

抽样调查的理论 及其应用方法

Sampling Survey Theory
and its Application

谢邦昌 原著 张尧庭 董麓 改编

中国统计出版社

抽样调查的理论及其应用方法

Sampling Survey Theory and its Application

谢邦昌 原著

张尧庭 改编
董麓

中国统计出版社

(京) 新登字 041 号

图书在版编目 (CIP) 数据

抽样调查的理论及其应用方法/ 谢邦昌著, 张尧庭、董麓改编.
— 北京: 中国统计出版社, 1998.3
ISBN 7-5037-2602-4

I. 抽 ..

II. ① 谢 ... ② 张 ... ③ 董 ...

III. ① 抽样调查—理论② 抽样调查—方法

IV. C811

中国版本图书馆 CIP 数据核字 (97) 第 20380 号

著作权合同登记: 图字 01-97-1845 号

中国统计出版社出版

· (北京三里河丹坛南街 75 号 100826) ·

新华书店经销

科伦克三莱印务有限公司印刷

*

850×1168 毫米 32 开本 16 125 印张 40 万字

1998 年 3 月第 1 版 1998 年 3 月北京第 1 次印刷

印数: 1—4000

*

定价: 35.00 元

(版权所有 不得翻印)

序

抽样的教本，极是难写；抽样的工作，也极是难做。吾友张尧庭教授及董麓先生二人很热心将在下所撰写之“抽样调查的理论及其应用方法”加以改编，融入大陆之特色，做得有声有色。

抽样的教本，因善本已多，要写出有特色的教本，当然不易。事实上，在国内外可以找到的一些教本的一般特色，就是它们都没有什么特色。而本书最大的特色即为结合海峡两岸抽样调查之现况及理论基础，实务与理论之密切配合，突显出抽样调查之务实性及其广泛之应用性。

张尧庭教授及董麓先生除了保留在下原著之特色外，还加入了符合大陆现况之实际内容，使本书增色不少。

希望此次与张尧庭教授及董麓先生之合作能为两岸学术界之互动及交流，尽一绵薄之力。

谢邦昌 敬上

台湾 辅仁大学统计学系 系主任

1997年7月1日

改编序言

谢邦昌先生是台湾辅仁大学统计系主任，在台湾主持过不少大型的抽样调查，效果很好，用很少的样本达到很高的精度。近来，他写了一本《抽样调查的理论及其应用方法》，充分体现了他的特点，既有理论，又着重于实际调查的方法，很系统地反映了他的观点和经验。考虑到目前我们这一方面的实际工作开展的历史并不长，他写的这本书对推动我们的调查工作会有帮助；由于两岸所用的名词术语并不相同，叙述、讨论的方式也有差别，为了使大陆的读者更便于了解此书的内容，经谢邦昌先生本人同意，我们进行了改编。在改编过程中，我们尽量保留原著的特点，补充了一些切合我们实际的内容，以不致影响原著的特色。

与常见的抽样调查的书籍相比，原著有以下几个特点：

1. 不是只讲理论、公式，还着重于实施。如全书的最后几章，专门介绍表格如何设计，哪些缺点应该避免；访查人员应注意什么，如何培训，遇到拒访该怎么处理，甚至连服色、穿着要注意什么，也都逐一提醒。这些具体的经验是付出了大量辛苦才能得到的。为了说明如何组织实施大型调查，有一章专门介绍了台湾地区进行过的三项调查，让读者从中学到组织实施、资料分析比较的具体经验。

2. 讲理论、公式也是由浅入深，用实例来说明，在体系的编排上也与常见的书籍不同。例如，对每一种抽样方式，先讲抽一个单位的样本这一非常特殊的情况，它既好懂又便于列出全部

可能情况，讨论抽样的特性，所以这一讲法是很有特点的。有时在一单位抽样后接着讲二单位抽样，然后到 n 单位抽样的一般情况。这种由简入繁、由显入潜的方法使一些公式便于导出和理解。原著还引用了不少印度调查的数据、台湾地区的数据，来分析印证一些公式的正确性、适用范围，或作各种比较，生动而具体地说明了各种抽样方式的优点、缺点和实施中的问题。

除了以上这两条外，还值得一提的是原著的附录，附录不仅提供了素材，还提供了不少有用的信息。印度在抽样调查方面是很有成就的，一些著名的抽样调查专家大都是印度人，台湾地区经济发展的需要也大大推动了抽样调查的实施和研究，他们的这些经验，对于我们从计划经济向市场经济过渡的转变，无疑是有借鉴作用的。在这一点上我们希望本书的出版能贡献我们微薄的力量。

