


印度馬哈拉諾比斯教授
訪華統計報告集

統計出版社



印度馬哈拉諾比斯教授
訪華統計報告集

(內部發行)

統 計 出 版 社

1958年5月

印度馬哈拉諾比斯教授：
訪華統計報告集
(內部發行)

*

統計出版社編輯、出版
(北京復興門外三里河)

北京市書刊出版業營業許可証出字第076号

國家統計局印刷廠印刷
新華書店北京分店內部發行組發行

*

787×1092^{1/32}·3^{11/16}印張·67,000字

1958年8月第1版

1958年8月第2次印刷

印數：1,301—2,350

書号：3006.112

定價：0.40 元

編輯說明

印度統計學院院長、印度政府統計顧問、联合国統計委員會主席 P·C·馬哈拉諾比斯教授，应我国國家統計局、中国人民大学和中国科学院經濟研究所的邀請，于1957年6月19日至7月12日在我国作了为期二十四日的講学訪問。

馬哈拉諾比斯教授在訪問我国期間，曾在國家統計局、中国人民大学、北京大學和中国科学院作了关于抽樣調查和数理統計的报告，并就有关抽樣調查的問題举行了几次座談。这本統計报告集，就是將馬哈拉諾比斯教授的这些报告和談話的记录或提綱收集起来編成的。

我国是一个社会主义国家，国营經濟是我国社会主义經濟的主体。国家要对社会主义經濟实行計劃管理，就需要經常地全面地了解社会主义經濟尤其是国营經濟的活动情况。因此，我国自从創立統計工作以来，就把全面調查作为对国民經濟进行統計观察的主要方法，把定期报表作为調查統計的最重要武器。这在过去和今后，都是正确的和必要的。

我国又是一个地大、人多、經濟还比較落后、經濟情况又相当复杂的国家。只靠全面調查，还不能滿足各方面的需要。經驗証明，統計的武器要多，除全面調查外，我們还必须采取各种非全面調查的方法。而在非全面調查中，抽樣調查是最完善、最有科学根据的方法。因此，抽

样調查在我国也是統計观察的一項重要武器。

印度应用抽样調查的規模很广，对抽样技术和調查方法都有所發展。馬哈拉諾比斯教授对抽样調查法素有研究，同时，他对我国的定期报表制度有很好的評价，对在我国应如何采用抽样調查十分注意。因此，这本統計报告集將对我们研究和运用抽样調查有所帮助。其中对有些具体問題的看法或有与我们不尽一致之处，也可以引起我們进一步的研究。

这些报告和談話，是由国家統計局、人民大学的一些同志譯成中文的。本社編輯部曾就譯文加以編輯整理。編竣后未經馬哈拉諾比斯教授看过。

和馬哈拉諾比斯教授同来我国訪問的拉希瑞教授，也参加了訪問期間举行的报告和座談。他所作的有关抽样技术的兩篇报告，也作为附录收入本集。

統計出版社編輯部

1958年2月

目 錄

抽樣調查的一般原則	1
有意抽樣和隨機抽樣	7
關於職工和農民家庭收支調查問題	13
誤差的計算和控制	22
根據印度的經驗，談談抽樣調查的幾個問題	27
(一) 講演提綱	27
(二) 講演記錄	33
(三) 問題解答	37
以交插抽樣控制抽樣調查的問題	44
(一) 講演提綱	44
(二) 講演記錄	54
(三) 問題解答	62
數理統計的發展和在社會科學上的應用 (講演提綱)	64

附 錄

一、拉希瑞教授在座談抽樣調查問題時的發言	84
二、拉希瑞教授關於“隨機抽樣的技术 (方法)”問題的講話	90
(一) 講話提綱	90
(二) 講話記錄	103

抽样調查的一般原則

(1957年6月24日)

这是 P·C·馬哈拉諾比斯教授来华訪問期間的第一次談話，参加这次談話的有：馬哈拉諾比斯教授，拉希瑞教授，国家統計局薛暮桥局長和王思华副局長。在薛暮桥局長就国家統計局成立以來所进行的工作以及我国統計工作的一般情况作了介紹之后，馬哈拉諾比斯教授开始了下面的談話。——編者

