

經濟論叢

第七十九卷 第四號

- 標本調査法の基本問題……………馬場吉行 1
- 『資本論』における『農民的分割地所有』範疇の
適用について……………福富正実 20
- 府県税・戸数割について……………北沢康男 36
- 木頭の森林組合……………山崎武雄 55
永尾誠之 輔
-

昭和三十二年四月

京都大學經濟學會

標本調査法の基本問題

馬場 吉行

一 はしがき

周知の如く、第二次世界大戦後、わが国統計調査界における大きな変革は、いわゆる「科学的」標本調査法が広く採用されたことであろう。しかしながらここに科学的標本調査法というのは、従来統計調査法の基本とせられて来た、全部調査（完全調査・悉皆調査）の補充・代用として登場したものであり、数学的確率論に基づく抽出理論によつて、調査全域に対する部分を抽出し、その部分の調査結果によつて、全域についてのありうべき結果を推定しようとするものである。しかも本来標本調査の目指す所は、全部調査にかえるものであるから、抽出された部分、いわゆる標本は、全域の縮図的表現である。しかるに数学的抽出理論においては、平均・総計・構成比率ないし比の推定に終つており、ここに層別任意抽出法その他の諸方式を生むこととなつたのである。

一方統計調査代用法としては、個別典型調査（標本詳査）があげられる。これは全域の典型とみなされる部分によつてこれを調査し、もつて全域の調査にかえようとするのである。¹⁾

私はこれらの二つを総括し、標本調査法とよぶ。その沿革をたずねると、一九二五年ローマで開かれた国際統計協会第一四回会議において取上げられた、当時の各国統計学者に対する共同研究議題、「統計学における代表法

(Representative method)の研究」を、現在への新しい出発点と考へうと思われ⁴⁾。この会議においてジェンセン (Adolph Jensen) は二つの報告をなし、統計調査法における代表法の学問的定義、その応用を論じている。ここで代表法とは、全域の代表標本を中心とする統計調査代用法の意味であつて、これによらない単なる見積り法とか、問合せ (アンケート) とかを除外したのである⁵⁾。

其の後標本調査法の発展は、主として数理統計学者の手により、数学的抽出理論の展開となり、現在の「科学的」標本調査法に到っている⁶⁾。しかも計画選出法 (Purposive selection) は、典型調査法の形をとることなく、調査項目と相関度のたかい補助的目的 (照準 Control) に関する既調査事実との関連において、数学的に問題にされたに過ぎなかつた。そしてこれも層別任意抽出調査法のなかに吸収されたのである⁵⁾。

一方ドイツ社会統計学派の主張した個別典型調査法は、その代表としてとる部分の選定が、恣意的・主観的であるとして、一般の数理統計学者からは等閑視され、その実質的意味の重要性については、深く顧みられては居らな⁵⁾。

私は従来こうした標本調査法の論理的性格、その適用可能性について研究して来たが、本稿においては統計調査法の論理的性格を反省しつつ、標本調査法のそれを明らかにし、標本調査法の問題をそのもつ代表性を中心として検討したい。そして標本調査法がどうして二つの型にわかれるか、それが現在の標本調査実施にあたって、どう利用されるべきかについて私見を述べようと思ふ。

(1) 拙著、社会統計学と抽出理論 (増補版) 有斐閣、昭和三〇年、第二章標本調査法について、参照。

(2) 同書、第二章補論第三、標本調査における代表性について。

(3) 嶋川虎三、統計利用の基本問題、岩波書店、昭和七年、第三章大量標本代用法の性質と問題、参照。

(4) 標本抽出調査法の発達については、

F. F. Stephan: "History of the Uses of Modern Sampling Procedures," Journal of the American Statistical Association, (March 1948) pp.12-39 を参照。

(5) 前掲拙著、第二章標本調査法について、および拙著、誤差法則と抽出理論、関書院、昭和二四年、第二編第一章、抽出調査の基本問題——ネイマンの所説を中心として——を参照。

