

イギリスにおけるセンサスマイクロデータの提供

—マンチェスター大学CCSRによる提供を事例として—

森 博美(法政大学)

目 次

- I はじめに
- II 提供センサスマイクロデータの特徴
- III SAR の提供条件
- IV むすび

I. はじめに

近年の欧米における統計の作成、提供について、2つの注目すべき動きが見られる。まず統計作成面では、レジスターベースの統計作成が北欧からそれ以外の地域への急速な広がりをみせている。一方統計提供面では、すでに多くの国でマイクロベースでのデータの提供ないしマイクロデータへのアクセスシステムが構築されている。この点でわが国の政府統計は、統計作成面ではなお調査統計が中心となっており、また調査結果の公表形態という点では集計表形式での提供という原則を堅持している。

社会における価値観の多様化は、これまでには考えられなかったような統計の利用ニーズを作り出すことになった。情報処理技術の飛躍的進歩の中でこのような多様化した統計利用ニーズは、これまで主として ready-made の集計結果(pre-determined tabulations)として提供されてきたわが国の統計の提供形態との間に次第に齟齬を表面化させつつある。独自の分析目的を持つ利用者そして特別なデータ解析手法ならびに処理技術をわがものとした利用者は、既存の公表結果に満足することなく、個票形態でのデータ(個体識別情報を削除するなどして匿名化されたデータは、マイクロデータと呼ばれる)といういわば record-based データへのアクセスを求めるようになってきている。このような新たなタイプの統計利用によって得られる新しい fact finding は、統計の潜在的利用価値を高めるのに貢献するものとして期待されている。

その反面でマイクロデータは、それが統計調査を通じて集められた個体情報の集合であることから、報告者のプライバシーあるいは秘密保護の問題に不可避免的に抵触することになる。マイクロデータでたとえ個体の秘密事項に対する秘匿措置が施されていたとしても、個別情報が露見するリスクは集計量によるデータ提供に比較すれば格段に増加する。そのことは、被調査者(調査客体)によるプライバシーないし秘密保護意識によるデータ公開への拒否意識を増幅することになる。このようにマイクロデータの提供について統計行政当局は、相互にトレードオフ的条件の中でいかに両者のバランスをとるかといった難しい政策選択を迫られることになる。

周知のようにイギリスにおいては、労働力調査、家計調査、総合世帯調査といった標本調

査に基づくいわゆるサーベイデータについてはすでにマイクロデータが公開⁽¹⁾され、一般の利用に供されている。イギリスは、同国におけるマイクロデータ提供の新たな展開として、1993年から91年人口センサスデータについても、一定の条件の下にデータの提供に踏み切った。同国がどのような手順を経て原データの提供者である国民のマイクロデータの提供についての合意形成に成功したか、またそれとも関係するが、どのような法的・制度的枠組みによってこの種の情報が提供されているかといったことは極めて興味ある問題である。

センサスマイクロデータの提供と関わるイギリスにおける経験は、わが国の今後の選択すべき政策のあり方を探る上でも示唆に富む内容を持っているものと考えられる。そこで以下では、同国におけるセンサスマイクロデータ提供の窓口の一つとなっているマンチェスター大学センサス・調査研究センター・センサスマイクロデータ部門(Census Microdata Unit, Centre for Census and Survey Research)⁽²⁾におけるセンサスマイクロデータの提供について概観してみたい。なお、イギリスにおけるセンサスマイクロデータの提供についての合意形成のプロセスあるいはその基礎となった露見リスク等に関する研究については、機会を改めて論じることにしたい。

II. 提供センサスマイクロデータの特徴

(1) 提供されているセンサスマイクロデータ

現在、研究センターのセンサスマイクロデータ部門(Census Microdata Unit,以下CMUと略称)を通じて2%個人匿名化標本レコード(Samples of Anonymised Records, 以下SARと略称)と世帯SARの2種類のfileが提供されている。なお、SARのデザインについては、その準備過程で地理学者が収録項目を基本事項に限定したとしても可能な限り小地域についての情報が得られるような個別レコードを要望したのに対し、社会学者や人口学者からは、地域区分の詳細情報を多少犠牲にしても人口及び世帯情報を生かすという異なった要望が出された(Dale, 1995, p.18)。その結果、このような2種類のSAR fileが作成されることになった。