本人感謝薛局長对中国統計工作的确切而明智的說明，使我很清楚地了解了中国的統計工作。

我所准备要談的話，可以分三个步驟进行。在第一个阶段获得你們同意之后，才进行下一步驟。如果第一、二兩阶段的談話都获得同意，我就准备取消到中国各地觀光的日程，同时緩期离此赴香港，以便与貴局合作，商討抽样調查的具体进行方法。

談話步驟

我准备的講演和談話的三个步驟是：(1)一般原則的介紹；(2)如果同意我所陈述的一般原則，就进一步較詳盡的討論这些原則；(3)大家又都取得同意，我將以抽样調查工作上最重要的一方面，即“抽样調查的設計”进行研討。

今天的談話，只限于第(1)項中的一些論點。

一 般 原 則

一个集中形式的機構(如一个工厂)，在各方面都有詳細的帳目記錄可查，在进行統計工作时，使用全面調查是可能的。但是，如果要調查的單位數目很多，地区分布很广，又沒有帳目記錄(如农户)，那末就應該用抽样調查搜集統計資料。因为在这种情况之下，抽样調查比較快，比較省錢，結果比較准确，而且又可以計算調查結果的誤差的限度。

例如，在中国的6万多个工矿企業單位中，使用全面調查是适合的，因为这些單位都有帳目記錄。但即使这样，以抽样調查来檢查工人數目或其他指标的数值的准确性、全面性等，或許还是必要的。

不論全面調查或抽样調查，都存在兩種誤差。第一种誤差是物理變異誤差。由于被調查对象随時間不同而有所變異，因之調查數字只能是被調查对象在調查阶段的平均數，而这个平均數决不会同于該阶段被調查对象物質本身的具体波動(物理變異，例如，谷物的重量受空气中湿度的影响)。当这种“誤差”仅有1%或2%时，可以不必管它。第二种誤差是由于人的因素所引起的調查誤差(例如，計算錯誤、謊報數字、对指标誤解或曲解以及在分析阶段所發生的錯誤等)。調查量愈大，調查誤差愈大；由于抽样調查比全面調查的工作量小，其調查結果的質量就高得多。另一方面，如样本加大，抽样誤差就会随之减小。在一般情况下，抽样誤差的减小率，总是大大低于調查誤差的增加率(当样本極小时，这种对比增減率的地位

就适得其反；但在这种情况下，由于抽样误差太大，调查结果所得数字，一般没有什么用处）。因此，总会到达这样一个阶段，这时样本的大小，正好使抽样误差小于全面调查的调查误差，一般地说，这个阶段的到达，总比调查误差“等于”抽样误差的阶段早得多。在抽样误差和调查误差大体“相等”的情况下，抽样调查的结果确较全面调查的结果好。

在经济落后的国家中，调查误差一般是比较大的。文盲较多是一个因素；交通不便也是一个因素。比如，交通不便时，渎职的调查员可能故意的漏掉一些调查对象。在这种情况下，全面调查实际上并不全面。

农业，家庭收支，人口等统计，用抽样调查会得到较准确的结果。国家统计局应当有自己控制的抽样调查工作人员，散布全国。当调查对象分散而数目众多或缺乏可靠记录时，国家统计局都应有自己直接控制的工作人员进行抽样调查。

根据印度的经验，如果只是要求所搜集的资料能够代表全国的情况，那末，1,200人就够了。但是，如果要求资料有分省的细数，那末，起码需要2,500人。至于最后所需人数的多寡，就得看统计项目的繁简如何而定。

抽样调查与全面调查

美国1950年的人口普查，耗资1.2亿美元，但经过抽样调查方法检查的结果，发现仍漏掉了约300万人口。

在苏联，以全面调查方法所取得的农作物产量数字，我所听说，还没有一定的把握。很多在苏联负责统计工作的人，都是我的好朋友，他们曾经来过印度，并住在我家作

容，因此我們有机会坦白的交換意見。我現在的談話，也同样根据这种直率坦白的精神。苏联在1928或1929年，就开始使用抽样調查。很自然的，在当时所用的方法是“有意抽样”，因为在那时，这是人們所知道的最好的方法。“随机抽样”在世界各国的大規模使用，只不过是近十年至十五年的事情。因之，我們發現，在苏联的新起部門中，如許多工業部門中，他們使用的抽样法，和印度及西方国家所用的毫无兩样。

