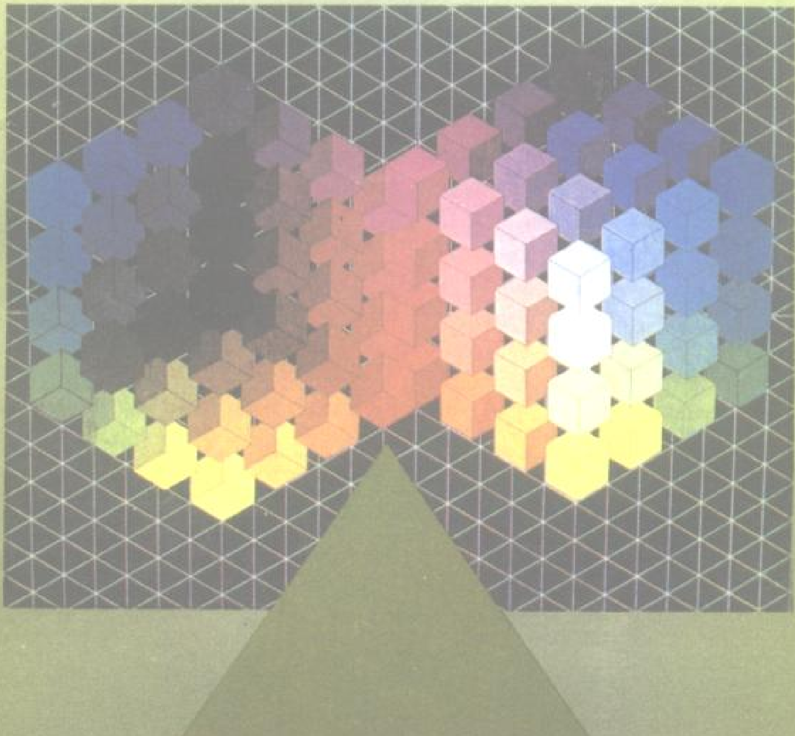
九三科学技术丛书

抽样调查与抽样检查

——理论方法与应用

杨义群　主编



Y30

九三科学技术丛书

抽样调查与抽样检查

——理论、方法与应用

杨义群　主编

学 苑 出 版 社

(京)新登字151号

抽样调查与抽样检查

——理论、方法与应用

主　　编： 杨义群
责任编辑： 裴雪重
封面设计： 邓中和
出版发行： 学苑出版社　邮政编码：100032
社　　址： 北京市西城区成方街33号
印　　刷： 北京彩虹印刷厂
经　　销： 全国各地新华书店
开　　本： 787×1092　　1/32
印　　张： 8.25　　**字数：** 174千字
印　　数： 0001—2600
版　　次： 1993年10月北京第1版第1次
ISBN7—5077—0588—9/O·6
定　　价： 4.60元

《九三科学技术丛书》编辑出版说明

九三学社是由科学技术和文化教育、医学卫生界高、中级知识分子组成的致力于社会主义事业的具有政治联盟特点的政党。九三学社成员一般具有较高的知识智能优势，他们在各自的领域为我国的社会主义建设已经做出了并继续做着贡献。

为了较客观地反映九三学社成员的科研现状，不断介绍、推广九三学社成员的科技成果，我们编辑出版了这套《九三科学技术丛书》，期望能为尽快实现科学技术向现实生产力转化，推进改革开放、加快经济建设这一目标服务。

《九三科学技术丛书》是一个开放的丛书系列。我们拟兼顾各学科选题比例，并视我们现有的经济承受能力，有计划地选取一些有特点的、有影响的科技著作陆续出版，坚持数年，形成阵容。

《九三科学技术丛书》仅收九三学社成员所著、所编著、编译的科技著作。九三学社成员所翻译的国外科技著作不收入本丛书。在对作者介绍时，对九三学社成员与非九三学社成员合作的著作，本丛书仅对九三学社成员作者加以介绍。九三学社成员所编著的社会科学著作，我们另编辑《九三社会科学丛书》收入出版。

由于我们水平有限，工作中定会有许多疏漏，敬希各界专家、学者和广大读者不吝赐教。

学苑出版社编辑部

1992年7月

前　　言

本书与一般数理统计教材或专著相比，有以下不同之处：一、一般数理统计教材用较多篇幅叙述概率论，而对实用统计方法只是罗列一下；本书略去对概率论的论述，直接介绍统计方法，内容多，范围广，实用性强。二、一般数理统计教材关于理论统计学的论述只注重经典的几个方面，对统计学新方法几乎均未涉及，而数理统计有关专著虽能深入探讨某些方面问题，但数学推理艰深难懂，不易普及推广。本书克服上述两方面的弊端，既论述了很多近年来理论界提出的新方法，又不拘泥于专业性推理论证，而是注重方法的介绍和应用的推广，较之传统教材从内容上有所突破。

