

抽样调查法



工 商 出 版 社





2 021 0137 7

抽 样 调 查 法

(日) 三 浦 由 己 著
井 出 満 译
李 士 瀛 译

工 商 出 版 社

抽 样 调 查 法

607-81/07 (日) 三浦由己 著
井出 满 译
李 士 瀛 译

*

工 商 出 版 社 出 版

北京三里河东路10号

北京市新华书店发行

北京市大郊亭印刷厂印刷

*

1982年12月第1版 1982年12月第1次印刷

787×1092 1/32 7.5印张 字数160,000

印数 00001—20,000

统一书号 4246·017 定价 0.85 元

出 版 說 明

本书根据日本统计协会编辑，一粒出版社1976年7月25日第1版第4次印刷的版本译出。为便于学习和使用，译者将原书卷末的公式集，作为附录分别插入有关各章。

这个译本在翻译过程中，承蒙国家统计局、北京财贸学院和北京市统计学会的支持，中国社会科学院世界经济与政治研究所郑光华同志校订并得到北京财贸学院怀军同志的帮助。

由于水平所限，译本中的不当之处在所难免，敬希读者批评指正。

工商出版社

1982年11月20日

积极学习和推广抽样调查

三浦由己和井出 满两位先生合写的《抽样调查法》一书，现在经北京财贸学院李士瀛同志翻译，由工商出版社出版发行。这个译本必将受到我国统计界广大读者的欢迎。

党的十一届三中全会以来，在党和政府的重视与关怀下，我国统计工作有了很大的恢复和发展，并在社会主义现代化建设中发挥了一定的作用。但就调查方法而言，我国的统计工作仍然是全面调查多于非全面调查。在非全面调查中，抽样调查方法用得更少。随着客观上对统计资料需要的增长，过多地布置全面调查，在人力、物力、财力上遇到的困难日益增大。这种情况迫切要求在可能范围内有效地运用抽样调查方法，及时地、节省地取得必要的全面统计数字。在这样情况下，这本书对于渴望学习、了解抽样调查方法理论和实践的广大读者会有一定的意义。

本书作者三浦由己和井出 满两位先生是日本总理府统计局的两位统计专家，在抽样调查方面既有理论修养，又富于实践经验。特别是井出 满先生曾于1980年秋应我国国家统计局的邀请，就抽样调查方法来华讲学，受到了听讲者的好评。井出满先生当时的讲课内容，主要取自本书。

关于抽样调查方法的著作，国内国外汗牛充栋。其在理论上的阐述，近年来虽有新的发展，但差异并不悬殊。而抽样调查方法在实践中的运用，则在不同国家不同时期，程度很不相同。如我国同日本比，日本在这方面的实践经验可能远较我国为多。本书正是作者根据日本统计工作的实践经验

写成的，目的是想让实际统计工作者和初学者了解抽样调查是什么，怎样设计，怎样使用抽样调查资料。因此，从理论上论证数学公式，不是本书的重点。本书对抽样调查的阐述深入浅出，通俗易懂，而且举例多取自实际统计工作。通过实际例子讲述抽样方法在统计工作中的应用，可能效果更好。

当然，作者在书中讲的抽样调查法的应用，都是根据日本统计工作的经验。这些经验不一定完全适用于其他国家，每一国家的统计学家和统计工作者都必须根据本国的具体情况考虑抽样调查方法的运用，不应生搬硬套。但这并不排除借鉴别国经验。

“他山之石，可以攻玉”。希望这本小册子同大家见面之后能起到它应起的作用。

铁大章

1982年国庆节

作 者 序 言

抽样调查一般都作为统计学或数理统计学的一部分予以说明，很少单独写成专著。

本书是以想要知道抽样调查是什么的人，或者计划进行抽样调查、利用抽样调查资料的实际工作者，不用太多的数学理论说明就可以看懂为目的而编写的。

因此，数学的证明不是本书的重点，只是为了参考，力求在读完全书以后，能对抽样调查法有所理解，并能根据需要使用有关方法。为了使用方便，特将有关“公式及其应用”汇总，作为附录放在卷末。

