

オケージョナル・ペーパー No.43

データ統合の視点から見た調査票情報の意味について

2014年8月

法政大学

日本統計研究所

データ統合の視点から見た調査票情報の意味について

森 博美

目次

はじめに

1. 表式調査と個票調査による調査票情報の情報形式の相違
2. 調査個票記載情報
3. 個体レコードデータとして見た調査票情報
4. 「統計＝集計量」という統計観の根拠
5. 調査票情報としての統計原情報を規定する時間的・空間的要因
6. 時間的要因の作用
7. 統計単位への場所的要因の作用
8. 場所的要因の面的取り扱いとしての地域表章
9. 場所的要因の点的性格
10. 調査票情報の拡張(1)－data carrier 情報によるデータの統合
11. 調査票情報の拡張(2)－経緯度情報による調査票情報への場所的特性変数の付加
12. 調査票情報の拡張(3)－経緯度情報によるデータの統合

むすび

はじめに

統計法(平成19年法律第53号)第2条11項は、「調査票情報」を、「統計調査によって集められた情報のうち、文書、図画又は電磁的記録(電子的方式、磁気的方式その他他人の知覚によっては認識することができない方式で作られた記録をいう。)に記録されているもの」と定義している。これによれば調査票情報は、第1に統計調査によって集められた情報であること、第2に記録媒体としては文書あるいは電磁的記録といった形態を持つものとされる。なお統計法の条文では、図画すなわち図面や絵・写真等が文書や電磁的記録と並列に掲げられている。

しかし、条文に並列表記されている図画と文書や電磁的記録とはややその性質を異にするように思われる。なぜなら図画がデータの表現形式でありいわゆるテキスト情報だけでなく図面や映像(絵、写真)なども調査票情報に含まれるとの整理であるのに対し、文書と電磁的記録は単に調査票情報の記録媒体について述べたものだからである。

また上記の条文は、統計調査によって収集される情報の形式、すなわちそれらが集計量かまたは統計単位毎の個体情報であるかは特に規定していない。なお調査票情報が統計調査によって集められた情報であるということは、情報の源泉が調査以外の例えば行政の遂行過程で作成、把握される行政記録から転記された行政情報は調査票情報の対象外であることを意味する。従って、仮に個体に関する記録であってもそれが行政的記録を情報源とする限り統計法での調査票情報には該当しない。

本稿でも統計法の規定に従い、調査票情報を統計調査によって収集される一連の情報(統計原

情報)として捉えることにする。そして本稿では、統計原情報を最初から集団に関する統計的記述として集計量というデータ形式で把握する表式調査から、個人、世帯、事業所あるいは企業といった統計単位を個票方式による非集計量というデータ形式で把握する調査方式への統計調査法の史的展開について、従来のような統計作成技術からの評価ではなく、統計情報論的観点から新たな今日的意義づけを試みてみたい。

1. 表式調査と個票調査による調査票情報の情報形式の相違

明治初期、わが国の黎明期の政府統計は、基本的に表式調査によって収集された統計原情報を資料として作成されていた。筆者がすでに〔森 2013g〕で論じたように、表式調査にも調査方式としての発展段階がある。わが国の生産統計の嚆矢とされる物産表を例にとれば、初期の調査では全国統一の報告様式もなく、単に調査事項のみを記した報告要請が政府から道府県等に対して命じられただけであった。そこでは、表式調査票への計上に際しての計量単位や把握期間などはすべて報告者の裁量に委ねられていた。統計基準の地域間での不統一は、町村、市郡、道府県という報告系統を通じての積み上げ集計の障碍となり、収集された統計原情報から道府県あるいは全国レベルでの集計ができないという統計の本質的欠陥を露呈することになる。こういった統計作成に関連する諸問題に対して、価額あるいは数量といった計量単位の統一、把握時点や対象期間の統一、報告期限の設定、さらには全国統一の報告様式の導入といった形で、表式調査においても次第に調査法として整備が進むことになる。

鮫島龍行は、明治16年12月の「農商務通信規則」(農商務省達第21号)を、「前近代的調査方式」としながらも「そのひとつの完成図」〔鮫島(1971) 57頁〕と評価している。同規則の施行により、表式調査票による統計原情報の収集が行われることになる。表式調査の調査票の最大の特徴は、調査票の設計が作成する統計の最終形としての集計表を最初から想定したものとなっており、調査票への記入者が調査事項に関して集落等境域内の計数を最初から集約値(集計量)として様式の各セルに記載する点にある。その意味では、表式という調査票を用いて遂行される統計調査による統計原情報の収集は、その後の集計過程とまさに一体化されたものとして行われている。

全面的に表式調査として実施されてきた物産表(その後、農産表)調査に部分的に「工場票」、「会社票」という小票(単記票)が導入されたのは、明治27年の「農商務統計報告規程」(農商務省訓令第17号)においてである。この報告規程による小票の導入について鮫島は、「この小票の採用は、明治期を通じて統計情報の支配的な収集形式であった表式調査から近代的な調査票形式、いわば点計調査形式への最初の移行を意味していた」〔同 59頁〕と評している。

それでは、表式調査の完成期に見られる表式調査票によって収集された調査票情報は、今日、統計調査において広く採用されている個票(小票、単記票、単名票)を用いて収集されたそれとどのような点で異なっているのであろうか。

上述したように、表式調査では集計表を想定して調査票が設計されている。このことは、収集される統計原情報が、個々の統計単位に関する個別情報ではなく、最初から集落等一定の地域集団に関する集計値として把握されることを意味する。このような集計量としての統計原情報の収集方式に対して、これまでは主として把握される統計原情報の精度あるいは多様な集計の可能性といった観点から表式調査が持つ限界性が指摘されてきた。

例えば、鮫島は表式調査について、「統計表の形式さえととのってれば、その数字がどんな手続きでえられたかを問わないし、また問うこともできない方式」〔同 27 頁〕であり、「調査の過程に誤りがあってもそれを訂正することはできない」〔同 27 頁〕調査方法であると評している。このことは、統計原情報が集計値として把握されることから、仮に集計処理の際に調査票の記載内容に疑義が生じたとしても、集団を構成する個体にまで遡って記入の適否を確認する術がないことを意味する。

また表式調査によって収集された統計原情報が集計面で持つ限界についても鮫島は、それが「歴史的には、表式調査は集計組織の整備されない初期段階に適応した方式」としながらも、「所定の表形式以外の組み替え表を作ることはできない」〔同 27 頁〕のものであり、「この方式では調査項目を各種の目的に応じて自由に製表することはできない」〔同 343 頁〕と指摘している。すなわち、表式調査が結果表の表形式に従って、しかも最初から集計量として統計原情報が収集されている故に、表式が許容する項目間の統合再集計以外の新たな結果表章を行うことはできないのである。

ところで、わが国における近代政府統計の高祖とされる杉享二は、明治 2 年に現在の静岡県で実施した駿河国人別調ですでに世帯票を調査に用いている。またわが国の国勢調査の先駆けとされる甲斐国現在人別調においても、家別表(世帯票) 11 万枚を印刷し各町村において戸籍に基づいて世帯構成員の人口に関する諸属性の把握が行われている〔高橋 267 頁〕。なお、甲斐国現在人別調の集計作業は、家別表から転記筆写して作成した単名票(小票)によって行われた〔同 268 頁〕。

筆者は〔森 2014c〕において、明治 31 年の人口動態統計への小票の導入の意義とその特徴について論じた。そこでも指摘したように、表式調査から個票調査への移行という統計調査方式の史的展開が持つ意義について、これまでは一般に把握の適否の事後的検証可能性あるいは多様な集計の可能性の確保という統計作成技術論的な視点から専ら評価されてきた。

調査個票は、個々の調査事項(項目)に関して、統計単位毎に統計原情報の収集を行うための様式である。その意味で個票調査とは、個人、世帯、事業所、企業といった個々の統計単位から個人票、世帯票といった様式によって個体ベースで統計原情報を収集し、それを集計することで統計を作成するものである。

個票調査の場合には、調査票を介して個々の統計単位に関する統計原情報が収集される。世帯票がその中に世帯ならびに個人に関係する調査項目を併せ持っている場合にも、世帯に関する調査項目に関しては各世帯と世帯票との間には 1 対1の個的対応関係が成立している。また世帯票に記載された個人に関係する項目については、それらを転記あるいはカード入力等の機械的操作によって、世帯情報を共通項として持つ個々人の情報に切り分けることが可能である。このような統計原情報の個体への切り分けは個票を用いた調査の場合に初めて可能となるものであり、表式調査票によって最初から集計量として収集された統計原情報がいかなる事後的操作処理によっても個々の統計単位に還元できないのとは本質的に異なる。

統計単位に関する統計原情報を統計単位に関して非集計的な情報形式で把握する個票調査による調査票情報には、統計の集計表としての結果表章すなわち「統計＝集計量」としてきた従来の統計(統計学)の在り方を根底から問い直す契機となる論点が含まれているように思われる。そこで以下では、表式調査から個票調査への統計調査法の展開について、調査の精度あるいは集計

といった統計作成技術的な観点からではなく、調査票情報が把握した統計単位に関する統計原情報の情報特性という視点から捉えることで統計の新たな展開可能性について考えてみることにしたい。

2. 調査個票記載情報

統計調査の実施者が調査対象となった統計単位から統計作成に必要な統計原情報を収集する媒体が調査票である。これまで一般に紙面や冊子体での調査票が調査に使用されてきたが、現在の統計調査では、電話、メール、ウェブと調査方法の多様化を受けて様々な形態の調査票が用いられている。メール調査の場合にはファイル形式での調査票が統計単位の記入担当者宛に添付送付され、ウェブ調査ではウェブ画面上に示された調査票に報告者側が必要な記入情報を直接入力して送信報告することになる。また、電話調査の場合にも、インタビューアである調査員の手元には調査票にあたる質問事項一覧資料が用意されている。電話調査やウェブ経由での調査の場合にも、統計単位からの統計原情報の収集という調査票の機能面に限れば、そこで調査に用いられている様々な形の調査票は、従来の印刷物としての調査票様式と本質的に何ら変わるものではない。

ここで、統計調査において一般に使用されている調査票の記載事項をながめてみよう。調査票には、統計単位から統計原情報を収集するための調査項目だけが列挙されているわけでは必ずしもない。そこには調査項目以外にもいくつかの記載欄が設けられている。

(1) 調査項目以外の記載情報

まず、調査実施者側で調査実施に先立ち記載を行う事項として、都道府県コード、市区町村コード、調査区番号といった地域コードがある。一方、調査票の報告者側でも、調査事項以外にもいくつかの記載欄が設けられている。事業所・企業調査の調査票には、事業所番号や企業番号、それに記入者情報として記入者名や連絡先電話番号の記入欄があり、一方、世帯調査の調査票には一連世帯番号などが設けられている。

調査票によって収集されるこういった統計調査項目以外の各種記載事項の統計作成面での用途は様々である。まず、市区町村コード、調査区番号、単位区番号は、都道府県や市区町村といった行政区による地域集計さらには小地域集計を行う際の境域を特定する際のキー変数として使用されている。

