

《論文》

etaddata, citation and similar papers at core.ac.uk

brought to you by

—国際マイクロ統計データベースの取り組みを中心に—

Current Efforts for Utilizing Foreign Official Statistical Micro Data
—From Efforts of the International Official Statistical Micro Database—坂田 大輔 Daisuke Sakata
鈴木 雄大 Takahiro Suzuki
櫻本 健 Takeshi Sakuramoto

Today, analyses of Asian countries' society and economy have become increasingly important; thus, the need for micro data on Asian countries' official statistics is also increasing. The International Official Statistical Micro Database is a data archive for micro data on some Asian countries' official statistics established by the Statistical Information Institute for Consulting Analysis (Sinfonica) and the Research Organization of Information and Systems in Japan. The database includes resampling official micro data on household income and expenditures for Southeastern and Southern Asian countries as well as a user's manual on these micro data. These micro data were supplied after resampling. Endeavoring to compile unique user's manuals is of particular significance. This article mainly provides a brief overview of the effort of this official statistical micro data archive.

Key words: Micro data, International Official Statistical Micro Database, Cambodia Socio-Economic Survey 2009, Thailand Household Socio-Economic Survey 2011, Nepal Living Standard Survey II 2003/04

キーワード: ミクロデータ, 国際マイクロ統計データベース, カンボジア 2009 年社会経済調査, タイ 2011 年世帯の社会経済調査, ネパール 2003/04 年生活水準調査

I はじめに

立教大学社会情報教育研究センター (CSI) 政府統計部会では、日本及び海外のマイクロ統計データの調査・研究に関心を払ってきた。本論文は 2015 年 12 月 16～21 日に開催された、第 7 回公的統計のマイクロデータ分析国際ワークショップ (以下、国際マイクロ WS) への坂田助教 (17～18 日)、鈴木学術調査員 (21 日) による出張報告がベースとなっている。

近年日本では国際マイクロデータの活用に研究者たちの注目が集まりつつある。特に、国際的にも注目されているアジア地域 (特に東南アジア) ではそのニーズが高まっている。

例えば 2015 年 4 月には、日本が援助してメコン川での建設を進めていたつばさ橋が開通し、ベトナム・カンボジア・タイが一つの地域圏になった。これまでメコン川沿いの国々は、農業を中心に別々の経済圏であったが、2015 年から一大地域圏として急速な経済成長が期待されている。地域圏としての成長が期待されるようになった結果、これまでデータが十分になかったメコン地域では統計データ整備の必要性が大いに高まっている。

大学共同利用機関法人情報・システム研究機構と (公財) 統計情報研究開発センターが連携協力協定を結び、2014 年 7 月から、日本国内で、上記のメコン地域諸国のマイクロデータをも含む国際マイクロ統計データベースが試行提供されるようになった。これは 2006 年に伊藤彰彦統計情報研究開発センター理事長、松田芳郎一橋大学・東京国際大学名誉教授らがアジア 7 か国と協定 (Charter) を交わしたことにより実現に向かった。この成果によって、アジア圏の公的統計のマイクロデータに対して、国内の研究者がこれまでよりも格段に容易にアクセスできる体制が整備された。

国際マイクロ統計データベースには、各国の政府統計のマイクロデータが提供されているという性質上、その使用に際して高いセキュリティレベルを確保するため、統計法第 33 条に

基づく政府統計の調査票情報のオンサイト利用も可能な情報・システム研究機構の統計数理研究所・オンサイト解析室内に限って利用が可能となっている。このため、分析に際して関東圏外の在住者等は、場合によっては泊まりも含めて統計数理研究所に通う必要があるという利用上の制約はある。しかしながら、これはデータの利用者が高度なセキュリティ環境を用意しなくても利用が可能ということでもある。

