経済統計学会編

社会科学としての統計学

第４集

2006

産業統計研究社

刊行のことば

経済統計研究会(1985年に経済統計学会へと名称変更)は,「社会科学に基礎を置く統計理論の研究」を目的として1953年に創立された。1955年6月に研究会の機関誌である『統計学』が創刊され,それ以来その発行は年2回を基本として続けられてきた。その間1976年8月に創刊20周年記念号(第30号),1986年8月に創刊30周年記念号(第49・50合併号),1996年3月に創刊40周年記念号(第69・70合併号)を発行し,それぞれの期間における学会の研究経過を振り返り,会員の研究成果と統計学各分野の動向をサーベイしてきた。今回の記念号は,創刊50周年という大きな節目に当るとともに21世紀最初の記念号であるということで,この10年間の会内外の業績を踏まえ,次の半世紀を展望するという視点から,今日の激動する世界と日本の社会経済状況が社会科学としての統計学に問いかけている諸課題を明らかにし,それに答えようという大志をもって取り組んだ。

経済統計学会は全国研究大会を1957年以来毎年開催しているだけでなく,月例研究会を関東,関西,北海道,九州などで,並行して50年以上に亘って続けている稀有な学会である。1つ1つの報告に充分な時間を使いじっくり討論する月例研究会は経済統計学会の特徴のある研究を形成する重要な場であり続けている。また最近21世紀に入ってからはミクロ統計部会,労働統計部会,日中統計部会,ジェンダー統計部会,健康・生活統計部会,政府統計部会などの研究部会を設けそれぞれの分野の統計の研究を深めようとしている。我々はこれらの現時点までの成果を出来うるかぎりこの記念号に盛り込もうと努力した。

50年という期間は相当に長い期間である。経済統計研究会の創立あるいはその基礎を築くうえで大きな役割を果たされた会員のうちには,現在もなお御元気で会の行く末を心配しながら温かく見守って下さっている方々もおられるが,既に亡くなられた方も多い。時代は変化し,統計研究の内容も方法も変わっていく。しかし我々は,長年の間に先輩会員の残して下さった貴重な研究成果を徹底的に大事にしてそれを踏まえて新たな研究に立ち向かいたいと考える。そうした時にのみ着実な前進が可能になると思うからである。10年ごとに記念号を編集し,また今回は50年の節目という観点も入れて,会と会員の研究活動を総括しているのはそのためである。

経済統計学の研究は,50年を経,取り巻く社会・経済と社会科学・思想状況の変化のなか,新しく入ってこられた若い世代の研究が中心になるにつれ,変化しつつある。しかし経済統計学会の先輩の形成してきた伝統は今も生きているし今後も生き続けていくであろうと思う。その伝統の一端を述べると以下のようになる。

第1に,経済統計学会は,社会・経済統計そのものを大事にし,それを研究することが誰よりも好きな人たちの集まりであるということである。我々は社会経済統計そのものを我々の研究対象と考える。社会経済統計がどのように作成されているか,社会経済統計をどのように作成すべきであるか,社会経済統計は何をどう反映しているか,必要であるにもかかわらず作成されていない社会経済統計にどのようなものがあるか,社会経済統計はどのように利用されているか,社会経済統計どのように利用すべきであるか等々,社会・経済統計に関するあらゆる問題を統計学の課題と考え,協力して研究している。この点で

我々の研究対象は，統計数理学の研究対象と重なる部分もあるが，それとは別の非常に広く複雑な内容を擁している。

第2に我々は，経済学をはじめとする諸社会科学の成果に依拠し，またその発展に貢献するということをめざして社会・経済統計を研究している。この50年において経済学・社会科学の潮流はかなり変化した。経済統計研究会の創立に重要な役割を果たされた会員のうちには社会科学としてマルクス経済学を想定されていた方々が相当数おられた。現在，若手・中堅会員の想定する社会科学には新古典派経済学をはじめいろいろなものがあり多様性に富んでいる。もちろん若手・中堅会員の中にもマルクス経済学を受け継ぎ新しい方法を取り入れ発展させようと考えている者もいる。我々は社会・経済統計を研究するという場で多様な社会科学の相互批判と相互協力を推し進めていこうと考える。社会・時代の変化に取り残されることなく，しかし流行に流されることなく，無用な対立はさけつつも曖昧な妥協はせず，真に学問的な研鑽を積んでいきたいと思う。

第3に，我々は，国民生活の向上と社会の進歩に役立ちたいという願いを心に秘めて，統計学の研究に取り組んでいる。統計学は，生活をまもり豊かにする国民の諸分野の活動に貢献すべきであるし，その学問的質を真に高めることを通じてそれが可能になると考えている。いろいろな分野で闘っている人たちに社会・経済統計と統計方法という優秀な武器を提供できたらと願っている。

経済統計研究会を創立した先輩から我々が受け継がねばならない最も大事な点は，なにものにもとらわれない鋭い批判精神であると思う。批判精神という牙をぬいてしまうと学会は魅力の乏しいものにおちぶれてしまうであろう。

現在の日本では，強行に推し進められている大学改革とも関係して，かつてないほどに業績主義が蔓延している。しかし我々は，学会を単にスマートにそつ無く業績を作っていく場ではなく，社会・経済統計を本気で研究する，あるいは統計を道具として社会・経済を本気で研究していく場としていきたいと志している。その点で，経済統計学会は，若い会員が成長しやすいよう配慮しなければと考えるとともに，試行錯誤をふくむ長期の泥臭い研究，なかなか形になる成果は得られないが本当に重要だと信じて日夜努力している会員の研究にも温かい眼を注いでいる。

現在日本でも世界でも政府統計・社会・経済統計をどう改革していくかという論議が盛んに行われている。第2次世界大戦後に作られた統計制度をその後の大きな社会経済の変化に対応するよう改革しようという議論である。本書で扱われている研究は，このような議論に貢献できる内容を含んでいると思う。そのような議論を深めていく手がかりとしても本書が読まれることを期待したい。そのような議論と切り結ぶ中で我々もさらに研究を深めたいと考えている。

この記念号は，全ての会員と社会・経済統計に関心をもつ多くの人々の座右におかれ，長期にわたって，今後の社会・経済統計研究の出発点，展開のヒントを得るための尽きせぬ泉となることをめざして執筆された。本書は，社会統計・経済統計の研究者のみならず社会科学の諸分野の研究者，社会統計・経済統計を使用する種々の分野の方々に参考してもらえることを期待するとともに，多くの方々から批判・コメント等をいただけると幸甚である。

経済統計学会会長
泉　　弘志

目　　次

刊行のことば

第１部　社会科学としての統計学 ―― その今日的課題 ――

第１章　統計の作成・公表・利用における公共性……………………金子治平（3）
　コメント…………………………………………………………山田満（14）
第２章　地域における統計の作成と利用…………………………藤江昌嗣（17）
　コメント…………………………………………………………菊地進（28）
第３章　個票データと統計利用……………………………………坂田幸繁（31）
　コメント…………………………………………………………岩井浩（42）
第４章　民間企業におけるデータの蓄積と利用……………………池田伸（45）
　コメント…………………………………………………………佐野一雄（57）

第２部　統計作成と統計制度をめぐる新たな展開

第５章　社会・経済の変容と政府統計の変化
　Ⅰ　センサスと統計調査の変容…………………………………西村善博（63）
　コメント…………………………………………………………岩崎俊夫（74）
　Ⅱ　産業・職業分類の変容………………………………………長澤克重（78）
　コメント…………………………………………………………松川太一郎（87）
　Ⅲ　政府業務記録と統計利用……………………………………岡部純一（90）
　コメント…………………………………………………………森博美（102）
第６章　民間統計の現状と利用可能性……………………………山田茂（104）
　コメント…………………………………………………………佐藤智秋（113）
第７章　統計制度改革の国際的動向と統計品質論…………………水野谷武志（116）
　コメント…………………………………………………………小川雅弘（128）

第３部　統計利用をめぐる諸問題

＜統計解析＞
第８章　計量モデル分析……………………………………………井口泰秀（133）
第９章　データ解析法………………………………………………田浦元（144）
＜個別領域＞
第10章　人　口………………………………………………………廣嶋清志（154）
第11章　産業・企業…………………………………………………藤井輝明・御園謙吉（164）

第12章　労　働 …………………………………………………………… 小野寺　剛(174)
第13章　家　計 …………………………………………………………… 大井達雄(184)
第14章　金融・財政 ……………………………………………………… 伊藤国彦(194)
第15章　国民経済計算 …………………………………………………… 金丸哲 光藤昇(204)
第16章　産業連関 ………………………………………………………… 朝倉啓一郎(214)
第17章　環　境 …………………………………………………………… 吉田央 光藤昇(224)
第18章　食料・農業 ……………………………………………………… 香川文庸(234)

第4部　部会における研究の成果と課題

第19章　ジェンダーと統計 ……………………………………………… 杉橋やよい(247)
第20章　労働と統計 ……………………………………………………… 福島利夫 村上雅俊(258)
第21章　中国統計 ………………………………………………………… 矢野剛(268)

第5部　社会科学としての統計学 —— その伝統と継承 ——

第22章　統計史 …………………………………………………………… 上藤一郎(283)
コメント …………………………………………………………………… 長屋政勝(289)
第23章　統計学史 ………………………………………………………… 芝村良(293)
コメント …………………………………………………………………… 木村和範(303)
第24章　人口センサスの方法転換問題と統計学研究の課題 …… 濱砂敬郎(305)
第25章　実質社会科学説の「勝利」とその後 ……………………… 大西広(318)
第26章　経済統計学会の歴史の四齣 ………………………………… 伊藤陽一(323)

『統計学』バックナンバー目次については，経済統計学会ホームページをご覧下さい。

第 1 部　社会科学としての統計学

――その今日的課題――

第1章　統計の作成・公表・利用における公共性

金　子　治　平

はじめに

周知のように，ある研究分野の主題・方法等は，社会的環境の変化によって，学問的伝統を保持しつつも変化していく。たとえば経済学・歴史学でいえば，戦前の日本資本主義論争で提起された視点は，戦後しばらくは主流にあったが，現在では多様な主題を対象とするように変化している。ここ十数年，政府統計をめぐる環境は，情報通信技術の発達を背景に，国際的には統計の品質論，国内的には現状把握が不十分な領域の統計整備などを内容とする統計整備構想や統計改革論議などの動きが見られ，2005年国勢調査では統計調査環境の悪化が新聞各紙をにぎわせるなど，激変しつつある。このような変化に対応して，経済統計学の領域においても，統計の真実性を主内容とする政府統計の批判的利用から一歩進んだ研究課題が提起されつつあるように思われる。

本稿では，このような問題意識のもとに，現在「公共財」として位置づけられつつある政府統計は，非競合性・非排除性をもつ公共財としてとらえるだけでは不十分であり，政府統計が持つ特質から公共性という視点を導入しなければならないことを主張したい。

1．政府統計と公共性

2005年9月に設置された統計制度改革検討委員会の第2回委員会において，ある委員から「標語的に言えば，今までの政府統計は，『政府の政府による政府のための統計』ということで，これを変える必要があり，特に『政府のための統計』が『国民のための統計』に変わることが重要と考える。」という発言がみられた。ここで指摘されているように，これまでの日本の統計が分散型統計制度のもとで，主として個別行政のために統計が作成されてきたことはしばしば指摘されてきた。

ところで，日本の統計が「行政のための統計」であったことの責任は，統計行政機関だけにあるのではなく，国家の政治・行政のあり方自体にある。村松岐夫(2001)によれば，戦前には明治憲法体制によって官僚の下に権力が集中し「弱い議会，弱い内閣，弱い政党システム」であったし，戦後においても，占領期にはGHQの権限のもとで官僚が大きな裁量を持っていた。1955年以降には，いわゆる五五年体制のもとで次第に政権党の優位が確定していくが，政権党の優位は必ずしも強い議会を意味しなかった。行政領域の拡大とともに，行政による許認可や行政指導を手段とする経済発展は，政府の市場への介入を重視するケインズ経済学の隆盛とも相まって，いわゆる行政国家として行政の裁量を維持・増大させた。このように行政の裁量が大きい場合には，個別施策を決定するための情報＝統計は行政内部に存在していればよいのであり，国民一般はもちろん，指定統計を除けば立法機関にさえ統計を公開しなければならな

いという必然性は持たなかったのであろう。付言しておけば，行政の裁量が大きかったことは必ずしも官僚だけで裁量が決定されていたことを意味しない。重要なことは，政権党の政治家が官僚の意志決定に及ぼす影響が，国会の場ではなく非公開の下で行なわれてきたことである。

しかし，1990年代半ばにおける55年体制の崩壊によって，日本においても小さい政府を目指し，官僚の裁量による行政から脱却しつつある。その背景としては，財政赤字を抱えるアメリカやイギリスのレーガンやサッチャーによって採用された1980年代の小さな政府への志向という実体的側面や，政府の失敗を重視するスティグラーのキャプチャード理論やボーモルのコンテスタビリティ理論などの新古典派経済学の研究が主流となったという理論的側面を指摘されている。

現在の日本の小さい政府論が，サッチャー・レーガン時代の小さい政府論と大きく異なっているのは，大住(1999，2003)などで紹介された1990年代以降のニュー・パブリック・マネージメント(NPM)論の影響を受けていることである。NPM論は，業績・成果による統制，市場メカニズムの活用，統制の基準を顧客主義へ転換，統制しやすい組織への変革などを特徴とするものである。つまり，国民を政府行政サービスの顧客とみなして，限られた行政リソースのもとで顧客の満足度を高めることを目的とし，そのために，効率性を求めて行政の階層構造を簡素化したり，行政活動を市場化したりすることを求めるものである。また，行政内部においてはモニタリング(監視・監査)を重視するのも，NPMの特徴と言われている。このNPM論のもと，欧米では電力・ガス・交通などの準公共財を供給していた公企業の多くが民営化された。

小泉内閣の構造改革の一環として，既存統計の抜本的見直し等，統計制度の充実をはかるため，2004年11月に経済社会統計整備推進委員会が設置され，数回の討議を経て2005年6月に『政府統計の構造改革に向けて』という報告書が提出された。この報告書の内容は多岐にわたり，経済センサスの具体化，GDP関連統計の整備，サービス分野の統計整備，ストック統計の整備，統計の体系化，「司令塔」機能の強化，行政記録の活用，統計情報の多様かつ高度な利用，統計調査の民間開放，法制度の見直しなどについて提言が行われている。

政府の統計行政の行方を示唆するものとして，1995年『統計行政の新中・長期構想』，2003年『統計行政の新たな展開方向』から，2005年『政府統計の構造改革に向けて』(以下，単に『委員会報告』と呼ぶ)において，統計の役割(意義)がどのように変化しているかをみておこう。『統計行政の新中・長期構想』では，「行政施策の企画・立案のための基礎的情報の提供に止まらず，広く国民一般の利活用のための情報提供という面についても十分配慮していく必要がある。また，国民の負担と協力によって得られる統計は，国民の共有財産として迅速かつ継続的に提供され，広くその利活用がはかられていくことが肝要である」と書かれており，政府内での利用を主とし，国民への情報提供を従とする位置づけとなっている。このような統計利用主体を主・従として位置づけているのは，『統計行政の新たな展開方向』でも同様であった。しかし，『委員会報告』では，基礎的な政策運営や行政施策の企画・立案目的，政策効果の事前・事後評価目的などの政府内での統計利用とならんで，事業者や個人の合理的な意志決定目的，学術研究目的などの政府以外での統計利用が並列的に位置づけられるように変

化している。すなわち，事業者や個人などによるニーズへの対応が強く打ち出されている。

このように『委員会報告』においても，統計の社会的ニーズを重視する顧客主義，「司令塔」の設置によるヒエラルヒーの簡素化，統計調査の民間開放による市場メカニズムの活用など，NPM論の影響を受けている。

ところで，政府による経済活動の主要なものは，公共財の供給である。『委員会報告』でも，「政府が作成する統計は，本来幅広い目的に提供されることを意図した『公共財』としての性格を持つべきであり，政府にはそれを供給する役割がある」（経済社会統計整備推進委員会『政府統計の構造改革』p.8）と書かれている。ここで用いられている「公共財」とは，経済学で非競合性・非排除性という特質を持つ財・サービスとして定義される公共財概念とは異なり，「行政以外の主体によっても幅広く利用されることを念頭に置いて作成され，社会が必要とする基礎的な情報を提供するという性格を持っていることをシンボリック（象徴的）に表現する語」（『経済社会統計整備推進委員会報告書（案）』2005年5月30日の第7回委員会での配付資料，p.8）として，独自に定義されている。つまり，公表されることによって多様な主体が利用できること，および社会が必要とする基礎的な情報であること，の二つを要件として「公共財」としての統計が特徴づけられている。前者は一般的な公共財概念に近いものを意味しているが，後者は一般の公共財にはない性質である。

さらに，「行政以外の利用が制限されている場合のように，ここでいう『公共財』に該当しない場合もあり得るが，それらの統計の中にも，高い公共性や公益性を有するものがあることを否定する趣旨ではない」（前掲『報告書（案）』p.8）とも書かれているように，公表されていなくても，公共性や公益性を有する場合もありうると主張している。ここでいう公共性とは，「公共財」の要件のうち，社会が必要とする基礎的な情報とほぼ同等のものであろう。

一般的な公共財として公共図書館を想定し，一般的な公共財と政府統計との相違を考えてみよう。図書館においては，図書を蔵書し，それを広く社会に公開することによって「国民の教育と文化の発展に寄与することを目的」（図書館法第一条）としている。「すべての国民は，いつでもその必要とする資料を入手し利用する権利を有する」（図書館の自由に関する宣言）と言われるように，その主たる利用主体は不特定多数であるし，図書・資料の提供自体が目的となり，その図書・資料を利用主体がどのように利用し，どのような文化発展に寄与しているかは問題とされない。つまり，一般的な公共財の場合には，政府はその供給のみに関与し，利用者としての国民がどのように利用するかは問題とされない。

しかし，政府統計は広く社会に公表されているが，実際的な統計利用者は不特定多数ではなく，政府，企業，政治的活動を行う市民，研究者などに限定される。したがって，統計を利用する事業者や個人を主たる利用者と位置づけることはできない。欧米で，統計作成者による統計作成目的に対応する統計利用を第一次分析（primary analysis），それ以外の統計利用を第二次分析（secondary analysis）と呼ぶように，政府統計は第一義的には統計作成者である政府による統計利用を前提に作成される。政府による統計利用は，現状を把握し，経済政策や社会福祉政策などを立案し，さらに政策の効果を評価するなど，国民へ大きな影響を持っている。すなわち，政府統計の場合，政府は供給者であるばかりではなく，その主たる利用者であり，その利用結果が国民

に影響を及ぼすのであり，一般の公共財とは異なっている。『委員会報告』において「公共財」の要件として，社会が必要とする基礎的な情報と指摘しているのは，この意味であろう。

以上のように，政府統計を公共財と理解するだけでは不十分なのであり，公共性を持つものとして理解する必要がある。一般に公共性という語は，善あるいは良というイメージでとらえられることが多いが，本稿ではさしあたり，善・良という価値観を持たないものとして公共性を考える。ところで，統計のもつ公共性を，上記のように「広く社会一般の利害に関わる性質」と定義するならば，その性質が成立するための条件は何であろうか。

広く社会一般の利害に関わるための条件として，法による支配・権力分立・参政権などを特徴とする自由民主主義国家においては，全体主義国家や権威主義国家とは異なり，対象となることがらは，特定の集団に限定されることなく広く社会に対して，公開される必要がある。なぜならば，あることがらが広く社会一般の利害に関わっていたとしても，仮に特定の個人や集団にだけ開かれている場合には，それ以外の個人・集団は，そのことがらが正当性を持つのかどうかを判断できないし，そのことがらに対して政治的な権利を行使することもできないからである。

ここで，主題を統計に限定すると，公開とは大きく二つの内容を持っている。一つは，個人ないしは法人が意志決定を行うための情報として統計の利用は重要な意義を持つため，プライバシー権を侵害しないように秘匿措置を施しつつも，統計データそのものが公開される必要がある，ということである。もう一つは，個人が政治的な権利を行使するための情報として，立法・行政・司法機関から構成される政府において統計がどのように利用されているか公開されていなければならない，ということである。つまり，政府の立法過程や行政過程におけるどの領域においてどのように統計が利用されているか，あるいは広く社会一般に利害を及ぼすものとして統計利用がどのように行われているか，が公開される必要がある。さらに後者の意味での公開は，統計の利用が正統性原理に合致しているかを公開の場で議論し判断すること，換言すれば，特定の集団に限定されることなく広く社会における討議の場が設定され，そこで統計の利用が適切であるかが判断される必要がある，という公開性を要請する。国連統計委員会『官庁統計の原則』では，「民主的な社会情報システムにおける不可欠な要素」「公的な情報に対する国民の権利」という言葉で，これらのプロセスとしての自由民主主義を保証する公開性を表現しているといえよう。このような公開性を，山田満(2000)は「公民権としての統計学」と呼んでいる。

さらに，社会一般の利害に関わる性質は社会の歴史的条件によって左右されるため，事前には措定できない，ということが問題になる。たとえば，1960年代以前にはジェンダー問題自体が社会問題となっていなかった。しかし，「女性問題」が1975年頃から国会でもとりあげられるようになり，経済統計学会内外の活発な研究や運動の成果として，ジェンダーの領域に関する統計は整備が進み，現在では国立女性教育会館の「女性と男性に関する統計データベース」として公開されている。したがって，社会一般の利害に関わるものとしての統計の主題も，特定の集団ではなく広く社会において討議され，合意形成される必要がある。

すなわち，自由民主主義国家においては，統計が広く社会一般の利害に関わる性質であるための条件として，(1)統計データの公開性，

⑵統計の利用や使用の公開性，⑶統計の利用や使用の正統性を判断する場の公開性，⑷統計の主題を決定する場の公開性，という四つの公開性が必要であるといえよう。統計の作成・公表・利用における公共性という本稿の題目に即していえば，さしあたり⑷は作成過程，⑴は公表過程，⑵および⑶は利用過程における公共性＝公開性と整理することができる。『委員会報告』で言及されている公共財概念のうち，社会が必要とする基礎的な情報は公開性を持つかどうかが不明であるので，『委員会報告』では基本的に⑴だけが考慮されていると考えてよいだろう。なお，統計の利用を通じてあるべき統計が吟味されるなど，当然のことながら作成・公表・利用はそれぞれ独立しておらず，相互作用を伴っている。

公開性の概念について，付言しておこう。瀧川裕英(2001)は，広義の公開性のレベルを，情報提供，情報公開(狭義)，説明責任にいう説明という3つのレベルに区分している。情報提供とは，情報の送り手によって解釈しフィルタリングされた事実を提供する場合であり，いわゆる一方的な広報活動が該当する。狭義の情報公開とは，あるがままの事実をそのまま提供する場合であり，議会・審議会等の議事録の公開が該当する。説明責任にいう説明とは，情報の送り手と受け手のコミュニケーションを通じ，受け手によって解釈が可能な情報として公開される場合であり，公共性を保証するための公開性とはこのレベルである。