これらのfileの主要な内容は、次の通りである。

まず2%個人SARは、センサスデータから抽出された個人の住居ならびに公共施設での常住者および滞在者約110万人のレコードからなる。なお個々のレコードの収録項目はセンサスでの全個人情報と世帯情報の一部からなり、サンプル全体は278の地域に区分されている。グレートブリテン島地域についての個人SARには世帯属性についての項目が含まれており世帯と個人を結合した分析が可能であるが、北アイルランドについての個人SARはこのような連結情報が不備である。

一方、1%世帯SARは、約216,000の世帯(世帯員数約50万人)からなる。なおデータとしては、センサスで調査されている世帯項目のすべてが収録されており、さらに世帯と家族についてはCMUが独自に作成した様々な導出変数(derived variables)が加えられている。なおグレートブリテン島地域のデータについては世帯と家族員をリンクして分析することも可能なよう

になっている。世帯 SAR は個体識別のリスクが個人 SAR よりも大きいいため、採用されている地域区分も大まかで、現在、イギリスの標準地域区分に Inner London と Outer London を加えた 12 区分で提供されている。(なお具体的な収録項目については、本稿末尾の付録参照)

(2) SAR の抽出方法

イギリスの人口センサスでは通常の項目については 100%処理されているが、職業や従事する産業等については全回収調査票の 10%だけがコード化され集計処理にまわされる。SAR はセンサスデータの中の 10%に相当する完全コード化サンプル(fully coded records)の中から作成される。個人、世帯 SAR での標本の重複を避けるため、最初に世帯 SAR を抽出し、これを除く個人データから個人サンプルを抽出することにより SAR を構成するといった2段階で両 SAR は作成される。

まず1%世帯 SAR の抽出にあたっては、England と Wales については州、調査区別にまた Scotland では地区や集計地域別に世帯を編成し、10 世帯からなる組が作成される。これらの組から各1世帯が無作為に抽出され、その後地域が識別されないように各地域が混ぜ合わ(scramble)される。また北アイルランドについては、センサス調査区および地区(District Council)内の世帯を 100 世帯の組に編成し、各組から1つの標本が抽出される。このような手順を経て、グレートブリテン島地区については 757,7111 レコード(215,789 世帯、541,894 人の世帯員)、また北アイルランドについては 20,833 レコード(5,255 世帯、15,578 人の世帯員)からなる世帯 SAR がそれぞれ作成される。

なお、提供される世帯データでは、地域別に世帯主の年齢(昇順)さらに職業コードによってソートされたデータが提供される。

続いて2%個人 SAR は次のようなステップを経て作成される。

まずグレートブリテン島地域については、世帯 SAR で抽出済みの 215,789 世帯を除く fully coded records の中から住居居住者については9人ずつの、また公共施設に居住する個人については5人ずつの組を作成する。これらの組の中から前者については2人ずつまた後者からは1人を無作為に抽出し、その後データを地域間で混ぜ合わせる。一方、北アイルランドについては、同じく世帯 SAR ですでに選ばれている 15,578 人を除く fully coded records を住居居住者については99人の組に、また公共施設の個人については50人ずつの組に編成する。これらの組の中から、グレートブリテン島の場合と同様に、前者については2名ずつ、後者については各1名を抽出し SAR を作成する。その結果、北アイルランドについては 10 地域 31,967 人、グレートブリテン島地域では 1,116,181 の個人が抽出される。これはセンサスの調査把握人口の約50分の1に相当する。

なお、現在 CMU では、連合王国(UK)全体の統一 SAR の作成作業に取り組んでいる。

(3) データの精度

SAR データ、センサスデータそれに調査そのものの把握対象としての人口の間には、次のような関係がある。

① センサスが把握した人口の中で SAR に含まれない者

居住者が居ることは調査員によって確認されているものの調査票が調査期間中に提出されなかった世帯、また調査期間中に調査員が接触できなかった世帯それに調査拒否世帯の世帯員の数については調査員が推測で記入し、その他の世帯属性については回答を得た隣家世帯のデータで帰属させて調査票の記入項目を埋める。このような世帯を留守世帯帰属世帯(wholly absent imputed households)と呼ぶ。センサス結果に含まれるこの種の世帯データは、SAR からは除外されている。ちなみにこの種の世帯は、Inner London boroughs(自治区)のいくつかでは 10%をこえ、グレートブリテン島全域では人口の 1.6%がこのような帰属計算によって把握されている[OPCS/GRO(S), 1994]。