上述した一連の記載事項のなかで、記入者氏名と連絡先電話番号は、収集した記入済調査票の審査過程で調査項目に対する回答に不備（欠測）、過去のデータからの著しい乖離、あるいは回答内容の間の論理矛盾等が発生した場合などの際の照会先情報として記入者に記載が求められているものである。こういった連絡先情報は、統計作成に直接結びついた統計原情報の収集ではなく、収集された統計原情報の品質の改善を通して作成される統計の質の向上を担保する目的で調査票に設けられているものである。

なお、わが国の場合には分散型統計制度の下で、主管の各府省がそれぞれの予算で統計調査を実施してきた。そのためこれまでは、統計単位である個々の企業や事業所について、調査毎に固有の事業所・企業番号が付与されてきた。それらはあくまでも個々の調査ベースではあるが、個々の統計単位を特定する識別情報ともなっている。なお、近年わが国でも、事業所母集団デー

データベースの構築を受けて、ようやく共通事業所コードを公的統計調査間で共有する方向での取り組みが開始されている。

(2) 統計調査項目

集計され統計として結果表章に用いられる統計調査項目の中には、調査員が統計単位からの聞き取りにより記入(他計)するものあるいは調査客体自らが記入報告(自計)する統計単位に関する統計原情報と、統計単位からの情報提供に依存することなく調査員(あるいは調査実施担当者)が目視等により独自に記入を行うことによって把握される情報とがある。

(i) 調査客体からの報告・回答事項としての調査項目

調査票に統計の調査項目として記載されている事項は、個々の統計単位についての諸属性に関係する事項と個々の統計調査に固有の調査項目とに大別される。このうち前者は一般にフェイス項目の名称で呼ばれるもので、事業所等の名称、所在地、電話番号、世帯員の氏名、個人の性や年齢、資本金や従業員数といった個々の統計単位の諸属性に関する統計原情報を収集するための調査項目がそれである。このうち事業所や企業の名称は、事業所データを名寄せすることによって企業データを編成するのに用いられてきた。このように、これらの情報は、いずれも結果表章の際に集計変数として使用されたりあるいは統計作成に際しての補助情報として用いられる。

一方、統計調査に固有の調査項目では、個々の統計単位の活動、行動、意識といった統計単位のパフォーマンスに関する統計原情報の収集が行われる。

(ii) 調査実施者側で独自に把握・記載する統計項目

集計処理され結果表として公表され一般の利用に供される統計の中には、調査客体側からの統計原情報の提供を受けることなく調査員等の調査実施者側で独自に把握した情報に基づいて作成されるものもある。例えば、住宅・土地統計調査には、世帯に配布する調査票(甲、乙)以外に調査員が独自に建物の外観等の目視により把握、記入する建物調査票がある。この他にも、この調査では住戸等の立地特性として調査区ベースでその地区の建蔽率、容積率、最寄りの交通機関や公園等までの距離などの情報が調査実施機関あるいは市区町村の統計所管部門で別途作成され、調査票情報に追加情報として付加され結果表章に使用されている。

3. 個体レコードデータとして見た調査票情報

現行の統計作成過程を見れば、調査によって収集された調査票情報はレコードファイルとして一旦格納され、論理チェックや異常値の点検処理等のデータクリーニングを経た後に集計処理過程にまわされる。そこでは、個々の統計単位に関する調査票情報は、それぞれが1レコードからなる個体レコードといったデータ形式を持つ。なお、調査記録情報として磁気媒体上で維持管理されている個体レコードの符号表(レコードフォーマット)によれば、そこには調査識別コードや調査実施年といった調査情報と調査票情報から転写された統計原情報、さらには調査従事者等による独自の付加情報などが記録されている。

公的統計の作成業務は、調査の企画から集計結果の公表までの一連の業務工程からなる。統計法第9条第2項ならびに第19条第2項は、基幹統計調査と一般統計調査の実施に際して集計事項を含む調査計画の総務大臣による承認を義務づけている。集計技術が未発達な段階においては集計処理に係る業務量や作業時間の関係で、調査実施規則として予め承認を受けた一連

の集計表だけが作成、公表されてきた。その後の情報処理技術の進歩により比較的容易に多様な集計が可能となったことから、結果報告書に非収録の閲覧用集計表、小地域統計の拡充、さらに最近ではオンデマンド型の集計処理サービスの提供や公的統計のポータルサイトとして開設されている e-Stat のように、データキューブの形でデータベース化され利用者が一定の範囲でクエリーによる再集計できるものもある。このように、多様な提供チャンネルによって各種集計表を提供することで利用者のいろいろな集計ニーズへの対応が図られている。

(1) 擬似的な集計レコードとして見た集計表

(i) 統計の公表形態としての集計表

印刷物として公表される個々の調査の結果報告書には、いろいろな統計項目を集計変数とした結果表が収録されている。また、『日本統計年鑑』等の総合編集統計書や『社会生活統計指標』等各府省が独自に編纂、公表している編集統計書にも、分野別に様々な集計表が収録されている。

個々の調査の結果報告書やこれらの編集統計書を眺めると、そこには例えば個人を統計単位とする調査の場合には、性や年齢階級、学歴等の、また世帯については、世帯規模、所得階級、世帯主の年齢階級等の、さらに事業所を統計単位とする調査の場合には従業者規模等の、そして企業については資本金階級等を表側の集計変数として持つ様々な結果表が収録されている。通常、マクロデータ(集計値)の利用者は、それぞれの分析目的に照らして個々の結果表の表頭、表側に注目して統計表中の各セルの結果数字を利用する。こういったマクロデータの利用者は、当該結果表と同じ表側変数を持つ他の表との間の関係についてふだん意識することはあまりない。

(ii) 擬似的集計レコードとしての集計表

上述したように、個々の調査票情報は個体レコードというデータ形式を持つ。このようなデータ形式についての考えを敷衍することによって、このように一般に提供されている一連の集計表についても、データ形式として集団に基づく疑似的な集計レコードとみなせないであろうか。すなわち、表側に同一のカテゴリー区分や階級区分からなる変数を持つ一連の集計表について、表側の各区分に該当する集団をいわば個体としての統計単位に擬制することで、種々の集計表を横方向にバーチャルな形で連結することができる。そうすることで、表頭の各変数に該当するセル値を擬似的な集団レコードに見立てることができる。属性変数値(あるいは属性変数階級値)として個体に擬制された集団を構成し、当該属性変数値を共通に持ち表頭の各変数が与える集計量としての変数値を統計原情報とするレコードを編成することによって、各種の集計表のセル値に新たな意味づけを与えることができる。なぜなら、そうすることによってそれまで単に個々の集計表の一構成要素に過ぎなかったセル値が個体に犠牲された集団を介して連結され、集団の諸側面を体現する変数値として相互に関係づけられることになるからである。

一連の集計表を集計量による疑似個体レコードとして捉えた場合、例えば男女、年齢階級、学歴等による集計結果表は、性、年齢階級、学歴といった集団から構成される疑似的統計単位を基盤情報として、各セル値が与える集計量ベースでの特性情報をレコード形式で連ねたものと解釈することができる。また都道府県、市区町村といった行政区分さらには町丁目、調査区、メッシュといったいろいろなレベルでの地域集計表も、それぞれの地域区分を擬制的な統計単位とみなせば、各レベルでの集計値をそれらの特徴づけ相互に関連性を持つ一連の特性情報からなるデー

タレコードとして捉えることができる。全国計もまた国を個体に擬制させれば、それらを構成要素に持つバーチャルなデータレコードが成立する。事実、地方財政や国際経済等の分野で広く行われているマイクロやパネル分析は、このような地域や一国を疑似的な個体とみなすデータ観に立脚している。

筆者は、個体レコードを構成する諸変数をそのデータ論上の特性の違いから、データの担い手 (data carrier) である統計単位ならびにその諸属性に関する一連の変数とそれによって担われるデータ本体 (data body) とに区分することを提案した〔森 2009a〕。上記の集計表のレコード的解釈を data carrier と data body といった観点から捉えれば、性、年齢階級、学歴、資本金といったフェイス項目は data carrier の属性情報を与えるものである。そのような属性によって集団構成された data carrier は、属性変数をリンクキーとして一連の集計表をデータ統合したバーチャルな統合集計表の各表頭変数に該当するそれぞれの変数値を構成要素とする data body 部を担っている。同様に地域集計についても、様々なレベルの地域がいわば個体として擬制され、地域という data carrier が一連の地域別集計の変数値から構成される data body を担っている。

集計表の個体レコード的解釈について、ここでは差し当たり以下の点だけを確認しておこう。

第1は、集計表から編成される疑似的個体レコードが、あくまでもバーチャルにデータ統合されているという点である。通常、集計表は統計調査毎に作成される。こういった集計表を共通の表側変数を用いて調査間で串刺的に相互にリンクさせ、いわばバーチャルにレコード編成したものがそれである。そこでは、data body を構成する各統計値は、それぞれ data carrier 情報を共有しリンク可能性を持ったリレーショナルな一連のデータとして存在する。

第2の特徴は、ここでの data body が集計量として与えられている点である。このことを data carrier との関係で付言すれば、data carrier の一属性としての表側変数は、その実現値を平均化あるいは統合した集計量というデータ形式でもって表側の各集団を構成する統計単位としての個体に係るそれぞれの data body 変数を担っていることになる。

第3に、調査毎に与えられる集計表を表側変数というリンクキー変数を用いてデータ統合することで、一連の集計表が持つセル値からなる疑似的個体レコードを編成できる。これによって、セル値という集計量によってではあるが、data body を構成する変数値の組を、集団としての data carrier を特徴づけししかも変数相互間に関数あるいは因果関係を持つ特性値ベクトルとして捉えることができる。

このように集計表をレコードというデータ形式で捉え直すことによって、あくまでも集計量としてではあるが、これまで各種の調査結果報告や編集統計書に特に関係づけられることもなく掲げられていた一連の集計表を、利用者は改めて共通の表側を持つ一連の表頭を構成する変数ならびに変数値として意識できるようになる。

(2) 個体レコードのレコード形式

上に述べた集計表の疑似的な個体レコードとしての解釈は、あくまでも調査票情報の個体レコードデータ表記の単なるアナロジーとして提案したものである。個々の調査票情報から転写作成される個体レコードの場合、それがまさに個体としての統計単位に帰属する統計原情報を反映したものであることから、data carrier と data body の関係は一層鮮明である。そこには data carrier に関する集団の個体としての擬制という要素は存在しない。統計単位である個体そのものと1対1の対応関係にある個体識別変数あるいはIDコードが、まさに data carrier として data body 部分

を構成する各種変数の実現値を担っている。そこでは、個体レコードの data body 部を構成する一連の変数群が、個体の持つ多様な側面を相互に連携関係にあるものとして統一的に表現している。その場合、data body 部を構成する各変数と data carrier との間には、まさに 1 対 1 の対応関係が成立している。それは、集計表の個体レコード的解釈がレコード表記というデータ形式の類似性にもかかわらず、個体に擬制された data carrier さらには平均化され集計量として統合された変数値との間に成立する多対多を基にした集計量に基づく集約的対応関係という点で本質的に異なる。以下に見るように、この点の確認は、集計量とは異なるまさに個体レコードとしての調査票情報の情報特性を明らかにする際の出発点となる。

