

オケージョナル・ペーパー No.24

統計の基本的要素との関連から見た 時空間個体データベースの可能性について

2011年4月

法政大学

日本統計研究所

統計の基本的要素との関連から見た 時空間個体データベースの可能性について

森 博美(法政大学経済学部)

要約

政府統計調査によって得られた調査票情報は、これまで基本的に集計結果を作成、提供するまでのいわば一過性的性格のものとして扱われてきた。広い意味での統計作成環境が悪化する中で諸外国では、統計調査に依存しない新たな統計情報獲得の手段としての既存情報の更なる有効活用のために、統計調査票情報の体系的保存というアーカイビングシステムの構築に向けての取り組みが、統計行政における重要な政策課題となってきた。本稿では、このような世界の統計の新たな展開方向を受けて、わが国における今後の政府統計データのアーカイビングシステムの構築に向けての準備作業の第一歩として、アーカイブにデータを格納する際のデータ形式のあり方についての基礎的、方法論的研究を行った。

本稿では、統計を規定する諸要因に関して、マイヤー、ジージェックそして蜷川の所説を中心として、その内容についての検討を行った。その結果、単位、標識、時間、そして場所といった諸概念が、マイヤーにおいては現実を反映した統計の作成の指針ならびに結果資料の中に貫く規則性の検出装置として、ジージェックにおいては比較可能な統計資料獲得の方法工程を導く基本的要素として、そしてまた蜷川においては、統計資料の吟味によるその批判的利用といういわば政府統計の二次利用の統計学の理論構築の基礎概念としての集団論(大量論)の視点から展開されていることが明らかにされた。

彼らの論点の中には、今日的視点からもなお意味を持つ点がいくつか含まれる。とはいえ、その後の統計の展開さらには本稿での問題関心である時空間個体データベースの見地から見た場合、いくつかの決定的な問題を内包しているのも事実である。本稿では、統計を規定する諸要素に関する彼らの指摘を批判的に紹介しつつ、時空間個体データベースにおける変数値の規定要因さらにはデータ形式についての試論を示した。

[キーワード]

アーカイブ、調査票情報、個体情報、時空間データベース、単位、大量の四要素、位置情報

はじめに

本稿は、政府統計データのアーカイビングシステム、特に、アーカイブにデータを格納する際のデータ形式のあり方に関する方法論的研究を中心的な検討課題としている。個体ベースでの政府統計データのアーカイビングについては、これまで、個体識別コードをリレーショナルキー変数とするリレーショナルデータベースが一般に考えられてきた。筆者はかねてから、個体識別コード以外にもリレーショナルキー変数となりうるものが統計情報においては存在するのではないかと、またそのような変数をリンクキーとした場合、どのような新たなタイプのデータセットを編成することができ、それはまた分析資料としてどのような利用可能性を持ちうるか、といった事柄に関心を寄せてきた。

特に、これまでの統計データベースにおいて、客体が把握される時点や場所が持つ情報が不当に過小評価されてきたのではないかということについての疑問を抱いてきた。そのような中、時空間データベースという発想に感化され、それに統計が従来からリレーショナルキーとして取り扱ってきた個体という要素を加えた、時空間個体データベースなるものを提起してみたいということが、その研究動機となっている。

このような今日の研究課題に対して、本稿で G. マイヤー、F.ジージェックそして蜷川虎三といった諸家における統計を規定する諸要素の吟味という、どちらかといえばやや訓詁学的とも受け取られそうなアプローチをとったのは、次のような事情からである。すなわち、蜷川が集団論を前面におし立てて構築した統計学のむしろその背景にある統計対象観、すなわち統計が反映すべき現実を彼がどのような視点から捉えていたかを、恐らく蜷川に少なからず理論的教示を提供したと考えられるドイツ社会統計学派の巨人、マイヤーとジージェックの所説などとも関係づけながら検討する中から、アーカイビングされる変数値を規定する要因について、もしかしたら有意義な示唆が得られるのではないかと考えたからである。本稿で取り上げる諸家は、指導的統計家あるいは統計研究者として、いずれも草創期の社会統計学の構築に中心的役割を果たし、それぞれ壮大な体系をもつ理論として構築されたものである。

それぞれの理論展開を総体として読み説きそれに対して論評を加えるのが目的ではなく、あくまでも上述のような筆者自身の今日の問題意識に照らし、それに対する思考をめぐらす上での示唆となりうる部分に焦点を当て、それを参考にしつつ自説の展開を試みることにしたい。

1. 統計を規定する諸要因に関する諸家の見解

蜷川統計学は、「其の存在が社会的に規定せられた集団」[蜷川(1933) 68 頁]をとくに「大量」と呼称し、それを安定的結果の析出のために意識的に構成された集団から明確に区別する。それは、これら性質を異にする二種類の集団から構成される独自の集団論に立つことによって、ドイツを中心に展開されてきた調査論を中心とする社会統計学と主として英米において展開されてきた数理統計学とを統合した理論体系の構築を目指している点にその大きな特徴がある。このような蜷川統計学の基底概念である彼の集団論について、その後、主として社会科学方法論説に立つ諸家から、蜷川集団論の二元論的性格を含め、様々な批判的見解が示される。しかしながら、冒頭で述べたような理由から、ここでは蜷川の集団論それ自体を巡ってその後わが国の社会統計学派の中で展開されることになる統計対象論、統計集団論そのものに立ち入ることはしない。

蜷川にとっては、社会的に規定された集団としての大量を語る数字こそがまさに統計である。また大量は、「大量の単位、標識、存在の時及び存在の場所」という四つの要素によって特徴づけられる。蜷川はこれらの要素を「大量の四要素」として、「大量に就いて不可欠の要素」[蜷川(1933) 99 頁]としている。その結果蜷川にとっては、統計による現実認識とは、「大量を認識し之を数量的に把握すると云ふこと」、換言すれば「大量の四要素を規定し、大量の大いさ及び部分大量の大いさを求めることを指すものに他ならない」[蜷川(1933) 99 頁]。このように、大量を統計による社会認識の基底概念として礎定していることから、蜷川は、統計数字を特徴づけるこれら四つの要素についても、あくまでも大量との関連において捉えている。

ところで、蜷川が大量を特徴づける四つの基本的要素として提起したものは、そのすべてが彼の知的創造によるものでは必ずしもない。マイヤー(Georg von Mayr)は社会集団の把握に単位ととも

に存在の時と場所という要因を挙げている。またジー・ジェック(Franz Zizek)には、現実を統計的に把握するための概念装置として、調査単位(Einheit)、調査標識(Merkmar)、それに群(Gruppe)と統計的表記(Aussage)という4つの要素を取り上げている。そのうちの前二者は統計資料の獲得過程(調査過程)に、また後二者は得られた調査資料の整理過程(集計過程)に関するものである。蜷川における大量の四要素は、一方の極として当時の世界の統計学をリードしていたドイツ社会統計学における到達点を部分的に継承しつつ、独自の唯物論的認識論という視角から定式化し直したものと見ることができる。

表1は、蜷川理論の形成過程において少なからず知的投入を与えたと思われるマイヤー、ジー・ジェックとともに、蜷川における大量の四要素を一覧整理したものである。

表1 マイヤー、ジー・ジェックそして蜷川における統計の諸要素

マイヤー	単位		時	場所		
ジー・ジェック	単位	標識			群	表記
蜷川	単位	標識	時	場所		

表1の整理からもわかるように、単位については、マイヤー、ジー・ジェック、そして蜷川が共通に統計を特徴づける要素とみなしている。また蜷川は、時と場所についてはマイヤーと、また標識についてはジー・ジェックとその要素を共有している。

以下に、諸家がこれらの要素をどのような内容のものとして捉えていたか、その異同も含め、統計の諸要素について概観してみよう。

(1) 単位

マイヤーは、指導的統計家として調査実施者の立場から、時間、場所とともに、統計の把握対象が持つ3つの要素のひとつに調査単位をあげている。

彼は、人間集団、人の行為および事件の集団現象、かかる行為および事件の集団結果[マイヤー(1914) 124 頁 なお、頁は邦訳書による。以下同様]を「社会集団」として定義し、大量観察によって社会集団に関する情報を収集することによって、その中から規則性を発見することを統計学の究極の目標として設定している。そのようなマイヤーにとって、統計学的方法的観察の前提となっているのが、分析目的に応じて社会集団を明確に限定すること[マイヤー(1914) 142 頁]であり、そのためには、「集団要素すなわち問題たる具体的統計単位を正確に一義的に規定すること」[マイヤー(1914) 143 頁]すなわち、当該調査において単位と看做すべきものとそうでないものを峻別することが何よりも重要となる。なぜなら、調査対象の明確な限定は部分集団の観察においても必要であり、そのことなしには、集団と部分集団の関係を明らかにすることができないからである。このようなことから、「統計単位概念契機を、根本的に極めて明確に規定することは実用統計学の凡ゆる部門に亘って極めて重要である」[マイヤー(1914) 144 頁]。

マイヤーは、調査単位の定義について、調査によって把握すべき集団要素を調査単位として規定し、例えば、人間が調査対象となる場合はその現在員数を、また建築物統計の場合には家屋を明確に定義づける必要があると指摘している[マイヤー(1914) 143 頁]。しかし他方でマイヤーは、単位を数えたり測ったりして積み上げることで直ちに集団(部分集団)が得られるという考えに立つ。

このような彼の主張からも明らかなように、マイヤーは個人や世帯、企業といった報告(回答)単

位を「単位」とみなしているのではなく、調査によって把握したい項目の数・量・額のうち、調査単位として規定した定義に該当する調査事項の数量そのものを単位と見なしている。例えば、穀物の生産量調査の場合、単位となるのは、穀物の生産農家という報告対象ではなく、その農家が生産した穀物のうち当該調査で把握対象として定義された穀物の数量そのものなのである。

すなわち、指導的統計家として調査の企画・実施者の立場に立つマイヤーにとっての調査単位という関心事項は、穀物の生産活動という人間の行為の結果に関する情報を持つ調査客体ではなく、穀物の収穫量についての比較可能な数量を調査によって直接的に把握する際の原単位情報に他ならない。言い換えれば、それは、種々の属性情報とともに穀物の生産量(生産額)という情報すなわち個体調査票情報の担い手としての調査客体ではなく、変数値である。