今後、本データベースを利用した統計分析事例がさらに示されるようになれば、アジア地域のマイクロデータに対する利活用を一層活性化させ、マイクロデータを用いた統計分析の高度化といった面でも好影響が期待出来る極めて画期的なマイクロデータアーカイブとなっていくであろう。

II 国際マイクロ統計データベース

1. 国際マイクロデータの利活用から見た日本の現状

マイクロデータの利用環境の整備は欧米が先行してきた。計 82 カ国から 1960 年以降のセンサスデータ 277 個（個体レコード数は約 6 億 1400 万）を収集しているミネソタ大学人口センター（Minnesota Population Center）の IPUMS-International (Integrated Public Use Micro-data Series, International) ¹⁾や 46 カ国から集められた所得支出関連マイクロデータを管理しているルクセンブルク国際データセンター（CROSS-NATIONAL DATA CENTER in Luxembourg : 旧称 Luxembourg Income Study）の LIS (Luxembourg Income Study Database) ²⁾、約 760 の大学、政府機関などが加盟している ICPSR (Inter-University Consortium for Political and Social Research) ³⁾などの大規模かつ国際的なマイクロデータアーカイブが多数存在している。

日本は欧米と比較してマイクロデータの利用環境の整備で後れを取っていたが、1998 年に東京大学社会科学研究所附属社会調査・データアーカイブ研究センターによって、SSJDA (Social Science Japan Data Archive) が構築されたのを皮切りに、複数の大学機関でマイクロデータのデータアーカイブが構築され、様々な研究機関により収集されたマイクロデータの蓄積が進んだ。公的統計の分野においても、2009 年の新統計法全面施行に伴い、公的統計の調査票情報に対して匿名化処置を行った匿名データの提供が開始された。統計法第 33 条により匿名化処置をされていないデータ（調査票情報）の利用も可能となっている。また、統計法の下で提供されているマイクロデータは中央省庁による統計調査に限定されているが、地方自治体による調査の調査票情報の二次的利用に対する試みも始まっている。神戸大学マイクロデータアーカイブ（KUMA : Kobe University Microdata Archive）では、神戸大学のオンサイト施設内での利用に限定されるものの、兵庫県が独自に行った統計調査の調査票情報を利用できるようになっている ⁴⁾。独立行政法人による調査データについても、厚生労働省所管の独立行政法人である労働政策研究・研修機構が、実施した調査の個票データに対して匿名化処置を施した上で、2009 年より JILPT データ・アーカイブを通じて提供する、といった試みがなされている ⁵⁾。

（公財）統計情報研究開発センターと情報・システム研究機構の協同によって進められている「国際マイクロ統計データベース」の整備もこうしたマイクロデータアーカイブ構築に関する取り組みの一つであるが、国内では珍しい海外のマイクロデータを収集管理しているデータアーカイブである。このデータベースの構築にあたっては、先述したように、一橋大学・東京国際大学名誉教授の松田芳郎氏、元統計局長の伊藤彰彦氏及び元総務省統計研修所長・国連アジア太平洋統計研修所(SIAP)講師の古田裕繁氏の尽力によるところが特に大きい。国際マイクロ統計データベースは、アジア版 LIS を目指し、複数の東南アジアおよび南アジアの国々から所得や支出に関する公的統計のマイクロデータの提供を受け、それらの利用環境を整備してきた。国際マイクロ統計データベースの大きな特徴として、各種メタデータ、すなわちマイクロデータを構成するデータの位置づけを示すデータ、に基づくユーザー・マニュアルの整備に力を入れている点が挙げられる。

国際マイクロ統計データベースに関する研究集会は、これまで何度か行われてきている。2015 年だけでも 3 月 15 日に統計数理研究所で行われた研究集会「アジア諸国の世帯統計マイクロデータの統合利用の研究」が挙げられるほか、9 月に行われた統計関連学会連合大会（岡山大学）でもセッションが組まれている⁶⁾。

