

政府統計匿名化データ利用に関する提言と要望

日本経済学会

2007年9月24日

1. 提言と要望の主旨と原則

政府統計は莫大な公費を投じて作成されている国民全体の財産である。そのデータは、政府の行う社会経済政策のあるべき姿を考え、政策評価を行う上でも、また社会経済の機能ならびに実態に関する理解を深めていく上でも、かけがえのない情報を提供する。このような重大な役割を担うデータの利用を、情報保護と矛盾しない範囲で促進することは、公正な政策評価を効率よく行うことを可能とし、わが国の民主主義の基盤強化と社会科学の進歩をもたらす。

以上の観点から、多くの方々の努力により、このたび統計法が改正された。新統計法がその本来の役割を十分に果たすためには、その運用の細部、すなわち、どのデータが、どのようなかたちで、誰に、どのような条件で利用できるのかが明確にされる必要がある。以下それらについての我々の提言と要望を記し、関係各位がマイクロデータ提供の仕組みを構築される際の参考とされることを希望する。

具体的な提言の前に、匿名化データの利用に関して定立されるべき原則を提示しておきたい。すなわち、第一に、利用許可は、一般に、利用するデータの秘匿処理の程度による匿名性確保及び利用目的の観点から判断されること。第二に、利用手続きや審査は公正且つ簡素であること。第三に、利用許可の審査を行政機関が直接行うことは学問の自由、思想の自由の観点から適切ではなく、公正な第三者専門家機関が審査するのが望ましい。

以上の三点を原則とし、改正された新統計法を念頭におき、以下の利用細則の具体例を提示する。ただし、欧米の現状を鑑みるに、この提言が理想的なデータ利用を意味するわけではない。たとえば、匿名性が担保されるのであれば、目的を問わず日本人のみならず万民が自由にデータを利用できるべきである。関係各位のより一層の努力・協力により、国民全体の財産としてのデータがより適切に利用されることを切に望む。

II. 匿名化データ提供に関する具体的な提言

1. 個票請求手続き・窓口の明確化

- a. 統計を所轄する省庁ホームページ等で、問い合わせ先・手続き、どのデータがどのようなかたちで、誰に、どのような条件で利用できるかを明示する。利用できないデータがある場合には理由を明示する。
- b. 日本語ならびに英語によるマニュアル類を整備する。
- c. 利用許可手続きは、第三者専門家機関が請求者の研究計画を受理し、審査によって決定するのが望ましい。

2. 提出すべき研究計画の内容

- a. 研究計画には、審査に必要な下記の情報を掲載する。全体として A4 版³ ページ程度で、審査機関が統一フォーマットを用意するのが望ましい。これらの要望を満たす研究計画書の一例を、本提言の付録として添付した。なお、これらの例は本提言における整備されるべきデータの優先順位を示唆するものではない。

- (1) 研究題目
- (2) 利用する匿名化データの名称
- (3) 利用者
- (4) 利用場所と秘密保持に関する安全性
- (5) 利用期間
- (6) 利用目的と学術的意義
- (7) 公表数表では不可能な理由
- (8) 匿名性確保への配慮
- (9) 結果の公表方法
- (10) その他必要な事項

3. 審査基準の明確化

匿名化データ利用の審査に関しては、審査の簡素化を旨とする。元来、政府統計のデータを利用する方法は、標本個体の識別に関して秘匿処理を施した匿名化データを利用する場合から、原データを利用する場合まで幅広く存在する。このとき、より厳重に秘匿処理が施されたデータについては利用可能性を広く設定し、秘匿処理の程度が軽いデータについては利用可能性を

より厳格に制限する必要がある。一般に、この利用制限は、利用するデータの秘匿処理の程度、利用目的の公共性の有無、匿名性確保の観点から設定され、利用が許可されるべきである。また、学問の自由、思想の自由の観点から、公共性の有無の審査を行政機関が直接行うことは適切ではなく、中立な第三者専門家機関が審査するのが望ましい。

