

◎ 高等学校统计学类系列教材

抽样： 理论与应用

◎ 金勇进



NLIC 2970624075



高等教育出版社
HIGHER EDUCATION PRESS

内容提要

高等学校统计学类系列教材
抽样：理论与应用

Chouyang: Lilun yu Yingyong

金勇进

ISBN 978-7-04-030081-9



出版发行 高等教育出版社
地址 北京市西城区德胜大街4号
邮政编码 100120
电话 010-58581118
010-82028818
010-810-0298
http://www.hep.edu.cn
http://www.hep.com.cn
www.landao.com.cn
www.landao.com.cn
www.widew.com



NLIC 2970624075



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

物料号 30081-00

内容提要

本书主要讲述如何进行抽样,才能得到一套科学的样本;怎样进行抽样设计,才能使抽样的效率最高;在不同抽样方法下,估计量的抽样误差如何计算以及怎样对抽样的效率进行评估;最后介绍抽样调查应用中经常遇到的问题。本书有理论阐述,也有实践中应用问题方法的讨论,主要特色表现在:抽样的方法体系完整;增添一些常用方法的内容;对复杂样本的方差估计方法进行了应用性讨论;增加统计分析的内容;对调查中的缺失数据进行了专门的讨论。

本书可以作为统计专业学生抽样调查课程的教科书,也可以作为非统计专业学生和各类人员学习抽样调查方法的教科书或学习参考书。

图书在版编目(CIP)数据

抽样:理论与应用/金勇进编. —北京:高等教育出版社,
2010.8

ISBN 978-7-04-030081-9

I. ①抽… II. ①金… III. ①抽样调查-高等学校-教
材 IV. ①C811

中国版本图书馆 CIP 数据核字(2010)第 125752 号

策划编辑	李蕊	责任编辑	李华英	封面设计	王凌波
责任绘图	宗小梅	版式设计	王艳红	责任校对	杨凤玲
责任印制	毛斯璐				

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120

购书热线 010-58581118
咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 蓝色畅想图书发行有限公司
印 刷 北京宏伟双华印刷有限公司

网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
畅想教育 <http://www.widedu.com>

开 本 787×960 1/16
印 张 24
字 数 450 000

版 次 2010年 8 月第 1 版
印 次 2010年 8 月第 1 次印刷
定 价 34.90 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 30081-00

前 言

抽样的理论与方法是统计学的一个重要分支。作为一门课程,它主要讲授如何进行抽样,才能得到一套科学的样本;怎样进行抽样设计,才能使抽样的效率最高;在不同抽样方法下,估计量的抽样误差如何计算;怎样对抽样的效率进行评估等。目前,抽样调查在我国的应用越来越广泛,已经成为人们获取信息的最主要方式之一。但我们也必须承认,在许多的抽样调查项目中,抽样方案的设计不尽科学,用样本数据推估总体方法不妥,更缺乏对抽样误差的计算,难以对推断结果进行评价。这说明,我们在抽样技术的应用方面,就整体而言,与国外相比还有不小的差距,这也从一个角度反映出,我们在抽样调查的理论与方法的知识推广方面还有许多工作可做。

教科书的主要功能是进行现代科学知识和技术的推广。在我国,有关抽样调查的教科书也有一些,这些教科书为传授抽样方法发挥了重要作用。但有相当一部分的教科书,过于追求数学公式的罗列和推导,缺乏结合应用背景的论述和讨论;还有一些教科书,沿袭多年传统的体系,缺乏现代新技术、新方法的引进,给人以知识老化之感。本书试图在原有教科书内容的基础上,努力做出一些探索。本书的特色主要表现在:

1. 抽样的方法体系完整。本书除第一章绪论外,从第二章到第九章分别介绍简单随机抽样、分层抽样、比率估计和回归估计、不等概抽样、整群抽样、系统抽样、多阶段抽样和二重抽样,这些内容涵盖了传统抽样理论中的经典抽样方法,也是一般教科书中所包含的内容。在对这些经典方法的介绍中,本书力求全面和精炼,既讲清楚每种方法的基本原理,又尽量避免复杂的数学公式推导,努力结合实际应用的背景进行论述。考虑到读者的不同数理知识基础,一些公式的证明放到附录中,使核心内容的介绍更紧凑。

