

【フォーラム】

匿名データの教育目的利用に関する一考察

小林良行*

はじめに

1947年3月に公布された統計法（以下、「旧統計法」）が60年ぶりに全面改正され、統計報告調整法の廃止とともに新しい統計法（以下、「新統計法」）が2007年5月に公布され、2009年4月より全面施行された。新統計法の中で公的統計は、「国民にとって合理的な意思決定を行うための基盤となる重要な情報である」とされ、社会全体で利用される情報基盤として位置付けられている。新統計法では、公的統計が一層活用されるように、学術研究・高等教育の発展に資すると認められる場合に一般からの委託に応じて既存の調査票情報から新たな集計表を作成・提供したり、匿名性を確保した調査票情報（匿名データ¹⁾）を提供する利用制度が新たに創設された。

ここで注目すべきは、高等教育の発展に資すると認められる場合に匿名データの利用が可能となったこと、すなわち大学等において講義や演習（以下、「講義等」）を行う教員又は当該教育機関が、教育目的で申請して承諾されれば匿名化されたマイクロデータを利用できるようになったことである。これは、旧統計法の運用において、当初承認された統計の作成以外の目的で調査票を使用（目的外使用）するには、例外的に、①行政上の基礎資料として利用する場合又は②高度に専門的かつ公益性が高い学術研究で利用する場合に限り、行政機関又は大学といった法人のみが申請者となりえたことと比べ大きな前進と言え

よう。

しかし、匿名データを用いてマイクロデータの利用方法を学生に教えようとする、実際的な面で課題と思われる点がないわけではない。

本稿では、まず我が国のマイクロデータの現状について整理し、次いで匿名データを教育目的で利用する際の課題と思われる点を考察し、次に匿名データなどの公的統計マイクロデータの利用方法を教えるための教育用データの可能性について述べることとする。

1. 我が国のマイクロデータ

本節では、匿名データの提供制度と我が国のマイクロデータについて概説する。

匿名データは統計調査を所管する行政機関等が作成し、その提供は独立行政法人統計センター（以下、「統計センター」）が提供の可否の判断も含めてすべての事務を受託できることとなっている²⁾。匿名データは、①学術研究の発展に資すると認められる場合（学術研究目的での利用）、②高等教育の発展に資すると認められる場合（高等教育目的での利用）、③国際社会における我が国の利益の増進及び国際経済社会の健全な発展に資すると認められる場合（国際比較統計利活用事業目的での利用）のいずれかに該当する場合³⁾に提供を受けることができる。実際の匿名データの提供に当たっては、総務省政策統括官（統計基準担当）が、匿名データの作成及び提供の事務処理の明確化、標準化を図るための「匿名データの作成・提供に係るガイドライン」（以下、「ガイドライン」）を定めている。

* 一橋大学経済研究所

〒186-8603 東京都国立市中2-1

2011年2月末現在で利用可能な匿名データは、総務省統計局が所管する全国消費実態調査、就業構造基本調査、社会生活基本調査及び住宅・土地統計調査の4調査（以下、「統計局4調査」）のみである。統計センター及びそのサテライト機関⁴⁾ではガイドラインに基づき、利用者向けのマニュアル（以下、「利用の手引」）を定め、提供事務を行っている。匿名データの提供を受けるには、利用目的その他の所定の要件を満たし、事前に申請して承諾され、利用に関して誓約をするといった一定の手続きが必要である。

表1は、新統計法の下で二次利用ができる我が国のマイクロデータについて、Wende (2004), Zwick(2007), Brandt et al.(2009) によるドイツのマイクロデータの例を参考に整理・比較したものである。ドイツを例として取り上げたのは、管見の限りでは作成されているマイクロデータの種類、利用目的、利用形

態と法制度上の規定が明確に関係づけられており、匿名化の程度の違いによりマイクロデータを体系的にとらえようとする上で一つの視座を与えるものと考えたからである。もとより日独では法制度やその運用に違いがあり、作成されているマイクロデータの種類などに完全な対応関係が成立するわけではないが、ドイツの例を通じて我が国の現状を把握することには意味があると言えよう。