本提言が前提としている新統計法第36条は、学術研究の発展に資すると判断されれば匿名化データを用いるための公共性条件が満たされると解されるべきなので、学術利用で(あるいは公共性)があること、匿名性が確保されることの2点について審査すれば足りる。原データの直接利用や匿名性確保が十分でないデータの利用については将来の課題であるが、秘密性保持とデータ流出を防ぐ技術的条件を備えた方法が構想されるべきである。この際、利用者の地理的な所在が過度の制約にならないようにする。但し、学問の自由、思想の自由の観点から、審査は学術利用の内容にはできる限り踏み込まないものとし、匿名性の確保を審査の中心とする。

<学術利用の定義>

- a. 文部科学省の認める大学に所属する助教以上の職位にあるものの利用申請は、学術利用の申請とみなす。また大学所属の研究者・大学院生についても同等の扱いとする。
- b. 大学以外の研究機関に所属するものに関しても、日本学生支援機構の第一種奨学金の返還特別免除制度が適用される機関に所属するものに関しては同等の扱いとする。
- c. 外国大学籍の者など、上記の条件に当てはまらないものによる利用申請については、その利用目的が学術利用かどうかの内容審査が必要となる。ただし、研究の高度化・専門化に伴い、学術性の有無の審査は専門家以外には不可能なのが現状である。また、学問の自由、思想の自由の観点からも、学術利用の審査を行政機関が直接行うことは適切ではない。よって、この場合の研究計画に関しては、学会や第三者専門家機関に審査を委託する。なお、この審査に関しては、学術利用かどうかは所属機関の適格性で行うのが望ましい。所属機関が民間コンサルティング会社など非営利研究機関と認められない場合に限り、研究の内容が学術利用であるかどうかを審査するものとする。

<匿名性確保>

- d. 政府データには、大きく分けて匿名化処理を済ませた匿名化データとそのような処理を施さない原データの2種類が存在しているが、匿名性確保の審査についてはデータの匿名化の度合いに応じた専門家による審査体制が必要である。事前に明示された、欧米で通常これまで用いられており、安全性に問題がないと思われる基準に従って、統計部局が行うのが望ましい。
- e. 単一の匿名化データを利用する場合には、匿名化データはすでに厳重に秘匿処理を終えているので、そのデータのみではどのような計算をしても匿名性は破られないと考えるべきである。よって、特に匿名性確保の審査は行わない。

- f. 原データを扱う場合、複数の匿名化データを接続あるいは同時に扱う場合、匿名化データに外部データを接続・併用する場合には匿名性確保の審査を要する。その際には第三者による個体識別の可能性が十分に低いことのみを利用許可基準とする。
- g. 原データを扱う場合、匿名性確保が十分でないデータの利用については将来の課題であるが、秘密性保持とデータ流出を防ぐ技術的条件を備えた方法が構想されるべきである。その際、研究者の所在場所の地理的条件が利用の過度の制約とならないよう考慮する必要がある。

4. 個票申請手続きの簡略化・迅速化

- a. 匿名性確保のために落とす必要がある変数を除いて、データセットに入っているすべての変数を一括して提供する。
- b. 1a および c で定められたルールに従い、いずれかの機関に利用申請が提出されてから 8 週間以内に許可・不許可の通知をし、通知から 4 週間以内にデータを提供する。各機関は利用申請書の提出日から起算して 1 週間以内に申請を受けつけ、その受理を申請者に通知する。利用を許可しない場合は、不許可の理由を明確にしたうえで申請者に通知する。
- c. 学会など第三者専門家機関が審査を担当する場合には、当該機関は期間内に手続きが済むようにすみやかに審査体制を整える。
- d. 申請が却下された場合の不服申し立ての制度を導入する。学会など第三者機関による制度運用が望ましい。
- e. 諸外国の官庁や研究機関では、インターネットやウェブサイトを用いた申請・利用も進んでいる。実際の仕組みを作る際には、こうした事例も参考にして、利用者にとって最も簡便な方法を検討すべきである。

