

抽样调查与推断

李金昌 著

The cover features large, stylized geometric shapes in green and blue. On the left, there are two overlapping green shapes that resemble a stylized 'S' or a series of parallel lines. On the right, there are two overlapping blue shapes that resemble a stylized 'Z' or a series of parallel lines. These shapes are set against a light yellow background.

中国统计出版社

抽样调查与推断

李金昌 著

中国统计出版社

(京)新登字 041 号

图书在版编目(CIP)数据

抽样调查与推断/李金昌著.-北京:中国统计出版社

1996. 3

ISBN 7-5037-2073-5

I. 抽…

II. 李…

III. 抽样调查-统计分析

IV. C811

中国统计出版社出版

(北京复外三里河月坛南街 38 号)

新华书店经销

双峰印刷厂印刷

850×1168 毫米 32 开本 11.625 印张 29 万字

1996 年 7 月第 1 版 1996 年 7 月北京第 1 次印刷

印数:1-3000 册

定价:19.80 元

(版权所有 不得翻印)

前 言

本书定名为《抽样调查与推断》，完稿之时正值抽样调查 100 周年纪念。100 年来，抽样调查的理论和方法得到了迅速发展，成为了现代统计学的基础。如今，抽样调查已被公认为最重要的统计调查方法并被世界各国广泛应用于各个领域。

我国的抽样调查历史也已有几十年，在理论方法研究和实践应用上都积累了不少成果和经验。尤其是党的十一届三中全会以来，抽样调查更加受到重视，在城乡住户调查、农产量调查、人口调查、市场调查、价格统计和社会调查等方面都取得了显著成绩。国家统计调查制度所包括的统计指标中，已有 1/3 通过抽样调查来取得统计资料。1994 年，我国进一步提出了建立以周期性普查为基础，以经常性抽样调查为主体，重点调查、科学核算等为补充的统计调查方法体系的目标模式，这标志着抽样调查将逐步成为我国的最主要统计调查方法，在应用的广度和深度上将进一步加强。在这种情况下，努力提高我国抽样调查技术的开发应用水平，积极开展关于抽样调查的研究，成了摆在我们面前的一项重要任务。这就要求我们从指导抽样实践应用的需要出发，去发现新内容、研究新问题、探讨新方法，并继续研究一些尚未很好解决的问题。正是在这样的背景和指导思想下，在他人研究成果的基础上，根据自己这些年从事统计专业本科生、硕士研究生

《抽样法》课程教学的积累和相应的研究所得，写成了此书。

本书具有以下几个特点：一是以社会经济领域作为特定应用对象，系统介绍和探讨了抽样调查与推断的理论、方法和技术，因此本书可作为高校统计专业师生、社会经济统计工作者和其它相关人员的教学、研究和参考用书。二是在体系安排上简洁明快，结构合理。本书以抽样调查概述、基本问题、各种抽样组织形式为顺序，把比率估计和回归估计看成与一般估计相并列的抽样估计方法纳入各种抽样组织形式之中，而把成数的抽样推断单列一章，把其它抽样方法和技术合为一章。这样全书共九章，既保持体系完整，又使各章相对独立。三是内容丰富，涉及面广。本书以如何根据抽样调查去推断总体数量特征为主线，尽量多地包括各种实用的抽样调查与推断的方法和技术（但限于篇幅，本书未包括假设检验方面的内容），尽量反映最新的研究动向和成果。作者本人的一些研究体会与所得也大胆地写进了各有关章节中，对一些问题作了力所能及的探讨。四是理论联系实际，注重应用。本书对每一种抽样方法、抽样组织形式和抽样估计方法的优缺点（特点）、应用条件等都进行认真的讨论；对有关公式和结论既力求阐述论证上的严谨、又不过分追求纯数学的推导；所举的例子贴近生活、确切恰当。五是深入浅出，通俗易懂。

本书承蒙恩师黄良文教授的指点并为作序，深感荣幸；承蒙中国统计出版社的大力支持，深表谢意。希望本书能对读者有所助益，并起到抛砖引玉作用。由于本人水平所限，书中挂一漏万及错误、不妥之处在所难免，恳请得到各位专家和读者的批评指正，谢谢！

