

国際マイクロ統計データベース作成を通して見た 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

(公財)統計情報研究開発センター客員上席研究員 古田裕繁

要旨 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共同で運営する『国際マイクロ統計データベース』を通して提供するリサンプリングデータの編成とユーザーマニュアルの作成により明らかになった、東南アジア各国統計局（NSO）から提供された家計収支調査のマイクロデータおよび調査方法の現状と問題点について述べる。次いで、東南アジア諸国のマイクロデータの利用分析を促進するために、提供するマイクロデータの編成とユーザーマニュアルの作成において考慮すべき事項を提言する。

キーワード：マイクロデータ解析、経済・経営統計、統計調査法・標本調査論

はじめに 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共同で運営する『国際マイクロ統計データベース』の構築にあたり、2013年から東南アジア各国マイクロデータのチェック・編集・リサンプリングと、ユーザーマニュアルの作成を担当している。その経験を通して、主に東南アジア諸国の家計収支世帯標本調査の調査方法の現状と、所得・消費分析でマイクロデータを利用する上で問題点を明らかにし、ユーザーマニュアルの編集方針について提言を行う。

なお、本稿の内容は、筆者の個人的見解であり、関連するプロジェクトの見解ではないことを申し添える。

1. 本研究の対象とするマイクロデータ

筆者が、国際マイクロ統計データベースのプロジェクトで扱ったマイクロデータは表 1 のとおりである。このうち、インドネシアの全国社会経済調査 SUSENAS 2006、タイの世帯社会経済調査 HSES 2004、ベトナムの世帯生活水準調査 VHLSS 2006、スリランカの世帯収支調査 HIES 2009 については、ユーザーマニュアルが（公財）統計情報研究開発センターのウェブ上で公開され、リサンプルデータも大学共同利用機関法人情報・システム研究機構のオンサイトで利用可能となっている。また、タイの世帯社会経済調査 HSES 2011、カンボジアの社会経済調査 CSES 2009、ネパールの生活水準調査 NLSS II

2003、ラオスの消費支出調査 LECS 2007 のマイクロデータに関しても、ユーザーマニュアルの素案を作成しているところであり、順次公開される予定である。これらのマイクロデータを中心に所得・消費の把握方法の現状と問題点を述べる。

表1 国際マイクロ統計データベース作成の進捗状況（順不同）

国	NSO から提供されたマイクロデータ	マニュアル・ ドラフト作成	マニュアル 公開 ・データ提供
インドネシア	SUSENAS 2000, 2003, 2006, 2011	✓(2000、2003)	✓ (2006)
マレーシア	HES 2004	✓(2004)	(注)
タイ	HSES 2007, 2011	✓(2011 作成中)	✓ (2007)
ラオス	LECS 2007	✓(2007)	
ベトナム	VHLSS 2002, 2004, 2006	✓	✓ (2006)
スリランカ	HIES 2006, 2009	✓	✓ (2009)
モンゴル	HSES 2007, 2012		
バングラデシュ	HIES 2000, 2005		
カンボジア	CSEC 2007-2012 (各年)	✓(2009 作成中)	
ネパール	NLSS 1995, 2003	✓(2003 作成中)	

(注) 協定からの脱退による公開中止

2. 所得・消費分析で期待されるデータと実際

マイクロデータの利用分野は種々想定され、提供するデータもそのニーズに対応できるものであるのが望ましい。ここでは、所得・消費の世帯分布の研究などで利用される世帯所得総額、世帯消費総額などの変数について述べる。

一般に、研究の対象である世帯所得総額、世帯消費総額などについて、エラーや異常値を含まないクリーンな変数がデータセットに存在することが期待されている。また、国際比較を行う際には、各国の所得・消費の概念定義、範囲、把握方法、推計方法等のメタデータが同時に提供されていることが期待されている。

表2は、各国 NSO から提供されたデータセットに、世帯所得総額、世帯消費総額などの変数が存在するか否かをまとめたものである。インドネシアやタイのように調査票上に世帯所得総額、世帯消費総額の記入欄がある場合は、当然に、データセットに当該変数が含まれる。

