

# 統計利用論からみたマイクロデータ

吉田 忠（京都橘女子大学・文化政策学部）

## 1. はじめに

本稿の目的は、いわゆるマイクロデータの利用をめぐる問題に関し経済統計学会が積み上げてきた統計利用論の視角から検討を加えようとするところにある。ここで取り上げるマイクロデータとはどのようなものか。それを、現在開示されつつある「マイクロデータ」に関してその関係者が与えている定義に見てみよう。まず工藤弘安氏によれば「統計調査または統計報告の目的に限って収集、保有される記録であって、個体に関する行政上の決定に利用されないもの（で）、通常は固体識別子が除去されている「統計記録」・・・と同等に扱（えるもの）」である。<sup>(注1)</sup> 一方、川上宏二郎氏は、一般に多様な意味で使われているが、「（調査票、申告書等の個々の資料たる）『原調査票』から個人識別情報をなんらかの形で除いたものを意味するよう」（だ）とされる。<sup>(注2)</sup> これらによれば、現在開示されつつあるマイクロデータとは、記入・点検済みの調査個票ないし行政記録個票が分類集計される前の転記された中間的形態のまま国に保管されているもので、それから個人識別指標を除去したものの、ということになる。確かにその通りであるが、マイクロデータ一般を統計利用論の視角から検討しようとするときにはやや形式的であるように思われる。例えばマイクロデータの具体的形態、すなわちそれは全数調査によるものかそれとも標本調査によるものか（現行のマイクロデータは標本調査資料からのリサンプリングデータであるが、全数調査資料の特定地域に関する全数データ、標本調査資料の全国・地域での全数データ等が考えられる。）、全国レベルか地域レベルか（後者は特に農林業・地域産業等の分析の場合に求められる。）、単年度利用に止まらず時系列としての利用が許されるかどうか、さらに進んで同一対象の時系列的利用は可能か（そのときいわゆるパネルデータ利用との間にどのような異同が生ずるか。）、等々の問題がそこに残されているからである。

しかしここではこれらの問題を措いたまま、現在利用可能なマイクロデータを取り上げて、それに統計利用論からの分析を加えることにしたい。

## 2. マイクロデータと統計利用論

マイクロデータを取り上げようとするときは、まずマイクロデータ利用が統計利用一般から区別される特質を確認する必要がある。それはなによりも、マイクロデータでは利用者自らが調査票のデータから（厳密には、点検済み個票を照査表等に転記したデータから）分類、集計、製表をなしうるところにある。そうだとすると、分類、集計、製表等の過程が統計作成全体の過程＝統計調査過程の中でどのような位置・役割を占めているかという問

題を検討する必要性が生じてくるであろう。すなわち、統計利用の見地からする統計調査過程の分析・検討が必要になってくる。このような統計利用のための統計調査論を最初に取り上げ検討したのは、故蜷川虎三氏であった。報告者も所与の課題に迫る方法論をまずそこに求めようとしている。

ただし統計を利用しようとしている者の関心や問題意識、視角と方法は多様である。そこで、とりあえず統計利用者を次のように分類し、それぞれの場合でマイクロデータ利用が持っている問題を検討することにしたい。

- |   |
|---|
| <p>A 市民の利用<br/>市民・勤労者が統計指標を通してその仕事や暮らしに関連する社会的集団現象を理解、判断したりする場合。</p> <p>B 研究者の利用 I<br/>研究者が主に統計的規則性の数理的把握の方法（＝数理統計学）を利用して社会現象のモデル分析を進めようとしている場合。</p> <p>C 研究者の利用 II<br/>研究者が、社会諸科学での実証的分析のために、そこでの分析成果と社会的集団現象の統計的規則性とを基に統計指標を作成する、あるいは既存の統計指標を検証、評価したりする場合。</p> <p>D 官庁の利用<br/>これは官僚が所管する問題の分析、立案、評価等を上記の諸方法で行う、あるいはさらに新規の統計調査の実施や既存統計の個票再集計を行ったりする場合。</p> |
|---|