李金昌

1995年秋于杭州商学院

序 言

抽样调查与推断是现代统计科学的基础，它的产生和应用大大提高了统计活动的认识能力，同时把统计理论提高到崭新的阶段。世界各国自本世纪30年代以来纷纷致力于抽样调查与推断的研究，也积累了丰富的经验，其中如尼曼 (*J. D. Neyman*)、汉森 (*M. H. Hansen*)、霍维茨 (*W. N. Hurwitz*)、马哈拉诺比斯 (*P. C. Mahalanobis*)、科克伦 (*W. G. Cochran*) 等等在社会经济抽样调查的理论、方法以及实践诸方面都作出突出的贡献。

我国抽样调查与推断工作起步也不晚，早在50年代就开始城镇职工和乡村农民的家计调查，60年代成立全国农产量抽样调查队，开展大规模的农产量抽样调查。但是，在传统的计划经济体制下，根据社会经济组织的特点，统计调查是以全面统计报表制度为基本的组织形式，抽样调查与推断在社会经济统计调查方法体系中只居辅助和补充的地位，这就限制了抽样方法的发展。

在社会主义市场经济条件下，社会经济环境发生了根本的变化。一是社会多种经济类型、多种经营形式、多种流通渠道并存，同时发展，经济成份复杂多了；二是经济经营分散，统计单位迅速增加，而且行业交叉变动很大；三是企业作为经济法人，独立经营、自负盈亏，经济决策多层次，经济利益多元化，国家、集体、个人之间、上下之间、调查与被调查之间的关系也更加复杂

了。所有这些都集中说明，依靠定期全面统计报表制度来掌握情况已经不能适应形势需要了。另一方面，基于国家改革开放和宏观调控的需要，要求统计扩大职能，发挥信息、咨询、监督的作用，对于统计的需求又大大增加了。抽样方法正是以自己的省、准、快、便的优势有广阔的用武之地，满足新条件下我国统计工作的需要。现在我国已确定建立以周期性普查为基础，经常性抽样调查为主体，重点调查、科学推算等为补充的多种方法综合运用的统计调查方法体系。国家城市、乡村抽样调查队逐步扩大调查内容，此外还成立企业调查队搜集我国工业、固定资产投资、商业、交通等基本统计资料。抽样方法将成为我国统计最重要的常规武器，它的发展前景无限广阔。

我们也应该看到，抽样思想和全面统计思想迥异，它的科学性为大众所接受亦非易事，掌握一套抽样方法更需艰苦的摸索。当前我们面临的任務就是大力学习、宣传、普及抽样调查知识，提高抽样意识，并且组织力量总结工作经验，引进适合我国实际情况的抽样方法，开展抽样调查推断理论研究，有效解决实际工作中存在的问题。

我很高兴看到杭州商学院统计系主任李金昌副教授所著《抽样调查与推断》一书为大家提供了很好的学习材料。作者长期从事抽样理论与方法的研究，有丰富的科研积累，本书就是作者多年在大学本科和研究生教学和研究基础上的系统总结。有明显的特点：

1. 内容全面丰富。本书在论述抽样调查推断一般理论与方法问题之后，再分别比较深入地讨论各类抽样形式如纯随机抽样、分层抽样、等距抽样、整群抽样、多阶段抽样，以及其他抽样方法和技术。讨论的起点较高，有一定深度，但又深入浅出、易学易懂。

2. 理论联系实际。各类抽样方法的研究注重解决实际问题，尤其下功夫总结我国抽样工作经验，围绕我国社会经济实际情况，

设例计算，给人深刻印象。

3. 系统性好。全书章节安排注意循序渐进，由浅入深，把工作的先后顺序和理论的逻辑衔接结合起来，形成自己的特色。

4. 可读性强。在论述方法上，作者注意先说明问题的背景、问题的产生、以及解决问题的方向，然后进一步分析问题的内在联系和数量关系。注意把直观说理和数学论证结合起来，这样使读者易于领会问题的基本观点，又不失讨论的严密性，同时可以满足不同层次读者的要求。

本书适宜作为大学统计专业的教材或教学参考书，也可供广大统计干部培训或自学之用。我希望《抽样调查与推断》的出版，将为促进我国抽样调查工作的统计方法改革事业作出自己的贡献。