本书的目的虽如上述，但力不从心，可能事与愿违，付梓在即，不禁汗颜。

本书倘能对初学统计者理解抽样调查法有些帮助，笔者则将喜出望外。

最后，对在编写本书过程中从多方面给予鼓励的永山贞则前辈和一粒出版社的彦板竹男先生谨致谢忱。

三 浦 由 己
井 出 满

目 录

出版说明

积极学习和推广抽样调查

作者序言

第一章 绪论

- 1.1 统计调查的计划与抽样调查方案..... (1)
- 1.2 全面调查与抽样调查..... (2)
- 1.3 随机抽样与目的抽样..... (3)
- 1.4 抽样误差与非抽样误差..... (4)
- 1.5 什么是最佳抽样方案..... (7)

第二章 抽样调查法的基础知识

- 2.1 简单的实验例..... (8)
- 2.2 概率样本..... (10)
- 2.3 样本分布..... (11)
- 2.4 根据样本进行推算..... (17)
- 2.5 抽样误差..... (22)
- 2.6 抽取样本的一些方法..... (26)

第三章 单纯随机抽样法

- 3.1 抽取方法..... (29)
- 3.2 推算公式..... (31)
- 3.3 抽样误差..... (35)
- 3.4 确定样本数目..... (40)

附录 I 单纯随机抽样法的公式及其应用

第四章 整群抽样法

4.1 抽取方法.....	(52)
4.2 推算公式.....	(53)
4.3 抽样误差.....	(56)
4.4 确定样本数目.....	(60)
4.5 与单纯随机抽样法的比较.....	(61)

附录 I 整群抽样法的公式及其应用

第五章 分层抽样法

5.1 抽取方法.....	(78)
5.2 推算公式.....	(81)
5.3 抽样误差.....	(81)
5.4 样本的分配.....	(83)
5.5 确定样本数目.....	(91)
5.6 分层的效果.....	(92)

附录 II 分层抽样法的公式及其应用

第六章 两阶段抽样法

6.1 抽取方法.....	(111)
6.2 推算公式.....	(112)
6.3 抽样误差.....	(118)
6.4 样本的分配.....	(121)
6.5 确定样本数目.....	(126)

附录 IV 两阶段抽样法的公式及其应用

附录 V 分层两阶段抽样法的公式及其应用

第七章 概率比例抽样法

- 7.1 抽取方法..... (156)
- 7.2 推算公式..... (158)
- 7.3 抽样误差..... (159)
- 7.4 与等概率抽样法的比较..... (161)

附录Ⅴ 概率比例抽样法（允许整群重复抽取）的公式及其应用

第八章 比率推算和回归推算

- 8.1 推算方法..... (167)
- 8.2 偏差..... (169)
- 8.3 抽样误差..... (171)
- 8.4 与单纯推算的比较..... (172)

附录Ⅵ 比率推算的公式及其应用

第九章 其 他

- 9.1 部分总体的推算..... (209)
- 9.2 二相抽样法或二重抽样法..... (212)
- 9.3 继续进行抽样调查..... (217)
- 9.4 用辅助样本推算的抽样误差..... (222)
- 9.5 由于拒不回答而引起的偏差..... (225)

第一章 绪 论

1.1 统计调查的计划与抽样调查方案

当准备进行统计调查时，必须先明确调查的目的，同时充分考虑搜集什么样的资料，为什么搜集以及如何使用这种资料。当然，也要考虑搜集这种资料是否必须进行实际调查，根据其他统计调查或现有业务资料能否得到。即使搜集这种资料的费用并不很多，但在保证精确度的前提下，如果可以从次一级或其他统计资料中整理出来，就没有重新进行实际调查的必要。

经过各种考虑以后，如果结论是必须进行实际调查，那么，就考虑花钱要尽量少，效果要尽量好。为此，在拟订调查计划时，必须明确下述各点：

(1) 确定调查的总体，确定调查的资料，即规定全及总体与有关样本数量。例如，调查全国15岁以上有劳动能力者的就业状况，调查全国制造业工人的工资分布状况，调查某地区具有某种属性的每户家庭平均收入等等。在确定调查对象总体的同时，也要规定调查资料的数量。

(2) 如何调查所需资料，即确定调查方法和汇总资料的方法。根据搜集和汇总资料的方法确定样本数目。

(3) 如何分析和使用调查所得的资料。

(4) 必要的精确度。

调查所得结果，一般与调查对象总体的实际数值有误差。这种误差在什么范围内才能不影响调查目的，应规定误

差的限度。

一般地说，如果用抽样调查法搜集某一全及总体的统计资料，就可以用很少的费用得到必要的精确度的结果。至于用什么方法、抽取多少样本、如何汇集根据抽样所得的结果、是否推算总体的推算值、推算值的误差限度如何等等，都属于抽样调查的理论。