2. 增添一些常用方法的内容。本书第十章“其他抽样方法”中,主要介绍了对总体单元数的估计,包括捕获再捕获法、逆抽样法;对敏感问题进行调查的技术,包括属性特征敏感问题的抽样设计和数量特征敏感问题的抽样设计;对稀有事件进行抽样的方法;如何进行自加权抽样设计及样本轮换设计。这些内容在实际调查中会经常遇到,但是在咱们已有的教科书中却很少提及。本书作者在

参考大量文献的基础上,对这些内容进行了提炼和归纳,希望这些内容对抽样调查的实践有所帮助。

3. 对复杂样本的方差估计方法进行了应用性讨论。本书第十一章是复杂样本的方差估计。目前许多抽样调查的教科书中,这部分内容是没有的。抽样调查的实际情况是,许多调查项目特别是大型的抽样调查项目,往往是复杂的抽样设计,从经典的抽样方法中,无法找到直接对应的估计量方差计算公式,这就需要针对特定的抽样设计,进行方差估计的公式推导,或者采用其他方法进行方差估计。在国外现代教科书中,这部分内容已经变得越来越不可或缺。国内一些学者意识到这个部分的重要,一些教科书中出现了这部分内容。本书在前人已经做出成果的基础上,试图有所前进。本章中,作者对复杂样本的方差估计方法进行了分类,对直接推导法进行了提炼式归纳,对不同方差估计方法的应用特点进行了比较性分析,希望能够丰富这个部分的内容。

4. 增加统计分析的内容。本书第十二章为“复杂抽样设计下的统计分析”。关于数据的统计分析方法常见于统计学的教科书中,本不属于抽样调查的教科书中的内容,所以,以往的教科书中是没有这部分内容的。但是,统计学教科书中所讨论的分析方法,都是针对简单随机抽样下产生的数据,而抽样调查的实践中,仅仅单纯使用简单随机抽样的情况是非常稀少的,实际常采用的是多种抽样方法的结合。这里把多种抽样方法结合产生的数据称为“复杂数据”。复杂调查数据通常很难满足观测值相互独立、服从于同一个总体分布等这些假定,因而这样的调查数据被录入计算机,不加任何处理,直接用标准模式的统计软件进行运算,无法保证结论的科学和正确。本章就是专门讨论如何对复杂数据进行统计分析的方法。这应该是抽样调查内容在数据处理、分析过程中的延续。过去,我们恰恰是忽略了数据采集和数据分析这个衔接点上存在的问题。

5. 对调查中的缺失数据进行了专门的讨论。本书第十三章是“调查中的缺失数据”,用一章篇幅专门讨论缺失数据是因为,缺失数据已经成为调查实践中面临的一个重要挑战,如果对缺失数据不进行任何有效处理,就有可能出现估计偏差和增大估计方差,如果问题严重,抽样调查的成果将付之东流。以往,我们对缺失数据问题重视不够,也研究不够,教科书中鲜有这方面的专门性讨论,现在看来,这个问题已无法回避,也不应回避。本书在介绍调查中缺失数据产生的原因、缺失数据的统计影响、缺失数据与样本量之间的关系之后,在缺失机制的基础上,重点讨论了处理缺失数据的两类方法,分别是加权调整法和插补法,以期对调查中的缺失数据起到补救作用。

本书可以作为统计专业抽样调查课程的教科书,也可以作为非统计专业学生和各类人员学习抽样调查方法的教科书或学习参考书。由于本书涉及内容较多,有理论阐述,也有实践中应用问题方法的讨论,不同背景的读者可以根据需