ここで重要なことは，説明責任にいう説明のレベルでは，情報の一方向の流れではなく，双方向のコミュニケーションが要求されることである。ある事物に対する理解はその事物の背景にある理論的枠組みや事実等によって左右されるが，情報の送り手が持つ理論・事実と，受け手が持つ理論的枠組みや事実は異なることが一般的であり，一致するのは送り手と受け手が同一の特定集団に属するなどの特異な場合に限定される。また，情報の受け手には多様性があり，送り手が事前にすべての受け手の理論的枠組みや事実を想定して情報を公開することは不可能である。したがって，情報の送り手と受け手のコミュニケーションを通じて，受け手が解釈可能な情報として整序していくことが要求される。

なお，本稿のようにアーレント，ハーバーマスを嚆矢として公共性を公共圏public sphereとして把握する以外に，種々多様の定義があり，作間(1998)のように「権力による自由」を保証するものとして公共性を定義したものもある。この定義に関しては，本稿で主張する公開性を保証した後のあるべき正統性原理を示しており注目されるが，あるべき正当性の原理そのものも公共圏の討議の場に置くことが必要であるというのが本稿の主張である。

2．統計データの公開性

一般的な公共財の定義から考えても，統計データは公開性を持っているし，『委員会報告』でも統計データの公開を強く推進しようとしている。その技術的な背景として，この十年間に，インターネットをはじめとする情報通信技術(ICT)が急速に進歩し，国民に普及したことを指摘しておこう。インターネットの利用者人口比率は1997年には1割弱であったが，2004年には6割を超え，しかも自宅でのインターネットへの接続は，ブロードバンド使用比率が2000年の7％弱から2004年の6割強へと増加した(総務省『平成17年版情報通信白書』)。このように国民へのインターネットの普及は量・質ともに急速に進展し，多くの政府統計が「統計データ・ポータ

ルサイト～政府統計の総合窓口～」などで公開されるようになっている。

ところで，統計データの公開には，公開される内容と，公開の程度が関連している。統計データの内容の公開性とは，統計表，メタデータ，マイクロ・データの公開に関する問題であり，統計データの公開の程度とは，統計の品質論でいう入手可能性（Availability）や容易性（Accessibility）の問題である。これらのうち，入手可能性については周知のこととして，メタデータとは何か，についてのみ説明しておこう。

統計表を正しく利用する場合には，統計表だけではなく，調査の方法，サンプリング方法，調査方法，集計方法，使用されている用語の説明，指標の算出方法，調査票の様式などの，いわゆるメタデータを十分に理解することが必要である。これらのメタデータは，一般的な統計書では，調査の概要，付録，利用上の注意等の名称で収録されている。メタデータの内容に関しては，国連欧州経済委員会がインターネット上に統計データを掲載する場合のガイドラインを，検索用のメタデータ，解釈用のメタデータ，ダウンロード後の処理のためのメタデータに分類して作成している（国連ヨーロッパ統計委員会，2000）。このうち，解釈用のメタデータに関しては，母集団，統計の四要素，原資料，問い合わせ先などの必須項目に加えて，他の資料との比較可能性，報道発表や要約へのリンク，収集・改訂・計算方法の記述や推計方法，誤差情報や正確性，統計の背景や目的の公開を勧告している。すなわち，メタデータには統計作成の作業過程に関連した情報が含まれており，その公開は統計の品質論での透明性（Transparence）や検証可能性（Verifiability）を高めることになる。このメタデータは，現在でも一部が「統計データ・ポータルサイト」などで提供されており，次第に入手可能性が高まっていくと予想される。しかし，現時点での大きな問題は，国民経済計算の推計方法が公表されていないことである。2005年4月には，1989年以降，輸入額に輸入関税を加えてから消費税率を求めるところを，輸入関税を加える前に消費税率をかけて消費税額を求めていたり，消費税率の3％から5％への変更後も3％で計算していたりする誤りを犯していたことが発覚した（内閣府経済社会総合研究所国民経済部「『平成15年度国民経済計算確報』の計数の修正について」2005年4月28日）。これが問題であるのは，推計過程に誤りがあったこと自体ではない。長期にわたる誤謬の直接的原因は，2003年12月に外部から指摘があったにもかかわらず1年間以上放置されてきたことであるが，間接的原因は，推計方法が公表されていないために推計を追体験して推計方法の正当性を検証することが不可能であったという点にある。この点に関して，『委員会報告』で「内閣府は，SNAの推計手法について情報公開を進める……べきである」ことが指摘されており，改善が予想される。このように統計の公開性においてメタデータの公開が重要であるのは，統計の品質論において定義される正確性（Accuracy）は，少数の統計をのぞいて生産された統計データそのものからは検証することが不可能であり，メタデータに記載されている統計作成の作業過程のチェックを通じてのみ行うことができるからである。

さて，統計表の公開の程度は，『委員会報告』の「公共財」概念でも「行政以外の主体によっても幅広く利用される」ことだと認識されている。日本の場合，業務統計や産業連関表作成の基礎資料となる一部の調査統計など公表されないものもあるが，原則的に統計表は公表されており，入手可能性が満たされており，

とくに指定統計については，統計法第16条第一項において「結果は，速やかにこれを公表しなければならない」と定められている。ここで，統計法にいう公表とは主管省庁で閲覧可能であることが要件であり，従来では，広くは公開されておらず閲覧のみ可能で，容易性を満たさない非収録統計表も多く存在していた。たとえば，2001年家計調査の結果表一覧によれば全55集計区分の結果表が作成されていたが，『家計調査年報』に収録されているのは8区分に過ぎなかった。しかし，前述のようにICTの発展・普及によってインターネットで結果表の公開は急速に進みつつあり，家計調査については2002年からはすべての結果表がインターネットで公表されるようになった。今後も統計表のインターネットでの公開は急速に整備されると予想される。

また，日本の標本調査においては，家計調査のように目標とする標本数に達するまで繰り返しサンプリングを行うような調査が多いこともあり，回収率や誤差情報が公表されないものが多かった。最近では，政府も公表を漸次進めようとしているが，「誤解のないように公表するというのは大変難しい」（第626統計審議会（2005年5月13日）での美添会長の発言）といわれている。公開が情報提供のレベルにとどまっている場合，提供されていない他の情報との矛盾がみつかると，誤解を発生させる可能性が高まる。このような誤解を生じさせないためにも，説明責任のレベルでの情報公開が望まれる。

統計データの公開に関するもう一つの問題は，いわゆるマイクロ・データの公開である。欧米では1960年代以降，カナダのようにオーダーメイド集計の場合や英米のように一定の条件の下にマイクロ・データそのものを研究者その他に提供している場合など多様性があるが，調査票を匿名化したマイクロ・データの公開が進められてきた（松田・濱砂・森，2000など）。このマイクロ・データの公開と理論研究の発展によって「欧米諸国における社会科学の実証分析は，1950年代には公表された集計表に基づく分析が主流であったが，70年代にはミクロデータによる分析に移行」（松井博，2004）した。しかし，日本では統計法第15条によって，社会全般の利益に貢献することを要件として総務大臣が承認する場合を例外として調査票の目的外使用が禁止されているために，欧米に比して学術研究が立ち後れているといわれている。2～3年をめどに専門的・技術的な検討を行うように指摘された『統計行政の新中・長期構想』の後，1996～98年の文部省科学研究費補助金（特定領域研究）「統計情報活用のフロンティアの拡大―ミクロデータによる社会構造解析―」（領域代表　松田芳郎），日本学術振興会科学研究費補助金「ミクロ統計データベース」（ミクロ統計活用研究会　代表2004年度まで井出満，2005年度から森博美），一橋大学経済研究所附属社会科学統計情報研究センター「学術研究のための政府統計ミクロデータの試行的提供」など，限定的・試行的なマイクロ・データの公開が進められてきた。しかし，行政内部での検討は進められているようであるが，統計法の改正がないために本格的なマイクロ・データの提供は進んでいない。

ところで現在，厚生労働省はニートと呼ばれる在学していない若年非労働力への対策に力を入れつつあるが，ニートが社会問題化し，それへの対策を企画･実現するためには，ニート人口の把握やその要因等の把握，すなわち事実認識が必須である。この事実認識は，労働経済白書や労働政策研究・研修機構（2005）に見られるように，労働力調査や就業構造基本調査のマイクロ・データの利用（白書等では特別集計と呼ばれる個票の集計）にもとづ

いている。このように，マイクロ・データの利用は，通常の集計による統計資料では把握できない事実を統計的に認識することによって，政策の企画立案や政策効果の測定にも有効な手段となる。このように，統計の作成過程で収集された情報を社会に有用な情報として還元するためにも，マイクロ・データは行政部内の利用に限定せず広く公開されることが望まれる。

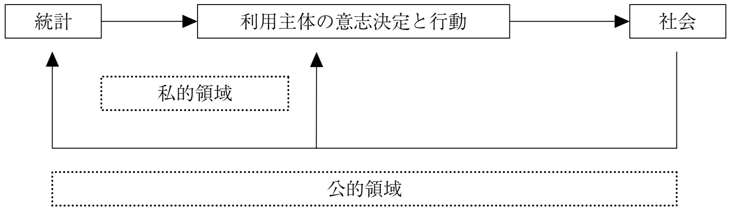
３．統計利用の公開性

前述したように，統計利用の公開性は，一般の公共財では要件とされないものであり，社会に必要な基礎的な情報として政府統計に固有のものである。

2004年11月に経団連がまとめた報告書では，「企業は，統計によって，景況をはじめ，市場環境，所属業界の生産動向や経営状況などを把握」しているのであり，政府統計は合理的な意志決定に必要であると書かれている（経団連，2004）。このような企業や個人の統計ニーズを重要視する視点は，政府統計の主たる役割が，戦後直後においては国民の生活レベルや地域間の経済格差の把握にあったが，高度経済成長期には景気判断や経済成長率の評価になり，現在においてはマーケット参加者が先行きの金融経済政策を判断していく上での材料に変化しているとみる元日銀統計局長村山昇作（座談会，1999）の視点と同様であろう。

下図に示したように，統計は情報の一つであるから，かならず利用主体の意志決定とその行動に何らかの寄与をする。しかし，その主体の意志決定と行動が，広く社会に影響を及ぼさない場合には，利用は単なる私的領域にとどまり，公共性を持たない。逆に言えば，利用主体の意志決定と行動が広く社会に影響を及ぼす場合には，その利用は公的領域に属し，どのように利用されたかが公開されることによって公共性を持つ。たとえば，研究者による統計利用は，その利用結果が論文として公開されるならば公開性＝公共性を持つが，企業に所属する研究者が利用結果を企業内部にとどめるならば公開性は持たない。また，この利用過程は，被調査者が公開された統計の利用過程を見ることにより，自分を取り巻く社会にとって統計がどのような影響を与えているかを判断することができることを示している。

このような公開性を持つ統計利用の典型として，立法過程，すなわち議事録の公開が原則とされる国会や地方議会での統計の利用があげられる。そこで，国会での統計利用の実態をさぐる第一歩として，機械的ではあるが国会会議録検索システムによって，指定統計について指定年から2005年4月までに国会で言及された回数をまとめてみたものが上表である。国勢調査と家計調査については1000回以上の言及が行われており，年間20回以上国会で取り上げられた。また，毎月勤労統計調査，労働力調査など14統計調査については100～999回，就業構造基本調査，全国消費実態調査など7指定統計については50～



国会で言及された指定統計の回数（指定年～2005年４月）

回数	指定統計数		指定統計の例
1000回以上	2	2％	国勢調査・家計調査
100～999回	14	15％	毎月勤労統計調査・労働力調査など
50～99回	7	7％	就業構造基本調査・全国消費実態調査など
20～49回	9	10％	科学技術研究調査・建築着工統計など
10～19回	7	7％	工業統計調査・自動車輸送統計など
1～9回	29	31％	民間給与実態統計・全国物価統計など
0回	26	28％	ガス事業生産動態統計・薬事工業生産動態統計など
合計	94	100％	

注１）国会会議録検索システムを使用して作成した。
２）単発で実施された統計調査は除くが，すでに廃止された指定統計も含む。
３）本会議のほか委員会での発言も含む。
４）一回の本会議・委員会で複数の言及があった場合は，１回と数えている。

99回言及されている。しかし，全く言及されていないものが26指定統計，９回以下のものが29指定統計あり，この２区分で約６割を占めている。すなわち，我が国で重要と位置づけられている指定統計においても，その半数以上がほとんど国会では言及されていない。もちろん，単に「政府統計」あるいは「統計」として言及された場合も多いので，一概には言えないが，全般的にみて立法過程での統計利用は少ないといえるであろう。

次に，行政機関での利用は広く社会一般に及ぼす影響は大きいが，影響の大きさだけが公開性の要件ではない。すなわち，どのように利用されているかが公開されていなければ，企業等の私的利用と変らず，理法方法が公開されて初めて公開性を持つ。国勢調査がどのように利用されているかを説明した総務省統計局のウェブサイトによれば，法令に基づいて利用されている事例として，衆議院議員選挙区画定審議会設置法による議員定数の決定，地方自治法による地方自治体の設置要件，地方交付税法による地方交付税の配分額の決定などがあげられている。これらの事例は，一票の格差をめぐって，議員定数是正が裁判で争われることがあるものの，法律によってどのように利用されるかが明らかになっている。一方，行政施策での利用の事例としては，少子高齢社会に対応した政策，住みよい街づくり，防災対策などがあげられており，具体的な一例をあげると，少子高齢化社会に対応した政策の一つである子育て環境の充実をはかるために，０～５歳人口に６歳人口の半分を加えた就学前の人口が使用されていると書かれている。この統計利用は，就学前の人口を子育て環境の充実という政策領域に利用しているという説明にはなっているが，どのように利用されているか，具体性を持った説明にはなっていない。このように，行政機関による行政施策での統計利用は，どこで利用されているかは明らかな場合があるが，どのように利用されているかは明らかにされていない。

『統計調査総覧』によって指定統計・承認統計・届出統計の目的をみても，「～行政のための基礎資料とする」という目的であるものが大半を占めており，どのように利用されているかが不明なものが多い。若干古いデータだが，1989年に総理府広報室が実施した統計調査に関する世論調査でも，統計調査の結果が行政に利用されていると思っているものは38％に過ぎず，そうは思わないもの32％，わからないもの30％であった（山田茂，1990）。

以上のように，どのように利用されているかが明らかにされない場合，当然，その統計

を利用して国民にとって有益な施策を講じて実施しているか否かを，明らかにすることはできない。つまり，行政内部の情報にアクセスできる特定の集団だけが統計利用の正当性を吟味できるのであり，統計の利用や使用の正統性を判断する場の公開性は保証されていない。

このように統計の利用における公開が進んでいないことは，国民にとって政府統計の利用結果の是非の判断ができないことを意味し，統計への無理解と拒否感を増長することになる。例えば，前述した1989年の世論調査によれば，統計調査に非協力な理由として，プライバシーと関連すると考えられる「個人の秘密を知られないから」が22％，「統計を作る目的以外に利用されるおそれがあるから」が4％に対して，統計の利用過程とその結果に関連する「どのように利用されているかわからないから」が30％，「調査に協力しても自分の利益にならないから」が12％と多く，さらには「面倒だから」が63％であった（山田茂，1990）。また，2005年国勢調査で，国勢調査の見直しを求める会が設置した電話ホットラインで受け付けた相談件数は1306件で，「国勢調査が役に立っているのか」「650億円を投入する意味があるのか」「なぜここまで聞かれるのか」「見直すべきだ」という意見が多かったという（「国勢調査を見直す会」ウェブサイト。同様の記事は2005年9月27日・10月8日付け毎日新聞朝刊など）。これらの意見は，プライバシー権を統計調査が侵害しているのではないかという問題ではなく，個々人にとって統計調査の有益性を見出すことができないという問題が底流にあることを示している。個人にとって統計調査が有益であると判断されれば，プライバシー意識は表面化しないであろう。たとえば，個人にとってセンシティブな情報といわれる所得に関する情報は所得税額の決定に使用されているが，政府によるこれらの情報取得がプライバシーを侵害しているとはいわれない。なぜならば，これらの情報は住民基本台帳制度によって教育や社会福祉政策などの権利とリンクされているからである。したがって，政府統計の政府による利用が公開され公共性をもつならば，統計への拒否感や「面倒くさい」という感情は減少するだろう。

ところで2004年6月の経済財政運営と構造改革に関する基本方針2004において「国・地方で，時代の変化を反映した的確な情報把握と迅速な情報開示のため，農林水産統計などに偏った要員配置等を含めて，既存の統計を抜本的に見直す。一方，真に必要な分野を重点的に整備し，統計制度を充実させる」ことが決定されている。実査機関である農林統計・情報センターの職員減少は農林統計の精度の低下を招くであろうが，戦後の食糧難時代とは異なって，ある程度の農林統計作成のリソースの減少はやむを得ないであろう。ところで，「真に必要な分野」の統計とは何なのであろうか。『委員会報告』によれば，早急に整備すべき統計として，経済センサス，GDP関連統計，サービス分野の統計，ストック統計などをあげている。これらの統計は早急に整備されるべき統計であることは疑いないであろうし，統計の必要の度合いに応じて政府統計のスクラップ・アンド・ビルドを行っていくことは必要である。これらは，行政の必要とする統計だけが作成され，政府以外の統計利用者のニーズを表明する場が設定されていなかったなかで，統計審議会の『統計行政の新中・長期構想』によって提言されてきたものである。しかし，統計審議会が政策審議・基準作成機能を失い法執行型審議会へ移行し，建議機能を失ってしまった現在，社会が必要とする統計とは何かを決定する場はど

こにあるのであろうか。そのため，統計データの公開・統計利用の公開とともに，統計体系の整合性を担保しつつ，どのような統計を整備することが必要なのかを，特定の集団に限定されることなく幅広い社会階層に開かれた討議の場が必要とされている。例えば，イギリスのNational Statisticsで設定されているユーザーズ・グループのような利用者会議のような場を設定することが考えられる。大切なことは，政府内での統計利用者，政府以外での統計利用者，統計作成機関が，統計データや統計利用の公開を通じて相互作用を及ぼし，より社会とって有益な統計を作成し利用していくことである。

おわりに

以上のような公開性の強調が，現在の政府統計がもつ諸問題にどのように対応しうるかという一つの含意として，被調査者の非協力意識からくる統計調査環境の悪化との関連について言及しておこう。

政府調査統計は，政府が作成しているが，被調査者の協力なしには生産することができない。その点で一般の純粋公共財とは異なり，環境汚染の削減投資のように，社会の参加者が自己所有の私的財を自発的に出し合う公共財であると考えられる。このような公共財の場合，ゲーム理論によれば，すべての人が協力したときに全員に対する便益が総コストよりも大きくても，各個人のコスト負担が自己の便益よりも大きい場合には，非協力的な態度，つまりフリーライドすることが最善の戦略になる（大岩雄次郎，2005）。統計調査に関していえば，被調査者が調査に協力する時間やプライバシー意識等がコストに相当すると考えられる。現在の統計調査環境の悪化は，このコストが政府統計から得られる便益を上回り，フリーライドすることが最善の戦略になっている状態であると推測される。このような社会的ジレンマを脱するには，調査負担の軽減によってコストを低減するとともに，政府統計から得られる便益を増大させ，自己の便益が自己のコストを上回るようにしてやればよい。そのためには，これまで非顕示的であった政府統計の便益を，政府統計そのものを有効利用し，公開することによって顕示的なものに変化させることが必要である。以上のような点からみて，とくに利用過程における統計の公開性は，統計調査環境の悪化に対応する一つの処方箋となりうるのではないだろうか。

そこでは，統計データのみならず，統計がどのように利用されているか，また，どのような統計が必要かについても公開性が要求されるし，真に社会に必要な統計が作成される必要がある。その際，何が社会にとって必要な統計であるかを見極め，政府において正しい統計利用が行われているかの責任は，統計利用者としての我々の課題となる。我々には，より積極的な政府統計の利用，政府での統計利用過程のチェック，そして社会に必要な政府統計に対する積極的な発言が求められるであろう。

参考文献

大岩雄次郎（2005）「なぜ囚人のジレンマが起きるか」加藤寛編『入門公共選択』勁草書房.

大住荘四郎（1999）『ニュー・パブリック・マネジメント』日本評論社.

————（2003）『NPMによる行政革命』日本評論社.

経団連（2004）『統計の利用拡大に向けて—景気関連統計を中心として—』11月.

作間逸雄（1998）「“ケイパビリティー”で考えてみよう」『専修大学社会科学研究所月報』No. 425.

座談会（1999）「変貌する日本経済と求められる統計改

革」『にちぎんクォータリー』秋季号.
瀧川裕英(2001)「公開性としての公共性」『〈公私〉の再構成』日本法哲学会.
松井博(2004)「政府統計ミクロデータの利用」21世紀COEプログラム「社会科学の統計分析拠点構築」第2回全体集会配布資料.
松田芳郎・浜砂敬郎・森博美編(2000)『講座ミクロ統計分析』日本評論社.
村松岐夫(2001)『行政学教科書　第2版』有斐閣.
山田茂(1990)「被調査者からみた統計調査」『政経論叢』国士舘大学政経学会 第74号.
山田満(2001)「『統計利用者のための統計学』から『公民権としての統計学』へ」杉森滉一・木村和範編著『統計学の思想と方法』北海道大学図書刊行会.
労働政策研究・研修機構(2005)『若者就業支援の現状と課題』労働政策研究報告書 No. 35.
Office for National Statistics (2002), *National Statistics Code of Practice*.
"United Nations Statistical Commission and Economic Commission for Europe (2000)" *Guideline for Statistical Metadata on the Internet*.