4. 「統計＝集計量」という統計観の根拠

近年、わが国でも、匿名標本データ(マイクロデータ)や統計法第 33 条による調査票情報に基づく学術分野での統計利用が支配的になりつつある。このような時代の流れを受けて統計作成機関側でも、提供する匿名標本データの種類の多様化、オンデマンド処理システムの制度化、さらには公的資金に基づく学術研究への調査票情報の提供といった対応がとられている。

とはいえ、これまで公的統計として一般の利用に供されてきたデータは、そのデータ形式としては集計表の形によるものである。このように、わが国で統計調査によって収集された統計原情報が個体ベースでの調査票情報としてではなく、集計表における集計量という形態で主に提供されてきたこと背景には「統計＝集計量」という伝統的な統計観がある。このような伝統的な統計観を支持しているのが、統計の安定性と統計に係る秘密保護という二つの要件である。

(1) 安定性要件

統計調査において中心的作業工程をなすのは実査である。それは、統計原情報の収集を目的に設計された調査票を媒介して個々の統計単位からそれを調査票情報として写し取る作業に他ならない。しかし、従来統計学で統計調査論が縷々論じてきたように、統計単位からの統計原情報の収集には様々な攪乱的要因が作用する。また仮に理想的に企画、実施された調査であっても、個々の統計単位から回答として提供される統計原情報は、多かれ少なかれ揺らぎの要素を持つ。このようなひとつひとつの回答に内在する情報の揺らぎの要素については、個体をある程度まとめた数の集団として扱うことでそれを相殺し、安定的結果(規則性)を析出することができる。古来、統計学が個体ではなくあくまでも集団を対象とし、大量観察による統計的規則性の検出を究極の目標として掲げてきたのもまさにこの点に関係している。

結果の安定性という統計に求められた要件は、利用に供されるデータの形式を自ずと規定する。個々の統計単位に関する調査結果は、調査票情報すなわち個体レコードそのものとしてではなく、より安定的な集計量としてのセル値を持つ集計表の形で提供されてきた。なお、公表用の集計表の作成に際しては、安定的な結果として提供できるか否かが集計変数の選定やクロス表の多重の程度を統計作成者側で判断するときの実質的な基準とされてきた。

(2) 秘密保護要件

調査結果がこれまで主として集計結果表として公表されてきたもう一つの有力な理由が、統計調査に係る調査客体の秘密保護の担保という要件である。業務統計と異なり統計調査によって作成される調査統計の場合、申告義務の課せられた基幹統計も含め、調査客体(被調査者)側から

の調査協力が調査実施の事実上の前提となる。そのために被調査者が調査への協力に伴う不利益を蒙ることがないように、調査実施者側では統計調査に係る秘密の保護を約束している。

個体レコードの場合、仮に様々な匿名化措置が施されたとしても、個体が識別されるリスクは集計量に比べて格段に大きくなる。その意味では、集計表による集計量形式での調査結果の提供は、統計に係る秘密の保護にとってより適格的なデータ形式であるといえる。なぜなら、個体から得られた個々の統計原情報は集計表中の各セルにおいて集計値として統合され、結果的に統計単位の個的存在に関する統計原情報は、セル値として完全に溶融しているからである。ちなみに、このような統計における秘密保護要件の充足は、集計結果表に希少事例が出現した場合の「X」表記による秘匿措置からも読み取ることができる。

5. 調査票情報としての統計原情報を規定する時間的・空間的要因

これまで統計は、調査票を介して統計原情報として収集された調査票情報を、個人や世帯、あるいは事業所や企業といった個々の調査単位そのものが有する諸側面をそのまま反映した情報と事実上みなしてきた。その結果、統計作成者は調査票情報がそのような情報特性を持つものとして調査結果を集計表として集約し、一方、利用者もまた集計量として提供されるマクロデータあるいは非集計データである匿名標本データや統計法 33 条等による調査票情報を、事実上そのような性格のものとして取り扱ってきた。

ところで、現実に統計調査の対象となる個々の統計単位は、何よりも社会的存在であり、時空間の内なる存在である。このことは、それらの存在の時空間的な座標としての場を規定する外的環境要素としての時間と空間による被規定性と無関係には存在し得ないことを意味する。このことの統計に対する含意としては、調査票情報として得られた個体レコードの **data body** 部を構成する個々の変数値が、実際には調査が実施された時点および統計単位の存在の空間的場所の要素との関連で成立する種々の時間的、場所的要因による作用というフィルターをくぐることで初めて具体的な形をとった諸変数の実現値に他ならないことを意味する。

以下の第6、7節では、統計あるいは調査票情報としての統計原情報に対する時間的要因と空間的要因の作用についてそれぞれ考察する。

6. 時間的要因の作用

調査票情報に内在する時間的要因としては、日、週、旬、月、四半期といった様々なタイムスパンに応じて発現する変動要因がある。統計調査によって調査票情報として収集された統計原情報には、実はこのような様々なタイムスパンを持つ変動要因が内在している。この事実は、例えば、年次データでは統合され潜在化していた季節変動が月次あるいは四半期データで初めて顕在化し、月次データに内在していた変動要因がより短いタイムスパンを持つ旬あるいは週ベースのデータでその変動が表面化するからわかる。曜日変動は当然ながら週以上の集計期間を持つデータからは確認できない。様々なタイムスパンと同期される変動要因は調査票情報に対して構造的に作用する。そのため時系列データの分析においては季節変動要因を特定しその系統的な除去が求められてきた。

このようないわば物理的な時間概念と結びついた周期的変動の他にも、特に経済統計分野での調査票情報には、不定期な周期性を持つ景気局面からの作用因、さらには非循環的な時代的要因もまた時間的要因として作用している。

わが国の黎明期の統計では、時間的要因は統計調査の企画者の意識に全くといっていいほどのぼっていなかった。そのため初期の統計調査では、調査時点や計上期間が明示されていない調査や行政の下部組織からの報告周期が統一的に定められていないものがむしろ一般的であった。その後、統計作成が近代国家の経常業務として認識されてくる過程で、統計調査における時間的要因は、まず統計調査における斉時性の要件として意識されることになる。それは、静態調査においては調査対象全域について統計単位の統計原情報の同時的把握のための統計調査の実施時点(reference date)として、一方、各種事象(event)の発生といった件数等の把握を行う動態統計調査においては、取り纏めの期限を定めた計数の集約結果の定期的報告として制度化される。

時間的要因は時空間の内なる存在としての個々の統計単位に対して、時々刻々異なる形で作用を及ぼす。これに対して統計調査の方は、時間軸方向に動態的特性を持つ個々の統計単位について、あくまでも時点ないし期間を定めてその統計原情報を切り取ることでその把握を行うものである。その結果、個々の統計調査については、静態統計調査によって収集される個々の調査票情報は調査時点情報を、また動態統計についても期間として時間情報を共有している。このような統計的把握の時間的斉一性の故に、一定時点(期間)を対象としたクロスセクション型データにおいては、様々なタイムスパンを持つ時間的要素の作用はコントロールされており、また動態データにおいても、期間として共通の時間的要因を持つ。

このように、静態と動態のいずれの統計調査もそれらが把握したデータにおける時間的要因が斉一的にコントロールされていることから、クロスセクションデータに基づく統計分析ではそもそも時間要素の作用は関心の外に置かれ、集計値を用いた時系列分析もまた時間要素をそれぞれの時点における集計データに斉一的に作用するものとして扱ってきた。時間的要因を取り込んだマイクロデータに基づくパネル分析もまた、個々の調査票情報はその時々々の時間的要因を斉一的にしか反映していない。その意味でこれまでの統計は、時間的要因の作用を統計単位毎に異なる特性としては反映してこなかった。これは、以下で見る場所的特性の調査票情報への作用が個別的であるのとは本質的に異なる。

7. 統計単位への場所的要因の作用

統計調査において、個人や世帯それに事業所や企業といった統計単位がしばしば取り上げられる。こういった統計単位については、その存在が少なからず特定の空間的な場所と結びついている。言い換えれば、好むと好まざるとにかかわらず、個々の統計単位はそれぞれの存在する場所が有する諸特性(場所的要因)の影響を受けているのである。個々の統計単位がそれぞれどういった場所的要因が作用している地点において存在するかは、その活動、行動、さらには意識などを自ずと制約する。このことは、調査票情報の data body 部を構成する変数値の中には、その所在あるいは立地の地点が有する場所的要因の作用に関係する部分もまた多かれ少なかれ反映されていることを示唆している。

(1) 個人、家計に対する場所的特性の作用

経済活動の主体であり統計調査において統計単位として取り扱われる個人や家計は、現実の空間の中にその存在の場を持つ。ここで、個人の就業行動の分析を例に取り上げてみよう。分析モデルの形式にはいろいろな創意工夫は認められるものの、それらは基本的に、就業、非就業という二値の従属変数を、性や年齢、学歴、その他の個人属性、親子関係といった世帯内の関連属性、さらには職業訓練等の経験の有無といった個体が獲得した事後的属性に対応する一連の独立変数によって説明するものとなっている。

ここで筆者は、この種のマイクロモデル分析に対して素朴な疑問を感じている。それは、個体属性が類似している場合、その存在の場所的要因とは無関係に、そのような個人はそれぞれの就業行動に対して単に確率的な差異しか持ち得ないのかという疑問である。なぜなら、調査票情報として得られる個体レコードに対して事後的に場所的特性ベクトルを付加、統合したデータセットがこれまで存在しなかったことから、従来のモデル構築に際しては、それぞれの個体の存在を規定し調査票情報に作用を及ぼし得る場所的要因は一切考慮されてこなかったからである。

この種のマイクロモデルは、モデル化の対象となる個体が空間的にはあたかも無機的世界の中に存在しているかのような個体観を前提している。言い換えればこの種のモデルは、ある個体が享受しうる利便性等に他の個体との間に本質的相違が仮にあったとしても、それらの要因はその者の就業行動には作用を及ぼさないと前提に立っている。同等の個体属性と等しい強度の就業意欲を持つ求職者でも、常住地の交通事情や通勤可能な近隣地域における労働需給という外部環境条件に顕著な差異ある場合、それらが当該求職者の就業確率に有意な作用を及ぼすことは現実に考えられる。

家計の消費パターンや消費行動などについても事情は同様である。委細は今後の実証研究の成果を俟つことになるが、商業施設の隣接地域に居住する世帯と遠隔地に立地する世帯とでは、おそらく家計消費のパターンや消費行動に少なからず差異があるものと考えられる。

これまでの個人や家計の就労や消費といった経済行動分析は、行動主体が具体的に置かれた場所的要因による被規定性という面を完全に考察の埒外に置いているという点で、共通にこの種の問題を内在させている。

