

Discussion Paper No.267

諸外国における政府統計データの提供の動向について

中央大学経済学部

伊藤 伸介

June 2016



INSTITUTE OF ECONOMIC RESEARCH
Chuo University
Tokyo, Japan

諸外国における政府統計データの提供の動向について

中央大学経済学部

伊藤 伸介

1. はじめに

政府統計(公的統計)データにおいては、秘匿性と利用者のニーズを踏まえた形で多様な提供形態が存在する。政府統計は、統計表およびマイクロデータという形で利用可能であり、政府統計のマイクロデータについては、主として、①匿名化マイクロデータ(個票データに匿名化処理が施されたデータ、anonymized microdata)の提供、②個票データ(original data, deidentified data)の提供、③オーダーメイド集計(tailor-made tabulation)、④オンデマンド型の提供サービス(リモート集計、remote execution)といった様々な形態による提供が進められてきた。諸外国では、このようなデータの提供形態を可能にするための統計法制度の整備と技術的な措置が展開されてきた。

わが国においても、政府統計データにおける利活用のさらなる促進に向けて、統計表(オープンデータ)とマイクロデータの両面から、様々な形態での提供が行われている。一方では、統計データにおける有効活用の推進という観点から、(1)API(Application Programming Interface)機能による利用環境の整備や、(2)統計 GIS 機能の整備によって、政府統計におけるオープンデータの高度化が推進されてきた。他方で、わが国においては、統計法(平成19年法律第53号)の全面施行によって、2009年4月より、政府統計のマイクロデータの提供が展開されてきた。国勢調査、就業構造基本調査等の7つの統計調査に関する匿名データが作成・提供され、様々な政府統計の調査票情報(個票データ)が広範に提供されているだけでなく、そのために、匿名データおよび調査票情報の利用におけるガイドライン等の法制度的な整備も図られてきた。さらに、第Ⅱ期「公的統計の整備に関する基本的な計画」(平成26年3月25日閣議決定)では、調査票情報の提供におけるリモートアクセスを含むオンサイト利用やプログラム送付型の集計・分析の実現に向けた整理・検討の必要性を指摘している。こうした状況を踏まえ、わが国では、諸外国で広範に展開されているオンサイト施設による個票データの利用とリモートアクセスによる個票データの提供サービスの可能性について、現在検討を進めている(伊藤(2016a))。

本稿では、諸外国における政府統計データの様々な提供形態についてその動向を明らかにする。とくに、個票データを用いて得られる出力結果の外部への持ち出し基準を踏まえつつ、個票データのアクセスの可能性について検討する。その上で、諸外国の事例をもとに、わが国における政府統計データの提供における将来的な方向性を探ることにしたい。

2. 欧米諸国における政府統計マイクロデータの提供の動向

欧米諸国における政府統計マイクロデータの主要な提供形態は、①匿名化マイクロデータ、②オンデマンド型の提供システム、③オンサイト施設による個票データのアクセス(on-site access)、④リモートアクセス(remote access)による個票データの利用であって、データの秘匿性と有用性の両面から、政府統計のマイクロデータの提供に関して多様なチャンネ

ルが存在する。その一方で、政府統計マイクロデータの提供における法制度的・技術的措置の観点から、各国における政府統計マイクロデータの提供状況は、個別具体的には異なる展開を示している。

本節では、政府統計マイクロデータの提供形態である、匿名化マイクロデータの提供、オンデマンドの集計形式による提供、オンサイト施設による個票データの利用、およびリモートアクセスについてその特徴を述べることにしたい。

諸外国では、人口・世帯系の統計調査を中心に、匿名化マイクロデータが作成・提供されている。匿名化マイクロデータの作成に関する特徴は、同一の調査に対して複数のファイルが作成・提供されることにある。例えば、ドイツ、イタリア、イギリス、オランダといった国々では、一般公開型ファイル(public use file)と学術研究用ファイル(scientific use file)の2種類のファイルが作成・提供されていることが知られている¹。一般公開型ファイルと学術研究用ファイルの作成においては、秘匿性に関する閾値を設定した上で、その閾値を超えない形で、様々な匿名化技法が適用されることが考えられる。一般公開型ファイルの場合、学術研究用ファイルと比較してより低い閾値が設定されることが想定される。その一方で、適用される匿名化技法については、データの特性や秘匿の程度に応じて、データの削除(レコード削除、変数の削除、suppression)、リコーディング(recoding)、トップ(ボトム)コーディング(top(bottom)coding)といった非攪乱的手法だけでなく、スワッピング(data swapping)、ノイズ付加(noise addition)、丸め(rounding)、マイクロアグリゲーション(microaggregation)、PRAM(Post Randomization Method)といった攪乱的手法が用いられる(伊藤・星野(2014))。