改编者 张尧庭 董麓

1997年6月26日

目 录

第一章 抽样调查的必要性	(1)
1.1 统计信息的重要性	(1)
1.2 资料的类型	(2)
1.3 普查	(3)
1.4 抽样的重要性	(4)
1.5 抽样与非抽样误差	(5)
1.6 成本问题	(6)
1.7 抽样调查与普查	(9)
第二章 定义与符号	(12)
2.1 总体和总体单位	(12)
2.2 样本、抽样单位和抽样框	(13)
2.3 概率抽样与非概率抽样	(14)
2.4 总体参数	(16)
2.5 误差	(20)
2.6 置信区间	(21)
2.7 交叉子样本	(22)
2.8 随机化的阶段	(22)
2.9 效率	(24)
2.10 成本	(25)
2.11 符号的说明	(25)
第三章 简单随机抽样	(29)
3.1 一单位抽样	(30)
3.2 二单位抽样	(31)

3.3	放回 n 单位抽样	(33)
3.4	\bar{Y} 的另一种估计量	(39)
3.5	二单位不放回抽样	(40)
3.6	不放回 n 单位抽样	(42)
3.7	总值的估计	(49)
3.8	比例的估计	(51)
3.9	置信区间	(54)
3.10	组合估计值	(57)
第四章 样本的抽取与样本量的估计		(61)
4.1	总体示例	(61)
4.2	变异系数的稳定性	(62)
4.3	样本量的作用	(64)
4.4	抽样误差与效率	(65)
4.5	抽样程序	(68)
4.6	抽样框的改变	(74)
4.7	随意抽样	(75)
4.8	样本量的决定	(77)
第五章 系统抽样		(96)
5.1	抽样程序	(96)
5.2	环状系统抽样	(101)
5.3	部分区间	(104)
5.4	系统抽样的实用意义	(104)
5.5	样本方差	(105)
5.6	系统法与简单随机抽样法的比较	(106)
5.7	样本方差的变化	(108)
5.8	样本方差的估计	(113)
5.9	直线趋势	(119)
5.10	周期变化	(127)
5.11	合适的排列	(127)
5.12	样本比例的分布	(128)
5.13	样本量的确定	(128)

5.14	二维系统抽样	(130)
第六章 不等概率抽样法		
6.1	单位大小的测度	(133)
6.2	一个单位的比例抽样	(133)
6.3	放回比例抽样	(135)
6.4	比例抽样法的效率	(136)
6.5	成本问题	(139)
6.6	经验研究	(141)
6.7	比例抽样法的优点	(144)
6.8	面积抽样	(145)
6.9	另一种估计量	(147)
6.10	比例抽样的程序	(148)
6.11	不放回抽样	(154)
6.12	调查的组合	(164)
6.13	比例抽样的推广	(166)
第七章 分层抽样		
7.1	为什么要分层	(171)
7.2	分层原理	(172)
7.3	分层变量与设计	(174)
7.4	样本量的分配	(175)
7.5	放回分层简单随机抽样法	(180)
7.6	不放回分层简单随机抽样法	(184)
7.7	比例的估计	(187)
7.8	分层比例抽样	(188)
7.9	示例	(192)
7.10	层的划分	(194)
7.11	层数的决定	(205)
7.12	互通子样本	(206)
7.13	复分层	(208)
7.14	事后分层的技术	(209)
7.15	控制选择	(210)

7.16	分层点的探讨	(213)
第八章	整群抽样	(217)
8.1	整群抽样的意义	(217)
8.2	等整群抽样	(218)
8.3	最佳群大小	(221)
8.4	比例的估计	(226)
8.5	群的大小不定	(227)
8.6	示例	(231)
第九章	多阶段抽样	(234)
9.1	抽样程序	(234)
9.2	估计和样本方差	(236)
9.3	简单随机抽样法二阶段抽样	(238)
9.4	放回比例抽样和不放回简单随机抽样	(244)
9.5	方差函数	(245)
9.6	成本函数	(249)
9.7	最佳 n 和 m 值	(250)
9.8	二阶段抽样的效率	(254)
9.9	示例	(258)
9.10	三阶段抽样设计	(259)
9.11	多主题调查	(261)
9.12	多级抽样	(263)
9.13	复合抽样设计	(265)
第十章	比率估计	(267)
10.1	比率估计的意义	(267)
10.2	比率估计量的偏差	(268)
10.3	均方误差	(272)
10.4	比率估计法	(273)
10.5	基本抽样方式	(275)
10.6	分层抽样	(279)
10.7	一个例子	(281)

