

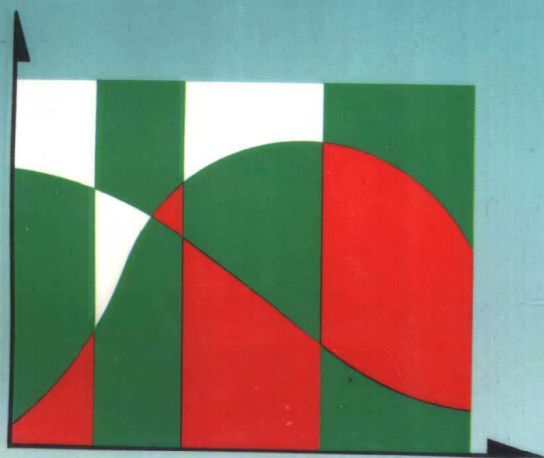
统计学专业教材系列

施锡铨 ● 主编

抽样调查的理论和方法

CHOUYANGDIAOCHADE

1992年7月第1版



上海财经大学出版社

本项目获上海财经大学资助出版

抽样调查的理论和方法

施锡铨 主编

上海财经大学出版社

责任编辑:董逸伟

封面设计:甘晓培

抽样调查的理论和方法

施锡铨 主编

上海财经大学出版社出版

(上海市中山北一路369号 邮编200083)

新华书店上海发行所发行 上海精英彩色印务有限公司印刷

开本 850×1168 1/32 印张 7.25 字数 184000

1996年7月第1版 1996年7月第1次印刷

印数 1—3000

ISBN7-81049-056-7/F·35

定价: 14.60 元

统计学专业教材系列编委会

顾问 孙祖尧 颜德伦 施锡铨 郑菊生

主任 徐国祥

委员 (按姓氏笔画为序)

卞祖武 刘汉良 苏均和 胡清友 徐国祥 董逢谷

总 序

随着我国向社会主义市场经济体制转轨,统计体制也在进行着相应的调整,使其适应市场经济发展的需要。过去主要为计划经济服务的财经类统计学教材的体系和内容也需作相应的变革。这一变革关系到是否能培养出适应社会主义市场经济要求的跨世纪的统计人才。因此,对传统的统计教材进行改革已势在必行,而且迫在眉睫。

同时,我国社会主义市场经济体制的建立也为财经类统计学专业的发展提供了机遇,为统计学专业服务于社会经济和管理提供了用武之地。统计学是研究如何有效地收集、整理和分析数据,并对所研究的问题作出预测或推断,直至为决策提供依据的一门方法论学科。在经济和管理现象日益复杂、市场信息瞬息万变的环境中,往往要在不完全观察数据的基础上,对所关心的指标作出可靠的估计,以便作出合适的决策。统计学学科正是为处理这些问题提供有效的工具和手段。因此,这门学科在我国社会主义市场经济体制下具有极其重要的意义。这也是我们组织编写这套系列教材的目的之一。

统计学学科有许多分支。对于财经类院校的统计学专业来说,不可能也没有必要把统计学的全部分支包括在内,我们强调厚基础、宽口径、重应用。我们把财经类统计学学科的培养目标定位在“经济和管理领域中善于在定性分析基础上从事定量分析的专门统计人才。”这一目标决定了财经类统计学学科的发展方向,这个

发展方向大致可分为应用数理统计和经济管理统计两个方面。为此,现阶段我们编写的统计学专业教材系列主要有以下几种,即《统计学教程》、《概率统计》、《抽样调查的理论和方法》、《多元分析》、《统计预测和决策》、《国民经济核算体系原理》、《市场经济统计学》、《企业统计学》、《经济计量学》和《计算机在统计中的应用》等。

为了适应社会主义市场经济体制的需要,使教学内容与改革开放的要求相适应,我们在组织编写这套教材系列时,强调了以下几个方面的特色:

1. 数理统计与社会经济管理统计并重。从世界范围来看,统计学的这两部分内容是客观存在的。强调两者并重,就是尊重客观实际,不以偏概全。数理统计方法通用性强,可直接借鉴国外的东西,并学以致用。而经济统计就需要考虑客观现实问题。因此,研究外国统计问题和研究中国统计问题都是非常重要的,这不但可以把握世界范围内统计发展的动向和趋势,而且也便于我们合理地借鉴外国的东西为我所用。

