规范、分析结果更为可信的话，也可以参考本书的这些内容。

对于本科生和其他缺乏基本统计知识、没有接触过 Stata 统计软件的读者而言，可以首先使用笔者的另一本教材：《社会统计分析 with 数据处理技术：Stata 软件的应用》——在这本旨在针对本科生的教材的基础上，再来学习研究生教材，将会极大地提高学习和研究效能。

二、本书的结构

全书共分为九章，主要内容包括三个部分。第一部分：数据质量控制与管理；第二部分：常规回归分析模型；第三部分：前沿回归分析模型。具体安排如下。

1. 数据质量控制与管理

定量研究人群中往往存在这样两类人：其一，一旦拿到数据，不管数据的结构，也不问其质量如何，就马上着手进行模型分析，恨不得一天之内做完整个数据分析。因为人们往往认为，所谓定量研究，不过就是做出几个模型而已。其实，这种看法是非常错误的，且如果数据质量本身存在问题的话，就可能产生十分严重的后果——它可能会得出错误的参数估计，从而误导他人，产生不正确的社会影响。模型越精确，他人受到误导的可能性就越大。其二，有的人虽然知道数据

的质量存在问题，却不知道如何处理。还有不少严肃的学者希望规范自己的研究，改进研究质量，但却苦于不知从何着手。对于后一种人，需要提高他们进行数据质量控制的能力；而对于前一种人，先需要加强其数据质量控制的理念，然后改善他们在这方面的能力。本书的这一部分就希望能起到这样的作用。

第一章 数据的基本结构。数据是对客观事物的符号表示，是统计描述和分析的载体，而数据结构是数据的存在形式。只有正确地了解了数据的结构及其测量水平，才能选择合适的数据处理方法和模型分析技术。但是，对于数据结构重要性的认识，学界还停留在刚刚起步阶段。本章描述数据结构的基本类型（如横向数据、纵向数据、回顾性数据和多层结构数据等）及变量的测量水平，介绍每种类型的优势与局限，阐释数据结构和类型的意义与作用，并利用 Stata 软件介绍实例加深读者对各种数据类型的认识。本章是本书后面模型分析章节的基础。

第二章 数据的缺失及其处理。数据分析结果的稳健性和可信性在很大程度上取决于数据的质量；很多可控和不可控的因素都会影响到数据的质量，但这恰恰是被研究者最经常有意或无意地忽视的问题。本章考虑影响数据质量的因素之一：缺失值。首先，描述缺失数据产生的机制，分析数据缺失对分析结果可能产生的偏差；同时，从理论上介绍数种处理缺失数据的方法，比较不同方法的优势和不足；然后，利用社会调查数据，借助实例分析，比较各种填补缺失值方法的异同。

第三章 数据的极值及其处理。数据分布中一些不符合数据一般特征的孤立点或极值普遍存在于调查数据中。极值是与正态分布相悖的一个概念，极值越多，异常程度越深，数据的偏态分布就越严重，犯 I 类错误或 II 类错误的概率就越大。本章介绍数据的极值及其应对方法，具体包括以下内容：极值产生的原因及后果，极值的类型和判别方法，极值的应对方法及各自的优势和局限，剔除极值与利用某种方法修正极值这两种选择的比较。本章还会通过实例，演示、比较不同应对极值方法的差异，借此唤起更多研究者对这一往往被忽视却十分重要的主题的关注。

第四章 数据的简化与因子分析。在社会科学、心理学及公共卫生等许多学科领域，学者往往利用许多不同的变量测量同一个问题，这就需要在实际数据分析中，对众多的变量进行整合，生成少数变量，从而使进一步的数据分析更为简明、合理。其中的方法之一就是因子分析。本章首先描述因子分析法的类型及原理，其次介绍提取公因子的方法，描述因子旋转及因子结构，再次通过实例进一步探讨因子分析法在实际研究中的应用，最后介绍因子分析的一些注意事项。

2. 常规回归分析模型

挖掘、处理数据的目的是为了更好地了解数据，并在控制其他干扰因素的情

况下，得出研究者最关心的一些因素之间的独立关系。但不同的研究问题、目的和数据结构需要采用不同的模型技术。在一般情况下，模型的选择主要是基于因变量的衡量方式，而如何衡量因变量则取决于我们的研究数据和研究问题；数据结构和类型决定模型选择，而不是相反。本部分根据不同的数据类型介绍几种社会科学领域最为常用的模型分析方法。