以下では、筆者たちが参加した 2015 年 12 月に開催された第 7 回公的統計のマイクロデータ分析国際ワークショップの内容も踏まえて、国際マイクロ統計データベースの取り組みを概観していくこととする。

2. 国際マイクロ統計データベースの概要

国際マイクロ統計データベースとは「アジア諸国の政府統計（主に、世帯の家計に関する統計調査）の個票データ（マイクロデータ）を、各国政府からの提供を受けて、マイクロ統計データ活用研究会が世帯単位で再抽出した匿名標本データを収録したもの」⁷⁾である。

すでに述べたように、国際マイクロ統計データベースの整備への取組は、LIS に対して、アジア版 LIS を構築しようとする試みから始まった⁸⁾。2006 年に東アジア統計局長会議等のため来日していたアジア諸国の統計局長に対して『研究目的のためのマイクロデータの統計利用のための実験的研究所についての協定』を提示し、マイクロデータの提供に関する協力要請を行った結果、インドネシア、スリランカ、ネパール、バングラディッシュ、ベトナム、マレーシア、ラオスの計 7 か国の統計局長から署名が得られたこと、そして、各国から提供された所得分布関連世帯統計のマイクロデータとメタデータ（データレイアウトフォーム、コード表等）間の整合性の点検および修正や新規の分析などを目的とした「アジア地域貧困問題研究用の世帯統計マイクロ統計データベースの編成とその解析」（研究期間：2009 年～2012 年、研究代表者：伊藤彰彦（統計情報研究開発センター））⁹⁾が科学研究費補助金基盤研究（B）として採択されたことにより、国際マイクロ統計データベース整備への取組が本格的に開始された。その後、協定にはカンボジア、モンゴルが新たに署名し、タイも署名はしなかったが事実上これを容認した。現在は、2014 年に残念ながらマレーシアが外れ、9 各国が参加している。国際マイクロ統計データベースでは、これらの国々から提供されたデータを整備し、公開を進めている¹⁰⁾。

2015 年末現在において、国際マイクロ統計データベースで利用申請が可能なデータは、2006 年インドネシア社会経済調査（National Social Economic Survey）、2009 年スリランカ家計支出調査（Household Income and Expenditure Survey）、2007 年タイ世帯の社会経済調査（Household Socio-Economic Survey）、2006 年ベトナム生活水準調査（Household Living Standard Survey）の計 4 調査についての匿名化されたリサンプリングデータである。これらのマイクロデータについて最終的な利用が認められた場合、統計数理研究所（立川市）内のオンサイト利用施設内で利用することが出来る。データベースの利用申請者にはテスト用データも提供されている¹¹⁾。統計情報研究開発センターホームページ上では、上記マイクロデータについての様々なメタデータをもとに編集したユーザー・マニュアル（英文）の提供も行われており、このユーザー・マニュアルの整備が本事業の極めて重要な特徴となっている。

3. リサンプリングデータの作成について

個人識別に対するマイクロデータの安全性を高めるため、提供されたデータに対して再抽出が行われている。再抽出における抽出率は、これまでの国内での事例や先行研究から 80% に設定されている。再抽出は世帯単位での系統抽出（systematic sampling）によって行われている。これは、地域ごとの抽出対象となった世帯数の構成が再抽出前と変わらないようにするためである。与えられている世帯ウエイトに比例して抽出確率を変える確率比例抽出（probability proportionate sampling）で再抽出を行ってしまうと、人口密度が低い地域では相対的に世帯ウエイトが小さいため、その地域でサンプルサイズの減少が起こる可能性がある。再抽出後の世帯ウエイトとしては、元々の世帯ウエイトを 0.8 で除した値が