2. 数理统计类教材要适应财经类学生的需要。在我国,迄今为止,有关数理统计教材不少,这些教材都各具特色,第一类是综合性大学或师范大学所编的数理统计类教材,大多偏重于基础、概念和理论,它讲究逻辑性和抽象性。第二类是工程或工科类大学所编的数理统计教材,它的侧重点是讲述统计方法在工程中的应用。这两类教材不能作为财经类院校学生的教材。而真正适合于财经类院校的数理统计教材却很少。因此,我们组织编写了《统计学教程》、《概率统计》、《多元分析》、《抽样调查的理论和方法》和《统计预测和决策》等教材。在编写中,我们在阐述基本原理与方法时,强调了各种方法在经济管理领域中的应用,尤其是强调各种方法的特点、适用条件、应用场合及其必要的评价,以便使读者在实践中注意对各种方法的选择。

3. 经济管理统计类教材应围绕社会主义市场经济,以经济运行和结构为框架,形成理论和方法体系。经济管理统计是以统计方法为主的边缘性和交叉性较强的学科。它与有关专业交叉,统计方法与经济统计指标(数据)紧密结合,共同用于实证分析。在该套教材中,主要有《国民经济核算体系原理》、《市场经济统计学》、《企业统计学》和《经济计量学》等。《国民经济核算体系原理》遵循社会再生产原理,通过五大核算,描述国内(民)生产总值的来龙去脉,描述和分析国民经济的总量和结构,它为市场经济条件下进行宏观调控,建立信息系统,提供系统的信息。市场有各个组成部分,有实体和体系。它们有运行规律,需要从数量上加以反映和作定量与定性相结合的分析。企业需要面向市场,市场需要宏观调控,所以,在《企业统计学》与《国民经济核算体系原理》之间,需要建立《市场经济统计学》。《经济计量学》是经济理论、数学和统计的结合,也是一门边缘学科,它在经济管理统计学科中,主要是提供用于分析的计量模型的方法论。《计算机在统计中的应用》主要是在大量数据和复杂的统计方法情况下,如何借助计算机来处理、加工和分析数据。

4. 财经类统计学专业教材的编写应充实案例。为了将每种具体的统计方法与我国社会主义市场经济实践相结合,应在对每一种方法作必要的阐述之后,都配有具体案例说明其应用过程。强调统计方法在市场经济和管理各个领域的应用,也是财经类统计专业教材的显著特色之一。

统计学科的内容非常丰富。鉴于国内对这类教材的急需,我们上海财经大学统计学系组织编写了这套教材系列。但限于时间和水平,这套教材中难免存在缺点和错误,恳切希望广大教师和学生提出宝贵意见,使其日臻完善。

徐国祥

1995年6月

• 3 •

KAI/01/02

序 言

在所有读过的用中文编写的抽样调查书中,我最喜爱且推崇的是由北京大学出版社整理出版的许宝騄先生的《抽样论》,薄薄的一本书,讲尽了抽样理论中主要的方法、手段以及估计所需要的公式,并且给出了严格的数学证明,论述了各种方法的特点,精炼之至,精采之极!然而,这本适合于数理统计学习和研究的好书对于财经类院校的学生是很有难度的。根据财经类院校的教学大纲以及学生的数学基础,他们需要的是知道如何抽样,抽样方法的基本原理以及抽样的推断与效果,繁复的数学证明实非他们之长。也就是说,为财经院校的学生讲抽样调查应当着重于应用而不是证明。因此,多年来很想编一本既讲清道理又介绍应用的教科书。上海财经大学统计学系计划在近年编著一套适合社会主义市场经济要求的财经类统计学专业教材系列,这一计划给了我们一个机会,于是就有了目前这本《抽样调查的理论和方法》。