求自行取舍。

本书是集体劳动的成果,由我和我的博士生团队共同完成。其中,我负责全书编写大纲的设计,并负责第一章和第十三章的写作;其余各章由博士生们分别提供初稿,具体为:艾小青编写第二、五章,谢佳斌编写第三、十一、十二章,汤琰编写第四、十章,陶然编写第六、七章,孙欣、汤琰编写第八章,孙欣、陶然编写第九章,各章完成后又经过集体讨论和修改,最后由我做全书的修订和总纂。在本书写作中参阅许多文献,主要观点和方法的直接引用在书中有注释,并列于参考文献中,特向有关作者和出版社表示谢意。

在本书出版过程中,高等教育出版社数学分社社长李艳馥女士、编辑李蕊女士给予了大力支持,在此对她们表示由衷的感谢。

尽管我们尽了努力,但书中一定会有不少错误和缺憾。对于书中的不足,恳请各位专家和读者提出宝贵意见。

金勇进

2009年7月

目 录

第一章 绪论	1
§ 1 调查与抽样调查	1
§ 1.1 调查	1
§ 1.2 抽样调查	2
§ 1.3 抽样调查的作用	6
§ 1.4 抽样调查与普查	6
§ 1.5 抽样调查应用领域	7
§ 2 基本概念	8
§ 2.1 目标总体与抽样总体	8
§ 2.2 抽样框与抽样单元	9
§ 2.3 总体参数与统计量	10
§ 2.4 估计量方差、偏倚、均方误差	11
§ 2.5 抽样误差与非抽样误差	13
§ 2.6 精度与费用	13
§ 3 几种基本的抽样方法	14
§ 3.1 简单随机抽样	14
§ 3.2 分层抽样	14
§ 3.3 整群抽样	15
§ 3.4 多阶段抽样	15
§ 3.5 系统抽样	15
§ 4 抽样调查步骤	16
§ 4.1 确定调研问题	16
§ 4.2 抽样方案设计	16
§ 4.3 问卷设计	17
§ 4.4 实施调查过程	17
§ 4.5 数据处理分析	17
§ 4.6 撰写调查报告	17
习题	18
第二章 简单随机抽样	20
§ 1 引言	20
§ 1.1 定义	20

§ 1.2	样本性质	20
§ 1.3	符号	21
§ 1.4	抽样方法	23
§ 2	估计量及性质	23
§ 2.1	总体目标量的估计	23
§ 2.2	估计量的性质	24
§ 2.3	放回简单随机抽样下的估计	29
§ 3	样本量的确定	31
§ 3.1	费用限制	31
§ 3.2	精度要求	31
§ 3.3	影响样本量的其他因素	34
§ 3.4	总体参数的预估计	34
§ 4	子总体的估计	35
§ 4.1	相关定义及符号	35
§ 4.2	子总体均值的估计	36
§ 4.3	子总体总量的估计	37
习题		38
第三章	分层抽样	40
§ 1	引言	40
§ 1.1	定义	40
§ 1.2	作用	40
§ 1.3	应用场合	41
§ 1.4	符号说明	42
§ 1.5	实施方法	42
§ 2	估计量及性质	43
§ 2.1	总体均值估计	43
§ 2.2	总体总量估计	44
§ 2.3	总体比例估计	46
§ 2.4	总体特征单元数估计	47
§ 3	样本量在各层的分配	49
§ 3.1	样本分配对精度与费用的影响	49
§ 3.2	比例分配	50
§ 3.3	最优分配	51
§ 3.4	奈曼分配	51
§ 3.5	某些层大于 100% 抽样时的情况	52
§ 4	总样本量的确定	54
§ 4.1	一般公式	54
§ 4.2	不同应用场合	54

