



格致方法·定量研究系列 吴晓刚 主编

计算机辅助访问

【荷】威廉·E. 萨里斯 (Willem E. Saris) 著
武玲蔚 译 周穆之 校

- ★ 革新研究理念
- ★ 丰富研究工具
- ★ 最权威、最前沿的定量研究方法指南

格致出版社  上海人民出版社

41

格致方法·定量研究系列 吴晓刚 主编

计算机辅助访问

[荷] 威廉·E.萨里斯 (Willem E.Saris)

武玲蔚 译

周穆之 校

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

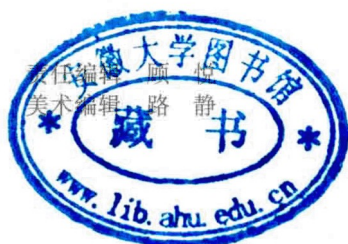
计算机辅助访问/(荷)萨里斯(Saris, W.E.)著;
武玲蔚译. —上海:格致出版社;上海人民出版社,
2014

(格致方法·定量研究系列)

ISBN 978-7-5432-2422-3

I. ①计… II. ①萨… ②武… III. ①数据收集-计
算机辅助技术-研究 IV. ①TP391.7

中国版本图书馆 CIP 数据核字(2014)第 145610 号



格致方法·定量研究系列

计算机辅助访问

[荷]威廉·E.萨里斯 著

武玲蔚 译 周穆之 校

出版 世纪出版股份有限公司 格致出版社
世纪出版集团 上海人民出版社
(200001 上海福建中路 193 号 www.ewen.co)



编辑部热线 021-63914988
市场部热线 021-63914081
www.hibooks.cn

发行 上海世纪出版股份有限公司发行中心

印刷 浙江临安曙光印务有限公司
开本 920×1168 1/32
印张 4.75
字数 92,000
版次 2014 年 9 月第 1 版
印次 2014 年 9 月第 1 次印刷

ISBN 978-7-5432-2422-3/C·106

定价:20.00 元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书,精选了世界著名的 SAGE 出版社定量社会科学研究丛书,翻译成中文,起初集结成八册,于 2011 年出版。这套丛书自出版以来,受到广大读者特别是年轻一代社会科学工作者的热烈欢迎。为了给广大读者提供更多的方便和选择,该丛书经过修订和校正,于 2012 年以单行本的形式再次出版发行,共 37 本。我们衷心感谢广大读者的支持和建议。

随着与 SAGE 出版社合作的进一步深化,我们又从丛书中精选了三十多个品种,译成中文,以飨读者。丛书新增品种涵盖了更多的定量研究方法。我们希望本丛书单行本的继续出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总序

2003年,我赴港工作,在香港科技大学社会科学部教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来,非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课,而第二门课“社会科学中的定量分析”为博士生的必修课(事实上,大部分硕士生修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候,根据社会科学研究生的数理基础比较薄弱的特点,尽量避免复杂的数学公式推导,而用具体的例子,结合语言和图形,帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上,即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”,学完这些课程后,我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章;另一方面,也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容,尽管在线性回归模型的内容上有少

量重复,但各有侧重。“社会科学里的统计学”从介绍最基本的社会研究方法论和统计学原理开始,到多元线性回归模型结束,内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型,以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法,将重点放在因变量为定类数据的分析模型上,包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型,包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生,一直鼓励和支持我将两门课的讲稿结集出版,并帮助我将原来的英文课程讲稿译成了中文。但是,由于种种原因,这两本书拖了多年还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩,每本书都写得通俗易懂,与我的教学理念是相通的。当格致出版社向我提出从这套丛书中精选一批翻译,以飨中文读者时,我非常支持这个想法,因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力,还要有对实质内容有较深的理解能力,而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容,只有语言能力是远远不能胜任的。在短短的一年时间里,我们组织了来自中国内地及香港、台湾地区的二十几位