10.8	乘积估计法	(283)
10.9	几乎无偏的比率估计量	(285)
10.10	无偏比率型估计量	(286)
10.11	无偏比率估计量	(289)
10.12	各种比率估计量	(291)
10.13	二级抽样	(297)
10.14	示例	(298)
第十一章 差与回归估计量		(300)
11.1	差的估计	(300)
11.2	回归估计	(301)
11.3	偏量和方差	(302)
11.4	简单随机抽样与分层简单随机抽样	(304)
11.5	二级抽样	(306)
11.6	接续抽样	(308)
11.7	相对变动的估计	(310)
11.8	多变量回归估计量	(312)
第十二章 自我加权设计		(315)
12.1	等权设计	(315)
12.2	分层单段抽样	(317)
12.3	分层二阶段抽样	(319)
12.4	几个例子	(321)
12.5	列表阶段自我加权设计	(323)
12.6	示例	(325)
12.7	样本观测值的重复	(328)
第十三章 非抽样误差		(330)
13.1	非抽样误差	(330)
13.2	非抽样误差的来源	(331)
13.3	非抽样误差的处理	(332)
13.4	非抽样偏差	(333)
13.5	非抽样方差	(334)

13.6	简单随机抽样	(335)
13.7	比例的估计	(338)
13.8	成本函数	(340)
13.9	不回答误差	(341)
13.10	误差的度量和控制	(348)
第十四章 抽样调查规划		(354)
14.1	讨论范围	(354)
14.2	资料需求问题	(355)
14.3	专案调查和经常性调查	(356)
14.4	资料搜集方法	(356)
14.5	问卷与填表	(358)
14.6	调查期、参考期和访问期	(359)
14.7	抽样框的问题	(359)
14.8	抽样设计的选择	(360)
14.9	试查	(360)
14.10	实地作业	(361)
14.11	调查资料的处理	(361)
14.12	报告准备	(363)
14.13	综合多主题调查	(364)
14.14	永久性调查机构	(365)
14.15	抽样调查的管理	(366)
第十五章 大范围抽样调查		(370)
15.1	人力资源调查规划	(370)
15.2	工商业调查规划	(377)
15.3	农林渔牧业调查规划	(384)
第十六章 抽样调查法		(397)
16.1	派员访查法	(397)
16.2	邮寄问卷调查法	(402)
16.3	电话调查法	(422)

第十七章 调查问卷设计	(428)
17.1 问卷的意义及功能	(428)
17.2 问卷设计的步骤	(429)
17.3 问项的种类	(431)
17.4 问卷设计的原则	(436)
17.5 问题排列的次序	(446)
附录一 随机数表	(448)
附录二 抽样调查国际互联网相关网址	(451)
附录三 学生身高体重的实际调查之应用	(493)

第一章 抽样调查的必要性

1.1 统计信息的重要性

进入二十世纪，科学和技术迅猛发展，人类的经济和社会生活以及其它活动都发生了巨大的变化。大规模的专业化生产、产品的标准化和劳务的综合利用，一切活动都以追求最大经济效率为目标。从事活动的规模越大，就越需要对发展计划进行周密的安排，而计划的设计、实施以及对未来成效的评估，不论工业、商业或政府活动，都不可缺少地要以客观资料（data）作为依据。各种决策性资料均需有健全的根据，因此就会用到各种统计信息。我们应该用科学的方法、客观的态度进行资料的收集、分析，加工，使之成为信息（information），从而具有使用价值。

统计资料的使用者大多为政府机构、工商企业、学术研究机构和国际机构。例如，政府要制定有关进出口贸易的政策，就需要各种产品的生产和消费资料；要制定工资或价格政策，就需要全国各地人民的家庭收支及家庭生活状况的资料。要了解和改善人民生活，就需要出生率、死亡率、人口增长率、人民受教育程度及生活状况等资料。又例如，在工商界，一方面需要成本、产量、产值、储量和供给状况的信息；另一方面需要消费者需求和爱好，消费行为及市场竞争、利润等信息。现代化有效的管理需要各阶段生产及销售活动方面的资料。

因此，为了执行种种任务，人们需要多方面信息，以作为决

策的参考。这些信息主要包括：国内及国际经济体系各种信息及变动资料，各民族及各地区分布与发展情况的信息，各种资源，如人力、耕地、森林、水资源、矿产、石油等。对于所收集的资料，应加以整理、保存，以备使用。

由以上叙述可知，在当今信息时代，统计信息扮演着重要的决策角色。没有完整正确的统计资料，就没有完善的决策。

1.2 资料的类型 (types of data)