§ 5 分层抽样的效率分析	56
§ 5.1 分层抽样与简单随机抽样的比较	57
§ 5.2 分层抽样的设计效应	59
§ 5.3 比例分配与奈曼分配的效率比较	61
§ 5.4 偏离最优分配时对方差的影响	61
§ 6 划分层的问题	63
§ 6.1 层的界限	63
§ 6.2 层数的确定	64
§ 7 其他分层技术	65
§ 7.1 目录抽样	65
§ 7.2 事后分层	66
§ 7.3 关于多指标分层抽样	68
习题	70
第四章 比率估计和回归估计	73
§ 1 引言	73
§ 1.1 定义	73
§ 1.2 应用条件	74
§ 1.3 符号	74
§ 2 简单随机抽样下的比率估计	75
§ 2.1 比率估计量及其性质	75
§ 2.2 样本量的确定	79
§ 2.3 比率估计量的偏倚及纠偏方法	81
§ 3 分层抽样下的比率估计	84
§ 3.1 分别比率估计	84
§ 3.2 联合比率估计	84
§ 3.3 分别比率估计与联合比率估计的比较	85
§ 4 简单随机抽样下的回归估计	87
§ 4.1 回归估计量及其性质	87
§ 4.2 差值估计——回归估计量的一种特殊情形	91
§ 5 分层抽样下的回归估计	92
§ 5.1 分别回归估计	92
§ 5.2 联合回归估计	93
§ 5.3 分别回归估计与联合回归估计的比较	94
§ 6 比率估计、回归估计的效率问题	95
§ 6.1 比率估计与简单估计的比较	96
§ 6.2 回归估计与简单估计的比较	97
§ 6.3 比率估计与回归估计的比较	97
本章附录	97

习题	100
第五章 不等概抽样	105
§1 引言	105
§1.1 定义	105
§1.2 种类	106
§1.3 区域抽样	107
§2 放回不等概抽样	107
§2.1 PPS 抽样的性质	107
§2.2 PPS 抽样的实施	108
§2.3 汉森-赫维茨估计量	110
§3 不放回不等概抽样	112
§3.1 π PS 抽样的性质	112
§3.2 霍维茨-汤普森估计量	113
§3.3 几种严格的 π PS 抽样	115
§3.4 几种非严格的 π PS 抽样	119
习题	120
第六章 整群抽样	123
§1 引言	123
§1.1 定义	123
§1.2 应用特点	124
§1.3 群的划分	125
§1.4 符号	126
§2 群规模相等的情形	127
§2.1 估计量及性质	127
§2.2 整群抽样效率分析	131
§2.3 最佳群规模的确定	133
§3 群规模不相等的情形	135
§3.1 等概抽样:加权估计	135
§3.2 等概抽样:比率估计	137
§3.3 不等概抽样	141
§3.4 群规模不等时三种估计方法的比较	143
本章附录	145
习题	147
第七章 系统抽样	153
§1 引言	153
§1.1 定义	153
§1.2 应用特点	154
§1.3 不同系统抽样方法	155

§ 2 等概率系统抽样	157
§ 2.1 系统抽样与整群抽样、分层抽样的关系	157
§ 2.2 估计量及性质	158
§ 3 不同总体排列方式的方差估计	163
§ 3.1 按无关标志排列的总体	163
§ 3.2 按有关标志排列的总体	164
§ 3.3 有周期波动的总体	169
§ 4 其他类型的系统抽样	170
§ 4.1 不等概率系统抽样	170
§ 4.2 分层系统抽样	173
§ 4.3 二维系统抽样	175
本章附录	176
习题	178
第八章 多阶段抽样	181
§ 1 引言	181
§ 1.1 定义	181
§ 1.2 应用特点	182
§ 1.3 推断原理	183
§ 1.4 符号	183
§ 2 二阶抽样初级单元规模相等的情形	184
§ 2.1 总体均值与总体总量的估计	184
§ 2.2 总体比例的估计	187
§ 2.3 样本的最优分配	188
§ 3 二阶抽样初级单元规模不等的情形	190
§ 3.1 按简单随机抽样抽取初级单元	191
§ 3.2 按不等概率抽样抽取初级单元	193
§ 3.3 总体比例的估计	196
§ 4 二阶抽样的扩展	199
§ 4.1 分层二阶抽样	199
§ 4.2 三阶及多阶抽样	200
§ 5 二阶抽样效率分析	203
§ 5.1 与简单随机抽样比较	203
§ 5.2 与分层抽样比较	204
§ 5.3 与整群抽样比较	204
本章附录	205
习题	208
第九章 二重抽样	213
§ 1 引言	213

