

抽样调查技术

樊鸿康 编著

市场调查与预测
人文调查与分析

信息搜集
民意测验
决策依据

随机抽样
分层抽样
多阶抽样
等距抽样
整群抽样

调查设计
抽样设计
问卷设计

南开大学出版社

抽样调查技术

樊鸿康 编著

南开大学出版社

439542

抽样调查技术

樊鸿康 编著

南开大学出版社出版

(天津八里台南开大学校内)

邮编 300071 电话 3508542

新华书店天津发行所发行

河北省邮电印刷厂印刷

1995年10月第1版

1995年10月第1次印刷

开本: 850×1168 1/32 印张: 10.75

字数: 265千 印数: 1-2500

ISBN 7-310-00800-6

O·89 定价: 12.20元

前 言

抽样调查是了解自然和社会、收集信息资料的一种科学方法和手段。几十年来，抽样调查越来越受到人们的重视，得到广泛的应用，并有了相当的发展。但是，到目前为止，关于抽样调查的既有系统理论又能指导实际运用的书籍却不多。

本书是根据著者在南开大学讲授《抽样调查技术》课程十几年的教学经验、讲稿、讲义，以及在参加和参与各重大抽样调查实践的基础上写成的。本书手稿的产生经历了以下两个阶段：（1）先是在对研究生讲授的《抽样技术》课程讲稿的基础上，降低难度充实内容改编成了针对大学本科生教学用的讲义。在对本科生的教学过程中，由于不断地增加内容而使讲义越来越厚。（2）在对实际工作者的培训中，特别是受国家统计局教育培训中心的委托，在南开大学举办的两期全国性培训班的教学中，根据实际工作者的需要和情况，着重研究了各抽样调查方法的内在联系，总结其规律，突出其共性，简化了公式，统一了格式，而使讲义变薄，使其基本方法简单易懂，易学易用。本书与其它抽样调查的书籍相比较有如下特点：

1. 本书为了适应实际工作者进行抽样调查和统计分析的需要，为了便于自学和实际操作，全书各章节基本上分三部分论述。第一部分以实例介绍抽样方法的特点、适用范围、具体实施办法，以及实际调查中可能出现的问题和解决的方法；第二部分是理论证明和习题；第三部分是公式一览表与实例。对数学比较生疏的读者和实际工作者，可以只阅读第一部分和第三部分，便

可掌握抽样调查的基本方法 and 应用。

2. 本书为了适应教学，在内容编排上完整系统、由浅入深、循序渐进；在理论证明上，力求严谨，推演连贯细致；在各种方法的比较中，做到全面深入，并配有习题，是一本理想的教科书。本书从基本概念入手，以第三章、第四章、第五章基本方法为“板块”或“菜单”，在以后几章中着重训练读者，针对实际调查的需要，灵活运用“板块”，巧妙地拼造出各种实用的调查方法，用基本“菜单”拼配出各种“美味佳肴”，从而培养读者分析问题和解决问题的能力，培养读者在实际调查中的应变、综合和创新能力。

3. 本书重视各种方法的特点和共性的研究。无论从内容上，还是从公式上，读者将会看到它们的基本思路，处理问题的方法，公式的格式与结构都大致相同。后一章是前一章的继续和发展。尽管本书内容多，范围广，抽样方法千变万化，但其基本思想，基本内容不变，一切都有规律可循，便于理解和记忆。本书将会使读者增强掌握抽样技术的信心，使您耳目一新！

编 者

目 录

第一章 绪论	(1)
§ 1.1 统计调查计划与抽样调查方案	(1)
一 统计调查计划	(1)
二 全面调查与抽样调查	(3)
三 抽样调查的优点	(4)
四 抽样调查方案	(5)
五 抽样技术	(6)
§ 1.2 抽样误差与非抽样误差	(7)
一 抽样误差	(7)
二 非抽样误差	(8)
§ 1.3 最佳抽样方案与设计效果 $deff$	(9)
一 最佳抽样方案	(10)
二 设计效率 $deff$	(10)
第二章 抽样调查的基本知识	(12)
§ 2.1 总体与样本	(12)
一 总体	(12)
二 样本	(13)
三 概率抽样与非概率抽样	(14)
四 抽样单位与抽样框	(15)
§ 2.2 抽样调查问题的基本提法与描述	(16)
一 总体数字特征	(16)
二 调查目标量	(18)
三 估计量及其评价	(19)
四 样本额 n 的确定	(20)

