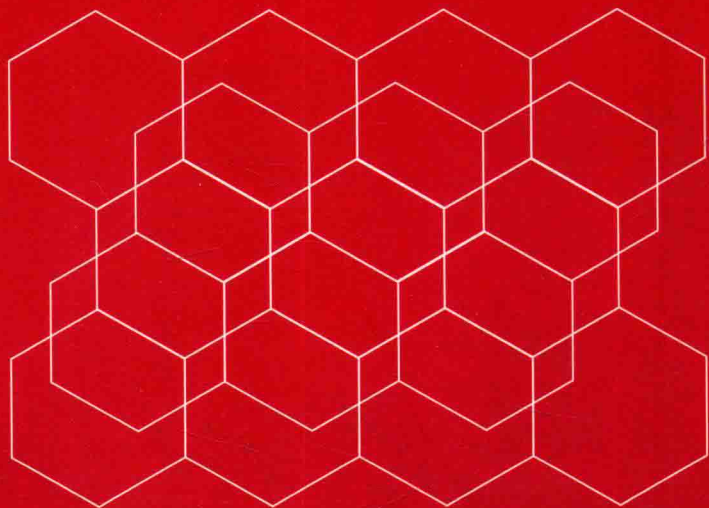


高等院校经济学管理学系列教材



基本有用的 计量经济学

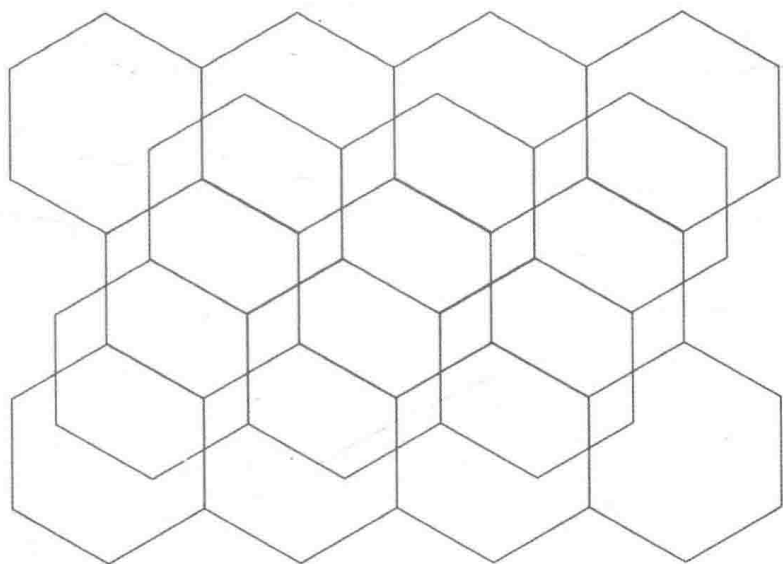
Mostly Useful Science of
Econometrics

赵西亮 著



北京大学出版社
PEKING UNIVERSITY PRESS

高等院校经济学管理学系列教材



基本有用的 计量经济学

Mostly Useful Science of
Econometrics

赵西亮 著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

基本有用的计量经济学/赵西亮著. —北京:北京大学出版社,2017.7

(高等院校经济学管理学系列教材)

ISBN 978-7-301-28481-0

I. ①基… II. ①赵… III. ①计量经济学—高等学校—教材 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2017)第 154391 号

- 书 名** 基本有用的计量经济学
JIBEN YOUYONG DE JILIANG JINGJIXUE
- 著作责任者** 赵西亮 著
- 策划编辑** 杨丽明 姚文海 吕正
- 责任编辑** 朱梅全 杨丽明
- 标准书号** ISBN 978-7-301-28481-0
- 出版发行** 北京大学出版社
- 地 址** 北京市海淀区成府路 205 号 100871
- 网 址** <http://www.pup.cn>
- 电子信箱** sdyy_2005@126.com
- 新浪微博** @北京大学出版社
- 电 话** 邮购部 62752015 发行部 62750672 编辑部 021-62071998
- 印 刷 者** 三河市北燕印装有限公司
- 经 销 者** 新华书店
- 730 毫米×980 毫米 16 开本 17 印张 296 千字
2017 年 7 月第 1 版 2017 年 7 月第 1 次印刷
- 定 价** 45.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱:fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话:010-62756370

