

高等院校统计学专业规划教材

# 抽样调查理论与方法

冯士雍 倪加勋 邹国华 编著

中国统计出版社

高等院校统计学专业规划教材

# 抽样调查理论与方法

冯士雍 倪加勋 邹国华 编著

中国统计出版社

(京)新登字 041 号

**图书在版编目(CIP)数据**

抽样调查理论与方法/冯士雍等编著

—北京:中国统计出版社,1998.3

高等院校统计学专业规划教材

ISBN 7-5037-2023-9

I. 抽…

II. 冯…

III. 抽样调查—概论—高等学校—教材

IV. C811

中国版本图书馆 CIP 数据核字(97)第 20292 号

中国统计出版社出版

(北京三里河月坛南街 75 号 100826)

新华书店经销

北京市兴华印刷厂印刷

\*

850×1168 毫米 32 开本 11.75 印张 29 万字

1998 年 3 月第 1 版 1998 年 3 月北京第 1 次印刷

印数:1—10100 册

\*

定价:11.20 元

(版权所有 不得翻印)

# 出版说明

随着社会主义市场经济体系的逐步建立,我国的统计观念、统计体制、统计标准、统计制度和统计调查方法都在发生着重要的变化。为了适应建立社会主义市场经济体制和统计现代化建设的需要,全国统计教材编审委员会制定了《1991—1995年全国统计教材建设规划》,并根据《规划》的要求研究制定了统计学专业主要专业课程的教学大纲。在此基础上,编委会采取邀请、委托等方式组织全国有关院校的专家、学者编写了这批统计学“规划教材”。

这批教材力求做到紧密联系社会主义市场经济和统计现代化建设的实际,充分反映当前我国统计改革的新发展,积极吸收国内外统计科学研究的新成果。相信通过这批教材的出版,对我国的统计教学改革将起到积极的推动作用,对我国的统计教材建设也将起到较好的示范、导向作用。

限于水平和经验,这批教材的编审、出版工作,还会有缺点和不足之处,诚恳欢迎教材的使用单位、广大教师和同学们提出批评和建议。

全国统计教材编审委员会

1994年11月

# 前 言

抽样调查是统计学的一门重要分支学科。随着我国经济体制从计划经济体制向社会主义市场经济体制的转变,抽样调查的作用显得日益重要。我国统计调查方法改革目标模式已明确我国统计调查“以经常性抽样调查为主体”,并列入修改后的《统计法》。在国际上,抽样调查早已成为世界各国普遍采用的一种统计调查方法。它的理论和方法研究发展很快。但是在我国统计学专业中,长期以来还缺少一本适合社会发展需要,建立在严格科学基础上的抽样调查教材。本书是根据1993年6月全国统计教材编审委员会讨论通过的《抽样调查理论与方法》教学大纲编写,作为高等院校统计学专业规划教材之一出版的。

根据大纲要求,本课程的教学目的是使学生能系统正确地掌握抽样调查的理论基础,掌握几种基本的抽样方法,培养学生初步具有能结合实际情况对具体项目进行抽样设计和对所获得数据进行处理和分析的能力。为此,在教学中要特别讲清各种概念和方法的统计思想,使学生牢固地掌握基本抽样技术的实施和分析方法,并通过基本训练有进一步学习文献、扩大知识面的能力。

本书共分十二章。第一章为概论,简述抽样调查的意义与作用,它的发展简史与主要应用。第二章是有关抽样的基本概念,为以后各章的学习打好基础。第三章至第十章是本课程的主体,逐章介绍简单随机抽样、分层随机抽样、比估计与回归估计、二重抽样、不等概率抽样、整群抽样、二阶与多阶抽样及系统抽样。详细介绍这些具体抽样方法的适用场合与实施方法,对总体目标量的估计,估计量的方差与方差估计。第十一章介绍非抽样误差。这些都是

大纲规定的内容(某些超出大纲要求的节或小节以“\*”号表示)。原大纲的第十二章是“案例分析”,但考虑到案例分析要占大量篇幅,在实际教学中也应根据教师与学生的不同需要而定,不宜作硬性统一规定,因此全国统计教材编审委员会考虑这方面内容以后另行编写出版。由于复杂抽样方差估计在现代抽样理论与实际应用的重要性,我们在最后一章中增加了这方面的内容。但这并非大纲所要求的,只作为补充教材使用。