コメント

山　田　　　満

1．統計をめぐる公共性とは何か

「公共性」という言葉には，ある種の曖昧さといかがわしさが付きまとう。「公的なもの」とは何かが明確に定義されないまま,「私的なもの」と「公的なもの」が対置され,「公(おおやけの利益)」の名の下に「私的なもの」が抑制され，統御されていく。今日のように「官」から「民」への仕事と権限の委譲が取りざたされる時代に, それに伴って「公共性」に関する語りが前面に出てくることは当然のことである。「事前規制」の対象から外れた「民」の活動における「社会性」が「公共性」という形で「事後的に」問われないはずはないからである。

金子は小泉改革を支える「小さな政府論」が1990年代以降のNPM論の影響を受けていることを指摘しているが，行政によるマクロな指標管理的な政策や行政指導を支えてきた「行政のための」(統計)情報システムが,「官」から「民」への仕事と権限の委譲のなかで,「官」のためだけでなく「民」の活動も支えるものとして再編されなければならないことは時代の趨勢というものだ。「官」の主要な仕事のひとつは，政府行政サービスの顧客である「民」の要求に応えて,「民」の自律的活動を下支えする情報基盤の整備を「民活」を含んで効率良く行なうことにあるというわけだ。ここでは「官」が領有するものを「民」へと開放するにあたって，統計の作成･公表･配布･利用の仕方のあらゆる側面で「公共性」が問われることになるのである。「官のもの」とは定義によって「公のもの」であるはずなのだから,「公のもの」である(統計)情報システムを「民」へと開放するには,「公共性」が先ず担保されなければならないからである。しかし，担保されるべき「公共性」とは何かを問うことは, 逆に,「公のもの」の「公共性」を問い返す契機となるかもしれないのである。先ず，問われるべきものは,「官」が作成し提供/秘匿してきた(あるいは作成せず提供してこなかった)統計情報システムの「公共性」なのかもしれないのである。

「政府，とくに諸省庁が作成してきた統計は「行政による行政のための統計」であり，その作成の論理も，その利用についての便宜も「民」の要求に十分応えていない。だから，政府統計を「公共財」として位置づけ,「パブリックな」要求に応えるものにしなければ

ならない」といったレトリカルな議論は(「民」を「パブリック」に置き換え,「官(政府)」を「非パブリック」の利権関係者にしてしまっているという点で)奇妙に聞こえるが, そこにあるのは行政の諸施策に対する「不信」であり, 行政とそれに係わる一部の者たちによる情報の独占に対する「恐れ」と「不公平感」であり, 多額の税金を投入して調査/収集された情報が有効に利用されていないのではないかという「もったいなさ」への感覚である。要するに,「公」の「公共性」への不信,「官」が十分に「公共的に」機能していないのではないかという苛立ちである。

２．ハバーマスの公共圏の理論とブレア政権の英国統計改革

「公共性」という言葉から抑圧性を取り除き, 開放的な空間として再定義しようとしたのは, フランクフルト学派の社会哲学者ハバーマスであった。ハバーマスにとって「公共性」とは,「公共圏 public sphere」のことである。例えば, それは公園であり, 市民が集い散策し時に論議し殴り合う「オープンな共有スペース」である。共有スペースとは, 排他的に誰かが場所を占有できない, 誰のものでもない空間であり, それゆえに誰もが「自由に」使うことができない空間である。そのスペースを「自由に」使うには, 定められたルールに従うことが要求されるのであり, それゆえに誰もがその空間の設置と空間の利用の仕方(その使用の規則, 管理運営の方法, 維持管理の資金, 等)に対し関心を持ち, 関与せざるを得ない性質の空間である。誰もが関心を持ち, 関与し, それらに「信頼」を寄せるからこそ, 公園は開かれた空間として,「市民の憩いの場」であり,「市民の自己表現の場」でありうることができるのである。

英国の政府統計/政府統計システムは,「中央統計局の主たる任務は, 中央政府の要求に応えることである」という基本的認識の下, 予算の削減と民間セクターに課された義務の軽減に取り組んだ1980年代前半のレイナー改革の推進によって決定的な打撃を被ったが, 1980年代の末以降, それを建て直す際に, 合言葉となったのは「公衆からの信任(public confidence)」と「他から干渉されない高潔さ(integrity)」であった。本格的な建て直しには1997年のブレア労働党政権の誕生を待たなければならなかったが, 失われた信頼を取り戻し, 建て直すには, 政府統計情報システムに自律性と専門性をもたせ, それを「公共性の空間(public sphere)」に置き, 公衆に信任され, 公衆に支えられるものとする必要があったのである。国家・政府の活動が政府・国家ということだけで「公のもの」として信頼され信任された時代が去った後に, 問われたのは, 政府･国家の活動の「高潔さ」と「公共性」だったのである。広範囲な公衆に信頼されない統計は利用されないし, 信用されない統計数字に基づいた国家の政策は公衆の信任を得られない。信頼されず使われない統計を作成する統計調査に公衆が協力するはずもない。統計作成に必要な情報がどこからどのように収集され, どのように蓄積されデータベース化され, どのような手続きを経て統計数字化されるのか, また, それがどのようにして公表・配布され, どのように利用され, その利用の結果, 社会経済生活にどのような影響がもたらされるのか等, が知ろうと思えば, 少しの努力を払えば, 分かるようになっていることが信頼を得るための制度的条件となろう。信頼の回復には, 中央統計局の機関としての「専門性」と「独立性」を確保した上で, 統計情報システムを「制御可能な」ものとして構築し, 統計の作成・公表・配布・利用の諸過程とそれに係わる諸論議を「公共

性の空間」に載せ，「公共的な」ものとして社会の組織体のなかに組み込み直さなければならなかったのである（re-articulation）。

金子は，「公共圏」の核心を，統計データ（メタデータを含む）の公開と（政府による）統計利用の諸過程の公開を主たる内容とする「公開性」に置いているが，公開性（あるいは，情報への「等アクセス性」）が公共圏の成立のための基礎条件であるとすれば，妥当な主張であろう。しかし，金子の論議は，公開性が如何に重要であるかということの論証に留まっているように思う。今後，さらに進んで公開性をキーコンセプトとした統計の制度設計の問題に踏み込むことが期待される。

第2章　地域における統計の作成と利用

藤　江　昌　嗣

はじめに－地状学への期待

統計学は国の状態を記述する学問である「国状学」として出発したとされるが，今や「地域」という地平で地状学（ちじょうがく）としてもその役割が大きく求められていると思われる。ここに「地状学」とは，「地域の状態を記述・分析する学問領域」を指す。もちろん，「地域」とは相対的な概念であり，例えば，地球規模で考えれば，それは，グローバリズム（Globalism）―地球主義―の対抗概念としてのリージョナリズム（Regionalism）との係わりで，国あるいは地理的領域を単位に措定できるし，国内に限定すれば，全国ベースのマクロ的単位（政府）に対するミクロ的な単位（地方自治体）として，あるいは，都市あるいは地方として措定することも可能となる。

その意味では，「地域とは何か？」という問いは最初に答えなければならない問いであるが，ここでは「地域」を差し当たり「国家」というレベルとは異なる「地方公共団体」という行政単位･行政区域もしくは「地域社会」とも呼びうる生活行動圏としておく。

さて，所謂「地方分権」の時代が強調され始めた1990年代以降，地域においては様々な変化が生じてきている。大きな変化の一つは，地域における経済・経営・労働・生活・健康・福祉・教育・安全等多くの分野での諸問題の累積である。所謂「市場メカニズム」の導入と「小さな政府」化，グローバル・スタンダードとの調和のための国内諸制度・諸規制・諸慣行の緩和は社会に「格差」「不安」「危険」等を生み出し，「自立・自助」という原則の下，地域レベルでその解決が求められてきている。そして，第二には，こうした諸課題の解決方法に係わり，その発生・増幅要因とも考えられる民営化・外部委託を軸に進められてきていた1980年代以降の「行政改革」―長期的には，「道州制」も視野に入れた国・地方の行政組織の再編成―の流れの延長線上に，1990年代後半には地方財政危機を直接の契機とした「地方行政改革」が進行し，これと連動しつつ地域における統治（ガバナンス）主体，行政サービスの担い手の多様化という変化が生じてきたことである。

こうした諸問題の地域における発生と統治（ガバナンス）構造の変化は，統計の対象・作成・利用にも影響を与え，新たな取組みを展開させてきている。すなわち，「調査対象」という点では，「地域」における住民の教育･福祉・健康などの生活環境や生活状況，中小企業・産業活動の状況（業況），そしてNPO非営利組織やボランティアを行う個人・団体，コミュニティビジネス・SOHO，また市民活動支援を行う行政組織・NPOなどの中間支援組織の活動，そして行政の提供する事業やサービスの内容，その事業パフォーマンスや財政パフォーマンスの評価等が重要な対象に加わってきている。

また，地域における統治主体や行政サービスの担い手の多様化との係わりでは，統計の

作成・利用主体が重要なものとなるが，地域において新たに登場した担い手であるNPO非営利組織やボランティアを行う個人・団体，コミュニティビジネス・SOHO，また市民活動支援を行う行政組織（担当部署）・NPOなどの中間支援組織などの活動は，地域における「統計の作成や利用」に変化を与えてきている。統計の作成を含め，利用における諸主体間の協働（Collaboration）が進んできたことはこの間の大きな特徴である。こうした統計の作成・利用主体における変化を把握することも重要な課題である。

「地状学」すなわち，「地域の状態を記述・分析する学問領域」は地域における統計「地域統計」の作成と利用を内包するのであり，一つの政治・経済・社会システムの複合体としての「地域」の把握と分析，問題の発見・解決のためにその役割が求められてきているのである。

また，忘れてならないことは，この地平に統計研究者の地域での役割が大きく開かれ，また，期待されていることである。

そして，統計学の「地状学」としての展開は情報通信技術（Information Technology）の発展を抜きにしては語れない。インターネットを代表とするIT（情報通信技術）の発展は，私たちの生活，経済・社会，そしてコミュニケーション（情報伝達）の範囲や手法を大きく変革しつつあり，“いつでも，どこでも，誰でもが情報の受発信をできる社会”であるユビキタス社会の登場として紹介されてもいる。「情報ネットワークを視野に含めることによってはじめて，統計学は現代の統計活動の特徴を十分にとらえ，現代の社会・経済に即した方法を発展させることができる」（野澤正徳，2002，p.8）のである。その意味で，統計学を「統計情報学」に広げる必要性は説得力をもつのであり，本論で展開する「地状学」において，統計は「地域において収集・整理・作成・公表・利用される情報（Information）の一部をなすもの」（同前）と考えられるのである。

1．「地域統計」―地状学―の展開

1.1　「地域統計」・「地域統計情報」の展開

「地域統計」あるいは「地域統計情報」については，統計学を始め，社会学，地理学，農学，環境学等諸分野の研究者がその作成・利用について説明を行ってきている。

例えば，地域統計の整備については，宍戸邦彦，川口清史他の論考があり，川口はすでに1980年代半ば，経済の構造変化が進む下，「高度成長末期から諸矛盾が地域レベルで現れるなかで，地域統計の整備が問題とされた」（川口清史，1986）と指摘している。

また，木下滋は同じ時期に「高齢化を中心とする人口構造の変化，コンピュータ化，コンピュータと通信の結合，サービス経済化，ソフト化，情報化，脱工業化……産業構造の変化，円の国際化をはじめとする経済の国際化……このような構造変化の方向を発見し，混迷する運動に指針を与えることは，あらゆる社会科学者の任務である。もちろん経統研もその任務からまぬかれることはない」（木下，1986，p.220）と，経済にとどまらない構造変化を予感しつつ，地域統計を課題解決のために利活用することの重要性を指摘していた。

また，藤岡光夫はその教育実践において「地域統計情報論」を開講し，地域統計情報としての社会統計情報の収集と作成，加工・利用方法について学ぶことを目的に，地域統計の中でも人口，労働，生活，健康，家計等に関する統計を中心に，⑴地域統計情報の収集－

政府統計や自治体統計の情報収集，⑵地域統計情報の吟味・検討，⑶データ加工・処理方法－それらの統計の作成，加工・利用方法（データ処理）について講義を行なっている。地域統計そのものと地域統計に関する情報という2つの構成部分を総称して「地域統計情報」としているようにも理解でき，地域統計に加え，地域統計情報も対象にしたより広い概念となっている。

また，宮内泰介（環境社会学）(2004)は，地域の問題，社会的な問題を的確に把握し，解決するための調査を念頭に置きながら，私たちの社会がどう成り立っていて，どこへ向かえばいいのかといった問題を考えるための調査について論じている。そして，「NPOの活動，市民活動は，正しい情報に基づき，今何が求められているのかを的確に把握して，初めて社会的に意味が出てくること」（宮内泰介，2004，p.9）を強調する。

こうした地域統計・地域統計情報へのニーズの高まりは，統治主体の多様化も踏まえた地域政策あるいは地域政策学との係りで理解することが必要である。

1.2 「地域統計」・「地域統計情報」の類型

このような考えを踏まえると，「地域統計」あるいは「地域統計情報」は，①小地域統計，②実態調査，③業務記録（第二義統計），④市民による調査　といった4つに分類して考えることが可能となる。以下ではいま少し，それぞれについて解説を行うことにする。

①　小地域統計

従来，地域調査は，「小地域統計」もしくは全体調査（悉皆調査）に対する部分調査（典型調査）と位置付けられてきた。全体調査が国（政府）のマクロ政策に資する調査であることを踏まえれば，「小地域統計」は，マクロ政策のための調査単位・報告単位における集計・分析として位置付けられるものである。全体調査（悉皆調査）であれ部分調査であれ，その報告単位は地方自治体（行政単位）であることに変わりはないが，仮に地方自治体を国（マクロ）に対してミクロと呼ぶとすれば，「小地域統計」はミクロレベルのマクロ政策に資する統計ということになる。小地域統計に関しては，その推計手法については美添泰人(2001)が，また，人口については上野健一(2004)がある。

②　実態調査

部分調査の中でも公害問題における被害実態調査などは上記のマクロ・ミクロ関係とは異なる独自の目的と価値をもつ調査として位置付けられるが，後に見るように，この間，地域において発生した諸問題―経済・経営・労働・生活・健康・福祉・教育・安全等多くの分野での諸問題―に係わる実態調査が困難な状況下にもかかわらず実施されてきたことは重要である。

③　業務記録（第二義統計）

地方自治体は，従来から，国の委任事務を始め，現物給付，人的サービスを行ってきた。こうした業務は，行政の組織ごとに行われるので，その行政記録・業務記録として蓄積されている場合が多い。また政府による*e-JAPAN*戦略Ⅰ，Ⅱ[1]の推進もあり，コンピュータネットワークを活用した行政活動や市民との双方向のコミュニケーションは，多くの情報記録を生み出している。こうした情報記録も，経済，財政，公営企業経営，労働，生活，健康，福祉，教育，安全等多くの分野にまたがるものであり，各分野での諸問題の実態把握や解決における役割は非常に大きい。

インターネットを活用した情報記録は，低コストで，行政の取組みや市民との協働を伝え，政策に反映させるための財産でもあり，業務記録（第二義統計）の意義・役割は益々大きくなっていくと思われる。

地方自治体レベルでのミクロデータの公開・利用という課題もあるが，個人情報の保護と統計調査における情報の保護を正確に区別し，かつ保証しつつ，積極的な統計情報・業務記録の公開がすすめられるべきである。

④　市民による調査

地域の諸政策における主体－統治主体の構造，すなわちガバナンス構造の転換が進行していることも1990年代以降の特徴である。そうした中で，市民やNPO，ボランティア組織などによる調査が増加してきている。

宮内泰介（2004）によれば，「市民による調査は，行政・専門家・マスコミによる調査の代替物，簡易版ではありません。たとえば研究者の調査の多くは，その研究者が所属する学会で評価されることを目的としています。また，調査に「学問的な厳密さ」を求めます。……それに対し，市民による調査は，具体的な問題発見・問題解決が目的であり，自分たち自身の状況を知って，何が望ましいのか，何をなすべきかを考えるために行われます。……有効な方法はいろいろと組み合わせて使ってみることができますし，学問的な厳密さよりも実践的な説得力に重きを置くことができます。さらに市民による調査は，調査をする人自身と，問題を解決しようとする人自身が同じである（あるいは，近いところにいる）という最大の特徴をもっています。調べて発表して終わり，ではなくて，調べる中でやらなければならないことが出てきたら，どんどん実践する，ということもできます。」（p.13～14）「厳密な「科学的」調査より，市民による調査の方が本当の意味で“有効”な場合が多い，ということです。専門家がちゃんとやればわざわざ市民が調査する必要はない，という話ではないのです。」（宮内泰介2004，p.14）

また，大阪経済大学地域政策学科（2003）は「地域主義」「現場主義」を前提に，商店街，伝統産業の都市，中小企業の都市，農村，経済，政治などを対象にした地域，自治体の調査事例を含め，フィールドワークの方法を体系的に紹介している。

近昭夫・岩井浩・木村和範・福島利夫編（2006）も市民による調査・結果の利用に資することを期待しつつ，現代経済の諸課題について統計・統計情報を用いた分析・考察の実例とその際の注意点に触れた書である。

2．地域における統計の作成と利用

2.1　地域における厳しい財政環境－戦後3回目の財政危機

地方自治体は戦後3回目の財政危機を迎えている。1回目の危機は，戦後復興の混乱期で，1954年度には赤字団体が34都道府県，2,247市町村という大規模なものであった。また，2回目の危機は，1974年の第1次オイルショック後の不況に伴うもので，1975年度には赤字団体が27都道府県，269市町村となった。しかしながら，1回目の危機は高度経済成長による税収増で，また，2回目の危機は，1980年代前半の対米輸出主導による景気回復と1987年からのバブル景気の税収増で解消された。福祉国家の追求との関連で，国の財政赤字が話題にされることはあっても，地方自治体は相対的に豊かで，仮に地方財政の赤字に陥ったとしてもそれは景気循環の中で解消されていくフロー面での財政危機（財政

赤字)であって，構造的な危機という認識は弱かった。こうした過去の2回の財政危機に比べ，現在の3回目の財政危機は，バブル経済最中のリゾート計画，その後のサッカー・ワールド・カップ誘致合戦におけるスタジアム建設に象徴される無駄な建設事業を含め，健康・福祉などの分野を除き，高度成長期以降の公共事業(ハコ物・道路)中心の歳出構造がそのまま維持され—むしろ，景気対策として地方の財源がフルに出動され—ため，地方自治体は財政調整基金を中心とするその貯金を吐き出し，フロー・ストック両面での財政危機に陥ったのである。

財政学者である重森暁は今回の戦後3回目の地方財政危機を，①国家財政危機との同時進行，②先進国中最悪の地方財政赤字，③大都市圏先行型の財政危機，④借金依存体質の蔓延として特徴づけているが，地域における統計の作成と利用という取組みにもこの地方財政危機はボディ・ブローのように効いてきているのである。

いち早く財政危機に直面した国から始まった行政改革は，現在，地方行政改革—「三位一体改革」,「地方分権」—という形で急ピッチに進められている。財政危機を理由にした合併自治体，合併しない地方自治体いずれの行政改革も，公務員制度の「改革」と相俟って進行している中央・地方官庁組織の改変を通じて「統計改革」「統計制度改革」にも連動するものである。

すなわち，国の行政改革は国家レベルにおける統計の作成・利用の見直しという動きを生み出してきているが，地方財政危機を背景にした地方行政改革は，地方自治体レベルでの統計の作成への抑制効果をもつことが懸念される。しかしながら，こうした厳しい状況の下でも，地域における統計の作成利用はさまざまな形，すなわち「協働」の形態をとりつつ実施されてきている。

ここでは前記4つの類型の枠を外して，地方・地域における会員の研究，新たな主体による，あるいは新たな協働の形態での地域における統計の作成・利用，そして地域課題の研究について紹介を行うことにする。

具体的には，地域景況統計，人口・労働・生活・健康・子育ての研究，費用便益分析・経済効果分析をテーマにした研究，自治体財政分析，自治体の行政評価，公会計制度の導入, 地域の構成員(ガバナンスの主体･構成員)を主体とした統計・調査などについて触れていく。

2.2　地域景況統計の研究

菊地進，坂田幸繁，佐藤智秋，鈴木幸明，田浦元らは地域経済・地域企業の分析，とりわけ「景況調査」の企画・実施・分析に取り組んできている(菊地1998，2000，2004，田浦2004)。菊地によれば，それは①中小企業家同友会の景況調査，②中小企業庁の景況調査の研究などからなり,「地域景況統計の総合的研究」として最終仕上げの状況となっている。

中小企業家同友会全国協議会(中同協DOR, 2004年)は特別調査『転換期における雇用と労働の変化』を実施し，非正規雇用増大の実態について解明を行った。

また，中小企業庁の景況調査の研究では，中小企業庁・中小企業基盤整備機構『中小企業景況調査』(商工会，商工会議所，中央会が実査)の100回記念事業として，調査に関する評価・分析も行なっている。

また，佐藤智秋も愛媛県を対象に「地域景況統計」の実施への協力並びに実施状況の研究を精力的に行っている。すなわち，愛媛大学法文学部総合政策学科のスタッフからなる総合地域政策研究会と愛媛県中小企業家同友

会(EDOR)が相互に協力し実施している愛媛大学・愛媛県中小企業家同友会「景況調査(EDOR)」が2003年度から公表されている。

地域景況調査は，1970年代以降，故木村太郎，故山田貢そして，岸啓二郎(後に藤江も参加)が四半期雑誌『埼玉県の経済』で先鞭をつけていたが，地方財政危機が進む中，民間委託事業化され，全国ベースでみても自治体による調査が減少してきており，菊地，坂田，佐藤，鈴木，田浦らの共同調査並びに研究は，貴重かつ重要なものとなっている。

2.3　人口・労働・生活・健康・子育ての研究

菊地らの作業が中小企業との協働作業であるのに対し，藤岡光夫は人口・労働(職業別死亡率調査)から生活・健康そして子育てへと研究の翼を広げつつ行政職員・市民との協働作業を実践し，ネットワークを着実に構築してきた。とりわけ，近年は精力的に保健師との協働作業を進めてきた。やや詳しく跡付けてみよう。

藤岡は，藤岡(1999)においてミクロデータを用いた職業別死亡統計比較方法の検討，また，同(2001)では，センサスデータによる産業・職業別労働移動の推計を行い，さらに同(2000)では職業別死亡統計の日・北欧比較と年齢別死亡分析を，また，同(2002)ではフィンランドと日本における年齢別・職業別死亡という国際比較を行ってきていたが，それは日本という国(地域)の労働現場・職業と健康・生活の特性の関係への注目と疑問そしてその改革を展望した実証研究である。

その藤岡は，近年，生活・健康・保健・子育てに研究と協働の領域を拡げてきている。たとえば，その健康問題・保健調査と社会統計学の役割に関する問題意識は藤岡(2002a)に示されており，経済統計学会第49回総会においては長崎原子爆弾被爆者の生活・健康調査(SPA法による社会調査と統計的研究)の報告(藤岡・吉峯悦子(2005a)を行っている。

また，渡辺訓子(2005)は個人情報保護法が全面的に施行された状況を踏まえた上での保健活動の中の調査研究と個人情報保護の問題について報告を行い，田中美穂・瀧口京子も横浜市鶴見区を例に，10年をかけた周到で丁寧な準備に基づく地域における母子保健・2歳児育児実態調査とその支援活動について報告を行った。

調査テーマの重要性や困難はもとより，個人情報やプライバシー保護の問題が指摘される中で，市民・行政(保健師等)・研究者の協働で行う統計調査の準備・利用過程での配慮と協同作業並びに結果の意義が明らかにされた。

2.4　費用便益分析・経済効果分析をテーマにした研究

1970年代から公共投資の経済効果について産業連関分析を用いて研究を継続してきた両雄が木下滋と土居英二である。フィールド・スタディを重視し，テクノポリスを含め「地域開発」に対する検証を続けてきた木下は，木下(1996)並びに絶筆となった木下(1997)で，都市間競争や日米比較を通して都市問題・都市の行く末について改めて考察した。**開発と都市**というフィルターを通して「人間」・「地域」・「社会」を考え続けた研究者の憤死は経済統計学会にとり大きな喪失であった。

他方の雄である土居は1989年の消費税導入前後における消費税の効果の研究に高い水準とアドボカシーを示し，また，地域産業連関分析の普及という裾野の広い重厚な取組みを行ったが，現在は，公共投資の経済効果の延長上に位置づけられる，しかし，新たな要

素を含む「政策評価」にかかわる研究・調査に取り組んでいる。ここに政策評価とは，ある地域政策の費用便益分析・経済効果分析を指す。その主な成果は，政策評価システムと統計情報については土居(1997)，公立病院の地域経済効果については土井・中野(2001)，また，CVM法を用いた自然系博物館整備の費用便益分析として土居・望月(2002)が，さらに熱海梅園を例にとり公共観光施設の整備と有料化の経済的影響について検証した論文が(土居2004)である。このほか，土居には熱海市の小売価格統計調査(1998)がある。

2.5　自治体財政分析

第3の地方財政危機が発生するまでは，市民・住民は，自らが住み，働く自治体の財政分析への関心は弱く，また，財政関係の業務記録･統計である『地方財政状況調査票』,『決算概況』,『決算カード』などが担当部署の非公開統計情報のような取り扱いであったため，市民，場合によっては議員でさえその所在すら知らず，ましてや，これらを用いて自治体の財政分析をするという状況には程遠かった。こうした状況に大きな変化が見られたのも，この間の地域における統計利用の大きな変化である。