第五章 连续数据与 OLS 模型。本章介绍最基本的线性回归模型，具体内容包包括相关与回归、假设检验、逐步回归、多项式回归、回归诊断、拟合曲线和稳健回归等，但重点关注回归模型基本假定的原理、具体内涵、违反每个假定的后果，强调对回归结果的解读与阐释，突出如何借助模型分析结果进行预测，以及各种事后检验，使读者不仅掌握如何进行线性回归分析，而且具有判断模型分析结果好坏的能力。

第六章 分类数据与 logistic 模型。本章与第五章的不同之处在于，因变量为分类数据。在一般情况下，分类变量不适宜采用 OLS 模型，而需要使用 logistic 模型或其他最大似然模型。同时，分类数据又包括二分类、序次分类、多分类三种情况，每种情况所对应的模型也各有差异。本章针对不同类型的分类变量，分别进行理论介绍及实际应用举例，具体安排类似第五章。

3. 前沿回归分析模型

社会现象是复杂多变的，随着研究者对各领域社会现象认识的不断深入和全面，一些常规性统计分析模型的适应性受到很大局限，支撑这些模型的一些基本假定往往由于研究问题的复杂性而被违反，这就要求研究者使用更合适的技术、采用更合理的模型分析实际数据，应对社会问题。

第七章 纵向数据与固定效果模型。准确衡量变量之间的因果关系是研究者面临的重大挑战。利用纵向数据进行固定效果模型分析可以较好地应对该挑战，得到更精确的参数估计结果。本章主要有三个目的：（1）介绍固定效果模型的基本原理及工作机制；（2）利用 Stata 统计软件，借助实际调查数据，描述线性固定效果模型和二元 logistic 固定效果模型的具体用法；（3）比较固定效果模型与随机效果模型的适用场合，从而更好地把握固定效果模型的优势及不足。

第八章 多层结构数据与多层模型。大规模的社会调查数据往往具有多层结构的特点，该特点使得传统的常规性模型的分析结果可能低估标准误，从而增加犯 I 类错误的可能性，而多层模型可以有效应对该局限。本章介绍多层模型的优势与适应性；描述多层线性模型（即线性增长模型）与非线性模型的技术原理；并利用 Stata 软件，通过对实际数据的分析，演示该模型的具体操作和应用，比较多层模型和常规模型分析结果的异同。

第九章 计数数据与相关模型。在社会科学的实际研究中，许多因变量的取

值范围是有限的，如二项式分布；因变量只能取正值，且介于一定的区间；因变量为非零、非负值的计数变量。本章主要关注第三种情况，而这种情况又可以细分为常规的计数数据、零膨胀数据（即因变量具有过多 0 值）、零截断数据（即计数的因变量完全没有 0 值）。本章的教学目的是，了解泊松模型、负二项式模型、零膨胀模型和零截断回归模型的基本原理，并通过实例分析，掌握它们在 Stata 软件中的应用。首先，介绍泊松分布和泊松模型的基本原理以及泊松模型的应用；其次，讲解负二项式模型的应用；再次，简要描述零膨胀模型的应用场合以及具体应用；最后，简要讲解零截断模型的应用。

三、本书的特色

如前所述，本书的主要目的在于架起研究理论与实际数据分析之间的桥梁，提高读者数据挖掘、管理和分析的动手水平，改善学生和定量研究者的定量研究能力。为了更好地起到桥梁作用，本书同时注重数据管理和数据分析。在全书的九章内容中，前面的四章关乎数据的质量控制，涉及高级数据的挖掘和管理技术；中间四章属于常规模型；最后 3 章是目前国内学者使用较少的前沿模型。这样的结构安排和内容选择使得本书既不同于现有的统计学教材，也有别于 Stata 软件（或其他诸如 SPSS、SAS 等统计软件）的使用教材。相反，它旨在在研究理论的指导下，介绍怎样使用 Stata 统计软件，恰当地完成一系列的数据处理与挖掘工作。因此，每一章的主题由研究问题驱动，对 Stata 命令的解释也是基于理论的需要和指导。