ジーゼックは調査単位を、「どの統計調査にとっても最も基礎的な概念」[ジーゼック(1922) 11 頁 なお、頁は邦訳書による。以下同様]であるとして、具体的な調査設計に際しての数え上げの単位であるとしている。従ってその場合、調査によって把握すべき現象については、「調査のさいにいかなる「要素」に分解されるべきか、いかなる「要素」が数え上げられるべきか、が明らかでなければならず、したがって「調査単位」(「数え上げの単位」)が定義されなければならない」[ジーゼック(1922) 6 頁]。

例えば、複合住宅を1戸として把握するかあるいは数戸とみなすか、現在人口と常住人口のいずれを把握するか、あるいは把握すべきストライキの定義を明確化しておく必要[ジーゼック(1922) 12 頁]があり、調査単位については、「多くの調査において、正しい統一的な手続きで調査が行われるためにも、また、調査結果の利用にあたって調査結果が本来何を意味するかが知られるためにも、全く厳密に定義されなければならない」[ジーゼック(1922) 7 頁]。ジーゼックがこのように住宅や人口それにストライキを調査単位すなわち「数え上げの単位」として例示しているの中には、単位に関して依然としてマイヤー的単位概念の痕跡を部分的にとどめているように思われる。

ところで、ジーゼックは彼の最後の著作である『統計数字の成立』[ジーゼック(1937)]において、調査単位の確定を、統計調査による確認のための基準として抽象的に調査単位の概念を行う行程と、この決定に基づいて調査単位を具体的に決定する調査単位の組織的・技術的決定という二段階からなる行程との関連で捉えるという立論を行っている。

抽象的理論的過程において抽象的に規定されていた調査単位概念は、その後継過程の中で、「申告者(Auskunftspersonen)」[ジーゼック(1937) 30 頁]として具体的に規定される。ここでは、それを集計することで直ちに(部分)集団の大きさを示しうる変数値(標識値)というマイヤー的調査単位概念からジーゼックが一步脱皮している事実を読み取ることができる。そこではジーゼックは、調査単位を報告担当者として捉えている。しかし、彼が言う「申告者」とは世帯や企業等で調査票への記入担当者に他ならず、企業や世帯といった調査客体そのものではない。

このようなマイヤーやジーゼックの調査単位概念に対し、蜷川におけるそれは一体どのようなものであったのだろうか。

まず、蜷川における統計の定義から始める必要がある。蜷川は、「大量観察の結果たる一団の数字としての統計」を、「大量の大いさ或は部分大量の大いさを示す数値の総称」[蜷川(1933) 99 頁]として捉える。そして、大量に関する個々の数値は「統計値」と呼ばれ、単なる個体の測定結果として得られる測定値あるいは統計値から導出される誘導統計値とは範疇的に区別される。

このように測定値と誘導統計値を統計値から明確に区別する根拠として蜷川は、測定値が大

量観察の結果ではないこと、また誘導統計値については、「一定の目的を以て統計値を結合したる結果」〔蜷川(1933) 99 頁〕であること、すなわち意識的に構成された集団がその背景にある点をその根拠としている。このことから蜷川は、統計を社会的存在としての大量という集団を直接的に集計量の形で数値表現したものとして捉えている。

蜷川においても、大量という社会的集団を特徴づける要素の一つとしての単位とは、いうまでもなく個数や人数、それに数量や金額といった計量単位などではない。しかし、「統計＝集団」説に立つ蜷川にとっては、主要な関心事はあくまでも大量という集団であり、単位については、統計が反映すべき「(大量が一引用者) 個別存在を其の組成因子とする」ことから「之を単位(Einheit)と謂ふ」〔蜷川(1934) 35 頁〕、とその定義については、さして踏み込んだ議論の展開は行っていない。

蜷川が例示的に「個々の失業者は…失業者なる大量の単位であり、此の単位の数即ち失業者数が先ず問題である。」〔蜷川(1933) 111 頁〕というとき、蜷川は、単位概念についてマイヤーと共通の理解に立っている。すなわち、蜷川の単位には、標識に従って部分大量を構成する単位の数または量という程度の意味しか与えられていない。

大量を理論構築の基底概念として据える蜷川の場合、四要素とはあくまでも大量の四要素であり、それが統計作成の後工程において大量観察の四要素として制度的・技術的に捉え直されるとはいえ、そのひとつを構成する単位概念についても、次に見る標識とともに、社会的存在としての大量を特徴づける統計値を形成する直接的構成要素として理解されているように思われる。そこには、集計処理される数や量という個々の統計情報の担い手としての個人や世帯、事業所それに企業といった報告の客体としての単位認識はない。それは、蜷川が統計を何よりも先ず大量という集団を表現する数字としたことの自然の帰結であるように思われる。

(2) 標識

マイヤーは、統計を特徴づける要素として標識概念を持っていない。そこで、ここでは、ジージェックと蜷川が標識に対してどのような意味づけを与えていたかを見ておくことにする。

ジージェックは、統計調査の企画から実施、さらには集計表の作成に至る一連の統計作成過程(統計数字獲得行程)の中で、標識に対する規定を調査単位の確定に続く作業行程として位置づけている。すなわち、調査単位の定義を行った後に「調査単位について観察されるべき「調査標識」が明らかにされなければならない」〔ジージェック(1922) 6 頁〕のである。

ジージェックは標識を、調査の結果得られた単位を集計表の形で整理する際の基準として認識している。なぜなら彼は、調査単位と標識の関係について、「調査単位は…単に数え上げられるだけではない。調査単位に対してそれが調査されるさいに一定の調査標識が確かめられる—これにしたがって単位は整理のさいに分類される。」〔ジージェック(1922) 13 頁〕としているからである。ここでは統計数字の獲得行程とは逆向きに、事前に想定した結果表の表形式のイメージに基づき、調査単位のどの側面を標識として切り取る必要があり、それに照らして具体的に設定された標識の基準に従って調査票における調査項目の確定が行われ、該当、非該当単位についての具体的な線引きが行われる。

耕作面積階級別に農業経営を表章した結果表の作成を目標とする統計調査の場合、ジージェックにとっては面積が調査単位であり、耕作面積が調査標識となる。調査標識の確定が統計調査企画の一作業工程となりうるのは、それに関して「さまざまな定義が可能な場合が非常に多い」〔ジージェック(1922) 13 頁〕からである。休耕地を耕作面積に含めるか否か、職業調査で自家の家事

を処理する妻を有業者として数え上げるか、住宅統計で台所を部屋数に含めるか否か、というように「調査標識もまた明確な概念の規定を必要とすることが多い」[ジー・ジェック(1922) 7 頁]。

調査実施に先立ち、調査単位のどの側面を調査標識として具体的に規定するかは、回答者が調査客体に関する調査事項に正確に回答できるためにも、また調査員者が誤りなく実査業務を遂行できるためにもその前提となるものである。仮に調査標識の定義が異なれば、当然、調査単位として数え上げられる単位のカバレッジも同一ではなく、結果的には比較可能性が担保されない統計しか得られないことになる。

大屋祐雪は、ジー・ジェックの調査標識の確定に際しても、抽象的過程と組織的・技術的過程とがあるとして、前者では「抽象的な調査標識のうち、どれとどれとを実際に調らべ、確認せねばならぬかを決定し、決定された標識については、その概念を精密に定めること」が、またその組織的・技術的決定に関わる過程では、「この決定によって調査標識はその確認にあたっての誤解や疑問の余地が少なくなり、標識申告を記録するための「調査票 (Fragebogen)」に、調査事項として」組み込むことがそこでの課題となるとしている[大屋(1967) 355,356 頁]。

以上のように、ジー・ジェックにおいては、調査標識は、彼のいう調査単位ある側面を、調査目的として設定された集計結果表の表章イメージに従ってより具体的に捉え直し、実査過程においていかなる曖昧さをも排除しうる形で概念化し調査票における調査事項として確定する一連の作業工程を根拠づける統計の要素概念となっている。なお、付言すれば、彼は、調査項目に対して記入式の回答も含めどのような回答肢を用意するかについては「群」という標識に続く第三の統計要素を設定している。

蜷川は、標識についても、単位の場合と同様、大量との関連でその性格規定を行っている。彼は、「大量に就いて予め規定する集団性の方向」[蜷川(1934) 36 頁]、すなわち、部分大量を構成しうる指標のことを標識 (Merkmale, Erhebungsmerkmale) [蜷川(1933) 98 頁]と呼ぶ。大量と部分大量、さらには個々の部分大量の大きさを計量し集団性の強度を評価する、そのような特定の集団特性を表現する指標として蜷川は、標識を大量の四要素のひとつとして捉える。

このように、独自の集団論に立脚して統計学の理論的再構築を目指す蜷川においては、標識もあくまでも集団論の基底概念である大量との関連で「大量の四要素」としてその位置づけが与えられている。そこでは、マイヤーよりもさらに現実の統計作成過程の実務との関連でその方法的展開を行ったジー・ジェックにおける標識規定が、調査単位のいわば標識単位化、すなわち、標識という調査技術的に操作化した概念を導入することにより、調査目的に照らして調査単位の特定の側面を実査の局面において選択的に把握するための概念装置として具体的に位置づけられていたのと較べれば、実際の統計作成との関連はやや希薄である。

ジー・ジェックと蜷川の標識に関する論点を、一方で調査過程においては厳密に定義づけられた調査項目として、他方、調査結果の表示形態として一定の定義に従って収集された情報の変数という視点から比較評価した場合、蜷川よりはむしろジー・ジェックの標識規定の方が実質的含意を持っているように思われる。

(3) 時間

上に見たように、統計の作成過程の中に標識をその不可欠の結節的概念として位置づけていたジー・ジェックではあるが、統計に対する時間的要素の関与については、余りにも自明のこととして殆ど関心を寄せていない。

蜷川は、大量が時に関する規定を持つのはその「社会的存在」によるもので、その結果として「一定の存在の時と存在の場所を有つことは云ふまでもない」〔蜷川(1933) 99 頁〕として、極めて限定的な意味づけしか与えていない。彼は時間という要素が「大量に就いて不可欠の要素」〔蜷川(1933) 99 頁〕の一つであるとしながらも、その持つ意味については、いまさら「云ふまでもない」として特に立ち入った言及は行っていない。

蜷川にとっての最大の関心事は集団論であり、時間要素についても大量との関連においてその言及が行われている。集団論視点に立つ蜷川にとって、時は、あくまでも静大量、動大量という集団としての大量の類別区分との関連で意味を持つものとして位置づけられている。そのような集団の区別に立ち、静大量集団については統計的把握の時点、また動大量集団については期間という時の選び方という統計作成技術に関わる事項がその主たる関心事である。