(6) 前掲拙著二冊、および拙稿、「標本調査法における代表性の問題」、滋賀大学経済学会、彦根論叢第三四号、昭和三一年一月。

二 統計調査法における問題

標本調査法は、本来の統計調査法の代用法として位置づけられる。そこで標本調査法の在り方を考察するに先だち、統計調査法のもつ性格について考察してみよう。

(一) 統計調査法は統計の作り方といえよう。ここに統計とは、社会的集団の数量的表示結果であると定義づけられている。しかしながら我々の求めようとするのは、実は変動し運動している社会現象を、数量的に表示した結果である。そのためこれを静止の姿においてとらえる必要がおこる。即ちある時点、あるいはある期間についておさえ、いわば断面 (Cross-section) においてこれを調査しようとする。この断面が統計調査の対象たる全域で、社会的集団といわれるものである。通常この四要素として、単位、標識 (特性)、存在する時、存在する所をあげている。いま社会現象のある断面において、社会現象を構成している運動している (生成・発展・消滅している) 単

位の断面が、右の四要素の一つである単位であり、断面はまさにこの単位の集団として理解されるのである。そしてこの単位あるいは全域を特徴づけるものとして、特性（標識）があり、この特性によって分類せられた結果が統計となる。

以上述べたように、断面的把握として時点的存在を得るが、これらはそれぞれ静態統計、動態統計を与えることとなる。しかし、実はいずれも運動している社会現象としてみて、はじめてその数量のもつ意味・性質を理解しうるのである。もちろんこれは調査対象についてであって、調査実施の時・期間についていうのではない。

次に常態 (Actual status) 調査の意味について考えよう。例えば常住人口は、現在地人口とは区別せられる。前者は運動している人口現象についてはじめて意味をもつのである。通常の統計調査において、この常態調査を問題にするのは、まさに統計調査が、社会現象の断面調査を通してその動ける姿、方向を見定めたいという意図のあらわれと解せられる。このことは失業者の実態調査、生産・消費の実態調査などにおいて意味をもって来る。

(二) 以上述べたように、統計とは断面としてとらえた社会的集団の数量的表示結果と考えられるが、次にいまましく立入って考察してみよう。先ずこの社会的集団（全域）は単位により構成せられている。例えば人口集団においては、個人が単位と考えられる。そこで例えば性別という質的特性をとれば、この集団は二つの部分集団にわたれ、人口の性別構成が得られる。総人口、男子人口、女子人口はいずれも統計値である。其の他の質的特性、たとえば職業をとるも同様である。次に年令という量的特性をとれば、総人口のほか年令階級別人口という統計値が得られる。いわゆる年令別人口構成であり、一つの度数分布を得る。人口調査の場合、各人の所得を特性に取り

うるが、別に国民の所得を社会的存在、従つて統計値とみることも可能である。

(三) 次にいわゆる誘導統計値⁽¹⁾について考察しよう。これには平均と比率とがある。

いま N 人からなる人口集団において、量的特性として各人の年令 (A_k) をとると、

$$A_1, A_2, \dots, A_N \quad (1)$$

なる特性値の集合を得る。これは丁度形式的には測定値の集合に照応する。通常年令階級により分類し、度数分布として表わされる。即ち

$$\text{(年令階級)} \quad B_1, B_2, \dots, B_N \quad (2)$$

$$\text{(人 数)} \quad f_1, f_2, \dots, f_N$$

この場合平均年令としては、計算的平均である相加平均(算術平均)、位置的平均である最頻値(Mode)などが採られる。もちろん実質的意味から言えば、この場合最頻値は意味をもつが、相加平均は意味をもたない。しかし一応の目安として、あるいはいろいろな利用面から必要なこともある。この場合かかる平均(統計的平均)は、誘導統計値とよばれているが、実は特性値から誘導された値である。