一般公開型ファイルについては、アメリカでは、人口センサス、アメリカン・コミュニティ・サーベイ(American Community Survey)、経常人口調査(Current Population Survey)といった様々な政府統計を一般公開型マイクロデータ(Public Use Microdata Sample)として無料でダウンロードすることができる。アメリカセンサス局は、非攪乱的手法として、地理的な閾値の設定(geographic thresholds)、カテゴリカルな属性における分類区分の閾値(categorical thresholds)、トップ・コーディングを行うだけでなく、攪乱的手法としてラウンディング(伝統的な丸め, traditional rounding)、ノイズの付加、スワッピングを行っている(Zayatz(2007), Lauger *et al.*(2014), 伊藤(2015))。さらに、2006年のアメリカン・コミュニティ・サーベイからは、施設等の世帯において一部合成データ(partially-synthetic data)の考え方に基づく秘匿処理が適用されている(Rodríguez(2007))。また、カナダにおいては、カナダ統計局が、主として非攪乱的手法を適用することによって、人口センサスや労働力調査等、数多くの政府統計に関する一般公開型マイクロデータファイル(Public Use Microdata Files=PUMFs)を作成・提供しているが、PUMFsの作成において、攪乱的手法が適用される場合もある(赤谷・荒川・伊藤(2014))。イギリスでは、一般公開型ファイルの作成・提供は最近まで行われていなかったが、2011年人口センサスの教育用マイクロデータが public use file として公開されている(伊藤(2014))。一方、オランダでは、 μ -Argusのようなマイクロデータの秘匿処理用のソフトウェアを用いて、オランダ

¹ ドイツにおけるマイクロデータの提供事例については小林(2011)を参照されたい。また、オランダにおける複数のマイクロデータファイルの提供については、Nordholt(2013)を参照。

統計局が **public use microdata files** を作成している(Nordholt(2013))。なお、Eurostat では、合成データ(**synthetic data**)の方法論を用いて **public use data** の作成に関する研究が行われている(de Wolf(2015))。

一方、学術研究用ファイルについては、イギリスのライセンス型マイクロデータ(**End User Licensed Data**)のように、一般公開型ファイルとは異なり、ライセンスの取得や誓約書を提出することによって、匿名化マイクロデータが提供されている事例がある。また、オーストラリアやニュージーランドでは、一般公開型ファイルが作成されておらず、匿名化マイクロデータファイル(**Confidentialised Unit Record Files=CURFs**)のみが提供されているが、オーストラリアの場合、秘匿の程度に応じて、基本ファイル(**Basic CURFs**)と詳細ファイル(**Expanded CURFs**)の複数のファイルが作成されているのが特徴的である。

世帯・人口系の統計調査に関する匿名化マイクロデータにおいては、世帯単位で抽出されるか、あるいは個人単位で抽出されるかによって、類型化が可能である。具体的には、世帯単位で抽出されたマイクロデータファイルは階層型ファイルになっており、世帯構成員の属性が利用可能であるのに対して、個人単位で抽出されたマイクロデータファイルにおいては、詳細な地域区分が利用可能になっているだけでなく、個々人の社会的・経済的な属性についてもより詳細な分類区分が提供されている。例えば、カナダの場合、2006年の人口センサスに関しては、カナダ統計局が、個人ファイル(**Individual File**)と階層ファイル(**Hierarchical file**)の2種類のPUMFsを作成している。こうした特徴は、カナダだけでなく、アメリカ、イギリスといった国々における人口センサスのマイクロデータにおいて指摘することができる(伊藤(2015), 伊藤(2016b))。一方、オーストラリアの場合、世帯単位で抽出された、抽出率の異なる(1%, 5%)人口センサスのマイクロデータ(**Census Sample Files (CSFs)**)が作成・提供されている。

他方、作成される匿名化マイクロデータの種類は、世帯・人口系のデータと事業所・企業系のデータに大別することが可能である。一般に、統計作成部局によって作成・提供されている匿名化マイクロデータの多くは、人口センサスや労働力調査等、世帯・人口系のデータであって、事業所・企業系の統計調査については、匿名化マイクロデータが作成されることは非常に少ない。それに対して、イタリアのように、企業データである **Italian innovation survey** が、一般公開型ファイルと学術研究用ファイルの両方で提供されている事例が存在する。さらに、Eurostatにおいても、企業データである **Community Innovation Survey** を学術研究用ファイルとして提供している。

なお、学術研究用ファイルの提供方法としては、匿名化マイクロデータが入った **CD-ROM** を配布する方法とダウンロードによって入手する方法が存在する。**CD-ROM** を提供している国としては、オーストラリア、ニュージーランド、カナダ、Eurostat 等がある。一方、ダウンロードによって匿名化マイクロデータが提供されている国としては、イギリスやイタリア等がある。例えば、イギリスでは、UKDA 内の組織の1つである **UK Data Service** が、政府統計のライセンス型マイクロデータ(**End User Licence Data**)に関する提供サービスを行っている。ライセンスを取得することによって、研究者や学生は、学術研究目的のために、労働力調査等のサーベイマイクロデータ、人口センサスの匿名化標本データ(**Samples of Anonymised Records**)といったライセンス型マイクロデータを **UK Data Service** のHP上でダウンロードして利用することが可能になっている。

諸外国では、オンデマンド型の集計システムの開発と実装が進められてきた。例えば、アメリカでは、センサス局が開発した American FactFinder と Data Ferrett が利用可能であることが知られているが、近年では、回帰分析の実行も可能な Microdata Analysis System というリモート集計システムの開発が進められている(Zayatz(2007), Lucero and Zayatz(2010), Lauger *et al.*(2014))。一方、カナダ統計局では、セキュアな場所に保管されているマイクロデータに対して、利用者がインターネットを通じてリモートでアクセスすることが可能な「リアルタイム・リモートアクセス (Real Time Remote Access)」の提供サービスを行っている(赤谷・荒川・伊藤(2014))。