(2) 企業、事業所に対する場所的特性の作用

企業や事業所についても同様に、マクロあるいはマイクロデータを用いた様々な実証分析が行われている。

ここで、企業の生産性分析を例に取り上げてみよう。物的生産性にせよ付加価値生産性にせよ、生産性は基本的に投入される労働や資本に対する生産物の価額によって評価される。また生産指数と投入指数との差によって定義される全要素生産性(TFP)も、長期的には生産組織も含めた技術水準の進歩、短期的には従業員の技能向上などとして説明されている。

現行の生産性分析では、個々の企業や事業所の立地条件という地点が有する場所的要因の作用の相違が全く考慮されていない。それはモデル構築の不備というよりはむしろ場所的要因の作用をモデルに取り込むのに必要な変数がそもそも未整備であるという分析データの制約による。仮に保有する人的・物的資源、経営ノウハウ、さらには技術力などがほぼ同等の事業所であったとしても、当該事業所が享受する立地環境がその事業活動のパフォーマンスに対して有意な影響を及ぼし得る。ちなみにここでいう立地環境を規定する要因としては、同業他社との競合の有無、物

資調達や輸送面での交通アクセス(港湾、空港、高速道路のICからの距離)面での利便性、享受しうるマーケットの規模や密度を示す各種の人口密度、確保可能な従業員の質や規模、関連業種の集積状況、さらには制度的要因としての経済特区としての指定の有無や誘致自治体による種々の助成措置の有無、といった様々なものが考えられる。

事業所の立地地点が持つ種々の場所的要因の作用の方向と強度は業種によって異なるであろう。しかし、特に事業所の場合にはその立地地点において事業活動を展開していることから、その地点が持つ場所的要因は、そこで展開される事業活動に少なからず影響を及ぼしているものと考えられる。ちなみに従来生産性分析は、少なくともモデル上は、このような場所的要因を技術進歩などとともに、事実上、全要素生産性に含めて取り扱ってきた。なお事業所の事業活動パフォーマンスに影響を及ぼしうる場所的要因が存在することは、確立した経営資源を持つ企業が事業所の新規開設場所の選定をめぐる、入念な事前検討を行っている事実からも明らかである。

8. 場所的要因の面的反映としての地域表章

(1) 地域表章単位としての調査区

国勢調査や経済センサスといった全数調査の場合、統計調査を実施するにあたって調査区が設定される。調査区は、実査の際の調査漏れや重複把握を排除することを目的に調査対象全域を網羅的に設定される。このことから明らかなように、調査区は本来的には統計単位の捕捉(把握)精度を確保する目的で導入された調査技術的な仕組みである。

調査区の境域設定は、行政区の境域と整合的な形で行われる。それには、統計調査の実施に関わる行政上の仕組みが関係している。なぜなら、国勢調査や経済センサスでは、都道府県、市区町村という行政組織におかれた統計主管部門が地方レベルでの調査業務を担当し、また実査担当者である調査員も市区町村が調達する仕組みとなっているからである。その結果、調査区の区割りも自ずと行政上の区画と整合性を持って設定され、行政界をまたぐ形で調査区の設定が行われることはない。

このように調査区が行政区と整合的な形でその境域設定が行われることから、本来的には統計単位の網羅的捕捉のための調査技術的仕組みとして導入された調査区が、調査票情報の集計処理過程においても地域集計の基礎的表章単位として用いられている。実査に先立って各調査票には、それぞれ該当する都道府県番号、市区町村番号、調査区番号が記載される。調査票情報が持つこういった地域コードを集計変数として用いることで、都道府県、市区町村別の集計表を作成することができる。なお国勢調査の場合、小地域集計として町丁目による集計表も作成されているが、これについても各調査区が町丁目区分と整合的な形で設定されていることから、調査区番号を用いて調査票情報を集約することでこのような集計が可能となる。

調査票に基づいて実査を遂行することにより、個々の統計単位別の統計原情報が調査票情報として収集される。統計調査への個票(小票、個別票)が導入されたのを契機に、それまでの表式調査の場合とは異なり、少なくとも統計原情報のデータ形式としては個体ベースでのそれが利用可能な新たなステージへと統計は踏み出したことになる。

このように、情報の形式としては統計原情報は個的なものとして存在するにもかかわらず、調査結果の表章においては、わが国の公的統計はこれまで一貫して調査区(国勢調査の場合には基

本単位区)を地域表章の基底的表章単位としてきた。すなわち、統計原情報が持つ点的表章可能性からではなく、調査区という実査実施上の把握単位から出発した調査結果の地域表章がこれまで行われてきた。個的性格を持つ統計原情報の調査区ベースでのこのような面的把握は、個体情報すなわち非集計情報として固有の情報量を持つ調査票情報を調査区ベースで一旦集約し集計量としてその後の集計過程へと送っていることになる。その意味では、これまでの地域表章に係る集計処理の方法は、少なくとも場所的要素の取り扱いに関しては、個票調査が持つ表式調査に対する優位性を生かしたもとはなっていない。換言すれば、点的な個体情報の面的集約に伴う情報損失の故に、これまでの統計は地域表章に関する限り、かつての表式調査が囚われていた情報制約をまさに共有しているのである。

(2) 場所的要因と行政区による地域表章

個別統計調査の結果報告や編集統計書等におけるこれまで統計の地域表章は、都道府県や市区町村といった行政区による結果表章が中心であった。これには以下のような事情が関係しているように思われる。

第 1 は、調査実施体制への自治体の関与である。すでに述べたように、調査区は本来実査遂行のための調査技術的枠組みとして設定されたものである。統計調査の実施に際して、調査員の募集、訓練、配置、調査督促、記入済調査票の審査といった調査実施の最前線の業務を各地方自治体が担っている。このことから、調査区の境域設定も自ずと自治体の行政区分と整合的な形で行われることになる。

第 2 の理由は、調査結果の利用に関係するものである。すなわち、調査結果の地域表章が都道府県や市区町村といった行政区分に従って行われてきたのには、中央政府や地方自治体が公的統計の主たる利用者であったことによるものと考えられる。行政は行政区を対象範囲として各種政策を企画、実施する。統計作成機関側でもこのような行政からの統計ニーズに応えるために、調査結果の地域表章も行政区分に沿った形で行ってきた。

なお、行政区分による境域とは異なる形での地域表章として地域メッシュ統計がある。これは、「国土を緯度・経度により方形の小地域区画に細分し、この区画に統計調査の結果を対応させて編集したもの」で、わが国では昭和 40 年代後半以降、国勢調査や事業所・企業統計調査結果等の小地域表章方法として採用されている。

地域メッシュ統計の場合、調査区ではなく空間を物理的に区切ったメッシュ(グリッド)という独自の境域設定が行われる。地域メッシュの利点として、つぎの4点があげられている〔総務省統計局 1 頁〕。

- ① 地域メッシュがほぼ同一の大きさ及び形状の区画を単位として区分されていることから地域メッシュ相互間の事象の計量的比較が容易であること
- ② 地域メッシュはその位置や区画が固定されていることから、市町村などの行政区域の境域変更や地形、地物の変化による調査区の設定変更などの影響を受けることがなく、地域事象の時系列的比較が容易であること
- ③ 任意の地域についてその地域内の地域メッシュのデータを合算することにより必要な地域のデータ入手が容易なこと
- ④ 地域メッシュは緯度・経度に基づき区画されたほぼ正方形の形状であることから、位置の表示を明確で簡便にすることができ、距離に関連した分析、計算、比較が容易であること。

地域メッシュ統計は、その面的表章単位の境界がメッシュという物理的に設定されていることから、調査区や行政区などに見られる境域変更の影響を受けないという統計利用面での長所を持つ。その一方で、現行の統計においては調査区が地域表章の際の基底的集計単位となっていることから、メッシュと調査区それぞれの境域の非整合性に起因するメッシュ統計独特の問題がその作成にあたって発生する。それぞれの境域が相互に非整合的であることから、メッシュ統計の作成の際には必然的にメッシュの境界を跨ぐ調査区が発生し、調査区ベースで集計量の形で与えられている調査結果を該当する各グリッドに配分する必要がある。調査区の同定と呼ばれている作業がそれである。このような同定作業が必要となるのは、個々の統計単位についての調査票情報が調査区という面的単位として集約され、それが地域表章の基底的集計単位となっていることによる。

9. 場所的要因の点的性格

(1) 点的空間存在としての統計単位

国民国家の場合、国境は国際的な人の移動、貿易や租税、金融その他の政策面で企業等の事業活動を制約する制度的枠組みを与える。また、政治的・行政的に設定される各種の経済特区等も、税制上の優遇や許認可要件など対象域内全域に等しく適用され域外とは質的に区別される制度的条件によって特徴づけられる。さらに、地方自治体による独自の産業誘致策や福祉行政等は、企業進出や世帯の居住地選択行動に対して有意な作用を及ぼし得る。この場合にもそれぞれの政策は当該行政区全域に適用され、その作用は隣接する他の行政区域から質的に区別される。行政行為が当該行政区域内に所在する全ての個体に対して均一に作用を及ぼす場合、場所的要因としての行政行為の域内の個体に対する作用に関しては、統計の行政区による結果表章は実質的な意味を持つ。

しかしながら、多くの場合、個人や世帯の日常の生活や行動、事業所や企業の事業活動において、行政区画による境界は現実にはあまり意味を持たない。人々は市区町村や都道府県といった行政区界を特に意識することもなく日常的な社会生活や事業活動に従事している。こういった活動の結果として作り上げられる地域経済圏や商圏は、市区町村界さらにはしばしば都道府県の境界を跨ぐ形でその圏域を形成している。公共交通機関や利用資格等に特に制約の設けられていない公共施設等の利用にあたっては、利用者にとってはあくまでもアクセシビリティが主たる関心事であり、それらの施設がいずれの行政区域に所在するかが利用者の意識にのぼることはない。

市区町村等の行政区以外にも、場所的要因が多かれ一定の広がりを持つ特定の部分領域(ポリゴン)あるいは部分領域の集合に対して同等の方向と強度で影響を及ぼすことがある。例えば都市計画法による用途地域区分は自治体の境域を土地利用に関して12の制度的区分を持ち、また建築基準法も建蔽率、容積率について地域のゾーニングを行っている。こういった行政区とは異なる制度的地域区分の他にも、地価、交通利便性、人口密度、事業所等の集積度といったそれぞれの地点が持つ場所的要因が与えるスコアの階級帯による地域区分などがありうる。

これらの場所的要因が個々の統計単位に対して一定の広がりを持つポリゴン(あるいはポリゴンの集合)として面的に作用を及ぼす場合も、それらによって形作られる境域が行政区分による境域区分と一致する保証はない。また上に例示したような種々の場所的要因のスコアの各階級帯が形作る境域も様々である。それぞれの場所的要因は全対象領域を排他的にゾーン区分したポリゴン