序

经过三十多年的努力,中国经济已经从计划经济模式成功转型为以公有制为主导的社会主义市场经济模式。在经济转型与发展过程中,从中央到地方的各种经济政策发挥了主导作用。在国家层次上,有“五年计划”、《中国制造2025》《新劳动合同法》《社会保险法》等规划或法律法规,国家支持新兴战略产业政策,京津冀一体化战略,长江流域发展战略,自贸区战略,“一带一路”愿景,当然也包括财政、货币政策等。在地方层次上,有地方“五年计划”、地方产业政策以及地方性财政政策等。

各级政府的各种经济政策,大都是在调查研究、智库论证、征求有关各方意见的基础上形成的。但是,不少调查研究、征求意见并不一定经过充分、仔细的论证。尤其是各级政府领导人政务繁忙,很多人没有充分时间进行系统的调研与征求意见,也有不少停留在召开座谈会或会议讨论,存在一定程度的形式主义。由于时间等因素的限制,很难充分征求社会各方面的意见。一些从属于各级政府部门的智库,本身可能并不具有很高的专业性,有时还需要揣摩上级主要领导的意图,缺乏必要的独立性。相反,身在中国高校与科研机构、立场比较超然的经济学者的经济政策研究,往往具有较高的专业性和独立性。但是,中国经济学者在从事经济政策研究时,经常是在经过比较简单的定性分析之后,即提出自己的观点、政策建议。这些观点与政策建议,大多不是基于经济数据进行分析,通过严谨的推断过程所获得的结论,因此不具备较高的科学性。诚然,除经济数据外,还必须对经济的现状、历史有深刻的理解与丰富的感性认识。但是,只有建立在经济数据基础上的严谨的实证研究,才能从复杂的经济现象中揭示经济变量之间的因果关系和内在规律,从而使结论以及建立在结论基础上的政策建议具有较高的科学性。

因此,中国经济学者关于经济政策研究的一个努力方向,是提倡以数据分析为基础的实证研究,特别是对各个时期的各种经济政策进行量化评估。最近二十年来,计量经济学出现了一个新的领域,称为政策(或项目)评估计量经济学(econometrics of program evaluation),并广泛应用于评估各种社会经济政策。

所谓政策评估计量经济学,是在经济数据的基础上应用计量经济学方法与

工具对社会经济政策进行量化分析,其主要目的是测度某个政策实施后对某个群体、某个行业或某个地区的“因果”影响。由于经济现象的不可实验性(如四万亿元刺激计划不可能再重复一次),政策评估的最主要困难是识别因果关系和测量政策效应的大小或强弱。政策评估计量经济学因此提出了各种条件下的量化评估方法,其核心是估计无法观测到的虚拟事实(counterfactuals),即没有实施某种政策时的效果,这样,某项政策的效应可测度为实际政策作用下的结果与虚拟事实之差。在国外,政策评估计量经济学已被广泛应用于经济学和社会科学很多领域,包括劳动经济学、工业组织、发展经济学、社会学等。例如,Card and Krueger (1994)运用双重差分法(difference in difference)考察了美国最低工资调整对就业的影响。Dehejia and Wahba (1999)使用倾向得分匹配(propensity score matching)估计就业再训练政策对于收入的影响。Hahn et al. (1999)应用断点回归方法评估了美国反歧视联邦法对少数族裔就业的影响。Abadie et al. (2002)则使用分位数处理效应模型估计美国大型的就业再培训计划对于收入分配的影响。Jacob and Lefgren (2004)利用美国芝加哥公立学校 1996 年教育改革数据评估了芝加哥公立学校所实施的设立暑期培训班以及学生成绩不好需要留级的教育政策对学习成绩的影响。Lalive (2008)评估了奥地利“区域扩展福利计划”(regional extended benefits program)实施后对男性与女性工人失业持续期的影响。