与えられている (Furuta, 2015a)。

4. メタデータとユーザー・マニュアルについて

本プロジェクトの最も重要な特徴であり、同時に最も困難な作業は、ユーザー・マニュアルの作成であろう。ここで重要になるのがメタデータの取得と整備である。マイクロデータを活用する上でメタデータと言え、データレイアウトと符号表 (data dictionary) がまず思い浮かぶ。しかし、実際にはこれのみで十分であることはまずなく、それ以外の幅広い情報が、マイクロデータに格納されている種々のデータの位置づけを掴む上で必要となる。

例えば、調査票である。調査票を通して、我々は調査の内容をかなり詳細に掴むことが出来る。しかし、大抵の場合は調査票が手に入っても全てを理解することは出来ないだろう。なぜならば、調査票の記入が他記式で行われている場合、記入の仕方は調査員に分かれれば良いわけであるから、この点についての記述が調査票には全くないか、必要最小限である場合が多い。したがって、調査票を理解するためには、調査員マニュアルが重要な資料となる。

調査員マニュアルが手に入れば、調査の仕方についての非常に詳細な情報を知り得るだろう。しかし、これは調査員マニュアルに限った問題というわけではないが、言語の問題がある。調査員マニュアルは、その目的上、一般に現地語のみで作成される。また、各国からの正式な提供は稀であり、概要的なものが英語で入手できた場合でも、英訳時に誤りが入り込む可能性は存在する。ユーザー・マニュアルの作成では、調査員マニュアルを含めた様々な資料がその編集においてメタデータとして利用されているが、何らかの誤りによって資料間での記述に矛盾が生じていた事例も指摘されている (The Institute of Statistical Mathematics and Statistical Information Institute for Consulting and Analysis, 2014 & 2015 and Furuta, 2015a)。

調査員マニュアルからもわからないことがある場合もあるだろう。ユーザー・マニュアルの作成においては、データチェックマニュアル、集計マニュアル等も重要な資料となるそうである。また、調査の実施側にとって自明であることでも、我々からは自明ではないといった事項が存在する可能性がある。Furuta (2015a) によれば、そうした暗黙知は日本の統計作成の現場においても数多く存在している。こうした問題についての解答は、資料のみから得ることは難しいだろう。次節で紹介する各国統計局の職員も参加してのワークショップは、こうした問題を解決する上で格好の機会を提供しており、ユーザー・マニュアルの質的向上に大いに貢献していると思われる。

Ⅲ 第7回公的統計のマイクロデータ分析国際ワークショップ

上述の「国際マイクロ統計データベース」整備の一環として、国際マイクロ WS が開催されてきた。今回著者たちは、2015年12月16日から21日に開催された第7回目の国際マイクロ WS に参加することが出来たため、以下では、第7回国際マイクロ WS での取組について紹介する。

1. ワークショップの概要

第7回国際マイクロ WS は、統計数理研究所において、情報・システム研究機構データ中心科学リサーチコモンズ「人間・社会データ」プロジェクトと統計情報研究開発センターの共催で開催された。国際マイクロ WS は、2010年3月に第1回が開催されて以降、毎年開催されてきた (2010年のみ2回開催)。第5回より、今回の第7回国際マイクロ WS と同様の形式で行われるようになっていく。

第7回国際マイクロ WS では、プロジェクトやマイクロデータ全般に関する報告、各国のマイクロデータを使った分析結果の報告などがなされただけでなく、招待されたカンボジア、タイ、ネパールの統計局職員より、提供されたデータの紹介、および事前に送付されてい

た質問状とフロアからの質問に対する回答が行われた。これにより、現在作成中の今回使用されたマイクロデータに対するユーザー・マニュアルの改善が図られるであろう。そしてさらに、今回のマイクロデータを実際に使用した国際比較分析がワークショップ開催期間中に各国統計局職員により行われ、マイクロデータ使用の困難性と有用性が認識されるとともに、マイクロデータおよびそのユーザー・マニュアルについて改善すべきいくつかの点が明らかにされた。最終日には国際比較分析の結果報告が行われた。