その他の国では、調査票には世帯消費の内訳や世帯所得の構成要素に関する質問事項はあるが、総額の記入欄はない。その場合、調査票のデータエントリを行い、NSO において、データチェックを行った上で、世帯所得総額、世帯消費総額の変数を作成している。結果表を集計し、調査報告書を公表しているのだから、各国 NSO は世帯所得総額、世帯消費総額の変数を持っている。ベトナムやスリランカは、当該変数が提供されたデータセットに含まれているが、ネパールやカンボジアは、提供されたデー

データセットに含まれていない。この違いは各国 NSO のデータ提供方針によるものである。

なお、各国ともに、原則として、調査項目に対応する原データはデータセットに含まれているので、そこから積み上げて世帯所得総額、世帯消費総額の変数を作成することは可能である。

表 2

世帯消費の調査事項			
世帯消費の内訳はあるが、合計欄は無し			
提供ファイルに合計値無し	Nepal	Cambodia	
提供ファイルに合計値有り	Vietnam	Sri Lanka	
世帯消費合計の調査項目有り	Indonesia	Thailand	
世帯所得の調査事項			
提供ファイルに含まれない	Indonesia		
世帯所得の構成要素のみで、合計欄は無し			
提供ファイルに合計値無し	Nepal	Cambodia	
提供ファイルに合計値有り	Vietnam	Sri Lanka	
世帯所得合計の調査項目有り	Thailand		

3. データベース構築にあたってのデータチェック

データベース構築にあたり、各国 NSO から提供されたマイクロデータについて、値の範囲チェック、コードのチェック、変数間の整合性チェックなど各種のデータチェックを行う。世帯所得・消費に関しては、次の3段階となる。

- ① 個票レベルの所得・消費の構成要素の変数から世帯所得総額、世帯消費総額の変数を試算する。
- ② 提供されたデータセットに世帯所得総額、世帯消費総額の変数がある場合、①での試算値と比較する。
- ③ 最後に、当該世帯所得総額、世帯消費総額の変数を用いて集計した平均世帯所得額、平均世帯消費額あるいは分布が、NSO の公表した調査報告書に記載された値に等しい、または近いことを追認する。

(第1段階)

合算のプロセスの概略については、調査報告書に記述されていることが多いが、その詳細は明らかではない。データチェックを行うにあたって、この合算プロセスが最も困難を伴う。表3に、提供されたマイクロデータファイルの概略を示している。インドネシアの場合は、1つの世帯ファイル内での変数の整合性のチェックとなる。タイも、世帯サマリファイルがあるので、サマリファイルの変数と各個票のデータファイル間の整合性チェックとなる。しかし、その他のベトナム、スリランカ、カンボジア、ネパールでは、多数のデータファイル进行处理する必要がある。

NSO における合算処理が不明確な例として、異常値の NSO での取り扱いが不明なことがある。また、農業所得や事業所得の計算で生じるマイナスの所得を、NSO がどのように処理しているかまでは、報告書に記載されていない。表4の「マイナス所得の取り扱い」は、ワークショップやメール等で各国 NSO 担当者から入手した情報を基にした。

(第2段階)

試算した世帯所得総額、世帯消費総額の変数について、NSO から提供された変数と最初から一致することは稀である。これは、NSO 担当者と議論する機会があれば、計算プロセスについての理解が深まり、ギャップは縮小させることができる。しかし、解消しないギャップもある。

一致しない原因として、「NSO が報告書の集計に用いた最終データチェック済のデータファイル」と「NSO から提供されたデータファイル」の内容が一部異なるケースがある（スリランカ）。

(第3段階)

世帯レベルの世帯所得総額、世帯消費総額の変数を集計して、報告書掲載の公表値と一致するかチェックを行うが、一部の国では、結果精度の関連から、実数の掲載は避けて、比率、構成比を中心に報告書をまとめている。この場合は、正確な比較に困難を伴う。