このようにジージェックと蜷川が時の規定について、全くあるいは事実上ほとんど特別な関心を払っていないのとは対照的に、マイヤーは、時について踏み込んだ議論展開を行っている。

蜷川の大量概念とは若干異なり、マイヤーは、「社会集団は時の流れのうちにある」〔マイヤー(1914) 146 頁〕として、社会集団への時間要素の関与をより明示的に意識している。このような視点は、「社会集団の観察に際しては結局、個々の時点が集団の限定にとって基準的であるか、あるひは、時の経過につれて社会要素の発現を観察せんとすることが問題であるかの何れかである。」〔マイヤー(1914) 146 頁〕としてマイヤーの統計調査論を特色づけている。

マイヤーの場合も、時間要素に関して時点と期間という側面から統計の対象集団を静態集団、動態集団とを区別している点は蜷川と共通である。また、マイヤー自身、時点と期間の規定については、「場所的限定に比し、きはめて截然かつ明白に…規定されうる」〔マイヤー(1914) 147 頁〕として、後述の場所と比較して、その規定はさほど複雑なものではないとしている。

しかしながら、上記のような「社会集団」観に立つことからマイヤーは、統計における時間に関して、様々な視点から立ち入った議論を展開している。

まず、静態集団を対象とする静態調査についてマイヤーは、把握時点の同時性が決定的に重要であるとする。なぜなら、「調査瞬間の斉一を欠く場合、いかなる同質的(比較可能な一引用者)結果をも得ることができない」〔マイヤー(1914) 148 頁〕からである。指導的統計家として調査の企画・実施者の立場に立つマイヤーは静態調査の調査時点の具体的な設定方法についても言及しており、社会集団につきそれが最も普通の状態であり(内部的理由)、しかも行政機関等が調査機関の意の如く動いてくれる時期(外部的理由)に設定すべきであるとしている。

マイヤーにおける時間規定で特に注目すべき点は、動態調査あるいは動態的記録の取り扱いについての指摘である。「静大量自体は時の流れの中に静止せる瞬間集団である」〔マイヤー(1914) 346 頁〕のに対し、動態集団の場合、取り扱いの対象となるものそのものが「時の関数である出来事」〔マイヤー(1914) 202 頁〕であり、「あらゆる変動現象の本質は観察要素の時間的差異といふ点にある」〔マイヤー(1914) 346 頁〕。つまり、動態集団は、それが観察者の意思によってはどうにもならない「正確な時間的特質」を帯びている点をその特徴とする。

動態集団の把握ならびにその統計的表章について、時間要素の場合、期間の設定が比較的容易であるとして、マイヤーは次のように指摘している。すなわち、「変動現象の場合には、観察ならびに整理活動を行ふためには、時の長さを標準として現象を限定せねばならないが、その際、場所的限定の場合に比して著しく有利な点として、大体において等質かつ明瞭な慣行的な時の長さが用ひられることである」〔マイヤー(1914) 346 頁〕と。

統計による把握の対象となる事象そのものの生起は瞬間的性格のものであるが、それをどう記録するかについては、統計作成者側の判断が入り込みうる。すなわち、記録そのものは、「出来事の発現した個々の暦日を査定するのが普通であって、とくに必要のある場合に限りその発現の時刻をも査定する」[マイヤー(1914) 202 頁]ことになる。

(4) 場所

調査単位の存在の場所に関するジーゼックの見解は、「場所的総括は、通例、政治行政区域に・・・による」[ジーゼック(1922) 8 頁]と、特段の考察を差し挟むことさえない極めてそっけない内容のものである。「通例」の適用対象外となる場所規定の存在の有無についても、これといった言及はない。

調査結果の地域的表章形態について、通例の政治行政区域でいわば事足れりといったジーゼックにおける場所の消極的な取り扱いには、統計作成がいわば日常的に反復遂行される経常的な行政行為となっているという当時の現実の統計の展開と関連した次のような事情がその背景となっているように思われる。

多くの統計調査が周期的に反復実施される中で、統計作成の目的も半ば所与のものとして前提されるとともに、地域表章についてもその慣行が尊重されることになる。このような現実の統計作成の展開の中で、消極的には統計調査の反復性によって、またより積極的には同一調査系列間の時系列的さらには地域間の比較可能性の担保という側面によって、通例外として取り扱われるべき場所的要素が、無視あるいはむしろ積極的に排除されるべきものとしておそらくジーゼックにおいて認識されるようになったものと考えられる。ジーゼックが統計の地域表章を基本的に政治行政区域によるとする彼の場所に関する観念は、このような統計作成の現実の反映でもある。その結果彼は、場所規定については、時間に関するそれと同様、そもそも統計を規定する基本的要素としては取り上げていない。

蜷川の場所に関する見解は、それを専ら統計作成技術の観点から展開したジーゼックとはやや異なる。なぜなら、時の規定の場合と同じく場所規定についても蜷川は、「大量の存在の場所も同じ意味に於いて重要である」[蜷川(1934) 38 頁]として、あくまでも大量という集団現象の統計的認識との関連でそれを論じているからである。

大量という集団論に立つ蜷川の場合、ジーゼックのように地域区分は必ずしも行政的区画とは一致しない。すなわち、「場所と云っても必ずしも単に地理的或は行政的区画を意味するものではなく、大量が存在する社会的領域をいふのである。普通に地理的或は行政的区画がとられるのは大量観察に於ける技術上の理由からで、必ずしもそれが客観的な大量の存在の場所を示すものでないことは注意すべきである。」[蜷川(1934) 38 頁]、また、「場所は、必ずしも行政的区画を意味するものでなく、要するに、単位の一団が大量として限定せらるゝ空間的存在条件たればよい。従って生産米を市場により、或は又、交通路により平野によって大量として限定することも、かく規定せざるを得ざる必然性を有つならば、その方が寧ろ正しいのである。」[蜷川(1933) 160 頁]として、行政的区画以外の統計の表章の可能性を示唆している。

それでは、どのような場合にジーゼックのいう「通例」でない地域区分を持つ統計の作成が可能であるのか。この点について蜷川は、大量という存在の特性をどのように捉えるかという社会科学の理論に裏打ちされた集団論から、その区分の方向は自ずと明らかになるとする。すでに(3)で論じたように、時の規定について蜷川は静大量、動大量という集団の特性の本質的差異と関連づけ

て時点と期間という理論展開を行っている。一方、場所については、地域区分の基準を統計作成者が持つ社会科学の理論に委ねているだけで、特に、場所という視点に立つ集団論を展開するまでには至っていない。

このように蜷川は、場所を大量存在の時とともに「大量に就いて不可欠の要素」〔蜷川(1933) 99 頁〕として大量の四要素のひとつとしながらも、その実質的な意味づけについては最終的には、大量の認識に関する社会科学の理論に委ねている。

それでは、蜷川とともに場所を統計の基本的要素のひとつとして位置づけているマイヤーは、統計における場所を一体どのように捉えているのであろうか。彼は、場所に関して、地理学の成果も取り込みつつ、空間的情報を用いた規則性の追求を軸として、実に多岐にわたる所説を展開している。

マイヤーが社会集団について、それが「国家および自治体区域とは無関係に、多種多様な分布状態で地表を蔽ふている」〔マイヤー(1914) 153 頁〕〕というとき、そこには、場所を「通例、政治行政区域」によるとしたジージェックとは明らかに異なる視点、すなわち、場所を「単位の一団が大量として限定せらるゝ空間的存在条件たればよい」〔蜷川(1933) 160 頁〕とした蜷川に近接した視点が認められる。ただし、蜷川がその空間の区分の基準を大量に対する社会科学的研究に求めるとし、場所規定についてそれ以上の追求を行っていないのに対し、マイヤーは統計地理法という統計学と地理学とを融合した方法による社会的合法則性の検出可能性という視点から独自の位置規定を論じている点で、両者の場所に対する取り組みには大きな違いが見られる。

マイヤーは、まず地理法について、「場所関係につき普通の行政区画によらず、地理学上分離している地域乃至地帯を標準として観察結果をグループ分けする場合である。その場合、さらに細かな観点からすれば、風土学上の、水勢学上の、地球構造学上の、その他等々の、地域構成が問題とされる。…この方法の特色は、所与の調査結果の総括が、その調査地域の不変の地理的状況と完全に一致しているといふ点にある。…1890 年度国勢調査においては、人口を 21 の風土学的大地域区画(沿岸沼沢地、大西洋側の平野、ニューイングランド丘陵地、草原地帯、等)によって区分し、さらに経緯度、高低、河川流域、湿度・温度・雨量状況によって区分している」と指摘し、このような「地理的要素が統計的研究のための部分地域限定の基礎となっている」〔マイヤー(1914) 338 頁〕とする。

このような地理的区分は、あらゆる統計標識に齊一的に適用される汎用的区分ではない。それについては、問題関心となるいわば標識毎に異なる地域区分がありうる。すなわち、理論的に確立された固定的な地理的地域区分ではなく、例えば児童死亡率の差異による地域区分といったように、マイヤーはいわば統計標識毎にその区分を異にするものとしてそれを想定している。このような立場からマイヤーは、「(原表の制約から一引用者) 地理学的区画を絶対厳密に境界づけうることは殆どない。地理学的地域や地帯の境界においては、諸現象に対する地理的影響を追求することは極めて困難である。(かかる場合はたとへば、地理学的境界線が市町村区を横断しているにもかかわらず、人口事情に就ては一般に市町村区全体としての原表のみしか利用し得ず…この市町村区を構成する個々の住居地の原表を全く利用し得ぬといったやうな場合に生ずるのである。)」〔マイヤー(1914) 339 頁〕とする。そこには、マイヤーの「大地域に対して示した社会集団に関する報告は、縮めた数値をその社会集団の状態や現象に就ての地理統計的細目を認識せしめうるがごとく注意が払はれている場合にのみ、その現実姿態を完全に体現しているものとして真の意義を持ちうる」〔マイヤー(1914) 337 頁〕という指導的統計家として政府統計の作成、提供の当事者とし

ての立場を超えた、マイヤーの社会集団についての合法則性の探求という社会学者としての立場を垣間見ることができる。それは、場所規定を単に通例による行政的区画を必要かつ十分とするジージェックとは明らかに異なる。

地理的区分はマイヤーにおいては、「一般行政区画によって定められた大地域の概括的結果を利用する代りに、一つの統計的に観察された具体的事実現象について、独自の地理的分布状態なかんずくその等級別の総括的分布状態を呈示する」[マイヤー(1914) 340 頁)]という統計地理法提唱のひとつの方法論的基礎概念として位置づけられている。