本书主要讲述抽样调查的各种方法,介绍了推断公式,导出了抽样误差的计算。主要特点之一是除了极个别地方之外,尽可能地用简单的盒子模型或直观理解以解释公式的来源以及它所表示的直观含义。本书的另一个特点是重视计算机在抽样推断中的作用,介绍了抽样调查误差的估算方法,尤其是计算机模拟方法。这部分内容以及非抽样误差的初步探讨得到了上海发展汽车工业教育基金会的资助,有关内容放在本书的第十、第十一章,这两章连同第九章是由范正绮同志完成的。虞定伟同志为本书的案例展示花费了大量精力,收集了可观的材料,并从中精选摘编了若干案例构成了本书的第十二章。为避免麻烦,我们只选择那些在杂志上公开发表过的案例进行摘录与介绍,对这些案例的作者表示深深的感谢。

作为教科书,我们也在几乎每章的后面附有少量习题,这些习题强调概念与思考。我们认为,要想学好抽样调查这一门课,真正的习题是结合教学实践踏踏实实地做一些抽样调查,从抽样方案的设计、回答卷的构思、逐个地上门调查以及用计算机处理数据等,从头到尾做一遍。

俗话说,说起来容易做起来难。我们的写作过程正应了这句话。要想介绍多种不同的抽样方法,将各种各样的公式推到读者面前,又想不作任何证明地使读者认为这些方法与公式是合理的,真是谈何容易!不过总得有人尝试并为此付出“代价”。本书肯定(不敢讲“难免”两字)有不妥之处,恳望读者指正!

施锡铨

1995 年 10 月

目 录

第一章 基本概念	(1)
§ 1.1 引言	(1)
§ 1.2 抽样调查的优越性	(2)
§ 1.3 抽样技巧广泛用于社会经济领域	(4)
§ 1.4 如何抽样	(6)
习题	(10)
第二章 抽样的基本步骤与数学准备	(12)
§ 2.1 调查的目的与要求	(12)
§ 2.2 抽样方案的设计与编制抽样框	(13)
§ 2.3 问卷设计	(14)
习题	(19)
§ 2.4 调查与数据处理	(19)
§ 2.5 调查总结报告	(20)
§ 2.6 若干数学准备	(21)
第三章 简单随机抽样	(26)
§ 3.1 简单随机抽样及实施方法	(26)
§ 3.2 总体平均数与总和的估计	(28)
§ 3.3 估计量的方差及其估计	(30)
§ 3.4 百分数的估计及其误差	(38)
§ 3.5 样本容量 n 的确定	(40)
习题	(43)

第四章	分层抽样	(44)
§ 4.1	分层抽样及其估计量	(44)
§ 4.2	比例分配及最优分配	(50)
§ 4.3	样本总容量 n 的确定	(55)
§ 4.4	分层的若干技术问题	(59)
	习题	(62)
第五章	比估计及回归估计	(64)
§ 5.1	比估计及其性质	(64)
§ 5.2	分层抽样中的比估计	(70)
§ 5.3	数值例	(72)
§ 5.4	回归估计量	(76)
	习题	(83)
第六章	不等概率抽样	(84)
§ 6.1	放回的不等概率抽样	(84)
§ 6.2	不放回不等概率抽样	(88)
	习题	(93)
第七章	整群抽样	(94)
§ 7.1	群大小相等的整群抽样	(95)
§ 7.2	群大小不等的整群抽样	(99)
	习题	(101)
第八章	二阶与多阶抽样	(102)
§ 8.1	初级单元大小相等的二阶抽样	(103)
§ 8.2	初级单元大小不等时的二阶抽样	(105)
§ 8.3	三阶及多阶抽样	(118)
	习题	(121)