このような利用形態を前提において、蜷川統計学での統計利用に関わる統計調査過程論を（現代的な解釈ないし吉田の理解に若干引きつけながら）見てみることにしたい。（注3）

### 3. 蜷川統計学における統計調査論と統計利用論

#### (1) 蜷川統計学における統計調査過程論

蜷川統計学ではダイナミックな過程として一般的に捉えられる社会集団を「大量」とよぶ。大量は集団であるからその構成要素を持つが、これが「単位」である。またそれは集団として集団的性質を持っているが、この集団性は大量の部分集団が共有する「標識」として表れる。人口という大量で言えばある領土に常住する個々の人間が単位であり、性別や年齢階層が標識である。さらに動的で一般的な大量を量的把握可能な集団とすべく地理的・時間的な限定を与えねばならない。これが「存在の時間」と「存在の場所」である。このように規定されたものを蜷川統計学では「大量の四要素」とよぶ。しかしこの大量の四要素は社会集団の理論的・実証的分析成果から理論的に規定されたものであって、そのままでは実際の統計調査設計の枠組みとはなりえない。そこで、現実の社会集団現象と統計調査の社会的条件とを前提に、実際の調査設計の枠組みとなりうるような「大量観察の四要素」が単位（調査対象）、標識（調査項目）、時間（調査の時点・期間）、場所（調査範囲）に関して具体的に構成されねばならない（括弧内は具体化された四要素）。こうして大量観察の四要素が具体的に構成されると、次はそれを基に調査票の作成や調査機構の確定に始

まり調査票の配布・回収、点検・集計、転記・製表・印刷をへて統計資料公刊に至る具体的な統計調査過程が進められる。

蜷川虎三氏は、大量の四要素の確定から大量観察の四要素の規定に至る過程を「大量観察の理論的過程」、大量観察の四要素の規定から統計資料の印刷・公刊に至る過程を「大量観察の技術的過程」とよんだ。

以上が蜷川統計学における統計調査過程論のあらましであるが、ではそれは統計利用者の問題意識や統計利用方法と如何に関連しているのだろうか。統計利用者には、調査対象や標識として統計調査が捉えたものが自らの問題意識において求めているものとどれだけ一致しどれだけ乖離しているかという問題と、統計調査が捉えたある標識で分類される対象の数量が真実の数量とどれだけ食い違っているかという問題がある。前者を蜷川虎三氏は「統計の信頼性」とよび、それは大量観察の理論的過程に起因するとした。後者が「統計の正確性」であって、大量観察の技術的過程に起因するとされる。まさに統計利用者のための統計調査過程論であることが見てとれるであろう。

## (2) 蜷川統計調査過程論への批判

このように蜷川統計学における統計調査過程論は、統計作成者ではなく統計利用者の立場からのものであった。加えてその統計利用者の問題関心は社会の長期動態的な問題を批判的に捉えようとするものに関わっている。だから蜷川統計学が体系化された直後、そこでは「大量のイデオロギー的性格を強調する」のあまり「其の大量観察法の理論は余にも『大量』的であって、方法的手続きの理論はその大量の理論の背後に押狭められて居る」という批判が杉栄氏からなされている。<sup>(注4)</sup> 一方 1960年代半ばには、大屋祐雪氏がご批判をふまえながら「方法的手続き」を含みつつもそれを越えた社会的な諸関係のなかに位置づけられた形で統計調査過程を客観的に捉えるべきであると主張した。<sup>(注5)</sup>

これらの批判は確かに蜷川統計調査過程論の弱点ないし不十分なところを突いている。例えば豊富な実務経験を基に執筆された藤田峯三『統計調査の実務』を見ると、統計調査の企画・設計から実施・結果の公表までが詳細に説明されているだけでなく「統計調査の誤差」や「統計調査環境」も詳しく述べられており、実際の統計調査の理解に資するところ大である。<sup>(注6)</sup> また長くわが国の官庁統計作成を指導してきた工藤弘安氏の『入門統計学—官庁統計の作成と利用—』は、統計制度に重点をおきながら統計調査過程を説明しているだけでなくその結果の集計過程が詳しく述べられており、現在開示されているマイクロデータの理解にきわめて有用である。<sup>(注7)</sup>