また同じく人口集団において、各人の所得を特性として選べば、前述(1)の如き系列を得る。しかしこの場合(1)又は(2)の総和、即ち

$$\sum A_k = A_1 + A_2 + \dots + A_N \quad (3)$$

$$\sum f_j B_j = f_1 B_1 + f_2 B_2 + \dots + f_N B_N \quad (4)$$

はこの人口集団においての所得総額を示すのであり、それ自身具体的な意味をもつ。即ち社会的集団とみてよい。

そしてこの場合、 A_i 又は B_j はこの社会的集団の構成単位とみてよい。そこで相加平均を取ってみると、それぞれ

$$\frac{\sum A_i}{N} = \frac{A_1 + A_2 + \dots + A_N}{N} \tag{5}$$

$$\frac{\sum f_j B_j}{\sum f_j} = \frac{f_1 B_1 + f_2 B_2 + \dots + f_m B_m}{f_1 + f_2 + \dots + f_m} \tag{6}$$

を得る。これは一人あたりの平均所得を示している。形式的には A_i 又は B_j の相加平均（単純または加重）とも見られる。または二つの照応している社会的集団の大きさの比とも見るのである。

この点を更に明らかにするために、次の例を考へよう。

米作農家番号 耕作面積(反) 米生産高(石)

1	2	3					
A_1	A_2	A_3	\dots	A_N	B_1	B_2	B_3
A_1	A_2	A_3	\dots	A_N	B_1	B_2	B_3

(7)

農家あたり平均耕作反別、平均米生産高を相加平均で示せば、それぞれ

$$\frac{\sum A_i}{N} = \frac{A_1 + A_2 + \dots + A_N}{N}, \quad \frac{\sum B_j}{N} = \frac{B_1 + B_2 + \dots + B_r}{N} \tag{8}$$

で与えられる。この場合 $\sum A_i$ 、 $\sum B_j$ はそれぞれ具体的な社会的集団の大きさを示している。次に各農家につき単位耕

作面積あたりの米生産高は

$$\frac{B_1}{A_1}, \frac{B_2}{A_2}, \dots, \frac{B_r}{A_r} \quad (9)$$

なる比率であらわされ、総農家に対しては米総生産高に対する総耕作面積の比としてあらわされる。即ち

$$\frac{\sum B_i}{\sum A_i} = \frac{B_1 + B_2 + \dots + B_r}{A_1 + A_2 + \dots + A_r} \quad (10)$$

これは(9)なる比率系列の、重みを A_i とした加重相加平均に外ならない。

こうした場合米作農家の集団を考えれば、 A_i 、 B_i 系列はいずれも、構成単位である農家の特性を与える数値の系列である。あるいは耕作地集団と生産米集団とを耕作農家集団により照応している。二つの社会的集団とみることが出来る。こうした考え方は、地域による選出(地域抽出 Area sampling)の問題を考える際に必要である。

次に統計的比率について考えよう。先ず社会的集団の質的特性をとるとき、構成的比率を得る。例えば人口集団において性別という特性をとれば、男 A 人、女 B 人、計 S 人となり

$$\frac{A}{S}, \frac{B}{S} \quad (S=A+B) \quad (11)$$

なる構成的比率を得る。職業別その他の質的特性を取るときも同様である。なお例えば年令別、所得(階級)別の如き量的特性については、前に述べたように度数分布を得るが、この場合にも各数値を全集団の大きさと割ることにより、(11)の如き比率による度数分布が得られる。

次に前述の如き一人あたり平均所得、単位面積あたり平均米生産高の如き相関的比率は、平均とみることが出来る。

る。また因果的比率は、原因・結果の關係にある二つの社会的集團の大きいさの比である。例えば出生率は人口集團（靜態）の大きいさと、そこから發生した出生兒集團（動態）の大きいさとの比として与えられる。

なお指数については、ある社会的集團を運動においてとらえようとする時あらわれる。そして例えば人口数の推移・物価變動の測定については、二時点におけるこの社会的集團の大きいさの比が問題となり、また例えば生産高の推移の場合には、二期間におけるこの社会的集團の大きいさの比として与えられる。