在印度，全面調查在原則上仅用于农作物播种面积的統計。从事調查的人員，約有5万之多。用抽样調查复核的結果，証明这些数字的誤差限度非常大。

这些事实說明独立的复查是必要的。在一个大国里，在农業中用抽样調查是必要的。

随机抽样与有意抽样

抽样調查的方法有二，即随机与有意抽样。

早在1925或1926年，英国統計学者A.L.波里在国际統計学会作过一次关于抽样法的报告，在这个报告中，他提到随机与有意抽样。在1930年，意大利統計学者卡瑞多·金里又作了一次关于有意抽样实验結果的报告，这次报告的根据是1927或1928年意大利利用有意抽样所进行的人口調查的結果。他的报告，說明了有意抽样是不能令人滿意的。

印度在1937年开始使用随机抽样法，美国在1929年經濟蕭条后，也于1936年左右开始用随机抽样来調查失業人数。从那时以后，美国使用随机抽样于工業、商業及劳动

* 原文为 Purposive Sampling——編者注。

統計方面有長足發展；而印度則多着重于抽样調查在食物、农業、人口、生活条件等統計方面的应用。

联合国抽样調查小組委员会討論了五年（1947—1951年）之久，我自己也参加了这些討論，結論認為随机抽样是比較优越的方法。

我曾就这个問題和苏联專家們討論过。

关于研究機構

統計研究学院直接隸屬於国家統計機構，本身并没有什么不好。重要之点，在于研究機構欲發揮其力量，必須具备兩個条件，第一是必須有独立的学术思想，第二是必須和实地工作有密切的联系，使其研究方向以解决实际問題为主。

現代統計工作，不仅需要能干的統計学者，而且需要数学与工程人才及其他科学家。因之，一个統計研究機構，必須和其他科学团体有密切联系。当然，統計主要是用于社会經濟調查，所以，这个研究機構必須和国家統計機構有密切联系，但它同时也必須和其他科学研究机关保持联系。在这一点上，印度对統計学的看法，是广义的看法。这就是說，我們相信統計工作虽然以社会經濟調查为主，但其范围决不仅此。生物学、工程学以及許多其他科学也需要統計。

这个研究機構可以不时担負一些日常的統計工作，但絕對不能要求它用全部的时间做这种工作。表示研究機構独立性的标志之一，就是研究工作虽是以解决实际問題为目的，但并不要求它任何时候都非如此不可。这个研究機構进行研究的題目，也許有很多对任何实际問題的解决并

未能立即有所貢獻，但仍然應該讓這些研究工作者有效地進行他們的研究，對他們研究的方法也不應予以干涉。

一個研究機構同時也應該訓練人才，但它不應擔負訓練大批中初級統計人員的責任，而應專門致力於統計教師和研究人員的訓練。訓練人才本身也就是一門科學。

一個統計研究機構在性質上是和一所醫學院一樣的。一所醫學院如果沒有醫院的聯繫，提供學生以實習的機會，是不能想像的。同樣的道理，一個統計研究機構如果不和實際工作機關聯繫，而想給學生以正確適當的訓練，也是不可能的。總之，一個統計研究院一方面要和實際統計工作有密切的聯繫，同時也應保持其在基本研究工作中的獨立性。

有意抽樣和隨機抽樣

(談話記錄，1957年6月25日)