前者で取り上げられている統計調査環境の問題、後者を特徴づける統計制度・統計政策の問題は、大屋祐雪氏の主張と関連するないしその展開として見ることができるが、いずれもその補正・補充として蜷川統計調査過程論に包み込むことは可能であるように思われる。そして、藤田峯三氏が分類した「統計調査の誤差の種類」の表につけられている「調査票の設計上の誤差はここでいう誤差には含めていない」という（注）は、蜷川統計学における「統計調査の理論的過程」、「統計の信頼性」という概念の有効性を改めて教えてくれるように思われる。事実、経済統計学会では、大屋批判を受け止めそれで蜷川統計調査過程論の内容を豊かにする方向が多く見られる。

それら諸見解の特徴をまとめて要約すると次のようになるであろう。<sup>(注8)</sup> イ、「大量」を

社会問題を内包した動的な過程である集団現象として捉える、ロ. それを既存の理論的実証的分析成果を通して社会構造 (=社会集団) として概念化し、そこでの社会問題を抽象する (調査目的の確定)、ハ. 調査目的を前提に社会構造に四要素を与え、統計集団を理論的に形成する (統計集団の四要素の確定)、ニ. 現実の人的・物的な調査環境・調査機構と統計調査の技術条件等を前提に統計調査集団を確定する (統計調査集団の四要素の確定)、ホ. 統計調査集団に基づき調査票の内容、(予備調査と本調査の) 実査方式、集計方式、表示方式を確定する、ヘ. (予備調査をふまえた本調査での) 調査票の配布・記入・回収・点検、集計・製表・印刷・公表。そして、イからホ迄の過程で生ずる統計の信頼性の問題と、への過程で生ずる統計の正確性の問題とを分け、さらに両者それぞれをそれが発生する個所に注目しつつ分類しようとするのである。

### (3) 蜷川統計学における統計利用論

蜷川虎三氏は統計調査過程の分析を通して統計の信頼性、正確性という概念を導出したが、蜷川統計学の統計利用論はそれとは別のところにあった。そもそも蜷川虎三氏の本意は、まずドイツ社会統計学を基盤に置いてそれに数理統計学を包摂させようとするところにあった。彼は、母集団 - 標本モデルのように方法的要請から構成される集団ではなく、存在する集団たる「大量」を対象に行われる「大量観察」とその結果たる「統計」から出発する。その統計の特定指標に関して時系列を構成し、もしそこに安定的な数量的規則性が表れたときそれを統計的法則として捉えることを統計学の課題とした。その結果、統計的法則を見出すために構成される集団、統計調査の対象たる存在する集団という二元論的構成に陥ることになったのである。<sup>(注9)</sup>

しかしこの体系とは別に、統計の信頼性、正確性の概念は統計の利用における武器としての役割を果たすことができる。これを、先に本稿(2)で分類した統計利用者のタイプごとに見てみよう。統計の信頼性、正確性が最も有効な役割を果たすのは、Aの場合、すなわち市民・勤労者がその生活・仕事のために統計指標を単独であるいは組み合わせて利用する場合であろう。統計の信頼性・正確性を基に官庁統計の統計指標を批判的に見ていくことの重要性は、何度繰り返しても繰り返し過ぎることはない。Bの研究者による数理的利用では、前段の信頼性・正確性は予備的なもの、後段の統計的法則が最終ゴールというように逆転させられたときの蜷川統計学の二元論的構成が方法として有効になるであろう。

問題はCの研究者の社会科学利用における有効性であるが、そこで信頼性・正確性は方法の重要な前提ではあっても決して方法の総てではない。経済統計学会は創立以来、そこでの方法としての体系化に全力を挙げてきた。そして、官庁統計での分類(分類基準、その組合せ、分類階層区分等)に対する批判、さらに公表された統計の組み替えによる再構成(大橋隆憲氏の階級構成表がその代表例)等で一定の成果が見られたが、体系化は未だしといわざるをえない。