- (1) 蜷川虎三、統計利用の基本問題、第一章統計の概念、参照。
- (2) 拙著、社会統計学と抽出理論、第二章補論第二、標本調査法の理論的性格、参照。
- (3) 蜷川虎三、前掲書、第一章統計の概念、参照。
- (4) (5) 同書、第四章統計利用の意義と問題、参照。

三 標本調査法の論理的性格

標本調査法が、いわゆる統計調査代用法において占める地位については、既に一言した。即ち社会現象の断面としての社会的集團を数量的に把握する場合、構成単位を選出してその調査を行い、これによって断面集團の全部調査にかえようとするのが、ここにいる標本調査である。従つて構成単位を選出によらない見積り（推算）とか、單なる問合せ（アンケート）の如きは、これに含めない。

さて断面調査に際し、通常いわゆる任意抽出調査と典型調査の二方法に別っている。次にこの必然性、さらに両者の関連について考察してみよう。

先ず運動している社会現象を数量的につかむために、その断面（時点的または期間的）について調査しようとい

うのが統計調査法の性格である。そこで断面を調査するとして、如何なる時において断面を取るべきかが問題であり、その調査結果は常にもとの社会現象の数量的把握として理解されねばならない。このことは卓近な例をとれば、写真の撮影による被写体の理解の如きものである。

次に標本調査法については断面の縮図構成に関する考察が必要である。先ず断面の縮図をとるためには、断面の構成単位を選出し、これを編成している。しかしながらこれは果して如何なる意味をもつだろうか。数量的な観点から考察すると、編成された構成単位の集り即ち標本は、断面（全域あるいは母域）に対比して、あらゆる特性について比例関係をもつものが、理想的な代表標本あるいは縮図である。かりに数量的にこの選出が可能であつたとしても、これを有機的の一体たる社会現象——その断面——の縮図と考へうるだろうか。即ち社会現象を構成している単位は、相互に結びあい意味づけあつてゐる。これを統計調査においては、仮りに個々の構成単位の集りに解し、かかる単位の調査の集大成として、数量的に再構成してゐるわけである。いま我々の代表標本といへば、右の理解のもとに個別単位の若干を構成してゐるのである。こうした見地から断面の縮図的表現の問題は、まずその基礎として完全調査（本来の統計調査）をもつてゐるといふ。断面に即していへば、まず断面（全域）についての知識を前提としてゐるといふ。もちろん実際の調査においては、全域について更に委しい数量的知識を得たのであり、全域における重要特性を媒介とし、あるいは全域と同系列にある他の時における断面についての調査結果、あるいは類似の他の集団現象の断面に結びつけてゐる。

さて本来の統計調査が社会現象の数量的把握であること、運動してゐる社会現象をとらえるには断面調査を必要とするとは、繰返し述べた所である。そこでこの社会現象の動きをとらえようとするとき、この社会現象を構成

している因子（単位）の動きが問題となる。既に述べたように全体として有機的に結ばれ、個体の単なる集りを越えた集団であるけれども、いま仮りにその構成している因子の活動により、全体の動きを理解しようとする時、必然的にその典型的因子を求めることとなるであろう。ここに典型とは数量的には最も多い型とみられようであろう。或いは社会の運動法則を端的に示す因子と解してもよい。このとき社会現象の全部観察を、構成している個（因子）の観察に帰させようとするのである。いまこれを断面における統計調査とみると、断面における典型的な単位の選出となる。しかもこれを運動している姿において理解しようとするのである。

最初に述べた代表標本は本来、断面の単位全部を意識しているものであり、後者の代表標本は本来個別的な単位に限られる。前者が数量的な平均単位群を指すというならば、後者は性質的な典型的な単位群を指すといえよう。いま前者を外延的（Extensive）標本調査と名付けるならば、後者は内包的（Intensive）標本調査といえよう。