また 12 人以上の世帯員数を持つ全国で 28 の世帯は、識別されるリスクを考慮して SAR から除外されている。

② SAR でもセンサスでも把握されていない者

グレートブリテン島全体で約 120 万人(2.1%)⁽³⁾と推定されている個人はセンサスデータでもまたそれから抽出される SAR データでも除かれている⁽⁴⁾。

このような SAR データの現実の人口からの標本バイアスを考慮して、SAR にはウエイト(POPWGHT)が設けられており、正確な性、年齢、人種別分布等を求める際には、このウエイトによるデータの修正が必要となる。

(4)匿名化措置

経済社会研究会議(Economic and Social Research Council, 以下 ESRC と略称)に設置された秘密保護に関するワーキンググループは、「SAR によって秘密が露見する危険性は無視できる」(Marsh et al. 1991)との結論を下した。また SAR 提供に関する秘密保護面での Registrar General のアドバイザーで、ESRC が人口センサス調査局(Office of Population Censuses and Surveys 以下 OPCS と略称)との連携の下に SAR の仕様を作り上げた際中心人物である Tim Holt 教授(現国家統計局長:chief executive of the Office for National Statistics)は、1992 年 3 月 11 日の議会答弁で SAR における個別情報の露見リスクについて、「しかるべき統計的検定の結果、露見のリスクは大多数の人口について無視できる程度のものであり、その他の者についても極めて小さいものと考えられる。国際的経験の示すところによれば、リスクのレベルは、SAR の提供に踏み切る決断を行うのにふさわしいものであり、私はこのように助言する」、としてリスクの水準が SAR 構築の妨げとはならないと主張している。

さらに SAR では、その創設に直接関わった学者による安全宣言をさらに補強する意味から、露見リスクをさらに限定されたものとするため、次のような 5 つの保護措置が採用されている。

第 1 は、SAR ではセンサスデータからの抽出率を世帯では 1%、個人については 2%に抑えることにより、データの個別情報が識別されるリスクが小さくなるようにしている。

採用されている第 2 の秘密保護措置は、データの圧縮(suppression)である。OPCS のセンサスデータですでに氏名、住所といった個人識別情報は削除されているが、希少事例あるいは

固有の事例が発生しないように、いくつかの情報についてはまとめられる。

第3の方法としては、一定水準以上あるいは以下の事例を「・・・以上(以下)」としてまとめる topcoding を含め、いくつかの調査事項について区分の統合が行われている。

まず、グレートブリテン島については、個人 SAR では 25,000 人以下の地域が、また世帯 SAR については 2,700 世帯以下の地域に対しては地域の統合が行われる。また住居での部屋数についても統合が行われており、14 室以上の住居についてはトップコード化により一括されている。年齢については、調査票に記入されている誕生日情報がまず各歳年齢に変換され、さらに 90 歳以上の年齢階級については、91,92 歳、93,94 歳がそれぞれ統合、95 歳以上はトップコード化されている。就業者の週労働時間については、71 時間以上 80 時間未満の層が一括され、80 時間を超える層についてはトップコード化されている。また産業、職業そして教育についても秘密保護の観点から分類の統合が行われている。

第 4 に SAR では、地域情報による識別化の可能性を小さくするために、センサスデータに含まれている様々な地域情報を限定することが行われている。

これまでの経験的研究に基づき、グレートブリテン島地域分の SAR データのうち 2% 個人 SAR については、広域地方自治体 (large local authority) に相当する 278 区分が採用された。(アルファベットと数字による 2 桁コード AA~TT, 01~52 の合計 278 区分 (LAD: Local Authority District) が基本分類として採用されている (AREAP: 001~278))。また、これを統合した Registrar General's 基本地域区分に Wales と Scotland を加えた 12 の REGIONP (Region of SAR area) があり、州 (counties) 別の 67 地域区分も作成されている。一方、北アイルランド地域では、1 地域の人口が最低 12 万人を超えるように属性の類似した地区をいくつか統合した 10 の地域区分となっており、連合王国 (UK) 全体では個人 SAR は、288 の地域区分を持つ。

また 1% 世帯 SAR については、露見のリスクが相対的に大きいことを配慮して全地域が 12⁽⁵⁾ に区分されている。従って、北アイルランドを加え、連合王国 (UK) 全体では、13 の地域区分が採用されている。