王思華副局長：馬哈拉諾比斯教授昨天談到，蘇聯在1930年左右，曾應用“有意抽樣”，請明確其具體內容。

它與“隨機抽樣”有什麼不同？

馬哈拉諾比斯教授：抽樣調查方法是從1920年開始大規模應用的；在1920年以前只有零碎的应用。自然，普通人買水果時，實際上也應用抽樣的方法，但我們這兒並不是指這種性質的抽樣。1926年，A·L·波里教授在國際統計學會作了一個科學報告，第一次討論了抽樣的方法。他在这次報告中，談到隨機抽樣和有意抽樣。這時候，許多統計學家對抽樣方法已開始有了一些模糊的概念。蘇聯於1927年在涅姆其諾夫院士（他現在已有七十多歲了）指導下，在農業生產方面做了許多抽樣調查工作。1930年以後，涅姆其諾夫院士對於抽樣調查的方法已較有把握，並將它應用在其他方面。但他所應用的主要是“有意抽樣”（或稱“代表性的抽樣”）。這種所謂代表性的抽樣方法，印度在1920年左右的作物調查中也曾應用過，即選擇一些有代表性的村子，根據它們的數字來推算總體。我是相信機率而不相信‘代表性’的。美國的“定額抽樣”方法，也相當於代表性的抽樣。定額抽樣方法的基礎是以主觀推測的總體組成來規定樣本的組成，這是很主觀的，因為首先要知道總體的組成後，才能確定樣本的組成。美國蓋洛

普民意測驗，就是应用定額抽样方法的，但它在1948年預言总统竞选結果时是完全失败了。美国政府举行調查时常用随机抽样；而私人企業举行調查时，則用定額抽样，因为定額抽样可以少花錢。苏联在工業产品質量的檢查方面完全用随机抽样的方法，1954年曾有一个报告。苏联所应用的这种方法和美国的質量控制檢查方法是差不多的，只是更簡單化一些（簡單的部分是全距）。美国現在每年有失業就業的調查和工業調查，每十年有一次人口調查。并用抽样的方法来檢查調查的結果。檢查共二次，如1950年美国人口普查后，就派了最优秀的統計人員采用抽样的方法去檢查普查的結果；又派了更高級的統計人員采用抽样的方法去檢查上述抽样檢查的結果。在檢查結果中，發現1950年的全面普查漏掉了300万人。英国在农業調查中也应用随机抽样的方法。瑞士在农業和林業的調查中也完全应用随机抽样方法。日本在战后也采用随机抽样的方法作了許多調查。所有这些国家进行抽样調查大都是最近十年的事。美国是1936年开始的，英国是1930年开始的，印度是1935年开始的。1947年联合国开始設立抽样調查小組，十年来一共开了五次会。第一次會議專門討論術語的标准化問題，这一工作現已基本上完成。我在1942年写了一篇关于抽样調查方法的論文，1944年在英国發表，文中举孟加拉省农業調查为例来討論抽样調查方法。四、五个月以前，国际統計学会出版的統計術語辭典，才对抽样調查的涵义作了解釋，并第一次介紹了交插抽样的概念。这一辭典以英、法、德、意四国文字印行。总的說来，現代的抽样調查是統計方面很新的發展。从1930年开始，苏联的統計工作比其他任何国家都做得先进，苏联的国民收

入統計和農業統計都是很有基礎的。蘇聯的統計組織是最集中、最大規模的，這點使我十分欽佩。但是，蘇聯在農業和家庭收支調查方面所應用的抽樣方法卻已落后了。我曾在抽樣調查方面和蘇聯的統計學家們交換過許多意見，據我所知，蘇聯的許多數理統計學家，如皮撒列夫等都曾主張採用隨機抽樣方法。1955年初，在蘇聯中央統計局方法處的一個報告中，曾提出了抽樣調查方法特別是在家庭收支調查中應用抽樣調查的問題。我相信，他們需要一段醞釀的時間，但如一經決定採用隨機抽樣方法，那末改變一定會很快的。至於蘇聯最近的情況如何，我還不太清楚。

現在我進一步解釋定額抽樣和分層隨機抽樣。如果事前對於客觀的事物沒有一定的知識就進行分層，那末這種分層是主觀的。如果事先已掌握了一定的知識再進行分層，那末這種分層是客觀的。層不是由調查員到目的地才分，而是事先確定的。分層後不是有意抽樣，而是隨機抽樣。分層有意抽樣與分層隨機抽樣的區別在於：分層有意抽樣是主觀揣測，分層隨機抽樣是客觀的。所謂分層，應有一定的標準的比例數字，否則就不能分。比方現在有許多民族的人士到北京來開會，我們如果要想作幾個民族的調查，這時只能根據各民族人數的比例作分層抽樣調查；如果不知道這一比例數字，那就不可能分層，即使是分層了，也是主觀揣測。再舉個例子，如果某一集會中有1萬人，我們想去調查他們的年齡、高度分配等。參加這一集會的有工人、農民、工程師、機關職員等，那麼應按什麼比例來抽樣呢？如果採用定額抽樣的方法，按照人口普查中這些人的比例去抽選樣本，顯然是錯的。如果調查員站