また、加重平均を求める際に利用する世帯ウェイトに関して、ウェイトの合計が推計世帯数と一致しないケースもある。

4. 国際比較を行う際の調査方法、推計方法上の論点

所得・消費の国際比較を行う際には、概念定義、範囲、調査票の設計、実査方法などの調査方法と、得られたデータから世帯所得総額や世帯消費総額の変数を推計した方法を明確にする必要がある。表4に論点となる項目をまとめた。

(東南アジア諸国の経済社会)

東南アジア諸国の経済社会の基本的な特性は、都市部では、非農業部門の雇用所得が主たる所得である世帯も見られるが、農村部では自給自足経済の状態にある世帯が多い。このため、家計収支調査による実態把握には種々の制約と困難がつきまとう。

(所得・消費の季節性の問題)

農業生産が中心であるために、季節性の問題を無視することはできない。このため調査方法の上では2つの工夫がされている。一つは、年間サンプル世帯を12等分し、各月に配分して実査を行う方法である。もう一つは、調査項目の所得や消費の参照期間を12か月とする方法である。

なお、ベトナムは調査の実査時期を5月と9月の年2回に分け、参照期間を12か月としている。また、カンボジアは調査票の上で、穀物生産の参照期間を雨季・乾季に区分している。

(消費の把握方法)

消費支出を把握する方法には、大きく分けて日記(diary)方式と想起(recall)方式がある。カンボジアは日記方式と想起方式を併用している。熟練調査員が調査世帯に頻繁に通うなどの支援があれば、多数の品目について正確なデータが期待できる。なお、カンボジアでは、消費推計には主に想起方式のデータが利用され、日記方式のデータは貧困分析に利用されている。

一方、想起方式は、実査は容易にみえるが、回答者の正確な記憶が頼りであるため、余り長い参照期間を設定すると、記入内容の信頼性が低下する。食料については、大部分の国が参照期間を7日としている。ベトナムは、すべての調査項目について参照期間が1年となっているが、食料の調査票を祝祭日用と平常日用に分けて、記入の容易さと正確性を確保するための工夫をしている。

(消費項目分類)

インドネシア、タイは、調査票に世帯消費総額の記入欄があるが、それが可能なのは、調査している消費項目数が少ないからである（インドネシアの食料4項目、タイの食料14項目）。データセットに世帯消費総額の変数があるので、消費の分布や格差を研究するためには便利であるが、消費の構造分析には適していない。

その他の国では、調査員が回答者の記憶を引き出し、できるだけ正確な情報を得るために、詳細な品目を調査票に記載していることが多い。

(自家消費)

自給自足的な農業経済のために、農作物の自家消費分を含めた現物の評価は、消費と所得の両面で重要である。調査票の上では、現物を区別できるような設計となっている国が多いが、記入は回答者任せであり、調査員がどのようにインタビューで回答を引き出しているかは不明である。

スリランカは、米の自家消費額の評価について、生産側から把握した額と消費側から把握した額の大きい値を、消費面、所得面の推計に用いている。

(帰属家賃)

帰属家賃についても、設問の仕方は様々だが、基本的に回答者の評価に基づいている。なお、ネパールは、不詳の回答、また、異常値と判断される回答については、調査の結果得られた家賃データを基にヘドニック法を適用して、補完をしている。しかし、異常値の範囲、家賃データの範囲等が不明なため、これを追認することは困難である。

(耐久消費財)

耐久消費財の評価について、大部分の国は、参照期間中の購入額を基にしているが、ネパールは、現在価格による原価償却分を推計して、耐久消費財の消費額としている。別途、提供データセットの中に、耐久消費財の価格上昇率に関するデータが含まれていないという問題がある。

(所得の把握方法)

世帯消費に比べ、世帯所得の調査はより一層の困難を伴い、結果精度は消費よりも低いと考えられる。したがって、所得データは、世帯属性間での所得格差の議論には適していても、家計収支バランスの議論をする際には、消費面、所得面のデータの精度について、比較可能なものであるか事前検証が必要である。

インドネシアは、調査票で所得項目の調査を行っているが、結果精度の点で問題があるので、所得に関するマイクロデータの提供はもちろん、所得に関する調査結果の公表も行っていない。