研究生参与了这项工程,他们当时大部分是香港科技大学的硕士和博士研究生,受过严格的社会科学统计方法的训练,也有来自美国等地对定量研究感兴趣的博士研究生。他们是香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛,硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊,应用社会经济研究中心研究员李俊秀;香港大学教育学院博士研究生洪岩璧;北京大学社会学系博士研究生李丁、赵亮员;中国人民大学人口学系讲师巫锡炜;中国台湾“中央”研究院社会学所助理研究员林宗弘;南京师范大学心理学系副教授陈陈;美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛;美国加州大学洛杉矶分校社会学系博士研究生宋曦;哈佛大学社会学系博士研究生郭茂灿和周韵。

参与这项工作的许多译者目前都已经毕业,大多成为中国内地以及香港、台湾等地区高校和研究机构定量社会科学方法教学和研究的骨干。不少译者反映,翻译工作本身也是他们学习相关定量方法的有效途径。鉴于此,当格致出版社和 SAGE 出版社决定在“格致方法·定量研究系列”丛书中推出另外一批新品种时,香港科技大学社会科学部的研究生仍然是主要力量。特别值得一提的是,香港科技大学应用社会经济研究中心与上海大学社会学院自 2012 年夏季开始,在上海(夏季)和广州南沙(冬季)联合举办《应用社会科学研究方法研修班》,至今已经成功举办三届。研修课程设计体现“化整为零、循序渐进、中文教学、学以致用”的方针,吸引了一大批有志于从事定量社会科学研究博士生和青年学者。他们中的不少人也参与了翻译和校对的工作。他们在

繁忙的学习和研究之余,历经近两年的时间,完成了三十多本新书的翻译任务,使得“格致方法·定量研究系列”丛书更加丰富和完善。他们是:东南大学社会学系副教授洪岩璧,香港科技大学社会科学部博士研究生贺光烨、李忠路、王佳、王彦蓉、许多多,硕士研究生范新光、缪佳、武玲蔚、臧晓露、曾东林,原硕士研究生李兰,密歇根大学社会学系博士研究生王骁,纽约大学社会学系博士研究生温芳琪,牛津大学社会学系研究生周穆之,上海大学社会学院博士研究生陈伟等。

陈伟、范新光、贺光烨、洪岩璧、李忠路、缪佳、王佳、武玲蔚、许多多、曾东林、周穆之,以及香港科技大学社会科学部硕士研究生陈佳莹,上海大学社会学院硕士研究生梁海祥还协助主编做了大量的审校工作。格致出版社编辑高璇不遗余力地推动本丛书的继续出版,并且在这个过程中表现出极大的耐心和高度的专业精神。对他们付出的劳动,我在此致以诚挚的谢意。当然,每本书因本身内容和译者的行文风格有所差异,校对未免挂一漏万,术语的标准译法方面还有很大的改进空间。我们欢迎广大读者提出建设性的批评和建议,以便再版时修订。

我们希望本丛书的持续出版,能为进一步提升国内社会科学定量教学和研究水平作出一点贡献。

吴晓刚

于香港九龙清水湾

序

当前,测量公共意见出现了革命性的进展,其中心就是计算机辅助访问。从 20 世纪 30 年代盖洛普公司声誉鹊起开始,科学的调查研究主要通过对受访者的面对面、纸笔采访进行。然而,在 20 世纪 80 年代,计算机辅助电访(CATI)开始替代这种面对面的方式。与之前在受访者家中进行采访不同,CATI 的访问员在办公室中,面对着终端,通过电话对受访者进行提问,并将答案输入电脑。在美国,当今大多数全国范围的调查都包含某些形式的 CATI,很少采用面对面访问。

在这本小册子中,萨里斯教授就对上面提到的 CATI 以及其他形式的计算机辅助数据搜集方式(CADAC)进行了介绍。比如,他在第 2 章讨论了计算机辅助个人访问(CAPI)、计算机化的自我管理问卷(CSAQ)、计算机辅助的面板研究以及按键式数据输入(TDE)。一些计算机辅助访问技术的支持者可能会草率地认为这一方法比传统方法更快、更经济。然而,萨里斯教授对此的意见则较为谨慎,他仅认为这一方式可以提高潜在的数据质量。比如,当受访者接受采访

