

社会学教材教参方法系列

SM

**量化数据分析：**  
Quantitative Data Analysis:

**通过社会研究检验想法**

Doing Social Research to Test Ideas

Donald J. Treiman ( 中文名：唐启明 ) / 著

任 强 / 译



社会学教材教参方法系列

SM

# 量化数据分析： Quantitative Data Analysis:

## 通过社会研究检验想法 Doing Social Research to Test Ideas

Donald J. Treiman (中文名：唐启明) / 著

任 强 / 译



社会科学文献出版社  
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

## 图书在版编目(CIP)数据

量化数据分析:·通过社会研究检验想法/唐启明著. —北京:  
社会科学文献出版社, 2012. 7

(社会学教材教参方法系列)

ISBN 978 - 7 - 5097 - 3178 - 9

I. ①量… II. ①唐… III. ①量化分析 IV. ①C8

中国版本图书馆 CIP 数据核字 (2012) 第 033039 号

社会学教材教参方法系列

### 量化数据分析: 通过社会研究检验想法

著 者 / Donald J. Treiman (中文名: 唐启明)

译 者 / 任 强

出 版 人 / 谢寿光

出 版 者 / 社会科学文献出版社

地 址 / 北京市西城区北三环中路甲 29 号院 3 号楼华龙大厦

邮政编码 / 100029

责任部门 / 社会政法分社 (010) 59367156

责任编辑 / 杨桂凤

电子信箱 / shekebu@ssap.cn

责任校对 / 张锦文

项目统筹 / 童根兴

责任印制 / 岳 阳

经 销 / 社会科学文献出版社发行部 (010) 59367081 59367089

读者服务 / 读者服务中心 (010) 59367028

印 装 / 三河市尚艺印装有限公司

开 本 / 787mm × 1092mm 1/16

印 张 / 28

版 次 / 2012 年 7 月第 1 版

字 数 / 504 千字

印 次 / 2012 年 7 月第 1 次印刷

书 号 / ISBN 978 - 7 - 5097 - 3178 - 9

著作权合同

登 记 号 / 图字 01 - 2011 - 2275 号

定 价 / 59.00 元

本书如有破损、缺页、装订错误, 请与本社读者服务中心联系更换

▲ 版权所有 翻印必究

Copyright© 2009 by John Wiley & Sons, Inc. All rights reserved.

This translation published under license.

---

社会科学文献出版社根据协议翻译、印刷、出版和发行此图书。

献给 John Pock, 是他引领我进入这个领域

## 英文版序言

这本书是关于如何开展理论导向的量化社会研究的——也就是说，通过社会研究检验想法。它源自我在加州大学洛杉矶分校（UCLA）30多年教授来自社会学、其他社会科学以及以社会科学为基础的专业学院（公共卫生、教育、社会福利、城市规划等）的研究生的课程。此课程是随着社会科学量化方法的发展演化而来的；早期课程只讲授本书的前一半内容（前7章），之后逐年增加新资料。有意思的是，从一开始我就一直保持同样的授课模式——在为期20周的学期内，每周授课三小时，并布置一次练习，期末要求学生在最后四周内提交一篇学期论文——以期不断提高社会科学领域的研究生在量化研究方面的能力。当然，本书也归功于多年来活跃的课堂讨论，尤其是在一些细微和复杂的方法问题上。

读完本书，你们或许能够知道如何从本质上理解量化数据。也就是说，你们能够在自己所学的领域很好地准备可发表的论文，以及一流的毕业论文章节。当然，总是有更多的东西需要学习。在最后一章（第16章），我讨论了一些高级方法，这些方法超出了数据分析初级课程所能涵盖的范围。

本书主要对明确定义的总体的代表性样本数据进行分析，当然也涉及一些特例。总体几乎可以包括任何东西——人口、正式组织、社会、职业、陶瓷碎片，或任何其他事物；分析的问题实质上是一样的。本书只是附带地提到数据收集方法；原因很简单：本书篇幅已经很长，没有足够的篇幅同时介绍数据分析和数据收集方法。因此，你们可以查阅其他的系统介绍数据收集方法的文献。我强烈建议你们应该在选修了数据分析课程之后而不是之前去做这件事情，因为在设计数

