

科学 专著丛书

抽样调查

——理论、方法与实践

SAMPLING SURVEYS
——THEORY, METHODS
AND CASE ANALYSIS

冯士雍 施锡铨 著

上海科学技术出版社

C811

F-767

科学专著丛书

抽 样 调 查

—理论、方法与实践

冯士雍 施锡铨 著

上海科学技术出版社

责任编辑 赵序明

科学专著丛书

抽样调查

——理论、方法与实践

冯士雍 施锡铨 著

上海科学技术出版社出版、发行

(上海瑞金二路 450 号)

新华书店上海发行所经销 上海商务印刷厂印刷

开本 787×1092 1/16 印张 28 插页 4 字数 442,000

1996年8月第1版 1996年8月第1次印刷

印数 1—1,200

ISBN 7-5323-3777-4/O·188

定价: 49.80 元

《科学专著丛书》序

如果说科学研究论文是创造性科学工作的发表性记录,那末科学技术学术专著则是创造性科学工作的总结性记录。前者注重的是优先权,后者注重的是系统化。

在大量科学研究的基础上,对一个专题或一个领域的研究成果,作系统的整理总结,著书立说,乃是科学研究工作不可少的一个组成部分。著书立说,既是丰富人类知识宝库的需要,也是探索未知领域、开拓人类知识新疆界的需要。特别是在科学各门类的那些基本问题上,一部优秀的学术专著常常成为本学科或相关学科取得突破性进展的基石。所以,科学技术学术专著的著述和出版是一项十分重要的工作。

近 20 年来,中国的科学事业有了迅速的发展,涌现了许多优秀的科学研究成果,为出版学术专著提供了坚实的基础。值此 20 世纪 90 年代,在出版学术专著方面,中国的科学界和出版界都在抓紧为本世纪再加些积累,为迎接新世纪多作些开拓。我高兴地看到,作为这种努力的一个部分,《科学》杂志的出版者——上海科学技术出版社推出了这套《科学专著丛书》。

上海科学技术出版社是科学技术界熟悉和信赖的一家出版社,历来注重科学技术学术专著的出版。《科学》杂志的编者组织编辑学术系列丛书,也不是第一次。在本世纪三四十年代,就曾推出过《科学丛书》,其中不乏佳作,对当时的学术研究起了很好的作用。

《科学》在中国是一份历史最长的综合性科学刊物,80 年来与科学技术界建立了广泛的密切联系。现在推出的这套《科学专著丛书》正是这种

联系的产物。我相信,加强这种联系,著者与编者、出版者,科技界与出版界共同努力,精心选题,精心编辑,精心出版,一定能使这套专著丛书反映出中国科学技术研究的最新水平,为本世纪多留下几本中国学者的优秀专著,为迈向新世纪多铺下几块引路的基石!

周光召

(《科学》杂志编委会主编)

1994年8月

本书序

1991年夏初，冯士雍教授应中国科学技术大学数学系之聘，为该系统计专门化学生讲授《抽样调查》课程。我当时适因事去合肥，与冯教授朝夕相见，因有幸拜读了他的讲稿，即此书的原胚，披阅之下，深感此稿取材精当，论述严谨而可读性强，尤其难得的是，其中包含了不少实例分析，实是一部极有出版价值的著作。后以此意与冯教授商议，知他也有这个打算，很是高兴。

光阴似箭，很快又过去了三个寒暑，在这期间，曾多次向他打听此书写作的进展情况。承告以此书系我国第一部抽样调查方面的大型著作，故在取材、编排和论述方面务求精当。因此颇费周折，加以科研事务繁重，故对进度有所影响云云。所幸经过几年的努力，这件精雕细刻的工作终于得以呈献于读者之前。作为本书的第一个读者，在感到欣慰之余，也深为作者这种勤勉刻苦、锲而不舍和精益求精的精神所感动。