黄良文

1995. 10. 于厦门大学



目 录

第一章 抽样调查概述	(1)
第一节 抽样调查的概念	(1)
第二节 抽样调查的特点	(3)
第三节 抽样调查的优点与作用	(7)
第四节 抽样调查的历史与趋势	(10)
第二章 抽样调查基本问题	(16)
第一节 总体与样本	(16)
第二节 抽样方法	(21)
第三节 样本统计量的抽样分布	(26)
第四节 抽样误差	(30)
第五节 抽样调查的一般程序	(40)
第三章 纯随机抽样	(46)
第一节 抽样形式	(46)
第二节 纯随机抽样的简单估计	(50)
第三节 纯随机抽样的比率估计	(57)
第四节 纯随机抽样的回归估计	(75)
第五节 样本单位数的确定	(90)
第六节 子总体抽样估计	(95)

第四章 分层抽样	(100)
第一节 抽样形式	(100)
第二节 分层抽样的简单估计	(103)
第三节 分层抽样的比率估计	(121)
第四节 分层抽样的回归估计	(128)
第五节 分层抽样估计的两种新思路	(134)
第六节 分层抽样的其它问题	(142)
第五章 等距抽样	(158)
第一节 抽样形式	(158)
第二节 估计量的方差	(166)
第三节 估计量方差的样本估计	(174)
第四节 等距抽样的比率估计与回归估计	(187)
第五节 等距抽样的其它问题	(196)
第六章 整体抽样	(201)
第一节 抽样形式	(201)
第二节 等群的整群抽样估计	(203)
第三节 不等群等概率的整群抽样估计	(209)
第四节 不等群不等概率的整群抽样估计	(218)
第五节 整群抽样的其它问题	(233)
第七章 多阶段抽样	(238)
第一节 抽样形式	(238)
第二节 抽样单位大小相等的多阶段抽样	(241)
第三节 一级单位大小不等的两阶段抽样	(253)
第四节 分层两阶段抽样	(267)

第八章 成数抽样推断	(275)
第一节 成数的纯随机抽样推断	(275)
第二节 成数的分层抽样推断	(282)
第三节 成数的整群抽样推断	(287)
第四节 成数的多阶段抽样推断	(292)
第五节 成数的比率估计	(298)
第九章 抽样的其它方法和技术	(306)
第一节 双重抽样	(306)
第二节 样本轮换	(317)
第三节 交叉子样本	(325)
第四节 无回答处理	(330)
第五节 敏感性问题的随机化回答	(336)
参考文献	(344)
附录一 随机数字表	(346)
附录二 正态分布概率表	(350)
附录三 t 分布概率表	(352)
附录四 F 分布概率表	(354)
附录五 χ^2 分布的上侧临界值表	(358)

第一章 抽样调查概述

第一节 抽样调查的概念

抽样调查是一种非全面统计调查方法。对于大量复杂的社会经济现象的调查，有时要用定期普查方法，有时则需或只能用典型调查、重点调查或抽样调查方法。其中抽样调查已是当今最主要的统计调查方法，广泛应用于社会、经济、科技、自然等各个领域。抽样理论和方法也已成了现代统计学的重要组成部分，并且成了现代统计学中发展最快、最为活跃的一个分支。

对抽样调查最通俗的理解就是抽取样本进行调查。由于抽取样本的依据不同，广义的抽样调查可以分为两类：一类是以调查者的主观判断为依据来抽取样本的有意抽样，也称目的抽样，主张“按照一个过去统计调查为基础的合理方法来选取样本单位。”在实际应用中，有意抽样又有以下几种形式：

典型抽样，即调查者凭借自己对调查对象的了解和主观判断能力，有意识地从总体中抽取具有平均水平或具有一般特征的单位作为样本。这种抽样的优点在于可以发挥调查者的主观能动性和充分利用已掌握的有关信息，避免产生极端的偏误。其缺点在于主观随意性较大，不同的人会有不同的判断，并且对调查结果的评价缺乏客观标准，对调查偏误也无法掌握和控制。事实上，这种抽样是单直型的典型调查。

随意抽样，即仅利用已有的名册、号簿和地图等资料任意确定调查单位。这种抽样的优点是简便易行，但却存在抽样所依据的总体不能涵盖所有被研究现象总体单位的缺点，所抽的样本经

常会产生大的偏差，从而使调查结论与实际情况不符，甚至相反。1936年，美国《作家文摘》杂志根据全国各地电话号簿对总统竞选结果所作的预测性民意抽样测验，就是一次失败的随意抽样。