§ 1.1	定义	213
§ 1.2	二重抽样与两阶段抽样	214
§ 1.3	二重抽样的作用	214
§ 2	为分层的二重抽样	215
§ 2.1	抽样方法	215
§ 2.2	估计量及其性质	216
§ 3	为比率估计的二重抽样	218
§ 3.1	抽样方法	218
§ 3.2	估计量及其性质	219
§ 4	为回归估计的二重抽样	222
§ 4.1	抽样方法	222
§ 4.2	估计量及其性质	222
§ 5	样本量的最优分配	224
§ 5.1	为分层的二重抽样最优分配	224
§ 5.2	为比率估计的二重抽样最优分配	225
§ 5.3	为回归估计的二重抽样最优分配	225
本章附录		226
习题		227
第十章	其他抽样方法	231
§ 1	总体单元数的估计	231
§ 1.1	捕获再捕获法	231
§ 1.2	逆抽样法	233
§ 2	敏感性问题的抽样方法	234
§ 2.1	属性特征敏感性问题的抽样设计	234
§ 2.2	数量特征敏感性问题的抽样设计	243
§ 3	稀有事件抽样	245
§ 3.1	按不等比例分配的分层抽样	245
§ 3.2	二重抽样处理	246
§ 4	自加权抽样设计	247
§ 4.1	概述	247
§ 4.2	主要抽样方法下的自加权设计	247
§ 5	样本轮换设计	248
§ 5.1	概述	248
§ 5.2	样本轮换下的总体均值的估计	249
本章附录		252
习题		254
第十一章	复杂样本的方差估计	256
§ 1	引言	256

§ 1.1	什么是复杂样本	256
§ 1.2	基本方法概述	257
§ 2	直接推导法	257
§ 2.1	精确公式	257
§ 2.2	近似公式	258
§ 3	随机组方法	259
§ 3.1	独立随机组情形	259
§ 3.2	非独立随机组情形	261
§ 3.3	随机组的划分和组数的确定	262
§ 4	平衡半样本方法	264
§ 4.1	方法简介	264
§ 4.2	用于多阶段抽样	268
§ 4.3	用于非线性估计	269
§ 5	刀切法	272
§ 5.1	方法简介	272
§ 5.2	有限总体的刀切法方差估计	273
§ 5.3	弃 d 刀切法方差估计	278
§ 6	自助法	281
§ 7	比较与总结	283
§ 7.1	直接推导法	283
§ 7.2	随机组方法	283
§ 7.3	平衡半样本方法	284
§ 7.4	刀切法	284
§ 7.5	自助法	284
	本章附录	284
	习题	286
第十二章 复杂抽样设计下的统计分析		289
§ 1	引言	289
§ 2	关于类别数据的分析	289
§ 2.1	卡方检验	289
§ 2.2	卡方检验的调查设计效应	293
§ 2.3	卡方检验的校正	296
§ 3	复杂样本数据的回归分析	301
§ 3.1	简单随机抽样下的回归分析	301
§ 3.2	复杂样本数据的回归分析	303
§ 4	方法的选择	312
§ 4.1	基于设计还是基于模型	312
§ 4.2	逆抽样设计算法	314

习题	315
第十三章 调查中的缺失数据	317
§1 引言	317
§1.1 缺失数据产生的原因	317
§1.2 无回答率的计算	318
§1.3 缺失数据的统计影响	320
§1.4 增大样本量有效吗	321
§1.5 提高回答率措施	322
§2 模式和机制	324
§2.1 缺失数据模式	324
§2.2 缺失数据机制	325
§3 加权调整法	329
§3.1 加权调整法的基本思想	329
§3.2 Politz-Simmons 调整法	331
§3.3 加权组调整法	332
§3.4 事后分层调整法	334
§3.5 迭代调整法	335
§4 插补法	338
§4.1 插补法的基本思想	338
§4.2 均值插补	338
§4.3 演绎插补	340
§4.4 比率插补	341
§4.5 回归插补	341
§4.6 最近距离插补	342
§4.7 热卡插补	343
§4.8 冷卡插补	346
本章附录	347
习题	348
习题参考答案	350
参考文献	366