つぎに、オンデマンド型の提供システムとしては、2つの論点を指摘することができる。第1の論点は、オンデマンド型の提供による出力結果についてである。オンデマンド型の提供システムの場合、リモート集計だけでなく、モデル分析も行うことができるかどうかによって出力結果の内容も異なる。例えば、オーストラリア統計局が開発した Table Builder は、集計表のみを出力するオンデマンド型のシステムである。それに対して、イギリスでは、オンデマンド型集計システムとして NESSTAR が用いられているが、NESSTAR では、変数を指定することによって、集計表を作成するだけでなく、重回帰分析といったモデル分析も可能である。第2の論点は、オンデマンド集計用システムのために保管されるデータのタイプについてであって、それは、「超高次元クロス集計表(データキューブ)」を含む集計表(伊藤(2009))とマイクロデータに大別される。前者について言えば、オランダ統計局が開発した StatLine では、保管用データとしてのデータキューブからオンデマンドで集計を行っている(森(2009,20頁))。また、アメリカセンサス局が提供サービスを行う American FactFinder は、保管用データとして集計結果表を備えている。他方、後者に関しては、例えば、アメリカセンサス局の Data Ferrett は、個票データを保管用データとして備えており、個票データに含まれる変数から集計表の集計事項を探索的に選択することが可能になっている。他方、イギリスで用いられている NESSTAR については、保管用データとして、人口センサスや標本調査の匿名化マイクロデータ(ライセンス型データ)を使用していることが知られている。

政府統計の個票データについては、インターネットが切断され、外部から遮断されることによって、セキュアな状態を保っているオンサイト施設の内部において、個票データの利用が可能になっている。具体的には、オンサイト施設のサーバに個票データが保管され、利用が容認された研究者(アメリカやカナダの宣誓職員(みなし職員)制度やイギリスにおける「承認された研究者」等)のみが、安全な環境のもとで個票データの利用サービスを受けることができる。例えば、イギリス国家統計局(The Office for National Statistics=ONS)の Virtual Microdata Laboratory(VML)のようなオンサイト施設において、政府統計の個票データが利用可能となっており、人口センサスの個票データや、1971年~2011年の人口センサスの個票データを対象に縦断的なリンケージを施した LS データ(ONS Longitudinal Study of England and Wales)等に、VML 内部でアクセスすることが可能になっている(伊藤(2016))。

オンサイト施設の場所については、イギリスの国家統計局やカナダのように統計作成部局に設置されている場合だけでなく、大学等(アメリカ、カナダ等)に設置されているリサーチデータセンター(Research Data Center)やその他の研究機関(ex, IAB(ドイツ))にオン

サイト施設が設置されている場合もある。例えば、アメリカでは、センサス局の経済研究センター(Center for Economic Study)を拠点として、全国 20 か所の大学等に連邦統計研究データセンター(Federal Statistical Research Data Centers=RDCs)が設置されている。また、利用可能な個票データの数と種類も国によって異なる。アメリカの RDCs においては、世帯・人口系の統計調査だけでなく、事業所・企業系の統計調査の個票データのアクセスが可能になっている。一方、カナダの場合、大学等の教育機関に設置されているリサーチデータセンターでは、世帯・人口系のデータしかアクセスすることができず、事業所・企業系のデータについては、カナダ統計局内部のオンサイト施設でのみアクセスすることが可能である。また、カナダやイギリスでは、オンサイト施設で行政記録データのアクセスが可能になっている(赤谷・荒川・伊藤(2014))。とくに、イギリスでは、2014 年に Administrative Data Research Network が創設されたことによって、学術研究や政策評価のための行政記録データの利用サービスが展開されており、イギリス国内にある Administrative Data Centre のオンサイト施設において、複数の行政記録データ、さらには統計調査の個票データと行政記録データにおけるリンケージデータの利用が可能になっている(伊藤(2016b))。

近年では、欧米諸国を中心に、リモートアクセスによる個票データの利用が広範に展開されている。例えば、エセックス大学の UK Data Service において、ESRC(=Economic and Social Research Council)の資金提供を受けて創設されたリモートアクセス施設である Secure Lab では、研究者が個票データに 24 時間リモートでアクセスすることが可能になっている(Afkhamai(2013), 伊藤(2014, 2016b))。なお、イギリスにおいて、大学の研究室から Secure Lab にアクセスする場合、大学の研究室にカメラは設置されていないものの、研究室で使用する PC は所属機関の IP アドレスで管理されている。また、Secure Lab においては、中間生成物(intermediate output)の持ち出しをすることはできないが、Secure Lab には Microsoft Office が利用可能であることから(UK Data Service(2014))、リモートアクセス上で分析結果に基づいてペーパーを書くことができる。そして、チェック済みの成果物(final output)については、持ち出しが可能であって、E メールで研究者に送付することも可能になっている。