定额抽样，即在对调查对象按一定标志进行分类，且根据有关历史资料确定每类应抽选的比例后，每类分别依主观判断抽取一定数量的单位作为样本。这种抽样由美国著名的专家盖洛普创立，比随意抽样有很多改进之处。其优点在于能使所抽的样本结构比较接近总体的结构，提高了样本的代表性，且可使调查单位数减少，因而具有一定的科学性，在实际应用中曾得到较快的推广。但它仍存在无法排除主观因素随意性影响和难以估计、控制抽样误差的缺点。1948年，盖洛普咨询公司用这一方法进行总统竞选的预测性民意测验，也失败了。事实上，这种抽样是划类选典式的典型调查。

另一类是以概率论和随机原则为依据来抽取样本的随机抽样。这种抽样能保证总体中每个单位有相等的被抽中的机会，哪个单位被抽选不受调查者主观因素的影响，从而能充分保证样本的代表性。本书所要讨论的抽样调查即指随机抽样调查。

由于抽样调查的最终目的是要从数量上认识调查对象总体的特征，因而严格意义上的抽样调查包含了抽样推断的内容，抽样理论本身就包括了抽样调查理论和抽样推断理论两部分。因此，我们可以给出抽样调查的完整概念为：抽样调查是按照随机原则，从研究现象总体的所有单位中，抽取部分单位作为样本，然后以样本的观测或调查结果对总体的数量特征作出具有一定可靠程度和精度的估计或推断的一种统计调查方法。例如，从某市的45万户居民家庭中随机抽取500户进行收入调查并据以推算全市居民家庭的平均收入和总收入；从某地区的20万亩小麦地块中随机抽取500个地块点进行实割实测并据以推算全地区的平均小麦亩产量和总产量；从某批1万件产品中随机抽取100件进行检测并据以判断该批产品的合格率等等，都是随机抽样调查。

随机抽样优于目的抽样，这是因为随机抽样能根据样本自身得出关于总体数量特征的估计量的精确度，而目的抽样却没有一种测定估计值的精确性或可靠性的客观方法。

第二节 抽样调查的特点

与典型调查、重点调查等非全面统计调查相比，抽样调查具有以下几个明显的特点：

一、在调查单位的选取上，遵循随机原则

典型调查以人的主观判断来选取调查单位，重点调查以在总体标志总值中所占比重大这一客观标准来确定调查单位，而抽样调查则以随机原则来抽取调查单位。

所谓随机原则就是等可能性原则，它必须保证总体中每个单位有同等被抽中的机会。根据这一原则，就可以使总体中各单位的中选或不中选都不受任何主观因素的影响，而完全取决于随机因素。也即，不论某单位在总体中分布何处，被抽中的可能性全由许多随机因素的实际情况而定。只有这样，才能有较大的可能性使所抽选的样本保持与总体相类似的结构，从而保证样本的代表性，提高对总体数量特征进行估计与推断的精度。因而，随机原则是抽样调查的根本原则。

但是遵循随机原则，排除人们主观因素对样本抽选的影响，并不等于在抽样调查中不要发挥人的主观能动作用。随机原则是取样的原则，指的是在抽取样本单位时要保证每个单位具有同等可能性，但怎样才能实现这个原则，怎样在遵循这个原则的基础上，使抽样的效果更好，正是需要发挥人们的主观能动性之处。例如，在对总体充分了解的基础上，对各单位进行排队、组合；利用有关辅助信息；设计最佳的抽样方案；选择最佳的估计、推断方法；尽量减少抽样中的误差等等，都离不开人们的主观能动作用，而

且越来越重要，成了抽样调查的一个重要研究方向。

二、在调查的功能上，能以样本的数量特征去推断总体的数量特征，即以部分推断总体

统计研究的目的是要认识现象总体的数量特征，但许多社会经济现象不能或不必要靠全面调查来取得所需的全部资料。如何只通过对总体中部分单位的调查，来达到认识总体数量特征的目的，正是抽样调查所有的独特功能。虽然样本的观测、调查结果，与总体的真实情况可能会有差异，或是正差，或是负差，但概率论与数理统计已严格证明了随机抽样下的样本与总体之间的内在联系和抽样分布规律。例如样本平均数的平均数等于总体平均数，样本方差的期望值等于总体方差，样本平均数会随着样本单位数的增加而更接近总体平均数，等等。这样，我们就可以用样本的数量特征去推断总体的数量特征。相反，重点调查不具备直接推断总体的条件；典型调查则由于无法阐明典型调查单位与总体之间的内在联系而使估计结果缺乏科学根据，具有较大的随意性和盲目性，因而严格地讲也不具备由部分推断总体的功能。由此可见，抽样调查既非全面调查，又不同于其它非全面调查，独具由部分推断总体的功能。抽样推断的过程，就是由部分到全部，由具体到一般的过程。