群を構成し、任意の地点を指定すれば、必ずいずれかのポリゴン値が対応している〔森 2013d 23-26 頁〕。

(2) レイヤーとしての場所的要因

ここで統計単位の存在する地点情報とその地点を特徴づける場所的要因の間の関係を GIS の図式に当てはめて解釈してみよう。それぞれの場所的要因はそれぞれ一つのレイヤーを構成し、空間的存在としての統計単位全体を重層的にオーバーレイしている。これを個々の統計単位の側から見れば、それらの空間的存在に係る地点(経緯度:X,Y)情報がリンクキーとして各レイヤー上のポリゴン値を串刺し的に統計単位に紐づけている。各レイヤー上のポリゴンの境域はレイヤー毎に異なる。その結果、個々の統計単位の存在に係る地点情報は、それぞれの地点に固有の場所的要因にあたる場所的特性変数の実現値を調査票情報に対応づけることになる。このことをデータレコードとして捉えれば、当該統計単位の空間的存在に係る固有の場所的特性変数ベクトルが調査票情報に対して、集団としてではなく個体ベースで対応していることを意味する。統計単位が所在する地点を特徴づけるベクトル要素としての場所的特性変数値は基本的に統計単位毎に異なる。

ところで、個々の統計単位は、所在地情報としての住所を持っている。住所には郵送用住所(postal address)と物理的住所(physical address)とがある。両者は通常は一致するが、私書箱や事業用の郵便物の配達先を自宅に指定している自営業などの場合にはそれらは異なる。このうち、統計単位としての事業所等が実際に所在する物理的住所を用いてアドレスマッチングによってその所在地点の経緯度座標を取得する。

アドレスマッチングによって経緯度情報を取得する場合、事業所や住戸の所在形態によっては、一つの物理的住所が与える地点に複数の統計単位が対応するケースがある。マンション等の集合住宅を住戸とする世帯あるいは複合商業施設のテナントなどがそれである。また、アドレスマッチングの方法によっては、異なる物理的住所に対して同じ経緯度情報が付与されるケースもある。街区レベルでのアドレスマッチングの場合、仮に住所が号による区別を持っていたとしてもマッチングソフト上ではそれらを区別できないことから、結果的に複数の物理的住所に所在する統計単位に対して同じ経緯度情報が付与されることになる。

こういった場合、複数の統計単位がその存在の場としての経緯度情報を共有することから、調査票情報と場所的要因を構成する場所的特性ベクトルの間にはn:1の関係が成立する。アドレスマッチングの結果複数の統計単位が経緯度情報を共有することになった場合、それらに該当する複数の調査票情報は同じ場所的要因の作用を受けているとみなされ、これらの個体レコードには同じ場所的特性変数値からなるベクトルが紐づけられる。

これとは逆に、仮に空間的に隣接して存在する統計単位であっても、その物理的住所が異なりアドレスマッチングが異なる経緯度情報を与えれば、GIS 上では異なる座標として認識される。その場合、それぞれの地点の座標が部分的に異なるポリゴン値を持つ場所的特性ベクトルを持つことから、各個体レコードにはそれぞれ異なった場所的特性値を持つベクトルがリンクされることになる。

(3) 地域集計による調査票情報からの情報損失

行政区内にも実際には様々な地域的差異があり、町丁目、地域メッシュといった小地域統計さらには調査区や基本単位区においても、個々の統計単位の存在を制約する場所的要因は必ずしも

同じではない。

個々の統計単位の空間的在り様は点的であり、境域内に様々な形で分布している統計単位はそれぞれの存在地点が持つ場所的特性変数値ベクトルによって特徴づけられている。データ上は場所的特性変数値からなるベクトルとして表現される一連の場所的要因が本質的に点的に作用しており、個々の調査票情報と直接関係づけられることは上述の通りである。このように調査票情報がまさに点的存在として把握されているにもかかわらず、それらを調査区さらには行政区等に従って集計値、平均値、あるいは比率等として面的に集約した集計量という情報形態を持たせた瞬間に、場所的要因の作用の統計データへの反映可能性に関して重大な情報損失が発生する。

場所的要因に関する点的情報の面的集約は、場所的要因がそれぞれ境域内で均一に平均的に作用しているとの認識を前提している。種々の場所的要因が実際にはそれぞれ異なる方向と強度でもって境域内の個々の統計単位に対して作用を及ぼしているとしても、収集された調査票情報を面的に集約処理した結果、それが境域内に存在するあらゆる統計単位に均一に平均的に作用しているものとしてデータ上は表現される。情報の面的集約に基づくこれまでの地域表章では、場所的要因は専ら域内均質性と域間差異性として取り扱われてきた。地域別集計表として集約されたセル値の中には、一方で個々の統計単位が本来持っている個体そのものの特徴が溶融・平均化されているだけでなく、個々の統計単位を規定する場所的要因の作用もまた平均的なものとして集約的に表現されている。

統計調査に個票が導入されたことによって統計単位に係る統計原情報は、事実論理としてはそれぞれの空間的存在を示す地点情報とともに収集されることになった。調査票情報はその空間的存在に関して新たな情報価値を潜在的に付与されることになったのである。にもかかわらず伝統的な地域表章では、調査票情報の面的取り扱いの故に、場所的要因を構成する場所的特性値ベクトルはそれぞれの地点に固有なポリゴン値としてではなく、あくまでも境域内の統計単位に関係づけられたそれぞれのポリゴン値の平均的特性値を該当する全ての個体情報に付与している。

10. 調査票情報の拡張(1)－data carrier 情報によるデータの統合

調査票への記載情報を data carrier と data body という視点から区分してみた場合、氏名や事業所等の名称は統計単位の個体識別情報である。調査票情報がこの個体識別情報を介して個々の統計単位に直接関係づけられるという意味で、それは統計単位が持つ諸側面の統計的反映である data body を担う data carrier に他ならない。

個体識別情報としての名称それ自体は非統計項目であり、集計され統計に編成されることはない。しかしそれが持つ data carrier としての特性は、統計におけるデータ統合を考える場合には特別の意味を持つ。なぜならそれは、まさに個体ベースでのリンケージによる調査票情報の拡張を行う際のキー変数に他ならないからである。

調査票に記入された氏名(名称)は、特に日本語によるテキスト情報の場合、種々の表記上の揺らぎを持つ。そのため、これまでの氏名を用いた調査票情報の名寄せ作業では、住所(所在地)や電話番号等も補助情報として併用されてきた。

(1)横断的リンケージ

同じ統計単位が他の調査でも調査客体となる場合、それぞれの調査資料は同一個体に関する

調査票情報を持つことになる。こういった場合には、それらを名称等で照合し、それぞれのレコードを連結することができる。このような源泉を異にする調査資料から得られた調査票情報をいわば横断面方向にリンクすることによって、個体ベースでのレコード変数の変数次元の拡張が可能である。こうして拡張された個体レコードを用いることで、新たに統計調査を企画、実施することなく、既存の単独の統計調査の結果では行うことのできなかつた新たな変数をクロスオーバー的に組み合わせた集計処理や回帰分析が可能になる。

(2) 縦断面リンケージ

センサスのような全数調査では、原則として全ての統計単位が調査把握の対象となる。このため定期的に実施されるセンサスの調査票情報を用いることによって、同一個体に関する個体レコードを時間軸方向に連結した個体追跡型のデータセットを編成することができる。センサス結果の事後的なマッチングによるデータ編成の他にも、そもそも縦断面調査として同じ調査客体だけを継続的に調査しているものもある。さらに、標本調査の中にも、例えば労働力調査のように、一定期間後に再度同じ調査客体を調査するように標本設計されている調査もある。このような同一個体を経時的に追跡するデータは、縦断面データ(longitudinal data)あるいはパネルデータと呼ばれる。パネルデータの場合、同じ調査客体が継続して調査されることから、集計量としての時系列データからは特定できない個体ベースでの動態変化の分析、グロスでの変化分析、動態的集団構成による分析が可能である。

(3) データ統合の視点から見た統一共通コード導入の意義

分散型統計制度を持つわが国では、各府省は統計予算として自ら確保した財源によってそれぞれ所管する統計調査を実施してきた。そのため従来は、それぞれの調査結果が他府省実施の調査とはもちろん、他部局主管の調査とも連携しておらず、それぞれが基本的に独立した調査として作成、使用されてきた。そのため統計単位としての世帯、事業所、企業等の個々の調査客体にはそれぞれ調査に独自の調査客体コードが付与され、それらを調査間で共有することは殆ど行われてこなかった。

第二次大戦後、公的統計に母集団概念が導入され調査の体系化がはかれることになる。そのような中、国勢調査や事業所統計調査は、それぞれ世帯あるいは事業所を調査対象とする他の標本調査のための母集団情報の提供という新たな統計体系上の役割を担うことになる。しかし、このような公的統計の体系的整備は、その後社会の変化がもたらす新たな脅威に曝されることになる。1970年代以降深刻化する統計の調査環境の悪化がそれである。

プライバシーや秘密保護意識が広まる中で、それまで国民の間での調査協力を支えてきた公民意識が後退し調査回答率が低下する。センサス型調査も例外ではなく、把握精度の低下は、センサスがこの間担ってきた母集団把握について、次第に機能不全を来すことになる。このような事態の推移を受けて、特に調査環境がわが国よりも一層深刻であった欧米各国は、1980年前後からセンサスに代わって行政情報を中心とした母集団の整備に着手する。ビジネス・レジスターと呼ばれるデータベースの整備がそれである。

わが国でも、2007年の統計法改正を受けて、遅ればせながら事業所母集団データベースという形でビジネス・レジスターの整備に向けた取り組みが開始される。税務情報を主要な情報源としている諸外国とは異なり、わが国の事業所母集団データベースは、経済センサスから得られる調査票情報を中心に労働保険契約情報、商業法人統計情報等の行政情報も用いることによって事業

所情報を収集し、その結果をプロファイリングという所在確認作業によって統計単位を確定することで母集団の整備を行うものである。

各国ではビジネス・レジスターは、調査実施の基盤情報として公的統計調査のための母集団情報や標本抽出のためのフレーム情報を提供することを主たる目的に整備されてきた。しかしこの制度的仕組みが持つ機能は、それにとどまるものではない。ビジネス・レジスターは、欠損データや欠損レコードの補完機能、さらには政府統計の調査票情報を個体ベースで統合、アーカイビングする際のプラットフォームとしての機能を同時に備えたものとして整備され、維持管理されている。この点をさらに敷衍するなら、それは単なる調査実施の基盤情報ではなかく、必要な制度的対応が図られれば、分散型統計制度の下においても収集された個別調査の調査票情報を相互に個体ベースでデータ統合することができる潜在的可能性を持つものである。

共通事業所・企業コードを装備した事業所母集団データベースがわが国でも2013年度から供用開始となった。それに伴って企業、事業所を対象とする基幹統計調査、一般統計調査については、いわゆる名簿情報として事業所母集団データベースを使用することが推奨されて、このデータベースを名簿情報として実施された主要21調査から得られる調査票情報はデータベースにフィードバックされデータベースの更新情報として活用される仕組みになっている。

こういった共通事業所・企業コードの制度化は、調査票情報のマッチングに際しての業務作業量を著しく軽減するとともに、その照合精度の向上に大きく寄与するものとして期待される。共通事業所・企業コードの導入により、わが国もようやく公的統計の調査票情報のデータ統合に向けての一步を踏み出すことになる。