用于计划设计及计划效率评估方面的资料大致可分两类：

① 已经存在但要经过观察或询问才能收集到的调查资料 (survey data)，② 用完善设计及控制而得到的实验资料 (experimental data)。调查资料又可以分为三小类：(1a) 在一固定时间或一段时间内的统计对象的数量特征及各种对象之间的关系，(1b) 统计对象及各对象之间的关系随时间变化的情况和趋势，(1c) 对时间相对稳定的资料。这几项资料可以简单的说明如下：

(1a) 这类资料大都和时间有关，例如各对象的数量特征和相关性等。举例而言，在资源方面有人口、不同利用方式的土地、农作物产量、工业品产量、工人、农民、医务人员、教师、律师、工程师人数等。彼此之间的关系如人口密度，单位面积农作物产量等；消费者需求及爱好，就业状况、教育程度、住宅条件、人口增长率、疾病发生率等。这类资料会随时间而改变、所以需要逐次收集。

(1b) 这类资料一般是对对象的变化及对象间的关系呈时间数列，要由定时收集的资料求出，可以用来预测未来的需要和资源。举例来说，分析社会经济因素间关系与人民生活状况，可以预测未来的需求（如对粮食、燃料、纺织品、建材、学校、医院等的需求）。

(1c) 这类资料一般不随时间改变。最常见的就是一国各地的地理和地质资料，如地形、气候条件、土壤类型、石油储量等。这种资料的收集需要的人力、财力较大，资料使用的时间较长。

上面所说的第②类资料只能由设计良好而经控制的统计实验获得。例如使某农作物产生最大产量的肥料用量，防治病虫害最有效的杀虫剂，使每单位成本产生最大产量的工业生产过程等都是实例。

本书后面几章将叙述与如何获得 (1a) 资料有关，同时也对 (1b) 资料提供相当重要的方法。这类资料多数和政府及工商业等有关，大致包含以下项目：

人口、出生、死亡、迁徙、收入与支出、就业与失业、资本形成、畜产、物价、建筑、居住条件、教育与保健统计、消费者爱好与需求、土地利用、灌溉、农业生产、耕作成本、生产、贸易、运输、百姓对政府现在或未来政策的意见和态度、消费者对市场产品或将上市的产品意见和态度。

1.3 普查 (census)

收集区域性或全国性的资料时，有时要对全部资料一一调查。这种单位如个人、住户、田块、工厂、商店等，研究对象的所有单位构成了该研究对象的总体 (population or universe)。收集调查总体内的每一个单位的统计调查方法就叫做普查。如果普查的范围较大 (如人口普查)，则需要花费大量的人力和财力，因此普查适用于必须收集每个单位资料的统计调查。普查必须弄清楚下列问题：

(1) 是要了解每一单位的情况，还是要了解全部单位的情况或某些个体的特性。

(2) 是否要获得完全的信息，并且不允许发生丝毫误差。

(3) 使用普查法是否能得到完全的信息。

(4) 在什么情况下，只抽查总体的一部分信息，可以代替普查。

第一个问题是指，如果研究必须涉及每个个体，就有必要进行普查。举例来说，所得税人人都要交纳，应征者人人都要体检。如果研究的目的是要了解大概情况，或某部分情况，就没有必要进行普查。

第二个问题要看研究者对资料正确程度的要求。一般来说对未来的规划都带有不确定因素，所以允许资料有一些误差，只要这些误差不会在决策中起决定作用。在许多社会经济问题中，还存在很多非统计性的不确定性，严格的规划通常不太实际。所以，实际工作中都允许存在误差，这种误差就是常说的允许误差 (permissible error)，请看下面例子：

比较各国或各地区人口时，常将单位记为千人或百人，而不是记到个位。例如印度 1951 年人口是 $36(10)^7$ ，日本 1950 年人口是八千三百万，而不是用 356 879 394 和 83 199 637 这样的细数。印度人口发布的允许误差为 $\pm 0.5(10)^7$ ，而日本是 ± 0.5 百万。

关于第三个问题，根据经验，普查也不一定能得到精确的数字。例如，印度 1951 年人口普查结果估计有 1.1% 的净低估 (Registrar General, 1953)，而美国 1950 年人口普查的净低估更达到 1.4% (Eckler 1953)。这就是说，普查只是得到一本总帐，但不能避免误差。大规模普查常常受度量及观察等因素的限制，而产生误差。法国在 1946 年做过一次城市工商业情况普查，结果竟低估约 24% (Chevry, 1949)，当然这属于极端的例子。

1.4 抽样的重要性 (need of sampling)

抽样调查的重要性主要体现在以下几个方面：