本书的第二、三、四、五、七、八、九诸章由中国科学院系统科学所冯士雍研究员执笔并任主编;第一、六、十、十一诸章由中国人民大学统计学系倪加勋教授提出初稿并经冯士雍研究员增删与改写;第十二章则由中国科学院系统科学所邹国华博士执笔。全书习题主要由倪加勋教授提供并经邹国华博士补充。书中的大多数例子都是经编著者精选的有实际背景的数值例子,个别例子(包括部分习题)选自所列参考书目,未一一注明,特向有关作者和出版社表示感谢。需要说明的是,由于篇幅所限有相当部分的数值例子的样本量都较小,这对实际应用是不够的。这些例子只能作为说明性的示例,请读者务必不要照搬。

北京大学概率统计系孙山泽教授对全书的手稿进行了仔细的审阅并提出许多宝贵意见,为本书增色不少。国家统计局统计教育中心与中国统计出版社的有关同志对本书的编写与出版给予了大力支持、鼓励与帮助,在此一并表示深深的感谢。

本书在编写过程中尽管对内容与结构经过反复的斟酌与多次修改,曾几易其稿,但由于我们学术水平与教学经验所限,书中错误及选材结构安排不当之处在所难免,恳请各方面专家及使用本教材的老师与同学们不吝批评指正,以便在再版时修改与完善。

编著者

1997年8月

# 目 录

<b>第一章 概论</b> .....	(1)
1.1 抽样调查的意义与作用 .....	(1)
1.1.1 什么是抽样调查 .....	(1)
1.1.2 概率抽样与非概率抽样 .....	(1)
1.1.3 抽样调查的作用 .....	(3)
1.2 抽样调查的发展简史 .....	(6)
1.2.1 国际抽样调查发展简史 .....	(6)
1.2.2 我国抽样调查的发展 .....	(11)
1.3 抽样调查的主要应用 .....	(12)
思考题与习题 .....	(14)
<b>第二章 基本概念</b> .....	(16)
2.1 总体与样本 .....	(16)
2.1.1 总体的概念 .....	(16)
2.1.2 抽样单元与抽样框 .....	(16)
2.1.3 总体指标 .....	(18)
2.1.4 样本的概念 .....	(20)
2.1.5 固定总体模型与超总体模型 .....	(21)
2.2 几种基本的抽样方法 .....	(21)
2.2.1 简单随机抽样 .....	(22)
2.2.2 分层抽样 .....	(22)
2.2.3 整群抽样 .....	(23)
2.2.4 二阶与多阶抽样 .....	(23)
2.2.5 系统抽样 .....	(24)

2.2.6	不等概率抽样 .....	(24)
2.3	误差与精度的表示方法 .....	(25)
2.3.1	抽样调查中的误差来源 .....	(25)
2.3.2	均方误差、方差与偏倚 .....	(27)
2.3.3	误差限与置信度 .....	(28)
2.3.4	精度与费用 .....	(30)
	思考题与习题 .....	(31)
<b>第三章</b>	<b>简单随机抽样</b> .....	<b>(32)</b>
3.1	概述 .....	(32)
3.1.1	什么是简单随机抽样 .....	(32)
3.1.2	简单随机抽样的实施方法 .....	(33)
3.1.3	简单随机抽样在抽样理论中的地位与作用 .....	(37)
3.2	总体均值与总量的简单估计 .....	(38)
3.2.1	简单估计及其无偏性 .....	(38)
3.2.2	估计量的方差 .....	(41)
3.2.3	估计量的方差估计 .....	(44)
3.3	总体比例(成数)的简单估计 .....	(48)
3.3.1	对总体的描述 .....	(48)
3.3.2	估计量及其性质 .....	(49)
3.4	样本量的确定 .....	(51)
3.4.1	确定样本量的原则与主要考虑因素 .....	(51)
3.4.2	估计总体均值或总量时样本量的确定方法 .....	(52)
3.4.3	估计总体比例时样本量的确定方法 .....	(54)
3.4.4	逆抽样方法 .....	(56)
3.5	放回简单随机抽样 .....	(58)
3.5.1	估计量及其性质 .....	(58)
3.5.2	设计效应与样本量的确定 .....	(61)
3.6	子总体的估计 .....	(63)
3.6.1	问题的提出 .....	(63)