政策评估计量经济学,渊源于统计学科的流行病学(epidemiological statistics)统计以及所谓的处理效应(treatment effect)统计。作为政策评估的重要方法论,政策评估计量经济学可用于对中国各种社会经济政策进行严谨的定量评估,这将大大提高中国社会经济政策研究的科学性。特别是现在已进入大数据时代,各种形式的经济数据非常丰富,也便于运用强大的计量机进行处理。事实上,一段时间以来,已有一些智库和学者开始运用计量经济学方法评估中国各种经济政策,并提出相关的经济政策建议。例如,中国科学院预测科学研究中心汪寿阳教授所带领的政策研究团队,多年来运用计量经济学方法,研究各种中国经济政策,很多建议获得政府采纳。美国南加州大学萧政教授等(Hsiao et al., 2012)提出了一种面板数据的政策评估方法,并评估了中国内地与香港的经济关系,他们发现虽然“九七”回归这一政治事件对香港经济增长没有显著影响,但是内地和香港关于建立更紧密经贸关系的安排(CEPA)对香港实际GDP增长贡献了约4%。韩乾、洪永森(2014)利用上海证券交易所提供的一组市场交易数据,评估了国务院2010年通过的支持七大新兴战略产业的政策对证券市场投资的影响,他们发现这个利好产业政策只是为少数机构投资者提供

了操作机会并造成财富从大量散户向机构投资者转移,并没有起到引导证券市场投资者向七大产业投资。Ouyang and Peng (2015)扩展了 Hsiao et al. (2012)的方法,使用半参数的方法评估 2008—2009 年实施四万亿元财政刺激计划的效果。他们的研究发现,财政刺激政策起初的确促进了经济增长,特别是在 2009 年第三季度左右,GDP 增速提高了约 5.4%,但这一政策效果随后迅速降低,在 2010 年第四季度之后甚至变成了负值。因此,四万亿元的财政刺激政策只是临时对经济有促进作用,并无长期影响。

与其他经济学研究方法一样,计量经济学并不是放之四海而皆准的“灵丹妙药”,它既有科学性,也有局限性。在应用计量经济学方法与工具评估社会经济政策时,首先需要注意,每种方法和工具都有其适用的范围与条件,特别因为经济现象具有不可操控实验的特点,很多方法与工具是建立在一些假设基础之上,在应用时需要注意这些假设条件的合理性。

其次,计量经济学的方法论基础是概率论与统计学,因此数据样本的选择是一个实际应用中必然遇到的重要问题。如何选取一个具有代表性的数据样本,避免样本选择偏差,与方法和工具的选择一样重要。另外,数据质量也是一个潜在问题。大数据时代,数据收集与处理成本越来越低,数据质量越来越高。但是在中国,各个地区特别是各地政府部门所收集的数据,可能会存在这样那样的系统性偏差。应该提倡随机抽样调查,或者收集整理数据,建立比较具有独立性和代表性的经济数据库。

再次,经济政策定量评估还必须与定性分析以及其他各种研究方法相结合,包括理论逻辑分析、历史逻辑分析、实地调查研究等,而不是去取代现有研究方法。中国高校的经济学者经常被批评不熟悉中国经济实际状况,因此必须通过调查研究等方法获得第一手资料,丰富自己的感性认识。但是,感性认识并不能确保政策研究的科学性,我们还需要科学的理论思维和分析方法。天文学家正是依靠科学的研究方法与工具探索天体物体的运动规律,这值得我们学习和借鉴。

最后,对经济政策的定量评估与分析,需要注意对量化结果进行经济解释。经济政策评估与分析不是以严谨的学术期刊论文发表为主要目的,应该用通俗的语言和直观的表达方式向政府部门和社会公众传播研究结果。当然,直观表达和通俗解释并不意味着降低量化评估分析的标准。社会经济政策评估是经济政策研究的重要组成部分和重要方法,是提升经济政策研究与政策咨询之科学性和有效性的重要途径。尤其在目前经济下行态势下,研究如何制定最佳经济政策组合促进中国经济持续稳定快速发展,具有重要的现实意义,而定量评估

各种社会经济政策的实际效果是制定最佳政策组合的关键环节。中国经济学者应该大力倡导应用计量经济学等现代经济学分析方法与工具对中国各种社会经济政策进行定量评估与分析,从而推动建设有中国特色的高水平的新型经济智库。