据收集时主要问题是决定收集什么，这意味着你们首先需要知道如何开展你们的分析。当然，你们也可以通过参与调查（可能无报酬）来学习数据收集的各种实用细节，并坚持参与每个过程——即使你们的参与毫无意义。

这本书涵盖了多种技术，包括表格分析（tabular analysis）、针对表格数据的对数线性模型（log-linear models）、各种形式的回归分析（regression analysis）、回归诊断（regression diagnostics）和稳健回归（robust regression）、处理缺失数据（missing data）的各种方法、逻辑斯蒂回归（logistic regression）、用于构建测度的因子分析技术（factor-based techniques）和其他技术，以及作为一种因果推论方法的固定效应和随机效应模型。但是，这不是一本统计学教科书；本书强调使用这些方法对社会如何运行进行研究，并得出相关结论。因此，本书是针对选修了第一年研究生统计课程之后的社会科学学生设计的。虽然书中有许多方程，但这是为理解统计方法的原理并灵活运用它们所必需的。本书的重点在于应用，因此书中提供了许多实际例子，且大多数引自我自己的研究。除了我自己收集的抽样调查数据之外，我也大量利用了美国综合社会调查（General Social Survey）数据——一项为研究和教学目的而设计的综合调查。附录 A 介绍了许多文献都使用过的主要数据集，并提供了如何获取数据的信息；它们都可以免费获得。

合理使用此书仅需的前提条件是，选修过一门初级研究生水平的社会科学统计学课程，认真思考和努力工作的意愿，以及具备高中代数知识的基础——无论是没有忘记还是重新学习。除了少数例外（涉及一点微积分和矩阵代数的内容），所有数学内容都没有超出高中代数知识的范围。如果你们对高中代数知识已经生疏了，可以参阅 Helen Walker 的 *Mathematics Essential for Elementary Statistics* 和 W. L. Bashaw 的 *Mathematics for Statistics*。这两本书很容易找到。虽然最近也有类似的书，但中学代数的内容没有变化，所以版本无关紧要。这些书可以从 amazon.com 网上和其他许多地方购买。

本书中使用的统计软件包是 Stata（版本 10）。用于计算的命令文件（Stata 的专业术语-do-文件）、结果文件（-log-文件）和辅助计算机文件可以从网站 [www.josseybass.com/go/quantitativedataanalysis](http://www.josseybass.com/go/quantitativedataanalysis) 下载。

某些计算的具体细节通常只能从下载的-do-和-log-文件中找到，所以务必要下载并仔细研究它们。这些文件会根据 Stata 的版本进行更新。

我在教学和此书中都使用 Stata，因为它已迅速成为一流社会学系和经济学系优先选择的统计软件包。这不是偶然的。Stata 是一个运算速度快和效率高的

软件包，它包括绝大多数社会科学家感兴趣的统计方法，并且能快速地添加新命令。虽然现在有许多统计软件包可供使用，但目前最有竞争力的三个软件包是 Stata、SPSS 和 SAS。其中，Stata 明显优于 SPSS——它速度快，更加准确，且应用广泛。虽然 SAS 很强大，但不如 Stata 那样直观，且更不易学习（和讲授）。此外，本书可以很容易地与 SPSS 或 SAS 结合起来使用，只需简单地转换 Stata-do-文件中的句法。〔我已经尝试做了一些，通过写出相应的 Stata 句法重现 Allison 基于 SAS 对固定和随机效应模型的完美指令（Allison, 2005）。〕

## 给教师的建议

当你们做课程设计时，我在讲课中是如何使用这些资料的一些经验或许对你们有帮助。

像前面提到的，以本书为基础的课程持续两个小学期（two quarters，共 20 周）。我每周讲一次 3 个小时的课，并布置一次作业。讲完课后，我亲自批阅这些作业；但随着学生人数的增加，我请助教（从以前课上表现好的学生中选择）协助我批阅作业，并解答学生在计算和统计学方面的疑问。最近几年，我两个小学期讲 17 周课，前 16 周每周布置作业，期末要求学生提交一篇学期论文。对学生论文的第一份草稿，我力争效仿期刊投稿的过程阅读并写出评论。因此，在我的课上，每个人都会收到“修改和再提交”的反馈。我鼓励学生根据我的反馈意见继续完善他们的学期论文，并在最后一次作业结束后两周提交第二份草稿。

