

社会科学が真の科学となる実証研究の推進のために

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
木下千大・北村行伸

社会科学分野が真の“科学”となるために

自然科学分野は、一般的に多くの観察・実験データを分析し、不変的法則を発見、予測して、これらを実験を通して実証的に分析し検証するという手法をたどって発展してきた。一方、社会科学分野では、知的考察によって構築された理論は、実験によって検証することが実生活に大きな影響を及ぼし、取り返しのつかない事態を招く恐れもあり、理論の検証という“科学的手法”が一般的に浸透していない。しかし、社会科学分野においても過去の事象を用いて検証することは可能であり、近年では自然実験とでも呼べるような状況を見つけて、そのエピソードに基づいて実証研究を行うという手法が用いられるようになってきた。これを可能にする良質なデータは政府が有する業務統計資料及び基幹統計調査の調査票データである。

事実、海外の社会科学分野の実証研究では、公的統計のマイクロデータによる分析が主流になっている。それは、マイクロデータを分析する統計解析理論の発展と、海外における公的統計のマイクロデータの提供が進んでいることが主因である。

一橋大学は、現在「社会科学の高度統計・実証分析拠点構築」事業をもって、各種の統計データベース、資料アーカイブを構築し、理論研究と実証研究の融合をはかり、そこで得られた予測を統計的に検証し、堅牢かつ現実的な政策提言に結びつける成果を得ること

を目指している。

社会科学分野の研究の重要性は、こうした実証研究に裏打ちされた“科学的”考察に基づく行政組織による政策立案、国会等における実証分析に基づく政策評価や予算審議に活用されることによって再認識され、社会科学は大きな発展を遂げるであろう。

統計法の改正により、公的統計は改めて国民共有の財産として位置づけられたが、正に国会や地方議会において、一部の集団の利益誘導型の結論にならないために“科学的根拠”に基づく透明性の高い議論を進めることが重要である。公的統計のマイクロデータによる実証分析によって導かれた、科学的・客観的資料をこうした場に提示するという点に、社会科学分野の研究者の大きな使命があるといえるのではないだろうか。

社会科学の研究環境の整備に向けて

公的統計のマイクロデータによる実証分析を進めていくために必要な研究環境は、いくつかの課題をクリアしなければ整わない。まず、制度的には、①統計法の目的外使用承認が得られにくく、②承認までに長期間を要し、利用期間も短い、③詳細な集計様式を事前に確定しておかなくてはならない、そのため④試行錯誤や追加分析は行いにくい、といった問題がある。実務的には、①提供用にデータが作成されていないことや、②公的統計データを扱うことに研究者が慣れていない

こと、といった問題もある。

先に述べたように、一橋大学の活動目的は、我が国の社会科学の実証研究レベルの向上に向けた支援であり、そのために必要な統計データの利用環境の整備である。まず、統計データの利用環境の整備に向けたこれまでの取り組みについて、あらためて振り返ってみたい。

統計審議会が打った布石

平成という新たな時代を迎えた頃、統計データの利用環境の整備に向けた取り組みが始まった。まず制度を変えることから始める必要があった。その動機付けの1歩は、調査票情報の二次利用は、指定統計を作成するという第1義的目的と比較しても十分大きな価値を生み出す資産であるということの理解を得ることである。

当時、政府が保有する統計のマイクロデータは、統計法第15条第2項の「統計上の目的以外の使用」により、総務大臣の承認を得て使用目的を公示して使用することができたが、使用目的が公共性の高いこと、使用者が原則として公務員（税務職員及び警察職員を除く）などの運用上の基準があった。あらかじめ承認された指定統計の作成以外には、たとえその調査の改善の目的で調査実施者が研究する場合であっても、総務大臣の承認が必要とされていたのである。そして、当時、大学研究者が公的統計の個票データを使用する機会や、指定統計調査を実施する府省が研究等の目的のために、研究を依頼された場合に制限されていた。

その後、大学の研究者による個票データの使用申請が認められるようになったが、それも国立大学の研究者に限られていた。公務員であれば職務の性格上、行うすべての研究は自ずと公益性を有しているはずであるという理由から、国家公務員である国立大学の研究者が行う研究については公益性基準を満たしていると

判断されていたのである。一方、私立大学の研究者が行う研究については、その研究成果が公表されることによって公益性を有するものもあり得るとして、文部省が国費を投じていることをもって公益的研究の証しとし、文部省科学研究費補助金を受けている研究に限って、私立大学の研究者にも使用が認められるようになっていった。