今回使用されたのは 2009 年カンボジア社会経済調査 (Socio-Economic Survey)、2011 年タイ世帯の社会経済調査、2003-2004 年ネパール生活水準調査 II (Living Standard Survey II) の 3 調査のマイクロデータである。これらの調査は、いずれも調査員による聞き取り調査形式で行われている。国際比較分析には、国際マイクロ統計データベースで提供されているマイクロデータと同様に抽出率 80% で再抽出されたリサンプリングデータが使用された。再抽出後のサンプルサイズは、2009 年カンボジア社会経済調査で 9,577 世帯、2011 年タイ世帯の社会経済調査が 33,667 世帯、2003-2004 年ネパール生活水準調査 II が 3,130 世帯であった (Furuta, 2015b)。

IV おわりに

国際マイクロ統計データベースは、LIS で手薄になっている東南アジア・南アジアの国々の公的統計のマイクロデータによって構築されており、世帯の所得支出に関する研究上きわめて重要なものである。近年アジア地域が経済成長を期待されるエリアとなり、統計に対するニーズも急激に高まってきている中で、国際マイクロ統計データベースが、こうした国々も対象に加えることが出来ているということは非常に高く評価されるべき点である。

国際マイクロ統計データベースの使用は、高いセキュリティレベルを確保するため統計数理研究所内に限定されるが、外部の研究者も利用可能となっている。また、利用申請者にはテスト用データが提供されるため、実際に統計数理研究所内でマイクロデータを利用する前にデータの扱いに慣れることが出来る。加えて、他国の研究者が利用することを想定したユーザー・マニュアルが提供されている点が、国際マイクロ統計データベースの価値を大きく高めている。すでに述べてきたように、海外のマイクロデータを用いる上での難問は、マイクロデータを構成している諸データの位置付けが明らかに出来ないことによって生じる部分が大きい。分析における大きな錯誤を防ぐ意味でも、十分に議論がなされたユーザー・マニュアルの重要性は極めて大きい。

今後、国際マイクロ統計データベースの利活用が進めば、アジア諸国のマイクロデータに対する利活用の活性化や分析の高度化にも繋がると期待される。同データベースのさらなる拡充と利便性の向上が待ち望まれる。

注

- 1) IPUMS-International ホームページ <https://international.ipums.org/international/overview.shtml> (2015 年 12 月 27 日閲覧) 参照。
- 2) ルクセンブルク国際データセンターホームページ <http://www.lisdatacenter.org/> (2015 年 12 月 27 日閲覧) 参照。
- 3) ICPSR ホームページ <http://www.icpsr.umich.edu/icpsrweb/landing.jsp> (2015 年 12 月 27 日閲覧) 参照。
- 4) KUMA ホームページ <http://www.rieb.kobe-u.ac.jp/kuma/hyogo/index.html> (2015 年 12 月 27 日閲覧) 参照。
- 5) 労働政策研究・研修機構ホームページ「JPILPT データ・アーカイブ」<http://www.jil.go.jp/kokunai/statistics/archive/index.html> (2015 年 12 月 27 日閲覧) 参照。なお、2007 年までは一部の調査について、データを SSJDA に寄託していた。

- 6) 2015 年度統計関連学会連合大会では、「「国際マイクロ統計データベース」の利用方法について」(岡本・山下, 2015) および「国際マイクロ統計データベース作成を通してみた東南アジア諸国の家計収支調査の調査方法の現状と問題」(古田, 2015) が報告された。
- 7) 統計情報研究開発センターホームページ「研究活動」<http://sinfonica.or.jp/information/research/index.html> (2015 年 12 月 27 日閲覧)。また以下の記述は、上記ホームページのほか、Ito (2015)、及び、『アジア地域貧困問題研究用の世帯統計マイクロ統計データベースの編成とその解析 2012 年度研究成果報告書』https://kaken.nii.ac.jp/pdf/2012/seika/C-19_1/82662/21330050seika.pdf (2015 年 12 月 27 日閲覧) を参考にしている。
- 8) 国際マイクロ統計データベース設立に至る経緯については、松田, 神谷, 白須 (2007) や伊藤, 中川 (2007) でより詳しく紹介されている。
- 9) 同研究の研究成果については、統計情報研究開発センター (2013) を参照されたい。
- 10) これらの国々の中にはマイクロデータ利用申請を直接受け付けている国もある。西 (2013) には、カンボジア、インドネシア、スリランカの 3 カ国における匿名データ提供についての紹介がある。
- 11) 手続きの詳細については、統計情報研究開発センターホームページ内の「国際マイクロ統計データベース利用の手引き」(<http://sinfonica.or.jp/information/research/index.html>) を参照されたい。