この他に SAR では、滞在者の常住地、勤務先住所、学生の学期中の住所、1 年前の住所に関する地域情報によって個人が特定されるのを回避するために、これらについては「標準地域」ベースあるいは「同一 SAR 地域」、「別 SAR 地域」という形で表示されている。

第 5 の秘密保護措置は、SAR の地域の並べ替えである。すでに SAR の作成手順の個所で述べたように、抽出された SAR レコードは、地域が特定されないように SAR 地域区分別に順番を混ぜ合わされる。その結果、SAR の地域順番は、OPCS が採用している地域序列と異なる。

(5) 露見リスク評価研究

センターは、欧州連合 (EU) からの資金提供を受けてヨーロッパの大学連合、研究機関、政府統計機関の参加機関の一つとして露見リスク評価モデルの開発に従事している。このワークショップでは、研究センターの Angela Dale 所長、サザンプトン大学の Chris Skinner、David

Holms 教授らが中心的役割を担っている。

ワークショップでの研究作業としては、イギリスのセンサスや標本調査のマイクロデータについて、識別リスクの評価やリスク観測指標の作成に関して、オランダ中央統計局が開発したソフトを用いた実験が試みられている。なお、この研究成果は、他のヨーロッパ諸国にも報告され、秘匿技術の相互交流に役立てられることになっている。

研究の具体的な内容としては、①イギリスの事情を考慮したデータベースや重要な項目を有するレジスターを公開する際の規模や内容、提供範囲の検討、②センサスマイクロデータの誤差の大きさを、非回答、欠損値・不完全回答、コード化・入力ミスによる誤差などについて評価すること、③大ブリテン島の2%個人 SAR を用いて地域分類の個体識別化への影響の研究(この結果は、統計開示の基準策定に利用される)、④国家統計局(ONS)との共同研究として、SAR データの個体が利用者がすでに保有しているデータベースからどの程度特定されるかの研究、といったものがある。

Ⅲ. SAR の提供条件

(1) SAR データの提供範囲

現在、SAR データの提供対象は、連合王国(UK)域内の研究者、政府機関勤務者、商用利用者そして学生に限られており、国外の利用者はこの file に直接アクセスすることはできない。現行の提供システムで海外の研究者が SAR にアクセスできるのは、次のいずれかの場合に限られる。まず研究者がイギリスの研究機関に客員研究員として滞在している場合、後に見る User License Agreement への署名等イギリス人と同じ条件の下で滞在期間中に限って利用できる。なおこの場合、帰国時にはデータを国外に持ち出すことは禁じられている。第 2 の可能性としては、イギリスの高等教育機関に所属する研究者との共同研究ということで海外の研究者に SAR 利用への道が開かれている。ただしこの場合にもデータの処理はイギリス側の共同研究者が行うことになっており、海外の研究者はその分析結果と集計表の提供を受けることができるだけで、データファイルそのものの転送を受けることはできない。

なお ESRC では、海外の研究者への SAR データの提供可能性について目下研究中であるとされており、検討の進行状況については SAR ニュースレターに掲載される。

(2) データの提供形態

SAR データは、磁気テープ、フロッピー、カートリッジ、カートリッジテープ、DAT テープといった磁気媒体で契約利用者に提供されている他、オンラインで CMU にアクセスできる他個人のサイトに転送を受けることもできる。また ESRC データ保管所では、CD-ROM 版の SAR データを作成中である。

利用者は SPSS, SIR, SAS, QUANVERT といったソフトを選択できるが、スプレッドシートでデ

ータが提供されているため,Excelのような通常の表計算ソフトで処理することができる。

(3)SARの学術的利用、商業利用

学術的利用者は SAR を無料で利用することができるが、商業(非学術的)利用の場合には有料である。使用料金は、下表のように定められている。なお、グループ購入の場合、SAR を購入したある小規模利用者は同じ地域や同一部門の別のユーザーに追加料金を支払うことなく file を転送することができるが、その場合でも当該利用者は End User License Agreement に署名する必要がある。

非学術利用の場合の提供料金(付加価値税を除く)

	2%個人 file	1%世帯 file	両 file
小口利用者 ^{a)}	1000(GB) 500(NI)	1000(GB) 500(NI)	1800(GB) 900(NI)
大口利用者 ^{b)}			1地点 500(GB) 250(NI)
グループ購入 ^{c)}	2000(GB) 1000(NI)	1000(GB) 1000(NI)	2600(GB) 1300(NI)