具体而言，本教材具有如下特色。

1. 视域开阔，针对性强

过往的统计学或 Stata 应用书籍多是从单一的统计学、计量经济学、医学等视角切入，范围比较狭窄，难以适用于其他专业的读者。本书以社会科学研究为基础，从更广的视角探讨 Stata 在众多学科领域中的应用，使更多的读者不仅既掌握必要的基本统计原理，又能了解并使用 Stata 工具软件进行数据处理和模型分析。

但是，本书不是针对所有层次的读者，而是针对具有初步的统计学理论知识、拥有一定的研究背景、掌握了基本的 Stata 软件知识的读者而编写的。其目的不在于扫盲，而在于针对目前社会科学及其他学科领域研究中存在的一般性问题，从多个途径入手，改善读者理论联系实际的能力，提高读者的定量研究水平。

2. 强调数据质量控制，突出数据管理能力

一个定量研究质量的高低往往不是看其使用了什么模型，而是看其对所用的数据是否有全面的把握，对数据的质量是否了然于心，对数据的处理是否得当，对模型的选择是否合适（例如，是否可以很好地检验理论假设）等。因此，了解

数据的局限，并在现有条件下采取适当的措施，是最大限度地应对这些不足、提高研究质量的保证。然而，同类教材很少涉及（或稍有涉及）这方面的内容，包括了解不同数据的结构与类型，如何填补数据的缺失值，如何有效地应对极端值，如何简明地整合数据等。因此，我国不少定量研究者都比较缺乏这些能力。

但是，这些都是数据管理和挖掘的基本技能，特别是对具有一定研究基础的人而言。事实上，这些技能也是体现一个定量研究者基本功的标准以及衡量某个研究质量的关键。本书详细介绍了几种高级数据质量控制和管理的方法、基本原理、分析步骤、注意事项等内容，不求全面，但求实用，以便有效地提高读者处理数据的能力和效率，为掌握高层次模型分析技术打下坚实的基础。高级数据处理技术是本书的重点内容，也是本书最具特色之处。

3. 关注基本统计理论，强化实际操作能力

只会做几个模型却不清楚为什么要做这些模型的研究者是难以产出高质量研究成果的。事实上，统计软件的推广使得模型分析变得十分便捷，即使是非常复杂的模型。但是，一个高质量的定量研究要求研究者掌握一定的统计学理论，并在该理论的指导下进行模型选择。因此，本教材对书中所包含的每一种模型都会比较详细地介绍其基本统计原理、主要假定、违反假定的后果、模型的适应性及需要注意的问题。但是，考虑到多数读者没有统计学背景这一特点，本书对统计原理的解释也不会像一般的统计学教材那样深奥，而是强调基本、实用、易懂；同样，对于每一个数据处理步骤，书中都会详细介绍其思路和缘由；对于每一个新出现的 Stata 命令和主要选项，书中也会对其用法逐项标注、仔细讲解，力争使命令简单易懂，读者乐于接受，知识掌握扎实。

这些特点使得本书既不同于一般的统计学原理教材，也与其他 Stata 应用教材明显有别。它一方面可以使读者在轻松愉快的心境下掌握一些必需的统计理论知识，避免知道运行一些模型却不知为何运行这些模型的局限；另一方面也能使读者在阅读了本书后，能尽快使用 Stata 软件进行数据处理与分析，从而达到理论与方法的有效结合。

4. 介绍常规模型，纳入前沿模型

本书既考虑到一般读者的需要，简要地介绍了一些常规性模型，也考虑到一些高端研究者的需求，讲解了社会科学研究中几个重要、实用、前沿的数据处理与模型技术，包括固定效果模型、多层模型、计数模型等。从数据结构、模型示范到 Stata 软件的具体操作使用，前后一体，相互补充，数据清晰透彻，模型阐述详细，操作简洁明了。同时，为了加深读者对这些前沿模型的认识和掌握，本书除了精选社会科学中常见的案例外，还特别选用了一些国际上通用的经典案例，结合每章的数据结构特征，详解如何选择合适的模型技术，使本书内容结合实际，

深入浅出，易于理解。

四、使用说明

本书是《社会统计分析与数据处理技术：Stata 统计软件的应用》（这是一本本科生教材）的延续与发展。对于缺乏一定 Stata 统计基础的读者而言，建议先了解一些简单的统计学知识和 Stata 操作步骤，再来研读本书，这样效果会更好。