

抽樣觀察

В. И. 格蘭科夫著

統計出版社

抽 樣 觀 察

B. II. 格 蘭 科 夫 著

張文華譯 王宗麟校

統 計 出 版 社

1957年·北 京

В. П. ГРАНКОВ
ВЫБОРОЧНОЕ
НАБЛЮДЕНИЕ

Государственное статистическое издательство
Москва 1955

抽 样 观 察

В. П. 格蘭科夫著
張文华譯 王宗麟校

*

統計出版社出版

(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075号

西四印刷厂印刷

新华書店發行

*

書号: 3006, 59.787X1062 页1/32. 3 $\frac{1}{4}$ 印張. 64,000字

1957年2月第1版

1957年2月第1次印刷

印數: 3,580 定价: (9) 0.36元

目 录

一	抽样观察的一般概念.....	(1)
二	应用抽样观察的原因和条件.....	(4)
三	机率理论在论证抽样法方面的作用.....	(7)
四	我国关于抽样方法理论根据的探讨.....	(10)
五	全及总体及其抽样.....	(13)
六	保证抽样具有代表性的抽选方法.....	(26)
七	抽样观察的误差.....	(29)
八	随机抽样的可能误差.....	(35)
九	如何确定必要的抽样数目.....	(42)
十	典型(分区)抽样.....	(49)
十一	机械抽样.....	(59)
十二	整群(成组)抽样.....	(62)
十三	苏维埃统计中抽样观察的应用.....	(67)
附录	评 B. П. 格蘭科夫著“抽样观察”.....	
	 П. 舒舍林	(92)

一 抽样观察的一般概念

統計学是社会的特殊部門。它密切联系着大量社会現象的質的方面来研究大量社会現象的量的方面，研究社会發展的規律在具体历史条件下的量的表現，並从生产力与生产关系的統一中研究社会生产的量的方面，同时研究文化和政治生活現象的量的方面。

一切統計研究都分为三个阶段：

1. 統計观察——搜集社会經濟現象的大量統計資料。
2. 按确定的綱要整理所搜集的大量統計資料（包括彙总及統計分組）。
3. 分析和綜合在彙总、分組基础上取得的統計数字。

統計观察是根据既定項目登記社会經濟現象或填写統計原始憑証，如报表，普查表和調查表、一覽表、登記卡片等等。在統計观察过程中，可以調查該总体的所有單位或者只調查它的某一部分。

可見，由於調查总体所包括的單位不同，統計观察可分为兩種：

1. 全面观察。就是該总体的所有單位都要調查，一無例外。在我們苏联，这种全面的統計观察就是：人口普

查、牲畜普查、播种面积总结調查、車輛普查、集体农庄播种进度五日报、集体农庄畜牧月报(24号表)等等。

2. 非全面观察。在这种情况下，所要調查的不是总体的所有單位，而只是它的某一部分。非全面的統計观察可用各种不同方法进行，最常用的是填表、通訊、專題調查和抽样观察等方法。

採用填表法时，为了取得必要的資料，要佈置專門制定的調查項目(即分發調查表)。不过調查表是自願填写的，所以有一部分調查表就可能送不回来，因而不能彙总。这样一来，所取得的資料往往帶有片面性，也就是說，所取得的資料沒有代表性，不能正确和切实地反映被研究的現象或該总体的标志。填表調查的实例有：铁路旅客調查(發出后送回来的調查表約佔原發出的5%)；为了檢查地方邮电機構工作質量而进行的邮电服务对象調查；出售消費品和文化用品时發給买主調查表来檢查这些商品質量等等。

通訊法与填表法有些相似。不同的就是通訊法是把調查表分發給專門选定的人，即所謂經常向統計機構提供必要資料的通訊員。通訊員的工作通常是有一定报酬的。这种通訊員制度，在資本主义各国，首先在資產階級統計統治一切的美国採用得最廣。这些通訊員是从“可靠的”的人当中选出来的，他們按照資本家的意旨和为了討好資本家而工作。他們所提出的統計資料帶有片面性和偏見性，这种資料也不能代表整个总体。在革命前俄国，以及在苏联建国初期，当社会主义企業統計报表制度还没有建立起来的时候，曾由富裕农民組成自願通訊員網。但他們所提供

的統計資料在很大程度上是歪曲实际的。后来对这些資料进行了專門的研究和整理，規定了訂正办法。在农民大批集体化时期，在富农作为一个阶级被消灭的时期，通訊員網所提供的資料是严重地歪曲实际的。1929年通訊員網撤廢，代之而建立的是社会主义企業报表制度。