第一章 绪 论

本书主要讲授抽样理论,这些理论为实践中的不同抽样方法奠定了良好的基础。而本章是全书的第一章,客观上起着导言的作用。在本章中我们首先需要对有关调查的一些知识性问题做简单的讨论,这些讨论集中在抽样调查方面,如抽样调查的类型,抽样调查与普查的关系等。接下来对抽样调查中的一些重要概念做稍微深入的分析,这些概念包括目标总体与抽样总体,抽样框与抽样单元,总体特征与估计量,方差、偏倚、均方误差,抽样误差与非抽样误差,精度与费用等。本章还对一些常用的抽样方法进行了简单介绍,目的是使读者在接触抽样理论之前,对其应用背景有大致地了解。本章最后介绍了抽样调查的基本步骤。

§1 调查与抽样调查

§1.1 调查

调查是获取信息的方式,信息可以有多种,信息的类型或对信息的兴趣点不同,对调查的定义就可能不同。本书关注的是数据信息,所以这里把调查定义为:通过使用明确的概念、方法和程序,依据专门设计的调查方案指导的方式,从一个总体全部或部分单元中搜集感兴趣的指标信息,并将这些信息综合整理成数据系列的有关活动。

上面的定义表述了调查的两个重要特征:首先,调查是一种有策划、有方法、有程序的活动;其次,调查的结果表现为搜集到的数据。这里还是需要指出,调查并不是获取数据信息的唯一方式。

数据有两种类型,实验数据和调查数据。搜集数据的一种方法是通过实验,这是指在实验方案的指导下,通过控制一个或多个变量,在有控制的条件下得到观测的结果。所以实验数据是在实验中控制实验对象,依据实验设计的指导完成的。在可控条件下得到的数据是实验数据的重要特征。例如,检验一个小麦的新品种是否具有更高的产量,实验人员将不同的小麦品种按某种设计方式分种在不同的地块中,控制某情景的所有相关方面,操纵少数感兴趣的变量,然后观察实验结果。其他如检验化肥的增产效果、防治病虫害最有效的杀虫剂、某种化学反应合适的温度等。这种类型的数据往往和实验的条件有关,改变进行实验的控

制因素,结果就会发生变化.实验的次数可以是无限的.

调查数据一般是指客观上已经存在,但需要通过观察或询问才能得到的数据.例如社会现象规模、水平、相互关系和发展变化的资料.具体说它有如下特点:首先,这类资料大多与时间有关,数据所展示的是特定时期内或时点上的结果,如一定时期内的生产量,一定时点上的人口数等.其次,这类资料会随着时间的变化而改变,因此定期的收集就非常重要,因为每次收集的结果不仅展示了研究对象目前的状态,而且把以往收集的资料汇集在一起,构成时间数列,可以据此分析事物之间的相互影响和发展变化,这就为信息的进一步开发提供了广阔的空间.最后,也会有一些数据,它们在短期内相对稳定、变化不大,最常见的就是一个国家或地区的地理和地质资料,如地形、气候条件、土壤类型、矿物储量等.这类数据的调查往往技术性强,需要相关专业人员使用专门的设备进行.这类调查的成本较高,但一旦取得这方面的资料就相对稳定,不需要经常更新.与实验数据相比,调查数据有一个显著特点,即一般情况下存在一定的误差,获得高质量的调查数据十分困难,这是因为调查对象通常是以人为主体的,虽然有些调查对象是客观事物,如土地资源,但也与人的活动有关.这种情况下利益关系就会卷入其中,调查结果就更富于变化.所以调查是一门与人打交道的艺术,好的调查会将人为干扰的因素降到最低限度,这不仅仅是一个调查的纯技术性问题,而且需要根据具体的调查内容、调查对象、调查方式和调查手段,发挥艺术的创造力.一个接近完美的调查一定是科学和艺术的结合.而且,在调查中获取数据的方式基本是询问,调查问卷是询问的依据,也是信息的载体.在有些调查中,调查者与被调查者面对面交流,如入户调查;在另一些调查中,调查者与被调查者不是面对面的,如电话调查、网络调查、留置问卷等;还有一些制度性的调查,要求被调查方填报.无论哪种调查,问卷(报表)都是必须的,所以问卷设计的质量就会对调查结果产生影响.本书是从统计学的角度讨论抽样理论,问卷设计并不是抽样理论中的内容,所以不在本书讨论范围之内,但需要指出,问卷的设计十分重要,好的问卷是搜集高质量数据的基础,问卷设计如何也是调查是否具有艺术性的一个标志.