ところで、リモートアクセスに基づく個票データの利用は大きく 2 つに類別することができる。第 1 の利用は、ルクセンブルク所得調査(Luxemburg Income Study)のように利用者が分析用のプログラムをサーバに送付し、プログラムに基づく個票データの分析結果が返ってくるタイプ(いわゆる「プログラム送付集計型」)である。これについては、利用者が個票データを直接見ることはできない。第 2 の利用は、大学の研究室からリモートアクセス施設のサーバに保管されている個票データにアクセスすることが可能であり、個票データを閲覧しながら、集計表の作成やモデル分析を直接行うタイプで(いわゆる「直接利用型」)、近年では、このタイプのリモートアクセスがヨーロッパ諸国の多くで広範に展開されている。こういったタイプのリモートアクセスが展開されている国としては、イギリス、ドイツ、フランス、イタリア、オランダ、スウェーデン、フィンランド、デンマーク、スロベニア、オーストラリア、ニュージーランド等のヨーロッパを中心とする多くの国を指摘することができる。なお、リモートアクセスの管理については、例えば、オランダやオーストラリアのように、統計作成部局がリモートアクセス施設を管理していることが少な

くないが、イギリスのエセックス大学にある UK Data Service の Secure Lab のように、大学がリモートアクセスを管理している場合もある。

3. 政府統計の個票データの提供と分析結果の出力について

近年、わが国ではオンサイト施設やリモートアクセス施設における政府統計の個票データの利用のあり方が議論されているが、個票データの利用後に利用者が「安全な分析結果」を得るためには、集計表や回帰分析の結果をどのようにチェックするかについての具体的なガイドラインが求められている。本節では、Eurostat が報告書としてまとめた『マイクロデータ研究に基づく出力結果のチェックに関するガイドライン(Guidelines for the Checking of Output Based on Microdata Research)』(以下『ガイドライン』と呼称)(Brandt *et al.*(2010))を例に、政府統計の個票データの提供と分析結果の出力のあり方について論じることにしたい。

統計作成部局が政府統計マイクロデータを提供するためには、マイクロデータの提供を可能にするための秘匿性の確保が要件として求められる。そのための枠組みとしては、「侵入者モデル(Intruder Model)」と「人間モデル(Human Model)」が考えられる。侵入者モデルとは、データに含まれる秘匿情報を故意に漏洩しようとする侵入者(intruder)を想定したモデルである。具体的には、侵入者が外部情報を用いて、政府統計マイクロデータに含まれる個人情報に漏洩する最悪なシナリオを想定した上で、露見リスクを回避するために、秘匿の程度を高めるための様々な匿名化技法をマイクロデータに適用することによって可能になる。統計作成部局が、匿名化マイクロデータを作成する場合には、このような侵入者モデル(Intruder Model)を伝統的に用いてきたと思われる。その一方で、侵入者モデルを適用することによって作成された匿名化マイクロデータは、利用者にとっては想定以上に有用性が低い可能性がある。そこで、近年、主にヨーロッパ諸国の統計作成部局で議論されている枠組みが、人間モデルである。人間モデルによれば、人間(human)は秘匿情報を漏洩する意図はなくても、誤って個人情報に露見される可能性が想定されている。人間モデルは、統計作成部局が個票データを提供するために適用される枠組であるが、この人間モデルを適用する上では、個票データの提供システムに関する法制度的な整備と個票データを利用する研究者の訓練が求められる。そのため、人間モデルは、Five Safes Model として体系付けられる(Ritchie(2008), Desai *et al.*(2016))。

Five Safes Model は、当時イギリス国家統計局の職員であった Felix Ritchie 博士によって 2003 年に考案されたモデルである。VML Security Model と呼ばれることもある。現在、主にヨーロッパ諸国においてオンサイト施設やリモートアクセスによる個票データの提供のために、Five Safes Model が採用されている(Ritchie(2008), Desai *et al.*(2016), 伊藤(2016b))。

Five Safes model は、個票データのアクセスに関する①安全なプロジェクト(safe projects)、②安全な利用者(safe people)、③安全なデータ(safe data)、④安全な施設(safe settings)と⑤安全な分析結果(safe outputs)の 5 つの基準から構成される。第 1 の安全なプロジェクトとは、妥当な統計目的のために個票データのアクセスが行われることである。第 2 の安全な利用者とは、個票データを適切な利用手続きに従ってデータを利用すること

によって、信頼された研究者だと認識されていることである。第 3 の安全なデータとは、データそれ自体が、個人情報が見えない安全なデータとして位置付けられることである。第 4 の安全な施設とは、個票データのアクセスに関して技術的な管理措置を施すことによって、容認されていないデータの移動が回避できることである。そして、第 5 の安全な分析結果とは、統計分析の結果に個人情報が露見されるような結果が含まれないことである。この分析結果に関する最終成果物については、オンサイト施設やリモートアクセス施設の審査担当者がチェックを行った上で公表することが可能になる(安全な利用(safe use))。