第九章 系统抽样	(122)
§ 9.1 系统抽样的若干习性	(122)
§ 9.2 估计量与方差	(123)
§ 9.3 方差与总体单元排列顺序的关系	(127)
§ 9.4 具有线性趋势的总体的抽样方法改进	(128)
习题	(131)
第十章 复杂样本的方差估计及计算机模拟方法	(132)
§ 10.1 随机分组法.....	(133)
§ 10.2 Jackknife 方法	(137)
§ 10.3 弃 d-Jackknife 方差估计	(140)
§ 10.4 半样本方法.....	(143)
§ 10.5 调查误差的统计计算机模拟.....	(151)
习题.....	(155)
第十一章 非抽样误差	(156)
§ 11.1 抽样方案及抽样框引起的非抽样误差.....	(156)
§ 11.2 无回答现象.....	(157)
§ 11.3 计量误差.....	(161)
§ 11.4 敏感性问题的调查.....	(165)
第十二章 抽样调查的若干案例	(168)
[案例一] 我国人口变动情况抽样调查.....	(168)
[案例二] 美国盖洛普民意测验机构的抽样设计.....	(179)
[案例三] 我国城市住户抽样调查.....	(181)
[案例四] 抽样调查技术在粮油质量管理中的应用.....	(184)
[案例五] 上海市专业技术人员使用情况抽样调查.....	(190)
[案例六] 1987 年全国残疾人抽样调查	(198)

[案例七]	江苏省 1986 年的库存商品抽样调查	(199)
[案例八]	中国学生体质健康的抽样调查.....	(205)
[案例九]	橡胶种植场对橡胶树的茎围作抽样测定的实例	(210)
附表 1	随机数表	(214)
附表 2	正态分布表	(217)
参考文献		(219)

第一章 基本概念

§ 1.1 引言

统计学是一门处理数据的艺术,它研究的是如何收集数据、整理与分析数据,由此进行推断,作出进一步的决策或预测。基于此研究内容,统计学在社会经济和自然科学等各个领域内正在发挥着越来越重要的作用。本书介绍的抽样理论是统计学的重要分支之一,它正是研究收集数据和分析数据,完全体现了统计学的研究思想。利用抽样理论及其方法,我们可以从局部得到整体的概况描述。用统计学的行话:即可以由样本来推断总体。

通常我们将待研究对象的全体称为总体(population),而那些单个的对象则称为个体(individual)。抽样调查中所面临的总体常常只是有限个个体,这是它与经典数理统计最明显的区别之处。总体这个概念总是与一些数字事实或称作指标的东西联系在一起,它们正是我们通过抽样调查希望对总体有所了解的东西,这就是常说的总体的参数(parameter)。譬如,研究全国劳动力市场,有若干参数是我们关心的,其中有两个是:

- ① 失业人员在全体劳动力中所占的比例;
- ② 失业人员年龄的中位数。

这里,总体是全国所有合格的劳动力人员的全体。要知道这两个参数,就应该作一次全面调查以了解合格劳动力的总数、失业人员的总数、各失业人员的年龄等等。然而,总体是如此地庞大,鉴于人力、物力与时间等方面的限制,对于大型的总体,每事必作全面调查显然是不现实的。容易做到的事情是从总体中抽取少部分个体,它们中的相

应指标完全可以计算得到。倘若这部分从形式上(或各个方面)与总体很相仿,顺理成章的事是将这部分的指标作为总体参数的替代。我们把采用一定手段从总体中得到的部分个体的全体称为来自总体的一组样本(sample),样本中个体的个数则称为样本容量或样本大小。从样本中计算而得的指标,称为统计量(statistics)。就是说,统计量是总体参数的估计量。为了估计上述研究中所想知道的两个参数,相应的统计量为:

- ①* 样本中失业人员在样本劳动力所占的比例;
- ②* 样本中失业人员年龄的中位数。

由此可以想象,有限总体的参数与样本统计量在结构上常常很相像,它们之间的差异是:参数是我们想知道的,而统计量是我们已经知道的。

直观地看,样本统计量作为总体参数的估计其合理性在于,样本是总体的一个有代表性的剖面。样本在成份结构上越接近于总体,那末统计量估计参数就越精确。因此,如何采用科学的方法来选取样本就成为抽样调查的关键之一。一般抽样理论需要解决的,也是我们在本书中将要展开的不外乎如下两个最基本的问题:

- (1) 如何抽样?
- (2) 在理想的抽样方案中统计量估计参数的机会误差有多大?

