

万卷方法

实用抽样方法

SHIYONG CHOUYANG FANGFA

■ 加里·T.亨利 著 沈崇麟 译



重庆大学出版社

<http://www.cqup.com.cn>

万卷方法

实用抽样方法

SHIYONG CHOUYANG FANGFA

■ 加里·T.亨利 著 沈崇麟 译

重庆大学出版社

Authorized translation from the English language edition, entitled PRACTICAL SAMPLING, by Gary T. Henry, published by Sage Publications, Inc., Copyright © 1990 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

CHINESE SIMPLIFIED language edition published by CHONGQING UNIVERSITY PRESS, Copyright © 2006 by Chongqing University Press.

实用抽样方法,作者:加里·T.亨利。原书英文版由 Sage 出版公司出版。原书版权属 Sage 出版公司。

本书简体中文版专有出版权由 Sage 出版公司授予重庆大学出版社,未经出版者书面许可,不得以任何形式复制。

版贸渝核字(2006)第 104 号。

图书在版编目(CIP)数据

实用抽样方法/(美)亨利(Henry, G. T.)著;沈崇麟译. —重庆:重庆大学出版社,2008. 8

(万卷方法)

书名原文:Practical Sampling

ISBN 978-7-5624-4487-9

I. 实… II. ①亨…②沈… III. 抽样调查—调查方法

IV. C811

中国版本图书馆 CIP 数据核字(2008)第 058461 号

实用抽样方法

加里·T.亨利 著

沈崇麟 译

责任编辑:雷少波 罗杉 版式设计:雷少波
责任校对:贾梅 责任印制:赵晟

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街 174 号重庆大学(A区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址: <http://www.cqup.com.cn>

邮箱: fxk@cqup.com.cn (市场营销部)

全国新华书店经销

四川省内江市兼升印务有限公司印刷

*

开本:940×1360 1/32 印张:4.75 字数:145千

2008年8月第1版 2008年8月第1次印刷

印数:1—4 000

ISBN 978-7-5624-4487-9 定价:23.00 元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书
制作各类出版物及配套用书,违者必究

作译者 简介

加里·T. 亨利 (Gary T. Henry) 在美国弗吉尼亚大学公共管理学教授研究方法、统计学、项目评估和政策分析等课程。曾担任弗吉尼亚公立学校教学评估指标体系研究的项目负责人和中国某大学统计学客座教授。此外,他也是国家议会全国会议的成员和演讲人。

早年在威斯康星大学取得博士学位之后,他进入了弗吉尼亚审计和评审委员会,在那里担任首席方法专家,用定量方法进行教育、惩处和交通问题的研究。随后在州内阁担任教育部副部长。他的研究兴趣涉及稳健估计、实地因果关系评估和公正评测等方面。亨利博士是《评估评论》(*Evaluation Review*)和《公共管理评论》(*Public Administration Review*)等杂志的非常活跃的撰稿人和评审人。他也是美国评估协会规范和道德委员会的主席。

沈崇麟 中国社会科学院社会学研究所研究员,中国社会科学院研究生院教授,“第二、三、四批百县市国情调查问卷调查”、“中国城乡居民生活调查”、“中国城乡家庭变迁调查”和“中国城乡社会变迁调查”等研究项目主持人。

致 谢

没有诸多同事和同行的指教和帮助,就不会有《实用抽样方法》一书的存在。玛丽·斯图茨曼(Mary Stutzman)、西摩·苏德曼(Seymour Sudman)和格里格·瑞斯特(Greg Rest)的评阅使我受益匪浅。感谢丛书的两位编辑列昂纳多·贝克曼(Leonard Bickman)和黛博拉·罗格(Debra Rog)的鼓励和盛情邀请。肯特·迪克西(Kent Dickey)和弗吉尼亚·黑廷格(Virginia Hettinger)认真校阅了文字、公式和计算结果。堤姆·亨德里克(Tim Hendrick)和帕特·斯特里诺(Pat Storino)费心为各种统计图润色。休·马汉(Sue Mahan)和安妮·科尔茨(Annie Kurts)做了大量的文字整理工作。最后,我还要感谢妻女的信任和支持。她们从不抱怨我将周末用来写作,而不是修理家中的庭院,她们也从来没有怀疑过,我的付出终将得到回报。