5. 利用期間等およびその変更に関する申請

- a. 利用期間を 1 年に設定する。
- b. あらかじめ定められた利用期間内に、利用目的を変更する必要があると判断されるときには、あらかじめ審査を受けるものとする。
- c. 利用期間の延長・再利用の申請について、たとえば学術雑誌の査読者への対応や既存研究結果の確認など、利用期間の延長あるいは再利用の理由の正当性のみを審査し、研究計画自体の再審査は必要としない。延長・再利用の回数に制限は設けない。
- d. 延長される利用期間、再利用の期間は 1 年以内とする。
- e. 利用延長・再利用の審査は、その専門性に鑑み、^{3c} で提言された第三者専門家機関に委託することが望ましい。また、その際には 4 週間以内に利用延長・再利用の許可・不許可の決定を申請者に通知する。不許可の場合には、^{4d} の不服申し立て制度を利用できるものとする。

6. 提供データの形態

- a. 個体識別の蓋然性を下げるための手法としてセンサスデータからのサンプリングデータまたはサンプリングデータからのリサンプリングデータを用いる際には、高いリサンプリング率を採用する。また、統計分析の結果の再現性を維持するため、リサンプリングされて提供されるサンプルは固定する。
 - b. 秘匿処理、サンプリング、リサンプリングの方法は公表する。
 - c. 個体識別の蓋然性が十分に低い水準にとどまり、研究上の必要がある場合、当該調査と他調査の接続を妨げず、接続に必要な情報を提供する。また、同一客体を継続して調査している調査については、同一客体の異時点間の接続を可能とするような情報を提供する。これら接続に必要な情報をデータセットに含めるよう設計段階から十分に工夫する。
 - d. 参考のために、賃金構造基本統計調査、労働力調査を題材に、匿名化データの作成基準の例を付録にまとめた。なお、これらの例は本提言における整備されるべきデータの優先順位を示唆するものではない。

7. 研究者の義務と罰則の明確化

- a. 利用に関しては、利用契約を締結することにより、利用者の義務を明示する。
- b. 利用契約は、国内研究機関に所属するものが匿名化処理の終了したデータを利用する場合には、研究者個人とデータ提供当局の間で結ぶものとする。それ以外の場合には、罰則の履行確保のため、研究者の所属機関とデータ提供当局の間で結ぶものとする。
- c. 提供を受けたデータを利用するコンピュータのセキュリティーの基準について、別途専門家の意見を参考に明文化し、利用審査の対象とする。
- d. 研究者の監督の下にリサーチアシスタントを雇用し、データ加工などの業務につかせることを妨げない。ただし、この際に雇用するリサーチアシスタントは申請の際に明記する。
- e. 利用期間経過後は原データは処分する。ただし、データ加工のためのプログラムファイルなどは研究者が自由に保持し、他の研究者に提供することを妨げない。
- f. 個人の特定を目的としたデータの利用を行ったものには、統計法に定められる罰則に服するほか、一定期間の利用停止の処分を行う。また、利用申請書と異なったデータの用い方を行ったものにも、利用停止の処分を行う。

8. その他

政府政策の立案・評価のためにマイクロデータは従来から利用されてきている。しかしながら、政策立案・評価のための研究が広く公募されることは少なく、質問項目や標本設計といった統計の設計についても広く意見が公募されることは少ない。これらは早急に第三者審査による公募

に基づくものへと変更されるべきである。

III. 付録

- ・ 匿名化データ使用申請書案
- ・ 賃金構造基本統計調査匿名化データ案
- ・ 労働力調査匿名化データ案

本提言の要望を満たす研究計画書の一例を、賃金構造基本統計調査を題材に提示する。他調査についても、同様の申請書案を審査機関が用意するのが望ましい。

匿名化データ使用申請書案

(1) 研究題目

(2) 利用する匿名化データの名称

賃金構造基本統計調査の匿名化データ(19XX年より20XX年まで)

(3) 利用者

〇〇大学経済学研究科教授 ××

△△大学経済学部教授 □□

(4) 利用場所と秘密保持に関する安全性

(1) 各所属大学の研究室内(下記参照)