§ 1.2 抽样调查

抽样调查是调查应用最常见的模式,是一种非全面的调查,它是指从研究对象的全体(总体)中抽取一部分单元作为样本,根据对所抽取的样本进行调查,获得有关总体目标量的了解.这是广义的抽样调查的概念.

从总体抽取样本的方法看,可以分为两类抽样:一类是非概率抽样,一类是概率抽样.

1. 非概率抽样

非概率抽样并没有严格的定义,其最主要的特征是抽取样本时并不是依据随机原则.这类抽样有许多不同的具体抽取样本的方法.下面列举一些.

(1) 判断选择.判断选择是指在抽取样本时,调查人员依据调查目的和对调查对象情况的了解,人为确定样本单元.实践中确定样本单元通常有几种情况,一种是选择“平均型”单元作为样本,选定的样本可以代表感兴趣的变量的平均水平,目的是了解总体平均水平的大体位置.另一种是“众数型”,即在调查总体中选择能够代表大多数单元情况的个体为样本.再一种是“特殊型”,如选择很好或很差的典型单元为样本,目的是“解剖麻雀”,分析研究造成这种异常的原因.

(2) 方便抽样.方便抽样是指在抽取样本时依据方便原则,以达到最大限度降低调查成本的目的.典型的形式是“拦截式”调查,如在街边或居民小区拦住行人实施调查,在商场门口或柜台向消费者进行调查等.方便抽样操作简单,能及时取得所需要的信息,节省调查费用.该方法的主要局限是样本信息无法说明总体状况,我们不能根据样本信息对总体进行数量特征的推断,因而样本不适合描述性研究和因果关系研究.它比较适合探索性研究,通过调查发现问题,产生想法和假设.它也可以用于正式调查前的预调查.

(3) 自愿样本.自愿样本不是经过抽取,而是由自愿接受调查的单元所组成的样本.比较典型的是网上调查,调查人员将调查问卷贴在相关内容的网页上,感兴趣的读者可以自行填答.将调查问卷刊登在报刊上,读者可以自愿参加等.自愿样本的特点是样本的结构具有独特性,接受调查的样本单元往往属于某些特定的群体,例如,他们必须有接触到调查问卷的条件,同时对这项调查感兴趣,愿意积极参与.参与调查的群体与没有参与的群体可能会存在很多差异,调查结果无法推断总体.但这种调查操作简单、方便,成本低廉,而且参与者对调查内容感兴趣,调查人员能够了解这个特定群体的观点和看法,这对于了解情况、分析问题也是必要的,特别是如果调查目标就是了解这个特定群体的信息,自愿样本是一个不错的选择.

(4) 配额抽样.配额抽样是将总体中的各单元按一定标准划分为若干类型,将样本数额分配到各类型中,从各类型中抽取样本的方法则没有严格限制,一般采用方便抽样的方法抽取样本单元.

在配额抽样中,可以按单一变量控制,也可以按多变量控制.表 1.1 是单一变量控制的例子.在一个城市中欲采用配额方法抽出样本量 $n=500$ 的样本,控制变量有年龄和性别,配额按单变量分配,如各个年龄段的配额和不同性别的配额.单变量配额简单,但可能出现偏斜,如年龄低的为女性多,年龄高的为男性多.表 1.2 是多变量控制的例子.