本书适用读者对象：大专院校学生，数学教师，统计工作者及其他科研工作者。既适合数理统计工作者进行理论研究，又适合自然科学与社会科学（包括农林、医药、生态、人口、经济、管理等）各领域中需要进行抽样的实际工作者，并对高年级学生、研究生及教师有应用价值。

主　编　杨义群

副主编　陆根尧　张荷观

　　　　陈茂海

序

抽样调查是一种了解社会情况的科学方法，几十年来越来越受到重视并有了相当的发展。然而适合中级以上技术人员阅读的书籍却不多。已出版的或是很简单，或是只讲原则不讲应用，有的又偏重于理论，只强调数学推导而缺乏应用的实例，因此适应中级以上层次的要求、比较实用的一本书就很需要。

抽样检查在国内已有不少教材出版，並在工业界得到了广泛的应用，在商业和其他方面近来也越来越受重视了。抽样检查和抽样调查实际上是联系密切的两大主题，作者在这两方面从理论到应用，尤其是在农业方面的应用，都做了不少工作，积累了经验，这使作者有条件写好这两个内容。

另一方面，作者阅读了国内外许多文献，本书把这方面的进展情况概要地介绍给广大的读者，这样本书在体系和写法上就要付出相当的努力。应该说这些努力是成功的，做到了在不大的篇幅内介绍较多的内容。

我衷心祝愿此书的出版能推动有关方面的教学和科研工作，祝愿此书产生较好的社会影响。

张尧庭

1990年10月于武汉大学

目 录

第一篇 抽样调查与估计方法

第二篇　抽样检验（抽样检查）

第一篇　抽样调查与估计方法

第一章　抽样调查的初步概念

1.1.　抽样调查方法的广泛实用性

在需要进行各种战略决策的领域，无论是社会科学或自然科学，还是生产建设或国防建设，例如：经济调查、人口调查、能源调查、作物播种面积估计、病虫害程度估计、森林、草原、畜禽与农田估产，以及产品，特别是药品与进出口商品质量检验等等，其决策所依据的信息绝大多数是来自于抽样调查的。又如数学规划或模型中的许多参数估计也都是来自于抽样调查。（杨义群等，1987）

抽样调查与全面调查（普查）相比，有以下优点。

（i）普查需要耗费大量人力、物力、财力与时间，这在实际中常常不可能或不合算。这里并不一概否定普查。例如我国的人口调查，是采用定期普查与抽样调查相结合的办法。1982年的全国人口普查中，使用了20多台电子计算机，花了近两年的时间，而普查的质量还是用了抽样调查的方法进行估计。

（ii）抽样调查的调查对象较少，可用较高级的人员与设备进行高质量的调查。因而在同样的费用下，抽样调查有时会比普查有更高的精度。

（iii）普查需要较长的时间，往往难于在战略决策前完成。

（iv）当调查或观测具有破坏性（也即个体在被调查或观测之后将失去其原有的价值，例如化学分析、军火试验或寿命试验等破坏性试验）时，就不能进行全面调查。

（v）当研究的总体中有无限多个个体（见例1.3.1）时，普查就根本不可能。

可见，抽样调查的应用范围要比普查广泛得多。而且，广义地说，普查是抽样调查中非常特殊的一种情形：样本容量n恰好等于总体容量N，也即$n=N$。而抽样调查一般有$n\leqslant N$。所以抽样调查方法比普查灵活得多。

因此，抽样的理论与方法是生产建设、国防建设、社会科学与自然科学研究中实际应用最广泛的基础学科之一。但是，现在我国大多数大学与系科尚未安排这门基础课程，这是很不合理的。例如，在进出口商品检验中，我国的有关科技人员由于不了解抽样理论与方法，而使我国吃了不少难言的亏。

1.2. 抽样的基本单元（个体）

抽样之前，总体必须划分成一个个抽样单元，即抽样的基本单元（最小单元），简称为个体。

个体有时可定为天然形成的。例如，国内的一个个省，省内的一个个县，县内的一个个农场或乡，乡内的一个个专业户或果园，果园中的一株株树，树上的一个个果子或一张张叶片（中的某一或某些指标值）等等。