しかも本来の統計調査の目標からいって、この両標本調査は互に経緯となって結びあい、調査の役割を果すべきものと考えられる。このようにして統計調査代用法としての標本調査法は、二つの型に大別されるのである。

更に進んでこの外延的標本調査法について考えよう。この場合断面の構成単位をどう取るかである。我々は一般に断面に関する何かの基礎的知識（既往の調査、あるいは推計）に依存して標本抽出をなすのである。即ち集団のもついろんな特性のうち、いま問題とする諸特性に最も関係の深い重要特性を基礎とする。あるいは現に述べたようにこの断面（全域）と関係ふかき基礎となつている他の集団（断面）に照応して代表標本を考えるのである。例えば耕作反別、米收穫高等の調査において、その耕作地に関する基礎調査を基盤として、代表標本を考える如きである。³⁾

そこでこうした特性より見て、断面（全域）が等質でなく、異質の部分に別かれていることが明らかであれば、勿論これらを互に区別していわゆる層別することになる。これは本来の統計調査において、既にその調査結果の作製、調査結果の利用において、重要な意味を荷っているのであって、外延的標本調査においてはじめてあらわれるのではなく、ただ抽出設計において先ず判然として、重要性をあらわして来るに過ぎない。次に全域が一段あるいは数段の層別によって、もはや同一層内においては、いまの重要特性につき、単位相互間に質的差違を認め難い場合、あるいは雑多な単位の集りではあるが、区別を認める必要のない段階に立ち到ると、構成単位（あるいはその群）を無差別的にみることとなる。即ち抽出に即していえば、構成単位（またはその群）を無差別的に選ぶこと、即ち任意抽出の構想に到る。しかも実際ににおいては、いわゆる系統抽出法を採ることが多い。即ち抽出さるべき単位（抽出単位）の集りに一連の番号をつけておいて、いわゆるランダム・スタート（Random start）たる番号を選び、抽出率の逆数を抽出間隔に取って、順次に抽出単位を抽出するのである。この方法は構成単位、ひいては抽出単位のもつ種々の特性について、その選出された単位を集計するとき、どの特性についてみても可なりよい近似を示す代表標本が得られるからであって、一般的には大数法則の一利用といつてよいであろう。

このようにして、この外延的標本調査法は、いわゆる任意標本調査法となった。ここに採用せられる数学的確率論に基づく抽出理論に即していえば、単純任意抽出理論が基本となり、層別任意抽出理論その他の諸方式は、この基本形式の発展した形態と考えられている。

なおここで、内包的標本調査法即ち個別典型調査法と関連して、いわゆる計画選出法（Purposive selection）の論理的性格について一言しておきたい。

上に述べた外延的標本調査法は、断面の縮図即ち代表標本を得るために、その構造に關し予め得ている知識を利用し、先ず層別しているのである。これと同様に、例えばある量的特性(A)について、総和・相加平均を求めたい場合、この特性と関連性のたかい(即ちいわゆる相関度の高い)他の特性(B)について現にこの集団の量的構造を知っておるとすると、このBについての集団の構造分布の知識を基礎として、構成単位をふりあてて、選出する構想が考えられた。しかしこれは先ず層別して、ついで各層内で任意抽出する層別任意抽出法に落付くのであって、ネイマンの有名な二論文⁵⁾は、まさにこの点を数学的抽出理論により、また具体例によって指摘したのであった。⁶⁾

(1) 本論文はしがき註(3)参照。

(2) 拙著、社会統計学と抽出理論、第二章標本調査法について、参照。

(3) 本論文第二節。

(4) 前掲拙著、第二章補論第三、標本調査法における代表性、および前掲拙稿一五六頁、参照。

(5) J. Neyman: "On the Two Different Aspects of the Representative Method of Stratified Sampling and the Method of Purposive Selection," Journal of the Royal Statistical Society, Vol. 97, 1934, pp.558-625.

J. Neyman: "On Statistical Methods in Social and Economic Research—Census by Sampling and Other Problems," Lectures and Conferences on Mathematical Statistics, 1937, pp. 89-108.