目录

第1章 导 论	1
定义样本	3
抽样和效度	4
为什么要抽样	6
本书概要	8
第2章 样本选择方法	10
非概率抽样	10
非概率样本的用途	17
概率抽样	19
结论	26
第3章 实用样本设计法	28
抽样设计中的误差源	29
实用抽样设计的框架	40
抽样前抉择	43
研究的性质究竟是什么——探索性的、描述性的还是分析性的	43
最感兴趣的变量是什么	44
研究的目标总体是什么	44
某些子总体或特定的群体是否对研究很重要	44
用什么样的方法收集数据	44
是否宜于抽样	45
抽样抉择	45
什么样的目标总体清单可作为抽样框使用	45
可容忍的误差或估计的效应的大小是多少	46
采用什么类型的抽样技术	47

选择的概率是相等的还是不等的	48
选入样本的单位是多少	48
抽样后抉择	51
如何评估无回答问题的影响	51
样本数据是否需要加权	52
研究变量的标准误差是什么	52
小结	54
第4章 实用样本设计的四个实例	55
北卡罗莱纳州居民调查	57
抽样前抉择	57
抽样抉择	58
抽样后抉择	60
佛罗里达高龄老人调查	63
抽样前抉择	63
抽样抉择	64
抽样后抉择	67
弗吉尼亚出院精神病人调查	68
抽样前抉择	69
抽样抉择	70
抽样后抉择	72
调查研究中心的美国本土住户样本	73
抽样前抉择	74
抽样抉择	75
抽样后抉择	77
小结	79
第5章 抽样框	81
一般总体的抽样框	82
特殊总体的抽样框	84
总误差和抽样框	85
结论	92
第6章 抽样方法	93
简单随机抽样	94
系统抽样	96

分层抽样	98
整群抽样	105
多级抽样	108
小结	114
第 7 章 样本容量	116
高效样本的容量	118
样本容量设计的实质	121
子总体分析	123
不合格和无回答的修正	124
费用	125
信度	126
小总体抽样	127
小结	129
第 8 章 抽样后选择	130
权的使用	130
无回答评估	132
陈述数据	134
结论	136
参考文献	137

大多数用于社会和政策科学的数据都收集自样本。公众舆论调查、社会实验和教育改革评估等方面的研究一般都要采用抽样方法。任何须将从调查的对象或单位得到的结论外推到更大的研究总体的规范的调查研究都要使用抽样方法。我们经常对样本进行调查,很少对整个研究总体进行调查。诸如十年一度的人口普查这样的总体调查是不多见的。收集用作项目评估的数据若不以抽样方法为依据,那么新的教学方法或社会服务设施的交付使用的风险和费用是很难估算的。同样,如果没有抽样方法,我们也无法对新项目的效力进行科学的评估。

在20世纪末叶,概率抽样技术的引进和使用极大地推动了经验的社会和政策科学研究的发展。尽管抽样方法是如此重要,但是社会和政策科学领域中的研究者却很少有这方面的实际经验,因而十分希望能得到有关抽样的实际应用的指导。有关抽样方法的一些假设常常构成了研究者希望采用的分析技术的基础。研究者经常都需要证明那些假定都是成立的。抽样常令我们处于两难的境地。一方面,限于时间和经费,研究者无法收集一个特定的研究所感兴趣的整个群体或总体的数据。然而另一方面,研究者和研究成果的使用者通常所感兴趣的是总体,而非总体的某一子集。正因为如此,克服时间和费用的限制与我们对于总体的信息的需求这一矛盾,即将来自总体的某一子集或“样本”的研究发现推广到整个总体便变得如此的关键和重要。而有关抽样的逻辑和方法的知识正是确保我们是否能将来自研究的参与者或被调查人的结论,合理地推广到他们所代表的总体的基础。

对抽样方法和它的基本含义有所了解,这一问题在政策研究中尤其重要。在政策研究的实施和使用过程中,所遇到的问题通常都与抽样设计方案的选择问题有关:政策和项目所定义的目标总体是否与研究所定义的相同?我们使用的选取对象和单位的方法是否会使决策者所需要的估计值产生偏倚?来自样本的估计值,就研究的目的而言是否已经足够精确?表 1.1 简要地列出了这三个问题,用于判断研究的有效性的标准和研究结果使用的潜在含义。