在大會的門口，每進去一個人就問一下他們是幹什麼的，這樣掌握了集會中工人、農民、工程師、機關職員等的比例數字，然後按這比例去抽樣調查，那是對的。至於隨機抽樣調查，則是事先不假定有這種比例，由調查員跑進去隨機地抽，抽出後再發現這個比例。如果你事先掌握了正確的定義，又知道了正確的比例，按照這種比例去抽樣，其結果和隨機抽樣的結果是會相同的。抽樣時，無論如何不能含有主觀因素，隨機抽樣就是要客觀地去進行。根據印度的經驗，任何一次抽樣調查，如果發現有主觀成份，就一定會發生偏差。印度測量小麥產量時，用一個圓圈進行實割實測，就是由調查員拿着圓圈跑到田里去丟，丟到那裏就測到那裏。如果調查員开着眼睛去丟圓圈，由於調查員潛意識的作用，他往往把圈子丟到吸引他的地方，結果圈子里的產量常常是比圈子外邊的產量多。如果調查員閉了眼睛去丟圈子，這種偏差就不會發生了。有趣的是，圈小會發生較多的偏差，圈大些發生的偏差少些。我有一篇歷年數字比較的論文談到這個問題，過去我們用方形的圈子會發生偏差，現在我們改用圓圈了。先用小圈確定地位，再在外面畫一大圓圈，割大圓圈內的作物。

王思華副局長：中國各縣的領導者，對農村情況較熟悉，都希望採用典型抽樣，亦即有意抽樣，對隨機抽樣在思想上也搞不通。如果對實際情況較為了解時，是否用“有意抽樣”較好？

馬哈拉諾比斯教授：如果對實際情況知道得很多，採用隨機抽樣也沒有損失；隨機抽樣能確定誤差，而有意抽樣則不能確定誤差。

隨機抽樣有二個優點：（1）比較簡單快速，（2）

有理論上邏輯上的根據。

有意抽樣只能根據一、二個指標來研究和確定其代表性，如果研究的指標較多，則代表性就難以確定了；而隨機抽樣則不受指標數目的限制。

孫治方副局長：純粹隨機抽樣好，還是分層隨機抽樣好？

馬哈拉諾比斯教授：如果分層不是分得太多，則分層隨機抽樣較好；如果分得過多，就失去分層的意义了。

我們進行抽樣調查時，如果同時採用兩種辦法，一是隨機的；一是有意的。如果二者的結果相同，那末兩種辦法都是正確的。按照機率理論，隨機抽樣的方法是可靠的。如果二者的結果不同，那末一定是有意的抽樣搞錯了。有意抽樣決不會比隨機抽樣好。

如果分層，每層分得很多很細，那麼用有意抽樣還可以。但假定從一萬個合作社中，每社選出一個代表戶，共計一萬戶，這一萬戶實際上是不能代表一萬個合作社總體的。因為從每個社中選出的代表戶可能是中等的戶，沒有最窮的，也沒有最富的。但如用隨機抽樣法，則最窮的最富的戶都可能抽到，都可以知道。又如在一个學校中每班抽出一個作為代表的學生，這個學生是按每班的平均成績抽出的，他並不能代表學校中學生成績的真實情形（真實的情形應是：有的高於平均成績，有的低於平均成績）。隨機的樣本則代表的不是平均成績，而是有高的和低的成績。這種隨機樣本是總體的縮影，而不是平均數。而且有意抽樣只能根據成績一個指標來確定作為代表的學生，縱使樣本在學業成績上能代表全校學生，但在身高、體重等方面就不能代表。至於隨機樣本，則不論學業成績、身高、體