(6) 拙著、誤差法則と抽出理論、関書院、昭和二四年、第二編第一章、抽出調査の基本問題——ネイマンの所説に關連して——参照。

四 標本調査法の問題

前節において、私は統計調査代用法としての標本調査法が、二つに大別され、一つは任意標本調査法となり、他

は典型調査法となる論理的必然性について論じた。

さて現在における任意標本調査法は、数学的確率論的抽出理論に基づいて設計せられている故、この数学的理論との関連において、具体的な任意標本調査法を取りあげてみよう。

外延的な標本調査法を問題とすると、代表標本という観点から、調査対象たる断面の層別が必要であることは既に述べた。そこである重要性からみて既に層別した後、或いはこの集団に対して何等の層別をなし得ないか、なさなかつたとき、任意標本抽出の操作の段階に入ることとなる。任意標本抽出は、層内の構成単位あるいは構成単位の群を、抽出されるべき単位（抽出単位）として、これらの集合から若干個の単位を無作為的に（At random）抽出し、この抽出された単位の集合（標本）について、母域のもつ各種特性について調査するのである。この場合私は母域（Universe）を右のよのように抽出母体と考えたときの断面と解し、各種特性をもつ抽出単位の集りとする。また母集団（Population）をある特性についてとらえた抽出単位の特性値の集合と解し、両者を区別している。¹⁾

数学的確率論的抽出理論においては、この母集団からの標本抽出、ついて抽出された標本における計算値から、母集団における統計値（母数）を推定しようとする。以下この見地から単純任意抽出法を考察してみよう。

先ず最も基本的なのは、大いさ N なる母集団

$$A_1, A_2, \dots, A_N \quad (1)$$

から任意抽出により、大いさ n なる標本

$$x_1, x_2, \dots, x_n \quad (2)$$

を得るとき、標本の相加平均

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n} \quad (3)$$

により、母集団平均

$$m = \frac{A_1 + A_2 + \dots + A_N}{N} = \frac{\sum A_i}{N} \quad (4)$$

を推定する場合である。このとき標本平均 (\bar{x}) は確率変数で、母集団平均 (m) の数学的期待値である。即ち m の不偏推定値になっている。そして \bar{x} の分散は

$$\text{Var}(\bar{x}) = \frac{N-n}{N-1} \cdot \frac{1}{n} \sigma^2 \quad (5)$$

と与えられる。ここに σ は母集団の標準偏差をあらわしている。

標本平均が母集団の不偏推定値であるという意味は次の通りである。即ち任意標本の平均は、抽出の都度その値を變ずるいわゆる確率変数であるが、しかもありうべきこれらの値の相加平均は、母集団平均と一致するのであって、これが抽出された標本の代表性を示している。即ち任意抽出においては、個々の抽出された標本について、その標本の代表性が云々されるのではなく、かかる抽出機構について全体的に問題となるのである。

次に分散 $\text{Var}(\bar{x})$ はその平方根 $\sigma(\bar{x})$ が、ありうべき標本平均の分布の標準偏差を与えるのであり、しかも σ の確率分布は、近似的に誤差法則(正規分布法則)に従うとみなし得るので、従って $\sigma(\bar{x})$ の三倍あるいは二倍は、 m の両

側に σ の落ちる「可能範囲」を確率的に示すのである。即ち σ が $m+3\sigma(x)$, $m+2\sigma(x)$ の範囲におちる確率は、それぞれ九九・九七%、約九五%である。これよりいまだ得た σ の値に対し、 m が $m+3\sigma(x)$, $m+2\sigma(x)$ の範囲に落ちる確率は、それぞれ上記二つの確率になると考えるのである。そこで任意抽出を行ったとき、その標本平均 \bar{x} は、いづれも母集団平均の代表と考えるのであるが、更に N , n , σ の大小によって、分散の大小を生じ、分散の大きさにより代表性の大小が抽出機構に即して論じうるのである。標本平均の不偏性ならびに分散の大小比較の問題は、単純任意抽出法を基本として、更に種々の方式を生むに到ったのである。