我们将用一个有关高龄老人的服务需求的评估的例子,来揭示不同的抽样设计对评估结果的影响:

- **总体的定义。**本研究旨在对某州全体高龄老人的服务需求进行评估。如果样本抽取自一个由目前正在接受各种公开提供社会和医疗服务的机构的服务的高龄老人组成的总体,那么那些目前没有接受这些机构提供的服务,但需要这些服务的老人将被排除在样本之外。这样的选择势必会导致对实际需求的低估。
- **抽样方法。**采用一种只关注居住在集体宿舍(在这里集体宿舍指养老院、疗养院等,而不是我们通常理解的那种集体宿舍,译者注)中的老人的抽样策略将会使结果有偏倚。在许多情况下集体宿舍提供的照料水平都是最高的。因而在将这样的需求的估计值外推到州的整个的老龄人口时,可能是偏高的。这一样本可能没有包括足够的独自居住的或与自己的家人生活在一起的自己照料自己的老人。
- **估计的精度。**来自样本的估计值,例如均值或比例数,即使在抽样方法是无偏的时候,也是会有一定的波动的。一个未被告知实情的研究结果的用户可能认为估计值是严格而精确的,因而对“精确”的估计值过于信任。例如,一个来自需求评估的估计值显示,有 63.4% 的高龄老人需要服务。而对于一个小样本来讲,研究者可能有理由认为真的均值在 54% 到 73% 之间,假定在一个 500 000 名高龄老人的总体中,服务人口数的估计值高低的波动范围,几乎达到了 100 000。这说明估计值是很不精确的,因而我们也无法据此来制定相应的政策和建立为高龄老人提供服务的项目。

表 1.1 政策研究中的样本设计问题

问题	判断标准	含义
总体的定义	目标总体和研究总体的 一致性	因为将那些并非目标总体的成员包含在内,或遗漏了目标总体的成员而导致的研究总体的偏倚
抽样方法	选择样本时,给研究总体 中的任何成员以相等的 选择的可能	如果研究总体中的某些成员的被选择的可能性比其他成员更高,抽样方法便将导致结果的偏倚
估计的精度	估计精确度足以满足政 策制定的需要	所有的样本产生的估计数都不是确定的数字。缺乏精确度势必影响政策的制定

定义样本

在科学和日常语言中,“样本”一词常以各种不同的方式使用。例如马克和沃克曼(Mark and Workman, 1987, p. 47)指出:“对化学家来讲,样本一词也许会令他们联想起一堆东西,一个盛有某种液体、糊状物的容器,或一堆难以言状,但成分却是确定的东西。”对于化学家来讲,样本具有内在的重要性。化学家手头得到的那些东西,例如一个犯罪现场或从某种较大的东西上选取的一部分可能就是全部问题之所在。对化学家而言,他们的根本目的就是确定样本的成分。化学家在法庭上的职责是发现茶中的砒霜,而不是茶的总体的代表。

化学家的样本可被视作一种标本,因而特例是很重要。相反,在研究文献和本书中,样本则是指总体中用来得到整个总体的信息的一个子集。在这个意义上讲,样本是总体的模型。一个上佳的样本将会完好地代表总体。社会或政策科学家的真正兴趣并不是样本——样本只不过是他们用来了解总体的工具而已。

以上的讨论自然而然地会产生以下两个问题:

- 如何选择一個用于代表总体的样本?
- 如何判断样本是否完好地代表了总体?