藤江昌嗣は1980年代の終わりに経済のサービス化と地方自治体の歳入・歳出構造の関係の分析を東北地方の主要都市を対象に行ったが,その際,『地方財政状況調査票』,『決算概況』,『決算カード』などの業務記録(第二義統計)を用いた分析を行った。その後，1990年代に入り，バブル経済が破綻した後，国のみならず地方自治体の財政危機も進行し，市民や自治体職員が『決算概況』あるいは『決算カード』を用いて行える平易な自治体財政分析の手法を提唱した。

また，企業活動の国際化―多国籍企業としての活動―が2国間の間に引き起こす新たな国際税務問題である移転価格問題の解決が，ある国(例えば，アメリカ)での企業の追加納税(追徴)を生み出す一方で，他方において第2国(例えば，日本)での企業の申告による国税・地方税の還付という問題を惹き起こし，この還付問題が地方自治体とりわけ企業城下町と呼ばれる地方自治体に深刻な影響を与えた問題をトヨタや日産という自動車メーカーをケース(オートケース)に分析したのが藤江(1994)である。ここでも，関係する自治体の多年度にわたる「決算書」,『地方財政状況調査票』,『決算概況』,『決算カード』，議会議事録などの業務記録(第二義統計)を含む多くの統計情報等が用いられた。

その後，市町村合併や政令都市化，中核都市化など地方自治体における変化は大きく，統計情報・会計情報という点でも貸借対照表(BS)や損益計算書(PL)の作成などが自治省(現総務省)の主導で推進されてきており,「公会計制度」の変更なども新たな目標になっている。地方自治体における業務記録を中心とした統計情報の作成・公開はIT(情報通信技術)の発展や*e-Japan*戦略Ⅰ，Ⅱの国による推進もあり進んできている。しかしながら，利活用という点ではまだまだ課題は少なくない。

市町村財政分析の手法については，千波主税(1997)，大阪自治体問題研究所(2003)，大和田一紘(2005)などがある。

また，決算統計を用いた自治体財政分析には，決算統計を用いて自治体財政分析とSWOT分析を行った藤江(1997)，新潟・中越地震前の長岡市を対象にその厳しい財政状況と「都市経営」を扱った藤江(1999)，また,「平成の大合併」の先鞭の一つとなった浦和，大宮，与野3市合併を例にその「合併と財政問題」を扱った藤江(2000)などがあり，財政危機の全体構造を解明した論考には「財政危

機の構造」藤江(2002b)がある。

2.6　行政評価・評価指標の作成

我が国では，1990年代以降，アカウンタビリティ(説明責任)を高めるという目的で，多くの地方自治体が企業と同様，貸借対照表(BS)，損益計算書(PL)，キャッシュ･フロー計算書等の財務情報を作成してきており，その様はまさに“ブーム”と呼びうるものであった。その後，三重県などを皮切りに事務事業を中心とした「事業評価」も行われており，国・地方政府・公団・公社・大学・協働事業等での評価の取り組みは弥増すばかりである。

こうした情報は，質的情報と量的情報を含んでいるが,とりわけ後者は業務上の記録(第二義統計)もしくはその加工物を多く含んでいる。限られた財源の「効果的･効率的活用」を目的に，納税者である市民(taxpayer)に対する，税の「費消者」であり，サービス提供者である行政の説明能力(アカウンタビリティ)が試されているのであり，予算制約下での統計・情報の利用形態が具体的に表れてくる場面といえよう。

その予算(策定)過程(*Budgeting Cycle*)—予算策定，実行，評価(決算・監査)，戦略策定というサイクル—並びに*Managing For Results Cycle*—に重要な役割を果たしている統計・情報が各事業の予算・決算数値であり，その効率性や有効性を測定する指標である。

予算・決算数値は毎年議会において審議・承認・議決を経て公表されていくものであり，すでに存在しているものであるが，財政危機を背景に行政組織等で少しずつ導入されてきたのが行政評価(Performance Measurement)，そのための評価指標の作成である。その特徴は，地方政府内・非営利組織内で作成され，その評価が翌年度の予算策定に反映されていくというPlan-Do-Check-ActionというPDCAサイクルに組み込まれるという点である。

行政評価(Performance Measurement)には，事業評価と財政評価があるが，初期には①バランスシート，②有形固定資産明細表，③土地明細表，④普通建設事業費に関わる補助金，⑤負担金の状況，⑥主な施設の利用状況などが用いられた。

現在は，行政組織の部門別に評価指標を作成している自治体も増加しているが，PDCAサイクルにうまく組み込まれているとは必ずしも言えない。こうした評価システムの導入は，人件費を中心とする財政支出の削減が目的であるため，もし，こうした問題が解消すれば，システム自体が有効に活用されるのかどうかも分からない。これも，ブームに終わる懸念がある。この分野での会員の研究としては，藤江(2003b)，藤江(2003c)，M.J. Smith著藤江・矢代訳(2006)がある。

このように見てくると，財政分析のみならず，行政評価・パフォーマンス評価・予算評価過程への市民参加が期待されるところである。

さて，こうした評価に関しては，大学評価学会の取組みもユニークである。

橋本勝が委員長である編集委員会は2005年に『大学評価を評価する』[2)](大学評価学会年報)を刊行したが，その創刊の目的は，「大学評価は，決して一元的な価値視点から行われるべきものでないこと」(編集委員会委員長　橋本勝)，また，「狭く限定的に議論されるべきものではなく，今日の大学のあり方を根源的に問い，また，国際的な視点から大学を捉える，そのようなものとして多くの人々によって多面的に議論される必要がある」(同前)という認識の下，その議論の場を提供しようという点にある。大学評価は「定性的目標の数値的評価」(質的データの順位尺度化)という課題を含む，データ分類の基本に係わ

る問題をも提起している(池内了2005)。

2.7　公会計制度の導入

財政分析，評価に加え，新たな公会計制度の導入の動きがある。すなわち「複式簿記・発生主義会計」の導入，統一的会計基準の導入，財務諸表(貸借対照表，行政コスト計算書，キャッシュ・フロー計算書)の導入である。東京都によれば，その狙いとして，従来の公会計制度において欠けていたとされる点，すなわち①自治体経営の視点の欠落，②説明責任の弱さ，③職員の意識改革(コスト意識)を変え，この改革により「都民サービスの向上を図ること」(東京都2005)にあるとされている。

とくに，新たな公会計制度の導入においては「財務諸表の活用」にウェイトが置かれており，マクロ面での「全体の財務状況の把握」，ミクロ面での「各事業部門における事業運営の評価に対する情報を得る」というその役割と組織別や事業別の財務諸表の活用方策の検討が強調されている。

財政赤字の規模が膨らむ中で，フロー面での財政バランスとストック面での資産・負債バランスを把握し，フロー面での組織，事業バランスの把握，活用すべき資産の搾り出しと評価，公務員のコストとパフォーマンスの評価を行う仕組みづくりと会計監査(監督)を行うことに行政(公務)の役割をシフトしていくことが志向されている。

予算・決算に関する現行の財務会計は，その作成単位が評価の単位と必ずしも一致せず，また，別組織であることも少なくないため，もし，目標達成度の評価のために数値化された指標として金額を用いるとすれば，現行の公会計制度の下では，評価指標に必要なデータが得られない可能性が出てくる。かかる事態を避けるために，行政評価の導入に伴い，公会計制度を見直すことが必要となるのである。

自治体の債務報告に関してはワシントン州バンクーバー市を紹介した藤江(2000)がある。

2.8　市民・NPO・ボランティア組織(地域ガバナンスの構成員)を主体とした統計・調査

1998(H.10)年12月に特定非営利活動促進法(NPO法)が施行されてから7年あまりが経過し，2万6千以上の特定非営利法人(NPO法人)が認証される一方で，他方においては，地方分権・地方行革の推進が図られ，地域における社会システムの維持やサービスの提供あるいはまちづくり(地域開発)といった領域において業務委託を含むさまざまな協働が進んできている。こうした協働の進行においてNPOと行政，企業，地縁組織，市民との協働において，コーディネーターとしての役割が期待されている。その一つが中間支援組織で，「市民活動の発展に向けた基盤整備のため，他の団体や個人に対して，組織の運営や活動にかかわる助言や相談，人材育成などのサービスを提供したり，人材，資金，または情報などの資源提供者との仲立ちをしたり，行政や企業などとのパートナーシップの構築やコーディネイトを行ったり，あるいは市民活動全般の振興を図ったりすることを主要な目的としている組織」のことを指す。

こうした中間支援組織を含めNPO(非営利組織)，ボランティアなどによる行政サービスの提供を含め，地域において新しい統治(ガバナンス)の構成員が登場してきている。こうした新しい主体による統計の作成と利用も広がりを見せてきている。

地域におけるNPO非営利組織やボランティアを行う個人・団体，コミュニティビジ

ネス・SOHOなどの新たな担い手の登場が，地域における統計の作成や利用にどのような変化をもたらし，また統計研究者のその関わり方にどのようなニーズをもっているのかを研究することも大きな課題である。

この関連では，藤江も加わっている明治大学経営学研究所及びアーバン・コミュニティプラットホームが独自の調査を行い，協働事業が各部署が所管する事業ベースで実施されるにとどまり，組織全体としての体系的な戦略を欠いている点を確認した（藤江昌嗣2005b）。

結　論

地域における統計の作成と利用―地状学への期待―という展望の下，この間の会員を中心とした取組みを検証してきたが，稿を結ぶにあたり改めて指摘しておかなければならないのは，2つの環境の変化である。

一つは，バブル経済が弾け，長期不況が続く中，国・地方において進行している市場化テストを含む行政改革，官庁組織改革という環境変化が統計改革や統計制度改革に偏りや歪みを与えるのではないかという環境変化への懸念である。この意味では，薮内武司，橋本勝が手がけた「統計調査と統計行政における中央に対する地方統計制度・統計行政」という問題設定は基本的に有効である。統計行革における「司令塔」の役割とも絡む重要なテーマである。

また，第二の変化としては，地域における統計の作成と利用を包み込む環境としてIT環境の発展がある。野澤の指摘する通り「情報ネットワーク化と統計学，統計情報学」という視座からの現代の統計学の新しい課題とは，「統計活動の三極面―統計の調査・作成，統計体系・データベース，利用・分析―のそれぞれにおいて，情報ネットワーク時代に対応して生じた新しい変化を特徴づけ，分析すること」（野澤2002，p.8）である。地域における統計の作成と利用，特に利用という点では，この環境変化は大きな意味をもつ。

こうした中で，本稿では地域景況統計の研究，人口・労働・生活・健康・子育ての研究，費用便益分析・経済効果分析をテーマにした研究，自治体財政分析，行政評価・評価指標の作成，公会計制度の導入，市民・NPO・ボランティア組織（地域ガバナンスの構成員）を主体とした統計・調査を取り上げたが，何れにも共通するのは，地域政策に反映させるべく，市民・行政・NPOと統計研究者が協働して統計の作成・利用等における作業を行っている点である。統計の作成・利用における**主体間の協働**という点で，**コミュニケーションとしての統計学**と呼びうる状況が現出している。

地状学の地平において，統計研究者という主体の役割の重要性と協働する相手・主体である市民の意識の高さを確認して結びとするが，地域において統計研究者も統計方法という道具を携えて白足袋（研究）と地下足袋（実践・協働）を履くことが期待されている。

注

1．IT戦略会議が2002年に策定した「2005年までに世界最先端のIT国家となる」を目標に進められた『*e-Japan*戦略Ⅰ』で，これに続き，2003年7月に発表されたのが『*e-Japan*戦略Ⅱ』である。*e-Japan*戦略Ⅰから2年半で各種施策を実施し，IT基盤がほぼ整備されつつあることを受け，第2期IT戦略としてIT利活用のための方策を示している。分野として((1)医療，(2)食，(3)生活，(4)中小企業金融，(5)知，(6)就労・労働，(7)行政サービス)の推進などの7分野と，新たなIT社会基盤を整備するための方策などが盛り込まれている。

2．同書は2004年3月28日に設立された大学評価学

会の学会誌『現代社会と大学評価』でもある。

参考文献

新田俊三・川口清史(1995)「社会的経済・協同経済(シンポジウム)」『生活協同組合研究』(237) 生協総合研究所.

池内了(2005)「法人化の影ひたひたと……」大学評価学会年報編集委員会編集『大学評価を評価する』(大学評価学会年報「現代社会と大学評価」創刊号)晃洋書房.

上田和宏(1994)「岐阜アパレル産業に関する若干の考察」『社会科学研究所年報』(8) 日本福祉大学社会科学研究所.

上野健一(2004)「小地域統計による人口分析－板橋区の一事例」『経済論集』(82) 大東文化大学経済学会.

愛媛大学・愛媛県中小企業家同友会(2003)「2003年度景況調査(EDOR)」

大阪経済大学地域政策学科『フィールドワークのすすめ　アクティブ地域政策学』法律文化社.

大阪自治体問題研究所(2003)『地方財政危機と住民生活』文理閣.

大和田一紘(2005)『これならできる市町村財政分析』自治体研究社.

川口清史(1986)「構造転換期における統計・統計学の課題」『統計学　社会科学としての統計学』経済統計学会 第2集.

————(1996)「オランダの福祉，医療改革と非営利組織」『政策科学』4(1) 立命館大学政策科学会.

川口清史・宮本太郎・北島健一(1998)「(特集)協同経済」『生活協同組合研究』(270) 生協総合研究所.

川口清史・池田直樹・横山寿一(1998)「(特集)What's NPO？」『経済科学通信』(88) 基礎経済科学研究所.

川口清史(1998)「イギリスの福祉改革とボランタリー組織－コントラクトカルチャーの展開と転機」『経済論叢』161(1) 京都大学経済学会.

————(2004)「日本型NPOと社会企業」『政策科学』11(3) 立命館大学政策科学会.

菊地進(1998)「平成不況下での中小企業財務－DOR企業の経営実態」『企業環境研究年報』(3) 中小企業家同友会全国協議会企業環境研究センター.

———(2002)「DOR特別調査による企業環境の変容」『企業環境研究年報』(7) 中小企業家同友会全国協議会企業環境研究センター.

———(2004)「「新卒」の採用・育成を重視する経営－若者の就職意識の希薄化が進む中で」『企業環境研究年報』(9) 中小企業家同友会全国協議会企業環境研究センター.

木下滋(1986)「構造転換期における統計・統計学の課題」『統計学　社会科学としての統計学』経済統計学会 第2集.

———(1996)「都市間競争の論理」『阪南論集(社会科学編)』阪南大学学会.

———(1997)「産業の構造変化と都市－アメリカと日本」『阪南大学叢書』(48) 産業統計研究社.

近昭夫・岩井浩・木村和範・福島利夫編(2006)『現代の社会と統計』産業統計研究社.

近昭夫・藤江昌嗣編著(2002)『日本経済の分析と統計』北海道大学図書刊行会.

坂元慶行(2002)「日本人の考えはどう変わったか」国民性調査委員会『統計的日本人研究の半世紀』.

坂元慶行・土屋隆裕・前田忠彦(2005)「(特集)日本人の国民性調査50年」『統計数理』統計数理研究所.

田浦元(2004)「中小企業の雇用と労働についての統計的分析－「転換期における雇用・労働の変化」特別調査をもとに」『企業環境研究年報』(9) 中小企業家同友会全国協議会企業環境研究センター.

高橋桂子・保坂仁美(2003)「地方の時代における「中間支援組織」の在り方に関する予備的考察」『新潟大学教育人間科学部紀要(人文・社会科学編)』6(1) 新潟大学教育人間科学部.

田中美穂・瀧口京子「横浜市鶴見区における母子保健・2歳児育児実態調査と支援活動」経済統計学会第49回総会報告.

千波主税(1997)『今日から始める市町村財政分析(改訂版)』自治体研究社.

土居英二(1997)「政策評価システムと統計情報」『経済研究』1(3) 静岡大学人文学部.

————(1998)「熱海市小売価格統計調査について」『経済研究』3(2) 静岡大学人文学部.

土居英二・中野親徳(2001)「公立病院の地域経済効果」『経済研究』6(3) 静岡大学人文学部.

土居英二・望月勝司(2002)「自然系博物館整備の費用便益分析とCVM」『経済研究』6(4) 静岡大学人文学部.

土居英二(2004)「公共観光施設の整備と有料化の経済的影響について－熱海梅園のケーススタディ」『経済研究』9(1) 静岡大学人文学部.

東京都(2005)『東京都の新たな公会計制度』8月.

野澤正徳「統計学の現代化と情報ネットワーク」(1995)泉弘志・木下滋・藤江昌嗣・大西広・藤井

輝明編著『経済統計の現代化』所収　晃洋書房.
橋本勝・大学評価学会(2005)大学評価学会年報編集委員会編集『大学評価を評価する』(大学評価学会年報「現代社会と大学評価」創刊号) 晃洋書房.
藤江昌嗣(1994)『移転価格税制と地方税還付－トヨタ・日産の事例を中心に』中央経済社.
————(1997)「埼玉県経済動向研究会の景気分析について」『企業環境研究年報』(2) 中小企業家同友会全国協議会企業環境研究センター.
————(1997)「決算統計を用いた自治体財政分析とSWOT分析」『経営論集』44(3) 明治大学経営学研究所.
————(1999)「岐路に立つ地方財政－長岡市の「都市経営」の示すもの」『経営論集』46(2) 明治大学経営学研究所.
————(2000)「合併と財政問題－浦和，大宮，与野3市合併を例に」『経営論集』48(2) 明治大学経営学研究所.
————(2000)「自治体の「債務報告」のあり方について－オレゴン州マルトノーマ郡のTSCCの「長期債務報告」を通して」『経営論集』47(4) 明治大学経営学研究所.
————(2001)「「IT革命」と地域産業政策－川口市“SKIPシティ”構想を例に」『経営論集』48(3) 明治大学経営学研究所.
————(2002a)「アメリカ地方政府のイントラネットワークと債務・財務情報公開の現状－ワシントン州ヴァンクーバー市を例に」『経営論集』49(1) 明治大学経営学研究所.
————(2002b)「財政危機の構造」『日本経済の分析と統計』北海道大学図書刊行会.
————(2003a)'*Financial Evaluation of Multnomah County*, OR and Citizen Participation Through Budgeting in Japan',『経営論集』50(3) 明治大学経営学研究所.
————(2003b)「ポートランド市(U.S.A. OR)の事業評価・予算関連の統計・報告書について」経済統計学会47回総会報告.
藤江昌嗣(2003c)「予算過程を通じた市民参加」『ニューヨークだけがアメリカではない』梓出版社.
————(2005a)「合併・政令指定都市化と財政の変容－静岡市・さいたま市を例に」『経営論集』52(3) 明治大学経営学研究所.
————(2005b)「自治体アンケート調査結果の分析と見解」明治大学経営学研究所『地域づくりにおけるNPO・自治体の協働の事業化・制度化・システム化の現状に関する全国自治体アンケート調査』.
藤岡光夫(1999)「ミクロデータを用いた職業別死亡統計比較の方法」『経済研究』3(4) 静岡大学人文学部.
————(2000)「職業別死亡統計の日・北欧比較と年齢別死亡分析」『経済研究』4(4) 静岡大学人文学部.
————(2001)「センサスデータによる産業・職業別労働移動の推計」『経済研究』5(4) 静岡大学人文学部.
————(2002a)「健康問題・保健調査と社会統計学」『統計学』82号 経済統計学会.
————(2002b)「フィンランドと比較した日本における年齢別・職業別死亡の特徴」『経済研究』7(2) 静岡大学人文学部.
藤岡光夫・吉峯悦子(2005)「長崎原子爆弾被爆者の生活・健康調査―SPA法による社会調査と統計的研究―」経済統計学会第49回総会報告.
M.J. Smith著　藤江昌嗣・矢代隆二訳(2006)『人的サービスにおけるプログラム評価』梓出版社.
宮内泰介(2004)『自分で調べる技術　市民のための調査入門』岩波書店.
山内直人・松永佳甫・高橋智子(2004)「非営利サテライト勘定によるNPO法人の統計的把握」*ESRI discussion paper series* 内閣府経済社会総合研究所.
美添泰人(2001)「小地域統計の推定手法と応用」『経済研究』52(3) 一橋大学経済研究所.
吉田忠彦(2004)「NPO中間支援組織の類型と課題」『経営学論集』44(2) 龍谷大学経営学会.
渡辺訓子(2005)「保健活動の中の調査研究と個人情報保護」経済統計学会第49回総会報告.