この点に関し、問題点を二つ指摘したい。一つは、官庁統計での分類の批判と再構成に対しては、将来もしその開示が十分に進められるならばマイクロデータの利用が、大きな役割を果たす、そして問題の解消に大きく貢献する可能性を持っていることである。二つは、蜷川虎三氏がその大量観察法の冒頭で大量を理論的に規定する枠組みを「社会科学の理論」としたことである。その結果、統計利用は、抽象的な理論で具体的な事実を説明する「上

向過程」において行われるものという理解が多くなった。しかし統計利用は、実証分析の中間段階で例えば発展段階や地域類型の検出を通してより一般的な理論に近づくというような「下向過程」においてより有効となるものではないだろうか。

#### 4. 統計利用論から見たマイクロデータ

##### (1) マイクロデータにおける統計の信頼性・正確性

以下、蜷川統計学の統計調査過程論ないし統計の信頼性・正確性の方法論から、現在わが国で開示されつつあるマイクロデータの利用に関してその可能性と問題点を指摘したい。

まずはっきりさせねばならないのは、統計調査の理論的過程の（統計調査集団の四要素の確定に伴う）企画・設計段階で、調査票の設計・実査方式の確定だけでなく、集計・表示の様式までが確定されること、そしてその上で実査過程が始められることである。このことは、先に取り上げたわが国の指導的統計家の著書が示す通りである。

公表された官庁統計における分類（分類基準とその組合せ、分類階層区分）は、経済統計学会会員の批判をまつまでもなく問題点をはらんでいる場合が多い。この統計の信頼性の問題は、その統計のマイクロデータが開示され、利用者がそれを用いて集計の「やり直し」をすることで解消される可能性がある。しかし、そこでの信頼性の問題の総てが解消されることはありえないであろう。なぜなら、調査・企画の段階においてある調査単位といくつかの調査項目が（統計調査集団の四要素の初めの二者として）確定されるが、それから除外された調査項目を集計「やり直し」の分類基準として取り上げることはいうまでもなく不可能である。ましてや取り上げられなかった単位に関しては、如何ともしがたい。また、調査が対象とする時期・期間や地域・場所においても程度の差はあれ同じ問題が存在するのである。

だから、開示されるマイクロデータを用いた集計の「やり直し」に際しては、どのような問題をどう深めたいのかという利用者の問題意識を明確にすることと共に、統計調査の理論的過程を分析して、その統計調査のどの項目をどのように再集計すれば利用者の問題意識の解明により有効であるかをはっきりさせることが重要である。そこにマイクロデータがあるからとにかく利用してみるということでは、既存公表統計の利用以上の成果を期待することは困難であろう。

もちろん、問題解明により接近するように見えた集計の「やり直し」を第一次的に試み、その成果の不十分さをふまえつつ別の分類基準や別の項目での再分類を繰り返すという方法もありうる。この繰り返しを通して、われわれは、統計調査集団の確定、すなわち調査単位、調査項目さらには調査の時期・地域の確定における不十分さ、問題点を明らかにできる場合がある。このとき、われわれは官庁統計（の企画・設計）に対してその是正につながるような有効な批判を加えることができるようになる。ただし、その不十分さや問題点が、統計調査の人的・物的な調査環境・調査機構、調査方法の技術的条件等に規定されている場合があることを忘れてはならない。

次に、統計の正確性の問題である。これに関しても、調査票の配布・記入・回収から点検・転記に至る統計調査の技術的過程で生じた正確性の問題（調査誤差）は、マイクロデータに

とつても所与である。集計の「やり直し」によって回復できる正確性は、分類集計・製表・印刷段階で生じるものに止まるが、それはコンピュータ集計のもとでは前者の正確性に比べてはるかに小さいであろう。