最初的练习在一定程度上是为引导学生进入分析过程而设计的，后来的一些练习也是如此。但是，后面的练习形式更自由，要求学生“像书中讲解的那样进行分析”。对于有明确答案的练习，我提供了完整而详细的答案——类似于统计学习题集。

课程大纲、每周的练习以及练习详解可以从网站 [www.josseybass.com/go/quantitativedataanalysis](http://www.josseybass.com/go/quantitativedataanalysis) 下载。

## 致谢

像前面提到的，这本书是在 UCLA 与许多不同年级的研究生互动的基础上不断完善起来的，他们对每一章都做出了重要贡献，并通过详细评论或提出疑问的

方式，指出讲解中存在的问题。如果没有他们，就不会有此书，因为我从来没有想到自己会写一本教材，所以我非常感谢他们。尤其是 Pamela Stoddard 真正促成了这本书以当前的形式出版，之前她在飞机上偶遇 Jossey-Bass 的图书编辑 Andrew Pasternack，谈到她的教授正考虑将授课的章节写成一本教材。Andy 联系我，后来的事情就尽人皆知了。

这门课程最早是与我的同事 Jonathan Kelley 合作讲授的，20 世纪 70 年代他是 UCLA 的一名访问教授。第一个练习就是他提供的，课程总的进展，特别是前半部分，大多数归功于他。

我的同事 Bill Mason 是我多年的统计学导师，他最近从 UCLA 社会学系和统计学系退休。我经常得益于他对疑难统计问题的洞察力，而且我学到的许多在研究生阶段没有学到的知识都来自旁听他的高级统计课程。另一位同事 Rob Mare 同样给我很多类似的帮助。我的新同事 Jennie Brand 在 2008 年秋接手我的量化数据分析课，她阅读了全部手稿，并提出许多有益的建议。Stata 公司作者支持项目部（Author Support Program）的 Bill Rising 检查了所有 Stata 的下载-do-文件，并且提出了许多改进和更新的建议，从而保证它们是目前最好的练习。最后，这本书极大地得益于约 100 名中国学生的仔细阅读，2008 年 6 月我在北京大学给他们讲授这门暑期课程。他们发现许多未被察觉的错误，并指出一些表述模棱两可的地方，促使我重新完善书中部分内容。

我对研究设计和统计问题的理解，尤其是对因果关系和因果推论的挑战，极大地得益于加利福尼亚人口研究中心（California Center for Population Research）每周举办的讲座。参加讲座的有社会学家、经济学家以及其他社会科学家，特别是来自其他学校的访问学者，大家一起聆听和评论，报告研究进展。这种生动的、跨领域的讨论有几分流动研讨班的味道，我想这应该就是我所想象的学术生活可能和应该的样子。

最后，我的妻子 Judith Herschman 一直表现出无尽的耐心，只是偶尔问我：“你何时出版你的方法书？”



## 中文版序言

在过去的15年里，我一直致力于对中国社会的研究。我与中美两国的同事一起在中国大陆组织了两次全国性的概率抽样调查：一次是在1996年，关注中国大陆居民生命历程中所遭遇的不平等情况；另一次是在2008年，关注的是中国国内人口迁移的原因、过程与结果（本书中的许多研究实例都来自1996年的调查数据）。我和一些同事目前正着手计划在中国进行第三次调查，主题是迁移对儿童的影响。自1995年以来，随着访问中国的次数日渐增多，我对中国文化的理解和与中国学生的交往日渐加深。2008年7月，我以本书为教材在北京大学—密歇根大学学院授课，正是在那里我认识了本书的译者任强教授。此后在2010年秋，我在香港科技大学再次用此书授课。两次授课使用的都是本教材的英文版。