マイヤーの統計地理法の核心は、分析目的に応じた適切な地域区分を導入することで、その地域的比較を通じて間接的に変数間の相関関係を明らかにできる点にある。このことは、彼のつぎのような指摘からも読み取れる。「(たとへば、諸宗派と自殺者の関係は示されていないが、諸宗派と人口一般との関係が示されている場合、各種地域、すなはち、同一宗派だけの諸地域あるひは種々の割合で宗派の混合している諸地域につき、無差別の一般的な自殺繁度を宗派別に研究するならば、自殺繁度と宗派との間に存する推測上の関係についての近似的な観念を得ることができる)」[マイヤー(1914) 157 頁)]と。

調査結果についてのこのような統計地理法による解析という分析目的に照らした場合、統計調査は、場所規定に関して一体どのような対応が必要とされるのであろうか。

社会集団を具体的な統計観察の対象とするためには、「社会集団につき具体的な場所的限定」[マイヤー(1914) 153 頁)]が必要である。そこでは、場所は二つの意味で規定されなければならない。その1は、調査企画段階における調査対象範囲の設定である。これについては、統計調査は一般に国家によって実施されることから、事実上、国という空間区画が重要な意義をもつ。その2は、実査レベルでの場所の規定である。これについてマイヤーは、調査区として、「調査機関が活動の手を指し広げ、観察すべき集団の要素をその範囲内で捉へねばならぬ調査区域の規定と限定にとっても重要」であり、「第一次統計にあつては、…一般的な大規模査定の場合にあつては、調査機関の観察分野の場所限定によって、観察すべき場所区域が特別に構成されねばならない。」[マイヤー(1914) 205,206 頁)]としている。

実査過程において調査員が収集した調査票情報は、その場所規定に関していかなる地域区分において適性に処理されるべきなのであろうか。この点に関してマイヤーは、一方で調査票情報に係る個々の位置情報の存在を認識しつつも、結果の安定性の視点から適切な微小レベルでの地域区分が妥当であるとして、つぎのような判断を示している。

まず、社会集団の分布特性の差異に従って、「社会集団が…集中して現はれるかあるひは普及(分散—引用者)して現れるかにより、観察客体としての社会集団に対しそれを適切に統計に捉へるための条件は異なってくる。場所的に集中せる集団にあつては、是非とも、その集団の所在地の個別性を重んぜねばならぬ。したがってこの場合、その客体は各個に異なる観察に付され、観察結果は場所または土地の特性によって別々に示されねばならない。場所的に普及せる社会集団にあつては事情はこれと異なる。もとよりこの場合にあつても、もし理想的な完全さを期さんと欲するならば、結局は統計調査に際し、できる限り個々の場所的個別性を重視すべきであらうが、しかしひとは、一切の捉へきれない場所的個別性を残らず区別しておくことは、第一次調査にとつてもすでに困難至極であるが、さらにその調査結果の整理の段となつては全く不可能である。それはあたかも変動現象を捉へるに際し、分とか秒とかいった極微の時間的個別性を残らず個々別々に区別しておくことが不可能であると同じことである。したがってかかる場合には、社会集団を適当な下

級小区画へ場所的に概括しておかねばならぬ。」「[マイヤー(1914) 155 頁]』と。

なお、このような小地域区画の設定に当たっては、「できるだけ小さな(しかし小さすぎない)場所区画を基準として総括を行はねばならない。この区画の大きさは、信用のおける比率計算を行ひうるだけの十分大きな基本数値を保持しておきたいといふ願望のために、制約されざるを得ない。」として、具体的には「小行政地区(県、地方裁判所管区等)による区画が地理的要素として利用されえよう」[マイヤー(1914) 340 頁]』としている。

なお、小地域区画の設定に関連してマイヤーは、「純幾何学的に規定された斉一的な面積単位」[マイヤー(1914) 155 頁]』あるいは「幾何学的見地(たとへば同心地帯設定)」[マイヤー(1914) 156 頁]』として、今日のメッシュ統計、あるいはGIS統計に相当する位置的総括にも部分的に言及している。しかし彼は、いずれも社会集団の場所区画の第一次的標準としては不適当で、「観察材料自体は上述の如き最小地域区画にしたがって区分されて居ることが望ましい。」[マイヤー(1914) 156 頁]』としてそれを退けている。

マイヤー自身、「整理目的上、ヨリ上級大区域単位にしたがった方が総括の集計技術がはるかに容易」であることは認めつつも、「観察材料自体は上述の如き最小地域区画にしたがって区分されて居ることが望ましい。」[マイヤー(1914) 156 頁]』とする。その理由は、その分析利用面での自由度の大きさにある。なぜなら、「統計観察を施す社会研究材料の場所的位相を省察し、具体的な社会関係を明確に認識するに最も役立つが如く何時なりと結果の地域別を編み直しうる様に予めしておくこと」[マイヤー(1914) 156 頁]』が何よりも重要であり、「社会集団をかくの如く小地域別に限定しておくならば、ひとは随時その欲するところにしたがひ現存統計資料を観察地域の地理的事情と結びつけることができる」[マイヤー(1914) 156 頁]』からである。

結果資料(観察資料)を適切に均質かつ安定的でしかも地域間で有意な差異性を持つ小地域に区分することができれば、その後の分析過程で適用される統計地理的方法による社会集団における合法則性の検出を行うことが可能となるというのが社会科学研究者としてのマイヤーの立論である。その一方でマイヤーは指導的統計家として、地域の地理的区分がいわば標識毎に異なりうることから、多様な地域区分に適用可能な小行政地区による区分を最小地域区分として提案している。

2. 時空間個体データベースの変数要素から見た単位、時間、場所

(1) 時空間個体データベースの変数要素から見た単位、時間、場所

前節では、統計における単位、標識、時間、そして場所規定について、蜷川と彼の理論形成に少なからず影響を与えたと考えられるマイヤーとジー・ジェックの所説とあわせて検討してきた。

マイヤーとジー・ジェックはいわゆる指導的統計家として、現実の統計調査の企画、遂行の当事者としての立場を踏まえ、マイヤーは社会集団としての現実存在の中に貫徹する合法則的規則性を追求し、またジー・ジェックは比較可能な統計数値の獲得のための統計調査の一般的定式化を試みている。一方、蜷川は、政府統計資料の批判的利用の際の規準として、これらの諸要素を導きの糸としつつ利用論の観点からしかるべき統計作成過程の理論化を行った。このように、それぞれの基本的問題意識は異なるものの、彼らにいずれも共通するのは、これらの諸要素が、統計調査過程あるいは調査結果として得られた調査資料の集計処理も含めた統計作成過程において、反映すべき対象を適切に統計数値化し、比較分析に耐えうる統計資料として調査結果を提供するた

めに満足すべき要件としてそれらを論じている点である。

本稿の冒頭にも述べたように、筆者の問題視角は彼らとは異なる。すなわち、統計調査さらには行政事務の遂行の過程でその副産物として収集、作成される行政情報も含め、調査個票あるいは個々の行政記録として収集された個体情報をどのようなデータ形式で維持、保存するのが、最も効率的でしかもそれらが潜在的に有している情報価値の新たな発掘を含め、その有効性を最大限に引き出しうるかというのが、筆者のここでの主たる関心事項である。

現時点では、統計調査票情報のアーカイビング形式としては、個体識別情報を潜在的なリンクキーとするリレーショナルデータベースが世界の標準的仕様であるように思われる。そのような中で、GIS(地理情報システム)の広範は普及を受けて、時空間データベースという情報の維持、管理方式が、近年、注目されつつある。

ところで筆者は、個々数年来、統計調査票情報の情報特性に関心を持ち、それに関する検討を試みてきた[森(2009)]。そこでは、調査票情報という個体レコードが持つ情報構造を、data carrier と data body 部分から構成されるものとして定式化することによって、個体識別情報あるいは位置コードが data carrier として種々の調査結果情報という data body 部分を担っている点を明らかにした。

個票方式による統計調査の場合、多次元の変数(標識)から構成される data body 部が個体と対応づけられているだけでなく、個体識別情報は、調査票情報の横断面(水平的)あるいは縦断面(垂直的)リンクによるデータ統合(data integration)のキー情報として機能しうる。筆者が、統計に固有のアーカイビング形式として、時空間に個体に関する情報要素を加えることで、時空間個体データベースの可能性についての検討を試みる基本的動機も、まさにこの点にある。

そのための第一次的接近として、本稿では時空間個体データベースのデータ形式、特にアーカイビングされるべき統計値が持つべき基本的要素についての方法論的検討を行った。アーカイビングされる個々のデータ本体(個体レコードにおける data body を構成する個々の変数値)は、蜷川における大量の四要素では標識に相当し、それらは個体、その存在の時、存在の場所情報を持つ。

ここで再び時空間個体データベースにおける変数の形式に立ち戻ろう。

$$V_{k_{its}}$$

ちなみに、変数番号 k は標識、個体コード i は個体、時点コード t は存在の時、そして地点コード s は存在の場所にそれぞれ対応する。データベースに格納される変数値すなわち標識値のあり方を制約する、単位、存在の時、存在の場所という三つの要素について、以下に、筆者なりの視点から論評を加えてみたい。

(2) 単位、時間、場所について

1) 単位

蜷川の大量概念について内海庫一郎は、日銀券の発行高あるいは国富を引き合いに出して、それが必ずしも個体であることを排除するものではなく、大量を「その存在が社会的に規定された集団」としてではなく、「社会的存在が、その一面において集団なる性質を持つ」[内海(1962) 117 頁]ものとして規定し直すべきであると批判している。このような内海の問題提起を受けて、大量が集団か否かということについて、蜷川統計学を継承するいわゆる方法論学派の中で、統計対象あるい

は集団論をめぐって論争が展開されることになる。しかし、そこで展開された論争それ自体は本稿での直接的関心事項ではない。むしろそれに付随する形で論じられた単位の性格の方が興味深い。なぜなら、諸家の主張から、時空間個体データベースにおける個体の意味についての有効な示唆が得られるからである。