三、在推断手段上，是以概率估计的方法对总体的数量特征进行估计和判断

抽样推断在方法上运用不确定的概率估计法，而不是运用确定的数学分析法。由于抽样的随机性，从总体中所抽取的样本是不确定的，在随机因素的影响下可以有各种各样的构成，因而样本的观测、调查结果本身也是随机变量，即不同的样本会得出不同的估计值。不确定的估计值与确定的总体数值之间并不存在象自变量与因变量那样的函数关系，因而当用样本的统计量来推断

084524

总体数量特征时,其可靠性只能以一定的概率保证程度来说明,而非完全肯定,例如,根据样本居民的收入情况来估计一个城市居民的收入情况,只能在一定的概率保证程度下(譬如 95%)作出具有一定区间范围的估计,而不能保证以 100%的把握程度给出唯一肯定的估计值。当然,概率保证程度的大小是可以根据实际情况和需要而变动的。

四、在推断理论上,以大数定律和中心极限定理为依据

大数定律也称大数法则,它论证了如果独立随机变量总体存在着有限的平均数和方差,则对于充分大的样本,可以近乎 100%的概率,期望样本平均数与总体平均数的绝对离差为任意小。也即随着样本单位数的增加,样本平均数将趋于接近总体平均数。这也表明,在抽样调查时,只有样本单位数足够多才能显示总体的规律性,因而对样本单位数的多少是有一定要求的。这一点也与典型调查和重点调查不同。但是大数定律并没有阐明样本平均数与总体平均数的离差有多大,离差不超过一定范围的概率是多少,这个离差的分布如何等问题。而中心极限定理则正好弥补了这一不足。

中心极限定理也称李亚普诺夫定理(1901年),它论证了如果总体存在有限的平均数和方差,那么不论这个总体的变量如何分布,随着样本单位数的增加,样本平均数的概率分布会趋向于正态分布,从而为抽样误差的概率估计提供了有效的条件。

正态分布是统计学上最常用最重要的概率分布。设 X 为连续型随机变量,则其概率密度函数为:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

其中 \bar{x} 为总体平均数, σ 为总体标准差。

正态分布的图形如下页图 1-1 所示。其中正态曲线与横坐标之间的极限面积等于 1。

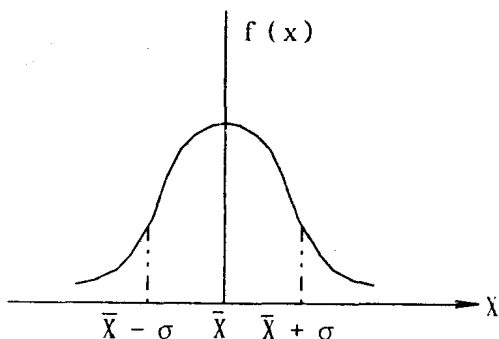


图 1-1

根据正态分布的有关特性，我们就可以求出某数值在任意一个区间内的概率的大小。例如，以 \bar{X} 为中心，变量取值在 $\pm\sigma$ 范围内的概率为 68.27%，在 $\pm 1.96\sigma$ 范围内的概率为 95%，在 $\pm 2\sigma$ 范围内的概率为 95.45%，在 $\pm 3\sigma$ 范围内的概率为 99.73%，等等，从而为抽样推断的概率估计和误差控制提供了科学依据。取值范围与概率之间的关系已编制成正态分布概率表，可供查阅。

五、在推断效果上，抽样误差可以计算并加以控制

抽样调查是非全面调查，根据部分单位观测、调查所得的数据，去估计与推断总体数量特征，不可能是绝对准确可靠的，只能是近似准确或相对准确，也就是说可能会产生一定的由随机因素引起的代表性误差，即抽样误差。抽样误差的大小，直接反映了抽样推断效果的好坏，因此我们总希望抽样误差尽量小并知道它有多大。做到这一点是可能的，因为样本统计量的抽样分布是可以描述的。根据前述的大数定律、中心极限定理及正态分布规律，我们可以估计出衡量抽样推断误差大小的指标——样本统计量的标准差，并且能通过各种有效的办法把抽样误差控制到某一