的时候,数据清理就可以进行;而当研究问题替换带来次序效果时,研究者可以方便地将这一替换进行随机化;同时,访问员的行为可以随时被检测,而编码异常值也可以随时处理。

然而,运用计算机辅助访问并不是没有困难的。一个主要的挑战在于问卷设计,对这一点,萨里斯教授进行了详细的讨论(第3章)。另一个重要的问题是如何在众多程序中进行选择。在第4章中,作者讨论了CADAC的一系列硬件和软件的选择。

尽管计算机辅助访问的出现并没有很长的历史,但它给传统方法中的一些问题提供了非常好的解决方案。比如,在处理面板研究中的高退出率(high-attrition problem)问题时,萨里斯教授使用了电访系统(tele-interview system),对荷兰家庭的一个随机样本的访问在其家中的个人电脑上进行,他们的回答会自动返回中央计算机中。有了这样的系统,面板研究就能够得到多期的数据,且退出问题不大。这个例子以及其他CADAC的优势,在萨里斯教授的这本小册子中都有讨论。

迈克尔·S.刘易斯-贝克

致谢

我想感谢我的博士生们在本书写作过程中与我不间断地讨论。我还想感谢我的同事迈克尔·S.刘易斯-贝克、罗伯特·格鲁夫、德克·斯科尔、米克·库珀、卡罗尔·豪斯、哈姆·哈特曼,以及两位匿名评审者对本书提出的意见。

前 言

任何参与过调查研究的人都知道,这项工作需要投入大量的人力、物力和财力。人们需要设计、撰写、打印并通过信件发送问卷,等待受访者回答并将其寄回,然后对回答进行编码,输入计算机并且对其进行检查。在这些步骤之后,数据的分析工作才能开始。

在数据搜集过程中使用计算机能够减少相当一部分工作量。在计算机辅助数据搜集(CADAC)过程中,访问程序能够在电脑屏幕上展现问题,并对受访者的回答进行即时记录。通过这种方式,我们可以省去打印和邮寄问卷、对答案的编码以及数据录入的麻烦。如果我们还能够对访问过程进行更精心的设计,数据核查与输入工作就可以得到相当的缩减。从这个意义上看,计算机辅助访问是更快、更经济的。然而,尼科尔斯和格罗夫斯(Nicholls and Groves, 1986)则指出,这一论点的论据并不充足。

CADAC 被推崇还有另一个原因。尼科尔斯(Nicholls, 1978)、格罗夫斯(Groves, 1983)、芬克(Fink, 1983)、德克尔和多恩(Dekker and Dorn, 1984)以及其他研究者都认为

(CADAC)会提高数据质量。他们认为,通过自动跳转和编码、一致性检查以及其他方法,数据质量能够得到提高。然而,这方面很少有相关的支持性证据(Groves and Nicholls, 1986)。

然而,CADAC 的使用每年都在增长。这一革命从 20 世纪 70 年代的计算机辅助电话访问(CATI)开始,现在已经在商业机构、大学和政府中被广泛使用(Nicholls and Groves, 1986; Spaeth, 1990)。CATI 的总数是未知的,但是在全世界范围内大概多于 1 000(Gonzalez, 1990)。在 1988 年,美国政府就已经有超过 50 处安装 CATI 了。现在,国家农业统计服务处每年进行超过 125 000 个 CATI 访问。

在此之后,人们又引入了计算机辅助个人访问(CAPI)技术。美国和欧洲的政府机构(Thornberry, Rowe, and Bigger, 1990; van Bastelaer, Kerssemakers, and Sikkels, 1998)和欧洲的市场调查公司都开始运用这一技术来搜集数据。在荷兰,荷兰统计局开发的 CAPI 程序发展很快,从五年前的接近于零到 1990 年每月增长 3 000 个用户。

所有大型商业公司也有了 CAPI 设备。我们可以从大多数西欧国家和美国看到相同的趋势。同样,我们也能在商业领域的实验中看到计算机辅助访问的应用,其中没有访问员的参与。所有这些都是为面板研究设计的。一些人使用了视讯系统(Clemens, 1984),其他人使用了家庭计算机(de Pijper and Saris, 1986a)或者电脑网络(Gonzalez, 1990; Kiesler and Sproull, 1986)。人们也开始尝试使用按键电话(Clayton and Harrel, 1989)和语音识别(Winter and Clayton, 1990)。后两个程序仅仅被小规模地应用于商业中。有了家庭电脑系统,也即电访,荷兰每年可以进行 150 000 次访问。