关于本书的内容与特点，作者在前言中已有充分的介绍。此外想特别提请读者注意的是书中的“案例分析”部分，在其中所汇集的一些大型的抽样调查应用实例，许多是由冯教授主持或参与的，有的并曾在全国性的媒体上报导过，冯教授详细介绍了这些项目的调查目的、方法、指标的选择、调查的组织实施、抽样方案的制定、数据处理方法以及所得结果的解释和应用等的全过程。在某种意义上，这部分可视为冯教授从事抽样调查研究和应用工作十余年的经验的结晶。现在他把这些总结在本书里以与读者共享，实在是弥足珍贵。不列颠百科全书把统计学定义为“收集和分析数据的艺术”。这“艺术”一词值得玩味。而冯教授所提供的这些案例分析，对在抽样调查的领域内这种“艺术”如何展现，提供了感性的范例。其启示的意义，实在是超出这些例子本身之外的。

数理统计学的快速发展，也使抽样调查这门相对说来较为经典的分支学科的面貌有了不少更新。这些在本书中不少地方有所反映。在此还特别要提到由施锡铨教授主持撰写的9、10两章。施教授是我国知名的

中年统计学家，其在 Jackknife 和 Bootstrap 方面的研究在国内堪称独步。由他来承担这类题目的主笔，可说是适人适选。

学术著作出版难，是当前困扰学界同人的一件憾事，而本书这样一部有很大学术和社会意义的著作，在上海科学技术出版社的大力支持下，得以顺利且迅速地问世，其扶持学术的远大眼光功不可没，作为本书读者及学界一员，愿借此机会表示个人的赞赏和钦佩。

陈希孺 1994年5月

前 言

抽样调查历来是应用统计的一个重要分支,在各个领域,特别是社会经济领域中有极其重要的应用。但直到十多年前,在我国应用面还很窄,而在学术界则可说是几乎一片空白。随着我国改革开放的不断深入及社会主义市场经济的初步形成,抽样调查在调查方法中将逐渐占据主导地位,随着它愈来愈广泛的应用,对方法及理论的需求也愈加迫切。但目前国内尚无一本理论、方法与实践这三个方面兼顾的抽样调查的学术著作。本书是为填补这一空缺所作的一个尝试。

由于工作的需要,十余年前,作者开始把注意力转向抽样调查这个领域,除进行理论研究,参加实际项目外,还自1985起先后在中国科技大学、华东师范大学、上海财经大学、上海交通大学、中国科学院研究生院等单位为本科生及硕士研究生开设了《抽样调查》课程。最初所用的讲义在框架上参考了William G. Cochran的*Sampling Techniques*。在本书写作过程中,参照本学科近年发展状况及实际需要,并结合自己一些研究心得,对本书的章节编排及叙述论证作了设计,特别是增加了许多我国的数字实例,其中不少是作者亲自参与的。考虑到新近发展起来的再抽样(resampling)等方法在复杂样本方差估计等方面的理论和实用意义,由在这个领域里富有研究成果的施锡铨教授撰写了第9、10两章,这两章使全书增色不少。

全书共分十一章。第1章介绍抽样调查的意义与作用、若干基本概念及其应用领域,从第2章至第8章详尽地介绍了几种应用中最重要抽样方法:简单随机抽样、分层抽样、不等概率抽样、整群抽样、二阶及多阶抽样、系统抽样。其中第4章介绍两种重要的非线性估计——比估计及回归估计。上述各章的重点是介绍各种抽样方法的适用场合与实施方法、样本量的确定及总体目标量的估计及其方差估计。第9章介绍复杂样本的方差估计方法。第10章讨论了抽样调查的误差来源,特别是非抽样误差及其相关问题的处理方法。最后一章是案例分析,介绍了多项实

际抽样调查项目的背景、目的、设计与分析,并对其进行分析评价。这一章的材料是经过精心选择的,为的是尽可能照顾到不同的应用领域和所用的方法。大部分案例采用作者及其同事们多年来承担的实际项目,另外一些则取自我国开展的其他有影响的全国性抽样调查方案,希望通过这些案例,使读者就一个实际抽样调查项目的目的、设计与分析的全过程有一个概要的了解。

本书主要读者对象是从事抽样调查的理论与方法研究的科研工作者、教师以及实际工作者。为尽可能照顾到多方面读者的需要,本书写作时尽可能做到简繁相间,难易结合。对于那些主要从事实际工作的读者,在初次阅读时可略去若干定理的证明,本书也适合作为研究生的教材、数理统计专业本科生学习抽样调查课程的主要参考书。