前節で紹介したように、蜷川の主要な関心事はあくまでも「大量」の性格規定であり、単位についてもその構成要素として、単位には量的差異をもたぬものと大きさを異にする単位がある、として、「数ふべき大量(例:人口)と測るべき大量(例:生産米、国富)」という大量の性格の違いをもたらすものという視角から専らそれに注目している。蜷川が Erhebungseinheit という場合の Einheit について、彼はマイヤーと同様、事実上、標識値が入る変数という意味に理解している。なぜなら、「例えば、国富に就いて、国家或は国民に属する各個の財産(所有する価値体現物の一体)が単位で・・・而して大量の集団としての大きさを大量の大きさといふならば、大量の大きさは単位の総数或は総量である。」「[蜷川(1934) 35 頁]として、単位の総和が大量に等しいとしているからである。「大量を構成する単位の数或は量は、大量の大きさを示すものに他ならない」[蜷川(1933) 98 頁]。すなわち、「大量の単位が、単位としての大きさを同一にするものであれば、それは量的差異をもたぬから、之を数へれば足りるが、若し単位が大きさを異にするものであれば、各個単位の大きさの総量が大量の大きさである」[蜷川(1934) 35 頁]。ここで蜷川が単位の量と言う場合、そこには、標識値の担い手という意識はない。

また、被調査者と単位の関係について蜷川は、つぎのように考えている。つまり、「被調査者は単位と同一である場合があり、また然らざる場合のあることは当然である。・・・被調査者は観察の対象たる大量の直接の関係者である。即ち、単位が人間たる場合には単位たる者或は其の関係者、単位が物である場合には、その物に対して一定の関係にある者、例えば所有者、管理者、売買取引の関係者、生産者、消費者の如きである」[蜷川(1933) 163 頁]と。ここでも単位は、あくまで人間あるいは物、すなわち特定の標識について定義された「数ふべき大量、測るべき大量」を構成する要素として捉えられており、被調査者は、それに関する報告担当者「一般的には単位の存在或は生起の直接の関係者」[蜷川(1933) 163 頁]なのである。

蜷川は、集団を構成する単位について、「相互に区別さるべきものでしかも特定の標識において同質性があればよい」[蜷川(1931) 132-3 頁]としている。これに対して内海は、「社会的存在とその数量的規定があれば事足」[内海(1962) 116 頁]りるとして蜷川に一方で賛意を示しつつも、日銀の卸売物価指数を取り上げ、「それはたしかに社会的量ではあるが、単位をもった大量とはいえない」として統計対象が集団でなければならない必然性はないと蜷川を批判し、最終的には「蜷川氏の所説から単位の概念をとりのぞいても支障はない」[内海(1962) 116 頁]と断言する。

田中章義は、内海が「統計調査は集団調査である。それは多数の個人の認識結果の総括、総合である」[内海(1962) 116 頁]というとき、統計単位とは、統計調査の際に調査票を配る単位として認識されているという[田中(1965) 297 頁]。内海にとっての単位とは、「各人の知識と記憶が相互に重複せず、統一をもって」さえいけば、国家、会社、家計といった「社会的個体の量的規定であれば足りる」[内海(1962) 116 頁]のである。

木村太郎は、大量観察という統計方法が対象とする社会的集団とは観察単位に他ならないとしてつぎのように指摘している。すなわち、「統計方法が対象とする社会的集団は、観察対象である観察単位集団であって、それ以外の社会的集団一般ではない。・・・統計方法が、観察単位集団を対象とし、また媒介として統計を生産するのは、かかる観察単位が社会的な集団として、しかも観察

し得る社会的形態において存在することを前提とするものである。・・・社会的な観察単位は、それ自体人間の社会的実践の単位」であり、労働者数や生産高を例に、それらが「観察単位としての工場の属性の総和として生産された統計である」と〔木村(1964) 5 頁〕。

このように木村は、単位を統計として編成される調査標識の原計数としてでもなく、また調査に対する報告担当者としてでもなく、個々の標識が調査事項として把握される個人、世帯、あるいは事業所や企業といった個体として捉えている。なお、蜷川は、『統計学概論』に先立って著した『基本問題』の中で、「(大量は—引用者) 集団なるが故に、個別的存在を其の組成因子とする。之を名づけて単位(Einheit、Erhebungseinheit)といふ」〔蜷川(1933) 98 頁〕として、単位として、Einheit とともに Erhebungseinheit という用語を併記している。しかしここでの Erhebungseinheit は、個体レコードを構成する原単位としての家計ではなく、例えばその一つの標識としての世帯所得が調査単位として意識されている。

本稿で紹介した論者の所説に共通して認められるのは、それらがいずれも統計資料をあくまでも単独の調査から得られる個々の集計表として捉えている点である。そこでは、目的とする統計資料を獲得するために対象をどう捉え、それをいかに調査標識にまで具体化するかという文脈の中で、調査過程の出発概念として単位が位置づけられている。その結果、個々の調査票情報そのものは、予め策定された集計計画に従って所定の集計表を編成するための素材としての意味しか与えられていない。伝統的な統計調査においては、集計計画の確定が調査票の設計に先立つ作業工程として存在し、統計の作成過程は、所定の一連の集計結果表の作成を目標として粛々と遂行される。そこでは、統計の二次的利用も、ジージェックが群として規定した要素に関わるグループ分け論として、あくまでも既存の集計表が許容する範囲でのみ許されてきた。

一方、時空間個体データベースの構築という視角から調査結果を捉え直した場合、調査資料として存在するのは一連の集計結果表ではなく、一群の個別調査票情報である。調査資料が仮に個体情報として格納、維持されている場合、これまでのような、所定の集計計画の達成という統計の一次利用と調査資料の二次的利用の間の絶対的な線引きは事実上消滅する。

このような集計表形式によらない個体ベースでの調査票情報のアーカイビングと結果原表に象徴される伝統的な保管方式との決定的な相違点は、ある種の統計情報については新規に統計調査を実施することなしに、アーカイビングされている既存の情報に基づいて創出できる点にある。このような統計展開の新たな局面においては、異種の情報源情報の個体ベースでの結合さらには同一個体についての縦断面結合中心的役割を果たすことになる。

調査過程を持たない行政情報のより広範な統計への利用も含め、既存の個体情報(調査票情報、個別行政情報)を連結、統合(data integration)することは利用可能な情報の外延的拡張だけにとどまらない。なぜなら、これによる既存の個体レコードが持つ変数次元の拡張は、現実についてのより包括的な分析資料として、その統計的認識の質的向上に寄与しうると期待されるからである。

このような問題関心から調査単位概念を捉え直した場合、単位とは、何よりもまず個々の個体に関する様々な変数値(調査標識値)すなわち data body の主体的担い手(data carrier)でなければならない。主として統計の発展段階に起因する時代的制約から、各論者に個体情報が持つ固有の情報価値としての意味、さらには調査単位に変数値の担い手(data carrier)としての役割を求めることはできない。しかし、あえて言えば、木村における単位＝観察単位が、本稿の問題意識から捉えた単位概念に最も近いように思われる。

2) 時間

ジーゼックは、統計数字の獲得過程に照らして、調査単位、調査標識、群そして統計的表記を社会構成体の量的規定を与えるにあたっての基本的要素としている。しかし彼の場合、単位、さらにはその具体的規定である標識そのものが存在するところの時間に関する規定は、統計を規定する基本的要素には加えられていない。[ジーゼック(1922)]は、統計の比較可能性という視点から統計の作成過程を上記の四つの基本要素との関連で論じたものであるが、そこでも時の規定については、特別な指摘はない。

有田正三も適切に指摘しているように、この点については、ジーゼックが統計数字を時空を超えた存在として捉えていたことを必ずしも意味しない[有田(1963) 185 頁]。むしろ逆で、空間的規定とともに、時間的同時性が統計比較の必須の要件であるということの自明さのために、彼はそれらを基本的要素の中にあえて加えなかったものと理解すべきであろう。

前節でも紹介したように、存在の時について、とりわけ動態現象の統計的把握との関連で最も立ち入った指摘を行っているのはマイヤーである。

動態集団の統計的把握ならびにその表章についてマイヤーは、場所規定に較べれば期間の設定は比較的容易であるとしながらも、時を期間として区切るに際しての三つの視点を提示している。統計の比較可能性、季節変動の検出、それに結果の安定性、がそれである。

まず、統計の比較可能性について彼は、「分割される各項が概して時の間隔につき相互に相等しいため、これを比較目的に利用して大いに効果があるといふ長所を持つ」[マイヤー(1914) 147 頁])としている。

次に、季節性の検出に関してマイヤーは、「社会現象の時間的経過を大期間だけにしたがって粗大に査定する場合には、時間的に一部著しく相違する要素を時間的に同価値であるかの如く取扱ふのであるから、それだけ社会集団の事実上の形相は歪められる」[マイヤー(1914) 151 頁])を立論の根拠として、「変動の特殊な年周期を示す幾多の社会集団に対しては特別な年区画を規定することが望ましい」[マイヤー(1914) 150 頁])として、期間区分の設定を行う場合には変動を析出しようような区分を行うべきであるとしている。マイヤーは、このような規則性の検出はより短期間の状況にも妥当するとしており、「1 日の時間上の配分の抽象的形相を認識しようとする特別目的がある場合には時刻別特定もまた必要」[マイヤー(1914) 149 頁])であると指摘している。

第三の視点は、公表に際しての結果の安定性に関係するものである。マイヤーは、期間の区切りについて、「時の関数である集団現象においては、それをより長い期間へ配分することが綿密であればある程、現象発現の時間上の繁度についての事実上の時間的形相への直接的な洞察はますます良好となる」[マイヤー(1914) 150 頁])ことから、「統計技術が発達している場合にあっては、時間的差異をほとんど平準化せずに最短期時別を保持するなら余りにも煩雑に分ち過ぎた数字表ができるといふ簡単な理由から、現象時点についての正確な観察があるにもかかわらず、実際上はこれに前述の如き平準化的総括を施す場合がある」[マイヤー(1914) 149 頁])と述べ、情報処理上の煩雑さを回避する上でもまた結果資料の「平準化的総括」によってより安定的な表章ができるように期間を設定するのが望ましいとしている。

このように、マイヤーにおける時間の規定、特に期間の設定については、一方で季節変動のような時間的規則性の検出を可能にするのに十分詳細な、しかし同時に他方で、区分された個々の期間内では十分に安定的な生起ケースが確保できるような区分方法が提案されている点が特徴的である。

これを多様な分析的利用の可能性という観点から格納されるべき変数値情報の時間規定を捉え直した場合、それについては、当然のことながら、動態集団事象の生起に関する記録がもともと保有する情報を減じることなく保管するのが適当であると考えられる。すなわち、日レベルで把握されるものについては生起の日が、また時間(時・分)についての詳細な記録を持つ情報については、生起の時点情報もそのまま維持した形で記録されるべきであろう。なぜなら、詳細な時間情報をより統合度の高い時間区分へと統合することは可能であるが、その逆は不可能であるからである。また詳細な時間区分を持つ動態情報は、日変化、週間変化といったより短周期の変動の分析資料としても有効である。