分析結果に関する最終成果物のチェックについては、①Rule-Based Approach と②Principles-Based Approach の 2 つの方法がある(Ritchie and Welpton(2015))。第 1 の Rule-Based Approach では、一連のルールにしたがって、分析結果が公表可能かについての判断を行う。そのルールについては、閾値ルール(threshold rule)や占有ルール(dominance rule) (ex. p%ルール等)といった集計表の一次秘匿(primary disclosure)や二次秘匿(secondary disclosure)に関する秘匿のルール(confidentiality rule)が用いられる。例えば、閾値ルールにしたがって、セルに含まれる度数を 3 以上にするといったルールがそれである。こうした Rule-Based Approach の利点としては、単純な方法であるが、分析結果のチェックにおいてあいまいさがないことが指摘される。こうしたルールに基づけば、分析結果のチェックに関する自動化も可能なように思われる。他方、Rule-Based Approach の欠点としては、ルールに従った結果、「秘匿の誤り(Confidentiality Error)(秘匿が十分でないこと)」や「有用性の誤り(Efficiency Error)(過剰な秘匿を行うこと)」を導く可能性が指摘されている。

一方、第 2 の Principles-Based Approach の特徴としては、原則として分析結果の公表の可否は、チェックを行う担当者に委ねられていることにある。研究者と分析結果のチェックを行う担当者との交渉によって、分析結果の公表の可否が決定される。具体的には、Principles-Based Approach の場合、分析結果が公共の利益(public benefit)に資するかどうかを考慮した上で、個票データの提供者側と利用者側の協力のもとで秘匿性が担保される。こうしたことから、研究者は良い分析結果を出す責任を有するが、その一方で、チェックを行う担当者も分析結果の意義を十分に考慮することが求められる。

イギリスといったヨーロッパ諸国のオンサイト施設やリモートアクセス施設では、Principles-based approach に基づいて得られた「安全な分析結果」を個票データの利用者に提供している事例が少なくない。こうした Principles-based approach における実践的原則として考案されたのが、rule of thumb モデルである。このモデルは、秘匿の誤りを減らし、効率性の誤りを減らすモデルとして設定される。『ガイドライン』によれば、rule of thumb モデルの全般的な基準(overall rule of thumb)として、つぎの 4 つの基準が提唱されている。

1. 全ての結果表において、セルの度数が 10 以上であること
2. 全てのモデルにおいて、自由度が少なくとも 10 あること
3. 全ての結果表において、特定のセルの度数が、それを含む行 (row)ないしは列(column)の合計の 90%を超えないこと
4. 全ての結果表において、特定のセルの度数が表全体の総計の 50%を超えないこと

これらの基準は、主として閾値ルールや占有ルールといった集計表の秘匿ルールに基づいていると考えられるが、rule of thumb モデルの場合、上記の基準が厳格に適用されるのではなく、あくまで分析結果の安全性を判断するための「目安」となっている点が重要である。

Principles-based approach の利点は、分析者が出力結果を持ち出すための交渉の過程であらゆるリスクを取り除くことが可能であり、かつ有用性を高めることができる点である。さらに、Principles-based approach によって、研究者に秘匿に対する認識を高めることが可能なことも指摘されている。他方、Principles-based approach の欠点としては、審査担当者によって、分析結果のチェックに不確実性があることから、チェックにおいて一貫性がないことが考えられる。また、Principles-based approach の場合、審査担当者がチェックを行うための技術と経験が要求されることから、分析結果のチェックに関するリソースが求められよう。具体的には、秘匿に関する専門的な知見と計量分析に関する技術的な知識の両方を備えている必要があることから、計量分析の経験をもつ職員がチェックを行うことが望ましいと思われる。一方、分析結果のチェックを行う担当者については、経験者であっても、秘匿措置に関する専門的な知識を習得するための訓練が必要である。そこで、Eurostat では、『ガイドライン』に基づいてチェックを行う担当者のための講習会が行われている。

ところで、『ガイドライン』では、集計結果表(度数表、数量表)、記述統計量(平均値、分散、最小値、最大値等)、相関係数や回帰分析の結果等、統計のタイプによって分析結果が「安全な」統計(safe statistics)かどうかについての類型化がなされている。「安全な」統計量とは、分析結果の安全性が保証されない事例が例外的であることと定義される。そして、理論的な観点よりも、むしろ研究活動の実践的な経験に基づいて、(個人情報漏洩されるような)特定の状況が「ありえない」かどうかの判断がなされる。したがって、出力結果が安全かどうかの判断は、露見リスクに関する理論に基づくというよりもむしろ主観的な判断(subjectivity)によって行われる。

表1は、統計のタイプごとの「安全な」統計と「安全でない」統計の類型を示したものである。分析結果のタイプによって「安全な」統計か「安全でない」統計を類別することが可能である。度数表や数量表のような集計結果表については、閾値を下回る度数を含むセルが存在する可能性があることから、「安全でない」統計と判断される。ただし、重み付きの度数表については、安全な統計として位置付けられている。平均値、最大値、最小値、メディアンを含むパーセンタイルといった代表値についても、データの分布特性や特異値によって、個体が特定化されるリスクがあることから、「安全でない」統計に位置付けられる。一方、『ガイドライン』によれば、分散、共分散、尖度や歪度といった高次のモーメントについては、「安全な」統計と判断されている。つぎに、相関分析と回帰分析について見ていくと、回帰係数、推定値に関する要約統計量(R^2 等)および検定統計量(χ^2 値)については、「安全な」統計だと考えられている。また、相関係数も「安全な」統計に該当する。それに対して、推定値の残差については、残差から個体の属性値が推測される可能性があることから、「安全でない」統計として位置付けられる。