(注)(a)小口利用者:地方自治区、州、Scotlandの州、衛生管区、郡行政委員会区
単独地利用者

(b)大口利用者:全国ネットあるいは大規模機関、大規模ユーザーのサイトの
リストが組織の end-user License の一部を構成

(c)グループ購入:地方自治区、Scotlandの州、衛生管区、郡行政委員会区

(4)User License Agreement

SARの使用については、まず End-User License Agreement(以下、EULAと略称)に SAR データの使用者が所属する機関が署名し、さらに利用者本人が User Registration Document に署名することが義務づけられている。ユーザーの登録文書への署名を受けてその者が利用許可者(authorized user)として登録されてはじめて SAR にアクセスする事ができる。

なお、ESRCとOPCS/GROとの間で合意されたEULAは、次のような内容を持つ。すなわち、① SARの第三者への提供の禁止、② SARは王室著作権(Crown Copyright)によりESRC/JISCの許可の下にマンチェスター大学CMUを通じて提供されるものであり、OPCS/GROに対する著作権支払は発生しないこと、③特定の個人や世帯に関する情報を得たり、そのような情報の入手を要求してはならないこと、④研究利用の場合、SARの使用状況について、印刷物あるいは機械可読な形での主要な分析結果を添付して6ヶ月目に報告を提出すること、⑤許可を得

て SAR の商業利用を行う者が SAR を無料で使用した者の結果報告の提供を希望する場合、使用料金を支払うこと、がそれである(詳細については the End-User License Agreement, the Individual User Registration Document 参照)。これらの条件に違反した場合、その者が所属する組織からすべての SAR データが引き揚げられる。

商業利用あるいは非学術的目的のために SAR を使用する場合(HEFC あるいは ESRC の資金によらない研究)には、利用者は Commercial End-User License Agreement に署名しなければならない。また次に見る教育面で SAR を利用する学生の場合にも、User Registration Document への署名が利用の条件とされている。

(5) 教育用 SAR データの提供

CMU は、利用機関(教育機関)による End-User License Agreement への署名、また利用者である学生及び指導教員の User Registration Document への署名を条件に、6~10 項目からなる SAR データを教育機関に対して主に統計学、地理学、社会科の教材として提供することになっている。その際には、地域比較が可能なように、グレートブリテン島、当該地域それぞれ 1,000 の標本からなる 2 組がセットで提供される。なお、北アイルランド SAR の使用に際しては、教育機関は Northern Ireland Academic End-User License に、また利用者個人は Northern Ireland User Registration Document への署名を求められる。

現在、CMU が教育用の提供を計画しているのは、次の 3 種の SAR データである。

① 雇用と学歴データ

これは 16~60 歳の 1,000 レコードからなる SAR データで、これには経済活動、学歴、社会階層、性、年齢、婚姻、持家の有無といった項目が収録されている。

この file は、学歴と失業、男女の経済活動の違い、社会階層上の位置、階層と持家の関係の分析などに利用できると考えられている。

② 就労データ

経済活動年齢人口の男女 1,000 名のデータからなるこの SAR には、職業、労働時間、通勤方法、職場までの距離、世帯の車保有台数、年齢、婚姻といった項目が含まれている。

この file を用いることによって、男女間の通勤距離の違い、男女職業の違い、男女間でのパートとフルタイムの違いが分析できると考えられている。

③ 住居と世帯データ

これは 1,000 の世帯から構成されるデータで、これには住宅の所有、集中暖房、駐車場のない自動車の保有台数、世帯類型、民族グループ、世帯内の就労者の数といった項目が入っている。