3.6.2	子总体均值的估计 .....	(63)
3.6.3	子总体总量的估计 .....	(64)
	习题 .....	(67)
<b>第四章</b>	<b>分层随机抽样 .....</b>	<b>(71)</b>
4.1	概述 .....	(71)
4.1.1	什么是分层抽样和分层随机抽样 .....	(71)
4.1.2	分层抽样的特点和适用场合 .....	(72)
4.1.3	记号 .....	(73)
4.2	简单估计量及其性质 .....	(74)
4.2.1	对总体均值或总量的估计 .....	(74)
4.2.2	对总体比例的估计 .....	(77)
4.3	各层样本量的分配 .....	(79)
4.3.1	样本量分配对精度的影响 .....	(79)
4.3.2	比例分配 .....	(80)
4.3.3	最优分配 .....	(82)
4.3.4	某些层需要超过 100% 抽样时的修正 .....	(86)
4.4	样本总量的确定 .....	(88)
4.4.1	影响样本总量的因素 .....	(88)
4.4.2	估计总体均值情形 .....	(89)
4.4.3	估计总体总量情形 .....	(92)
4.4.4	估计总体比例情形 .....	(93)
4.4.5	给定总费用时样本量的确定方法 .....	(94)
4.5	分层随机抽样效果分析 .....	(95)
4.5.1	分层随机抽样与简单随机抽样的比较 .....	(95)
4.5.2	最优分配在精度上的改进 .....	(97)
4.5.3	估计总体比例时分层和最优分配精度上的得益 .....	(98)
4.5.4	从分层样本估计分层抽样的效果 .....	(99)
4.5.5	分层标志的选择及最优分层 .....	(102)
4.5.6	层数的确定 .....	(105)

* 4.6	若干进一步问题 .....	(105)
4.6.1	偏离最优分配时对方差的影响 .....	(105)
4.6.2	多指标情形样本量的分配 .....	(107)
4.6.3	层权误差对估计量的影响 .....	(109)
4.6.4	事后分层 .....	(111)
习题	.....	(115)
<b>第五章</b>	<b>比估计与回归估计</b> .....	(120)
5.1	概述 .....	(120)
5.1.1	问题的提出 .....	(120)
5.1.2	比估计与回归估计的作用与使用条件 .....	(121)
5.2	比估计 .....	(122)
5.2.1	定义及基本性质 .....	(122)
5.2.2	方差估计及置信限 .....	(127)
5.2.3	比估计与简单估计的比较 .....	(133)
* 5.2.4	乘积估计 .....	(133)
5.3	回归估计 .....	(134)
5.3.1	定义 .....	(134)
5.3.2	$\beta$ 为设定常数情形 .....	(135)
5.3.3	$\beta$ 为样本回归系数情形 .....	(136)
5.4	分层比估计与分层回归估计 .....	(139)
5.4.1	分别比估计与联合比估计 .....	(139)
5.4.2	分别回归估计与联合回归估计 .....	(141)
5.4.3	各种估计量的比较与选择 .....	(144)
* 5.5	比估计与回归估计及其方差估计的偏倚 .....	(149)
5.5.1	有关简单随机样本中心矩阶的基本结果 .....	(149)
5.5.2	比估计及其方差估计的偏倚 .....	(150)
5.5.3	比估计偏倚的消除或减少办法 .....	(151)
5.5.4	回归估计及其方差估计的偏倚 .....	(155)
习题	.....	(156)

<b>第六章 二重抽样</b> .....	(162)
6.1 概述 .....	(162)
6.1.1 什么是二重抽样 .....	(162)
6.1.2 二重抽样的作用 .....	(162)
6.1.3 二步抽样估计量均值与方差的一般公式 .....	(164)
6.2 为分层的二重抽样 .....	(164)
6.2.1 二重分层抽样的样本抽选方法 .....	(164)
6.2.2 二重分层抽样的估计方法 .....	(165)
6.3 为比估计与回归估计的二重抽样 .....	(168)
6.3.1 二重抽样的比估计及其性质 .....	(168)
6.3.2 二重抽样的回归估计及其性质 .....	(170)
* 6.4 二重抽样样本量的最优分配 .....	(171)
6.4.1 二重分层抽样样本量的最优分配 .....	(172)
6.4.2 二重比估计与回归估计时样本量的最优分配 .....	(173)
* 6.5 连续抽样中的样本轮换及其估计 .....	(174)
6.5.1 样本轮换的必要性 .....	(174)
6.5.2 样本拼配与二重回归估计的应用 .....	(175)
习题 .....	(178)
<b>第七章 不等概率抽样</b> .....	(182)
7.1 概述 .....	(182)
7.1.1 不等概率抽样的必要性及优点 .....	(182)
7.1.2 不等概率抽样的主要分类 .....	(184)
7.2 放回不等概率抽样 .....	(185)
7.2.1 多项抽样与 <i>PPS</i> 抽样 .....	(185)
7.2.2 多项抽样的实施方法 .....	(186)
7.2.3 汉森—赫维茨估计量及其性质 .....	(187)
7.2.4 数值例—职工人数调查 .....	(190)
7.3 不放回不等概率抽样 .....	(193)
7.3.1 包含概率与 $\pi PS$ 抽样 .....	(193)