コメント

菊　地　　進

記念号第4集では，本学会でこれまであまり取り上げられてこなかった領域にも研究面で踏み出すことが期待された。その一つに本章の「地域における統計の作成と利用」があ

る。むろん，これを政府統計における地域集計の利用と解せば，これまでも対象の範囲内でなかったわけではない。しかし，地域には地域独自の課題があり，全国一律の調査票では捉えられない問題が少なくない。ここに地域独自の調査が必要とされる理由がある。

かつて，地方自治体が独自に地域企業の調査を行い，地域の産業振興策を講じてきた。しかし，そうした取り組みは年々弱まってきている。その背景には，地方自治体における統計力の後退がある。政府統計は国が企画立案し，地方自治体の統計担当課が実査を担う。そのことによって効率的に政府統計が作成されてきた。そして，それに基づき都道府県別，市町村別集計がなされ，地域のデータとして公表されてきた。さらには，それらを加工することにより，独自の景気動向指数や産業連関表が作られてきた。多くの自治体でここまではかろうじて進んできた。

しかし，決定的に変化してきているのは，地域企業の動向をはじめ，継続的に独自に調査する力である。地域経済の再生は地域自らの手でといわれながら，継続的な調査がなされなくなっているのはなぜか。わが国の統計制度の中では，国が企画，地方が実査という関係のみが規定され，地方での統計調査の企画力を位置づける枠組みがもともとなかった。統計法上も何の位置づけもない。そうしたなかで，地方自治法の改正により機関委任事務が廃止され，法定受託事務として政府統計の実査事務が行われるようになってきている。すなわち，依然として事務は義務付けられているが，対応する体制は自治体に任されるということである。

こうした状況下で地方財政の制約が進むとなると，地方自治体において統計の担当者を継続的に養成する余裕がなくなり，独自の継続調査を実施することが困難になる。実施予算も制約される。さらには，アドホックな調査ですら，独自調査が困難になってくる。そうなると，政府統計の実査の力の後退へとつながることになる。

藤江の指摘するように，地方財政危機が急速に進行しており，破産が懸念される自治体すらある。高度成長期以降の公共事業中心の歳出構造を切り替えられないまま財政危機に陥っているケース，大企業の海外展開により大工場が撤退する中で地方財政構造が急速に悪化するケース，人口流出による過疎化の中で財政が悪化していくケースなど様々ある。こうしたなかで，少子高齢化の進行をも見据えながら，自治体職員の大幅削減計画が競うように策定されてきている。統計調査が法定受託事務として義務づけられているとはいえ，これが実査の質に影響を与えないとは考えられない。藤江は，「地方財政危機を背景にした地方行政改革は，地方自治体レベルでの統計作成への抑制効果を持つことが懸念される」としているが，実態は「懸念」というレベルにとどまらないほど深刻であるように思われる。

バブル崩壊後の経済社会の変化に政府統計ならびに統計制度が対応しきれなくなってきているとの認識から，今世紀に入って以降，『統計行政の新たな展開方向』（各府省統計主管部局長会議の申し合わせ，2003年3月），『政府統計の構造改革に向けて』（経済社会統計整備検討委員会報告，2005年6月），『統計制度改革検討委員会報告』（2006年6月）と矢継ぎ早に，答申・報告がまとめられ，統計法改正を射程に入れた検討が進められている。

地方統計を見るとき，こうした改革の中で，地方での実査の力の後退がどのように認識され，どのような解決方向を見出そうとしているのか，この点の検証・検討がまず第1に行われなければならないであろう。本学会会

員・森博美が委員として加わった統計制度改革検討委員会は,「統計制度改革」の狙いを,「行政のための統計」から「社会の情報基盤としての統計」へと切り替え，作成・利用双方の視点に立った情報を提供することにより,「国民経済の健全な発展・国民生活の向上に寄与」するよう統計の役割を改めることであると定式化した。まったく異論のないところである。

問題は，その位置づけを可能とする実査の体制と情報提供の体制が本当に作りうるかという点である。このことが，最も先鋭的に問われてくるのが地方統計である。藤江は，この点についてすでに先を見越して考えているのか，地方自治体レベルの統計作成への抑制効果の懸念があるとしながらも,「しかしながら，こうした厳しい状況の下でも，地域における統計の作成利用はさまざまな形，すなわち『協働』の形態をとりつつ実施されてきている」として, 地域景況統計の研究, 人口・労働・生活・健康・子育ての研究，公共投資の費用便益分析・経済効果分析，自治体財政分析，行政評価・行政指標作成の研究，公会計制度の導入の研究，地域の構成員を主体とした統計・調査などについて近年の成果をサーベイしている。これらの研究の重要性についてもまったく異論のないところである。

しかし，ここで挙げられた調査研究が進んだとしても，政府統計の実査の力の後退を克服し，統計情報提供が十分なされるような状況には直ちには結びつかないように思われる。藤江の紹介した研究を引き続き重視しながらも，政府統計とこれらの研究を結びつける力を身につけ，さらには，政府統計の実査の力を後退させないための関わり方，適切な情報提供がなされ，統計を名実ともに「社会の情報基盤」としうるための関わり方，本学会においては，これらの点にもっと関心が寄せられてよいように思われる。社会統計学とは，統計の作成からその利用まで，そしてその背景たる社会経済構造にまで思いを寄せて研究を深めようとするところにその核心がある。藤江は,「統計研究者も統計方法という道具に加えて白足袋(研究)と地下足袋(実践・協働)を履くことが期待されている」と指摘しているが，社会統計学の研究にはとりわけそのことが求められているように思われる。

統計学を作り上げてきた先達はすべて実践の学として統計学との関わりを持ってきている。このあり方はこれからも変わらないのではなかろうか。藤江の言に付け加えるならば,白足袋と地下足袋を履くには，社会経済構造についてのなにがしかの理解，これが不可欠であるという点である。統計学研究のためには社会経済構造についての理解が不可欠であり，研究を数理の及ぶ範囲内に止めてはならないというのが本学会の創立の精神であり，それは今でも間違っていないように思われる。そのことを最も考えさせられるテーマの一つが地域統計の作成と利用の問題である。

第3章　個票データと統計利用

坂　田　幸　繁

はじめに

ミクロデータをはじめとする個票データの利用・分析は，欧米に大きく水をあけられながら，日本においても『統計行政の新中・長期構想』（1995年），それを受けた90年代の文部省科学研究費補助金による特定領域研究（旧称重点領域研究）「統計情報活用のフロンティア：ミクロデータによる社会構造解析」（1996-1999年）さらにはそのフォローアッププロジェクト等を契機に広がりをみせ，世紀変わりを挟み，これまでの研究蓄積が一挙に刊行物などの形で公開，公表されつつある[1)]。とくにこのような変化を顕著に示してくれるのは，『経済財政白書』（内閣府）に示される統計分析の形態と特徴である。ちなみに2005年度版でみると，主に地域別データ，ミクロデータ，パネルデータ，個票データなどの使用を前提とした分析手法が顕著であり，50年代，60年代を席巻したマクロ集計値による計量分析のアプローチは影を潜めた観がある。もちろん，その背景にはケインズ理論・政策＝集計値分析から恒常所得仮説やトービンのq理論などをキーワードするマクロ経済理論への乗り移りがあり，期待形成を分析変数として含むミクロレベルの個体行動に基づく理論的実証的分析と経済政策の評価が必要とされていることと密接に関連している。いずれにしても，日本の分析実態は，欧米レベルの計量分析の潮流をようやく認識し，その追体験と検証作業が終わったという段階かもしれない。

このような潮流に対して，社会統計学の立場からも本格的に検討を加える時期が来ているように思われる。計量分析や統計利用の実態にはさまざまなバリエーションがみられるが，その核には個票データが直接間接利用されるか，もしくは強く意識（前提）された分析が想定されている。本章では，そのような個票データとその統計利用について考察を加え，社会統計学の課題を考えてみたい[2)]。なお，対象として取り上げる個票データの範囲は，政府統計を中心に，統計作成を目的として収集された個々の単位情報の集合とする。つまり，統計作成の経路にかからない情報，さらには個票概念が必ずしも明瞭でない業務統計の該当情報は，今回の考察の対象外とする。

1．社会統計学と調査個票

蜷川をはじめとする社会統計学の系譜において，統計利用の問題は，調査過程において獲得された統計数字の利用法をめぐって展開されてきた。ここで統計数字とは，集計計画に基づいて，実査を通じて回収された調査個票（記入済み調査票）の集計値である。調査事項は集計事項へ，調査標識は表章形式へ昇華されて統計化するため，利用者は統計作成者の調査・集計の全過程を利用上の制約条件として背負わざるを得ない。そして，利用者の社会認識が統計作成者のそれと異なる場面で

【統計作成】　　　　＜理論的過程－信頼性＞
社会科学の理論A－調査目的a－調査計画・集計計画－調査票*
（社会認識）　　　　*－実査－調査個票－分類・集計－[統計表]
＜技術的過程－正確性＞

【統計利用】
社会科学の理論B－利用目的b
（社会認識）　　非解析的－[統計表]－批判，吟味－加工・組替・編成統計表
　　　　　　　　解析的－[統計表]－批判，吟味－統計値集合

図－1　統計利用のプロセス

社会科学の理論（社会認識）－認識（分析・利用）目的
非解析的－調査計画・集計計画－調査票*
解析的－調査計画・解析計画－調査票**
*－実査－調査個票－分類・集計－統計系列
**－実査－調査個票－分類・整理－解析値集合

図－2　統計的研究のプロセス

は，統計作成の理論的過程に関する信頼性批判，技術的過程の正確性吟味が統計利用の前提をなしていた。これを形式的に図式化すれば図－1のようになる[3)]。

統計作成と利用の主体が異なり，その社会科学理論（A，B）や目的（a，b）の違いや隔たりの程度により，批判と吟味の濃度は当然異なる。利用のための基礎データが集計表に記載される統計数字に限定されているかぎり，個々の統計表は統計的概念＝集計表章概念と指標複合として形成され，統計表の全体は作成主体がもつ対象の理論的モデルに規定されてしまうからである。そのため利用者の社会科学の理論（B）が作成主体のそれ（A）と大きく異なる場合には，実際には統計利用の途は断たれ，その場合批判だけが残ることになる。それでも分析を進めようとすれば，強引な仮定に基づく推計や利用者自ら調査を企画・実施するといった方策を検討するしかなく，一般の利用者にとって多くの場合，それは実行困難であるか，正確性を犠牲にせざるをえない。

ところで，個票データとは，直接には実査を経て回収された記入済み調査個票（の集合）である。大屋に倣い，これを統計単位情報と呼べば，統計作成の過程は，単位情報が「調査目的に適合するように設計された一連の統計表の様式に従って集計され，表示あるいは表章される。」[4)] このような統計単位情報が利用可能であるとすれば，統計利用はどのように変りうるであろうか。

いまあるべき統計的研究（統計利用）のプロセスを図－2のように模式化しよう。

理論的過程はともかくとして，実査を中心とする技術的過程の遂行を考慮すると，研究の全体は一般には実行が困難なプロセスである。このとき，統計作成機関の集計表しか使えないとすれば，研究を断念するか，先の既存統計を利用する図－1の【統計利用】のプロセスに戻るしかない。これに対して，個票データの利用はもう一つの可能性を与えてくれる。

図－2の過程で獲得されるべき調査個票の統計単位情報が，既存の統計作成過程で得られる統計単位情報と同じであるか，その近似として代用できる，あるいは加工によって再

構成できるとしよう。このとき，統計作成者ではなく実際にはその利用者であるにもかかわらず，個票データの利用は，調査を含む全過程を利用者自ら企画・実施するかのような統計的研究のプロセスを実現させる。逆に，個票データがそのように機能するとすれば，研究計画，分析計画を許容できる範囲で修正することによって対応できることもあり，また計画の一部だけでも実現できるかもしれない。統計表形態での統計情報の提供と比べ，個票データの提供によって分析者は次のような追加的な利用可能性を享受できるからである。

ａ．集計過程における調査者の理論的，技術的過程の制約が弱まる。例えば，集計過程では理論規定による統計的表章概念(労働力概念など)を中心とする集計計画のもとに集計表が作成されるが，個票データの利用によってそのような理論的制約からある程度の自由度が保てる。また集計・公表上の技術面の制約から，すべての意味のある集計表が作成されるわけではなく，目的と重要性に応じて集計範囲(集計表の数やクロス表の次元数など)は限られている。この点についても，利用者は個票データを用いることで自らの分析目的に応じた多次元クロス集計を行うことができる。

ｂ．調査実施者，調査目的の如何にかかわらず，調査事項は調査票に具体化され，調査票は被調査者の回答(実査)とその後の運用を円滑に行うことができるよう設計されねばならない。そのためには，質問内容と回答項目(質問事項)は，被調査者はもちろん社会的に了解可能な形で設けられねばならない(現実適合性－技術的制約，社会的制約)[5)]。その結果，程度の差こそあれ調査者の理論的規定性は薄められる。また，調査事項をさらにいくつかの質問事項に分解して調査するなどの工夫もされる。このような制約は，利用者が想定する統計的研究プロセスでも同様に背負わざるを得ないものであり，その結果，統計単位情報として同一，同種・類似のものである可能性をもたらす。

２．個票データの類型と特徴

個票データが問題となるのは，当初個票が獲得されたときの調査目的とは異なる利用の局面(二次利用)である。調査論の視点から，個票データはその作成主体，作成過程(調査，集計)，および調査体系(方法)によってその形態と性格が大きく規定される[6)]。

⑴　統計作成主体とデータアクセス

統計は，いわゆる政府統計と非政府統計とに大別される。その区別は統計の公表性とあいまって個票データへのアクセスにかかわる制度的保証(不保証)という点で重要である。民主社会における政府統計は，その公共性を基本性格の一つとすることから，民間統計とは一線を画する。民間統計の個票データは，ビジネスベースで作成されているものを除いて，原則的に個票データへのアクセスを可能とする公共性原則はない[7)]。

政府統計については，統計法などの網がけられていることから，統計データとして一定の質を確保することが法制度上要求される。当然，それに応じて個票データレベルでも申告やその内容・形式について一定の質は確保されているとみてよい。なお政府統計の中でも，指定統計と任意統計の区分は調査標本の協力と調査事項への回答状況の点で個票データの制度的品質に影響する。

民間統計(業界団体，企業，調査･研究機関，個人)にはその種の法規定はないが，調査主体と調査客体に信頼関係が形成されている，

もしくは共通の利害関係が形成されているなどの要因から，実質的な個票データの質が確保されている場合もある。ただし，民間統計は，政府指定統計にみられるように母集団－標本設計を含めて厳密な統計調査の手続きを経るものばかりではない。その個票データの利用については，別途検討が加えられねばならない。

(2)　調査体系(方法)

調査類型により個票データとしての性格が異なり，その利用目的や利用可能性が規定される。とくに政府統計の調査体系という観点からセンサスと標本調査の区別が重要である。さらに標本調査ではセンサスに準じた内容と精度をもつ構造統計作成のための大規模サンプルサーベイ(サンプルセンサス)と経常的水準把握，速報型の標本統計が区別されねばならない。個票データには，標本抽出などの調査法上の情報も集計に必要な記録として保持されている。そのような情報も二次利用のための源泉となる。単一調査の個票データの利用という側面から特徴を整理しておこう。

全数調査という性格を持つセンサスの個票データは，調査対象の全構成要素を把握するという特長とともに，センサス目的に規定され調査項目が基本的属性や事項に絞られてしまうという短所を持つ。また調査間隔も一般に5年前後と長い。このため，あまり複雑な関係分析には適していないが，全数調査のためかなり深い多重クロスでの部分集団，小集団の大きさの確定を可能にする。またセンサス個票データセットの逐次系列は，縦断面調査としての性格をもつので，各調査時点の個体をマッチングできるとすれば，個体レベルでの生成発展消滅といった長期の動態面の統計的把握に適したデータを得ることができる[8)]。

サンプルセンサスは，特定分野に絞った構造的情報の把握にねらいがあり，そのため相対的に調査事項も多岐にわたり，分析変数も多い。変数間の関係を分析するなどの二次利用には最もニーズが高いタイプである。しかし標本調査である以上，計画精度を超えるような多重クロスや多変数の分析には限界がある。さらに，個体レベルでの完全マッチングは個人や世帯調査については，標本設計上無理があり，パネル的な利用には限界がある[9)]。しかし企業や事業所対象の調査については，抽出率の高さ(階層の一部については全数の場合もある)などから一定数のマッチングサンプルが予想され，分析上有用な情報が獲得できるものと思われる[10)]。

経常的標本調査は，月次，四半期，年次と定期的に実施され，その目的も水準変動や特定の構造情報の把握とさまざまであるが，いずれにしても，サンプル数の関係から，詳細なクロス分析などには限界があり，集計・分析変数の増加は精度の犠牲につながる。しかし，労働力調査などかなりの調査が1年間の固定サンプル方式やサンプルローテーション方式を採用しており[11)]，このような調査については，該当サンプルをリンクすることで，短期的なパネル的情報が抽出できる場合がある。また，構造把握と動向調査(速報統計)を組み合わせて実施する方式の調査[12)]もあり，これは構造特性を考慮した時系列パネル的処理の可能性を与える。

この他，不定期もしくは1回限りの特殊なテーマや時宜的な問題を捉える調査についても個票データが存在する。これらは一般に規模も小さく行政や経営上の必要から実施されたという意味で，結果利用面では汎用性に欠けるが，特定の分析関心には貴重な情報源となりうる。

⑶　作成プロセス

現在，学術面での利用に提供されている個票データは，一般に個体（人）識別が困難であるように何らかの匿名化処理が施されている。いわゆるミクロ（あるいはマイクロ）データと呼ばれるものがそれである[13]。調査はミクロデータを作成するために実施されているわけではないので，公表される正規の結果表とは異なり，ミクロデータはこれまでのところあくまで統計作成の副産物，派生品に留まっている。したがって個票データは，本来の統計目的に規定された実査から集計にいたる過程のなかにしばられた中間形態のデータとして存在する。

個票データとして情報量が最も多いのは，実査され審査終了後の個体識別子を含む調査個票の集合である。これが，前節でみた研究過程にもっとも適した有望な形態の個票データといえる。これを個票原データと呼んでおこう。本来，統計単位情報は統計目的で収集されているのであり，個票データの取得が目的ではない。むしろその利用は統計調査が成立する前提条件にかかわるため，極めて狭く限定されるべきものと考えられている[14]。個体識別を前提とした特殊な使用を除いて，このような個票原データの利用は現実的ではない。

調査個票は，集計目的に適合する形式と内容で維持・管理される。内容検査，コーディング，データチェック（論理整合性など）を経て，（集計時も含めて）必要な修正を施されるとともに，個体識別子は抹消される。このようなデータセットは，統計処理のみを想定していることから，統計的個票データともいうべきものであり，個票原データとは区別されねばならない。当然，統計的個票データから個票原データに戻ることはできない。一般に個票データの利用は，この種の統計的個票データを出発点とする。

ところで，個票原データと統計的個票データとの差は，個体識別子の有無であり，また正確性という点を除けば実質的に差は無いように思われる。指定統計に見られるようによく練られた統計調査では，両者の差はほとんどなく，区別する意味は少ない。しかし，調査の性質上複雑な構造をもつ調査票（企業・事業所統計，農家統計，調査単位と統計単位の関係が複雑な調査の多く）については，統計的個票データに個票原データのデータ構造や情報が十分汲み尽くされているかどうかの検討の余地がある。なぜなら，集計目的に沿って，それに必要なデータ構造と情報だけが少なくとも統計的個票データに組み込まれれば，調査のプロセスとして十分事は足りるからである。そのため二次利用に対して，このような意味で調査過程の理論的制約が課されてしまう。それにしても政府統計における統計的個票データへのアクセスは，いまのところ目的外利用の枠内できわめて限定された形で利用されるにすぎない。

統計的個票データに匿名化処理を徹底し，一般利用に供するものがミクロデータである。直接的な個体識別子を抹消しても，調査項目への回答特性によっては間接的に個体識別の可能性が高まる。このようなリスクを除くため，高度に数理的なものまでさまざまな匿名化手法が議論されているが，実際には，リサンプリング，地域区分範囲の広域化，トップコーディングなどの処理法が多用されている[15]。ミクロデータでは統計的個票データにさらに匿名化措置という利用上の制限が付加される。

ところで個票データの利用に際しては，方法的に以下の点のチェックが不可欠である。

ａ．本来，異なる分析目的に基づく（頭脳上の）調査法と調査票に対して，既存の統計

調査の個票原データがどの程度近いのか，どのように転用できるか。

b．実際には統計的個票データが利用できるのであるから，転用すべき情報・関係はそこから再抽出できるのか(保持されているのか，歪められているのか)。

c．ミクロデータによる匿名化処理で分析上支障がないのか，支障がある場合回避策はあるのか。

3．個票データ利用の特殊性(制約)

利用者の分析に必要な個票データが，既成の調査の個票データで代用できるとしても，利用者にとってその作成プロセスは関与できない外的なものである。集計値とは異なる個票データ固有の作成事情と性格を意識する必要がある[16]。

(1)　調査票設計の技術的，社会的制約

個票データの利用可能性は，言うまでもなく調査者の調査票設計の論理に本質的に規定される。調査目的に照らして不可欠な特性だけが調査事項に選ばれ，利用者が求める特性に関する情報がない虞はある。さらにこのような理論的制約に加えて，調査事項が利用者の必要とするものと同じか，もしくは類似していても，調査票設計の技術的，社会的制約から生じる問題もある。

例えば，質問は一般に選択肢法によるものが多い。個人や世帯を対象とするのか，企業や事業所を対象とするのか，調査対象によっても異なるが，本来計数，計量ベースで測定可能な場合でも，しばしば階級区分による回答肢が使用される。利用者のあるべき調査票においても選択肢法になるかもしれず，選択肢法の採用が問題なのではない。問題は利用者の階級区分と個票データのそれとが同じである保証がない点である。異なる場合には何らかの処理や工夫が必要となる。

同様にして特性がそもそも質的である場合でも，用意されたカテゴリー区分が利用者のそれと一致しない場合が多い。また利用者には複数の標識(質問項目)の組合せで捉えたい事象が，個票データでは単一の質問項目で集計・分類表章に直結する複合標識型の選択肢で調査される場合もある。また調査票の質問法によっては，特定標識をもつ集団だけに付随質問という形式で調査が実施されることがある。付随質問に利用者の関心がある場合，当該集団以外についての情報は得られない。このような調査票設計にかかわる問題については，分析計画の変更や修正を含めて，個票データに特有な処理を要求する[17]。

(2)　調査無回答および項目無回答(あるいは欠測値)

全数調査であれ，標本調査であれ，実査の場面では，本来調査対象となるべきものの協力が得られない事態が生ずる。調査環境の悪化はこのような無回答サンプルの増加を招いている。無回答サンプルが発生した場合どのように処理されたのか，例えば，再調査，あるいは追加サンプルといった処理がなされたのか，利用者にはそれらの情報が個票データとともに提供されねばならない。利用者は無回答によって分析にバイアスが生じないか検討する必要がある。多数の標識によるクロス集計など特定の小集団の構造把握を目的とする場合もあり，無回答との相関が結果にバイアスをもたらす可能性があるからである。また，無回答がかなりの数にのぼるときには，確率的標本抽出とは異なるサンプル評価の基準から分析対象について代表性などの検討も考えざるを得ない。

さらにサンプルの中には，調査項目の一部について回答が得られていないものも存在し，また回答内容に論理的に矛盾したものが含まれるケースもある。本来ならば，これらは利用者の目的にとって分析に偏りをもたらさないように処理すべき性格のものである。しかし，これらについては，統計作成の過程ですでに特定の値が補完されていたり，修正や不詳といった処理が施されている。このような処理後の個票データが利用者には提供されることから，利用者側でそれらの処理法についての選択の余地はない。この点については，むしろ，このような処理の規則，補完の方法などの情報が開示され，利用者側でその吟味を行えるようになることが望ましい。

⑶　外れ値

集計値を利用する限りはさほど問題とならないが，個票データでは必ず外れ値の存在が分析結果に大きく影響する。外れ値は測定や記録上のミスや入力ミスなどに起因する場合もあれば，異質な単位が混入している，あるいはそうでなく正しいにもかかわらず極端に大きな値をとるケースなどがある。前者の場合には修正が必要であり，後者であれば解析上結果に大きく影響を与える場合があるので，注意が必要となる。

各変数の値ごとの外れ値，あるいは異常値については，実際には集計時にすでにチェック済みであると考えられ，また集計計画にかかわる変数間の関係からみた外れ値についてもほぼチェックが終わっていると予想される。したがってその点では，個票データの利用者が改めて判断を下し，処理する機会はほとんどない。しかし，利用目的によっては，統計作成時のチェックにかからない変数間で外れ値が新たに問題となることもあり，利用者はその事前点検を含めて，分析を進めねばならない。

⑷　標本抽出法

全数調査であるいくつかのセンサスを除き，大部分は調査対象（目標母集団）の一部（調査単位）を抽出し調査される。抽出方法には，確率的な無作為抽出と有意抽出法があるが，二次利用にあたって，抽出に関する情報の開示と，利用可能な抽出情報が個票データに記載されていることが前提となる。調査対象（特に個人・世帯と企業・事業所）や目的によって調査方法や標本設計も異なる。さらに，現実の調査では予算，人員，組織など技術面の制約の下で調査精度を追及する。標本調査の理論面では，層化，集落抽出，比例抽出，層化多段抽出など確率数理の方法規定を論ずればよいが，現実は複数の関心のある対象範囲，集団レベルで特定の推計精度を確保するように標本調査が具体化される。

当然，このような個票データの集合に対しては，無限母集団からの無作為抽出を想定した教科書的な推定や分析では，推定値に偏りをもたらす。そのため標本設計を考慮した分析と標本誤差の評価が不可欠となる。

また層化基準や抽出クラスターの定義など標本設計にもよるが，分析目的によっては，抽出方法，抽出単位レベルの操作が必要となることも想定される。例えば，分析に必要な推定対象の区分や範囲（地域）を事後的に変更するために，他の情報を用いて抽出クラスターを再抽出（リサンプリング）することなど考えられる。加えて，複数の調査個票データのリンケージが必要な場合には，調査区コード，世帯番号，事業所番号など抽出情報を用いてマッチング作業を実行しなければならない。