§ 2.3	几种常用的抽样方法	(21)
一	简单随机抽样	(21)
二	分层抽样(类型抽样)	(22)
三	整群抽样	(22)
四	多阶抽样	(23)
五	系统抽样	(23)
六	二重抽样或两相抽样	(23)
第三章	简单随机抽样法	(25)
§ 3.1	概述与使用	(25)
一	定义与概念	(25)
二	实施办法	(26)
三	估计量及其误差	(26)
四	样本额 n 的确定	(27)
五	实例说明	(28)
§ 3.2	简单估计法(SE 法)	(31)
一	简单估计量	(31)
二	简单估计量的无偏性	(32)
三	简单估计量的方差(均方误差)	(33)
四	方差(均方误差)的估计量	(36)
五	有放回等概率的随机抽样	(37)
六	根据精度要求确定样本额 n	(38)
七	对方差预先估计后, n 的确定方法	(41)
§ 3.3	关于比例(成数)的简单估计	(43)
一	问题的提法	(43)
二	关于比例 P 的估计	(43)
三	样本额 n 的确定	(45)
四	样本额 n 的确定的进一步探讨及逆抽样方法	(47)
五	置信区间	(50)
§ 3.4	部分估计	(52)
一	问题的提法	(52)

二	部分估计	(54)
三	两种方法的比较	(56)
四	考虑调查费用决定样本额的举例	(57)
§ 3.5	小结与公式一览表	(59)
一	估计量	(59)
二	均方误差及其估计量	(60)
三	样本额的确定	(61)
四	公式一览表	(62)
§ 3.6	比率估计法 (简记为 RE 法)	(65)
一	问题的提法	(65)
二	比率估计法	(66)
三	如何评价有偏估计量及比率估计量的合理性	(71)
四	比率估计法与简单估计法的比较	(73)
五	比率估计量成为最佳线性无偏估计量的条件	(75)
六	比率估计量的偏差	(78)
七	置信区间与 n 的确定	(80)
八	无偏的比率估计量	(82)
九	部分估计中的比率估计法	(85)
十	部分之部分的比率估计问题	(88)
十一	公式一览表及实例	(92)
十二	两个比率的比较	(95)
§ 3.7	回归估计量	(96)
一	当 b 预先确定情况下的回归估计量	(96)
二	从样本计算出 b 时的回归估计量	(99)
三	与比率估计、简单估计的比较	(103)
四	线性回归模型的线性回归估计量	(103)
习题 3		(105)
第四章	分层抽样法	(112)
§ 4.1	概述与使用	(112)
一	概念与特点	(112)

二	如何分层	(113)
三	估计方法	(114)
四	样本容量 n 的确定与分配	(115)
五	举例	(123)
§ 4.2	分层抽样的简单估计法	(127)
一	简单估计法	(127)
二	按比例分配样本额	(129)
三	最优分配样本额	(130)
四	某些层抽样比超过100%时的修正	(131)
五	分层抽样在估计精度上的收益	(135)
六	公式一览表	(140)
§ 4.3	分层抽样的部分估计法	(142)
一	问题的提法	(142)
二	估计法	(143)
三	样本额的确定	(144)
§ 4.4	分层抽样的比率估计法	(146)
一	比率估计法	(146)
二	SRE 法与 SSE 法之比较	(148)
三	组合比率估计量	(149)
四	组合估计量与分层比率估计量的比较	(151)
五	分层比率估计法的样本额分配	(151)
§ 4.5	分层抽样在实际应用中的问题	(153)
一	偏离最优分配对精度的影响	(154)
二	层含量的偏差对精度的影响	(155)
三	当调查项目超过一个时, 样本额在各层之间的 分配问题	(158)
四	小样本情况下的两种分层方式	(164)
五	抽样后分层(事后分层)	(166)
§ 4.6	分层抽样中的回归估计量	(168)
一	分层抽样的回归估计量	(168)