关于政策评估计量经济学,目前国内暂无详细介绍的计量经济学教材,最近,厦门大学经济学系赵西亮博士写了一本《基本有用的计量经济学》,该书利用潜在结果语言,从总体视角出发,集中探讨了政策评估过程中常用的因果效应识别策略。全书以随机化实验为基础,首先介绍了潜在结果和分配机制的概念,并利用潜在结果定义因果效应。然后介绍了随机化实验,一种特殊的分配机制,可以消除选择偏差,成为观测研究中各种识别策略的基础。所有的识别策略都是通过一定的设计模拟随机化实验,从而得到可信的因果效应估计。另外,作者还简要介绍了因果图方法,它是与潜在结果框架完全等价的因果模型,但是更加直观,容易使用。在这三个理论的基础上,该书介绍了线性回归、匹配方法、工具变量法、面板数据方法和断点回归设计等几种在观测研究中常用的因果效应识别策略。全书的主要特色是用潜在结果语言和因果图介绍各种识别策略。对于每种识别策略,作者还利用具体实例讲解各策略在 Stata 软件中的实现。当然,该书介绍的方法也有一定的局限性,正如我上文及该书结语中所述,各种识别策略模拟随机化实验,可以得到更为可信的因果效应,但是往往不能回答具体的内在影响机制,读者在使用的时候需要了解各识别策略的假设前提和内在限制。

赵西亮博士嘱我为该书作序,我欣然命笔,并认为该书是有关政策评估计量经济学的一本很好的入门教材,推荐给大家。是为序。

洪永森

康奈尔大学 Ernest S. Liu 经济学与国际研究讲席教授
厦门大学经济学院和王亚南经济研究院院长

2017年2月6日

前 言

目前,国内计量经济学的训练(无论是本科生还是研究生)主要集中于统计推断(statistical inference),即如何利用样本信息获得总体信息的估计以及如何进行假设检验以判断估计结果的统计显著性。在经济学实证中,我们拿到的往往是总体的一个样本,利用样本信息进行的估计是否能够反映总体,是实证研究中一个非常重要的问题。但是,统计推断本身往往很少能够给出因果关系的信息。随着信息技术的发展,数据获取的成本越来越低,我们开始进入一个大数据的时代,这意味着我们可以获得的数据样本容量越来越大,甚至可以获得总体信息。这样,统计推断的作用可能就越越来越小,比如,如果拿到了总体数据,那么传统意义上的统计推断就没有用武之地了。但是,就算我们有总体数据,也不能回答因果关系的问题。比如,假设我们有中国人口普查的数据,想考察教育如何影响个人收入,仍然是没有办法获得因果效应的知识的。因而,因果效应无关样本大小,对于因果效应的探讨是更加底层的问题,是任何科学获得知识的关键。所以,要获得变量之间因果效应的知识,必须要进行因果推断(causal inference)。

目前,经济学的经验研究正在经历一场研究范式的转变(paradigm shift)(Panhans and Singleton, 2016),从统计推断向因果推断转变。越来越多的实证研究开始探讨如何才能科学地识别经济变量之间的因果影响,而非集中于估计量的统计显著性问题,统计推断往往是相对次要的问题(second order problem),而因果推断才是获取知识的首要问题。这种研究范式的转变被 Angrist and Pischke (2010)称为经济学经验研究的“可信性革命”,其关键特征是引入潜在结果框架清晰定义因果,利用随机化实验的思想作为因果效应识别的基础,因而新研究范式有时也被称为“以实验设计为基础的计量经济学”或计量经济学的“实验学派”(Angrist and Pischke, 2017)。尽管其他计量经济学家并不一定完全认同“可信性革命”的说法,但是,“实验学派”的计量经济学确实使经济学经验研究获得很多“基本有用的经验知识”(Rust, 2016)。最近二十多年来,“实验学派”计量经济学方法在经济学经验分析中的影响越来越大,研究范式的“技术进步”也不断在其他学科中“技术扩散”(Bowen et al., forthcoming),但是在经济学教学中,本科生和研究生所使用的《计量经济学》教材仍然沿

用了老的研究范式,在经济学经验研究文献中广泛采用的因果推断方法仍然没有进入《计量经济学》教科书。Angrist and Pischke (2017)指出,传统计量经济学教材中的很多指导是过时的,比如有关异方差、序列相关等问题的讨论,这些问题都不会影响因果效应的识别,而这些问题的解决只需要利用 White (1980)的异方差一致性标准误差或 Newey and West (1987)的序列相关及异方差一致性标准误差进行修正。Angrist and Pischke (2017)提议新的研究范式更加有趣、相关性更强、识别结果更加令人满意,为什么不能让我们的学生也获得这些技能呢?