参考文献

- 岡本基・山下智志, 2015, 「「国際マイクロ統計データベース」の利用方法について」統計関連学会連合大会報告資料。
- 統計情報研究開発センター編, 2013, 『国際マイクロデータ・ラボラトリー研究報告・統計データ解析の国際協力への新たなアプローチ』統計情報研究開発センター。
- 西文彦, 2013, 「東アジア諸国における政府統計データの提供事情」http://www.stat.go.jp/training/2kenkyu/pdf/ea_data.pdf。
- 伊藤彰彦, 中川雅義, 2007 「シンフォニカにおける国際統計協力」『ESTRELA』2007 年 4 月 (No.157) pp.26-27。
- 古田裕繁, 2015, 「国際マイクロ統計データベース作成を通してみた東南アジア諸国の家計収支調査の調査方法の現状と問題」統計関連学会連合大会報告資料。
- 松田芳郎, 神谷傳造, 白須孝, 2007 「統計データ解析における国際協力の新しい試み: ミクロ統計データ解析実験国際研究所 (An International Experimental Laboratory to Analyse Micro Data Database for Comparison of Income Distribution and Related Topics) 設立」『ESTRELA』2007 年 4 月 (No.157) pp.2-7。
- Ito, Akihiko., 2015, “Laboratory Project” (第 7 回公的統計のマイクロデータ分析国際ワークショップ報告資料)。
- The Institute of Statistical Mathematics and Statistical Information Institute for Consulting and Analysis., 2014, *Users' Manual for Handling Resampled Micro Data of Indonesian National Social Economic Survey SUSENAS-Overall and Survey Process (Version 1.2)* <http://sinfonica.or.jp/information/research/index.html> (2015 年 12 月 27 日閲覧)。
- The Institute of Statistical Mathematics and Statistical Information Institute for Consulting and Analysis., 2015, *Users' Manual for Handling Resampled Micro Data of Sri Lanka Household Income and Expenditure Survey HIES 2009/10 (Version 1.1)* <http://www.sinfonica.or.jp/information/research/index.html> (2015 年 12 月 27 日閲覧)。
- Furuta, Hiroshige., 2015a, “Outline of the International Micro Database” (第 7 回公的統計のマイクロデータ分析国際ワークショップ報告資料)。
- Furuta, Hiroshige., 2015b, “Guideline for the workshop of international comparative study” (第 7 回公的統計のマイクロデータ分析国際ワークショップ報告資料)。

ホームページ

情報・システム研究機構新領域融合研究センターホームページ「公的統計のマイクロデータ利用」<http://www.rois.ac.jp/tric/micro/>

統計情報研究開発センターホームページ <http://www.sinfonica.or.jp/>

ルクセンブルク国際データセンターホームページ <http://www.lisdatacenter.org/>

労働政策研究・研修機構ホームページ <http://www.jil.go.jp/index.html>

ICPSR ホームページ <http://www.icpsr.umich.edu/icpsrweb/landing.jsp>

IPUMS-International ホームページ <https://international.ipums.org/>

KUMA ホームページ <http://www.rieb.kobe-u.ac.jp/kuma/hyogo/index.htm>