なお任意抽出調査の実際においては、右のような抽出機構自体の代表性、その大小の問題のほかに、なるだけ推定値が真実値に近いようにとらうという消極的工夫が取られるのである。この点については既に別稿「標本調査法における代表性の問題」において論及したので、ここでは割愛しておこう。

以上述べた標本平均による母集団平均の推定理論は、これを総和・構成比率・度数分布の推定に應用しうるのであって、従つてまたこれらの統計値の代表性、その大小を抽出機構に即して考えうるのである。以下既に考察した統計値の意味に照応しつつ、順次この問題を考えよう。

先ず社会的集団の大きさ、部分集団の大きさという意味での、本来の統計値の推定を考えてみる。例えば人口調査において総人口、年令別人口、男女別人口の推定である。この場合人口調査が全部調査として行われ、単に年令別、男女別人口を得たいような場合には、総人口(N)は母域あるいは母集団の大きさを表わして既に既知である。また質的特性である男女別については、男を1、女を0なる特性値で置きなせば、男 A 人、女 B 人とすればこれらの特性値の相加平均は

$$\frac{1+A+O+B}{A+B} = \frac{A}{A+B} = \frac{A}{N} \quad (6)$$

となつて、男の占める割合は、かかる特性値の相加平均に帰せしめられる。従つて N を乗ずることにより A が出る。即ち標本平均から母集団平均を推定する問題に帰せられる。^①このことは質的特性として職業の如き三つ以上にわかたれる時にも拡張して用いうる。次に住居地域、あるいは世帯を構成単位とし、それに属する人数を特性としてとると、普通の平均の問題となる。そして総人口、男女別人口、年令別人口の推計はいずれも総和の統計の問題に帰せられる。またある地域における総耕作面積、米收穫高の推定の如きは、たとえば各耕作者を単位にとれば、特性値の総和の推定の問題となる。

統計的平均ならびに比率については、既に述べたが、いまこれらの推定方法について考察しよう。

例えばある所得調査において一人あたりの平均所得を推定する場合、個人を抽出単位にとり、特性を各人の所得額とすれば、この問題は全く前述の抽出理論にあてはまる。しかしまた世帯ごとに調査する場合には、 A_i 、 B_i をそれぞれ第 i 世帯での総所得額、世帯人員とすれば、これら N 世帯につき一人あたりの平均所得額は

$$\frac{B_1+B_2+\dots+B_N}{A_1+A_2+\dots+A_N} \quad (7)$$

で与えられる。いまかかる N 個の世帯から n 個を任意抽出したとすると、それらの所得・人数の系列を

$$y_1, y_2, \dots, y_n; x_1, x_2, \dots, x_n \quad (8)$$

とすれば、この標本における一人あたりの平均所得は

$$\frac{y_1 + y_2 + \dots + y_n}{x_1 + x_2 + \dots + x_n}$$

(9)

て与えられる。(9)の分母、分子はともに抽出の都度、その値を變ずるものであり確率変数である。この場合(9)なる標本平均(標本比)は(7)なる母集団平均(母集団比)の推定値としてとりうるであろうか。これはまた指數の推定についてもおこる。

この問題は数学的抽出理論において、近年展開せられた比推定(Ratio estimate)の理論に帰せられる。この推定値はいわゆる偏りのある(Biased)ものであつて、あらゆる可能な標本比の算術平均、即ちその数学的期待値は母集団比に一致せず、ある偏りをもつのである。しかし一方分散を計算してみると、平均推定の場合に比し、その値を相当小さく出来る特質をもっている。