本书顺应了 Angrist and Pischke (2017)的号召,将经济学经验研究新范式介绍给我们的学生。本书分成两大部分:理论基础和识别策略。理论基础部分,首先介绍潜在结果框架。潜在结果概念的引入,便于清晰定义因果效应,从而避免 Lord 悖论。然后,介绍随机化实验。随机化实验是所有识别策略的基础,本书介绍的所有识别策略在一定的识别条件下都可以看作是一种随机化实验。另外,我们还介绍了因果图。因果图是与潜在结果框架完全等价的因果模型(Pearl, 2009),但是更加直观。

第二部分主要介绍了经济学经验研究中常见的几种识别策略。首先是线性回归,主要关注在什么样的识别条件下,线性回归系数可以解释为因果效应参数。我们强调在线性回归中,我们关心的解释变量和其他控制变量的地位是不同的,其他控制变量的引入是为了识别我们关心的解释变量的因果效应,为强调这一点,我们将关心的解释变量称为原因变量或干预变量。然后,我们讨论了匹配方法的识别条件。线性回归和匹配方法有着密切的联系,具有相同的识别条件。我们强调回归和匹配的识别条件都是 CIA,都不能解决内生性问题。接着是工具变量法。工具变量法是经济学中一种相对比较成熟的方法,但是利用潜在结果语言和引入异质性之后,工具变量法有了新的内涵,工具变量所能识别的因果效应参数是依从者的因果效应,并且在异质性框架下,不同的工具变量识别不同的因果效应参数。在有多期数据的情况下,可以利用多期数据的特点,消除不随时间变化的混杂因素的影响,从而使双重差分法模拟增量上的随机化实验。另外,我们还讨论了可以允许时变混杂因素的合成控制法和回归控制法。最后,我们讨论了最接近完全随机化实验的一种识别策略——断点回归设计(RDD),它利用断点附近个体具有高度相似性的特点来识别因果效应。另外,对最新发展的弯折回归设计(RKD)也进行了简要介绍。

本书主要根据我为厦门大学经济学院研究生和高年级本科生开设的“应用微观计量经济学”课程讲义修改而成。本书的主要特色是利用潜在结果语言和

因果图讲解各种识别策略,另外辅以具体案例讲解各种识别策略在 Stata 软件中的实现。适合从事经济学经验研究的学者和本科生、研究生使用,也适合从事社会学、政治学、流行病学等相关学科研究的学者作为因果推断方面的参考书。限于作者水平有限,时间紧迫,定有很多错漏之处,欢迎读者批评指正。

赵西亮

2017年2月20日

致 谢

首先感谢北京大学出版社杨丽明编辑,没有她两年前的邀请,可能就没有本书今天的出版。另外,在本书的出版过程中,杨编辑做了大量细致的工作。特别感谢洪永森教授在百忙之中为本书作序。洪永森教授是国际计量经济学领域的著名学者,一直致力于中国的计量经济学教育事业,对国内年轻学者也给予了很多帮助。非常感谢洪永森教授对笔者的鼓励和对本书的评价。

感谢上海交通大学安泰经济与管理学院朱喜教授与笔者的有益讨论,感谢上海交通大学安泰经济与管理学院博士后梁文泉博士为笔者提供的相关资料,另外,梁文泉博士还通读了初稿,指出了文中的多处错别字,在此一并感谢。感谢厦门大学研究生院副院长方颖、厦门大学经济学系主任龙小宁、厦门大学经济学系副系主任赵建等教授对笔者的帮助和支持。

非常感谢选修笔者课程的学生,自2013年开课以来,已有四届7个班(4个本科生班、3个研究生班)近200名学生选修了以本书为基础的课程,与学生的讨论使本书内容不断完善。本书初稿完成后,部分学生分别阅读了部分章节,并修改了明显的错别字及语句问题,他们是厦门大学经济学系博士研究生范振中、彭骏,硕士研究生陶宇、张惠杰、唐梦莹等,在此表示感谢。

目录

Contents

第一章 绪论	(1)
第一节 引言	(1)
第二节 因果推断简史	(3)
第三节 经济学研究的基本问题	(5)
第四节 本书安排	(6)

第一篇 理论基础

第二章 潜在结果框架	(9)
第一节 潜在结果	(9)
第二节 稳定性假设	(12)
第三节 分配机制	(13)
第四节 潜在结果和 Lord 悖论	(15)
第五节 因果效应参数	(19)
第六节 总结	(24)
第三章 随机化实验	(25)
第一节 随机化实验的作用	(25)
第二节 随机化实验与选择偏差	(28)
第三节 随机化实验的分类	(30)
第四节 随机化实验的分析	(32)