こうした我が国の現状が、国内の社会科学分野の研究が海外の研究に遅れをとる要因であるとした識者の声が高まり、平成7年3月10日に統計審議会が答申した「統計行政の新中・長期構想」に、「内外の研究者や国際機関等から、標本データ（個票データから必要に応じて抽出を行い、地域区分や世帯番号等の個体の識別子を消去するなど個体の識別を不可能にしたもの）の提供を求める要望が高まっている。……このため、標本データの提供については、個体の秘密保護の担保方策を中心に、外国の制度及び提供例、国内外におけるニーズの実態、現行制度との関係、具体的な提供方策について、おおむね2〜3年を目途に専門的・技術的な研究を行う必要がある。」と記述された。

この答申が書かれたとき、既に政府マイクロデータ利用の実現に向けた課題は見えていた。そして、答申における課題という布石を打ち、官僚を動かすための準備が整えられたのである。

しかし、このプログラムは、遅々として進展しなかった。なぜなら、各府省の統計作成担当は、調査の目的である指定統計の作成が本務であり、それが実現できれば職務は全うしたことになる。すなわち、当時の法令のもとでは、調査実施者は「指定統計を作成するための調査」を実施しているのであって、その調査票情報を二次的に利用することについては、あくまで例外であり、「目的外」と考えていた。例外規定の運用は、行政上の要請からどうしても指定統計作成以外の利用が不可欠である場合に限るべきというのが当然の考え方なのである。法令が、“目的外利用”を禁じ

ている状況下で、統計作成部局が二次的利用の枠組みを積極的に検討することにはならないのである。したがって、このプログラムは制度官庁が主体となって進めていくことになる。しかし、統計データの取り扱いに慣れていない制度部局では、技術的検討は難しいため、同官庁の実施部局である統計局・統計センターが技術的な研究を行っていくことになったのである。

個票データの提供の有用性の立証に向けて

旧統計法の基本的な考え方は、「正確な統計を作成するためには、国民の理解のもとに設計どおりの情報を得ることから始まる。そのためには、集められた調査票の情報は統計の作成以外には使用しない」というものであった。

調査票情報の二次的利用の実現に向けた次の1手は、その資産価値を認識させることにあった。そのための活動が、平成8年度～10年度にかけて実施された文部省の科学研究費補助金による特定領域研究「統計情報活用のフロンティアの拡大—マイクロデータによる社会構造解析—」(研究代表者 一橋大学経済研究所教授松田芳郎)である。すなわち、当時の法令(旧統計法)にのっとり、「統計行政の新中・長期構想」の検討課題についての研究ということで、目的外使用の承認基準である“高度の公益性”を満たし、使用者の範囲を大学研究者に限定することで守秘義務を担保しての研究活動として実施したのである。

この研究では、まず、研究参加者が研究のための集計表、関数などを事務局(一橋大学経済研究所)に提出する。当時の法的枠組みである統計法第15条2項の規定に基づき、調査票の目的外利用申請は事務局が一括して行う。目的外の利用が許可されたものについて、参加者が集計のためのプログラムを作成し、そのプログラムを使用して事務局が集計し、その結果を申請者に渡し、参加者は、その結果を分析するという仕組み

をとり、参加者は、当時の法令に抵触しないように一切マイクロデータに手を触れずに研究を進めたのである。

そして、この研究活動は具体的な3つの課題であるマイクロデータに関する①統計的技法の研究、②社会制度上の問題の研究、③利用分野の拡大において、大きな成果をあげた。

社会科学分野において、公的統計のマイクロデータの利用に広く関心がもたれるようになったのは、この特定領域研究「統計情報活用のフロンティアの拡大—マイクロデータによる社会構造解析—」において、実際にマイクロデータを用いた多くの研究活動が契機になっている。特定領域研究者は、全国で約130人に達した。

しかし、研究参加者から、試行錯誤して説明変数を決めるような関数を求める場合、研究者自身がデータそのものを扱いたいとの要望多く出され、この要望を受けて、研究活動は平成12年度から16年度の日本学術振興会の科学研究費補助金による「マイクロ統計データ活用研究会」(研究代表者 松田芳郎、その後井出満)に引き継がれた。この研究会ではマイクロ統計データからのリサンプリング及び一部の調査項目の削除やトップコーディングなどの秘匿処理を施したデータを用いて、利用者と事務局が共同で研究する方式をとった。こうした一連の研究活動によって、調査票情報の資産価値は、指定統計を作成するだけにとどまらないことを実証し、匿名データの提供の枠組み構築に結びついたのである。