1.2 全面调查与抽样调查

进行统计调查搜集资料，究竟用全面调查为好，还是抽取调查对象的一部分，进行抽样调查为好，要从各种角度比较两者的利弊再作判断。

一般地说，抽样调查比全面调查有下述优点：

(1) 经济。全面调查要调查全部对象，费 力 大、花钱多。抽样调查只抽取一部分对象，省力、也省钱。

(2) 可以提早汇总调查结果，及时利用调查结果。

(3) 便于对只能用抽样的对象进行调查。例如，对农作物的收获量预测调查，就不能在收获前，把农作物预先收割进行调查。

(4) 可以保证调查结果的质量与准确性。如果从事调查的人员少，可以集中水平高的调查人员进行全面训练，以便提高调查质量和效果。

至于全面调查的优点可列举如下：

(1) 能得到较小单位的全面情况。如要取得较小地区的资料 and 把全区细分各组的资料，由于抽样调查的精确度较低，一般都用全面调查。

(2) 容易得到调查对象的协助。全面调查对全部对象都

进行调查，一视同仁。抽样调查则不然，往往引起调查对象的疑虑，如说：“为什么不调查他而调查我？”。

根据调查的性质，有时不可能进行全面调查，在这种情况下，就不得不借助于抽样调查。但如认为抽样调查可以代替全面调查，也是不妥当的。即使可以进行全面调查，在抽样调查与全面调查任选时，既要求迅速、经济还要保持资料的精确度。这样就必须考虑哪种调查计划最好。经过研究，有时抽样调查最好，有时全面调查最好。

1.3 随机抽样与目的抽样

调查对象的总体，即由其搜集资料的总体，在抽样调查中称为全及总体。抽样调查就是从构成全及总体的单位中抽取一部分单位作为样本，根据所抽样本进行调查的结果，获得有关全及总体的知识。

从全及总体抽取样本有两种方法：一种是按照随机原则、利用随机数，从构成全及总体的各个单位中抽取样本的方法，叫做随机抽样；另一种是有意识地选择若干具有代表性的单位为样本进行调查的方法，叫做目的抽样。

目的抽样一般由专家考虑抽取某些有“代表性的”或“典型的”单位为样本。各个单位有没有代表性？能否被抽取为样本？由专家个人判断决定。因此，各单位被抽取为样本的可能性并不一样，其概率也不清楚，而且由样本获得调查结果的可靠性亦由专家来判断、评价，这不是客观的测定方法。所以目的抽样的最大缺点就是对调查结果的可靠性不能客观地评价。

与目的抽样相反，随机抽样无论反复抽取多少次，按照

规定的抽样方法所得推算值的偏差程度，借助于概率论都能了解。采用这种方法，根据调查的结果，就可以客观地计算调查结果的可靠程度。这是随机抽样的特点。如果用不同的调查方案分别调查，就可以相互比较调查结果的精确程度，据以考虑效果最好的调查方案。

能客观地评价调查结果的精确度，是随机抽样的最大优点，并且可以按照要求的精确程度考虑样本数目和抽样方法。至于目的抽样的调查结果是否具有符合调查目的可以利用的精确度，在客观上不能保证。所以，准备进行统计调查时，应充分理解随机抽样的意义，最好采用随机抽样法。

如同准备大规模调查而进行的试点一样，在各种条件下对居民区、商业区、工业区、农村等拟订调查方案时，按照随机抽样法确定地区，不如根据调查目的有意识地选择具有各种特征的地区。为了特殊目的，用目的抽样比随机抽样有时更合适。一般的统计调查，还是以采用随机抽样调查为宜。

1.4 抽样误差与非抽样误差

统计调查的结果常常出现各种误差。抽样调查因只调查构成全及总体的一部分单位，根据调查结果推算全及总体的一般情况，所以推算值与实际调查全部单位所得结果不同。抽样调查的推算值与调查全及总体全部单位所得结果之差，即为抽样误差。因此，抽样误差是因只调查全及总体的一部分单位产生的，而调查全及总体的全部单位的全面调查就没有抽样误差。