以上は断面において特性を一つとつた場合であるが、実際の標本調査では断面を構成している単位は、多くの特性をもつものとされ、従つて母域から抽出単位を選ぶことになる。例えば抽出人口調査において、年齢別・性別・職業別人口の推計が同時におこなわれるのである。そこで平均の推定に関する限り不偏性は失われなくとも、分散の大小は大小区々となり、得られた平均値、度数分布の各区域(階級)の値、構成比率の値は、推定値としての信頼区間を異にし、精度は互に異なることとなる。

次に任意抽出調査においては、いわゆる系統抽出が實際上多く用いられる。これは任意抽出のうち無難な一方法とみるのであり、特に抽出母体(母域)を地域に求める場合(地域抽出)、標本は地域的にはらまかれることとなり、一つの層別をおこなうことともなるのである。他方任意抽出というのは、あらゆる可能な抽出の場合を含む

のであるから、抽出単位を可成適当に選択する余地もあるわけである。このことは個別典型調査をわざわざ用いることを示している⁶⁰⁾

- (1) 拙稿、「標本調査法における代表性の問題」、産経論叢第三四号、昭和三十一年一二月、参照。
- (2) 拙著、社会統計学と抽出理論、教学附録Ⅰ、参照。
- (3) 例えば、M. H. Hansen, W. N. Hurwitz, W. G. Madow: Sample Survey Methods and Theory, Vol. I&II, John Wiley & Sons, Inc., New York, 1953 参照。
- (4) 註(1)
- (5) これに関連して、わが国国勢調査における抽出集計の問題がある。昭和二五年国勢調査報告第八卷最終報告書、総理府統計局、昭和三十〇年、三六一五二三頁参照。
- (6) 拙著、誤差法則と抽出理論、関書院、昭和四四年、一〇四頁。
- (7) 比推定の数学的理論およびその応用については、例えば W. F. Deming: Some Theory of Sampling, John Wiley & Sons, Inc., New York, 1950 および註(3)の書を参照。
- (8) 前掲拙稿参照。なお地域抽出についての代表標本の作製については、米国における Master sample の研究・設計が注目される。これについては、A. J. King & R. J. Jessen: "The Master Sample of Agriculture," Journal of the American Statistical Association, March 1945, Vol. 40, pp. 38-56 を参照。
なお地域抽出の実例については、L. G. Allbaugh: Crete, A Case Study of an Undeveloped Area, Princeton University Press, 1948, Appendix: Statistical Methods Sections by R. J. Jessen & N. V. Strand を参照。抽出集計における系統抽出については註(5)の報告参照。
- (9) 拙著、社会統計学と抽出理論、一四〇頁、および前掲拙稿を参照。

五 五 五 五 五

以上において私は標本調査法を代表標本を基軸とした統計調査代用法とし、その基本問題について考察した。最初に統計調査法の性格を論じ、それが運動している社会現象の数量的把握のために、その社会現象の断面を考え、構成単位の集団とみてその数量的表示を求めらるるものであることを指摘した。そうしてこの観点から断面の縮図的表現——代表標本——を求めようとする標本調査法が、外延的標本調査法と内包的標本調査法に分たるべき必然性を論じた。前者はいわゆる任意標本調査法であり、後者は個別的典型調査法である。ついで任意標本調査法の基礎となつてゐる数学的確率論による抽出理論を検討し、これが實際調査とどう結びついているかを考察し、いわゆる代表性の質的・量的意味を、標本平均の不偏性と分散に求めたのである。

こうして標本調査法は一応二つに分たれるのであるが、本来の統計調査の目標より見ると、この両者は互に経緯となつて密接に結びつくべきものである。これを実際の調査に即していえば、数次の層別の後においては、任意抽出の一つとして系統抽出をとることが多いが、その選ばれた単位についてこれを個別的・典型的に調査すればよい。あるいは層内では任意抽出であり、その限りでは適当な抽出単位（従つて構成単位）を選んで一向差支えがないのであり、これによつて任意標本調査としては、問題とする推定値を信頼限界内において求めうるし、他方選ばれた単位の動態的な調査により、その運動法則を量的に求め、あるいは更に質的に理解しうることとなるのである。

(一九五七・二・二〇)