有了在北京以及在香港面向中国学生的授课经历，当任强教授提议将这本教材翻译为中文时，我真的非常高兴。这次与任强教授的合作也相当愉快，在此我表示衷心的感谢。同时，我还要感谢香港科技大学的一些研究生（李骏、李兰、宋曦、叶华、张卓妮和郑冰岛），他们校对了全部译稿并在中文措词方面提出了很好的修改意见。虽然我自己不懂中文，但是熟悉汉语的朋友告诉我，在任强教授和这几位同学的辛勤工作之下，翻译工作相当出色。

本书的英文版填补了美国研究生定量社会研究方法出版物中的一个重要空白。尽管在美国，针对社会科学研究的统计教科书很多，但大多数都集中讲述统计的数理基础，很少关注如何在实践中运用这些统计方法来做研究。本书的宗旨

就在于弥补这一缺憾，它所面向的是那些已经学习了统计学的入门课程并想将这些统计知识应用到严谨的实证研究中的学生。以前由于缺乏这样的书籍，所以我就自己为学生写讲义材料，后来将这些讲义扩充为一些章节，最后就形成了本书。

我希望本书在中国也会起到同样的作用，使学生作为社会科学研究者掌握或提高定量研究的方法和分析技术。由于中国学生的数学训练比美国学生好，此外，同欧洲的大学一样，中国的大学在本科生一入学就开始提供专业化的学科训练，所以我希望本书除了适用于研究生之外，同样也适用于主修任何一门社会科学的本科生。此外，本书还适用于那些已经完成了专业训练但或许还没有机会接触本书所展示的各种分析技术和策略的研究人员。我以前的许多学生——如今已是美国的大学教授或研究人员——告诉我，每当碰到新的研究问题时他们都会继续从本书中寻找解决的办法。其实，每当要使用一种因日久而生疏的分析方法时，我自己也会这样做。因此，本书既可用作学生的训练手册，也可用作研究人员的参考工具书。

在中国，定量社会研究方法的训练和应用的质量正在迅速提高，这从中国学生在申请美国大学的研究生项目时所提交的论文质量不断提高这一点上可见一斑。但中国地域辽阔，许多学生因英语能力有限而在阅读英文的技术性文章时面临困难，也有许多学生可能不具备购买昂贵的外文书籍的条件。因此，我希望这本教科书中文版的发行能有助于定量研究方法在中国继续推广——社会科学研究者们现在正需要用这些方法来做严谨的分析工作，这对了解一系列涉及中国社会特征的重要问题来说是必需的。本书不仅对学术界的社会科学研究者有用，而且对应用社会科学研究者也应该有用——不管他们是社会学家、经济学家、地理学家、流行病学家或其他公共卫生研究者，还是城市规划者、计划生育专家、社会服务行政人员，以及其他隶属于政府机构的研究人员；此书对市场分析员和其他在私人公司的从业者可能也同样适用。

在我看来，用社会研究检验各种想法充满了乐趣，希望本书能带领你——本书的读者，得出相同的结论。

Donald J. Treiman (中文名：唐启明)

2011年10月5日

## 译者序言

在我上研究生阶段和查阅文献的过程中，曾读过不少 Donald J. Treiman 教授的研究文章——他的朋友们都亲切地称呼他 Don。因为没有系统地受过统计学训练，所以我将他的文章作为理解统计方法和指标解释方面的范本，在不理解一些模型的原理的时候，也常常重读他的文章以寻求答案。可以说，他的文章是我学习统计学的指南。

第一次见到 Treiman 教授是在 2008 年夏天，当时我与谢宇教授一起负责主持北京大学—密歇根大学学院社会调查与定量分析实验室的暑期课程，邀请 Treiman 教授讲授“量化数据分析”课程，我同时作为助教之一辅助他的教学，与其说是助教，还不如说是学员——我完整地听完他 4 个星期的讲课，并认真完成每次作业。当时，课上有 98 名学生，加上旁听的有 100 多人。自己也做教师多年，但 Don 驾驭课堂的技巧和能力依然令我十分佩服。他在课堂上讲解清晰、逻辑性强、循序渐进，且说话风趣，课堂气氛十分活跃。他一丝不苟，对学生提出的每个问题都认真思考，给予解答，并常常提供丰富的补充资料。他的讲解就像他的文章一样，清晰、简洁、透彻，容易理解。这次暑期课的讲义用的就是已出版的本书英文版的校对稿。当上课前拿到 Don 发给我的校对稿时，我用了 2 天的时间将它通读了一遍，感叹不愧是凝聚了他几十年教学、研究心得的一本好书，当时就下决心再进一步精读此书，并翻译成中文，让更多的中国读者受益。