ところで、個票データの分析結果についてチェックを行う担当者は、具体的にはどういった考え方に基づいてチェックを行うのであろうか。それを具体的に図示したものが、「安

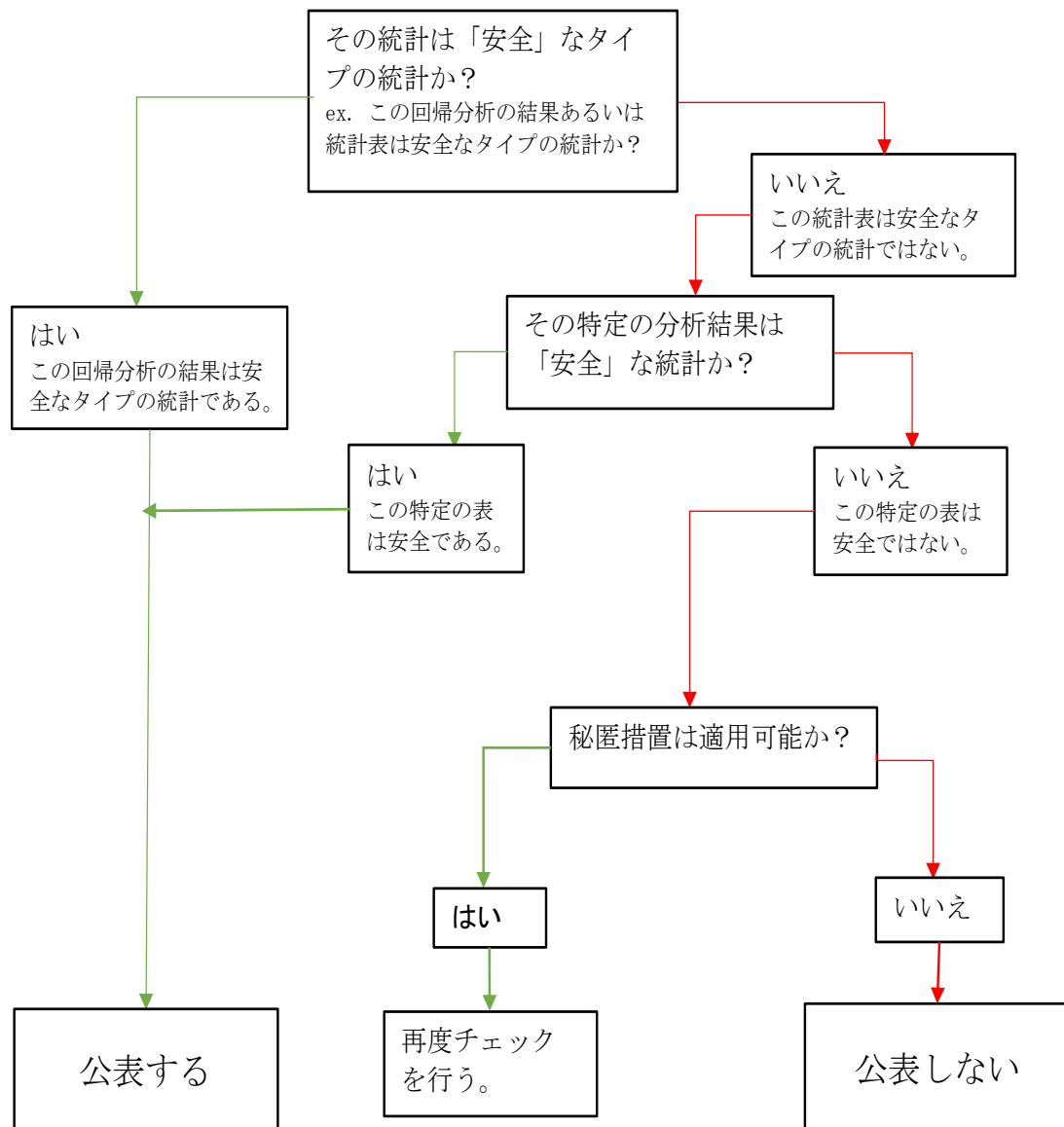
表1 「安全な」統計の類型

統計のタイプ	分析結果のタイプ	安全か安全でないか
	度数表 (Frequency tables)	安全でない
	数量表 (Magnitude tables)	安全でない
	最大値, 最小値, パーセンタイル(メディアンを含む)	安全でない
	最頻値	安全
	平均値, 指数, 比率, 指標	安全でない
	集中度 (Concentration ratios)	安全
	分布に関する高次のモーメント(分散, 共分散, 尖度, 歪度を含む)	安全
	グラフ, 実際のデータに関する絵入りの表現 (pictorial representation)	安全でない
相関係数と回帰分析	線形回帰モデルにおける回帰係数	安全
	非線形回帰モデルにおける回帰係数	安全
	推定値の残差	安全でない
	推定値に関する要約統計量および検定統計量 (R^2 , カイ2乗値等)	安全
	相関係数	安全

出所 Brant *et al.*(2010)