本书成稿过程中得到了作者的老师和许多同行的多方鼓励和支持,其中有陈希孺、成平、项可风、张尧庭、汪仁官、孙山泽、茆诗松和梁小筠诸位教授。陈希孺教授在百忙中为本书专门作序,汪仁官教授认真地审阅了全书并提出不少中肯而宝贵的意见。对此作者谨表示深切地感谢。作者还要感谢我的学生邹国华,他十分仔细地阅读了原稿,并对原稿进行了若干校正。最后作者还要特别感谢上海科学技术出版社为编辑出版这本书所作的努力。由于作者水平所限,书中一定存在不足之处,望请有关专家及广大读者惠予批评指正。

冯士雍

1994年4月于中国科学院

系统科学研究所

内 容 提 要

本书所反映的研究工作系国家自然科学基金重点资助项目之一。

全书共分 11 章。第 1 章介绍抽样调查的意义和作用、若干基本概念及其应用范围。从第 2 章至第 8 章详尽地介绍了几种应用中最常用的抽样方法: 简单随机抽样、系统抽样、分层抽样、不等概率抽样、整群抽样、二阶及多阶抽样, 其中第 4 章介绍了两种重要的非线性估计——比估计及回归估计。上述各章的重点在于阐述这些抽样方法的适用场合、实施方法、样本量的确定与总体目标量的估计及方差估计。第 9 章介绍几种复杂样本的方差估计方法, 第 10 章讨论了抽样调查的误差来源, 特别是非抽样误差的处理方法。最后一章案例分析介绍多项实际抽样调查项目的背景、目的、设计与分析, 并对其进行分析评价。

本书主要读者对象是从事抽样调查理论与方法研究的科研人员及实际工作者。也可供数理统计及经济统计专业学生或相关专业研究生作教材或主要参考书。

目 录

《科学专著丛书》序

本书序

前言

第1章 引论	1
§1.1 抽样调查的意义和作用	1
§1.2 若干基本概念	3
1.2.1 总体与样本	3
1.2.2 概率抽样	4
1.2.3 抽样单元与抽样框	5
1.2.4 总体参数的分类	6
1.2.5 误差来源与精度表示	7
1.2.6 样本量、费用与效率	9
§1.3 几种基本的抽样方法	10
1.3.1 简单随机抽样	10
1.3.2 分层抽样	10
1.3.3 整群抽样	11
1.3.4 多阶抽样	11
1.3.5 系统抽样	11
§1.4 抽样调查的步骤	12
§1.5 抽样调查的应用与历史发展	14
1.5.1 主要应用领域	14
1.5.2 国际发展简史	15
1.5.3 我国抽样调查的应用与发展	18
第2章 简单随机抽样	22
§2.1 定义及实施方法	22
2.1.1 从一个有限总体中抽取所有可能的样本	22

2.1.2	两个等价的定义	23
2.1.3	简单随机抽样的实施方法	24
§ 2.2	估计量及其性质	25
2.2.1	简单估计及其无偏性	25
2.2.2	估计量的方差与协方差	28
2.2.3	方差与协方差的估计	31
2.2.4	简单估计的优良性及可以进一步改进的途径	33
§ 2.3	总体比例的估计与对子总体的估计	36
2.3.1	总体比例(百分率)的估计	36
2.3.2	子总体的估计	38
§ 2.4	样本量的确定	41
2.4.1	确定 n 的一般原则	41
2.4.2	总体参数为 \bar{Y} 或 Y 的一般情形	42
2.4.3	估计总体比例 P 的情形	43
2.4.4	总体方差的预先估计	45
§ 2.5	放回简单随机抽样	46
2.5.1	抽样方法及基本特征	46
2.5.2	总体平均数 \bar{Y} 估计量 \bar{y} 的性质	47
2.5.3	设计效应 (deff)	49
2.5.4	\bar{Y} 的另一种估计量	49
§ 2.6	利用随机数骰子和随机数表进行随机抽样的方法	50
2.6.1	随机数骰子及其使用方法	50
2.6.2	随机数表的使用方法	52
第 3 章	分层抽样	54
§ 3.1	一般描述	54
3.1.1	定义与记号	54
3.1.2	分层抽样适用的场合和优点	55
§ 3.2	估计量及其性质	55
3.2.1	估计量的构造	55
3.2.2	基本性质	56
3.2.3	比例分配及自加权样本	57
3.2.4	一个简单的实验例子	59