表1にあるPUF (Public Use File) とは、1981年にドイツの連邦統計法改正で規定された絶対的な匿名性⁵⁾を持つマイクロデータのことで、利用するために特別な資格や手続きを必要としないものである。SUF (Scientific Use File) とは、1987年の連邦統計法改正で規定された事実上の匿名性⁶⁾を持つマイクロデータのことで、利用するためには一定の資格や手続きが必要で学術研究目的のみに利用できるものである。また、CAMPUS Fileとは、大学に

表1 匿名化の程度からみた我が国のマイクロデータードイツとの比較

日本のマイクロデータ		ドイツのマイクロデータ				
マイクロデータの種類	利用目的	匿名化の程度	情報損失	情報の有用性	マイクロデータの種類	利用目的
調査票情報	(注1)	非匿名化	↑	↑	アクセス不可	
		形式的な匿名化 ^(注2)			CRDPによる利用 ^(注3)	科学研究目的
匿名データ (国外での利用可)	・学術研究目的 ・高等教育目的 ・国際比較統計利活用事業目的	事実上の匿名化			↓	↓
-		絶対的な匿名化	PUF (国外からの購入可)	一般汎用目的		
			高	低	CAMPUS Files	教育目的

(注1) 新統計法第32条及び第33条によれば、行政機関等での二次利用（調査実施者内部で調査票情報を二次的に利用すること）及び行政機関等での利用と同等の公益性を有すると認められる学術研究目的に利用することができる。

(注2) 氏名、住所のような直接的識別子をなくしたもの

(注3) CRDP (Controlled Remote Data Processing) は、我が国のオーダーメード集計に相当

における統計手法の教育訓練を専らの目的として作成されたマイクロデータであり、インターネットを通じて無料でダウンロードできるPUFである。PUFは学術研究に用いることができるように作られているが、CAMPUS Fileはそのように作られてはいない。

表1を見てわかるとおり、新統計法の下では現在のところPUFに相当するマイクロデータはない。したがって、ドイツのようにPUFの一種として専ら教育目的で利用することを想定して作成された匿名化データ⁷⁾は存在していない。しかし、日本学術会議(2005)及び松田(2008)によると、統計法の改正前にも①マイクロデータの使用方法を教えるのに利用でき、②実際のデータに基づいているが個別の回答者とはリンク不能で、③特別な手続き(目的外使用申請)なしに、④大学の教員クラスの研究者だけでなく大学院生にも使用できるといった性質を備えた、「レプリカデータ」と呼ぶ教育用データの作成が提案されていたことがわかる。「レプリカデータ」は、その性質からすると表1にあるCAMPUS Fileに相当するものと考えられる。これに対して、匿名データは、事実上の匿名性を備えているマイクロデータと考えてよく、利用に当たって一定の手続きが必要である点からすると、SUFに近いものと考えられる⁸⁾。また、調査票情報には、個人の氏名や会社等の名称、住所を含んでいるものと氏名や名称、住所を含まないものの二種類がある。前者のタイプの調査票情報は、紙媒体や画像に記録された調査票及び事業所・企業名簿のもとになる調査の電磁的記録が該当し、表1では非匿名化データのカテゴリに分類されるものである。一方、電磁的記録媒体に記録されているほとんどの調査票情報は後者のタイプである。これは表1の形式的な匿名化データに相当するものと考えられる。

2. 教育目的での匿名データの利用と課題

本節では、匿名データを教育目的で利用する際の課題について考察してみたい。

まず1点目は、多人数の学生(学部学生及び/又は大学院生を指すものとする)を対象とした講義等で学生に直接匿名データを利用させる場合の課題である。ガイドラインや利用の手引では匿名データを高等教育目的で利用する場合の運用上の考え方として、利用者の範囲により、提供された匿名データを用いて教員が講義等の資料を作成・配付する場合と提供された匿名データを教員が各々の学生に利用させて講義等を行う場合という2つの利用形態が想定されている。一方、同じくガイドラインや利用の手引には、①事前に、受講するすべての学生の氏名を明らかにすること及びすべての学生の誓約書の提出が必要であること、②同一の匿名データを同時に複数のコンピューターにより複数の学生が利用する場合は、利用する人数=台数に応じたファイル数が必要となること、③匿名データの格納媒体はまとめて指導教員に提供すること、④匿名データの利用期間中は匿名データとその格納媒体を指導教員が適正に管理する義務⁹⁾があることが定められている。