3) 場所

単位存在の場所的規定については、ジーゼックと蜷川においては、ほとんど特筆すべき記述はない。それは、彼らにとって場所情報が統計を規定する要素として無価値だからではなく、時の規定と同様、その重要性はあまりにも自明であると思われるからである。

統計作成の一般的行程の定式化が主要な関心事であったジーゼックにとっては、調査結果の地域表章は、行政的区画という通例を示せばそれで良かったのである。彼の理解によれば、作成される統計は、その場所的区画を異にする場合には比較可能性を失うことになる。空間を超越した統計などそもそもありえないのである。従って地域表章については、通例に従い、行政的区画による結果表章が、統計の横断的あるいは時系列的比較可能性の面からも最も適わしいことになる。ジーゼックの場合、存在の時とともに場所の規定については、社会的構成体の統計的把握にとっての基本的要素とした単位、標識、群、統計表示とは明らかに次元を異にする要素であったものと考えられる。

一方、蜷川は、大量の四要素の一つとして場所を掲げている。しかしその根拠は、単に大量が社会的存在であるからである。

蜷川は大量の空間的存在について、行政的区画と必ずしも整合的でない場合が存在することを、生産米の大量としての地域的限定を例にあげて示している[蜷川(1933) 160 頁]だけで、存在の場所的規定が持つ意味については、それ以上の踏み込んだ指摘は行っていない。

前節で詳論したように、存在の場所規定に関するマイヤーの指摘は、地理といった統計の隣接領域との学問的連携も含め、極めて多岐にわたっている。その中には、統計に関するつぎのようないくつかのマイヤー固有の発想を垣間見ることができる。

第 1 に、時の規定における期間に関する主張にも同様に認められる発想であるが、地域の区分にあたってマイヤーは、地理学の成果などを踏まえ、地域内等質性、地域間異質性を充足するように類別すべきとしている。これは、今日的表現をすれば、地域区分について、域内分散を最小化し域間分散を最大にするような区分とすることと同義である。第 2 に、仮に調査票情報が詳細な場所情報を持つ場合にも、結果の安定性の見地から最下位地域区分は決定されなければならないとしている。そして第 3 に、地域は、既存の集計結果に統計地理的分析を加えることによって新たな統計的知見が獲得できるように区分されなければならない。

ここで、最後の2つの論点について若干付言しておこう。

まず第 2 の論点についてマイヤーは、「この区画の大きさは、信用のおける比率計算を行ひうるだけの十分大きな基本数値を保持しておきたいといふ願望のために、制約されざるを得ない」[マイヤー(1914) 340 頁]として、場所についての最小区画は結果の安定性の視点から自ずと決定されるとしている。

また第3の論点に関してマイヤーは、例えば宗派と自殺の関係についての直接的集計結果表がない場合にも、「諸宗派と人口一般との関係が示されている場合、各種地域、すなはち、同一宗派だけの諸地域あるひは種々の割合で宗派の混合している諸地域につき、無差別の一般的な自殺繁度を宗派別に研究するならば、自殺繁度と宗派との間に存する推測上の関係についての近似的な観念を得ることができる」[マイヤー(1914) 157 頁])としている。このような宗派と自殺の関連の分析は、今日的に言えば、一次統計の二次的利用にあたり、既存集計表にはない結果表の作成は、オンデマンド集計あるいはマイクロデータによる独自集計として広く行われている。しかし、当時、調査票情報を個票ベースでアーカイビングし、事後的な二次的利用にも対応するという統計のあり方は、当然のことながら指導的統計家としてのマイヤーにおいても、全く意識の外にあった。しかしながら、マイヤーの卓越は、社会集団の存在の場所区分と関連づけて、すでに統計の二次的利用への対応も想定していた点である。

このようにマイヤーは、設定の仕方によっては、二次的利用も含め多様な分析的潜在性を持つ地理的区分を想定している。しかしそれは、その内容からして決して汎用的区分とはなりえない。なぜなら、マイヤーには社会集団を特徴づける諸要素として標識概念こそないものの、いわば標識毎にこの地理的区分は異なりうるからである。そこである程度の結果の安定性を持つと同時に多様な地域区画編成にも対応できるものとして、彼は県や地方裁判所管区といった「小行政区画」による調査結果の地域表章を提唱している。

そこで、このようなマイヤーからの知見を参考にして、時空間個体データベースにおける空間情報のあり方について考えてみよう。

わが国の統計を例にとれば、現行の調査票情報が持つ場所規定としては、一般に都道府県コード、市区町村コード、調査区コードが用いられてきた。この他にも国勢調査など一部の調査では、町丁字コード、あるいはメッシュコードなども場所に関する情報を提供する。

ところで、場所情報については、特に市区町村コードの場合、合併などに伴いしばしば行政区界が変更される。このような区界の変更は、地域表章結果の時系列的比較可能性を損なうことになる。なお、市区町村といった行政区画が合併などの影響を受けやすいのに対し、町丁字の境界は比較的安定的であるとされており、このことが国勢調査への町丁字による結果表章の導入契機のひとつとなっている。調査の実施にあたっての最小の地域区分単位である調査区についても、建物や道路等の構築物の改廃などにより現状に大幅な変更が生じた場合、調査区は再編され、それに伴い区界も変更される。

ところで、わが国では、総務省統計局によって、平成 2(1990)年国勢調査実施の際に、町丁字よりさらに下位の地域区分単位として、おおむね調査区を二分割した「基本単位区」が導入された。これは、国勢調査の調査区設定の基礎単位地域とするとともに、国勢調査結果の集計上の恒久的かつ最小の集計単位として利用するために導入されたもので、防災を初めとする様々な小地域統計ニーズへの対応がその導入の背景となっている。

基本単位区は「固定的な境域として恒久化されて意味を持つ」[統計局(2007) 5 頁]ことから、その区画は、住居表示に関する法律(昭和 37 年法律第 119 号)に基づく街区又は街区に準じた境域として設定され、街区方式による住居表示を実施している地域については原則、街区を、それ以外の地域については、街区に準じた小区画が基本単位区とされている。これについては、市町村境界の変更、街区方式による住居表示の実施、町・字界の変更等による市町村内の住居区画の変更、それに区画整理、地域再開発、水面埋立て、道路・河川等の新設・改修、災害等による地

形・地物の著しい変化がない限り将来においても変更しないものとされている〔統計局(2007) 5頁〕。

このように、基本単位区は、調査結果についての時系列的比較可能性の要件を前面に出した利用視点からの地域区画概念である点で、専ら実査のための調査技術上の区域概念である調査区とは明らかに一線を画している。しかし、他方で調査個票との関係で言えば、調査区よりもさらに小地域化されたものであるとはいえ、一個の基本単位区コードは依然として複数の調査単位に関する調査票情報と対応しており、両者の間には1対nの関係が維持されている。

周知のように、行政区の再編に左右されず時系列的な比較が可能なより頑健な性格を持つ地域表章形態として、地域メッシュ統計がある。どのような論拠からその実践的有効性の有無を評価したのかは定かではないが、「純幾何学的に規定された斉一的な面積単位」についてマイヤーは、「これを社会集団の場所区画の第一標準となすことは不適當である。」〔マイヤー(1914) 155頁〕として、)この種の表章形態に対して否定的である。

メッシュ統計については、現実にもいくつかの問題を抱えている。まずその作成面では、方形として与えられるメッシュ(グリッド)の境界が行政区だけでなく調査区さらには基本単位区とも整合的でないことから、調査区等が境界をまたいで位置する場合、その中に存在する世帯や人口についての同定による各メッシュへの配分作業が必要となる。また、メッシュをプラットフォームとしてその上にオーバーレイされる各種情報の中には行政区レベルでしか利用できないものも多く、防災マップ等のメッシュ統計の作成に際しては、多くの事務的作業を必要とする。一方メッシュ統計の利用面では、政策の多くが行政区単位で策定、遂行されることから、行政的利用には未だ十分な広がりが見られない。また、民間の市場分析において多用されているバッファリング分析への利用面でも、その利用は限定的である。

調査の実施の際に実査のための調査区が設定されるが、それらは、あくまでも重複把握と把握漏れを回避するために調査対象地域全体を排反的に区分した実査業務遂行上の制度的装置に他ならない。個票方式による調査の場合、調査区は集計上の最小地域単位である必要は必ずしもない。なぜなら、実査は調査区毎に遂行されるものの、個々の調査票は、基本的に個人、世帯、事業所、企業といった単位毎に作成、報告されるからである。

わが国に限らず、これまでの統計では、一般に調査結果の地域表章は調査区という最小地域単位からの積み上げによって行われてきた。そこでは、調査票情報が潜在的に保有する調査単位の個々の位置情報が調査区コードによって代表され、集計結果に反映されてきた。

もし、個々の調査単位が持つ位置情報を場所情報として取得し、保管することができれば、上述したような統計の地域表章にかかわる一連の問題は氷解する。それは、行政区界の如何なる変更にも容易に対処でき、メッシュ統計の作成の際に不可避とされていた同定に関わる作業も完全に自動化することができる。さらに、本稿で取り上げた各論者の時代には統計の利用形態として想定もされていなかった様々なバッファリング的利用にも対応可能である。なお、このような調査単位の場所に関する情報の喪失の統計作成、利用面での帰結、さらにはその解決策については、すでに〔Mori(2011b)〕で論じた通りである。調査単位の場所情報を調査区という実査のための制約から解放することによって、調査票情報が本来潜在的に保有していた情報価値を飛躍的に拡張させることができるように思われる。

統計の二次利用の視点も含め、マイヤーの場所規定には多くの示唆的内容が含まれている。しかしその一方でマイヤーの場所規定は、集計による安定的結果を理論的前提としたこれまでの統

計学が共通に持つ発想の枠内に留まっている。この限界からさらなる一步を踏み出すためには、何よりもまず統計を集計量という軛から解放することが必要であった。

3. 時空間個体データベースの情報形式について

(1) 時空間個体データベースにおける単位、時間、場所規定

前節では、その後の統計における個体情報の利活用さらには GPS (Global Positioning System) による位置情報取得をはじめとする情報技術の展開を踏まえて、時空間個体データベースに統計調査あるいは行政業務の遂行過程で得られた個体に関する各変数値 v_k が関係づけられるべき単位情報 i 、その存在の時情報 t 、そして場所情報 s のあり方について考察してきた。行論を進めるにあたり、そこでの論点を要約的に総括しておこう。