§ 3.3 最优分配	60
3.3.1 最优分配的定义	60
3.3.2 主要结果	61
3.3.3 Neyman(最优)分配	62
3.3.4 某些层需要超过100%抽样时的修正	64
§ 3.4 分层随机抽样在精度上的得益	64
3.4.1 与简单随机抽样的比较	64
3.4.2 何时分层及最优分配的精度得益最大	65
3.4.3 分层随机抽样精度反比简单随机抽样差的情形	66
3.4.4 从样本估计分层随机抽样精度的得益	68
3.4.5 数值例子——关于职工月平均奖金额的调查	69
3.4.6 偏离最优分配时对方差的影响	72
3.4.7 多指标情形样本量的分配	73
§ 3.5 样本总量 n 的确定	75
3.5.1 估计的总体参数为 \bar{Y} 的情形	75
3.5.2 估计的总体参数为 Y 的情形	78
§ 3.6 对总体比例(百分率)的分层随机抽样	78
3.6.1 估计量及其方差	79
3.6.2 最优分配	79
3.6.3 分层和最优分配精度上的得益	79
3.6.4 样本量的估计	80
§ 3.7 分层技术的充分利用	81
3.7.1 层的构造	81
3.7.2 多重分层	85
3.7.3 每层只抽一个单元时的方差估计	87
3.7.4 事后分层	88
3.7.5 定额抽样	91
§ 3.8 用于分层的二相抽样	91
3.8.1 层权误差对分层估计的影响	91
3.8.2 二相抽样及估计量均值与方差的一般公式	93
3.8.3 用于分层的二相抽样的估计	96
第4章 比估计与回归估计	100

§ 4.1 比估计及其基本性质	100
4.1.1 定义	100
4.1.2 基本性质	101
4.1.3 方差的估计	103
4.1.4 置信限	104
4.1.5 比估计与简单估计量的比较	105
4.1.6 数值例子——小麦估产调查	105
4.1.7 乘积估计	107
§ 4.2 比估计的偏倚及其均方误差和方差估计的阶	108
4.2.1 关于有限总体样本中心矩阶的基本引理	108
4.2.2 比估计的偏倚与均方误差及其阶的估计	111
4.2.3 均方误差或方差估计的偏倚	115
§ 4.3 分层随机抽样中的比估计	117
4.3.1 分别比估计	118
4.3.2 联合比估计	118
4.3.3 分别比估计与联合比估计的比较	120
4.3.4 分层比估计时的最优分配	121
4.3.5 数值例子——耕地面积核实调查	121
§ 4.4 消除或减少比估计偏倚的方法	124
4.4.1 无偏的比类型估计量	125
4.4.2 减少比估计偏倚的方法	127
§ 4.5 回归估计量(β 设定时的情形)	129
4.5.1 回归估计量的一般形式	129
4.5.2 β 设定情形的一般结果	129
4.5.3 差估计量	131
§ 4.6 回归估计量(β 取样本回归系数的情形)	131
4.6.1 表达式及若干引理	131
4.6.2 基本性质	135
4.6.3 回归估计量与简单估计量及比估计量的比较	138
§ 4.7 分层随机抽样中的回归估计	139
4.7.1 分别回归估计	139
4.7.2 联合回归估计	140
4.7.3 数值例子——专业技术人员总数的调查	142