有关第一个问题的基本要点我们将在下一章进行讨论,此后,这一讨论将贯穿全书。在第3章,我们将对样本和样本设计过程中选择的抽样框中可能存在的误差源进行扼要的介绍。

在我们转入这些问题的讨论之前,有必要对我们使用的“有代表性”一词做一些解释。一个样本是用来代表总体的,因此它是总体的一个模型或代表。我们将“有代表性”这一术语加在样本上,就如我们日常所言的“有代表性的样本”一样,并未给样本提供任何额外的信息。形容词“有代表性”并不是技术上的定义,只不过是这一术语的使用者的一种主观判断而已。我们并未因此而建立确定一个样本究竟是有代表性的还是没有代表性的标准。尽管如此,“有代表性的样本”仍然是人们给自己所提供的样本所做的最为常见的描述。为了避免不必要的误解,我们建议在样本和总体之间的对应问题的描述上去掉这一形容词,改为采用一些能对样本选择过程进行描述的词。有关这样的描述的重要性,我们将在随后的一章进行讨论。

抽样和效度

很少有研究者能在一个特定的研究中,收集所有自己感兴趣的对象的数据。而样本则为我们提供了切实可行和有效地收集数据的工具。样本好比总体的模型,然而研究者若要将研究的发现推广到总体,则必须要使模型是总体的精确的代表。

研究者或某一项研究成果的使用者将研究的发现推广到样本个体、时间和地方之外的能力称作“外部效度”(external validity)(Campbell & Stanley, 1963; Cook & Campbell, 1979)。库克和坎贝尔认为外部效度的关键问题是“如果从建构A到建构B可能存在某种因果关系,那么在跨越不同的人、场合和时间时,这种关系究竟在多大程度上可以被普遍化呢?”(Cook & Campbell, 1979, p. 39)例如,一些研究者发现,在三年级阅读课程中使用计算机辅助教学软件包,能使位于市中心的学校的学生的词汇量 and 理解力有所提高。这一研

究成果的用户有理由提出一连串的问题:诸如这样的教学方法是否在农村地区的学校也同样有效?在四年级阅读课中是否也有效?学习成绩的提高是否源自计算机使用的新奇?一旦学生不再对计算机感到新奇,学习成绩便不再会因为它们的使用而提高?

对研究的发现进行概括的能力是数据实际从中收集的样本的一个功能。抽样设计和实施都会对诸如这样的概括能力有所影响。本书所介绍的实用抽样方法对设计和实施二者予以同样的重视,因为这二者都可能对研究的效度或总误差有所影响。样本设计的内容包括:挑选恰当的选样方法,如随机数码拨号法;确定研究所需的个案数目等。将一个设计付诸实施则包括:得到有关研究总体的完整的清单;收集真实可靠的数据;确保调查答案的确收集自一个其组成能精确地代表总体的群体。任何有可能影响数据实际收集的群体的组成的计划或行为,都关系结果的概括与推广。正因为如此,切实可行的抽样设计必须与整个研究设计和实施整合为一体。

除了外部有效性之外,我们还必须考虑其他两种统计效度问题。这两种统计效度会直接对样本设计有所影响,反过来,样本设计也会对它们有所影响。这两种统计效度问题的讨论见诸于库克和坎贝尔的有关著作(Cook & Campbell, 1979)。统计结论的效度是一种得出有关出现在样本数据中的关系,即所谓的共变的结论的能力。通常我们用统计检验来考察观察到的关系是否由变化引起的,或者如克雷默和西曼所言:“在将统计检验运用于数据时,它定义了一条规则,一条是否能拒绝零假设(null hypothesis),即证据是否是不容置疑的规则。”(Kraemer & Thiemann, 1987, p. 23)因为这些检验对关系的大小(效应的大小)和样本的大小二者都十分敏感,所以样本的大小可能对避免得出“伪的共变结论”是至关重要的(Cook & Campbell, 1979, p. 37)。

在进行统计检验时,小的样本量可能导致某种保守的偏倚(第二种类型的误差)。当实际上零假设是不成立的,但却未能被否定的时候,第二种类型的误差便会发生。在遇有这样的情形时,一个被检验的项目或干预政策将被认为是无效的,尽管实际上它是有效的。然而“合理的怀疑”标准可能无法满足给定的期望效应量和实际的样本量。在发生保守的偏倚时,较小的效应或共变虽然的确存在,但是样本量却不足把效应记录到统计显著性的界限之内。在对一个参与人