当我向他表示希望翻译此书时，Don 非常高兴，并主动联系美国的出版社，希望给予版权等方面的帮助；随后又将他的详细讲义和资料发给我，以便我透彻

理解此书。在书即将出版的时候，当他得知我在密歇根大学人口研究中心访问时，就马上给编辑写信，请编辑将书直接寄给我。离开的前一天晚上才收到美国出版社寄给我的书。于是，在回北京的飞机上，我又一次通读了此书。

2011年初，我完成了译稿，并且前后校对了三遍，反复斟酌一些中文用词的习惯。为了理解和翻译得更加准确，尤其是考虑到更加符合中文的表达习惯，Don 建议再多请几位修习过此课程的同学帮忙校对（尽管我与社会科学文献出版社签订的出版合同要求5月底前交付书稿，但出于保证翻译质量的考虑，出版社同意延期半年）。这些同学不仅参加过2008年北京大学—密歇根大学学院的暑期课程，而且分别在香港科技大学和加州大学洛杉矶分校聆听过 Don 亲自授课。我们采取每章轮流由两人校对的工作方式，最后由我统一审校。无论在翻译还是审校的过程中，当遇到不理解的内容和有疑问的地方时，均得到 Don 详细、耐心的讲解和回复，这使我进一步加深了对本书的理解。

当翻译、校对工作基本完成时，鉴于 Don 对中国文化的了解和对中国社会问题的研究兴趣，我也希望 Don 借此书中文版出版之际，跟中国学生、读者谈谈他多年做中国研究的体会。当我约他为中文版写序时，他欣然同意。同时，我们在这本书的封面上看到 Don 第一次有了正式的中文名字。虽然 Don 从事中国研究十多年，他的中译名也曾以多种形式（如特雷曼等）出现过，但多是英文直译，没有得到 Don 的认可。然而，从此书在中国面世开始，以及在未来以中文发表的文章、著作中，Don 决定正式使用他的中文名字——唐启明。

组织、协调工作是很繁琐的工作，在这里我要特别感谢宋曦同学在校对过程中出色的组织工作。参与本书校对工作的有：宋曦（加州大学洛杉矶分校博士研究生，负责校对第1~4章、第8章、第10章、第12章，以及英文版序言、导言、附录和作者介绍）、李兰（香港科技大学硕士，负责校对第1~4章、第8章、第10章、第12章，以及英文版序言、导言、附录和作者介绍）、李骏（香港科技大学博士候选人，负责校对第5章、第7章、第9章、第11章、第13~14章）、叶华（香港科技大学博士候选人，负责校对第6~8章、第13章、第15~16章）、郑冰岛（香港科技大学博士候选人，负责校对第6章、第9章和第11章）、张卓妮（香港科技大学博士，负责校对第14~15章）。最后，李骏负责翻译了 Don 为中文版写的序言，张卓妮和宋曦对此做了校对。在此向参与校对的6位同学表示衷心的感谢。

最后，在此感谢社会科学文献出版社的编辑杨桂凤女士给予的理解和帮助。

同时感谢谢宇教授、郭志刚教授和郑晓瑛教授的支持和鼓励，并再次向 Don 表示感谢，感谢他在联系版权的过程中给予的无私帮助，在翻译和校对的过程中给予的耐心、细致的解答。