二	用样本估计回归系数	(170)
习题 4		(172)
第五章	二阶抽样法	(179)
§ 5.1	二阶抽样问题的一般提法	(179)
§ 5.2	二阶抽样法的估计法则	(181)
一	抽组为随机抽组时的估计法	(181)
二	抽组为有返回 ($P_1, \dots, P_k$) 概率抽组的估计法	(182)
§ 5.3	组内为随机抽样时的估计法	(182)
一	抽组为随机抽组时的估计法 (TSE法)	(183)
二	以 ($P_1, \dots, P_k$) 有放回抽组估计法 (TPE法)	(188)
三	TSE法与SE法的比较	(190)
§ 5.4	样本额的分配与确定	(192)
一	费用固定的条件下, 使均方误差最小的	
样本额的分配		(192)
二	满足给定的精度, 使调查费用最小的样本额的分配	(195)
§ 5.5	二阶抽样的部分估计	(198)
§ 5.6	公式一览表及其应用	(203)
一	均值与总量的估计	(203)
二	公式一览表与例题	(206)
三	部分估计中的公式一览表	(212)
5.7	混合抽样	(215)
一	对一级抽样单位分层 (分类) 的二阶抽样法	(215)
二	对二级抽样单位分层的二阶抽样法	(220)
§ 5.8	二阶抽样的比率估计法	(223)
一	等概抽组 (随机抽组)	(223)
二	以 ($P_1, \dots, P_k$) 有放回抽组	(226)
§ 5.9	多阶抽样	(228)
习题 5		(234)

3
第六章 整群抽样法与系统抽样法(等距抽样法).....(238)

§ 6.1 整群抽样法.....(238)

一 整群随机抽样法的简单估计法(CSE法).....(238)

二 整群以($P_1, \dots, P_k$)概率有返回抽样的估计法...
(CPE法).....(240)

三 整群随机抽样法的比率估计法(CRE法).....(241)

四 公式一览表及实例.....(245)

6.2 系统抽样(等距抽样)法.....(250)

一 系统抽样的估计法(SYE法).....(251)

二 个体标志值与其次序有某种线性关系.....(253)

三 个体标志值与其次序有某种周期性关系.....(257)

四 个体次序随机排列.....(258)

五 自相关总体.....(260)

六 分层等距抽样.....(265)

七 二阶等距抽样.....(265)

习题6.....(266)

第七章 二重抽样(两相抽样).....(270)

§ 7.1 为分层抽样进行的二重抽样.....(270)

一 抽样与估计法.....(270)

二 关于样本额的最优分配问题.....(277)

§ 7.2 为回归估计进行的二重抽样.....(281)

一 估计法.....(281)

二 样本额的确定与比较.....(284)

三 二重抽样的比率估计.....(286)

§ 7.3 对同一总体的反复抽样.....(287)

一 在两个不同时间上的抽样.....(289)

二 在两个以上不同时间上的抽样.....(291)

三 简化与进一步发展.....(294)

习题7.....(296)

5. 整群抽样	
6. 二阶段抽样	
附录一	(297)
附录二	(303)
附录三	(310)
参考文献	(312)
附表	(313)

7. 回归抽样

8. 比合

9. 成数抽样

10. 双重抽样

11. 不等概率抽样

12. 非概率抽样

第一章 绪 论

§ 1.1 统计调查计划与抽样调查方案

现代社会从某种意义上讲，是一个信息社会。由于现代科学技术的高速发展，特别是以计算机为核心的信息处理技术的推广，使得社会各个领域的知识和技术更新速度加快，对各种信息需求量加大。信息资料的收集、开发和应用，对经济建设，对社会的发展和进步，起着越来越大的作用。信息本身就是一种财富，这一观点已被越来越多的人所认识。