こうして、マイクロデータ利用における信頼性・正確性の問題は、主に集計の「やり直し」＝再分類をめぐる信頼性の回復に関わっている、ということができよう。

## (2) まとめ

マイクロデータ利用による集計の「やり直し」によって期待される効果と問題をまとめてみる。まずA市民的利用であるが、もし集計の「やり直し」によって、人々がより強く求めている新たな統計指標を導出することができた場合、あるいは重要な統計指標に関して、それと関連するの指標を検出しないしそれを規定する要因指標に分解することができたりした場合、人々の統計指標を通した社会的集団現象に対する理解はより深くなりまた判断もよりの確になるであろう。ここでは統計の信頼性の検討が重要な前提をなす。

B研究者的利用 I では、重要な多数の変量を持つ膨大な数の標本が利用可能となり、そこから要因間の新たな関連の検出と結びつくモデルや分析方法の開発が可能となる。現在、マイクロデータ利用の直接的な成果が最も期待されるのはこの分野であろう。C研究者的利用 II であるが、当面、マイクロデータ利用による集計の「やり直し」を進めることで、社会の実証分析における新たな類型・段階等を示すような統計指標体系を構築していくことが期待される。しかしこの方法が最も有効なのは、社会の構造的な変動そのものを統計指標体系において捉えようとする場合である。だからマイクロデータも時系列としての利用、さらに言えば同一標本の時系列的利用すなわちパネルデータとしての利用が望まれる。このパネルデータとしての利用は、B研究者的利用 I ではさらに強く求められるであろう。しかしパネルデータとして利用しようとする、個人識別情報の消去というマイクロデータの基本前提との矛盾が生じてくるから、それはおそらく不可能であろう。しかしいずれの場合も、今後、どのような形でマイクロデータの開示が進むかということが問題である。

最後にD官庁的利用であるが、上に述べた各種利用形態で指摘した集計の「やり直し」が持つ可能性と問題点はそこに共通して存在している、といえるであろう。加えて、現在、各省庁の企画・調査部門で行われている調査個標の再集計の結果が十分に公開されていない、という問題点を指摘しておきたい。

## [注]

- (1) 参考文献 (1) [I]、169-70 ページ。
- (2) 同上、309 ページ。
- (3) 以下の蜷川統計学の内容については、主として参考文献 (2)、(3) によった。
- (4) 参考文献 (4)、引用は 42 ページ。
- (5) 参考文献 (5)、(6)。
- (6) 参考文献 (7)。
- (7) 参考文献 (8)
- (8) 参考文献 (9)、(10)、(11)、(12)。
- (9) 参考文献 (2) の内海庫一郎氏の「再刊にあたって一本書の特長と歴史的意義」

及び参照文献(13)の「統計学の古典 蜷川虎三『統計学概論』(吉田忠)」参照。

[参照文献]

- (1) 松田・浜砂・森編著『講座ミクロ統計分析』[1], [2], [3] 日本評論社, 2000
- (2) 蜷川虎三『統計学概論』(第12刷) 岩波書店, 1985、なお第1刷は1934
- (3) 蜷川虎三著・横本広現代語訳『統計利用における基本問題(現代語版)』産業統計研究社, 1988
- (4) 杉栄『理論統計学研究』立命館出版部, 1940
- (5) 大屋祐雪編『現代統計学の諸問題』産業統計研究社, 1990
- (6) 大屋祐雪『統計情報論』九大出版会, 1995
- (7) 藤田峯三『統計調査の実務—企画から実施まで—』(株)ぎょうせい, 1986
- (8) 工藤弘安『入門統計学—官庁統計の作成と利用—』(財)全国統計協会連合会, 1997
- (9) 大橋隆憲・野村良樹『統計学総論(新訂版)』有信堂高文社, 1980
- (10) 伊藤陽一『統計学』法政大学通信教育部, 1981
- (11) 大屋・野村・広田・是永編著『統計学』産業統計研究社, 1984
- (12) 吉田忠『農業統計の作成と利用』農山漁村文化協会, 1987
- (13) 吉田忠編著『現代統計学を学ぶ人のために』世界思想社, 1995