「安全な」統計の決定に関するチャート(decision tree)(図1)である。最初に、チェックの担当者は、利用者が個票データを用いて得られた統計に関する出力結果(集計表あるいは回帰分析の結果)が「安全」なタイプの統計かどうかを判断する。出力結果が回帰係数といった回帰分析の結果であれば、それは、表1に示されるような「安全な」統計のタイプに属することから、公表可能であると判定される。その一方で、出力結果が統計表である場合には、「安全でない」統計のタイプに属することから、つぎに、その特定の統計表のセルに含まれる度数や量的な数値が、「安全な」統計かどうか判断される。特定の集計表が「安全な」統計であれば、公表することが可能になるが、特定の集計表が「安全でない」統計と判定されると、集計表に対して秘匿措置が適用されているかどうか確認される。分類区分の統合といった非攪乱的手法やノイズの付与といった攪乱的手法の秘匿措置が適用されている場合には、再度集計表の公表が可能かどうかについてチェックを行うが、秘匿措置が適用されていない場合には、チェックの担当者は、最終的に出力結果の公表をしないという判断を行う。そして、担当者は、利用者に集計表の公表が可能なレベルまでの分類区分の統合等の秘匿処理に関する指導を行った上で、集計表の再作成を要請する。一方で、利用者側にとっては、出力結果の公表が容認されなかった場合には、集計表を再作成した上で、外部に公表するために再度チェックを依頼するという追加的な手続きが必要になる。個票データの使用期間の制約もあることを勘案すると、個票データの利用者は、出力結果の公表が可能なレベルを事前に想定した上で、集計表を作成することが求められよう。

図1 「安全な」統計の決定に関するチャート(decision tree)



出所 Ritchie and Welpton(2015)

4. むすびにかえて—わが国における公的統計データの提供における将来的な方向性—

本稿では、諸外国における公的統計に関するデータ提供の動向を洞察した。これまで、欧米諸国では、匿名化技法が適用された匿名化マイクロデータが広範に作成・提供されてきたが、ヨーロッパを中心に **Five Safes Model** が展開されていることは、個票データの利用促進を図るための法的制度的な枠組が整備されたことを意味している。

一方、わが国では現在、リモートアクセスを活用したオンサイト施設における個票データの利用が議論されている。それは具体的には、統計法 33 条申出による個票データの提

供において、オンサイト施設やリモートアクセスにおける事後チェック型の調査票情報の利用方式(小林(2012))を検討することである。調査票情報の利用に関する現行のガイドラインのもとで政府統計の個票データの利用申請を行う場合、実証分析を行う上で最低限必要な調査事項(変数)のみの利用が想定されている。事後チェック型の調査票情報の利用によってさらなる探索的な研究が可能になれば、個票データの利用者側だけでなく、提供者側にとっても有益であると考え、集計表や回帰分析の結果についてどの程度持ち出しが可能かについては、わが国でも今後議論が必要になる。その意味では、主にヨーロッパで展開されている **Five Safes Model** は、わが国においても大いに参考になる有益なモデルだと言えよう。また、**Five Safes Model** の具体的な形として、ヨーロッパでのオンサイト施設やリモートアクセスにおける個票データの利用においては、『ガイドライン』をもとに、「秘匿の誤り」や「有用性の誤り」を回避することを目指した **principle-based approach** および **rule of thumb** モデルが採用されている。このことから、わが国においても、事後チェック型の調査票情報の利用可能性を追究するために、出力結果の有用性と秘匿性の両面から **principle-based approach** の適用可能性を議論することが望ましいと考える。その場合には、諸外国でも議論されているチェック担当者の審査技術の養成だけでなく、個票データを分析する研究者の側においても秘匿措置に関する基本的な知識の習得が必要になるだろう。

ところで、わが国でリモートアクセス型の個票データのオンサイト利用を議論する上では、現時点では、24時間リモートアクセス施設が使用可能になる状況は想定されていないと思われる。一方、イギリスといったリモートアクセスが展開されている国では、中間的な成果物(中間生成物)については、24時間サーバ上でいつでも閲覧することが可能である。さらに、研究プロジェクトのメンバーであれば、サーバ上に保存されている中間的な分析結果をメンバー全員が閲覧することができる。したがって、わが国におけるリモートアクセスの整備における前提条件は、諸外国とは大きく異なると言える。このことは、わが国においてどういったレベルの中間生成物が利用者に提供可能かに関して、その法制度的な条件を含め、今後検討する必要があることを示唆している。

他方、わが国においてオンサイト施設やリモートアクセスによる個票データの利用についてさらなる展開を図ろうとすれば、「安全な」統計に関する基準を具体的に設定することが求められる。ヨーロッパの事例を踏まえると、一定のサンプル数があれば、回帰分析の結果は、残差を除けば安全な統計だと考えて良いと思われる。他方、記述統計量や集計結果表については、「安全な」統計の決定に関するチャート等を踏まえつつも、個別のケースに即した上で安全性のチェックの考え方を議論することが今後必要になるだろう。

謝辞

本稿の内容の一部は、筆者がイギリスの **University of the West of England** 等で実施したヒアリング調査(2015年8月25日~28日)の内容に基づいている。ヒアリング調査で貴重な情報をいただいた **Felix Ritchie** 准教授(**University of the West of England**)と **Richard Welpton** 氏(**Valuation Office Agency**(当時)、**Cancer Research UK**(現在))に謝意を表したい。