1) 単位

単位とは、もちろん人数や重量、価額といった計量単位ではなく、また多くの理論家が主張するように、その集計が直ちに集団に関する統計情報を与える変数値としての原単位情報あるいは統計調査における報告者でもない。それは、木村が「観察単位」として表現した個人や世帯さらには事業所や企業といった調査客体、すなわちひとつのまとまりを持って統計的に把握されるべき個体 (individual entity) として捉えるべきであろう。なぜなら、そうすることによって初めて、単位をデータベースの中に格納される個々の変数値から成る data body という一群の情報の最も重要な担い手 (data carrier) として位置づけることができるからである。

調査票に係る単位情報としては、世帯調査の場合には世帯符合 (世帯コード) が、また事業所や企業を対象とした調査ではそれぞれ事業所一連番号 (事業所コード) や企業コードが、個体識別コードとして転写される。わが国においては、これらの個体識別コードは、これまで一般に同一系列の調査においては整合的に設定されてきた。しかし、典型的な分散型統計制度の下で、個々の調査をそれぞれ独立した単体 (stand alone) 的性格のものとして捉えるという伝統的な統計観が根強く、調査単位に関する府省横断的な統一的コード設定は未だに実現していない。

単独の統計調査から得られる調査結果の場合、仮に集計表ではなく調査票情報を転写した個体レコードをその構成要素として持つデータセットを想定したとしても、単位が持つデータの担い手 (data carrier) としての意義はあくまでも潜在的である。その価値は、同一個体に関する他の源泉から得られる情報とのデータ統合 (data integration) によって初めて顕在化する。すなわち、同一の単位に関して同時並行的に実施あるいは収集される他の源泉からのデータとリンケージすることにより、新たな調査を実施することなく変数次元の拡張を実現することができる。単位をキー情報とするレコード・リンケージは、このような横断面での情報価値の拡張だけでなく、縦断面でもパネルデータの編成という形で既存の情報に新たな付加価値を与えることができる。

2) 時間

時間については、当然、静態調査と動態調査とでその取り扱いが異なる。

まず、静態調査については、調査時点は調査企画時点においてすでに確定している。このため、年次以上の調査周期を持つ調査では年を、また月次や四半期を周期とする調査の場合は、年ならびに月、四半期といった情報から調査時点を把握することができる。月次調査の中でも例えば労働力調査のように、原則として月末の 1 週間についてのアクチュアルな就業状況についての回答

を求める調査の場合、年・月情報が得られれば、具体的に調査の対象期日についての確定的情報を入手することができる。

一方、動態現象の中には、裁判や結婚、離婚事例のように日単位で届出や処理が記録されるものもあるが、出生や死亡あるいは交通事故や違反事件の場合には、把握の対象となる事象の生起は瞬時的である。後者の場合、例えば、厚生労働省人口動態統計においては出生、死亡の原情報となる出生届(出生証明書)、死亡届(死亡診断書)について、年月日だけでなく発生の時・分までの報告が義務づけられている。その結果、それらに基づいて統計目的で作成される人口動態出生票、死亡票にも発生の時点(分)まで転記され、情報そのものとしては、可能性としてありうる日変動の解析にも耐えるものとなっている。

現在、人口動態統計として公刊されている年報告書(上・中・下巻)に収録されている時間の要素を持つ表章結果表の大半は年計で、一部の表には月別の表章もある。また報告書には報告書非掲載の閲覧用結果表(保管統計表)の一覧も示されており、特にその中の「出生」第7表には、出生年月日時・出生の場所別出生数として、時間レベルでの結果表の表章が行われている。

時空間個体データベースに格納されるべき変数値の時間要素 t については、静態統計値の場合は、年あるいは年月との関連づけが必要である。一方、動態統計値については、現に上記のような事例もあることから、原情報が保有している情報量をできる限り喪失することなく、調査票情報をより多面的に総括することでそれが持つ潜在的活用可能性を發揮できるために、時点情報については原情報が持つ限りの詳細な情報が転写、格納されることが望ましい。

ところで、時間規定については、従来、静態統計の把握時点あるいは特定の歴史上の期間に係る動態現象の生起を支配する情報として統計においては主として位置づけられ、調査結果の表章もそのような形で行われてきた。注目すべきは、時間という要素を標識に相当する変数情報、さらには単位あるいは場所に関する情報と連動利用することによって、例えば、複数の個体の同時行動あるいは同行動といった個体と個体間の関係を析出するリンクキー情報として利用できる可能性を持っている点である。このような時間を一種の data carrier として捉える視点からの研究としては〔栗原(2011)〕があるものの、ようやくその存在が認識されただけで、多くが未開拓の領域として残されている。今後の理論的、実証的研究の更なる蓄積に注視したい。

3) 場所

場所情報については、わが国の場合、これまで都道府県コード、市区町村コード、さらには調査区コード等が調査票情報から転写、保管されてきている。最小の地域区分単位とされる国勢調査の基本単位区でさえも、一般に20~30世帯を域内に含む地域規模である。その結果、本来、それらの世帯の存在の場である個々の住戸が持っている位置情報は、基本単位区コードによってひとつの集団として一括表示されている。

調査によっては、調査票情報の中に住所や所在地として位置情報につながる情報を持つものもある。センシング技術の最新の展開により、利用可能な精度を持つ位置座標情報の取得も現実のものとなりつつある。近代統計の確立以来、統計調査の中心的方法とされてきた個票方式の調査意義をこのような状況を踏まえて捉え直した場合、位置情報の調査区あるいは基本単位区コードとしての統合表示は、調査票が本来収集し得たはずの単位の所在に関わる個体位置情報の明瞭な情報損失が発生していることを意味する。

結果的にこのような情報損失をもたらす取り扱いの背景には、つぎのような三つの事情が関係しているように思われる。

その 1 は、地域情報に関わる統計の原情報の収集過程と得られた原情報からの統計の編成過程という相互に全く性格を異にする二つの過程の相違についての無理解に起因する。すでに論じたように、調査区という地域の区分は、重複把握と把握漏れを回避しより正確な観察単位を捕捉する目的で設定された調査実施の制度的装置として導入されたものである。当時の情報処理技術の未発達という状況もあり、集計過程においても調査区は、位置情報の潜在的価値に思いをめぐらすことなく、そのまま地域の積上げ集計の基礎的単位として用いられてきた。

その 2 は、すでに前節において指摘したように、統計を個体としてではなく集団として、個体に関わる調査票情報としてではなく結果表示として集計表として捉える統計観に起因するものである。このような伝統的統計観に依拠する場合、個体は個体そのものとして意味づけられるのではなくあくまでも集合体の個々の構成要素として捉えられる。そこには、個体に関する情報には常に不安定要素が付きまとい、集団として認識することでそれらを相殺消去することができ、そうすることによって得られた相対的に安定的な結果に基づいて結果表の積み上げを行うというのがそれである。

そして第 3 は、上記の各論点とも部分的に重なることであるが、集計結果に較べて個体情報の場合には識別されやすく、統計を個体として捉えることで統計に係る秘密保護が侵害されやすくなるとの認識がそこにはある。なぜなら、調査客体に関する所在地情報は、直ちに当該個体の識別につながりうるからである。ここでも、原情報の収集とそれによって得られた情報の解析処理とが、同一次元で捉えられている。統計データのアーカイブの観点からこの点を捉え直せば、要するにこの問題は、データの格納、維持管理とその利活用システムとを異次元の事柄であるとの認識に立ちそのシステム設計あたるということである。

個体が保有し調査票情報が潜在的に持ちうる位置情報を調査区等の位置コード情報として希薄化することなくその情報価値の最大限活用を目指すものとして、近年、世界各国の統計作成機関で GPS 座標情報の利用可能性が注目されつつある。これには、衛星情報だけでなく様々な既存情報の統合利用による測位精度の飛躍的向上と端末の急速な低廉化がその追い風となっている。GPS が与える詳細な位置情報の利活用を契機とする個体情報の新たな情報価値の拡張可能性については、位置の座標情報を備えた個体情報に基づく様々なデータセットの編成可能性も含めて[森(2011a)]が多角的に論じている。

これまで調査票情報については、観察単位としての調査客体の属性、行為、活動さらにはそれらの結果に関する記録とみなされ、調査によって収集された個体情報も、主としてそのような方向で処理、活用されてきた。この調査票情報を視点を変えて捉え直した場合、世帯や企業それに事業所といった単位に関する調査票情報は、ある特定の地点において存在し活動している単位についての記録として見ることもできる。調査票情報に対する位置情報視点からのアプローチは、統計情報の担い手(data carrier)として data body 部分を構成する諸変数値を相互にリンクするキー変数として捉える方法論的立場に立つ。その意味では、この視角は、すでに上述した時間情報を統計情報の担い手とみなして同時行動の分析視角とその基本的立場を共有している。

このように、調査客体が把握される地点情報そのものに焦点をあてた場合、それから編成されるデータセットは、既存の統計とは全く異質の情報を提供するものと考えられる。この点について筆者は、すでに[Mori(2011c)]において、調査票情報による地域の潜在力の検出等への利用可能性などを論じた。

(2) 時空間個体データベースの情報形式について

ここでは時空間個体データベースの情報形式に関して検討を試みるが、その際の参照資料として、デジタル画像情報のデータ形式について簡単に見ておこう。

1) デジタルの画像データ形式

デジタル画像情報は、(0,0)から(m-1,n-1)によって与えられるm×n 個のピクセル(picture cell)と呼ばれるラスタ型ラスター型の微小な面に対してその濃淡(tone)を与える情報によって構成される。モノクロ画像の場合は各ピクセルに該当する濃淡を示すスカラー情報によって画像がデジタル表現され、カラー画像については、原則的には赤(Red)、緑(Green)、そして青(Blue)の三原色三原色が、それぞれについての濃度情報を与える。個々のピクセルに対して、赤、緑、青の濃度を示す3次元の濃度情報を重ねることによって各ピクセルの色彩が数値化され、m×n 個のラスタ情報を用いて画像がデジタル表現される。なお、カラー画像を合成する場合には、三原色情報に透明度情報(Alpha)を加えた RGBA が用いられることもある。