由于其他原因而产生的误差叫做非抽样误差。非抽样误差种类很多，仅将主要的列举如下：（1）因调查计划不周

而产生的误差；（2）划分调查对象范围的误差；（3）调查中或回答中发生的误差；（4）汇总或计算时发生的误差。

（1）调查计划不周所产生的误差

由于调查计划不周，即使完全按照计划进行调查，其结果也不能完全切合调查目的。例如调查所用的概念与调查目的要求有出入，因而调查结果与调查目的也会不一致。

（2）划分调查对象范围的误差

在统计调查中无论抽样调查或全面调查，都必须明确规定调查对象的总体。从全及总体抽取调查对象时使用某些形式的调查表格（在抽样调查中称为范围[frame]），由于这个范围不完全，则构成全及总体的各个单位，如有的未包括在表格中，就绝对没有被抽取的机会。例如利用在调查时点前拟订的调查表作为取样范围，从编成调查表到调查时点这段期间内调查单位的变动情况，如未采取适当方法补救，样本就会选偏，因此调查表一经编成必须及时利用。这一点是很重要的。

划分调查对象范围的误差，除因范围不完全引起的以外，还有在调查阶段发生的误差。例如人口普查，本来规定在某一时刻全部无遗地调查全国人口，但实际上常有遗漏或重复的错误，特别是青年人流动性大，难免发生某种程度的漏查，这种倾向在以人口为对象的抽样调查中也不例外。

至于以企业基层单位“事业所”为对象进行调查时，虽然明确规定以属于生产企业的“事业所”为对象，但在实际调查阶段常常遇到性质尚未确定应否作为调查对象的问题。由于调查对象的定义不明确而发生遗漏或重复也屡见不鲜。

（3）调查误差或回答误差

回答误差有两种：一种是报告单位（人）故意不如实回答

而形成的误差,另一种不是故意的,而是由于回答错误造成的误差。

报告单位(人)故意不如实回答的事例,在调查经济情况时常常遇到。在家计调查中,一般有把收入额报得过少的倾向。在以“事业所”为对象的商业调查中,一般都倾向于把销售额和收入等缩小。在人口调查时也有假报的倾向。在调查最后学历或职业时,报告单位(人)往往又有夸大的倾向。

与此相反,有的由于不注意而漏写回答,也有回答错误的,有弄错提问意义,也有不知如何回答而发生的误差。

(4) 汇总时的误差

在汇总资料过程中发生错误的可能性也是常有的,如注错符号、卡片穿孔错误、机械使用错误和计算错误等。即使数字计算无误,在印刷时也会造成错误,发生误差。

上述各种非抽样误差并非抽样调查所特有,在全面调查中同样存在。就是说在抽样调查中有抽样误差与非抽样误差两种,在全面调查中,只有非抽样误差。

在考虑调查结果的正确性问题时,如果只考虑抽样误差是不妥当的,应该综合抽样误差与非抽样误差一併考虑。关于抽样误差,由于应用概率论随机抽取样本,所以抽样误差的限度可以推算出来。

至于非抽样误差,目前还没有一个知道误差范围的方法。一般用调查结果与另一种调查方法得到的同类资料比较,或者采用事后复查方法来评价调查误差。尤其在大规模调查中,一般都把事后调查作为调查体系的一个环节。

在制订统计调查方案时,为减少非抽样误差,重要的是尽量采用减少调查各阶段误差的方法。诸如研究调查表的格式,

推敲提问的措词，严格训练调查人员等都可以减少回答误差。如果加强各阶段的工作管理，汇总的误差也是可以减少的。

1.5 什么是最佳抽样方案

抽样方案是统计调查方案的一部分。当确定调查目的、调查方法、调查组织时，应考虑下述五个重要因素：(1) 自全及总体抽取样本时，用什么样的调查表较好；(2) 用什么方法抽取样本；(3) 抽取多少样本；(4) 根据抽样调查结果用什么算式推算全及总体的数值；(5) 确定用什么算式计算推算值的精确度。应该指出，抽样方案并不是仅仅确定抽样方法。

那么，什么是最佳抽样方案？当有两个方案可供研究时，说这一方案比另一方案好，究竟好在哪里？

不能完全达到调查目的要求的精确度的抽样方案，不能称为好方案。抽样的样本范围如果过小，利用有限的调查结果进行分析判断或作决定时，势必影响精确度，甚至判断错误。

反之，如果超过调查目的所要求的精确度，势必花费太大，也不是好方案。

总之，精确度和调查费用是判断抽样方案好坏的两个重要因素。当然，针对实际情况，还应考虑调查方法、调查组织等各种条件进行综合判断。但最基本的是兼顾精确度和费用两个因素。因此，本书根据下述原则判断抽样方案的优劣，最佳的方案是：

- (1) 在核定费用范围内达到最高精确度；
- (2) 以最少的费用达到要求的精确度。