(6) 利用申請手続き

SAR を学術研究目的に使用する場合、以下のような使用申請を経て使用が許可される。

ステップ1:申請者は自らが所属する機関が SAR へのアクセスを許可されているかどうかを点検する。

ステップ2:自機関内の CMU 連絡職員から SAR 個人登録書類1式を入手する。

ステップ3:SAR の使用手引きと学術使用の際の条件に目を通す。

ステップ4:商用ライセンス契約が必要な場合には CMU と契約を結ぶ。

ステップ5:ユーザー登録文書の第1部に所定事項を記入し、署名を行う。

ステップ6:ユーザー登録は、所属機関の CMU 連絡員によって確認されなければならない。

ステップ7:学生登録は指導教員の署名を必要とする。

ステップ8:登録文書第1部を CMU に提出し、1部を自ら保管する。

ステップ9:登録文書の注文票により SAR のユーザーガイドとコードブックを注文する。

ステップ10:CMU は登録文書を点検し,SAR へのアクセスを公認する。

ステップ11:ユーザーは、CMU 連絡員に返却された登録文書を申告し、機関の記録として処理する。

ステップ12:所属機関でのあるいは MCC 国立オンラインデータセットサービスを経由して SAR へのアクセスが許可される。

ステップ13:ユーザー登録の詳細情報が CMU の学術利用者登録データベースに入力される。

なお、付録3は、MIDAS の利用も含めた SAR の利用に至る流れを図解したものである。

IV. むすび

本稿では、1991 年センサスを契機に匿名マイクロデータの提供に踏み切ったイギリスについて、その提供の窓口となっているマンチェスター大学センサス・調査研究センターセンサスマイクロデータ部門(CMU,CCSR)を事例として、提供されているマイクロデータ(SAR)の特徴ならびに提供システムについて概観してきた。

SAR もセンサスそのもののイギリスの特徴、すなわち連合王国(UK)が England の他に Scotland、Wales さらには北アイルランドといった相互に行政的にも独自性の強い地域の連合体として構成されていることを反映している。すなわち、統計上の地域区分も行政区画の相違を反映して地域によって微妙に異なる。この点は、全国的に統一された地域区分を採用しているわが国と大きく異なっている。この点 SAR で特筆すべきは、それがグレートブリテン島地域と北アイルランドとでそれぞれ別のファイル編成となって提供されていることである。現在、統一ファイルの作成に向けての作業が進行中とのことであるが、当面は世帯、個人 SAR のいずれも国全体を2種類の異なるレコード編成からなる file でカバーするという状態は解消されない。

このような特殊イギリス的制約はあるものの、採用項目や秘匿措置といった SAR の編成、さらには提供の技術的・法制度的側面といった提供形態のいずれについても、事前の綿密な研究さらには海外の選考事例などを参考にしながらこれは作り上げられている。これは、わが国が今後マイクロデータの提供という政策を選択することになった場合、一つの有力な先行事例として政策的に示唆に富む内容を持っているものと考えられる。

いうまでもなく本稿で紹介したのはイギリスにおけるセンサスマイクロデータの提供システムであり、これをそのままの形でわが国の統計「風土」に移植するには様々な困難が予想される。そこでこれを一つの参考事例として適切に評価する上で必要であると思われる今後の検討課題を2、3指摘しておくことにより、ここでのむすびとしたい。

イギリスでは家計調査や労働力調査などいくつかの標本調査データについては以前からマイクロデータが提供されている。このような実績があったとはいえ、一国の最も基本的な統計である人口センサスデータをマイクロデータとして提供するには被調査者である国民の側から個人の秘密保護の立場から異論も少なくなかったはずである。この点で、イギリスがどのような手続きを経て国民各層との間に合意を取り付けることができたか、特に Marsh らによる研究 (Marsh C. et al. 1991) や 1991 年～92 年にかけて立法府での審議も含め、どのようなプロセスを経て合意が形成されたかを確認しておくことは、わが国における合意形成のあり方とも関係して極めて興味ある検討課題であると考えられる。

第2に、マイクロデータ提供の法制度的枠組みに関して、End-User License Agreement や User Registration Document の内容を含めたマイクロデータ提供に関わる法体系についてさらに検討しなければならない。その際には、この制度運用に関しての問題点なども押さえておく必要があると考えられる。さらに、制度発足以来の SAR データの具体的な利用実績⁶⁾ならびにそれによって現実についての認識としてどのような新たな知見が得られつつあるかについてのサーベイも興味ある課題である。

これらの課題については、今後引き続いて検討してみることにしたい。

(1996 年 12 月 25 日)

(本稿は、1996 年 12 月 22 日に京都で開かれた A02 (ウ・エ) 班合同研究会での報告に加筆したものである)

注

(1) イギリスにおけるサーベイマイクロデータについては、エセックス大学に設置された ESRC (Economic and Social Research Council) の Data Archive がその提供窓口となっている。

(2) CMU, CCSR: Census Microdata Unit, the Centre for Census and Survey Research, Faculty of Economic and Social Studies, the University of Manchester.