参考文献

- 赤谷俊彦・荒川智浩・伊藤伸介(2014)「カナダ統計局における政府統計データの提供の動向について」、『ESTRELA』No.241, (財)統計情報研究開発センター, 2～9 頁
- 伊藤伸介(2009)「匿名化技法としてのマイクログリゲーションについて」熊本学園大学『経済論集』第15巻第3・4号合併号, 197～232 頁
- 伊藤伸介(2011)「わが国におけるマイクロデータの新たな展開可能性について—イギリスにおける地域分析用マイクロデータを例に一」明海大学『経済学論集』第23巻第3号, 36～54 頁
- 伊藤伸介(2012)「政府統計マイクロデータの提供における匿名化措置—イギリス統計法における法制度的措置と攪乱的手法の適用可能性を中心に—」明海大学『経済学論集』Vol.24, No.3, 1～14 頁
- 伊藤伸介(2014)「イギリスにおける政府統計データの二次的利用の現状」『ESTRELA』No.241,10～20 頁
- 伊藤伸介・星野なおみ(2014)「国勢調査マイクロデータを用いたスワッピングの有効性の検証」『統計学』107号, 1～16 頁
- 伊藤伸介(2015)「人口センサスにおけるマイクロデータの作成状況について」、『統計』2015年1月号, 8～13 頁
- 伊藤伸介(2016a)「わが国における政府統計のデータシェアリングの現状と課題」『情報管理』, Vol.58, No.11, 836～843 頁
- 伊藤伸介(2016b)「政府統計におけるリモートアクセスと秘密保護について—イギリスを例に一」, 『経済学論纂(中央大学)』第56巻第5・6合併号, 1～19頁
- 小林良行(2011)「匿名データの教育目的利用に関する一考察」『統計学』第100号, 100～105 頁
- 小林良行(2012)「公的統計マイクロデータ提供の現状と展望: 一橋大学での取り組みをもとに」『日本統計学会誌』第41巻, 第2号, 401～420頁
- 森博美(2009)「オランダの社会統計データベースSSDについて」『経済志林』第76巻第4号, 5～28頁
- Afkhamai, R. (2013) “Statistical Disclosure Control Practice in the Secure Access of the UK Data Service”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-7.
- Brandt, M., Franconi, L., Guerke, C., Hundepool, A., Lucarelli, M., Mol, J., Ritchie, F., Seri, G., Welpton, R. (2010) *Guidelines for the Checking of Output Based on Microdata Research*, Final Report of ESSnet Sub-Group on Output SDC, Eurostat
- Desai, T., Ritchie, F., Welpton, R.(2016) “Five Safes: Designing Data Access for Research”, *Economics Working Paper Series* 1601, University of the West of England.
- De Wolf, P. P. (2015) “Public Use Files of EU-SILC and EU-LFS data”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Helsinki, Finland, pp.1-10.
- Lauger A., Wisniewski, B., McKenna, L. (2014) “Disclosure Avoidance Techniques at the

U.S. Census Bureau: Current Practices and Research”, Research Report Series (Disclosure Avoidance #2014-02), U.S. Census Bureau, pp.1-13.
https://www.census.gov/srd/CDAR/cdar2014-02_Disc_Avoid_Techniques.pdf

Lucero, J. and Zayatz, L.(2010)“The Microdata Analysis System at the U.S. Census Bureau”, Domingo-Ferrer, J. and Magkos, E.(eds) *Privacy in Statistical Databases UNESCO Chair in Data Privacy International Conference, PSD 2010 Corfu, Greece, September, 2010 Proceedings*, Springer, pp.234-248.

Nordholt, E.S.(2013) “Access to Microdata in the Netherlands: from a Cold War to Cooperation Projects”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-11.

Ritchie, F.(2008) “Secure Access to Confidential Microdata: Four Years of the Virtual Microdata Laboratory”, *Economic & Labour Market Review*, Vol.2, No.5, pp.29-34.

Ritchie, F. and Welpton, R. (2015) “Operationalizing principle-based output statistical disclosure control”, mimeo

Rodríguez, R.(2007) “Synthetic data disclosure control for American Community Survey Group Quarters” paper presented at Proceedings of the Survey Research Methods Section, American Statistical Association, pp.1439-1447.
<https://www.amstat.org/sections/srms/proceedings/y2007/Files/JSM2007-000430.pdf>

UK Data Service (2014) User Guide: For Your Secure Lab Account.

Zayatz, L. (2007) “Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update”, *Journal of Official Statistics*, Vol.23, No.2, pp.253-265.

【マイクロデータの提供に関する以下のHPについては2016年6月28日現在】

欧州統計局(Eurostat) <http://ec.europa.eu/eurostat/web/microdata>

イタリア統計局(ISAT) <http://www.istat.it/en/archive/public-use-micro.stat-files>

<http://www.istat.it/en/archive/microdata-for-research-purposes/>

オーストラリア統計局(ABS)

<http://abs.gov.au/websitedbs/D3310114.nsf/home/About+CURF+Microdata>

ニュージーランド統計局

http://www.stats.govt.nz/tools_and_services/microdata-access/confidentialised-unit-record-files.aspx#methods