(4) 調査票情報と行政情報のデータ統合

統計単位の個体識別情報をキー変数として調査票情報を個体ベースでリンケージすることは個体レコードの視点からデータの統合拡張を捉えたものである。このことを現在の情報処理技術を前提としてデータのアーカイビングの形態という視角から捉えた場合、それは各種調査の調査票情報に共通事業所・企業コード等の個体識別情報をリンクキー変数として持たせることによってそれらに相互の連結可能性を保持させたリレーショナル型のデータベースに他ならない。

行政における単位は統計単位とは必ずしも一致しない。しかし、行政においても個人や世帯、事業所あるいは企業といった単位で行政情報が維持管理されているケースは少なくない。このような場合、原理的には統計における共通事業所・企業コード等の個体識別情報をキー変数として統計情報と行政情報とをデータ統合することも可能である。特に行政においては単位を悉皆的に把握する機会が多いことから、調査票情報に各種行政情報を事後的に付加することで、一般に統計調査によっては入手が困難な変数も組み込んだ新たな拡張データセットを作成し、それをを用いることで行政的に意味のある分析結果や政策上の示唆が得られるであろう。

このような個体識別情報をキー変数としたデータ統合の発展系として、全住民位置情報や全事業所位置情報という経緯度情報を変数として組み込んだデータ統合形態も想定することができる。このような経緯度情報を変数に組み込んだ統合データの場合、これまでの統計には無かったような多様な空間集計が可能であるだけでなく、交通事故多発カ所や災害避難時の被支援者の所在など個々の地点と結びついた政策対象の特定といった行政施策の策定、実施にとって有用な情報を得ることができる。さらに経緯度情報の履歴情報を格納できる形でデータベースを設計しておくことができれば、個々の統計単位の移動という空間的要素も持つ新たなタイプの拡張された縦断

面情報の整備も可能であるように思われる。

11. 調査票情報の拡張(2)－経緯度情報による調査票情報への場所的特性変数の付加

第5節でも指摘したように、調査票情報は個々の統計単位に帰属する個体情報が場所的要因の作用というフィルターをくぐって実現したものとして捉えることができる。このことは、個体に関するいわゆる data body 部分を構成する変数によっては、その実現値に対して多少なりとも場所的要因の作用をうかがわせるものである。

場所的要因の個体への作用を織り込んだ実証分析のためのデータの整備は、今後の統計に課された重要な課題である。その場合、様々な場所的特性変数をレイヤーとして一つ一つ整備していく地道な取り組みが求められる。個々の場所的要因を特徴づけるレイヤー上では任意の地点に対していずれかのポリゴン値が定義されており、経緯度情報として地点が特定されれば、その座標に対して各レイヤー上のポリゴン値を要素として持つ一連の場所的特性変数値からなるベクトルが紐づけられる。このことは、事業所等が立地する地点(所在地)を示す経緯度情報がひとたび座標として与えられれば、当該統計単位の調査票情報に対して場所的特性ベクトルを事後的に付加できることを意味する。

経緯度情報をリンクキーとして事後的に場所的特性変数が付加され拡張された個体レコードを使用することで、既存の調査票情報の変数値に対する場所的要因の作用の有無やその強度を定量的に評価できるであろう。このことは、面的集約として捉えられた集計量としての地域特性からは決してなしえなかった実証分析への新たな接近を意味する。

12. 調査票情報の拡張(3)－経緯度情報によるデータの統合

前節では、統計単位に関する個体識別情報を用いた調査票情報のリンケージによる異種調査間の横断面方向での拡張ならびに同一調査の縦断面方向でのデータ統合による個体レコードの拡張について取り上げた。

統計調査が収集する世帯、事業所、企業といった調査票情報の中には、住所(所在地)情報を持つものも少なくない。これまで住所情報は名称を用いたマッチングの際の個体識別のための補助情報として用いられることはあっても、住所情報そのものが集計変数として使われ、統計が編成されることはなかった。

統計単位の空間的存在地点を示す物理的住所というアナログ情報を経緯度情報という空間データ処理が可能なベクター型変数へと変換することによって空間集計が可能となるだけでなく、さらにはそれをキー変数とする調査票情報の事後的な変数拡張によって調査票情報が持つ潜在的可能性の開拓が可能になる。以下ではアドレスマッチングによる経緯度情報の取得とそれを用いた調査票情報の新たな展開可能性を持つ統計的意味について考察する。

(1) アドレスマッチングによる所在地情報の経緯度情報への変換

2000年の米国大統領命令による政府による地理測位システム(GPS: Geographical Positioning System)の選択利用(SA: Selective Availability)規制の解除、その後の観測精度の向上を受けた端末の広範な普及により、民生利用も含めて経緯度情報の各方面での利用が

進むことになった。このような情報技術の進展を受けて、経緯度情報の統計への利用可能性も近縁注目されている。

地点情報を空間データとして処理するためには、それに経緯度情報を付与する必要がある。Geocoding と呼ばれる変換の中でも特に住所情報の経緯度情報への変換は、アドレスマッチング用データベースでのデータ照合によって行われる。アドレスマッチング用ソフトウェアは住所データベースとそれにリンクされた経緯度情報を持ち、照合元 (donor) の住所情報をデータベース上の照合 (recipient) 用住所データと照合することを通じて照合元住所に経緯度を付与するものである。なお、すでにふれたように、経度情報の付与方式には、街区と号レベルによるマッチングがある。

近年、諸外国では経緯度情報を変数として装備した建物レジスターを政府が国家的事業として整備しつつある。例えばフィンランドでは建築着工の届出様式に経緯度情報の記載が法的に義務づけられており、その記載情報はビジネス・レジスターに変数として格納されている。またフランス政府が整備している建物登録 (RIL: le répertoire d'immeubles localisés) やカナダ統計局の住所レジスター (AR: Address Register) では、街路の始点と終点の経緯度情報を取得し、両端点間に建物が等間隔に配置されているとの仮定の下に、経緯度情報の推計値がデータベースに格納されている。さらに、アメリカセンサス局では、2010 年人口センサスの実施に先立って 2009 年に全住戸について、canvasser と呼ばれる現場作業員が、GPS 端末装置を用いて経緯度情報を取得している。このように諸外国では経緯度情報の取得が組織的に行われており、それが一種の社会基盤情報として整備されつつある。

一方わが国では、住所情報について、「住居表示に関する法律」(昭和 37 年法律第 119 号、最終改正:平成 11 年 12 月 22 日法律第 160 号)の第 2 条に、(住居表示の原則)として、市街地にある住所若しくは居所又は事務所、事業所その他これらに類する施設の所在する場所を表示するには、都道府県、郡、市、区及び町村の名称を冠するほか、下記の二種類の住所表示方式による旨規定されている。

街区方式	市町村内の町又は字の名称並びに当該町又は字の区域を道路、鉄道若しくは軌道の線路その他の恒久的な施設又は河川、水路等によって区画した場合におけるその区画された地域(以下「街区」という。)につけられる符号(以下「街区符号」という。)及び当該街区にある建物その他の工作物につけられる住居表示のための番号(以下「住居番号」という。)を用いて表示する方法
道路方式	市町村内の道路の名称及び当該道路に接し、又は当該道路に通ずる通路を有する建物その他の工作物につけられる住居番号を用いて表示する方法

わが国の住所は、地名の漢字等による表記や地番の表記形式など日本語に固有の情報の揺らぎを持つ。そのため上述したアドレスマッチングソフトも、正規化という広範な表記の揺らぎについての処理機能を内蔵したものとなっている。それでもなお、京都市など地域によってはかなり低い照合率に留まっているところもある。このような機械的照合による経緯度情報の取得ができなかったケースについては、Google マップあるいは実地調査による直接取得などの方法を併用することで、統計の空間分析が要求する精度に十分適合した経緯度情報の取得を行うことができる。な

お、現行のアドレスマッチングが持つ経緯度情報取得に関する問題点を考察したものとして〔長谷川・卯田 2012〕がある。

(2) 経緯度情報による地域表章の自由度の拡張

調査票情報が持つ住所情報を経緯度情報に変換し個体レコードに変数として装備することの統計上の意義は大きい。調査票情報に経緯度情報を付与することによって、利用者は、従来の調査票情報が持っていた調査区コードや基本単位区コードに起因する様々なデータの利用制約から完全に解放されるだけでなく、経緯度情報をキーとした調査票情報の拡張、さらには行政情報も含めた情報のリンク可能性など統計の新たな展開が可能となる。

(i) 時系列比較可能性の担保

情報処理能力の向上を受けて、例えば国勢調査の地域別表章においては、従来からの都道府県や市区町村といった行政区レベルでの域集計表の作成に加え、集計項目によっては人口集中地区さらには町丁目、調査区、調査単位区といった様々なレベルでの小地域統計が現在作成されている。

こういった地域表章は、地域区分が変更された場合に時系列的な比較可能性が失われるという宿命的な難点を持つ。都市再開発、道路等の新設といった開発行為の結果、調査区はその境域に変更が加えられ、また市区町村の合併等により行政界に変更が加えられるケースも少なくない。境域が変更された場合、旧地域区分による結果表章、あるいは新地域区分による過去の調査結果の遡及再集計には多大な事務的、機械的作業を必要とする。そのため、国勢調査を例にとれば、それらは特別集計として、限られた調査年次のしかも一部の調査項目について行われているだけである。

世帯に関する調査票情報についても事業所の場合と同様に、各住戸の住所に基づいて経緯度情報を付与した形で世帯に関する調査票情報をアーカイビングする体制が整備できれば、行政区や調査区の境域の分割や統合さらには部分的拡張や縮小といった種々の改変に起因する時系列的な地域表章に関する制約は一気に解消される。調査票情報に経緯度情報が装備されてさえいれば、座標情報を持つ境界データを準備するだけで任意時点での地域区分に従った時系列表を、遡及集計も含め機械的に行うことができる。ちなみに、事業所データについてはすでに個々の事業所の所在地情報を日本測地系の住所辞書データから緯度、経度情報を取得することによってメッシュコードを付与するアドレスマッチングによる同定(「アドレス同定」)が行われている。

(ii) バフアリング等による多様な地域表章の可能性

地域経済圏、商圈、ニュータウンのような居住圏は行政境域を跨がった形で広がっており、上下水道、電力といったライフラインや鉄道・バスといった公共交通機関が形成する通勤圏、河川の流域圏なども行政区分による境界を越えて、線あるいは面的に広がっている。他方で、広域避難所や地域コミュニティ施設が提供するサービス圏域、学区、町内会、郵便番号、商業施設や公共交通機関への利便度といった既存の行政区分による境域の一部を切り取った形で設定される区域による結果数値についても様々な統計ニーズが存在する。これらの中には町丁目、あるいは調査区レベルまで遡及することで対応できるものもあるが、調査区と必ずしも整合的でない境域設定が必要なケースも少なくない。