一 统计调查计划

统计方法是取得信息的重要手段，统计资料就是信息的一个重要组成部分。统计资料的收集主要来自两个渠道：一个渠道通过对原有资料的汇总整理，得出所需要的资料。另一渠道而且主要是靠统计调查获得资料。当着手准备统计调查时，必须明确调查的目的，同时要考虑到收集什么样的资料，为什么要收集以及如何收集和利用这些资料。

统计调查计划就是根据调查目的和经费，设计调查项目和指标，确定调查总体、对象以及调查的方式，决定采取收集、汇总、整理和分析资料的方法，并指出误差和可靠性（信度）。所以在调查前必须明确以下几点：

1. 确定调查的总体，即调查对象的全体，确定调查项目和

指标。例如对某市居民的心态调查，要确定的总体是，调查对象的范围，要确定的调查的项目和指标是：对政策的看法，对廉政、社会风气的看法，对物价及社会变革的承受能力，对社会保障和公共设施以及教育事业的想法，影响心态变化的因素与年龄、文化、性别、收入、职业等有关，所以调查的项目必须包括这些项目和指标。

2. 如何调查收集所需要的资料，即确定调查方法。通常要考虑是全面调查，即普查；还是抽样调查。如果是抽样调查，进一步就要考虑如何选出最能代表总体的样本，即根据经费及要求的精度和信度，确定出抽样的样本额和抽取的方法，给出汇总、整理和分析资料的方法，根据数据处理的结果，给出统计推断和决策。

3. 调查表的设计，即问卷设计。设计统一的问卷，可以使调查内容标准化和系统化，便于收集、汇总和整理所需要的调查资料。问卷的设计，既要包括所要调查的内容，又要尽量简单，能用别的变量或标志得到的信息，绝不列入表中，措辞明确，回答不模棱两可，回避过敏性问题，不刺激回答者，不引起回答者反感或抵触。

问卷由以下几部分构成：

(1) 被调查者的基本情况。主要包括性别、年龄、民族、文化程度、职业、家庭人口等以便对资料进行分类和统计分析。

(2) 调查内容。这是问卷的主要组成部分，是通过若干问答的项目实现的，它要求既要简练，又要有一定的系统性和逻辑性。提问和回答要求准确，又要讲究艺术和趣味性。轻松容易回答的问题放在前，困难的问题放在后，切记问题不易过多。

(3) 问卷填写说明。这部分是填写问卷的要求和方法，包括目的要求、项目含义、调查时间、被调查者填写应注意事项，调查人员应遵守的事项等。

(4) 编号。问卷的问题要编号，以便分类和计算机录入与处

理。

问卷的形式主要有两种：

(5)卡片式：是把许多个调查单位和调查项目依次在一张问卷上登记调查；或把一个单位的所有调查项目登记在一张问卷上。

(6)开放式和封闭式：开放式问卷，是指调查的问题不列出所有的答案，而是由被调查者自由回答的问卷。封闭式问卷，是指在问卷中提出的问题，已拟定了各种可能的答案，被调查者只需从中选定答案。