(農業生産、収入の参照期間)

穀物生産の季節性の問題を避けるために、各国ともに参照期間を1年としている。なお、カンボジアは前述のように1年をさらに乾季と雨季に分けている。

(農業所得の把握)

各国の所得源泉別所得は、農業所得の割合が高い。農業所得の推計においては、農業生産額の推計が中心となる。スリランカ、ベトナムは金額ベースで調査を行っている。タイは、生産量と生産額の両方を調査している。ネパール、カンボジアは、生産量と販売価格から生産額を計算している。

各国ともに、生産を販売分と自家消費分を区別している。なお、ネパールの推計方法は複雑である。調査項目には、収穫量、販売数量、販売価格がある。販売金額の推計には、各世帯の販売価格を適用する。しかし、「収穫量－販売数量」＝自家消費分の評価に適用する価格は、上位地域レベルの最頻値を採用するとしている。ここでのデータ処理上の問題点は、価格単位として、例えば重量のKgの他に、ローカルの単位を数種類認めていることである。

(マイナス所得の取り扱い)

雇用者所得には、その推計プロセスから、マイナスの所得が出ることはあり得ないが、農業所得、事業所得は、「収入－費用」として推計されるので、マイナスの所得が発生することがある。このため、世帯所得総額でみてもマイナスとなる標本世帯があり、所得分布、ローレンツ曲線、ジニ係数等の分析の際に問題となる。

タイは、マイナスの所得項目があっても、そのままにしている。一方、スリランカ、ベトナムは、マイナスの所得項目を0で置き換えている。また、カンボジアは、真の所得0との区別をつけるために、マイナスの所得項目には1US\$相当の金額で置き換えている。

5. ユーザーマニュアル作成および提供マイクロデータセット編成に当たっての対策

以上、例として世帯所得、世帯消費を取り上げたが、マイクロデータのユーザーが、分析に取りかかる前に、片付けなければならない課題が多々ある。特に、世帯消費総額や世帯所得総額の変数がデータセットにない場合には、ユーザーが自ら積み上げのプロセスを実践しなければならない。

そこで、多様な研究者による国際マイクロ統計データベースの利用が促進されるように、この場合「国際マイクロ統計データベース」では、筆者がマイクロデータから積み上げて試算した世帯収入合計や世帯消費支出額合計の変数をデータファイルに含めて提供し、また、計算プロセスも「ユーザーマニュアル」に記載している。これは、マイクロデータのユーザーの利用の便を図るとともに、ユーザーによる検証を容易にし、マイクロデータ編集の責任の所在を明確にするためである。

まとめ 結論として、データベース作成およびマニュアル作成に当たっては、次のような方針で臨むことが重要である。

1. 国際マイクロ統計データベースのファイル作成にあたって考慮すべき事項
 - 1) エラーのある変数については、修正過程が明らかとなるように、元の変数を残すとともに、訂正候補の変数を合わせて提供する。
 - 2) 世帯所得総額変数や世帯消費総額変数が NSO から提供されていない場合、その試算値を推計し、提供ファイルに含める。併せて、推計プロセスをユーザーマニュアルで明らかにする。
2. マニュアルの作成にあたって考慮すべき事項
 - 1) NSO にとって暗黙知とされている事項をできるだけ文書化する。
 - 2) 調査票レベルのデータから個人、世帯の上位の単位に合算するプロセスをできるだけ明らかにする。特に、世帯所得総額や世帯消費総額の推計プロセスを明らかにする。
 - 3) 州・省 (Province) などの地域区分・地図、都市・農村 (Urban/rural) の定義を明確にする。
 - 4) ユーザーが追認できるように、エラーチェック、合算プロセス等のスクリプトを明らかにする。