由于译者的能力和知识有限，在翻译过程中难免会有理解不当或错误的地方，恳请读者和同行批评指正。

任 强

2011年9月于北京大学人口研究所

## 导 言

在社会科学领域，研究生所学的统计课程常常被当作数学课程来对待，过多地强调数学推导和证明，这是很常见的，甚至在进行实证研究的时候——实证研究之所以经常被称作统计学，是因为要概括数据，描述变量之间的关系，要依据样本对总体做推论——学生们对所学的知识该作何用，经常表现得很迷茫。也就是说，不知如何将他们所学到的统计方法运用到实际研究中去。

本书的核心思想就是学习应用统计技术。我设想你们已经掌握了研究生一年级的统计课程，最好是社会科学系或专业院所设的统计课程。它包括基本的统计学概念，至少涉及常规最小二乘回归，最好有一些逻辑斯蒂回归知识。但是，现在你想成为一名应用定量分析研究者，分析抽样调查或其他定量数据集，以支持你所涉足领域的大量文献知识，最后发表你的研究结果。这本书正是为你而设计的。它是基于两个小学期的课程和我在加州大学洛杉矶分校过去 30 多年的教学经验，并跟踪社会学和其他社会科学领域定量分析方法最新的发展成果写成的。

当今发表在社会学和其他社会科学顶级刊物上的文章所使用的各类度量或统计方法大大不同于多年以前。他们大量借鉴计量经济学、心理学、生物统计学，甚至数理统计学的最新进展——每年都有新进展。因此，我过去认为一些内容在第一年的定量方法课程中讲授太难（例如，固定效应和随机效应模型），但现在在发表的社会学文章中却很常见。这使我觉得非常有必要将它们也纳入我的课程教学和这本教材中。

## 各章介绍

本书从分析非实验数据最基本的方法（百分数表）开始介绍。第1~3章阐述列联表的基本逻辑，并介绍关于如何制表的许多技术细节（使得此类表简单易读）。在这三章中有两个核心思想：如何确定百分数表的方向问题和统计控制的概念。根据我的教学经验来看，第一个问题对一些学生来讲是困难的——甚至要比处理后面章节介绍的复杂数学公式困难得多。因此，即使你们认为自己已经掌握了百分数表，我仍希望你们认真学习这几章。这样做将会使你们受益匪浅。

第4章是关于计算的介绍。我会介绍怎样管理数据以便于计算机分析，怎样使用统计软件进行分析。本书将使用Stata软件包，我会提供使用Stata的一些建议。本章中的介绍也同样适用于其他统计软件包，比如SPSS和SAS。

第5~7章讲解常规最小二乘相关和回归，这是社会科学中统计分析的基础。这些过程提供了一种量化某些数量结果与其决定因素之间关系的方法。例如，在其他影响因素保持不变的情况下，我们期望在多大程度上人们收入的不同是由他们的受教育水平不同导致的？它们同样提供了一种如何评估预测结果的方法。例如，收入差异在多大程度上能被归结为教育、性别、种族等方面的不同。第5章主要讲解双变量相关和回归，让读者对相关和回归统计有逻辑直觉，并在解释相关和回归统计结果时注意一些常见错误。第6章介绍多元回归，用于当一个因变量有几个自变量的情形，同时介绍“虚拟”变量或二分变量的概念，这需要特别的处理方法。通过使用虚拟变量和“交互项”，我提出一种思路来评价不同人群的社会过程是否不同这个社会科学中常见的问题。第7章介绍了在回归框架内检验相关假设的多个“窍门”。

社会科学家在使用大多数数据时会受到“缺失数据”的困扰——某些个人在某些变量上缺失信息。第8章回顾了处理缺失数据的方法，并在最后部分以目前最新的方法为例，讲解如何进行缺失数据的多元估算。

接下来的第9章讨论抽样及其在统计分析中应用的问题。尽管前面各章都假设是简单随机抽样，但是，多数常规人口抽样实际上是复杂的、多阶段抽样。要正确分析这些抽样数据，我们在计算标准误的时候就需要考虑到观测值之间的“聚类”（clustering）特征。本章介绍处理这种特征的抽样估计（survey estimation）方法。

处理回归问题时常常容易误入陷阱。在第5章中我们曾（简单地）讨论过该问题。在第10章中，我们将通过介绍回归诊断（regression diagnostics）更全面地阐述该问题。这些步骤可以避免依据回归结果做出错误推断的可能性发生。