问卷设计的程序：

(1)明确设计主题。根据调查目的，经过调查研究，突出问题，明确设计主题。

(2)设计问卷初稿。按照主题要求，确定问卷形式，列出调查项目，编写提问命题和填写说明，设计出问卷初稿。

(3)进行试验性调查，选少数单位或个人试填，分析是否有问题，是否便于回答，能否达到原设计要求。

(4)设计正式问卷。把试验性调查中出现的问题，进行整理和分析，对问卷初稿进行修改和补充，设计出正式的问卷。

4. 提出调查结果的精确度和可靠性（信度）。

二 全面调查与抽样调查

统计调查又分全面调查与非全面调查。全面调查是对调查对象全体进行逐个调查，又称为普查。

其优点是：

1. 能消除被调查对象的疑虑，一视同仁，有时可以得到被调查者的配合。

2. 能了解到总体的各种指标，不需要统计方法进行估计和预测。

全面调查的资料往往会给人产生一种信任感，觉得这样得

到的资料一定全面、准确、可靠。其实并不能一概而论，全面调查的资料往往受到各个方面或渠道的干扰，而使资料自相矛盾，明显失真。为了排除干扰，有时不得不派专门的调查人员深入基层，进行实地的抽样调查，依据抽样调查的资料，对总体进行估算。例如为了偷税漏税，为了争取政府补贴，有些单位的上报材料就有意失真。随着科技的迅速发展，技术更新和产品更新的速度加快，市场需求变化剧烈，要求收集信息的手段及时准确，就必须采用科学的抽样调查技术。

非全面调查是指从总体中抽取其一部分调查单位或对象进行调查的调查方法。用非全面调查的资料可以对总体进行估计和分析。非全面调查又分典型调查、目的调查和抽样调查。为了某种需要，有意识地选某些样本进行调查，通常称为目的调查；选一些典型代表进行调查称为典型调查。目的调查和典型调查对分析一些问题是有用的，但是它们的代表性、可靠性是很难客观评价的。人们习惯上称那种不是有目的选取样本的调查叫做**抽样调查**。抽样调查从抽样方式上分为**概率抽样**和**非概率抽样**。概率抽样是根据各调查对象在总体中的比例、作用、影响以及差异性大小，产生一个被抽取的比例即抽取概率所进行的抽样，这种抽样，以概率论为理论基础，可以进行各种统计分析、误差估计和可靠性分析。有时我们对总体中的调查单位一无所知，所以采用等可能的抽取，例如抽签、抓阄就是等可能抽样，即简单随机抽样。抽样调查应理解为概率抽样调查及一些近似为概率抽样的调查，绝不能把为了某种目的，有意选一些有利的样本进行的调查，冠以“抽样调查”。本书研究的抽样调查均指概率抽样的抽样调查。关于概率抽样的详细介绍见第二章 § 2.1 三。

三 抽样调查的优点

抽样调查是一种概率抽样，其最简单的是等可能抽样——简

单随机抽样。抽样调查的理论基础是概率论与数理统计。它是根据调查目的、调查费用、精度和信度设计的一种最能代表总体的抽样方式,设计一套估计、预测、数据处理和统计分析的方法,并给出估计和预测的精度和可靠性。它的优点是:

1. 收集资料快,能及时捕捉信息。在市场经济的形势下,及时得到市场信息,进行科学的决策是竞争取胜的关键。对于市场调查与预测,产品开发前景分析与论证,证券行情分析与预测,抽样调查都是其必不可少的调查手段。

2. 收集的资料准确可靠。抽样调查是根据用户提出的精度和信度设计调查方案,给出估计和预测的方法,提供估计和预测的精度和信度。由于资料的收集是在预先设计好的方案中进行的,所以资料准确可靠。

3. 抽样调查以概率论和数理统计为理论基础,它所得的资料可以用计算机实现各种数理统计方法的分析,便于深入地开发资料,得出比较全面和深刻的结论。通常可以作相关分析、方差分析、假设检验、因子分析、主成份分析、判别分析、聚类分析、层次分析和各种统计决策。

4. 抽样调查可以节省人力、物力、财力。用比较少的钱办比较多的事。由于调查人员少而精干,便于培训提高调查人员的素质,从而提高了调查质量。

5. 由于抽样调查建立在概率论的理论基础上,所以它可以对各种抽样方法和抽样结果进行比较科学的评价和分析,通过对各种方法的比较和分析,从中选出最优的调查方案。其它非概率抽样调查,不能给出精度和信度,没有比较客观的评价标准,有的估计和预测甚至提不出其相应的误差,所以只有参考意义。

四 抽样调查方案

由调查的目的,确定调查的项目和范围(总体),根据已掌