(5)　ミクロデータ

個票原データや統計的個票データはすでに述べたように集計までのプロセスで生成されるファイルであり，それが利用できるのであれば，上記のような個票データの特殊性を考慮（逆用，利用）して分析を計画することが望ましい。しかし実際には，これらの多くは詳細な情報が開示されているわけではなく，当該統計の作成当事者以外の者は，このようなタイプの個票データの特性を利用することができない。実際この種の個票データに関しては利用者も狭く限定されており，これまでのところ，指摘の多くは作成機関の内部的利用のための留意点というにとどまる。これに対して利用者の制限を緩め目的外利用を実現する方法のひとつが，すでに述べたミクロデータであり，個体識別を困難にする秘匿処理が施されている。

しかしミクロデータにおいては，匿名化処理の技術的制約だけでなく，実は上記項目(2)から(4)に関わる問題は回避されており，その特徴を考慮した分析は多くの場合不可能である。それが，一般利用というミクロデータの目的に適合したデータセットとして作成されるためである。ミクロデータの基本型というものがあるとすれば，それは，調査無回答に関する処理の情報は含まれない（回答集団だけのデータである），調査の基本となる調査事項に対する無回答サンプルは含まれない，調査目的に対して基本的な外れ値，異常値は処理済のデータであり，多くは個票データからのリサンプリングデータであり復元乗率を用いた集計・解析ですむようなデータセットである。言い換えれば，ミクロデータでは現実の調査過程の痕跡は消去されている。

統計的個票データを基礎にする以上，ミクロデータにおいて無回答や外れ値の問題が解決したわけではない[18]。また標本抽出情報などの活用についても，ミクロデータでは匿名化のため一般には限定（消去）されている。しかし他方で，ミクロデータのデータ処理は単純で，容易である。限界はあるものの，このような長所を活かした分析にミクロデータの意義がある。

4．個票データの利用可能性

社会的に重要な基本情報は主に政府統計に依拠せざるを得ず，それは利用者の目的に応じて企画・調査されているわけではないから，個票データの利用には集計値利用に共通する制約がある。しかし，調査者の集計計画に基づくレディメードの集計表（値）だけに依存してきたこれまでの分析限界を考えると，利用者が享受する利益は大きい。以下に，その代表的な利用形態を整理しておこう。

(1)　集計表にない集団構成・分布の把握（再集計）

推定精度の問題を度外視すれば，個票データセットに盛り込まれた調査事項（質問項目）の範囲内で，集団を細分化し，その大きさや特性，さらには分布を比較することができる。それは，既存集計表の下位集団である場合（詳細集計）もあれば，それとは異なる分類標識からの集団の再構成であることもある[19]。とくに地域区分の自由度は，小地域集計を含めて，地域統計分析に貢献する[20]。また，調査によっては，現在の状態とともに過去の状態についての質問を設けている調査票がある。このような個票データについては，現在と過去の状態とをクロスしたいわば動態類型により，静態的集団構成だけでなく，動態的集団構成を分析できる場合がある。なお，調査者とは異なる理論に基づく概念規定による集団構成の把握についても，必要な調査事項と分

析標識が個票データとして提供されているならば，統計表の組替え推計ではなく再集計の論理で処理することになる。

⑵　利用者によるミクロ計量分析（統計解析）

集計値によるマクロデータ分析に対して，個票データではn個のサンプルに基づく個体（ミクロ）ベースの統計解析を適用できる。集計値分析が個体レベルの均質性（同質性）を仮定としていたとすれば，個票データ分析は個体レベルでの異質性を前提として分析にとりこむことができる。また現在のマクロ経済理論が企業や家計の個別主体のミクロ的最適化行動を基礎におくことから，計量経済分析やそれに基づく政策評価分析には，不可欠なデータリソースである。分析モデルに使用できる変数は，個票（調査票）の調査事項の質問項目に基本は限られ，標本設計も単純無作為抽出というケースは少ない。また非標本誤差など現実の調査過程では固有の処理が施されており，解析上注意すべき点も含まれている。何より政府統計は因果的研究を目的に作成されているわけではない。しかし，個票データの提供により，他の情報を利用して追加的な変数を導入したり，次項で述べるデータリンケージを行うなど，利用者の理論や分析目的に応じたデータセットを作成し，自らミクロ計量分析を実行できる可能性は高まる。いくつか利用の特徴点を指摘しておこう[21)]。

ａ．集団の異質性を前提とした解析である。男女別など属性や要因によって関係が異なることを想定したモデルが多用される。そのため多くのダミー変数などがモデルに組み込まれる。なお多重クロス表により，複数の要因を固定して特定の関心のある変数の特性（規則性や傾向）を比較，検討する場合にも，回帰などのこのような解析的アプローチが有効な場合が多い。この場合は，いわば要約的，記述的な利用と位置づけられる[22)]。

ｂ．個体ベースであるため目的変数が質的であるケースが多く，ロジット，プロビットモデルの使用が不可欠である。また，計数データも多く，ポアソンタイプのカウントデータモデルが用いられる場合もある。

ｃ．目的変数が質的側面と量的側面を併せもつ場合がある。株式保有高など，保有，非保有（0）があり，保有の場合にその保有高が観測されるというケースである。この場合，切断分布モデルやトービットモデルが利用される。

ｄ．目的変数（賃金）がある特定の標本（就業者）だけしか観測されない場合がある。他の標本を無視するか，あるいは標本の選別プロセスを考慮する（標本選別モデル）などの工夫が必要となる。

ｅ．個票データ分析の異時点間の結果系列は，比較静学の観点から集団的構造の変化を分析できる。しかし，個票レベルで時系列リンクしたパネルデータについては，個体レベルでの動的変化を利用したモデル分析が可能となる[23)]。

⑶　データリンケージ（新規調査　パネル調査の代用）

異なる統計調査間での個票データの結合，および同一調査の個票データの異時点間の結合は，統計利用の大きな可能性を与える。事実上前者は，関心のある諸変数間の関係を把握する新規調査を実施したことと同種の効果を与えるとともに，後者については，パネル化までに要する時間を待たず，個票データを遡及的に結合することでパネルデータを構築できる。その結果，既存の調査から新しい追加的な情報セットが構築され，分析に使用できることになる。

データリンケージには，個体識別子による完全照合の方法と，いくつかの変数セット(キー変数)の一致に基づく統計的マッチングの手法がある[24)]。統計的マッチングの手法については理論的にはともかく現実のデータに適用するにはまだ検討の余地があり，実際的な方法としては完全照合による結合に頼らざるを得ない。しかし，政府統計において，仮に個体識別子情報が維持・管理されているとして，そのようなリンクが可能な場面は限られている。全数調査と全数調査(一部階層についての全数も含む)，標本調査とその抽出フレームとなるセンサス型調査，同一の抽出フレームをもつ標本調査同士，(標本交代方式の)固定サンプル型標本調査同士などである。標本設計にもよるが，標本調査において2つの調査に同時に含まれる標本数は限られることから，実際にリンケージが有効なのは，センサスやサンプルセンサスなどの大標本を前提とする。

個体識別子の管理問題を考えると，このような作業は現状では統計作成機関内部で行われるほかなく[25)]，また秘匿性の観点からも，一般の利用者が個票データを完全照合する状況は考えられない。しかしその統計情報としての重要性を考えると，個票データ公開の促進とともに，リンケージデータの作成と利用を可能にする制度と仕組みが早急に実現される必要がある。

むすび－個票データが提起する課題

「統計利用者のための統計学」を標榜してきた本学会では，統計生産者と利用者との分離を前提に，政府が最大の統計生産者であり利用者であるという状況のなかで，統計調査の過程を明らかにし，統計の性格を批判・吟味し，統計利用の実際や統計解析の手法を批判的に検討し，統計利用の道を示すことに多くの努力が積み重ねられてきた。しかし個票データの利用は，これまでの本学会での統計利用の議論の枠組みには想定されていない新しい問題状況である。もちろん，個票データが統計作成過程のなかで生まれる中間産物である以上，これまでの本会の研究成果は個票データの利用にも妥当し，それを援用することもできる。しかし，統計(＝集計値)の利用とは本質的に異なり，個票データの利用者には利用面でのより大きな自由度が与えられている。ここに二次利用の特質があり，個票データ利用の良さがある。反面，それを十全に活かすために，個票データの生成過程を明らかにし，個票データの性格を批判・吟味するという形で，従来，批判者の立場から正確性を論じてきたものを利用主体として自ら遂行しなければならない。本稿はそのための論点提起の試みである。

ところでこのような利用者＝作成者となる関係では，統計的研究方法論の立場，いわゆる社会科学方法論説が改めて見直されねばならない。個票データの利用はいわば統計利用の限界的形態であり，そこでは対象反映性が重要な評価の基準となり得る。そもそも集計値分析からミクロデータの利用やパネルデータ志向への方法論の拡大にはこのような対象反映性が大きな契機となっていたと思える。個票データの解析モデルに関する数理と事物論理の明確化(有効性とその限界)，有限母集団への確率モデル適用の是非とその解釈，パネルデータの現実反映性－非反映性(内海科学方法論との対比[26)])など，検討すべき素材は現実化している。

注

1．特定領域研究の成果は，松田(1999)や松田他(2000，2003)の『講座ミクロ統計分析』(1-4巻)としてまとめられている。内閣府経済社会研究所をはじめ，労働政策研究・研修機構や経済産業研究所など政府系の研究機関を中心に，かなりの数の個票データを用いた分析が進められており，ワーキングペーパーなどの形で研究成果が公表されている。またその成果を踏まえた清水谷(2005)などもある。政府内部や独立行政法人における利用は活発化し常態化しつつあるように見受けられる。

2．本学会におけるミクロデータ利用をめぐる議論としては，森(2004)，吉田(2004)があり，本稿の問題意識と密接にかかわっている。なお法政大学日本統計研究所『研究所報』ではミクロデータ利用に関する本学会員を中心とした研究成果が精力的に刊行されている。例えば，『研究所報』No. 25(1999)では「ミクロ統計データの現状と展望」，No. 32(2004)では「ミクロデータとその利用」，No. 34(2005)では「政府統計の二次的利用」について特集テーマが設定されている。

3．本学会には統計利用をめぐる豊富な議論の蓄積があるが，以下の整理に当たっては蜷川(1932)，大橋・野村(1980)，大屋(1995)に拠るところが大きい。

4．大屋(1995)，p.78。

5．調査票設計に関する論理については，大屋(1995)を参照されたい。

6．政府統計の作成と利用に関する詳細は，例えば工藤(1997)参照。

7．二次利用の促進の観点から，民間統計を含む個票データのためのデーアーカイブの構築が進んでいる。詳細は，佐藤・石田・池田編(2000)，前田幸男(2004)参照。

8．イギリスのLSデータ(longitudinal study data)など。森(2004)参照。

9．統計的マッチング，あるいはサンプル移動が少ない小集団に分割できれば擬似パネルなどの可能性は残されている。

10．事業所や企業を対象とする場合，大規模な事業所では極端に分布が歪んでおり，全数的把握を必要とすることが多い。また動態調査においてもそのため特殊な抽出法が用いられる。

11．例えば，毎月勤労統計調査(厚生労働省)では3年，法人企業統計調査(財務省)では1年となる。

12．例えば，個人企業経済調査(総務省)では，1年の固定サンプルで動向調査票(年4回)と構造調査票(年1回)に分けられ実施される。

13．近年，一橋大学経済研究所社会科学統計情報研究センターにおいて，秘匿処理を施したリサンプリング形式で，学術研究のために政府統計ミクロデータを提供するシステムが試行的に運用され始めた。

14．「個体識別子を含め個々の申告内容が明らかとなる調査個票の使用は，秘密の保護や調査対象者の信頼の確保という統計法制度がこれまで保護してきた利益にかんがみれば，公益性が高いと認められる場合に限って例外的に認められるべきものと考える。」(総務省「統計法制度に関する研究会報告書－中間とりまとめ」2005年12月，p.21)

15．統計調査制度とミクロ統計の開示については松田・濱砂・森(2000)参照。

16．松田・伴・美添(2000)においてはミクロ統計固有の集計解析にかかわる問題が詳細に議論されている。

17．労働力調査や就業構造基本調査におけるミクロデータ利用の問題について，例えば坂田(2005)を参照されたい。

18．指定統計についてはそのような制約は分析に大きな歪みを与えないと想定できるが，任意統計については深刻である。

19．例えば，日本統計研究所『研究所報』No. 34(2005)では，この種の再集計に基礎を置くミクロデータ利用がみられる。

20．地域分析へのミクロ的利用の詳細と分析事例については，松田・垂水・近藤(2000)参照。

21．ミクロデータの解析モデルについては，制限従属変数モデルを中心に山下(2004)による整理がある。

22．説明変数を固定して，言い換えれば説明変数のセットに意味があり，その目的変数への効果に関心がある場合。例えば，多重クロスなどの再集計により細分化したグループ間の特性値から傾向を読み取りたい場合がある。しかし，グループ数やそのサンプルサイズの差による確率誤差などの要因から，一般に傾向把握が困難な場合が多い。使用した変数の数やその標識数に比例して困難は増していく。このとき，細分化に用いた質的変数についてはダミー変数などを用い解析的に処理する

ことは傾向把握の助けとなる。

23. パネルデータ分析の詳細については北村(2005)参照。

24. 統計的マッチングについては松田・伴・美添(2000)参照。

25. すでに試行的にそのようなリンケージはかなり進められており，研究成果も公表されている。最近では，早見(2005)において「事業所・企業統計調査」，「商業統計調査」，「賃金構造基本調査」のリンケージが行われている。

26. 内海(1962)参照。

参考文献

内海庫一郎(1962)『科学方法論の一般規定からみた社会統計方法論の基本的諸問題』(限定出版・非売品).

大橋隆憲・野村良樹(1980)『統計学総論(新訂版)』有信堂.

大屋祐雪(1995)『統計情報論』九州大学出版会.

工藤弘安(1997)『入門統計学―官庁統計の作成と利用―』全国統計協会連合会.

北村行伸(2005)『パネルデータ分析』岩波書店.

坂田幸繁(2005)「就業構造の変容と労働統計ミクロデータ」『中央大学経済研究所年報』第35号.

佐藤博樹・石田浩・池田謙一編(2000)『社会調査の公開データ　2次分析への招待』東京大学出版会.

清水谷諭(2005)『期待と不確実性の経済学―デフレ経済のミクロ実証分析―』日本評論社.

蜷川虎三(1932)『統計利用に於ける基本問題』岩波書店.

早見均(2005)「リンクしたマイクロデータによるパネルの作成」『リサーチペーパー』総務省統計研修所　第1号.

前田幸男(2004)「世論調査データの行方―データ・アーカイブの役割―」中央調査報(No. 558).

松田芳郎編(1999)『統計情報活用のフロンティアの拡大：ミクロデータによる社会構造解析』財団法人統計情報研究開発センター.

松田芳郎・濱砂敬郎・森博美(2000)『講座ミクロ統計分析①』日本評論社.

松田芳郎・伴金美・美添泰人(2000)『講座ミクロ統計分析②』日本評論社.

松田芳郎・垂水共之・近藤健文(2000)『講座ミクロ統計分析③』日本評論社.

松田芳郎・清水雅彦・船岡史雄(2003)『講座ミクロ統計分析④』日本評論社.

森博美(2004)「ミクロデータの利用特性と統計利用論」『研究所報』法政大学日本統計研究所 No. 32.

山下裕歩(2004)「ミクロデータのモデル分析」『研究所報』法政大学日本統計研究所 No. 32.

吉田忠(2004)「統計利用論からみたミクロデータ」『研究所報』法政大学日本統計研究所 No. 32.

コメント

岩　井　　　浩

1　歴的経緯

記念号第3集(1995年)の「情報化と統計」(金子)で，ミクロデータの提供と利用の検討の必要性が言及された。その後の10年の経緯において，この課題をめぐる状況は一定の前進がみられ，本記念号の坂田論文では，社会統計学の統計の信頼性と正確性の視点から統計作成の一過程の産物である個票データとその統計利用について，理論的方法論に検討が加えられるに至った。その歴史的理論的意義はおおきいと言えよう。

諸外国における個票データ(ミクロデータ)の提供と利用が大きな蓄積と広がりを見せているのと対比して，わが国のそれは限定されたものに止まっている。しかし各界の個票データ利用の要請の広がりの中で，政府の統計改革をめぐる論議，『統計行政の新中・長期構想』(1995年)，最近の統計制度改革委員会の「中間整理案」(2005年12月)等において，個票データの提供と利用が政策的課題とされている。統計利用者の側からも積極的な調査研究と問題提起(1996年「統計情報活

用のフロンティアの拡大」，『講座ミクロ統計分析』）がなされ，その成果として2000年〜2004年「ミクロ統計データ活用研究会」（統計情報開発研究センター），2005年以降，一橋大学経済研究所社会科学研究センター「政府ミクロ統計の提供の試験的システム」が運用されおり，個票データの目的外使用によるリサンプリングデータの提供・利用が限定的だが継続している。先の「中間整理案」に対して，ミクロデータの提供と利用の積極的推進の意見が各界から要請されている。

２　社会統計学と個票データ利用

個票データ（ミクロデータ）の提供と利用の広がりは，これらのデータの作成と利用の諸過程と諸形態の考察を課題とさせている。以下，統計の作成過程と個票データの作成・利用に係わる論点にコメントを限定する。

個票データは，統計調査の企画（調査目的，集計計画）に規定された調査結果としての調査個票の内容検査，データチェック等を経た集計のための統計記録であり，最終集計の統計表作成にいたる過程で中間的に作成されデータである。特に匿名化処理された個票データがミクロデータとされている。ミクロデータ利用によって，統計調査の枠組み（調査項目とその構成）に規定されて集計された既存の統計表（一定の既存の標識による分類･クロス統計表）の限界に対して，個票データの再集計によって標識選択による自由な分類・組み合わせが可能になり，制約はあるが，統計対象の新たな枠組みと構成を明らかにできるようになった。この意味で，個票データの作成がいかなる対象の反映なのか，一定の制約における個票データの利用が社会的現実のいかなる分析・利用でありうるのか。歴的社会的に規定された「大量」（社会的集団）を調査者の目的，理論と社会的技術的制約で，「大量の四要素」から調査可能な「大量観察の四要素」に具体的に規定（調査票の設計）し，調査票の運用による調査の実査，調査結果の集計の諸過程を個票データの作成・利用の諸関係において，社会統計学の統計の信頼性と正確性の視点から考察する課題が提起されている。これらの考察・分析には，統計調査の諸過程，特に調査結果の調査個票が，集計され，統計表として作成される具体的なプロセスの考察が対象となる。

これらの視点からの最初の問題提起は吉田によってなされた。統計の信頼性・正確性の視点から，統計調査の諸過程，特に調査結果の点検・集計・公表過程において作成されたミクロデータの対象反映性と利用可能性について，統計利用論からの独自な検討の必要性を提起された。森は，大屋の統計単位情報を基礎に，個別識別情報を含めた調査個票の情報を統計原単位情報，統計に編成される個体情報を統計単位情報と区別し，統計単位情報とミクロデータの関係を検討し，ミクロデータで利用が，「統計作成・提供の基本的在り方とされてきた集計表によるデータ提供と統計利用面でどのような違いがあるか」を考察した。統計の二次的利用としてのミクロデータ利用が，既存の集計表からは得られなかった「数多くの新たなfact findingsが得られる」ことを明らかにしている。

坂田は，さらに社会統計学の信頼性，正確性の視点から，統計作成の諸過程における個票データの作成と利用に焦点を当てて考察を加えている。⑴個票データの類型と特徴を個票データの作成主体と作成過程（調査，集計）に即して，統計調査の主体とデータアクセス，調査体系（方法），作成プロセスの三つの視角から考察し，その特性を分析する。個票データの匿名化処理されたものがミクロデータ（マイクロデータ）であり，ミクロデータは統

計作成の中間物，派生品であると規定する。調査結果の記入済み調査個票の集合(大屋の統計単位情報)を，森の規定に従い個票原データと統計的個票データに区分し，後者は調査個票から，内容検査，コーデイング，データチェック，必要な修正を経て，匿名化処理されたものであると規定する。⑵また個票データ利用の特殊性(制約)として，「統計集計値とは異なる個票データの固有の作成事情と性格」を視点に入れて，①調査票設計の技術的，社会的制約，②無回答(あるいは欠損値)，③外れ値，④標本抽出の諸論点を考察し，目的外利用としての調査個票から匿名化処理されたミクロデータでは，⑵～⑷に係わる項目は回避されており，それの特徴を考慮した分析はできないとされる。「ミクロデータでは現実の調査過程の痕跡は消去されている」ことが指摘される。

3　統計作成過程とデータ・エディティング

坂田は，無回答や外れ値等の処置を論じているが，「個票原データと統計的個票データの差は個体識別子の有無，正確性の視点をのぞけば，両者の差はほとんどない」と規定している。調査済み個票から集計目的のための統計記録(統計単位情報)が作成される過程は統計作成者によって人力また一定の数式的に処理される内容検査，データ・チェックの過程であり，統計調査の技術的過程，統計の正確性の視点に係わるとされる。しかし例えばこの過程にはいわゆるデータ・エディティングの問題がある(国連『データ・エディティングに関する用語集(対訳)」http://www.nstac.go.jp/services/pdf/skk-yogosyu1.pdf，参照)。例えばこの用語集にある「14歳以下であるが，結婚している」などの調査個票内の項目の無効・矛盾について，データ・エディティングやimputationを施すことは，正確性の視点からのみで処理できない論点が含まれている。上記の矛盾項目も，法律婚としては14歳以下の有配偶はありえないが，事実婚ならありえるケースである。法律婚を前提にしているのも，一定の社会認識を前提しているといえる。このように調査結果の集計過程の諸論点も，単に技術的視点に止まらず，理論的視点が内在しているといえる。集計過程が統計利用論の一過程になってきたのがミクロデータ利用の特徴であるが，それによって「集計過程は，統計調査論に入るのか，統計利用論に入るのか」，あるいは，より根本的に「調査論と利用論を時間的流れの中で二分できるのか，できないとすれば，調査論と利用論を区別するものは何なのか」という基本問題を派生していると思われる。これらの視点からの個票データの提供・利用と社会統計学と諸論点の考察も，研究課題の一つであると言えよう。

(本文中に言及した文献は，坂田論文の文末の参考文献，参照。)

第４章　民間企業におけるデータの蓄積と利用
――マーケティングリサーチ，データマイニング，統計――

池　田　　伸

はじめに

本稿では，近年の経営実務における統計的方法の適用に係わり，とくに新しい手法であるデータマイニングをとりあげる。従来，政府作成・公表統計が社会統計学の前提とされていたが，経済のデジタル化の著しい進展によって民間企業（おもに公開会社）が自社業務にまつわる記録をほとんど網羅的にデータベースDB化することが行われるようになり，統計に類似した現象が政府の制度外で広範に見られるようになった。さらに，そのようなDBに対し，従来の統計的方法よりもふさわしいと思われる解析方法とされるデータマイニングが勃興し経営実務への導入が始まってきた。

ここでは，おもにマーケティングに関連した顧客との取引記録に係わるデータのDB化およびそれに対するデータマイニングの考え方と実践とについて，従来のマーケティングリサーチや統計的方法と比較対照しながら考察する。それによって，データマイニングをめぐる統計現象について経営統計学・社会統計学上の位置づけを試みる。