実際に授業でデータを使わせようとなると、教員は、まず匿名データ格納媒体の保管場所から媒体を取り出し、教室で出席者に配付し、講義等を終えたら学生から媒体を回収して、学生が利用したPCにデータが残留していないか確認し、回収した媒体を再び保管場所に保管するという行為を授業のたびに行わなければならない。少人数の場合はそれほどでもないが、多人数の講義等で同じことをやろうとすると手間と時間がかかり、講義等の時間を圧迫しかねない。このように、匿名データ提供制度の運用の中で、想定はしているものの現実にはうまく機能しないことがある。この点については、法制度の運用面で何らかの改善が必要であろう。

2点目は、マイクロデータの知識をもたない初学者に匿名データの適切な使用法を教育する場合の課題である。新統計法の下で、教育目的に利用できる公的統計のデータは、公表されている統計表と匿名データの2つのみである。統計局4調査の匿名データは、各調査とも大量データと多くの変数(調査事項)が特徴¹⁰⁾であり、データ構造も単純ではない。匿名データを利用する際には、もともなる統計調査の調査設計に関する基礎的知識、調査票の調査事項とレコード上の変数の対応、公表値の集計方法などについて十分理解した上で、適切なデータの使用方法、集計結果や解析結果の解釈のしかたを習得するのが望ましい。しかし、匿名データの初学者がそこまで至るにはかなりの時間を要するのは否めない。匿名データの利用を学ぶ前提知識として、公表されている統計表の利用法や調査設計に関する基礎知識を教えることも必要であろう。大学における教育を考えると、統計表のようなマクロデータとその利用に関する知識の習得から匿名データ、調査票情報のようなマイクロデータとその利用に関する知識の習得まで、公的統計とその利用を段階的に学んでいけるようなカリキュラムの検討が必要であろう。

3. 教育用データの可能性

前節のような点を解決する方法の一つとして、匿名データの入門用として使える日本版CAMPUS Fileの可能性を考えてみたい。ここで、日本版CAMPUS Fileと仮称したのは、教育目的での利用に限定したマイクロデータという意味合いである。

初めて匿名データを扱う研究者の訓練や大学生の教育に利用できる教育・訓練用データについて、統計委員会(2009)は、一般研究者用の匿名データとは別に利用者の習熟度に応じた簡易な匿名データを作成し、簡便な手続きで提供することの必要性に関して議論を

行っている。議論の中ではこの簡易な匿名データのことを「いわゆるレプリカデータと呼ばれている」ものとしている。ここで提案されている簡易な匿名データは、研究目的に利用のウェイトを置いている現行の匿名データに比べ、より十分な秘匿を施した匿名データとされており、文脈上からはSUFに相当すると受け取れる。星野(2010)は、「レプリカデータ」を日本版PUFと解釈し、日本版PUFの議論は未熟であると指摘したうえで、匿名化措置の一種である模造(synthetic)に基づく日本版PUFの可能性を提案している。いずれにせよ統計委員会(2009)の中で指摘されているように、匿名化の程度が異なる匿名データを作成するには法令の改正が必要となるので、現行法令の下で教育用データを実現するとしたら匿名化とは違う方法を取る必要がある。