§ 4.8 多变量比估计与回归估计	145
4.8.1 多变量比估计	145
4.8.2 多变量回归估计	147
4.8.3 数值例子——农作物估产调查	147
§ 4.9 二相抽样中的比估计与回归估计	151
4.9.1 比估计	151
4.9.2 回归估计	152
第5章 不等概率抽样	153
§ 5.1 一般描述	153
5.1.1 不等概率抽样的必要性	153
5.1.2 不等概率抽样的分类	154
§ 5.2 放回不等概率抽样与 Hansen-Hurwitz 估计量	155
5.2.1 多项抽样、PPS 抽样及其实施方法	155
5.2.2 Hansen-Hurwitz 估计量及其性质	157
5.2.3 数值例子——职工人数的调查	159
§ 5.3 不放回不等概率抽样与 Horvitz-Thompson 估计量	161
5.3.1 不放回不等概率抽样与包含概率	161
5.3.2 Horvitz-Thompson 估计量及其性质	162
§ 5.4 几种严格的不放回 π PS 抽样方法	164
5.4.1 $n=2$ 的情形	164
5.4.2 $n>2$ 的情形	167
§ 5.5 其他不放回抽样方法及其相应的估计量	169
5.5.1 Yates-Grundy 逐个抽取法及 Das-Raj-Murthy 估计量	169
5.5.2 Rao-Hartley-Cochran 方法及其估计量	173
5.5.3 Poisson 抽样	174
5.5.4 配置抽样	176
5.5.5 不同抽样或估计方法性质的比较	177
第6章 整群抽样	179
§ 6.1 引言	179
6.1.1 定义	179

6.1.2	适用场合及实施理由	179
6.1.3	群划分的原则	180
§ 6.2	群大小相等的情形	181
6.2.1	记号	181
6.2.2	总体与样本平方和的分解	182
6.2.3	群内相关 ρ_c	183
6.2.4	估计量及其方差	187
6.2.5	设计效应	187
§ 6.3	对比例估计的整群抽样	188
6.3.1	群大小相等的情形	189
6.3.2	群大小不相等的情形——比估计	189
§ 6.4	群大小不等的一般情形	191
6.4.1	按简单随机抽样抽群——简单估计	191
6.4.2	按简单随机抽样抽群——比估计	192
6.4.3	对群进行不等概率抽样	194
6.4.4	数值例子——对交通运输量的调查	196
第7章	二阶与多阶抽样	199
§ 7.1	引言	199
7.1.1	定义及适用场合	199
7.1.2	实施方法及同其他抽样方法的关系	200
§ 7.2	二阶抽样——初级单元大小相等的情形	201
7.2.1	记号	202
7.2.2	估计量及其方差	203
7.2.3	最优抽样比例	205
7.2.4	数值例子——生猪存栏量的调查	206
7.2.5	关于比例的估计	208
7.2.6	分层二阶抽样	209
§ 7.3	二阶抽样——初级单元大小不等情形($n=1$)	211
7.3.1	一般说明与记号	211
7.3.2	等概率抽取初级单元	212
7.3.3	不等概率抽取初级单元	213
§ 7.4	二阶抽样——初级单元大小不等的一般情形($n>1$)	215

7.4.1	按多项抽样抽取初级单元	216
7.4.2	不放回抽样时的一般结果	220
7.4.3	按简单随机抽样抽取初级单元	223
7.4.4	按不放回不等概率抽取初级单元	224
§ 7.5	三阶及多阶抽样	227
7.5.1	各级单元大小相等时的三阶抽样	227
7.5.2	多阶抽样中不等概率抽样的应用	229
第 8 章	系统抽样	232
§ 8.1	一般描述	232
8.1.1	定义及实施方法	232
8.1.2	系统抽样与整群抽样和分层抽样的关系	233
8.1.3	系统抽样的优缺点	234
§ 8.2	等概率系统抽样(等距抽样)	235
8.2.1	估计量	235
8.2.2	估计量的方差——用样本(群)内方差表示	235
8.2.3	估计量的方差——用样本(群)内相关表示	236
8.2.4	数值例子	237
§ 8.3	方差与总体单元排列顺序的关系	239
8.3.1	随机排列	239
8.3.2	线性趋势	242
8.3.3	单元指标呈周期性变化的情形	243
8.3.4	单元指标呈自相关的情形	243
§ 8.4	具有线性趋势的总体的估计量与抽样方法的改进	245
8.4.1	首尾校正法	245
8.4.2	中位样本法	246
8.4.3	对称(平衡)系统抽样法	246
8.4.4	回归估计量的应用	247
8.4.5	数值例子——部门职工总人数的估计	248
§ 8.5	不等概率系统抽样	250
8.5.1	概述及实施方法	250
8.5.2	估计量	252
§ 8.6	系统抽样中的方差估计	252