專題調查。就是对总体的專門选定的个别單位进行調查和詳細記述。在苏联的条件下，對於好的和先进的企業要进行專題調查，以便詳細研究和推广它們的工作經驗。在我們蘇維埃統計中，工厂、国营农場、机器拖拉机站和集体农庄的專題調查具有很大意义。但是，只搜集先进企業的統計資料並不能反映所有企業的工作。因此，除了研究先进企業的經驗之外，統計还担負着另外一个任务，就是搜集足以說明所有企業工作的綜合資料。为了达到这一目的就要採用抽样觀察。

抽样觀察是非全面觀察的最完善、最有科学根据的方法。其目的在於根据調查总体的某一部分来推断整个总体。此时，总体單位不能根据主觀意圖抽选，而要用随机方法抽选調查單位。抽样觀察的特点就是被調查的單位是由总体中預先确定的部分（如 $\frac{1}{20}$, $\frac{1}{10}$, $\frac{1}{5}$ 等等）。已确定（已选定）的調查單位，都一定要調查，不得遺漏一个、也不得用另外一些單位代替。

由於選擇單位的随机性，抽样觀察要应用大数定律和可以說明抽样調查誤差界限的机率理論定理。

列宁再三指出应用抽样觀察的必要性和合理性。1921年9月1日列宁在写給中央統計局局長波波夫的信里，把为了研究而選擇为数不多的典型企業（工厂、国营农場）和

机关……列为“最必要的工作”。

1934年，在領導机关關於畜牧業統計問題的指示中着重指出了科学地組織抽样調查的必要性，指示中指出，牲畜普查后进行的抽样复查的范围太小（1.5—2%），就不能据以了解小的地区，例如行政区的牲畜总头数。

在我們国家統計工作中，广泛採用抽样觀察。在苏联所实行的統計抽样調查計有：牲畜普查后进行的抽样复查；集体农庄市場价格登記；工人、職員和集体农庄庄員家庭收支調查等等。

在工業企業和農業企業中，例如，在鑑定产品質量，如鑑定乳类含脂量，种籽發芽率，谷物实重，紗的强力，电灯泡和电子管的耐用時間等等时，都广泛应用抽样觀察。

二 应用抽样觀察的原因和条件

組織和进行統計觀察的形式和方法取決於工作任务和具体情况，取決於合乎目的的条件，人力財力的合理使用以及与进行該項統計工作有关的其他因素。

某些統計調查可以用任何一种方法来进行，用全面觀察或者非全面觀察的方法都可以。例如：計算工人的平均产量和平均工資，計算平均單位面积产量以及計算本区良种牲畜比重等等就是一些实例。在实际工作中也会遇到这种情形，就是只有用全面調查才能确定被研究現象的数值。例如，全国人口总数、播种面积、牲畜总头数等只能根据全面調查(普查)来确定。

也可能有相反的情況，所要研究的標志的平均值只能根據非全面調查來確定。譬如試驗燈泡和電子管的耐用時間就是一例。為了確定它們的平均使用時間只能把一部分燈泡和電子管拿去作試驗。在實驗室里將它們通上電流來確定試驗開始和燒壞的時刻。顯然，不能把所有燈泡和電子管都送去作這種試驗。

當統計調查既可以根據全面觀察也可以根據抽樣觀察進行時，則往往要進行抽樣觀察，因為它和全面觀察比較有許多好處，如可以縮短調查時間，大大減少物力和財力，需要的調查人員（調查員、計算員和登記員）比較少，用抽樣觀察有時可能得到比全面觀察更正確的結果；在其他條件相同的情況下，抽樣觀察可以比全面觀察大大增加統計觀察項目。所有這些因素，在組織和進行統計觀察時具有重要的而且往往是帶有決定性的意義。

在偉大的衛國戰爭時期，統計觀察的期限具有決定的意義。如果在正常條件下進行某項普查需要幾個月，那末在戰時條件下就要在幾天之內作完這一工作。

如果事前已經曉得全面觀察不能保證在規定期限內提供必需的資料，那末為了迅速搜集和彙總必需的統計資料，就要組織抽樣調查。例如，根據所有國營農場和集體農莊的報表可以取得關於谷物收穫量的正確資料，而這些資料都是在全部作物收割和打谷以後才報來的。但是在我們社會主義經濟條件下，為了編制糧食計劃和進行糧食再分配，應當尽早地掌握單位面積產量的資料。