匿名データ提供の実現に向けて

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(以下「センター」と呼ぶ)は、平成13年の日本学術会議「情報化社会における公的統計の一次データの提供形態のあり方について」の報告を受け、公的統計の一次データ(マイクロデータ)を提供することを目的として、平成14年に日本経済統計情報センターを改

組して発足した。

その目的は、我が国の実証分析の水準を向上させるためには、公的統計のマイクロデータを多くの研究者が利用できるようにすることが喫緊の課題であるとの認識のもと、全国の研究者が従来よりもより広く公的統計のマイクロデータを利用できるようにするための組織・施設を構築する活動を新たに行おうというものであった。センターのこのような活動は、21世紀COEプログラム「社会科学の統計分析拠点構想」の研究活動（平成15～19年度）としても位置付けられた

「統計行政の新中・長期構想」の提言を受け、学界では特定領域研究による学問的な研究を進めることとなったことについては、すでに述べたとおりである。一方、答申を受けた行政の側では、関係府省による「統計行政の新中・長期構想推進協議会」が設置され、この協議会の下に検討委員会、ワーキンググループ、研究会を設置して具体的な検討が進められた。委員会では標本データ等に関するニーズへの対応が検討され、「標本データの秘密保護に関する研究会」において、秘匿措置の方法や秘匿したデータの実用性などが研究された。研究成果として、①統計調査によっては、グローバル・リコーディング、サブ・サンプリング（標本の再抽出）の秘匿措置を組み合わせることにより、個体識別を防止することができ、かつ、有用性が見込まれるデータを提供できることが技術的に可能であり、②秘匿措置を施したデータの実用性は見込まれると報告されている。

こうして、標本データの提供についての専門的・技術的な見地からの検討は終了したが、標本データに関する社会的コンセンサスの必要性、法令上の諸問題については、引き続き検討すべき事項として、新たな指針として、平成15年6月の各府省統計主管部局長等会議申合せ「統計行政の新たな展開方向」に引き継がれ、改めて検討されることとなった。

「統計行政の新たな展開方向」の中で述べられている

る個票データ提供の在り方の検討は重要な政策課題とされた。匿名標本データの作成・利用における具体的方策として、「匿名標本データの作成・利用のためのガイドラインの作成など実用化に当たっての基本的な枠組みを検討・構築する。」また「匿名標本データのニーズや導入に当たっての課題等を把握するため、必要に応じて試験的な匿名標本データの作成を行う。」としている。

総務省統計局は、この課題について、実際に提供を行ってみることにより、提供方法などの問題点を把握することを検討し、試験的な形での提供を行うことを計画した。そして、試験的な形での提供を行うために、学問的な利用の問題点の把握には学界側の協力が必要であるとのことから、特定領域研究での実績があり、マイクロデータの利用促進活動を行い、21世紀COEプログラム「社会科学の統計分析拠点構想」の研究も行っていった、全国共同利用施設である当センターに協力を依頼することとなった。

我が国の旧統計法の下でも、秘匿処理を施したデータを提供することにより、従来よりも利用しやすい提供方法を構築することが可能ではないかと考えられた。そこで、当センターは、平成16年には、総務省統計局から「匿名標本データの作成・利用の試行的運用」の研究を委託され、公的統計の秘匿処理済マイクロデータを、全国の大学研究者が、学術研究目的で利用できるシステムを構築することとなった。

試行的提供開始

試行的提供の目的は、匿名標本データ提供のガイドラインの検討に資するために、提供システムの構築や運営に関する問題点を把握することであった。

試行的提供の意味するところは、匿名標本データの作成・提供が試行的であることと、旧統計法のもとで、ユーザ側から見て簡素化された提供事務を試みたとい

うことにある。しかし、提供事務の簡素化は、旧統計法に則った調査票の目的外使用の包括的承認によって実現されたに過ぎなかった。匿名標本データに関しては、データの匿名性の検証というより、匿名化によって情報が制約されたデータの有用性についての検証に重きがおかれたといつてよい。

試行的提供は、旧統計法の下で、「匿名標本データの作成・利用に関する研究会」における研究の一部として実施するという方式をとり、その指揮・監督の下でシステムを構築・運営することになっていた。

この間、就業構造基本調査、社会生活基本調査、全国消費実態調査、住宅・土地統計調査の4つの調査の匿名標本データを大学に籍を置く研究者に提供し、その件数は、述べ132件にのぼった。