この研究センターは、1992 年に故 Cathie Marsh が 91 年センサス匿名サンプルレコード (SAR: Samples of Anonymised Records) の管理、普及及び研究の実施のために

ESRC(Economic and Social Research Council)の助成を受けてマンチェスター大学内に設置した CMU をその組織上の前身としている。

CUM はその後、センサスペースでの研究活動を行ってきたが、95 年はじめ、研究活動の拡大とともに、Cathie Marsh 記念 Center for Census and Survey Research(CCSR)へと発展的に改組され、同年 11 月に本格的にその活動を開始した。その際に CMU は新しい研究センターの中核部門として存続することになった。

また、CMU は、マンチェスター大学コンピューティングのグループからなる the National Database Team(MIDAS)と連携し、センサスや調査データの分析の他、SAR についてのトレーニングコースなどのサービスを提供している。なお、コースの開催日程やカリキュラムの内容などについては、SAR ニュースレターや CCSR のホームページに掲載されることになっている。

(3) 帰属計算による把握の問題とともに、1991 年センサスにおける把握漏れの急増が統計関係者の間で大きな問題になっており、一部には、イギリスでの今後のセンサスそのものの実施を再検討する時期にきているとの議論も出されている[Dale, 1995]。

(4) この調査漏れの規模については、センサス終了後に実施された事後調査でその約 3 分の 1 に相当する部分が調査漏れ、残りが把握世帯の中での世帯員の不十分な把握によるものとされている。なお SAR から除外されているいわゆる帰属サンプルの特性としては、若年の青年、特に 20 歳代が圧倒的に多いことが知られている。また上記②については、政府推計人口との比較の結果、センサスデータも SAR も子供、高齢者にも相当数いるものと推測されている。このような調査漏れは地域的には、都市部、特に Inner London で大きくなっている。

(5) England 地域については、Registrar General's Standard Regions が全地域を 8 に区分している。SAR データではこのうちロンドンが属する South East 地区を Inner London, Outer London そしてそれ以外の South East の 3 地域に分割し 10 区分となっている。これに Scotland と Wales を加え、グレートブリテン島全体では 12 地域区分が採用されている。

(6) SAR のユーザーとしては、現在のところ約 250 名の社会学、人口論、地理学、保健学、社会政策研究者、相当数の地方政府、中央省庁それに市場・社会調査会社の社員が登録されている(Dale, 1995, p.18)。

参考文献

Dale, Angela, (1995) *The Decennial Census of Population: Do we still need one?* Manchester Statistical Society.

Marsh, C., Skinner, C., Arber, S., Penhale, B., Openshaw, S., Hobcraft, J., Lievesley, D. and Walford, N. (1991) *The Case for Samples of Anonymised Records from the 1991 Census*, *Journal of the Royal Statistical Society, A*, 154, Part 2, pp. 305-340.

OPCS/GRO(S), (1994) *1991 Census User Guide 58, Undercoverage in Great Britain.*

付録1 SAR 収録項目

グレートブリテン 2%個人 SAR 項目

<個人項目>

SAR 区(278)、SAR 区地域(region of SAR area:(グレートブリテン 12、北アイルランド1)、年齢、公共施設の状態、公共施設のタイプ、出生国、移動距離(移住者)、職場までの距離、経済的地位(第1次)、経済的地位(第2次)、人種、家族のタイプ、ゲール語(Scotlandのみ)、通常の労働時間、産業分類、非活動長期疾病、配偶関係、以前の居住地、職業分類、資格の数、最終学歴、資格名、世帯主との続柄、在留資格(resident status)、性、社会階層、社会・経済グループ、学期中の住所、主たる通勤手段、常住地(訪問者)、ウェールズ語(Walesのみ)、勤め先名

<世帯項目>

風呂/シャワー、集中暖房、内部 WC、車の保有台数、住居の最下床のレベル(Scotland)、居住面積、1室当たりの居住者数、居住空間の所有、世帯の居住者数、世帯内の扶養子供数、世帯内の非活動長期疾病者数、世帯内の年金年齢者数、世帯内の就業者数、世帯主の経済的地位、世帯主の性、世帯主の社会階層

<導出変数>

(個人レベル)

州(SAR areas を 67 に統合)、資格群、産業分類、職業(中分類)、職業(大分類)

(その他の導出変数)

ケンブリッジ・スコア、配偶者のケンブリッジスコア、人口ウエイト、NES 平均時給、NES サンプル数、NES 標準偏差、NES 非労働力スコア

グレートブリテン 1%世帯 SAR 項目

<世帯レコード>

SAR 区、風呂/シャワー、集中暖房、内部 WC、車の保有台数、住居の最下床のレベル(Scotland)、居住面積、居住面積、居住空間の室数、居住空間の所有、世帯人員数、移動世帯