日本版CAMPUS Fileとして望ましい性質はどのようなものであろうか。まず、匿名データのような特別な手続きを必要とせずに自由に使用できることがあげられよう。実際の統計調査のデータに基づいて作成され、匿名データに比べて学習者の理解が複雑にならない程度にデータ量や項目数を絞ったものであることも匿名データの訓練用として望ましい性質である。また、一種類のデータで様々な問題に利用でき、教育に必要なすべての用途を満足するようなものを想定するのは困難なので、教育内容に応じた複数種類の教育用データを作成するのが現実的である。集計結果や相関構造などが実際のマイクロデータにできるだけ近い方が望ましいが、教員があらかじめ教育用データの特徴を知った上で学生の指導に利用するならば、必ずしも実際のマイクロデータに近似していなくても一定の教育効果を得ることは可能であろう。

むすびにかえて

我が国の教育用データの作成は、最近に

なってようやく統計センターの試みや立教大学の試みが緒に就いたところである¹¹⁾。そこで採用されている作成方法は、個別データをいったん集計し、それを加工することで疑似的なマイクロデータを作る方法である。現行の法令の下では、匿名化データを作らないこのような方法が現実的な方法であろう。将来的には法令等の改正により、“本物の”マイクロ

データから作成される日本版PUFの実現も望まれる。今後、匿名データを利用した論文作成など大学院における教育だけでなく、教育用データによる学部学生の教育を進めることで、公的統計マイクロデータの適切な扱いを身に付けた人材を育成できるようになることを期待したい。

注

- 1) 「調査票情報」及び「匿名データ」の定義は新統計法第2条に規定されている。
- 2) 匿名データの作成については新統計法第35条に、提供については同36条及び37条並びに統計法施行令第12条に規定されている。
- 3) いずれの場合も実際には統計法施行規則第15条で規定する付帯条件にもすべて該当しなければならない。
- 4) サテライト機関とは、統計センターと公的統計の二次的利用に関する連携協力協定を結んだ組織のことである。
- 5) たとえば濱砂(2000a)、濱砂(2000b)に詳しい解説がある。
- 6) たとえば濱砂(2000a)、濱砂(2000b)に詳しい解説がある。
- 7) 匿名化措置を施したマイクロデータを新統計法の「匿名データ」と区別するため、このように呼ぶことにする。
- 8) SUFは科学研究目的の利用に限定されているが、利用目的の範囲の広さからすればイギリスのSARsに近いとも言える。
- 9) 匿名データの適正管理義務は新統計法第42条、ガイドライン、利用の手引などを参照。
- 10) たとえば全国消費実態調査(平成16年)では約5万レコード、1780項目となっている。
- 11) これらの試みについての報告は、経済統計学会第54回(2010年度)全国研究大会報告要旨集(pp.50-53)を参照。

参考文献

- [1] 統計委員会(2009)『第20回統計委員会議事録』
- [2] 日本学術会議(2005)『政府統計・世論調査等の一次データ(含む個票データ)の体系的保存と活用・公開方策について』, 学術基盤情報常置委員会報告(平成17年9月15日)
- [3] 濱砂敬郎(2000a)「ドイツ」『講座マイクロ統計分析(1)統計調査制度とマイクロ統計の開示』(松田他編), pp.109-128, 日本評論社
- [4] 濱砂敬郎(2000b)「事実上の匿名性の原則」『講座マイクロ統計分析(1)統計調査制度とマイクロ統計の開示』(松田他編), pp.196-224, 日本評論社
- [5] 星野伸明(2010)「公的統計マイクロデータ提供制度の課題」『日本統計学会誌』, 第40巻, 第1号, pp.23-45, 日本統計学会
- [6] 松田芳郎(2008)「日本におけるマイクロ政府統計活用の新しい夜明け」, 『統計』2008年12月号, pp.2-9, 日本統計協会
- [7] Brandt, M, Crössmann, A.&C. Gürke(2009), “Harmonization of Statistical Confidentiality in the Federal Republic of Germany”, Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Bilbao.
- [8] Wende, T.(2004), “Different Grades of Statistical Disclosure Control Correlated with German Statistics Law”, *Privacy in Statistical Databases, Proceedings of PSD2004*, LNCS3050, pp.336-342,

Springer, 2004.

- [9] Zwick, M. (2007), “Campus Files-Free Public Use Files for Teaching Purposes”, *Schmollers Jahrbuch-Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 127(4), pp.655-668.