〇〇大学経済学研究科 ××研究室内

△△大学経済学部 □□研究室内

(2) △△大学経済学部計算機サーバー

なお、(1)および(2)は別途定められている「匿名化データ使用環境基準(仮称)」を満たしている。

(5) 利用期間

承認の日から1年

(6) 利用目的と学術的意義

わが国における労働基準法制の主要な柱のひとつとして最低賃金制度がある。労働者は、使用者と比較して本来的に交渉力が弱いという一般的な認識の下、劣悪な労働条件が成立すると思われる職種や地域について、最低賃金法をもって賃金水準の

最低限を規定するのがこの制度の目的であり、経済学でいうところの価格規制の典型である。同様の制度は多くの国で採用されており、わが国でも1959年に最低賃金法が成立・施行された。

ところが、最低賃金制度が価格規制である以上、当該制度は労働者保護を謳いながらかえって雇用を減少させるとして長い間批判にさらされてきた。労働市場が完全競争的であると考え、均衡賃金以上に最低賃金額を設定すれば、その分労働市場で成立する雇用量は減少するからである。ただし、使用者に市場独占力が存在する場合には、均衡賃金が社会的に効率的な賃金水準を下回る可能性があり、この場合、最低賃金制度をもって賃金率を上昇させることは、かえって雇用を増大させる可能性がある。このように、最低賃金の雇用に対する効果を巡っては、理論的には大きく二つの可能性が存在する。それゆえ、実証によって決着がつけられるべく、諸外国では盛んに実証研究が進められてきた。

たとえば、米国では Card and Krueger (2000) など多くの論争が生まれ、いまだに決着が付いていない。また、これらの論争の過程で、独自のデータソースがいくつか開拓されており、労働政策の経済的効果を調べるうえで貴重なデータが蓄積されるのに大きな役割をになった。日本においても、とりわけパートタイム労働との関連で、最低賃金制度が雇用量に与える効果が議論されてきた。代表例は 永瀬 (1998) や小原 (2000) である。これらの研究では、最終的な雇用量に対する含意は定かではないが、概ね最低賃金の下支え効果が認められている。しかしこれらに対して、安部 (2001) はパートタイム実態調査の個票を用いた分析の結果「日本のパート労働市場において最低賃金ときわめて近い賃金を受け取る労働者が少ないということは、パート労働者の賃金決定に対して、最低賃金は直接的な影響は持っていないことを意味する」と結論している。結局、最低賃金のもつ経済的効果についてはわが国についてもよくわかっていない。

その一方で、規制改革の一環として最低賃金制度も見直しが議論されるようになり、2005年3月³¹日厚生労働省に設置された「最低賃金制度のあり方に関する研究会」が報告書をまとめた。そこでは「最低賃金制度の第一義的な役割は、すべての労働者を不当に低い賃金から保護する安全網としての「一般的最低賃金」としての役割であ」とし、最低賃金の決定方式のひとつであった産業別最低賃金を見直すように提言した。ただし、最低賃金の経済的効果については一般論を紹介するのみで、日本における現状は討議されていない。

本研究では、賃金構造基本統計調査(以下賃金センサスと表記)を用いることで、前年の最低賃金水準の変動が次年の雇用・賃金分布の変動と関係するかを都道府県レベルで観察する。このような直接的な把握を通じて、最低賃金のもつ経済的効果について一定の実証的見解を提出できる。

具体的には、各年各都道府県別の所定内時給分布を計測し、パーセンタイルなど

分布の形状を示す特性値を把握する。次に地域別最低賃金と比較し、最低賃金以下の所定内時給にとどまっている標本割合を確認する。そして、翌年の各都道府県における所定内時給分布を計測し、前年と分布の形状、特性値の変化を考察する。

(7) 公表数表では不可能な理由

公表数表では最低賃金以下の標本割合がわからない。また、賃金センサスの公表数表は賃金月額の絶対水準を明らかにすることに重きをおいているため、(所定内)賃金額の平均と(所定内)労働時間の平均が別個に記載されている。したがって、(所定内)時給の正確な分布を知るためには調査個票を用いる必要がある。