この活動に対しては、平成18年度には日本統計協会から「統計活動奨励賞」、平成19年度には日本統計学会から「統計活動賞」を受賞している。

	提供件数	利用者数
平成16年度	6件	7人
平成17年度	16件	20人
平成18年度	29件	42人
平成19年度	27件	39人
平成20年度	54件	87人

こうした試行的提供の成果は、平成19年に行われた統計法の改正、匿名データ作成のガイドラインに貢献するものとなった。

そして、この試行的提供を通じて、匿名データでも十分な有用な分析結果が得られることも確認できた。すなわち、世帯や世帯員に関する統計分析であれば、目的（地域間比較をしないなど）によっては現在提供されている匿名データで十分ということが言えるのである。匿名データについては、高齢者や規模の大きい

世帯の情報がないなどの指摘がなされるが、若年や中高年世代の分析を目的とする場合には全く影響しないし、特殊な世帯類型が削除されていることなども、仮にそれらのデータを含めたところで、もともと希少データなので統計的には誤差が大きく有意性が低いものである。場合によっては、特異値として除去した方が分析しやすいことさえある。

統計法の改正と匿名データの提供

当センターが行った研究の一環としての匿名データの試行的提供は、平成20年度末で終了した。現在は、平成21年3月31日に締結した独立行政法人統計センターとの連携協力協定に基づいたサテライト機関として、平成21年6月22日から、新統計法に基づいた匿名標本データ提供の窓口業務を行っている。

当センターは、「匿名標本データ提供の窓口業務を、単なる独立行政法人統計センターのサテライト機関としての委託業務とは考えていない。一橋大学の目的である「社会科学の高度統計・実証分析拠点構築」事業をもって、各種の統計データベース、資料アーカイブを構築し、理論研究と実証研究の融合をはかり、そこで得られた予測を統計的に検証し、堅牢かつ現実的な政策提言に結びつける成果を得ることに資する事業と認識しているからである。

これまで、当センターを通じての匿名標本データの利用状況は、平成24年12月末現在、次表のとおりである。

	提供件数	利用者数
平成21年度	11件	32人
平成22年度	18件	34人
平成23年度	4件	10人
平成24年度	9件	24人

新統計法は、公的統計を作成するための調査票情報の二次利用に関して間口を広げた。しかし、大学等の研究者が調査票情報そのものを使用できるのは、法第33条第2号の規定に基づき公益性を有する統計の作成等として承認された、さらにその実施に要する費用の全部又は一部を公的機関が公募の方法により補助する調査研究に係る統計の作成等に限られている。このような、制約は、公的統計の性格上必要であることは理解できるが、統計委員会が公表した「公的統計の整備に関する基本的な計画」（平成21年3月）では「研究者等による、より高度かつ多様な研究分析等を通じて、統計が学術研究はもとより社会の一層の発展に寄与することが強く期待されていることから、秘密の保護に配慮しつつ二次利用に係る事務処理を適切に実施していくことが必要となっている」と明記されていることを鑑みると、より広範な研究者・分析者が公的統計を利用できる道を探ることも同時に考えなければならない。現状では、法第33条第2号の利用が出来ない研究者には法第36条の規定に基づく匿名データを使用してもらうという道が開かれており、当センターでは単なる匿名データの提供窓口業務に終わらず、調査のしくみ等について、利用説明会を開催するなどして、統計の二次利用が社会発展に貢献することを支援している。

改めて実証研究の環境支援を

匿名標本データは提供が開始されて3年を過ぎたところである。この間、匿名データを使用しての実証研究の件数は期待している程には増えていないのが実態である。その要因は、一つには匿名データの有用性が十分に評価されていないこと、そして存在自体が認知されていないということがあるだろう、もう一つは、公的統計データを使った実証研究を行う研究者の後継者が十分育っていないということも考えられる。

こうした状況を踏まえ、当センターは、平成21年度

から23年度までに4回の匿名データの利用説明会を開催してきた。平成24年度は、これまでの調査の概要や利用手続きの説明に加えて、若手研究者の育成への活用事例や匿名データを利用する上での分析上の留意点についても紹介する内容とした。

その他の研究支援活動

当センターは、これまで述べてきたような匿名データの提供以外にも、実証研究を行う研究者を支援する活動として、調査実施者が集計を行っていない結果表のうち、研究者のリファレンス資料として重要であり、活用が見込まれる結果表を“独自集計”として行い公表している。また、独立行政統計センターからオンサイト施設として認証された一橋大学小平国際キャンパス内の1室を用いて、総務省統計局との共同研究「オンサイト利用施設を用いた調査票情報の使用の試行運用」を行い、その機能・運用形態等について課題の洗出しを行っているところである。