CADAC 的持续发展表明下列三个优势中至少一个是成立的:速度的加快、成本的减少以及数据质量的提高。对于不同的项目,它们可能还会同时成立。一方面,如果人们并不想提高数据质量,CADAC 无疑也会快速便宜地实现这一目标。另一方面,通过使用 CADAC 的新技术,复杂研究可以更容易地进行,但是可能并不会比传统项目更迅速、更便宜。对于本书而言,许多问题都是很有意思的。比如:

- 为了维持电脑系统,我们需要考虑额外设备和额外雇员的成本收益对比。
- 在访谈过程中需要的额外技术,以及对访问员的培训时间以及无应答的处理。
- 程序的弹性。
- CADAC 系统的引入对研究机构的影响。

然而,在本书中,我们并不想讨论这些问题。我们希望集中讨论如何提高数据质量,因为这一点是所有社会科学家在搜集数据的过程中所关心的核心问题,而使用 CADAC 的确可以提高数据质量。一个这方面的例子是由托尔托拉(Tortora, 1985)给出的,他说明在普通调查中,有 77% 的数据错误需要通过再次联系受访者确认,而使用 CADAC 进行在线核查的话就可以解决这些问题(具体来说,该程序在访问过程中检查并要求改正不一致性)。然而,这一优势也并不容易获得:在完成问卷设计之后,还有很多的工作需要进行。在这一方面,研究者需要考虑以下几个方面(House and Nicholls, 1988):

- 跳转指令。
- 指定填充。
- 在线核查一致性和异常值。
- 帮助界面。
- 输入问题答案。
- 记录答案的方式。

在这一过程中,我们可以用到更多的方法。但有一件事是肯定的:计算机辅助访问不仅需要很多教科书中提到的问卷设计技术(比如,Converse and Presser, 1986; Sudman and Bradburn, 1974),还需要很多新的技术。这本书的目的就是提供给读者有关这些技术的一些概念,以及在这一过程中会遇到什么问题的指示。

在第1章中,我们介绍了计算机辅助访问的概况。在第2章中,我们讨论了计算机辅助访问在访问员管理的访问中、受访者自主管理的访问中以及面板调查中的技术运用。在第3章中,我们讨论了CADAC问卷的设计。在第4章中,我们对计算机辅助访问程序的主要特征进行了介绍,特别是当人们希望买入这些程序的时候。

我们还给出了研究问卷的一些例子。这些例子并不局限于某个特定的访问程序。我们主要关心的是它的逻辑,而不是具体内容。在本书中,我们对这些例子给出了相应的注释(Baker and Lefes, 1988)。在这种问卷中,它们与普通纸面问卷的关系是明确的,同时我们也额外给出了针对计算机辅助访问的一些批注。有很多软件用到了非常不同的程序,但理解本书中讨论的程序对于问卷设计者而言已经够了。

目 录

序	1
致谢	1
前言	1
第 1 章 计算机辅助访问	1
第 1 节 CADAC 作为纸面问卷的替代	5
第 2 节 CADAC 作为对访问员的部分替代	11
第 3 节 CADAC 可以比访问员做得更出色	18
第 4 节 CADAC 的缺点	23
第 5 节 总结	25
第 2 章 计算机辅助访问的应用	27
第 1 节 访问员管理的调查研究	35
第 2 节 自我管理调查研究	48
第 3 节 计算机协助的面板研究	60

第 3 章	问卷编写	75
第 1 节	问卷的模块化	81
第 2 节	模块的设计	83
第 3 节	问卷设定	87
第 4 节	测试代码和问卷分流	91
第 5 节	在预调查中测试问卷	94
第 6 节	测试数据	97
第 4 章	CADAC 的硬件与软件	99
第 1 节	CADAC 的硬件	101
第 2 节	CADAC 软件	106
附录		117
参考文献		120
译名对照表		128