在進行統計工作時，關於如何縮減經費，如何合理而適當地利用經費的問題，也具有極大的意義。用抽樣觀察

代替全面觀察，就可以大大縮減開支。抽樣調查所需要的調查表較少，可以減少登記員、督導員和統計資料彙總人員的工資總額，也可以減少郵電、交通費、出差費等等。

抽樣調查的數目比全面調查少得多。因此，進行抽樣調查所需要的調查員、計算員、登記員也較少。所以進行這種工作，可以選出比較熟練的人員，可以更好地訓練他們和易於檢查他們的工作。結果就可以提高工作質量，提高統計資料的正確性和可靠性。因此，用抽樣觀察有時甚至可以得到比用全面觀察更正確的資料。

例如，蘇聯每年進行牲畜普查。這是全面的統計觀察。在普查結束以後，為了驗證普查所取得的牲畜總頭數資料，為了訂正可能遺漏的牲畜數，要立即進行更詳細的抽樣觀察，即所謂抽樣復查，就是在所選定的居民點里，挨戶檢查和計算牲畜。由此可見，抽樣觀察是全面觀察的補充：根據抽樣復查可以修正普查的結果。

使用同樣多的物力和財力，經過同樣多的時間，搞抽樣調查就可以增加觀察的項目，也就是說，可以根據更多的指標和標志搜集更詳細的資料，可以更詳細和更全面地說明被研究的總體。

有時，因工作量太大不能進行全面觀察。例如，確定糧食實重（每公升的平均重量）時，如用全面觀察，就得調查一個國營農場或集體農莊的所有糧食，或者全區的所有糧食調查的總重量（公斤）和容積（公升）。至於根據全面調查確定谷物雜質百分比，確定蘇聯所有森林地區的木材儲存總量，以及確定某一地區礦產的儲藏量，那就更困難了。在所有這類情況下都要採用抽樣觀察。最後，已如上述，

在下列情況下，例如試驗燈泡和電子管的耐用時間，試驗種子發芽率，試驗紗的強力，以及嘗試食品工業和調味品工業的產品（紙煙、茶、酒及罐頭等）味道時，就不能採用全面調查，因為採用全面調查就會毀壞該總體的調查單位，使其不能使用，在所有這類情況下，只有採用抽樣觀察。

三 機率理論在論証抽樣法方面的作用

為什麼抽樣觀察資料有時比全面觀察的結果還正確呢？要回答這個問題，我們必須明確統計觀察誤差這一概念。統計觀察的誤差有兩種：1. 登記的誤差，2. 代表性的誤差。

登記誤差無論在全面觀察或非全面觀察中都可能發生。登記誤差就是統計原始憑證中所記載（登記）的資料與該總體實際具有的資料之間的差別。登記誤差的產生原因如下：

1. 登記員的工作態度馬虎，缺乏某項統計觀察方面的足夠的知識和工作經驗；
2. 計量用具不准，例如，在測量土地面積時現有的卷尺、米尺、雙米尺不准，在秤量產品重量時秤具不准等等；
3. 登記員或資料提供者故意歪曲統計資料，提供錯誤資料。

我們蘇維埃統計的特點，它與資產階級統計的原則區別，就在於我們蘇聯的所有統計數字都必須絕對準確和客觀。因此在統計工作的所有階段，首先是在統計觀察過程

中，坚决反对統計資料的各种錯誤和歪曲实际。

克服登記誤差的办法如下。統計觀察必須挑选通曉調查統計工作的認真負責和最熟練的工作人員进行。對於調查幹部要認真培养和訓練，要在統計觀察过程中进行檢查並就地改正工作中缺点。在統計工作中要使用精确的、經過仔細檢查的計量用具。对羣众要普遍进行宣傳，並講解和說明統計工作的目的，說明調查的任务是取得編制国民经济計劃及管理社会主义国家所必需的正确統計資料。

抽样觀察的工作量較少，因此，在幹部比較熟練、訓練得很好、而且对其工作的檢查比較認真的条件下，登記誤差就会少於全面觀察。由此也就可以說明，在进行抽样觀察时，例如在进行牲畜普查后的抽样复查时，会取得比全面觀察更为准确的資料。