(8) 匿名性確保への配慮

匿名化データを単体で用いるので、分析の段階で特に匿名性確保へ配慮する必要はない。

(9) 結果の公表方法

分析結果は XX ディスカッションペーパーシリーズとして公表する。

(10) その他必要な事項

- (1) 利用者連絡先
- (2) 所属先ネットワーク責任者連絡先

賃金構造基本統計調査の匿名化措置に関する提案

1. 賃金構造基本統計調査の提供の重要性

- 個別賃金を広範囲に収集し、賃金格差や賃金水準などの議論に欠かすことのできない重要な資料である。
- 賃金額は、世帯調査では、収入のうち労働所得として質問される。通常、被調査者本人の記憶・記録などの観点から、把握精度は落ちると考えられる。本調査では、賃金額は事業所の人事担当者が賃金台帳をもとに答えており(月額 100 円単位)、少なくとも所定内賃金については、本調査以上の精度をもった統計は存在しないと思われる。
- 同時に前年度ボーナスや労働時間も質問しており、時間あたり賃金率に関するもっとも信頼できる指標を提供している。
- また、毎年 100 万人を超える標本を確保しており、諸外国の賃金統計とは比較にならないほど大きなサンプルを有し、最低賃金など通常の世帯調査では正確に捕捉できない部分なども検討対象とできる点も特徴のひとつである。

2. 匿名化の基本方針

- 本調査は事業所調査であるが、本質的には個人調査であり、個人票を匿名データへの変換・提供は可能であると考ええる。
- 本調査のメリットを生かすためには、賃金構造の全体的な把握を行えるように、賃金額や内訳の精度とサンプルサイズを確保することが重要である。

3. 具体的な匿名化措置

1. 再抽出確率

- (a) 90%抽出(できれば全数)

2. 事業所情報

- (a) 都道府県(47 分類)
- (b) 産業大分類(9 分類)
- (c) 企業規模(8 分類)または事業所規模(10 分類)
- (d) 公営民営の別

3. 個人情報

- (a) 性別(2 分類)

- (b) 労働者の種類(2分類)
- (c) 雇用形態(2分類)
- (d) 就業形態(2分類)
- (e) 学歴(4分類)
- (f) 年齢(17歳以下、65歳以上をコーディング)
- (g) 勤続年数(42年以上をコーディング)
- (h) 実労働日数
- (i) 所定内実労働時間数
- (j) 超過実労働時間数
- (k) 決まって支給する現金給与(0.2パーセンタイル以上でコーディング)
- (l) 超過労働給与
- (m) 所定内給与(0.2パーセンタイル以上でコーディング)
- (n) うち手当で総額
- (o) 賞与等特別支給額(0.5パーセンタイル以上でコーディング)

労働力調査特別調査(平成13年以前)及び労働力調査(平成14年以降)の匿名化

措置に関する提案

1. 労働力調査の提供の重要性

- 家計調査であるため、非就労者も含めた人々の情報が入手できる。これ自体で就労・非就労の意思決定に関する分析が可能である。少子高齢化社会の到来により、労働力人口の減少が懸念される中、女性、高齢者、若年者の就業行動がどのように決定されるかを分析することは重要である。
- 15歳以上の家計の構成員の情報が入手可能であり、家計構成が時系列でどのように変化したかの分析が可能である。
- ローテーションサンプリングとなっており、同一世帯が2ヶ月調査され、その後10ヶ月空き、さらに2ヶ月調査されているため、2年分のパネルデータとすることが潜在的には可能である。
- 毎月約4万世帯を超える標本を確保しており、昭和42年以降、ほぼ一貫して同一の質問項目が含まれているため日本経済の長期にわたる変化を捉えることができる。

2. 匿名化の基本方針

- 本調査は世帯調査であり、匿名データへの変換・提供は比較的容易に可能であると考ええる。
- 本調査のメリットを生かすためには、世帯構成やパネル構造を生かせるように、世帯ごとに4調査時点をひとつの観察値として取り扱ったりサンプリングを行う必要がある。
- パネル化のメリットは大きいため、異時点間の情報をつなぎ合わせるができるような情報提供が行われるべきである。

3. 具体的な匿名化措置

1. 再抽出確率

- (a) 90%抽出(できれば全数)

2. 各変数の秘匿処理

別紙のとおりであるが、年齢のトップコーディングを施すことで米国の Current Population Survey 並みの秘匿処理を施すことができると考えられる。