7.3.2	霍维茨—汤普森估计量及其性质	(195)
7.3.3	$n=2$ 的严格 $\pi PS$ 抽样	(196)
7.3.4	$n>2$ 的严格 $\pi PS$ 抽样	(200)
7.3.5	$n>2$ 的两种非严格 $\pi PS$ 抽样	(201)
	习题	(204)
<b>第八章</b>	<b>整群抽样</b>	(208)
8.1	概述	(208)
8.1.1	什么是整群抽样	(208)
8.1.2	整群抽样的特点及适用场合	(209)
8.2	群大小相等情形,对群进行简单随机抽样时的 估计量及其方差	(210)
8.2.1	记号	(210)
8.2.2	估计量及其性质	(211)
8.2.3	群内相关系数与设计效应	(214)
8.2.4	整群抽样效率分析及群的划分原则	(218)
8.3	估计总体比例的整群抽样	(223)
8.3.1	问题的提出	(223)
8.3.2	群大小相等情形	(224)
8.3.3	群大小不等情形	(225)
8.4	群大小不相等的一般情形	(229)
8.4.1	记号	(229)
8.4.2	按简单随机抽样抽群	(230)
8.4.3	按与群大小成比例的不等概率抽样抽群	(233)
	习题	(236)
<b>第九章</b>	<b>二阶及多阶抽样</b>	(241)
9.1	概述	(241)
9.1.1	多阶抽样的定义及与其他抽样的关系	(241)
9.1.2	多阶抽样的特点和作用	(242)
9.2	初级单元大小相等时的二阶抽样	(243)

9.2.1	记号	(243)
9.2.2	总体均值 $\bar{Y}$ 的估计量及其性质	(244)
9.2.3	关于总体比例的估计	(248)
9.2.4	最优样本量 $m$ 与 $n$ 的确定	(250)
9.3	初级单元大小不等时的二阶抽样(I)	
	——对初级单元进行多项抽样情形	(253)
9.3.1	一般说明及记号	(253)
9.3.2	总体总和 $Y$ 的估计	(254)
9.3.3	估计量是自加权的条件及对初级单元的 PPS 抽样	(255)
9.4	初级单元大小不等时的二阶抽样(II)	
	——对初级单元进行不放回抽样情形	(259)
9.4.1	用简单随机抽样抽取初级单元	(259)
9.4.2	用不放回不等概率抽取初级单元	(261)
9.5	三阶及多阶抽样	(263)
9.5.1	各级单元大小相等时的三阶抽样	(263)
9.5.2	各级单元大小不等时的多阶抽样	(265)
	习题	(267)
<b>第十章</b>	<b>系统抽样</b>	(271)
10.1	概述	(271)
10.1.1	系统抽样的定义及实施方法	(271)
10.1.2	系统抽样的特点及局限性	(273)
10.2	等概率系统抽样——等距抽样	(274)
10.2.1	估计量及其性质	(274)
10.2.2	估计量方差的不同表示形式	(278)
10.2.3	估计量的方差与总体单元排列顺序的关系	(282)
10.3	线性趋势总体抽样方法的改进	(286)
10.3.1	中心位置样本与首尾校正法	(286)
10.3.2	对称系统抽样	(288)

10.4	等概率系统抽样的方差估计	(290)
10.4.1	方差估计的形式	(290)
10.4.2	各种估计量的比较及适用场合	(292)
10.5	不等概率系统抽样	(293)
10.5.1	$\pi PS$ 系统抽样及其实施方法	(293)
10.5.2	估计量及其方差与方差估计	(295)
	习题	(296)
<b>第十一章</b>	<b>非抽样误差</b>	<b>(301)</b>
11.1	非抽样误差的主要来源与分类	(301)
11.2	抽样框误差	(303)
11.2.1	抽样框误差及其影响	(303)
11.2.2	丢失总体单元引起的抽样框误差	(304)
11.2.3	多重抽样框的使用	(306)
11.3	无回答误差	(308)
11.3.1	无回答的不同类型及产生原因	(308)
11.3.2	无回答的影响	(309)
11.3.3	降低无回答的措施与方法	(312)
11.3.4	对无回答的调整	(315)
11.4	敏感性问题调查与随机化回答技术	(319)
11.4.1	敏感性问题调查	(319)
11.4.2	沃纳随机化回答模型	(320)
11.4.3	西蒙斯随机化回答模型	(322)
11.4.4	采用随机化回答技术应注意的一些问题	(323)
11.5	计量误差	(324)
11.5.1	设计引起的误差	(324)
11.5.2	由调查员引起的误差及降低的措施	(326)
11.5.3	由被调查者引起的误差及其预防与调整	(326)
	习题	(327)
<b>* 第十二章</b>	<b>复杂样本的方差估计</b>	<b>(330)</b>