数少和样本量比较小的试行项目进行评估时,发生这样的现象尤其令人感到沮丧。项目产生的虽然不太大,但却是很有意义的效应,可能会因为缺乏统计的显著性而被吞噬。这样,评估者可能就会因此而错误地认为项目是无效的。我们将会在第7章对样本容量和统计结论的效度问题进行讨论(希望对这一问题有更多了解的读者,可参见 Kraemer & Thiernann, 1987; Oakes, 1986; or Lipsey, 1989)。

有关统计结论的效度的第二个方面的问题——量度的信度(reliability of measures)是又一个我们在抽样设计和实施时应该考虑的问题。数据中观察到的变差越大,我们所使用的工具的信度就越低(Cook & Campbell, 1979)。每当观察到的变差加大的时候,尽管真关系的确存在,虚无假设的否定也会因此变得更加困难。假如我们使用的工具是无偏的,那么比较大的样本量可在一定程度上对加大的变差有所补偿。然而为了补偿因为工具的信度的缺乏而导致的方差的波动,我们必须在设计阶段的初期就对这一问题有所了解。

为什么要抽样

抽样涉及的信度问题既然如此复杂,那么研究者就可能会因此而提出“我们为什么要抽样?”这样一个问题。其实,抽样无非是达到我们最终目的的一个切实可行的手段。研究者的工作通常都始于一个目标总体。而这一总体一般都取决于一个政策或项目。研究者通常都会对有关政策或项目提出一个问题。例如,一个研究者可能会提出“为处于风险中的四岁儿童(at-risk 4-year-olds)建立一个学前教育项目是否会提高这些儿童的认知能力,从而减少他们在以后的学习中对特殊教育援助的需求?”这样一个问题。如这样一个由制定政策的人确定的处于风险中的四岁儿童的目标总体将被纳入我们研究的问题中。

通过抽样,研究者将这一问题转变成一个可行的经验项目。不言而喻,我们不可能为所有处在风险中的四岁儿童提供这样的发展项目,并在他们身上进行数年的测试,以确定项目究竟有什么影响。限于资源,我们无法这样做。在这一例子中,我们受到了来自两个方面的条件的限制——项目方面和研究方面。一方面,寻找项目发展

所需的资助、设备和训练有素的工作人员就是很困难的。另一方面,得到数据收集与分析和随后进行的评估等需要的投资同样也不是一件容易的事。而耗费很多的公共资助,却未能对项目的影响做出评估的做法显然是很不妥当的。

抽样使我们得以使用总体的一个子集来对一个项目进行测试。我们之所以要进行抽样的主要原因固然是由于研究项目的资源是有限的,但是,同时也因为抽样也能使研究的质量有所改进。例如,由于能完全胜任三、四岁儿童的预调查的训练有素的研究人员人数有限,我们不得不聘用训练不足的工作人员,或对风险状态进行过于简单的测试,因而无法得到可靠的结果。而抽样则可以使我们得以集中使用资源,从而提高每一个工作人员收集的数据的数量和质量,同时最大限度地减少数据的丢失。

在有些场合,研究人员可能会发现抽样方法是不可取的。这种场合可能有很多,我们无法一一列举,但其中两种情况我们希望读者能够记住:小总体和有可能降低结果的可信性的抽样。在处理小总体时(成员数低于50),从整个总体收集数据会提高数据的可信度和可靠度。小样本数据中的单个极端个案或离群数据(outlier)的影响尤其显著,因而用总体数据来检验假设会使问题变得更加简单。同样,如果研究项目的用户了解到一个“独特的个案”已从样本中删除,那么结果的可信性同样会受到影响。诸如这样的问题大多发生在小总体中。因为这时研究项目的用户对总体单个成员的信息有着比较详细的了解。

在一个有关如何合理分配公共基金的研究中,如果抽样方法不当,有可能对研究结果的可信性产生不良的影响。例如,使用一个行政区域的样本——城市和郡县的样本来检验地区性特征之间的关系和对基金的需求,虽然可能有助于提高统计的效率,但在政治上却是不可取的。不仅如此,如果在分析中没有体现研究项目的用户(在本例中是立法者)所在的地区,那么研究结果的可信性在他们的心目中就会大打折扣。因为研究者无法明确地阐述,不会因立法者的家乡未在研究中得到体现而导致研究结果的可信性下降。在本例中,由于某一地区的缺失而使某些人趁机对项目提出质疑,导致研究结果可能因此而被弃之不用。