第五节	随机化实验的缺陷	(37)
第六节	总结	(38)
第四章	因果图	(40)
第一节	基本概念	(40)
第二节	因果图和边际独立性	(41)
第三节	选择偏差和因果效应识别	(45)
第四节	选择偏差的处理	(50)
第五节	总结	(52)
第二篇 识别策略		
第五章	线性回归	(55)
第一节	条件期望函数和线性回归	(55)
第二节	线性回归和因果效应	(66)
第三节	回归方法的软件实现	(72)
第四节	总结	(75)
第六章	匹配方法	(77)
第一节	协变量匹配方法	(77)
第二节	倾向指数匹配方法	(80)
第三节	匹配方法的基本步骤	(83)
第四节	倾向指数匹配方法的软件实现	(97)
第五节	总结	(117)
第七章	工具变量法	(119)
第一节	同质性工具变量法	(119)
第二节	异质性工具变量法	(130)
第三节	工具变量法的软件实现	(140)
第四节	总结	(145)

第八章 面板数据方法	(147)
第一节 固定效应方法	(147)
第二节 双重差分法	(149)
第三节 合成控制法	(162)
第四节 回归合成方法	(171)
第五节 面板评估方法的软件实现	(176)
第六节 总结	(185)
第九章 断点回归设计	(186)
第一节 断点回归设计	(186)
第二节 断点回归设计的图形分析	(194)
第三节 断点回归设计的估计	(199)
第四节 弯折回归设计	(208)
第五节 断点回归设计的软件实现	(217)
第六节 总结	(229)
第十章 结语	(231)
附录 Stata 数据处理编程简介	(234)
第一节 常用命令	(234)
第二节 do 文件	(235)
第三节 结果呈现利器:estout 软件包	(240)
第四节 练习	(246)
参考文献	(250)

第一章 绪 论

物有本末，事有终始，知所先后，则近道矣。

——《大学》

第一节 引 言

在经济学研究中，我们非常关心经济变量之间的因果关系。尽管其他相关关系也很重要，但是在政策评价中，因果关系是更根本的。关于因果关系的讨论，可以有两个方向。一个是看到结果，寻找结果背后的原因，即考察“结果的原因”(cause of effect)。这种研究非常重要，往往是科学研究的起点。但寻找结果背后的原因，非常复杂。某一种结果产生的原因可能有很多，需要通过详细的调查、深入的分析才能找到。另一个是看“原因的结果”(effect of cause)，主要关注某一干预对结果的影响。对于政策评价而言，往往仅关注某一项原因，考察这一原因变动时，结果会有怎样的变动。相对而言，这一方向更加简单，可操作性更强。事实上，两个方向的分析是密切相关的，搞清楚了“原因的结果”，才有利于考察“结果的原因”以及每一种原因在产生结果中的作用。因而，很多情况下，我们更关注“原因的结果”。比如，在微观经济学中，我们关心某些措施或干预对个体行为的影响，比如班级规模对儿童学习效果有什么影响？安全带是否可以增加乘客在车祸中的生还机会？教育对个人就业及未来收入有什么样的影响？助学金是否可以增加贫困大学生的升学率？对农民工进行职业培训是否可以增加他们的就业机会和收入水平？在宏观经济学中，我们关心大学生扩招是否增加了大学毕业生的就业难度？货币政策调控是否能够抑制通货膨胀？限购政策是否能够抑制房价的持续上涨？2008年的4万亿刺激方案是否促进了我国的经济发展和就业扩张？这些问题都是在探讨“原因的结果”，也是本书的主要研究对象。

一项干预^①对结果变量产生的影响通常称为因果效应(causal effects)或干

^① 干预可以指一项措施、政策、行动或暴露，在英文中对应词有 manipulation, treatment, intervention, exposure 等，本书中我们通称为干预，对应的变量我们称为干预变量或原因变量。

预效应(treatment effects)。^①类似于在医学领域对病人的治疗,如果采用某一种新药或治疗措施,会对病人的康复有什么样的影响?在医学领域称为治疗效果(treatment effects),在经济学及其他社会科学中也经常遇到类似的问题,对这些问题的考察称为因果效应分析。下面看两个简单的例子,了解因果推断的困难。