代表性誤差，是一种反映上的誤差，是一种一致性上的誤差。这些誤差只在非全面觀察中發生，因为在进行非全面觀察时，我們是打算根据总体的已調查部分来推断整个总体。代表性誤差是指非全面(特別是抽样)觀察指标或标志与整个統計总体的同一指标(标志)之間的差別。發生这种誤差的原因就在於所調查的不是該統計总体的所有單位，而仅仅是該总体的某一部分。該总体已調查的部分，就其成分，即就其構成來說，可能与整个总体有差別。例如，在一个托拉斯的工厂里有25,000工人。全部工厂的工人每月平均工資为898盧布。用抽样方法調查了900工人。根据这个調查，我們确定工人每月平均工資为900盧布。按我們的計算，非全面觀察所取得的平均工資(900盧布)与整个总体的同一指标(898盧布)之間的差額，就是

代表性誤差，可用 Δ 代表。就本例來說，代表性誤差 $\Delta = (900 - 898) = 2$ 盧布。

在某些非全面觀察當中，代表性誤差，通常偏於一個方面；也就是說，根據這些觀察所取得的指標的誤差，通常偏於一個方面。例如，向買鞋的人發調查表，調查鞋的穿用時間，那末很明顯，送回來的調查表，大部分都是短時間內就把鞋穿壞了的買主送來的。因此，根據這個調查表來估計鞋的平均穿用時間就會偏低。在這種情況下，誤差可能很大，而且不能用增加總體的調查部分的方法來減少誤差。即使再增加總體的調查單位，代表性誤差，仍將是一貫偏向於一個方面。

至於進行抽樣調查則情況就不同了，抽樣調查的調查單位不為調查者的意志和願望所左右，而是隨機選出的。在這種情況下，代表性誤差帶有偶然性。這些誤差既可能是正差也可能是負差，兩種誤差都同樣可能。同時誤差的絕對值也有變動，一個調查者的代表性誤差可能大於或小於另外一個調查者的代表性誤差。因此，機率理論定理、大數定律都適用於抽樣觀察，而對其他各種非全面觀察來說就談不上適用了。

大數定律的本質就在於抽樣數目越大，根據抽樣觀察取得的綜合指標與整個總體的同一指標的區別就越小。運用大數定律我們可以增加抽樣數目，以控制代表性的可能誤差的範圍，使其盡量小（小到容許的程度）。知道了容許誤差（公差——譯者）的範圍，我們就可以根據大數定律確定抽樣的必要數目。在統計學中，大數定律提出了抽樣法的數學根據。而資產階級統計學家所持觀點就不同，他們

認為大數定律是抽樣法和作為科學的整個統計學的基础。從這一點出發，他們認為無論對社會生活現象，也無論對任何自然現象，統計學都同樣適用。他們認為統計學的主要條件，就是觀察的大量，同時並不根據人們對社會發展客觀規律的認識把這些觀察加以綜合。上述提法根本不對。任何一門科學，包括統計學在內，其方法都是馬克思主義的辯證法。在資產階級社會，統計方法和其他社會科學一樣都是形而上學的方法，都是形形色色的實證主義的方法。

資產階級統計學，被單純經驗主義所統治，迷戀並且濫用數學和數理統計學。與此不同，蘇維埃的統計學家認為統計研究的基础應當是歷史唯物主義和馬克思列寧主義政治經濟學，認為大數定律，在不同的程度上適用於自然現象和社會生活現象。現象愈複雜，那末應用數理統計方法就應當愈加慎重。

四 我國關於抽樣方法理論根據的探討

抽樣觀察法老早以前即為俄國人所熟知。他們在進行抽樣觀察（統計觀察方法之一）方面已立下了豐功偉績。

三百多年以前在俄國就採用了抽樣法，此事到現在尚有文件可查。

在馬羅卓夫大貴族（1648年）的莊園“記事”中，曾指出如何利用抽樣法來估算收穫率。關於抽樣打谷的作法，寫在馬廐辦公室的章程（1733年頒佈的“馬廐章程”）中。當然，這些抽樣觀察，是相當原始的方法，不過在當時却