§ 1.2 抽样调查的优越性

抽样调查既然是从“局部”来推断“整体”,它的优越性十分清晰地体现在下述几个方面:

1. 节省费用

全面普查将对总体中的每一个体均付出费用,抽样调查显然大大地缩减了开支。世界上许多国家,尤其是一些人口较多的国家诸如美国,就规定10年才进行一次人口普查。我国亦是如此。两次人口普查之间的时间间隔很长,其间人口资料发生的变化常通过人口抽

样调查给予解决。1995年国务院及国家统计局决定在全国实施1%人口抽样。将比全国普查节省相当可观的费用。

2. 快速

抽样收集的数据远远少于全面普查数据,需要处理的数据自然就比较少。这样,由抽样调查提供的数据资料和决策预测等往往比全面普查更为及时。例如,如果想核对一下商行库存货物的总价值,通常的会计审计工作就是核查全部存货。然而这样的工作常常需要快速完成才有现实意义。全面盘货无法在短时间内获取存货总价值这个急需信息。此时抽样审计将不失是一种理想的选择。

3. 较高精确性

由于抽样调查的工作量远远少于全面普查,习惯于全面调查的人们自然会担心其精确性或有效性。经验表明,来自较少样本的推断与全面调查的计算结果,在精确性与有效性方面基本上具有异曲同工之妙,有时候,抽样推断的结果甚至更为有效一些。关于此点用不着任何大惊小怪。设想一个大公司——庞大的总体——需要盘存,必定需要大量受过培训的工作人员。但是全面培训大量人员实在是件困难的工作。缺乏训练有素的工作人员在全面调查中就不能保证盘存工作的质量。倘若采用抽样盘存,小批工作人员完全可以进行培训,从而保证了核查工作的质量。只要抽样方法科学得当,由抽样推断的结果比起启用大量未得到全面培训人员而使核查质量得不到控制的全面盘存自然更精确一些。在这里,由于抽样误差所引起的损失往往比由于精干的人员和得到控制的核查质量所产生的效益小一些。

当然,抽样调查还具有诸如灵活性等各种优点。譬如有些调查需要有相当熟练技巧的调查人员和精巧昂贵的仪器设备,这使得大规模的全面调查几乎不可能,此时也只有抽样才可能解决问题。

应当强调,抽样调查所有的优越性的共同基本点是:科学的抽样手段、确保质量的调查数据以及精确严格的统计计算程序。

§ 1.3 抽样技巧广泛用于社会经济领域

只需部分地调查一些数据,做到省时省力又省钱,且以较好的精确度反映总体的有关信息的抽样技巧,自然可以在各个领域发挥作用。尤其在社会经济领域,抽样技巧已经成了预测、咨询和决策的必不可少的工具。

人口普查的重要性已经众所周知,它可以为政府有关部门的决策、立法等提供依据;用于对政府计划的制定、实施与评价;为工业开发、商场销售等提供信息;……一句话,如果没有人口普查所提供的信息,许多社会经济研究就无法进行。前面提到的两次人口普查之间进行的人口抽样调查既及时获取有关信息又节省了费用,充分发挥了抽样技巧的优越性。其实,即使在每一次的人口普查中也缺少不了抽样技巧的应用。这是因为两次人口普查之间间隔实在太长,其间有许多情况将会发生变化。因此在新的一次普查中肯定不能沿用上一一次的调查表,必须对一些问题进行修改并且提出一些新的问题。例如,随着社会主义市场经济体制的逐步完善,竞争体制的引进,出现了在计划经济体制下所没有的现象:下岗或失业。因此在新的人口普查中也许会添上以前不会出现的“就业与失业”问题。应当增添些什么样的新问题;或者删去什么样的不合适的过时问题;或者对原来的一些问题的提法作出怎么样的校正,这都可以通过少许的抽样来获取信息并加以完善。

人口普查与其它大规模的全面调查一样,工作量大,历时长,情况比较复杂。各种各样的原因会最终产生差错。如何评估人口普查(或其它大型调查)的质量是一个极需解决的问题。当然再去独立地重做一次普查是绝对不现实的。适度的较小规模的抽样将有助于评估工作。我们可以将抽样样本的数据与普查中相对应的个体所得数据互相作一比较,它们之间差异的大小是对普查质量的有效评估。而且我们可以从样本出发对总体作出估计,这个估计同全面普查所得