第11章介绍如何及怎样构建多题项测度（multiple-item scales），不仅主要集中在讨论以因子为基础（factor-based）的测度法，而且介绍效应比例测度法（effect-proportional scaling）。我们经常想研究一些用问卷中的一个题项无法充分测量的概念问题，如“生活水平”、“自由主义”、“A类性格”和“抑郁”等。基于多个题项的综合测量（summary measures）或测度（scales），与单个题项的测量或测度相比，通常会提供更可信（reliable）、更有效（valid）的变量指标。本章介绍如何构建和使用这类测量指标。

第12~14章介绍受限因变量（limited dependent variables）的分析方法。常规最小二乘回归是为处理连续型因变量而设计的，如收入、受教育年限等。但社会科学家感兴趣的许多因变量或是二分变量（如人们是否投票，是否结婚，是否受到犯罪的伤害，等等），或是多项变量（如在多党派社会中的政治派别、职业类别、大学类型等）。对数线性分析（log-linear analysis）和逻辑斯蒂回归（logistic regression）是处理此类受限因变量的方法。第12章介绍对数线性分析，它是对一组多项变量之间的关系做严格推论的方法，也就是说，对列联表中各变量关系的程度和模式做推论。从这个角度讲，对数线性分析提供了一种对我们在第1~3章所讨论的各种表做统计推论的方法。第13章介绍二项逻辑斯蒂回归，它是一种适合分析二分结果变量的方法，进而讨论如何使用此方法来处理各种特殊情况：递进比，即所研究的问题是什么因素导致人们是否一步步地发生变化，例如，从某一受教育水平到更高水平的影响因素；离散时间风险率模型，它所研究的问题是一个事件（如初婚）在某一给定时点（如特定年龄）发生的可能性；案例—对照模型，它提供了一种研究罕见事件发生可能性的方法，如传染病，获得精英职业，等等。第14章进一步介绍如何研究其他受限因变量：非序次多项变量，如居住地方的类型，可以通过多项逻辑斯蒂回归（multinomial logistic regression）来分析；序次变量，即分类是有等级次序的，但分类之间不等距，例如一些态度测量（你是否“很高兴”、“有点高兴”、“不太高兴”），可以通过序次逻辑斯蒂回归（ordinal logistic regression）来分析；“删失”（censored）变量，即某一测度的范围被截断，例如，一个收入变量的最高分类是“每年100000美



元及以上”，可以通过 tobit 回归 (tobit regression) 来分析。

当使用非实验数据时，通常很难明确地建立某一变量导致另一变量的因果关系，因为它们两者可能共同依赖于第三个变量，而此变量常常是未被测量的。第 15 章介绍了一些处理这类问题的方法，称为固定效应 (fixed-effects) 和随机效应 (random-effects) 模型，前提是要有合适的数据——面板数据 (同样的个体在多个时点被观测) 或者整群数据 (在一个家庭、学校、社区等中多个个体被观测)。在可以获得合适数据的情况下，这是非常有说服力的方法。

最后一章 (第 16 章) 介绍一些高级的、我在本书中所不能涉及的方法，通常超出一年级研究生所学定量数据分析课程的范围。目前，其中的许多方法被广泛应用于经济学，它们主要处理各种各样的内生性问题 (endogeneity problem)，即不可观测变量同时影响自变量和因变量，从而导致有偏估计。固定效应和随机效应模型提供了一种解决此类问题的方法，但还有许多其他可用的方法，这将在第 16 章中介绍。我也会简要介绍结构方程模型，它是一种应对复杂社会过程的方法，即一个因变量是另一个因变量的自变量。例如，在社会地位获得研究中，我们想研究父母的社会地位如何影响个人的受教育水平、父母的社会地位和个人的受教育水平如何影响个人的第一份工作等类似的问题。简要介绍这些高级方法是为了引导学生在量化分析方面继续往更高层次学习。本章以如何做好研究的建议结束，旨在暗示如何提高你们的研究质量并节省时间和精力。