これまでの統計における地域表章にはないこのような多様な統計ニーズに対応するには、調査区という面的地域表章単位を改め、個体情報としての調査票情報が潜在的に保有するまさに点

的な位置情報を活用する必要がある。個体情報として収集されている調査票情報を調査区という調査技術的枠組みによって集約するのではなく、統計単位の空間的存在をあるがままの個的な点的存在として情報化することによって、多様な集計ニーズへの対応が可能となる。

(iii) 防災や被害推定への利用

大雨、高潮、津波などによる浸水・水没被災地域、火災による延焼地域の広がり、海抜高度や堤防の状態、降雨量や津波の高さ、風向きや風力等によって異なる。このような被災地域の広がり、調査区あるいは行政区の境域とは自ずと独立であり、またメッシュによる方形とも異なる境域を持つ。迅速な被害把握や事前の被害予測のためには、個々の事業所や住戸等と関連づけられた全住民位置情報のような非集計量によって与えられる情報が不可欠である。

2005年8月に合衆国南部を直撃した巨大ハリケーンのカトリーナは、ニューオーリンズ市を中心に各地に甚大な人的、経済的被害をもたらした。ニューオーリンズ市における水没による被害について、米国商務省センサス局は、一部推計も含め個々の事業所の経緯度情報を取得し、事業所個体情報とリンクさせることで、水没によって失われた雇用数、取引額等の経済的被害額の推計を行った〔Jarmin and Miranda 2006〕。この研究は、その推計の迅速性ととも、経緯度情報と調査票情報とを結合した被害評価分析の嚆矢として注目されている。ちなみにわが国では、総務省統計局が2011年3月の東日本大震災に伴う津波による被災人口、世帯数を公表している。これは、空撮により把握した津波水没地域に国勢調査の基本単位区情報を重ね、境界に位置する単位区については同定により水没罹災人口・世帯数を推計したものである。建物登録による全住民位置情報が整備されていれば、より迅速でしかも正確なその確定ができると考えられる。

(iv) 自治体等による独自境域設定による利用

統計利用者はそれぞれの利用目的に従って独自に境域を設定し、分析を行うことがある。特に自治体の場合には、行政サービスの提供にあたり、地域住民の利便性に配慮した地域コミュニティ施設や公共避難場所等の最適な配置が求められる。施設の設置場所の選定に際しては、徒歩あるいは乗用車等でのアクセス可能境域を設定しサービス享受者数の把握が必要となる。また人口減少局面においては、既存施設の統廃合を含む再配置や急増する空家や空き地の発生状況の把握といった行政課題も存在する。さらに、人口の高齢化の進展に伴い、いわゆる買い物難民や医療施設難民の規模やその地域別分布の解明もまた重要な行政課題となっている。

一方、民間においても、流通事業者は公共交通機関による交通移動時間等も考慮した潜在的集客能力や既存の他の店舗等との競合度あるいは集中立地のメリットといった様々な要素を考慮した上で出店を決定し、製造事業者も工場の進出、移転場所の選定にあたって、企業側では実に多様な要因を考慮してその決定を行っている。

こういった分析ニーズは、それぞれの分析目的によって異なる固有の設定境域を持つ。そこでは、都道府県や市区町村といった行政区による境界区分はもとより、町丁目や調査区、あるいは地域メッシュ統計といった小地域統計で採用されている境域も、こういった利用者の利用目的に応じた境域設定とは一般に対応していない。

既存の地域区分を分析目的に対応した境域に調整するためには同定のための作業が必要となる。その点では、経緯度情報と個々の調査票情報とが関連づけられてさえいれば、強力な空間集計、空間解析機能を用いることで、専ら技術的なデータ処理としてそれを行うことができる。

(3) 経緯度情報付き調査票情報の分析的利用可能性

(i) 事業所・企業に関する調査票情報への経緯度情報の付加

①経緯度情報付クロスセクション型調査票情報(単独調査時点データ)の利用

特定時点のクロスセクション型の経緯度情報付の調査票情報を用いることによって、業種別の産業立地・集積状況、その規模(生産額、従業員規模)、集積密度等を集積地域別に把握することができる。

②経緯度情報付クロスセクション型調査票情報(複数調査時点データ)の利用

複数時点でのクロスセクション型の経緯度情報付の調査票情報を用いることによって、業種別の産業立地・集積状況、その規模(生産額、従業員規模)、集積密度等の地域別の時系列的な変化を追跡することができる。

経緯度情報は、事業所の交代に関して、事業所の立地地点そのものの特性分析にも使用することができる。単一の事業所が一つの物理的住所に立地している場合、複数の調査時点の調査票情報を経緯度情報を用いてマッチングした際に、調査時点によって事業所の名称(あるいは事業所識別番号)が異なることがある。これは、異時点で別の事業所が同一の場所に立地していること、すなわち二時点間で事業所の交代が発生したことを意味する。このような照合結果を用いることで、業種毎の交代確率や業種間の交代率といった事業所の特性分析だけでなく、例えば地域別の交代確率といったような空間特性の分析を行うこともできる〔坂本他 2014〕。

③経緯度情報付パネル型調査票情報の利用

わが国の場合、事業所や企業を調査対象単位とする統計調査の多くでは、これまでそれぞれ独自の事業所コード等の個体識別番号が付与されてきた。それを用いることで各調査について調査票情報を縦断的に接続したデータセットの編成が可能である。これらは、企業・事業所パネルデータとして各種の分析に用いられてきた。

パネルデータに各調査時点における事業所や企業の所在地に関する経緯度情報を履歴情報として持たせることで、事業所等の経時的な空間的移動を把握できる。このような空間的移動状況を業種別、規模別等に分析することで、企業や事業所の立地の再編あるいは集積状況の動態的变化がわかる。さらに、移動に関するOD(Origin-Destination)行列を作成することで、移動方向の特徴を把握することもできる。また、それまで参入数、退出数の中に混在していた事業所等の移動数を特定することによって、近年、OECD を中心に広く展開されているビジネスデモグラフィーに与える主要関心事である正確な新規開設(起業)数、閉鎖(廃業)数すなわち事業の出生率や死亡率を把握することができる。

(ii) 世帯に関する調査票情報への経緯度情報の付加

経緯度情報の付加による調査票情報の利用可能性の拡張は、住戸の所在地情報が与える経緯度情報を用いることによって、世帯に関する調査票情報に対しても適用可能である。それは、ハザードマップ等と連動させた被害予測、被害推計といった防災上の利用可能性だけでなく、いわゆるバッファリング分析による小地域マーケティングへの利用可能性を拓くものである。そこでは、地域メッシュ統計の場合と異なり、任意の地点を中心とする任意の地域形状による地域の切り取り(バッファリング)に対応しうることで、その利用可能性は飛躍的に拡充される。

経済統計の基盤情報としてビジネス・レジスターがあるように、人口統計について北欧諸国やオランダではすでに世帯レジスターが整備されている。世帯レジスターが履歴情報として経緯度情報を持つようになれば、世帯についてのデモグラフィックな動態分析や地域間移動などの追跡が可

能となる。

むすび

表式調査の場合には集計表の作成が究極の目的であり、集計論理が表式による統計原情報の収集という調査過程を支配している。そこでは統計は集計表毎にそれぞれ調査過程を従えた閉じた世界を形作っており、統計間の連結利用などは統計の作成者の意識にはそもそも登っていない。

また、調査結果報告書や編集統計書等に掲載されている各種集計表の場合、本稿では疑似的個体概念を導入することで、集計表の各セル値のレコード的解釈の可能性を提案した。これはあくまでも個体レコードのアナロジーとしての解釈に他ならず、その意味では集団の個体としての擬制さらには集計表のバーチャルな形でのレコード的整理そのものが、データ統合という問題視角からの新たな論点提供である。集計表として提供される統計の実際の利用者は、時系列比較の場合を除き、個々の集計表をそれぞれ解釈しデータの利用を行う。

地域メッシュ統計の場合、疑似的個体の立場に立つのが個々のメッシュである。ただし、地域メッシュ統計の場合それが地図上に表記されることから、それぞれの統計がいわばレイヤーとしてメッシュ上にオーバーレイされているとの解釈が成り立つ。このように解釈することによって、レイヤー相互間の分析利用が可能となる。

統計をレコードとしてデータ統合という視点から捉えた場合、表式調査が与える統計から集計表、そして地域メッシュ統計へと、個体に擬制された集団を軸とした諸統計のレコード的解釈が次第に現実的意味を持ってくる。ただし、集計表や地域メッシュ統計のデータ統合的解釈はあくまでも個体に擬制された集団とその集団に係る集計量というデータ形態によるものである点で、以下に述べる個体レコードによるデータ統合とは本質的に異なる。

個票調査によって収集された調査票情報は統計単位と 1 対1に対応づけられた統計原情報で個体レコードというデータ形式を持つ。そこでは統計単位についての集団の個体としての擬制もなければレコード値についても集計量としての統合による平均化・相殺作用の介在する契機もない。調査票情報の担い手はまさに個体としての統計単位そのものである。個体を特徴づける各レコード値は個体に直結したものであり、個体に関して諸変数が相互に関係づけられている。

このような個々の統計単位毎に切り分けられた形での調査票情報の取得は、統計原情報を最初から集計量として把握する表式調査では決して行い得えなかったもので、統計調査への調査個票(小票、個別票)の導入という統計調査法の転換がはじめてそれを可能としたのである。

本稿では、データのアーカイビング技術、経緯度情報の観測精度の向上や操作可能な情報形態への転換を含む情報処理技術のその後の展開を踏まえ、データ論として調査個票導入の意義の再検討を試みた。その結果、個票調査によってはじめて収集できる個々の統計単位に関する調査票情報について、それ自身のデータ形式が初めて可能にする統計の新たな展開可能性が浮かび上がることになる。

画像や音声といったデジタル情報は、情報の担い手である **data carrier** とそれによって担われ個別 **carrier** を特性づける **data body** 部とから構成される[森 2009a]。統計もまた同様であり、調査個票によって収集される調査票情報は、個々の統計単位に関する統計原情報を与える。

筆者は調査票情報から作成される個体レコードを **data carrier** とそれによって担われている **data body** から構成されていると捉える。このように調査票情報に含まれる情報要素をその機能に従って **data carrier** と **data body** とに明確に切り分けることによってはじめて、われわれは集計量としての統計にはない個体レコードが持つ個体ベースでのデータの拡張可能性という個体情報としての調査票情報に固有の情報特性の存在を確認することができる。

統計単位の個体識別情報を **data carrier** として位置づけることで、調査票情報を **data carrier** としての識別情報とそれと個体ベースで結合された **data body** というデータ論的に異なる機能を持つ情報要素の組み合わせとみなすことができる。調査票情報をこのような理解の自然の帰結として、既存の調査票情報の事後的マッチングによる調査票情報の横断面、縦断面方向への変数次元の拡張が可能となる。このことは、既存の調査への調査項目の追加や新たな調査を企画、実施が予算や報告負担の面で現実的でない中、既存の個々の統計調査の結果からは得ることのできない新たな変数の組み合わせによる分析を可能にする。