1．経営におけるデータの獲得・蓄積

1.1　OLTPによる業務記録の蓄積

近時デジタル化の飛躍的な進展にともない，企業等組織における情報処理の能力が飛躍的に向上している。ITを利用した経営情報処理が企業の競争優位の源泉となってきている。たとえば，大規模小売店では商品マスターのデータベースDBと照合された1件1品SKU[1)]毎のアイテムが最小単位として識別され，仕入からの全取引がトラッキング可能となる。このシステムの出口を担うPOSは販売の商品情報と顧客情報とをマスターと照合しつつ業務用の基幹系DBに反映する。このようなサイクルでDB管理システムに支援された単品の管理OLTPが行われ，業務の合理化や品揃えに貢献している。

他方，OLTPは記録として蓄積される。従来，これらの業務記録は日次や月次決算のための帳簿システムへの入力とされているが，会計情報としては基本的に勘定・科目毎に集計値として表章されることになる。また，OLTPの記録は，証憑という点を除けば消去（揮発）してもかまわない。むしろ，ディスクスペースが貴重であれば，無駄な記録を止めておく理由はない。

ところが，前世紀末から情報処理のハードウエア資源が安価になり，TCP/IPによるネットワーク化が促進され，システム的にも並列処理が行われるようになると，情報処理の資源が急速に安価かつ大規模に利用可能となった。このことは当然OLTPを含めた基幹系の経営情報システムそのものを進化させることとなり，ますます大量の記録がとどまるよう

になった。また，近年web上の閲覧のログやクリックストリームもOLTP同様に情報源とみなされるようになっている。膨大になるこれらデジタル記録を処理することによって経営上有益な情報が得られるのではないか，と大いに期待が高まった。

しかし，そのためには二つの大きな課題が残されていた。すなわち，大規模になるデータにふさわしい蓄積と解析とについての各方法論が要求された。

1.2　データウエアハウスによるDB管理

一つめの課題であるデータ蓄積の方法論は，OLTPの記録をデータとしていかにして利用可能とするかであった。すでに基幹系の情報システムにはOLTPの記録がTB単位で蓄積されているが，業務用DBのままではデータ解析に利用することはできない。迅速･確実･非公開に取引を処理すべき基幹系DBの上で，対照的な性格のデータ解析を行うことは望ましくなく，またOLTPの生データのままでは解析に耐えうるデータではない(図－1の左側。ただし，本図は一つのモデルである)。

そこで企業の情報システムにおいて，基幹系とは別に解析用DBをデータウエアハウスDWHとして構築するというアイディアが生まれてきた。このシステムを提唱したインモン(1999)によると，DWHの基本アーキテクチュアとして，1)サブジェクト型データの格納，2)統合・単一化，3)時間的変化が追跡可能，4)不揮発性，5)詳細データにサマリーを付加したデータの形態，の特徴があげられている。つまりDWHでは，業務記録であるOLTPを揮発させることなく蓄積し，あまり処理は経ないが解析に耐えうるようには洗浄され正規化されたデータとして，メタデータが整備され統合されたDBに格納されることになる。さらに，実際の解析にあたっては，研究目的に応じてDWHから抽出されたサブセットのDBであるデータマートを直接の対象とする(図－1のなかほど)。

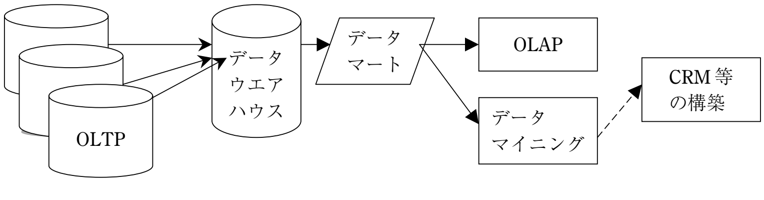
そもそもDWHに入るデータは，顧客IDが付されたものも(たとえば，クレジットカードでの支払や通信販売)，そうでないものも(POSによるSKUの管理)ある。顧客の購買等の情報はまずこのいずれかでデータとして記録される。これらに外部で購入された種々の情報が付加されて利用されることもある。

2．経営における大規模DBの利用

2.1　OLAPによるリポーティング

DWHが構築されると，デーマートのDBを端末から操作することで，これまでは，「オフライン」で専門的に扱われていたさまざまな情報が，エンドユーザであるマネジャー自身によって文字通り「オンライン」で入手することができる[2)]。蓄積されたデータに対し，このような利用・解析法をOLAPという。

たとえば，マーケティング担当者は日付-商品-店舗等からなる多次元のデータ(3次元



図－1　データマイニングにおけるワークフローの概要

以上も含め「キューブ」と称される）の様子を瞬時に知ることができる。さらに関心に応じて，データを対話的に処理するクエリを送ることができる。すなわち，データの断面を変えて再集計したり一部だけを呼び出したり（キューブのスライス＆ダイス），より詳細を掘り下げたり（ドリル）するような指示や，その結果を視覚化することができる。これを従来の非同期的オフラインの固定された製表によるリポートとは大きく異なる優位点を持つ（ベリー・リノフ(2002)，鈴木(2003)などを参照）。

2.2　データマイニングによる解析

OLAPは以前には非常な投資を伴うシステムであったものの，原理的にはクエリによる集計のような通常のDB処理であった。OLAPのためだけにDWHへの投資は正当化されず，オフラインにおいてより集中的・本格的にDWHのデータを解析するための手法が開発されねばならなかった。大規模DBの解析においては，データの「汚染」は別にして，通常の統計処理における小標本のようにデータを節約する必要はなかったが，見通しが悪いデータの中を組合せによる爆発を避けながら新たな発見に至ることができるようなアルゴリズムが求められたからである。

この課題と可能性とを一般化していうと，大規模DBからの知識発見であり，データマイニングとはKDDのための手法の総称である。KDDは必ずしも経営データのみに適用されるものではないし，データマイニングの諸手法は経営分野で発達したものではない。また，経営実務においても種々の部門で用いられている。しかし，本稿で対象とするようにマーケティング関連の分野は，データマイニングの一つの主要適用分野といえる。（以上，図－1の右側）。

2.3　CRMへの実装

これらKDDを競争優位に寄与させるためには，得られた知識を何らかの政策に反映されねばならない。DBから発見されたパターンから，たとえば電話会社の「乗り換え」(churn)防止や併売（クロス/アップセリング）に役立てられて十分な効果を上げれば，KDDへの少なからぬ投資に対するリターンを生み出したことになる。

データマイニングによって得られた知見は，CRMとよばれる経営手法に活用される。顧客データのRFM解析などによって，市場や顧客を発見的にセグメンテーションすることができ，セグメント毎に柔軟に顧客に対応することが可能となる。一般に，顧客をあるモデルに基づきスコアリングすると，スコア十分位（デシル）毎にリフト（後述）を求めることが可能となる。各モデルによってデシルのリフトが異なるので，ゲインが大きくなるような手法を選択することが有効である。たとえば，DMへの反応に対するスコアリングを行うと，上位デシルのリフトが最も高くなるようなモデルが望ましく，それらのデシルにDMを送付することになる（マルトハウス(2003)参照）。オンラインでの利用法として，コールセンターにおけるオペレータを支援するためのCTIにKDDの成果を反映したモデルによる情報処理エンジンを構築することがあげられる。顧客とのインバウンド/アウトバウンドのテレマーケティングに際して，会話中のキーワードや顧客属性を入力して，データマイニングによって確立されたルールに基づくリコメンデーションによる併買促進や「乗り換え」防止に必要な手段をオペレータに指示することを自動化することができる。このように，デジタル化されたCRMは，マス市場での平均ではなく，個別の顧客への対応を実現し，その効果や効率を高める手段で

ある[3)]。

3．データマイニングの代表的手法

ここではデータマイニングの代表的手法の概要について紹介する（全般に，ベリー・リノフ（1999），豊田（2001），月本（2004），山口・高橋・竹内（2004）などを参照している）。

3.1　連関規則

もっともよく知られたデータマイニング手法かもしれない。「Aを購入する顧客は，Bも購入する割合が高い」のような，条件Aが与えられたなら結果Bが成立しがちであるという規則抽出を連関規則（association rule）といい，POSのようなOLTPに基づく点でマーケットバスケット分析ともいう。見かけは単純であるが，アイテム数が増えると組合せが簡単に爆発してしまいがちなので，大量のデータからすばやく連関を見つけるためにはハードウエアの性能向上とともに，アルゴリズムを工夫する必要がある。

連関規則の実例として紹介されているものとしては，たとえば

① 木曜日には食料雑貨店でビールと紙おむつとが買われる
② 家電の製品保証契約をつけるときは高額のものが買われる
③ 荒物店の新規開店ではトイレットリング（便器の部品）がよく売れる

がある。上記①は経営上有益，②は自明，③は説明困難の各例とされている（ベリー・リノフ（1999）pp.17-8）。規則が発見されたからといってすべてが有効なわけではない。

連関規則を定式化するために，使用される用語を定義しよう。「サポート」とはAとBとを事象とする同時確率$p(A, B)$をいい，Aが与えられたときにBが生じるであろう「信頼度」はBのAという条件付き確率$p(B|A)=p(A, B)/p(A)$をいう。つまり，サポートを条件の妥当する割合で除すと信頼度が求められる。また，リフト$=p(B|A)/p(B)$は条件Aの有無によるBにおける効果である。AとBとに関して，信頼度は非対称であるが（有名な例では，紙おむつ売場でのビールの訴求であって，逆のビール売場に紙おむつ，ではない），リフトは対称的である。規則においては，信頼度が高いことは必要条件であるが，サポートがある程度大きくてリフトも高くないと経営上の意義は薄い。

実際の計算にあたっては，単品（SKU）から商品グループまでアイテムの取り方によって組合せ数や発見が期待される規則が変わってくる。とくに，A，Bそのものが組合せになるので，計算量の指数的増大が生じる。それを避けるために，アイテムでグループ化を行ったり，些少ではあるが重要な兆候の見落としのリスクを勘案しつつサポートに最小閾値を設けたりする。

3.2　決定木

基準（被説明）変数がカテゴリーの場合（たとえば，購入する・しない），それに一番大きな影響を与えるのはどのような要因（説明変数）であろうか。また，どのような要因の組合せが購入に結びつくのであろうか。これらの問に答えられるようにデータを分類するのが決定木（decision tree）である（基準変数が量的な場合は回帰木）。決定木が図示される場合，通常上にルートとして基準変数の平均値が置かれ，その値に最も大きな影響を与えるように集団を分割するように変数（またはその水準）が選択されると分岐のノードから

「枝」が生じる。これを次々繰り返すと，分割終了時にはターミナルノードを「葉」とする樹状の分類系統図ができる。

この分類は複数の有力そうな説明変数が多数あり交互作用が予想される場合に，予想外の有力要因を発見し見かけの関連を排する点で有効である。分割のためのアルゴリズムは，情報エントロピー，ジニ・インデックス，対数尤度比カイ自乗に基づくモデルが代表的である。いずれも，ノードでの分割を行うばらつきの指標であり，これらが最大になるようなグループ分けを行う。これらのアルゴリズム毎の特徴には，初期の尤度比モデルやジニ・モデルのように説明変数の水準で2分割を行うものや，ノードを最大まで広げてから過学習を回避するために枝の分岐にペナルティを与えて「剪定」を行うか(情報エントロピーおよびジニ・モデル)，有意水準を先に与えて分岐を早期に終了させるか(尤度比モデル)などがある。いずれの方法においても，リフトを上げるようなグループ分けをルール化して，予測に用いられる。

決定木は応用範囲が大きいが，一例として携帯電話会社は,顧客の他社への「乗り換え」を予測して防止策を考案できるかがある。ベリー(2002「事例編」，第3章)では，加入者の契約の継続・乗り換えを基準変数として，そのリフトに影響力を与えそうな情報(クラブへの加盟，年齢等の顧客データ，料金プラン等のサービス内容，通話回数・料金等の請求システムからのデータ)を説明変数としてモデルが構成され，モデルの比較・評価からCRMへの実装までが述べられている。また，表－1は，尤度比モデルに基づく決定木を用いて，都市化の下で属性や意識に応じた国勢調査員のグループ分けの一部である。ここでは，次回国勢調査においても再度調査員を引き受ける意向があるかを調査員に尋ねて再任率に影響を与えそうなルールの抽出を試みている。その結果，再任を忌避するものは全体で約半数であったが，動機が依頼による場合は忌避感のリフトが高くなり，それが特定の地区に限ると忌避率は実に9割近くに達する一方，再任忌避率がとくに低くなるような属性の組合せも識別された。

3.3　ニューラルネットワーク

生体の脳の神経細胞(ニューロン)において，他のニューロンから樹状突起を経由してもたらされた神経伝達物質は一定の条件で細胞体のパルスを発生させ，その信号は軸索を通りシナプスからまた他のニューロンに伝播される。これらニューロン相互のシナプス結合は学習によって強化されることでネットワークが形成され，脳における認知等の機能を果たしていると考えられている。これを一般化し

表－1　1995年国勢調査員調査における調査員の再任の条件*

もし（条件）	ならば	再任忌避率	n	リフト
データ全体（平均）では,	→	53%	1,839	1.00
調査員になるのが，頼まれたから仕方なく，	→	70%	811	1.33
かつ，地域が福岡市	→	89%	249	1.69
調査員になるのが，頼まれたから仕方なく，ではなく	→	39%	1,028	0.74
かつ地域が長崎県である	→	21%	178	0.40
調査員になるのが，頼まれたから仕方なく，ではなく 　かつ，地域が町田市および福岡市， 　かつ，社会奉仕のため，ではない	→	57%	359	1.09

＊）調査，調査票，解析方法等に関して詳しくは資料を参照。
資料）池田（1998），p.123から作成。

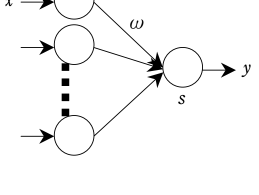
て(機械)学習のモデルとしたのが(人工的)ニューラルネットワーク(artificial neural network)である。なお, ニューラルネットワークの種類としては構造上フィードフォワード型と相互結合型とに分けられ, また学習方法上は教師付きかどうかで分類することもできる。本項では, 理論史上および応用上の観点から代表的といえるパーセプトロン(perceptron)をおもに取り上げることとする。

初期のモデルである単純パーセプトロンは, 図－2aにおいて$s=\sum_i \omega x$のような形でユニットへのn個の入力が重みづけられ, $s \geq 0$のとき$y=1$(さもなくば0)となるような関数を表している。ここで, 一般性を失うことなく重み係数ωのうち一つを閾値とすることができる。信号は図の左から右に流れ, モデルの出力yをデータである教師信号y^*と照合し, 正誤(二乗誤差)に応じて逐次ωを変更することが学習の課題である。

単純パーセプトロンはニューロンのよいモデルであり, その後の「教師付き学習」(supervised training)の基本となった。しかし, このままでは通常の回帰と同じく線形分離できないようなデータには不適合であった(たとえばXOR回路)。その克服は入力層と出力層の間に中間層を設けることで前進を見た。中間層は隠れ層ともいわれ, 見えない内部構造を持つことで非線形のデータを処理できるようになった。図－2bは, 2つのユニットの入力層, 3つのユニットの隠れ層が一つ, そしてユニット2つの出力層からなる多層(階層型)フィードフォワード・パーセプトロンを表している。なお, 出力はステップ関数から微分可能なように自由度を持たせたシグモイド関数$y=1/(1+\exp[-\alpha s])$等に置き換えられるが, 基本的に定義域・値域とも0から1の範囲である。中間層数やユニット数を与えるモデルのトポロジーはデータとの適合性を考慮しつつ試行錯誤的に設定することになる。

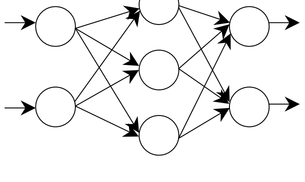
多層パーセプトロンにおける効果的な学習方法として考案されたのが, 「δルール」または「誤差逆伝播法」(error-backpropagation)である。出力の誤差が直前の中間層のωを勾配法によって修正し, 順次入力層の係数まで遡行して修正し, 誤差が収束する(か局所解に達する)まで繰り返しこれを行う。多層パーセプトロンにおけるフィードフォワードされる信号とは逆方向で係数が修正されるのでこの名称がある。

多層パーセプトロンモデルは与えられたデータに対して学習によって出力の関数を適度に組合せることで, あらゆるパターンに適応することが保証されているため汎化性が高い。とくにパターン認識や画像処理など工学上の成果があげられる(手書きの郵便番号の認識や*NETalk*)。経営での応用事例は金融の予測や不正使用防止, 与信管理, 医療診断等



図－2a　単純パーセプトロン

入力xiが重みωi付きで合計され, 値が閾値を超えると出力$y=1$, さもなくば$y=0$ ($i=1,2\cdots,n$)。



図－2b　多層パーセプトロン

左から入力2, 中間3, 出力2の層を持つ

多岐にわたる。たとえばベリー・リノフ(1999)ではボストンの不動産鑑定が取り上げられ，ある物件の評価額を出力するために，入力として居住面積，車庫の広さ，築年数などを与えている。ある条件の組合せや一定の値を超えると評価額が変わるような場合にとくに有効と思われる。もちろんリフトの高い顧客層のセグメンテーションにも用いられる[4)]。

データマイニングについて既存分野へのあるいは自立した分野としての位置づけが議論されてきた。たとえば，同時確率をサポートというように統計学と等価な用語が増殖している。他方，脳の認知活動の模擬から始まったニューラルネットワークの研究であるが，実際のニューロンは決して階層型パーセプトロンやδルールのようには作動しないので，「ニューラル」という形容は議論の対象となる。多層パーセプトロンは入力を伴うとはいえ，波形を周波数に分解するフーリエ解析のような関数近似法でもある[5)]。

3.4　テキストマイニング

テキストマイニングの対象とするデータは企業に寄せられた自然言語によるある程度非定型的な発話・記述を対象とする。具体的には，CTIに記録されたクレームや故障情報，web掲示板・ログ，アンケートの自由記述欄のような文章に対し，キーワード化や分類検索にはじまり，有用な知識を得ようとするものである。コールセンターの例では，顧客の発するキーワードがテキストマイニングによって処理され，従来は組織的な知識の開発・利用は難しかった。その結果に基づきオペレータはCTIを通じて瞬時に経験を要さずに対応することが可能となる。

統計データはコード化によってあらかじめ指定され類型化された回答や測定値だけから構成されるが，文章はそうではない。テキストデータの解析に先立って自然言語処理を行われるが，日本語などの文章表記では分かち書きを行わないため，予めテキストの品詞分解を行う形態素解析が必要となる。形態素や構文は解析することができても，発話の趣旨をかなりの程度再現するのは困難である。意味を特定するためには直接書かれていない(話されていない)文脈を取り扱わねばならないからである。

それでも，ある程度定型的な発話状況を前提して語の類似性から検索・FAQや自動分類を行うことをテキストマイニングとして，メーカー(コールセンターや営業日報)，webのフィルタリング，大学の講義(試験・授業評価の自由回答)などの実践例がある(石井，2002参照)。

3.5　その他の手法

何がデータマイニングに属するかがあらかじめ決められているのではなく，KDDに用いられうる方法が結果的にデータマイニングとされる。上記以外の方法としては，ここでは単に名称をあげるに止めるが，たとえばベリー・リノフ(1999)にしたがえば，記憶ベース推論，クラスタ分析((*k-means*法)，リンク分析，相互結合型ニューラルネットワークである自己組織化マップSOM，コホーネンネット，遺伝的アルゴリズムGAなどが紹介されている。これらの手法は単独で使用されるだけでなく，同一課題に対してパフォーマンスが比較されたり組合せたりされて柔軟に用いられる。

4．データマイニングと統計学

まとめとして，まず従来の統計的方法とを比較してデータマイニングの特徴を明らかにし，次いで社会統計学においてKDDのよう

な統計的実践がどのように位置づけられうるかを考察する。経営におけるデータマイニングの特徴について理解するために，表－2のように統計的方法やマーケティングリサーチと比較対照しながら以下の三点について整理を行った（芝村（2006），Hand（1999）参照）。

4.1　解析の目的

マーケティングリサーチや他の工学的手法同様，経営におけるデータマイニングについては，解析の目的が経営上の課題に従属している。このことは解析の方向性や結果の求め方に基本的な相違をもたらしているので，両者の相違は認識論上の差異に還元できない。一般に，統計的学術研究の目的は，自立的もしくは当該学科あるいはパラダイムの内で法則定立志向的に設定される。具体的には，過去の事実に関するある言明が一定のケース数の関連するデータに照らし矛盾していないかについて過誤確率付きで判定される。統計学はきわめて普遍的な方法論という形態をとる。

しかし，このような方法論は，「誰がこれを買いそうか」というような経営実務上の意思決定にふさわしいとは限らない。マーケティングリサーチにおいては通常の統計的方法が適用されるが，外部にある収益等の経営上の目的が当該研究計画に付与されていると考えられる。そのため，マーケティングリサーチであれ経営におけるデータマイニングであれ，当面の行動の指針に役立つような予測やパターンであるルール，つまり試す価値のありそうな仮説の導出の発見が課題である。経営上の意思決定における「仮説検証」と法則定立的な学術的な仮説の「推定・検定」とは用語ほどには似ていない[6)]。

マーケティングリサーチやデータマイニングによって得られた結果はルールにまとめられ，当該企業においてたとえばCRMに反映されるなどして公開されずに私的に利用されることになる。これらルールは，おそらく局所的，非永続的，発見的，非理論的であり，普遍的な「法則」や真理（が存在するとして）の探究とは異なる。手法としては，経営以外の分野においてもたとえば遺伝子のマイクロアレイの解析などにも用いられるが，発見的な使用法といえる（あるいは工学上の応用である）。

表－２　データマイニングと統計的方法との比較

方法*		統計学（統計的方法）	マーケティングリサーチ	経営におけるデータマイニング
目的	基本価値	科学	→	収益
	目標	検証	→	予測・発見
	成果	科学的知見	（業務に反映）→	CRM等に反映
	結果の公表	あり	→	なし
データ	ソース	調査または統計	←	OLTPのDB等
	パラダイム	母集団/標本	←	記録
	サイズ	小	←	（きわめて）大
	獲得形態	一括	←	逐次
モデル	目的	パラメータの推定	←	パターン検出・識別
	条件	一意性・最適性	←	安定性
	解析主体	手動	←	自動
	チェック	モデルの検定	←	過学習
	チェック方法	有意性検定	←	クロスバリデーション
	理論	分布論	←	機械学習

＊）典型的な場合を想定しているので，個々の手法・条件では妥当しないこともある。

4.2　データの性質

データに関しては，制度的な官庁統計にせよ学術的な質問紙調査にせよ，また経営上のマーケティングリサーチにせよ，調査目的にしたがってデータが収集される[7)]。それに対し，データマイニングの対象は逐次更新される大規模な業務の記録であるOLTPが正規化されて蓄積されたDWHである。企業と顧客との接点の痕跡から何か有用な知識獲得を行うことがデータマイニングの動機であった。統計調査では調査(設計)段階と後の解析段階とでのカテゴリーの理論的同一性(蜷川の「信頼性」)が強く要請または観念されるが，データマイニングの場合はそもそもどのような性質のデータかがほとんど不明である。