例 1.1(新药的治疗效果) 假设现在想考察一项新药的治疗效果,通过调查获得了某医院的新药使用情况,见表 1.1。

表 1.1 Yule-Simpson 悖论

总样本	生还	死亡	生还率
干预组	20	20	50%
控制组	16	24	40%
男性样本	生还	死亡	生还率
干预组	18	12	60%
控制组	7	3	70%
女性样本	生还	死亡	生还率
干预组	2	8	20%
控制组	9	21	30%

从总样本看,共调查了 80 个病人,40 个病人使用了新药,另外 40 个病人没有使用新药。使用了新药的病人归为干预组,没有使用新药的病人归为控制组。结果显示,干预组生还率为 50%,控制组生还率为 40%,干预组生还率更高。从总体样本上看,新药采用与生还率正相关,但这并不一定说明新药真的有治疗效果。表 1.1 的第 2 部分和第 3 部分分别按照性别列出男性和女性病人的结果。可以看出,对于男性而言,干预组生还率低于控制组;对于女性而言,也是干预组生还率低于控制组。从而出现,按照性别分层后,男性和女性群体中,均出现了新药使用与生还率的负相关关系。这种现象在统计学中称为 Yule-Simpson 悖论。

上面的例子来自 Pearl (2009),是人为构造的,下面看一个真实的案例。

例 1.2(美国加州大学伯克利分校研究生录取存在性别歧视吗?) 1973 秋季,美国加州大学伯克利分校研究生招生录取工作中,有 12763 名学生申请,男

^① 本书关注因果关系,所关注的变量具有明确的干预,从而关注的结果变量会有多个状态的水平。实验研究中干预状态分配是由研究者控制的,观测研究(observational studies)中干预不是由研究者控制的。另一类研究,没有干预,从而结果变量只有一个水平,这种研究称为描述性研究(Holland and Rubin, 1982),比如研究男性身高比女性身高高多少,今年我国的宏观经济形势如何等。

生 8442 人,女生 4321 人。最终录取 5232 人,其中男生录取 3738 人,女生录取 494 人,男生录取率为 44%,女性录取率为 35%。男生录取率高于女生,因而一些人认为伯克利分校研究生录取中存在性别歧视。真的存在性别歧视吗? Bicke et al. (1975)对此进行了研究,发现尽管总体上,男生录取率高,但如果按照专业划分之后,几乎每个专业中均是女生的录取率更高。表 1.2 给出了两个专业的假想数据,^①总体上看,男生录取率为 44%,高于女生的录取率 38%,但是每个专业中,女生的录取率都更高。为什么总体而言,男生录取率更高呢?因为男女生的专业分布不同,男生大多数集中于申请录取率相对较高的(或是冷门)专业 2,而女生却集中于相对录取率较低的专业 1。

表 1.2 伯克利分校的研究生录取率

专业	男生			女生		
	申请人数	录取数	录取率	申请人数	录取数	录取率
专业 1	200	40	20%	400	120	30%
专业 2	800	400	50%	100	70	70%
合计	1000	440	44%	500	190	38%

上面两个例子告诉我们,简单的直观印象往往是错误的,真理的揭示是非常困难的,需要更加深入地分析,才有可能找出现象背后的真相。同时也说明,寻找真理的过程是非常吸引人的。

第二节 因果推断简史

在哲学领域很早就有关于因果的讨论,但真正可以操作的因果概念还是来自于统计学领域。统计学领域流行的观点是统计工具只能考察和识别相关关系,不能考察因果关系,相关关系不是因果关系。Neyman (1923, 1990)在研究农业实验中首次提出了潜在结果的概念,但由于其论文是用波兰语写成的,并没有引起学术界的关注。^② Fisher (1935)提出,随机化实验(randomized experiments)是因果推断的基础。20 世纪 70 年代以来,Rubin (1974, 1977, 1978)的一系列论文,重新独立地提出了潜在结果的概念,并将之推广到观测研究,从而构造出适用于随机化实验和观测研究的基本分析框架,Rubin 提出的潜在结果框架被称为 Rubin 因果模型(Holland, 1986)。^③ Rubin 因果模型(Rubin Caus-

① 详细数据请参考 Bickel et al. (1975)。

② 直到 1990 年,该文才被译成英文,发表在 Statistical Science 上。

③ 有时也称为 Neyman-Rubin Model (Pearl, 1996)。