是統計研究的新方法。

在俄国地方自治局統計中广泛地使用了抽样法（十九世紀末——二十世紀初）。地方自治局統計虽然在工作中有許多严重錯誤和缺点，但是对統計工作的理論研究和組織，特别是在搜集俄国革命前农民經濟的大批优良統計資料（农民經濟按戶普查）方面作了很多工作，地方自治局統計在工作中經常採用抽样觀察並研究和改进了抽样調查的方法。1900年在伏雅特省第一次採用所謂典型抽样，把整个調查总体預先划分为大致相同的組。这种抽样法比一般的、亦即純随机抽样可以提供更准确的結果。1896年，加路格地方自治局統計人員第一次採用了机械抽样，这种机械抽样比純随机抽样具有組織上的优点，而且可以保證遵守不帶偏見的抽选原則。

偉大的十月社会主义革命胜利以后，我国开始建立了国家統計。1918年建立了中央統計局。为了搜集統計資料，中央統計局开始进行了專門統計觀察（全面普查）和抽样調查。特别是在农業統計方面进行了許多抽样觀察：农户的春秋兩季抽样复查、农民家庭收支調查、打谷抽样复查等等。苏維埃統計学家在科学的基础上，制訂了使用公尺方框和五公尺長杆实测而进行的糧谷和寬壠作物的單位面积产量調查以及其他許多抽样觀察（牲畜普查以后的抽样复查、工人、職員及集体农庄庄員的家庭收支調查等等）。

如上所述，抽样法是以大数定律为根据的。大数定律的基本的、最一般的原理，为我們俄国的数学家 П.Л. 車貝舍夫、А.М. 李亞普諾夫和 А.А. 馬尔科夫等院士所确

定。苏維埃的大数学家院士 C·H·別尔恩什坦、院士 A·H·柯尔莫果洛夫和 A·Я·辛欽的著作对机率論作了进一步的探討。

就抽样来講，車貝舍夫定理可表述如下：設机率愈接近於 1，就可以肯定，独立观察的次数足够多时，那末根据抽样而取得的被研究标志的平均值与整个总体同一标志的平均值的差別就尽小。

讓我們回想一下数理統計学教程中車貝舍夫定理所講的有名的机率基本概念。机率就是在一定的、可能重复無限次的条件下，一定事件出現的客观可能性的程度。成功的(順利的)机会数(M)对全部机会数(N)的比率叫做事件(A)的机率。可列式如下：

$$P(A) = p = \frac{M}{N}$$

机率的范围是从 0 到 1

$$0 \leq P(A) \leq 1。$$

如果事件 A 的机率小(換句話說，接近於零)那末这个事件 A 出現的机会就少。这种事件叫做可能性小的事件。实际上可以把它看做是不可能的。如果机率接近於 1，那末这个事件出現的次数就頻繁(在大多数情况下)。 $P(A)=1$ 时，事件 A 就是真实可靠的。可以指望它确实出現。

П·Л·車貝舍夫的功績就在於他在 1867 年証明了自己的定理，确定了大数定律的一般形式。当然在更早以前(1713 年)瑞士数学家亞科布·白努意發表过类似的定理，但这个定理仅就个别情形，即仅就被研究标志的比例說明

了所謂頻率和機率的相互關係。就抽樣觀察來說，亞科布·白努意定理可以表述為：設機率愈接近於 1，可以肯定獨立觀察次數足夠多時，根據抽樣觀察所取得的被研究標誌的比例與同一標誌在整個總體中的比例的區別就盡小。因此，白努意定理所講的是抽樣中標誌的比例（頻率）和全部總體中同一標誌的比例（機率）的互相關系，舉例來說，就是根據抽樣觀察所確定的種子發芽率（百分比）和全部種子發芽率之間的比例關係。

世界最著名的數學家（普阿森、拉普拉斯、高斯等等）繼亞科布·白努意之後，在機率論方面進行了長期的、頑強的工作，取得了許多成就，但卻沒有總結白努意的定理。這個任務是由俄國數學家車貝舍夫輝煌地完成了。他提出了關於大數定律一般形式的證明。從車貝舍夫定理中可以很容易地看出白努意定理是前者的一種局部情形。

關於車貝舍夫定理，等到研究了統計觀察中的基本指標以後，還要詳加說明。

五 全及總體及其抽樣

進而談抽樣觀察的組織問題時，我們應當明確統計觀察的某些基本概念和某些指標（標誌）。如果全部單位都調查，那末我們可以觀察包括全部單位的整個總體即所謂全及總體。全及總體中的抽樣調查部分叫做抽樣總體。全及總體的單位數目（數量）用 N 表示，而抽樣總體的單位數目（抽選的數或量）用 n 表示。

進行統計觀察可以提出兩個任務：1. 確定被研究標誌