このようなデータ取得の経過から，データマイニングにおいてデータの質の確保には固有の課題がある。一つは，取引その他の記録が残らないようなケースが完全に脱落することである。統計調査の意義は母集団の性質の解明にあるが，データマイニングにおけるOLTPには半分(以上)の取引のないデータがそもそも欠落している[8)]。さらに，データのDB化にあたっては，正規化や洗浄に多大のコストや情報流出・悪用のリスクが発生する。OLTPはもともと解析を予定する調査データとは異なる2次データである[9)]。

データのサイズはマクロ経済時系列の数十からDWHのTBのオーダーのスケール差があり，そのため小標本のような統計学の基本的な考え方が不要になることはよく知られている。軽微な差異が膨大なケースによって検出されて帰無仮説が棄却されがちになるが，その結果経営上の意思決定(あるいは学術的にも)に有益な情報がもたらされるとはいえない。データマイニングの対象は，単に大きなデータセットというだけでなくDBであることからその性質の予測がつかないのでそこに記されたテキストそのものであり，また集計量ではないので製表された「統計」とはほとんどかかわりがなくなっている。

なお，大規模DBを前提として，データマイニングでは，データを背後に実際もしくは仮設的な母集団から抽出されたサンプルとはみなさず，直接テキストそのものからパターンを検出しようとする。そのため，注目する現象を含むケースがデータ全体に僅少な場合，そのようなケースとそうでないケースとがバランスするように，オーバーサンプリングを行って感度を増そうとすることもある。パラメータの推定のためのランダムサンプリングとは異なり，あえて偏ったサンプリングが行われる。

4.3　モデル化

近代の統計的経済モデルでは，変数(しばしば集計量)および関数型(しばしば線形)と誤差項とについて真の関係や分布を想定し，それにデータをあてはめてパラメータを推定し，またそのような想定が無矛盾かを検定する，というパラダイムが卓越している[10)]。このモデルは実証主義的な真理発見という学術的枠組みに適したものでありつつ，要求される条件はほとんど満たされない点で特異といわざるをえない。これに対し，データマイニングでは，モデルが実用的な範囲で結果を返すことが重要である。

実用性には，一つひとつのデータセットにはそれほど敏感でない安定性が必要である。反応が安定していて，また，目的に応じて設定されるパラメータやランダムシードがモデルの根幹に少なからぬ影響を与えるが(ニューラルネットワークの重み係数や決定木のノード)，形式的な一意性や最適性よりも微少な差を無視した結果の安定性が優先する。たとえば，階層型ニューラルネットワー

クにおいては勾配法が局所解によって支配されてしまう可能性があるため，同一データセットにおいても，境界条件を変化させて複数回推定が繰りかえされ，個々のモデルの各係数値やその一意性については問題としない。またモデルの人間的理解も困難なので入出力関係はブラックボックスとして扱われる。モデル化が統計的現象の理解ではなく，それが何であれパターン検出のために行われる。モデルの評価は結局経営上の目的に照らして効果的かどうかで行われることになる。

実用性のチェックとして，データマイニングにおいては学習用のデータセットにのみフィットネスが高い過学習（over training）を避けることを重視している。経済時系列データなどではデータセットは原理的にサイズの小のものが1セットのみしかないが，他方膨大なDWHやデータマートのデータをすべて学習させるのは不合理である。そこで，複数のデータセットに分割し，その一つでモデルを構築・学習させ，もう一つでそれをチェックする。その結果を，残りのデータセットについて予測するようにさせることで，実用的なモデルを作り上げる。データマイニングにおいては，通例このようなブーツストラップ的なクロスバリデーションがケースを節約せずに行われる。

通常の学術研究においては研究主体である研究者がデータの設定とモデル化・解析とを自らが一回の作業で行うことが想定されるが，データマイニングではこれらの過程の自動化が志向されている。見通しの悪いデータからの知識の発見を逐次的に行うためには，機械学習を用いざるをえないためであるが，あるパターンがジャンクルールではなく経営上有益な知識と判断するのは人間の役割である。

結　論

社会統計学が広い意味で人びとの統計実践を対象としているなら，企業における統計実践も含まれると考えられるが，データマイニングは非確率的基礎付け・非国家・非公表・非科学的真理追究性等の点で従来の統計概念からははずれている。ここまで，データマイニングの実態と位置づけについて，どのような統計的実践かについて検討を行ったところ，統計学の想定とは異なるデータの状況に対処するための統計的実践であるとの結論を得た。

経営上のデータマイニングの目的は究極的には収益や株主価値への貢献にあり，そのための意思決定の支援への有効性で効果は測られる。情報諸資源の活用によって単なる個人的経験主義的でない（ではできない）意思決定の費用対効果からDWH，データマイニングからCRMに至るシステム導入が考量される。このような経営上の統計実践としてのデータマイニングは学術研究ではない。

大規模で見通しが悪く，調査目的で収集されていないデータにおけるパターンを探索的に逐次的自動的非確率的に処理しておもにルールの発見（それに基づく個別ケースの判定）を行う点で，データマイニングを通常の統計学よりも機械学習や脳科学，EDAの伝統に近づけている。このことは，観測数の少ないマクロな集計量による経済時系列の解析や母集団-標本パラダイムにおいては事実上解の一意性がシステムの安定性よりも重視され，データセットが比較的小さいことや順序づけられていたこともありクロスバリデーション等による過学習のチェックは等閑視されてきた。このように方法論は革新的であるが，経営におけるデータマイニングの成果は発見された知識が適用されるCRMの有効性

に大きく依存している。現在までのところ，KDDとしては理論的，実践的に限定された効果しかえられていないようであるが，データの蓄積は進行しているといえる。

この点をデータマイニングに関連しつつより広い領域にかかわって述べると，「統計」からデータやDBへ統計的実践の対象が，したがってそれに応じた方法論も移ろうとしているように見える。19世紀に成立した近代統計が統計の調査と公開とを通じて国民国家を表象していたが，資本主義のデジタル化，グローバル化によってその意義が変容してきているためと思われる。反対にICTの発展によって個人情報が各所で記録されDB化されることが可能となり，実際にますますそうされてきている。KDDの対象は，統計ではなくこのように社会情報化された個人情報の集合である。データマイニングの結果である条件付確率（信頼度）のルールに基づくプロファイリングによって，集団や個人への予断や監視に結びつく可能性が検討されなければならないであろう。このことは，民間企業はもちろん，それら情報を名寄せして管理しようとする国家についていっそうあてはまる。

謝　辞

本稿の執筆にあたっては大学・研究機関や関連企業の方々の情報提供や意見交換に多くを負っている。記して感謝申し上げる。なお，本稿は2003-2005年度科学研究費補助金による研究成果の一部である（課題番号15330059，代表：古川彰立命館大学経済学部教授）。

注

1．本文中の経営管理および経営情報に関する略語については付表を参照。

2．戦後日本の総合的品質管理TQCにおいて，オンラインとはワークフローに従ったラインメンバーによる同期処理を指し，オフラインとはスタッフメンバーによる非定型的非同期的でより研究的な処理を指す。以下もこの用語法を借用しているので，通信状態をさすのではない。

3．Berson, Smith and Thearling（2000）参照。たとえばラブマン（2003）によると，デジタル化されたカジノではカード利用と併せデータの収集・解析とCRMにはかっこうの分野であった。

4．中敷領（1995）は計量経済学の手法と比較して少なくとも同程度の結果を得たように報告して，解析手法としての位置づけの検討の必要性を説いている。

5．Cheng and Titterington（1994）およびそれへのBienestock E. and S. Geman, *Comment*, pp.36-7参照。フーリエ解析との相似については熊沢（1998），4章参照。ここでは，ボルツマンマシンのような「教師なし学習」モデルについてはふれていない。また，学習理論での位置については，渡辺他（2005）参照。

6．何が売れそうかという実務上の「仮説検証」が行われるとき，通常の統計学的仮説検定がほとんど行われない理由である。他の例として，池田（2000）では，経営実務としての品質管理における統計的方法の認識論的還元に反対している。

7．官庁統計においてもレジスターベースに基づく作成やマイクロデータ利用，また各種社会調査結果のアーカイブ化やメタアナリシスなどの2次利用の諸形態についても検討が必要であろう。

8．もっとも，統計調査においても悉皆調査であれ標本調査であれデータの汚染や欠落は大きな問題である。

9．Hand et al.（2000）ではさまざまな分野から取られたデータについて吟味しつつ，データマイニングにおいてとくに解析の前提となるデータの特質を考慮にいれることが強調されている。

10. マクロ経済学以外の経営経済の研究やマーケティングリサーチでは非集計量が扱われる。その場合でも以下の記述はおおむね妥当するものと思われる。

参考文献

エイドリアン，P.，D. ザンティンジ（1998）『データマイニング』共立出版.

甘利俊一，麻生英樹，村田昇（2003）『パターン認識と

学習の統計学：新しい概念と手法』統計科学のフロンティア　第6巻　岩波書店.
麻生英樹(1988)『ニューラルネットワーク情報処理：コネクショニズム入門　あるいは柔らかな記号に向けて』産業図書.
ベリー, M. J. A.(2002)『マスタリング・データマイニング：理論編－CRMのアートとサイエンス(理論編)』海文堂出版.
――――(2002)『マスタリング・データマイニング：CRMのアートとサイエンス(事例編)』海文堂出版.
ベリー, M. J. A., G. リノフ(1999)『データマイニング手法』海文堂.
ビーガス, J. P.(1997)『ニューラル・ネットワークによるデータマイニング』日経BP社.
キャベナ, P. 他(2000)『データマイニング活用ガイド』改装版　星雲社.
カリー, J., A. カリー(2001)『カスタマー・マーケティング・メソッド』東洋経済新報社.
福田剛志, 徳山豪, 森本康彦(2001)『データマイニング』データサイエンス・シリーズ　共立出版.
ヒューズ, A.(1999)『顧客生涯価値のデータベース・マーケティング』ダイヤモンド社.
市川紘(1993)『階層型ニューラルネットワーク－非線形問題解析への応用』共立出版.
池田伸(2000)「工程の声をきく：統計的品質管理の勃興と衰退」, 杉森滉一, 木村和範(編著)『統計学の思想と方法』(統計と社会経済分析Ⅱ)　北海道大学出版会　第3章所収.
―――(1998)「クロス表の視覚化による国政調査員の属性および意識の分析」『研究所報』法政大学日本統計研究所 No. 24.
インモン, W. H., R. ソーサ, C. インホフ(1999)『コーポレート・インフォメーション・ファクトリー：企業情報生態系の構築と管理』海文堂出版.
石井一夫(2004)『図解よくわかるデータマイニング』B&Tブックス　日刊工業新聞社.
石井哲(2002)『テキストマイニング活用法：顧客志向経営を実現する』リックテレコム.
熊沢逸夫(1998)『学習とニューラルネットワーク』電子情報通信工学シリーズ　森北出版.
ラブマン, G.(2003)「顧客データこそサービス向上のカギ」『ダイヤモンド・ハバード・ビジネス・レビュー』2003年7月号.
ライアン, D.,(2002)『監視社会』青土社.
マルトハウス, E. C.(2003)「スコアリング・モデル」, イアゴブッチ, D., B. J.カルダー(編著)『統合マーケティング戦略』ダイヤモンド社　第11章所収.
村田久(2004-2005)「ニューラルネットワークによる統計分析：非線形解析への応用」『ESTRELA』統計情報開発センター No. 129-134.
中敷領孝能(1995)「ニューラルネットと経済統計学」泉弘志他編著『経済統計学の現代化』晃洋書房所収.
オハロー, R.(2005)『プロファイリング・ビジネス：米国「諜報産業」の最強戦略』日経BP社.
大滝厚, 堀江宥治, D. Steinberg(1998)『CARTによる応用2進木解析法』日科技連出版社.
佐藤雅春(2001)『個客行動を予測する「データマイニング」』日刊工業新聞社.
芝村良(2006)「数理統計学とデータマイニング」『九州経済学会年報』第43集.
宍戸周夫(1996)『データウエアハウス』B＆Tブックス　日刊工業新聞社.
杉浦司(2001)『データサイエンス入門：経営に活かすデータ解析の基礎知識』日本実業出版社.
鈴木みゆき(2003)「意思決定支援システムとしてのOLAPシステムの導入」『経営論集』明治大学経営学研究所 第50巻第3号.
田辺和俊(2003)『NEUROSIM/Lによるニューラルネットワーク入門』日刊工業新聞社.
戸田山和久, 他(2003)『心の科学と哲学：コネクショニズムの可能性』昭和堂.
豊田秀樹(2001)『金鉱を掘り当てる統計学：データマイニング入門』講談社ブルーバックス.
――――(1996)『非線形多変量解析：ニューラルネットによるアプローチ』統計ライブラリー　朝倉書店.
月本洋, 松本一教(2004)『やさしい確率・情報・データマイニング』森北出版.
牛田一雄, 木暮大輔, 高井勉(2003)『SPSSクレメンタインによるデータマイニング』東京図書.
ワン, J.(1999)『戦略的データベース・マーケティング：顧客リレーションシップの実践技法』戦略マーケティングシリーズ　ダイヤモンド社.
ウエスターマン, P.(2003)『ウォルマートに学ぶデータウエアハウジング』翔泳社.
山口和範, 高橋淳一, 竹内光悦(2004)『よくわかる多変量解析の基本と仕組み：巨大データベースの分析手法入門』秀和システム.
吉冨康成(2002)『ニューラルネットワーク』シリーズ非線形科学入門　朝倉書店.
渡辺澄夫, 他(2005)『学習システムの理論と実現』森北出版株式会社.

Berson, A. S., Smith and K. Thearling (2000) *Building Data Mining Applications for CRM*, McGraw-Hill.

Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*, Oxford University Press.

Cheng, B. and D. M. Titterington (1994) Neural Network: A Review from a Statistical Perspective, *Statistical Science*, Vol. 9, No. 1.

Fayyad, U., et al. (eds.) (1996) *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press.

Hand, D. J. et al. (2000) Data Mining for Fun and Profit, *Statistical Science*, Vol. 15, No. 2.

――――(1999) Statistics and Data Mining : Intersecting Disciplines, *ACM SIGKDD*, vol. 1, Issue 1.

Kleinberg. J., C. Papadimitriou and P. Raghavan (1998) A Microeconomic View of Data Mining, *Data Mining and Knowledge Discovery*, Vol. 2, Issue 4.

Rosset, S., et al. (2003) Customer Lifetime Value Models for Decision Support, *Data Mining and Knowledge Discovery*, Vol. 7, Issue 3.

付表　略語リスト

CTI	Computer Telephony Integration
CRM	Customer Relationship Management
KDD	Knowledge Discovery in Databases
or	Knowledge Discovery and Data Mining
OLAP	On-line Analytic Processing
OLTP	On-line Transaction Processing
RFM	Recency, Frequency and Monetary
SKU	Stock Keeping Unit

コメント

佐　野　一　雄

社会科学は可能か？

情報技術革命は人間社会に計り知れない影響を与えつつある。とりわけ営利を目的とする民間企業にとって，顧客情報と取引記録の管理と分析は経営の死活問題であろう。他方で，個人情報の漏洩による犯罪被害なども後を絶たない。国家もまた，膨大な情報を適切に管理することなしには成り立たない。池田論文は，民間企業における情報の蓄積と利用を概観することにより，従来の統計概念の限界を明らかにしようと試みている。池田が指摘するデータマイニングの「非科学的真理追究性」は，社会科学の可能性と限界についての問いでもある。そこで，社会統計学の伝統でもある社会科学方法論という観点から，池田論文の提起する問題について考察してみよう。

1　データマイニングとは何か？

池田によれば，データマイニングとは大規模データからの知的発見のための手法の総称である。何らかの原理に導かれた演繹的な知的発見のプロセスとは対照的に，データ解析の現場では，帰納的な試行錯誤のプロセスを辿るのが普通である。商用のデータマネージメントソリューションにおいては，メタデータと呼ばれるべき不定形なデータベースがデータウエアハウスとして再構築され，データマイニングの対象に変換される。分析可能な多次元データはキューブと呼ばれ，ネットワークを介して任意の解析手法によって切り刻まれる。このような概観は，原データが不定形かつ大規模であること，ネットワークを利用したフレキシブルな分析が可能であること以外に，格別に革新的な理論や方法を含むものでないことを示す。条件的確率を基礎とする連関規則の発見，決定木やニューラルネットワークの応用，テキストマイニング，その他の手法は，いずれもデータ主導型の分析方法であり，帰納的な知的発見の道具である。ではなぜデータマイニングの「非科学的真理追究性」が，社会科学の可能性と限界に

ついての問いとして意味をもつのか。この点について，池田は，データマイニングと従来の統計方法とを比較して考察している。

２　従来の統計方法とどう違うのか？

池田の論点は，解析の目的，データの性質，モデル化の三点である。池田は「経営におけるデータマイニングについては，解析の目的が経営上の課題に従属している」ことを最初に指摘する。この点は，従来の統計学における理論主導型とデータ主導型の対立が，あくまでも科学的真理追究を基軸としていたのに対し，収益等の経営上の目的が真理追究に優先し，池田の主張する「非科学的真理追究性」の有力な根拠となっている。続いて「統計調査では調査（設計）段階と後の解析段階とでのカテゴリーの理論的同一性（信頼性）が強く要請または観念されるが，データマイニングの場合はそもそもどのような性質のデータかがほとんどまったく不明である」と指摘する。さらに池田は，「モデル化が統計的現象の理解ではなく，それが何であれパターン検出のために行われる。モデルの評価は結局経営上の目的に照らして効果的かどうかで行われることになる」と結論する。

経営の現場は常に動いており，想定外の事象を経験することも少なくないであろう。状況をリアルタイムに記録するメタデータとデータウエアハウスの必要性も，経営上の重要な課題として認識されることも想像に難くない。実際，データマネジメントソリューションやCRMソリューションに対する需要は強く，様々なベンダーがソフトウエアを提供している。

結局，池田の論点は，いずれも経営上の課題からもたらされる必然的帰結であり，データマイニング固有の本来的な特性とは言えまい。官庁統計においても，民主的であるか否かを問わず，国家による支配という目的が優先されることを考えれば，データマイニングそのものが従来の統計方法からの大きな逸脱であるとも言えないのではなかろうか。

３　社会科学は可能か？

それにもかかわらず，池田論文は馴染み深い古典的な問題をわれわれに再認識させる。それは自然と社会の違い，さらに科学と工学の違いの問題である。池田が「経営上のデータマイニングの目的は究極的には収益や株主価値への貢献にあり，そのための意思決定の支援への有効性で効果は測られる」と述べ，「経営上の統計実践としてのデータマイニングは学術研究ではない」と言うとき，その言外にあるのは社会科学の可能性に対する疑問ではなかろうか。社会研究において科学的真理の追究は可能かという問題である。

自然と社会という対立軸において，最もわかりやすい例は，天気予報と景気予測の違いである。バリ島の祈祷師などを除けば，人間の意識が気象に影響を与えると信じる者は少ないであろうが，景気や相場となると話は別であろう。およそ社会現象に関して，人間が意識するか否かに関わらず客観的に存在する真理を，科学の名において追究し確立することは極めて困難である。それは形而上学が担うべき課題であろう。

このような社会科学の本質的な困難性に対し，経営工学の可能性は無限の広がりと奥行きを感じさせる。営利企業に限らず，国家においても，支配目的を達成するための有用な方法であるからこそ，官庁統計，業務統計，業務記録が利用されているのであり，経営におけるデータマイニングが経営上の課題に従属するのは当然なのである。したがって，官庁統計を利用する場合と同様に，データマイニングに科学的な価値があるかどうかは，そ

の目的とはまったく別の問題であると考えるべきであろう。政治や経営においては「勝てば官軍」なのであり，正しいものが勝つとは限らず，勝者が正しいとも言えない。

それでは，経営工学の可能性と有用性に対して，社会科学は，専門家集団による言語ゲームのフィールドとしてだけの存在意義しか持たないのだろうか。実は「科学」に対する拘りが，このような疑問そのものを生んでいることは明らかである。数学の定理や物理の法則に対するプラトニックな憧憬が，社会研究に過度の科学性を要求させるのである。しかし，仮に社会現象に真の法則と呼ぶべきものが発見されたとしても，人々はその法則を学習し，行動を変化させるだろう。つまり，社会現象の研究は，どこまで進んでも人間自身との対話でしかありえない。その意味で，経営におけるデータマイニングは，健全な明るさを持っているようである。

最後に，池田が指摘するように，プロファイリングによる集団や個人への予断や監視は，おそらく不可避であろう。すでに特定の犯罪前歴者に対する監視や，出入国管理の現場がそれを示している。真の問題は，プロファイリングにあるのではなく，民主的支配における合意形成のプロセスにある。情報技術やデータマイニングには何の罪もないのであり，まさに社会に関する形而上学が必要とされる所以である。

執筆者紹介

（50音順，所属は2006年7月1日現在）

朝倉啓一郎	（流通経済大学）
井口泰秀	（愛知大学）
池田伸	（立命館大学）
泉弘志	（大阪経済大学）
伊藤国彦	（兵庫県立大学）
伊藤陽一	（法政大学）
岩井浩	（関西大学）
岩崎俊夫	（立教大学）
上藤一郎	（鈴鹿国際大学）
大井達雄	（藍野大学）
大西広	（京都大学）
岡部純一	（横浜国立大学）
小川雅弘	（大阪経済大学）
小野寺剛	（法政大学・非常勤）
香川文庸	（京都大学）
金丸哲	（鹿児島大学）
金子治平	（神戸大学）
菊地進	（立教大学）
木村和範	（北海学園大学）
坂田幸繁	（中央大学）
佐藤智秋	（愛媛大学）
佐野一雄	（福井県立大学）
芝村良	（日本大学）
杉橋やよい	（お茶の水女子大学）
田浦元	（拓殖大学）
長澤克重	（立命館大学）
長屋政勝	（京都大学）
西村善博	（大分大学）
濱砂敬郎	（九州大学）
廣嶋清志	（島根大学）
福島利夫	（専修大学）
藤井輝明	（大阪市立大学）
藤江昌嗣	（明治大学）
松川太一郎	（鹿児島大学）
水野谷武志	（北海学園大学）
御園謙吉	（阪南大学）
光藤昇	（松山大学）
村上雅俊	（関西大学・大学院生）
森博美	（法政大学）
矢野剛	（徳島大学）
山田茂	（国士舘大学）
山田満	（高崎商科大学短大部）
吉田央	（東京農工大学）

社会科学としての統計学　第４集［創刊50周年記念号］
統計学　第90号

2006年８月１日

編　　者	経　済　統　計　学　会
	会　長　　泉　　弘　志
	〒194-0928　東京都町田市相原4342
	法政大学　日本統計研究所
発 行 者	品　川　宗　典
発 行 所	㈱ 産 業 統 計 研 究 社
	〒102-0072　東京都千代田区飯田橋3-7-3
	Tel　03-3230-0731
	Fax　03-3237-9287

Social Statistics as a Social Science: the 50th Anniversary Special Issue, *Statistics*, No. 90.

1st August 2006

edited by	The Society of Economic Statistics, Japan
	president　Hiroshi Izumi
address	Hosei University, Japan Statistical Research Institute, 4342 Aihara, Machida-shi, Tokyo 194-0928
published by	Sangyo Tokei Kenkyusya
	president　Toshimori Shinagawa
address	3-7-3 Iidabashi, Chiyoda-ku, Tokyo 102-0072
	Tel　03-3230-0731
	Fax　03-3237-9287