<個人レコード>

年齢、出生国、移動距離(移住者)、職場までの距離、経済的地位(第1次)、経済的地位(第2次)、従業上の地位、人種、世帯主、家族数、家族のタイプ、ゲール語(Scotlandのみ)、通常の労働時間、産業分類、非活動長期疾病、配偶関係、(移住者)以前の居住地、職業分類、資格の数、最終学歴、資格名、世帯主との続柄、在留資格(resident status)、性、社会階層、社会・経済グループ、学期中の住所、主たる通勤手段、常住地(訪問者)、ウェールズ語(Walesのみ)、勤め先

<世帯と個人を関連付ける項目>

世帯識別子、世帯内の個人番号

<世帯レベルの導出変数>

世帯内の居住者数、世帯内の扶養子供数、最年長扶養子供年齢、最年少扶養子供年齢、世帯内の成人居住者数、世帯内の16歳未満居住者数、世帯内の年金者数、世帯内の非活動長期疾病者数、世帯内の有業者数、世帯内の経済活動居住者数、世帯内の失業者数、世帯内の退職者数、世帯内の不治傷病者数、世帯内の経済的非活動居住者数、その他の非活動居住者数、学期中の住所で調査を受けた世帯内の学生数、世帯内の扶養者数、世帯内の最高齢扶養者数、世帯内の最若年扶養者数、学生のみからなる世帯(学期中の住所)、年金生活者のみからなる世帯、成人のみからなる世帯

<世帯主に関する導出変数>

世帯主の経済的地位、世帯主年齢、世帯主の性、世帯主の社会階層

<個人レベルのレコード変数>

最終学歴科目群、産業分類、職業(小分類)、職業(中分類)、職業(大分類)

<家族レベルの導出変数>

家族内の居住者数、家族内の扶養子供数、家族内の最高齢扶養子供の年齢、家族内の最年少扶養子供の年齢、家族内の成人数、家族内の16歳未満居住者数、家族内の年金者、家族内の長期疾病者数、家族内の有業者数、家族内の経済活動者数、家族内の失業者数、家族内の退職者数、家族内の不治傷病者数、家族内の経済的に非活動居住者数、家族内のその他の非活動居住者数、家族内の学期中の住所で調査された学生数、家族内の扶養者数、家族内の最高齢扶養者の年齢、家族内の最若年扶養者の年齢、世帯主の経済的地位、世帯主の年齢、世帯主の性、世帯主の社会階層

<その他の導出変数>

世帯構成のタイプ、ライフステージ変数、扶養世帯、ケンブリッジスコア、配偶者のケンブリッジスコア、ゴールドスロープ階級、女性・就業調査で使用されている階層分類、SOC 単位グループ、NES平均時給、NES標本数、NES非労働力スコア、NES標準偏差、国際標準職業分類、標準国際職業威信度、従業上の地位の国際社会・経済インデックス、最小世帯単位のタイプ、最小世帯単位の地位、最小世帯単位の組み合わせ

付録2 Academic End-User License

<使用条件>

- (1) SARあるいはその加工データは、学術的教育ないし研究にのみ使用すること。
- (2) データは、有料コンサルタント、商用あるいは政府ないし地方当局の資金提供を受けた非学術的研究に使用してはならない。
- (3) 特定の個人や世帯に関する情報を入手したり導出したりするためにSARを、使用しても、使用を試みても、また使用したと主張してはならない。
- (4) SARに基づく刊行物や報告には王室著作権によることを明記すること。
- (5) 許可を受けた学術機関に所属する他の登録ユーザーを除き、SARおよびその導出データの写しを他に譲渡してはならない。
- (6) SARに基づくあらゆる刊行物や報告に ESRC/JISC/DENI および CMU の役割を明記すること。
- (7) SARの使用に関する年次報告をCMUに行うこと。
- (8) 刊行物および導出データセット(コードブックを含む)の写しをCMUに預けること。
- (9) 所定の登録機関終了時あるいは組織を離れるときにSARと導出データのすべてを削除すること。
- (10) 別な高等教育機関に移動する際には新規にユーザー登録文書を作成すること。

付録3 SAR 利用申請－承認手続き