12.1	随机组方法	(330)
12.1.1	独立的随机组情形	(331)
12.1.2	非独立的随机组情形	(333)
12.2	平衡半样本方法	(336)
12.2.1	基本方法	(336)
12.2.2	用于多阶抽样	(341)
12.2.3	用于非线性估计	(342)
12.2.4	数值例子	(343)
12.3	刀切法	(346)
12.3.1	刀切法简介	(346)
12.3.2	有限总体的刀切法估计	(347)
12.3.3	自助法方差估计	(350)
12.3.4	数值例子	(351)
12.4	泰勒级数法	(353)
12.5	各种方法的比较	(356)
	<b>主要参考书目</b>	<b>(358)</b>

# 第一章 概 论

## 1.1 抽样调查的意义与作用

### 1.1.1 什么是抽样调查

统计调查是获取数据(资料)的一种重要手段,它又分为全面调查与非全面调查两类。抽样调查(survey sampling)也称样本调查(sample survey)是非全面调查中的一种重要方法,它是按一定程序从所研究对象的全体(总体)中抽取一部分(样本)进行调查或观察,获取数据,并以此对总体的一定目标量(参数)作出推断(例如估计)。

在日常生活中人们经常自觉或不自觉地在应用抽样方法,例如到市场上去买花生、瓜子,总要先抓几粒看看是否饱满、干燥;在厨房做菜,在炒菜过程中往往要取一点尝尝菜的咸淡如何。工厂在生产过程中以及商家在进货验收过程中也常抽取一定数量的产品,检验其质量并以此判定整批产品质量的优劣。在统计工作的资料收集和分析研究中也常在研究对象中抽取数量有限的个体或单元来作调查研究,其目的是掌握全面情况,这些都是抽样的例子。

### 1.1.2 概率抽样与非概率抽样

根据样本抽取方法的不同,抽样可分为概率抽样和非概率抽样两类。非概率抽样(non-probability sampling)是相对于概率抽样(probability sampling)而言的,它并无严格的定义,如我国所谓



典型调查和重点调查,在西方国家称为有目的的抽样(purposive sampling)或判断抽样(judgement sampling)等都属于非概率抽样。它们的一个共同特点是样本的抽选(selection)是根据主观判断有目的有意识地或根据方便的原则进行的,而不是按随机原则来抽选。因此这种抽样效果的好坏在很大程度上依赖于抽样者的主观判断能力和经验,且不能计算抽样误差,不能从概率意义上控制误差并以此来保证推断的准确性。

概率抽样也称随机抽样。与非概率抽样不同的是概率抽样是严格地按照给定的概率来抽取样本。具体地说,概率抽样具有以下特点:

(1)按一定概率以随机原则抽取样本。所谓随机原则就是在抽选样本时排除主观上有意识地抽选某些调查单元,使每个单元都有一定的机会被抽中。这里有几点需要加以说明:一是随机并不是“随便”,随机是有严格的科学含义的,可用概率来描述;而“随便”仍带有人为的或主观的因素,它不是一个科学的概念。其次随机原则并不等同于等概率原则,有些教科书把随机抽样简单地解释为“使总体中每一个单元都有相同的概率被抽中的抽样”,这并不很确切。严格地说,这种说明应称为随机抽取(draw an item at random)。等概率抽样确实是一种随机抽样,但随机抽样并不限定必须是等概率抽样。因此随机原则也不等同于等概率原则。最后需要指出的是随机原则一般的要求总体中每个单元均有一个非零的概率被抽中。若主观上要求某一单元一定被选入样本,则可规定其入样概率为1;但若要求某一单元一定不被选入样本,就要采用某些特殊的方法。

(2)用概率抽样抽取样本单元需要按照给定的入样概率通过一定随机化程序来实现。

(3)当用样本来估计总体目标量(参数)时,要考虑到该样本(或每个样本单元)被抽中的概率。这就是说,估计量不仅与样本观测值有关,也与其入样概率有关。