図1は、カラー画像をデジタル化する際のデータの読み込み配列例を示したものである。

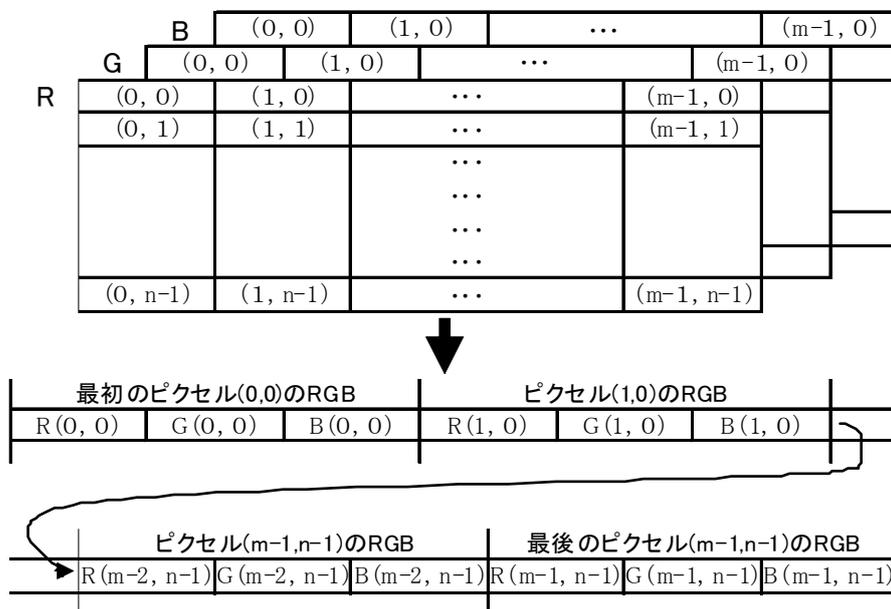


図1 デジタルカラー画像のデータ配列例

例えば、ピクセル座標 (i, j) としての画像素子情報は、 T を濃度情報 $T_{ij} = (T_r, T_g, T_b)_{ij}$ によって与えられるが、これを個々の情報が格納される形式として表現すれば、

$$T_{cij}$$

$$\text{ただし、} c = r, g, b \quad i = 0, \dots, m-1 \quad j = 0, \dots, n-1$$

となる。すなわち、色彩と位置に関する情報を持つ濃度情報 T が、1枚の画像データを構成していることになる。

2) 時空間個体データベースのデータ形式

統計データをアーカイビングする場合、格納する情報の形態としては原情報を格納するのが最も効率的である。なぜなら、各種の集計値さらには指標等の編成統計量は、目的に応じて適宜、原情報から容易に導出することが可能だからである。以下に、データ形式についてのイメージ図を例示的に掲げることとする。

ていることがそれである。それでは、単位情報だけでなく、時間や場所も含めた多面的利用に対応できるためには、データベースに格納される変数値はどのようなデータ形式を備えておくべきであろうか。

すでに見たように、画像情報の場合、各ピクセルの色は、色彩の濃淡を与える三種類(あるいは高々四種類)の濃度情報によって決定されていた。[森(2009)]においてすでに論じたように、統計の場合、このような画像情報のデータ形式とは異なり、数百以上の変数からなる極めて多次元の情報によって現実の統計把握空間の諸側面を切り取っている。個々の調査票に記載されたある一定の時間と場所において収集された単位に関する原情報を変数の実現値とみなすとき、その変数はその種類を示す添え字 k を持ち、各変数の存在が単位 i 、時間 t 、場所 s によって規定された

$$V_{k_{its}}$$

といった形で表現できると考えられる。言い換えれば、時空間個体データベースの構成要素となる変数値は、 $\mathbf{V}_k(i, t, s)$ という形で、個体、時間、そして場所といった規定情報を持つ変数のそれぞれ実現値として格納されることになる。

むすび

本稿では、統計データのより有効なアーカイビングシステムの構築という今後、政府統計において重要な意味を持つようになると思われる課題に向けての取り組みの第一歩として、時空間データベースの基本的アイデアからの知的投入を受け、またその発想の統計への適用に関しては先人たちの知的到達から多くを学びつつ、その方法論的検討を試みてきた。

もちろん、その後の世界の統計ならびに統計理論の展開は厳然として存在し、またデータベースの構造設計という今日の問題意識から彼らの到達点を評価することは、学問的にはどちらかといえば公平性を欠く行為である。従って、本稿で各論者の主張を取り上げるにあたっては、彼らの所節に対する今日的視角からの批判というよりはむしろ、筆者自らの問題意識を具体化していく上での検討の手掛りを求めるというスタンスを心がけたつもりである。幸い彼らから得られた知的刺激は、事前に期待していた以上のものであった。

本文での考察から、統計データのアーカイビングにあたっては、変数値に対して時と場所規定のみを付与してそれらを格納しただけのいわゆる時空間データベースでは不十分機能することができず、統計データに固有の規定要素である単位情報が不可欠であることが明らかにされた。現実を統計調査等によって統計的に写し取るにあたっての単位の重要性については、各論者がすでに 1 世紀近く以前に指摘している通りである。また、時さらには場所の規定についても、諸家の研究成果の中から多くの示唆を得ることができた。

本文でもすでに指摘したように、情報技術の現段階における到達点を所与としてより有効な統計データのアーカイビングという視点から各論者の所説を改めて読み直してみると、統計データの捉え方について、いくつか決定的と思われる問題点が散見される。もちろん、その大半は当時の統計の発展段階による時代的制約に起因するものであるが、そのうちのいくつかはすでに技術的には対処可能であるにもかかわらず、今日の統計実践においてもなお継承されているものである。

本稿で取り上げた各論者と筆者の間の統計のあり方をめぐる本質的な見解の違いは、統計を本来的に集団を反映した集計量と見るかあるいはその構成要素としての個体情報に固有の情報価値を認めるかという一点に集約される。この点に関して、個別情報に不可避免的に含まれる観察(観測)誤差を集団として捉えることでその中に貫く安定性あるいは規則性の析出が可能となるというのが、集団説を支持する理論的根拠となっている。これに対して本稿での筆者の主張は、集計に伴う誤差の相殺効果の存在は認めつつも、統計＝個体情報はそれをも包摂した新たな統計の展開の契機を有する点についての問題提起にある。そこでの筆者における中心的関心事となっているのは、単位だけでなく場所あるいは時間についても、アーカイブされた既存情報のデータ統合(data integration)による新たな統計情報獲得の可能性、言い換えれば、調査票情報が潜在的に保有していた情報価値の新たな開拓可能性の追求という統計の将来展開方向である。

本稿でその検討課題として掲げた時空間個体データベースのデータ形式については、時の規定について、静態データと動態データについてのより効率的な格納はいかにあるべきか、その他にも変数値の情報源情報の取り扱いや天文学的規模に上ると想定されるデータ量に対する情報技術面での対応可能性など、多岐にわたる難問が数多く残されている。これらについては、今後も引き続き関心を持って取り組むことにしたい。

〔参考文献〕

- Mayr, Georg(1914) *Statistik und Gesellschaftslehre* 大橋隆憲訳(1943)『統計学の本質と方法』
- Žižek, Franz(1922) *Fünf Hauptprobleme der statistischen Methodenlehre*, München und Leipzig, Verlag von Ducker & Humbolt「統計方法論の5つの主要問題」有田・足利・松井訳『フランクフルト学派の統計学』晃洋書房 1987 年
- Žižek, Franz(1937) *Wie statistische Zahlen entstehen. Die entscheidenden methodischen Vorgänge*. Leipzig, Hans Buske
- 蜷川虎三(1931)『統計学研究 I』岩波書店
- 蜷川虎三(1933)『統計利用に於ける基本問題』岩波書店
- 蜷川虎三(1934)『統計学概論』岩波書店
- 内海庫一郎(1962)『科学方法論の一般規定からみた社会科学方法論の基本的諸問題』
- 有田正三(1963)『社会統計学研究』ミネルヴァ書房
- 木村太郎(1964)「統計と社会集団」『統計学』(経済統計研究会)第12号
- 田中章義(1965)「統計対象に関する諸家の見解について」『東京経済大学創立 65 周年記念論文集』
- 大屋祐雪(1967)「F.チチェックの統計調査論」『九大経済学研究』(40周年記念経済学論文集)
- 総務省統計局(2007)『平成 17 年国勢調査 調査区関係資料 利用の手引き』
- 森 博美(2008)「情報資産としての統計と政府統計データアーカイブ」『統計学』(経済統計学会)第94号
- 森 博美(2009)「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
- 栗原由紀子(1011)「TUD による時空間共有行動の分析」『生活時間データによる時空間共有行動』

の統計的解析 - 社会生活基本調査マイクロデータを利用して』(学位請求論文)所収

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』法政大学経済学会 第78巻第3号

Mori,Hiromi(2011b)“Constraints in Use of the Data Due to the Insufficient Obtaining of Location Information and a Breakthrough in Statistics”,*KEIZAI-SHIRIN(The Hosei University Economic Review)*, Vol.78, No.4

Mori,Hiromi(2011c)“The GPSed Datasets and the Possibility of Exploring the Micro-based Concept of Regional Potentiality”,*KEIZAI-SHIRIN(The Hosei University Economic Review)*, Vol.79, No.1

河合基伸、高橋史忠、Phil Keys(2011)「決め手は位置情報」『日経エレクトロニクス』(日経BP社) No.1051

[謝辞]本論文は、平成23年度日本学術振興会科学研究費補助金 基盤B「政府統計データのアーカイビングシステムの構造と機能に関する国際比較研究」(課題番号 22330070)による研究成果の一部である。

オケージョナル・ペーパー(既刊一覧)

号	タイトル	刊行年月
1	EC 標準産業分類(NACE)	1990.07
2	On Numerical Calculation Programs of American-type Options Using GAUSS Codes	1998.09
3	わが国の統計体系の現状と課題(I)	2000.03
4	ICD10 における自動車事故による死亡者数の試算	2000.04
5	人口動態統計における交通事故死亡統計の特徴について	2000.05
6	Trends in U.S. Working Hours since the 1970s	2001.07
7	わが国における外国人の国籍別出生率について	2001.09
8	東京の消費構造—東京都生計分析調査	2002.10
9	Wide Variations in Statistics Data Sets on the Same Subjects—Reconsidering the Report of the Indian National Statistical Commission	2003.12
10	日中 1995 年産業別購買力平価の推計	2004.04
11	日本における「統計法」の成立	2005.06
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

オケージョナル・ペーパー No.24

2011 年 4 月 1 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