个体有时需要人为确定。例如，在估计亩产量或估计害虫在田野中的密度时，虽然还有着一株株植株或天然的一片片田野，但是以此作为个体往往太小或太大，这时可抽取其中一小块一小块田地，观测其中的产量或害虫数。而这些“一

小块”（称为样方或抽样框）的面积与形状是人为确定的。我们应当根据已有的信息或经验知识，设计出尽可能合理的样方（参见8.1节）。这时，一个样方（中的某一或某些指标值）就是总体中的一个个体。

1.3. 有限总体的基本参数

定义1.3.1 用N表示总体中个体的个数，称为总体容量。$N<+\infty$的总体称为有限总体。$N=+\infty$的总体称为无限总体。

例1.3.1 在研究生物中的某一品种时，因为这一品种的繁殖，从长远的观点来说，这一品种的个体是无穷无尽的，所以要研究的是一个无限总体。类似地，在研究某种工艺下的产品时，这种产品的全体是一个无限总体。

在实际中，总体的容量N大多是有限的。而且，即使遇到无限总体，一般来说，总可以把无限总体看成容量N充分大的总体，从而利用有限总体抽样的各种理论、方法及其分析计算公式。要精确的话，还可取$N\to+\infty$时的极限公式。

因此，为讨论的一般计，在无特别声明时，总假设总体容量N是有限的。现把研究的总体记为Y，总体Y中的N个个体依次记为

$$Y_1, Y_2, \cdots, Y_N \quad . \tag{1.3.1}$$

总体Y的总和数$\widetilde{Y}$，平均数$\overline{Y}$以及两种标准差σ与S依次记为

$$\widetilde{Y} := Y_1 + \cdots + Y_N, \quad \overline{Y} := \widetilde{Y}/N \tag{1.3.2}$$

$$\sigma \triangleq \sigma(Y_1, \cdots, Y_N) \triangleq \left[\frac{1}{N} \sum_{i=1}^{N} (Y_i - \overline{Y})^2 \right]^{1/2} \tag{1.3.3}$$

与

$$S=S(Y_1,\cdots,Y_N)\triangleq\left[\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\overline{Y})^2\right]^{1/2}. \tag{1.3.4}$$

它们都是 $(Y_1,\cdots,Y_N)$的函数。

例1.3.2 若记 $Y_1,\cdots,Y_N$ 为 N 株植株中的虫口数，则$\widetilde{Y}$就是这N株植株中的虫口总体，而$\overline{Y}$是这N株植株中平均每株植株中的虫口数。

例1.3.3 若把某一区域划分成N个样方，记$Y_1,\cdots,Y_N$为这N个样方中的产量，则$\widetilde{Y}$就是该区域的总产量，而$\overline{Y}$是该区域中平均每个样方的产量。

注1.3.1 当容量N已知时，由（1.3.2）式知，这时估计总体平均数$\overline{Y}$与估计总体总和数$\widetilde{Y}$同样难易。而当N未知时，总体平均数$\overline{Y}$一般比总体总和数$\widetilde{Y}$容易估计。除了总体平均数与总和数外，实际中有时也需要估计其它总体参数，例如分位数（冯士雍等，1989a）等。

第二章 简单随机抽样

2.1. 定义

简单随机抽样简称为简单抽样，主要有放回的与不放回的两种。

定义2.1.1 放回的简单抽样是：先确定样本容量，记为n，然后连续抽取n次，每次从总体中抽取一个个体并观测其值后就放回总体中去，而且每次抽取时，总体中每个个体都有同等的机会(相同的概率)被抽取。

定义2.1.2 不放回的简单抽样是：先确定样本容量，记为n，然后连续抽取n次，每次从剩下的总体中抽取一个个体，而且每次抽取时，总体中剩下的个体都有同等的机会（相同的概率）被抽取。

定义2.1.3 简单抽样取得的样本简称为简单样本。第i次抽得的个体记为

$$y_i = Y_{k_i} \qquad (i=1,\cdots,n), \qquad (2.1.1)$$

其中k_i称为第i次的入样号码。

简单抽样时，上述随机变量y_i与k_i的概率分布显然为

$$P(y_i = Y_j) = P(k_i = j) = 1/N(j=1,\cdots,N;\ i=1,\cdots,n). \qquad (2.1.2)$$

样本来自于放回的简单抽样时，上式是显然的，而且还可知随机变量（2.1.1）互相独立。样本来自于不放回的简单抽样时，由机会均等与对称性也易理解上式成立（可用排列组合工具加以证明），但这时随机变量（2.1.1）不互相独立。