在为——一个其结果将在政治环境中使用的研究项目进行抽样时,

我们还应在考虑科学的信度问题之外,做深一层的考虑,即,考虑一下有关政治的信度问题。我们这样说,并不意味着政策研究是不需要抽样的。

在不进行抽样的场合,仍然会涉及一些与抽样设计有关的问题。例如,要得到一张目标总体的全面的清单便是抽样涉及的首要问题。对目标总体的所有成员进行调查的普查同样也需要一张全面的清单。研究中与目标总体的覆盖程度和无回答(即数据丢失的个案)有关的问题,与使用抽样方法和不使用抽样方法这两种研究都有关。

本书概要

本书使用的“实用抽样设计”一词,其含义已经超出抽样理论的范围。实用抽样设计包括抽样理论、设计的逻辑和设计方案的实施。样本设计的逻辑和它的实施贯穿整个研究的所有方面和全过程。研究的性质、测量指标和测量工具、数据收集方法、研究总体的定义和数据收集的方法等都会受到抽样方法的影响。实用抽样设计必须与整个研究方法整合为一体,以便尽可能地提高研究结果的效度。

重要的问题在于,对实用抽样设计问题的思考,必须联系目前人们对抽样设计问题的具体认识。我们将用在抽样顾问和研究团队共事之初,比较普遍发生的一些问题来对这一问题加以形象的阐述。通常研究团队一方开始都会问:“为了将样本推广到总体,我们需要一个多大的样本?”而抽样顾问则会反问:“你们研究的问题是什么?研究的对象是谁?”研究团队就抽样所提出的问题很多时候仅限于样本的大小。但为了提高效度和降低总误差,我们有关抽样问题的考虑决不能仅限于这一点,而必须设计研究的所有方面,贯穿整个研究过程。

本书所介绍的样本设计的思路,为我们提供了在整个研究过程中如何在若干备择设计方案中进行抉择的基础。我们将这一思路称之为实用抽样设计方法。我们之所以使用“实用”一词是因为我们为读者介绍的框架,强调的是各种可供我们选择的方案和如何在这些备择方案中进行选择的操作要领而非抽样的理论。本书将在概念上对这样的框架进行阐述的同时,还辅之以来自实际的抽样实践的详

尽实例。虽然本书并不打算过多地从理论和数学的角度来介绍抽样问题,但它介绍的内容都以以往的理论 and 数学的抽样著作为基础,而这些著作将为那些有志于对这方面文献作更深入了解的读者提供了一个完整的参考书目。

本书的主要阅读对象是那些将抽样作为自己的研究工具的研究人员。同样,本书也可作为社会和政策科学的本科生的方法课的补充教材,以帮助那些有志于从事研究工作的学生学习有关抽样的知识。除上述两个用途之外,本书也可用作那些在计划自己的研究项目时,在抽样方面需要一些咨询的研究人员的参考书。而那些计划进行抽取大的、复杂的样本的研究人员,则最好去寻求有经验的抽样专家的帮助。

本书第2章主要介绍了两种选择样本的方法——非概率抽样和概率抽样。我们通过若干种基本的设计方案对每种方法进行了介绍。第3章介绍实用抽样方法。它包括两个相对独立的部分:一部分讨论概率抽样中的总误差,另一部分则概要介绍了实用抽样方法的基本路数。这两个部分合在一起则会使我们进一步了解,为什么抽样设计必须要贯穿于一个研究的设计和实施的全过程。实际可供我们选择的方案,在这些方案中进行抉择,以及考虑到其他的抉择的可能性都是实用抽样设计方法的不可或缺的组成部分。

第4章详细介绍了摘自研究文献中的四个例子。这些实例是在第3章介绍的实用抽样方法的框架下组织起来的。这些例子介绍的各种研究,在总体类型、数据收集方法和样本设计上显示出了很大的差异。

其余三章则通过具体例子对实用抽样方法进行了进一步讲解。第5章主要介绍抽样框的问题,第6章则主要介绍抽样技术,而第7章讨论的问题是样本的容量。最后一章则对有关抽样后选择(postsampling choices)问题进行了讨论。