以上

表3 各国世帯収支統計調査のマイクロデータのファイル概要

国	世帯収支統計調査	年	調査票 頁数	ファイル数 (ウェイト を含む)	ファイル 形式	ユーザー マニュアル
Bangladesh	Household Income and Expenditure Survey	2005	40	30	DBF	
Cambodia	Cambodia Socio-Economic Survey	2009	27	42	STATA	作成中
Indonesia	National Socio-Economic Survey (SUSENAS)	2006 Core 調査票	8	2	DBF	公開
Lao PDR	Expenditure and Consumption Survey (LECS4)	2007	40	6	SPSS	作成中
Malaysia	Household Expenditure Survey	2004	9	3	SPSS	作成済 (注)
Mongolia	Household Socio-Economic Survey	2004	68	22	STATA	
Nepal	Nepal Living Standards Survey (NISS II)	2003 Cross-section 調査	74	79	TXT	作成中
Sri Lanka	Household Income and Expenditure Survey	2009	32	25	TXT & CSV	公開
Thailand	Household Socio-Economic Survey	2007	32	19	SPSS	公開
		2011	65	22	SPSS	作成中
Vietnam	Household Living Standard Survey	2006	85	49	STATA	公開

(注) 協定からの脱退による公開中止

表4 各国調査での消費・所得推計の概要

項目	Nepal	Cambodia	Sri Lanka	Vietnam	Thailand	Indonesia
	2003	2009	2009	2006	2007	2006
世帯消費総額変数の有無	無（試算）	無（試算）	無（試算）	有（8大区分も）	有（詳細票以外に世帯サマリ調査票がある）	有（世帯サマリファイルが支出に関するマイクロデータのすべて）
消費の把握方法	想起(Recall)	日記(1 か月)：貧困推計用と想起：消費推計用の併用	想起(注1)	想起。食料は、祝祭日と平常日に区分して把握。	想起	想起
消費の参照期間	食料：12 か月、購入頻度の高い非食料：30 日と12 か月	食料（想起）7 日、非食料1 か月－12 か月	食料7 日、非食料1 か月－12 か月	すべて12 か月	食料7 日 非食料：1 か月と12 か月	食料7 日 非食料：1 か月と12 か月
消費分類項目数	食料 68	食料（日記）4 桁 食料（想起）20	食料 252	食料（祝祭日）24 食料（平常日）57	食料 14 非食料 121	食料 4 非食料 16
自家消費	自家生産、現物を区分	現物と自家生産は別計	自家生産、現物を消費支出の内数として計上(注2)	購入・交換 (bartered)と自家生産・現物を区分	購入、自家生産、現物の合計	購入、自家生産、現物の合計
帰属家賃	回答者の評価 (注)不明、異常値に対してはヘドニック法を適用	回答者の評価 (注)所得推計では調査項目から複雑な計算を行う	回答者の評価	調査票の設計（家賃・水道光熱の総額計算）では考慮していない。 所得面でも同様。	回答者の評価	調査票には明示なし
耐久財	現在価格による減価償却を推計 (注) 価格データが欠所	支出額	購入額	購入額	購入額	購入額

項目	Nepal	Cambodia	Sri Lanka	Vietnam	Thailand	Indonesia
	2003	2009	2009	2006	2007	2006
世帯所得総額変数	無 (試算)	無 (試算)	無 (試算)	有(9大分類も)	有(世帯サマリ調査票がある)	無(注)
農業生産・収入の参照期間	農業年(agricultural year)	雨季・乾季に分けて年間分を把握	耕作年(cultivation year)	12 か月	12 か月	—
農業所得の把握	収穫量、販売数量、販売価格から自家消費額を推計。販売金額には販売価格を適用するが、自家消費分に適用する価格は、上位地域レベルの最頻値を採用する。 (注) 価格単位は、Kg 以外にローカル単位も認めている。	生産量と販売単価	生産額と自家消費額	農業生産収入額、費用額	販売・自家消費別に生産量、金額	—
マイナス所得の取り扱い	不明	1US\$相当金額と置き換える	0 で置き換える	0 で置き換える	マイナスはそのまま	—
その他			(注1) 使用人(domestic servant)の 外食、非食料支出は、別調査票で把握 (注2) 米の自家消費額は、生産側からの把握と、消費側からの把握の大きい方を用いる。	(注)収入・支出調査と収入調査の標本世帯数は1:4		(注)所得を調査しているが、マイクロデータの提供、調査結果の公表はない