行政情報の中には、対象を悉皆的に把握されているものも多く、また統計単位と共通の個体識別情報を持つものもある。こういった種類の行政情報については、個体識別情報をリレーショナルキー変数とした調査票情報との潜在的リンク可能性を持っていることになる。このことは、従来、統計情報と行政情報というそれぞれ異なるセグメントの情報として取り扱われてきたものが、相互にデータ統合の余地を持っていることを意味する。事実、統計作成に係る予算制約が厳しくしかも調査環境がより深刻な海外では、行政情報が持つ情報価値に注目し、その統計への広範な利用がすでに制度化されている。そのような中、行政情報の統計への組織的活用をはかるために行政情報に対しても統計と共通の識別コードの付与を義務づけている国もある。このような国では、行政情報も含めた統計・行政情報統合システムがすでに構築されている。

集計量としての統計の場合、個々の統計単位に正に個別的に作用を及ぼしている場所的要因は、その域内に所在する全統計単位の間で平均化されている。これを個体レコードとしてみれば、平均化され等しい場所的特性値を持つベクトルが調査票情報にそれぞれ付加されていることになる。

空間的存在である個々の統計単位を規定している一連の場所的要因は、おのおのが多様な境域を持つポリゴン群から構成され、個体の存在を規定している。個々の場所的要因がそれぞれ一つのレイヤーを構成し、個々の統計単位に作用を及ぼしている。

このことを個々の統計単位側から見れば、調査票情報がそれぞれ異なる場所的特性値を要素とする場所的特性ベクトルと1対1(あるいはn対1)として関連づけられていることを意味する。さらにそれを個体レコードというデータ形式として見れば、統計単位の空間的存在の場を示す経緯度情報をキー変数として、調査票情報に対して場所的特性変数ベクトルを事後的に追加できることを意味する。

調査票情報を変数として経緯度情報を持つことの意義は大きい。結果の表章面では、それが個体ベースでの経緯度情報を持つことで、これまでの地域集計にはない多様な空間集計が可能となる。また、統計の作成面においても、場所的要因を層化変数として持つ標本設計の可能性なども今後の課題となりうる。

こういった統計の作成、利用面での可能性に加え、個体レコードが経緯度情報を変数として持つことにはデータ論としても意味がある。なぜなら統計単位の空間的存在の場を規定している場所

的要因に関する変数を経緯度情報をキー変数として調査票情報を表す個体レコードに事後的に付加でき、そのようにして編成された拡張レコードを用いることによって場所的作用も含めた要因分析が可能となるからである。

個票調査が公的統計に導入されてすでに 1 世紀半ほど経過した。個票調査の表式調査に対する最大の優越は、まさに個体情報として調査票情報が収集できる点にある。しかしながら、それは専ら統計作成技術面における進歩として捉えられ、「統計＝集計量」という伝統的統計観もあり個体情報が持つ潜在的価値の側面にはこれまで関心が向けられてこなかった。その結果、地域表章においても個体情報として現実には存在する調査票情報が調査区という面的集団情報として集約され地域表章が行われてきた。

このような個体情報あるいは点的情報が集団的面的情報として取り扱われることによって、調査票情報はそれが本来有していた個体ベースでのデータ統合に際してのキー変数として機能する個体識別情報や場所的要因のデータ化に不可欠な統計単位の空間的所在情報を喪失したのである。

21 世紀の現在、世界の統計は行政情報も含めた個体ベースでのデータ統合へと大きく舵を切りつつある。個票調査によって収集される調査票情報が本来的に有していた個体レコードの拡張、個体ベースでのデータ統合という統計の新たな展開可能性に改めて注目すべき時代であるように思われる。

【文献】

鮫島龍行・相原茂(1971)『統計日本経済』筑摩書房

高橋二郎(1905)「明治 12 年末甲斐国現在人別調顛末」(統計集誌第 228-9 号)『明治大正期統計学雑誌論文選集』(日本統計協会)所収

長谷川普一・卯田強(2012)「災害時における全住民位置情報の重要性(その2)―アドレスマッチングに係わる諸問題と解決策について―」GIS システム学会予稿集 c-1-3.pdf

森博美(2008)「政府統計体系について」『統計』(財)日本統計協会 2-7 頁

森博美(2009a)「統計個票情報の情報特性について」『経済志林』法政大学経済学会 第 76 巻第 4 号 403-427 頁

森博美(2009b)「統計個票情報の情報特性とその潜在的利用可能性」『日本人口学会第 61 回大会報告予稿集』 37-38 頁

森博美(2010)「ジオコードによる統計の情報価値の拡張可能性について」『経済統計学会第 54 回研究大会報告予稿集』 22-23 頁

森博美(2011a)「統計調査における地点情報の把握による統計の情報価値の新たな展開可能性について」『経済志林』法政大学経済学部学会 78-3 249-281 頁

森博美(2011b)「位置情報を用いた調査票情報の情報価値の拡張とその分析的意義について」『オケーショナルペーパー』法政大学日本統計研究所 No.25 1-25 頁

森博美(2011c)「駿河国人別調沼津・原政表―調査票から見た調査の性格規定を中心に―」経済統計学会政府統計研究部会ニュースレター No.15 30-38 頁

森博美(2012a)「駿河国人別調沼津・原政表再論」『オケーショナルペーパー』法政大学日本統計研究所 No.28、 1-21 頁

森博美(2012b)「データ統合の観点から見たデータキャリア情報の統計的利用可能性について」

- 『経済志林』法政大学経済学部学会 79-3 185-217 頁
- 森博美(2012c)「人別改帳情報の情報特性について-寛政 8 年人別書上改正を中心に-」『経済志林』法政大学経済学部学会 79-4 203-245 頁
- 森博美(2012d)「場所特性変数の付加による個体レコードの拡張について」『オケージョナルペーパー』法政大学日本統計研究所 No.36、1-30 頁
- 森博美(2013a)「行政記録情報の情報形態と表式調査」『ディスカッション・ペーパー』法政大学日本統計研究所 No.1、1-10 頁
- 森博美(2013b)「統計の社会的性格と調査票情報について-蝮川における統計の場所的規定性を中心に」『ディスカッション・ペーパー』法政大学日本統計研究所 No.2、1-12 頁
- 森博美(2013c)「昭和 15 年農林統計改正と調査票情報について」『オケージョナルペーパー』法政大学日本統計研究所 No.38、1-26 頁
- 森博美(2013d)「調査票情報論の視点から見た D. Harvey の空間言語について-相対空間の実質言語化による調査票情報の拡張-」『経済志林』法政大学経済学部学会第 81 巻-1 号、1-35 頁
- 森博美(2013e)「統計調査票情報と場所的特性について」『経済統計学会第 57 回全国研究大会報告要旨集』59-60 頁
- 森博美(2013f)「統計の調査票情報とデータインテグレーションによる情報の拡張」『日本公衆衛生雑誌』第 60 巻第 10 号特別付録、61-62 頁
- 森博美(2013g)「わが国農業生産統計における表式調査の展開-府県物産表から昭和 15 年農林統計改正まで-」『ディスカッション・ペーパー』法政大学日本統計研究所 No.3、1-11 頁
- 森博美(2013h)「明治期における個票による農村実態の統計的把握の試み-余土村是調査における下調べ個票様式をめぐる-」『ディスカッション・ペーパー』法政大学日本統計研究所 No.4、1-12 頁
- 森博美(2014a)「村是調査における調査様式の展開-余土村是調査「統計調書」と『町村是調査指針』の雛形小票の調査項目比較を通じて-」『オケージョナルペーパー』法政大学日本統計研究所 No.41、1-26 頁
- 森博美(2014b)「『町村是調査指針』における小票論」『ディスカッション・ペーパー』法政大学日本統計研究所 No.5、1-9 頁
- 森博美(2014c)「明治 31 年内閣訓令第 1 号乙号と調査票情報」『オケージョナルペーパー』法政大学日本統計研究所 No.42、1-20 頁
- 坂本憲昭・森博美・高橋朋一(2014)「事業者の退出、参入に伴う事業所の業種間交代について-東京都八王子市を事例として-」『研究所報』法政大学日本統計研究所 No.44、31-42 頁
- 総務省統計局「地域メッシュ統計の特質・沿革」<http://www.stat.go.jp/mesh/pdf/gaiyo1.pfd>
- Jarmin S. Ron and Miranda, Javier, (2006) The Impact of Hurricanes Katrina, Rita and Wilma on Business Establishments: A GIS Approach, *Working Paper*, pp.8-26, Center for Economic Studies, U. S. Census Bureau.

日本統計研究所

オケージョナル・ペーパー(既刊一覧)

号	タイトル	刊行年月
12	「統計法」と法の目的	2005.07
13	諸外国におけるマイクロデータ関連法規の整備状況とデータ提供の現状	2005.09
14	統計に係る個人情報の秘密保護について	2006.08
15	若年層における雇用状況と就業形態の動向—『就業構造基本調査』のマイクロデータによる実証分析	2006.12
16	社会生活行動から見た若年層の不安定就業化・無業化の分析	2008.03
17	国勢調査による従業地把握の展開と従業地別就業データの意義	2009.06
18	無償労働の評価と世帯生産サテライト勘定	2009.10
19	エンゲルとザクセン王国統計	2009.12
20	第一次統計基本計画と政府統計の直面する課題	2010.01
21	エンゲルとプロイセン統計改革	2010.02
22	エンゲルと1875年ドイツ帝国営業調査	2010.03
23	調査形態論再論	2011.03
24	統計を規定する諸要因との関連から見た時空間個体データベースの可能性について	2011.04
25	位置情報を用いた調査票情報の情報価値の拡張とその分析的意義について	2011.06
26	ジオコード情報の活用による統計の把握精度改善の試み	2011.09
27	統計的マッチングによる疑似パネルデータの作成と精度検証	2011.11
28	駿河国人別調沼津・原政表再論	2012.01
29	ザクセン王国統計協会(1831-50年)	2012.01
30	ザクセン王国における初期人口・営業統計	2012.02
31	フィンランドのビジネス・レジスター	2012.03
32	エンゲルのザクセン王国統計局退陣をめぐって	2012.04
33	フランスのビジネス・レジスター	2012.05
34	タウンページ情報を用いた事業所の自然・社会動態の把握	2012.07
35	疑似景況パネルによる予想パフォーマンスの計測	2012.11
36	場所特性変数の付加による個体レコードの拡張について	2012.12
37	フランスの新人口センサスにおける詳細な統計結果の推計方法	2013.03
38	昭和15年農林統計改正と調査票情報について	2013.04
39	1855年ザクセン王国営業調査について	2013.07
40	Estimation of the Start-up, Closure and Relocation Rates of Local Units	2013.09
41	村是調査における調査様式の展開	2014.01
42	明治31年内閣訓令第1号乙号と調査票情報	2014.05

オケージョナル・ペーパー No.43

2014年8月30日

発行所 法政大学日本統計研究所

〒194-0298 東京都町田市相原4342

Tel 042-783-2325、2326

Fax 042-783-2332

jsri@adm.hosei.ac.jp

発行人 森 博美