

# 統計学

第 104 号

---

## 論文

教育用擬似マイクロデータの作成

—平成16年全国消費実態調査を例として—

..... 山口 幸三・伊藤 伸介・秋山 裕美 (1)

Analysis of IO-based Annual Supply and Use Tables for the Development of QNA

..... Takeshi SAKURAMOTO (16)

## 研究ノート

Introduction of the Theory of Correlation into Russia and E. Slutsky

..... IRINA ELISEEVA (41)

## 本会記事

支部だより..... (52)

経済統計学会内規・編集委員会規程・投稿規程・執筆要綱・投稿原稿査読要領..... (57)

---

2013年3月

経済統計学会

## 創刊のこ と ば

社会科学の研究と社会的実践における統計の役割が大きくなるにしたがって、統計にかんする問題は一段と複雑になってきた。ところが統計学の現状は、その解決にかならずしも十分であるとはいえない。われわれは統計理論を社会科学の基礎のうえにおくことによって、この課題にこたえることができると考える。このためには、われわれの研究に社会諸科学の成果をとりいれ、さらに統計の実際と密接に結びつけることが必要であろう。

このような考えから、われわれは、一昨年来経済統計研究会をつくり、共同研究を進めてきた。そしてこれを一層発展させるために本誌を発刊する。

本誌は、会員の研究成果とともに、研究に必要な内外統計関係の資料を収めるが同時に会員の討論と研究の場である。われわれは、統計関係者および広く社会科学研究者の理解と協力をえて、本誌をさらによりよいものとするを望むものである。

1955年4月

## 経 済 統 計 研 究 会

## 経 済 統 計 学 会 会 則

第1条 本会は経済統計学会（JSES : Japan Society of Economic Statistics）という。

第2条 本会の目的は次のとおりである。

1. 社会科学に基礎をおいた統計理論の研究
2. 統計の批判的研究
3. すべての国々の統計学界との交流
4. 共同研究体制の確立

第3条 本会は第2条に掲げる目的を達成するために次の事業を行う。

1. 研究会の開催
2. 機関誌『統計学』の発刊
3. 講習会の開催、講師の派遣、パンフレットの発行等、統計知識の普及に関する事業
4. 学会賞の授与
5. その他本会の目的を達成するために必要な事業

第4条 本会は第2条に掲げる目的に賛成した以下の会員をもって構成する。

- (1) 正会員
- (2) 院生会員
- (3) 団体会員
- 2 入会に際しては正会員2名の紹介を必要とし、理事会の承認を得なければならない。
- 3 会員は別に定める会費を納入しなければならない。

第5条 本会の会員は機関誌『統計学』等の配布を受け、本会が開催する研究大会等の学術会合に参加することができる。

- 2 前項にかかわらず、別に定める会員資格停止者については、それを適用しない。

第6条 本会に、理事若干名をおく。

- 2 理事から組織される理事会は、本会の運営にかかわる事項を審議・決定する。
- 3 全国会計を担当する全国会計担当理事1名をおく。
- 4 渉外を担当する渉外担当理事1名をおく。

第7条 本会に、本会を代表する会長1名をおく。

- 2 本会に、常任理事若干名をおく。
- 3 本会に、常任理事を代表する常任理事長を1名おく。
- 4 本会に、全国会計監査1名をおく。

第8条 本会に次の委員会をおく。各委員会に関する規程は別に定める。

1. 編集委員会
2. 全国プログラム委員会
3. 学会賞選考委員会
4. ホームページ管理運営委員会
5. 選挙管理委員会

第9条 本会は毎年研究大会および会員総会を開く。

第10条 本会の運営にかかわる重要事項の決定は、会員総会の承認を得なければならない。

第11条 本会の会計年度の起算日は、毎年4月1日とする。

- 2 機関誌の発行等に関する全国会計については、理事会が、全国会計監査の監査を受けて会員総会に報告し、その承認を受ける。

第12条 本会会則の改正、変更および財産の処分は、理事会の審議を経て会員総会の承認を受けなければならない。

付 則 1. 本会は、北海道、東北、関東、関西、九州に支部をおく。

2. 本会に研究部会を設置することができる。
3. 本会の事務所を東京都町田市相原4342法政大学日本統計研究所におく。

1953年10月9日（2010年9月16日一部改正[最新]）

# 教育用擬似マイクロデータの作成

— 平成16年全国消費実態調査を例として —

山口幸三\*・伊藤伸介\*\*・秋山裕美\*\*\*

## 要旨

公的統計のマイクロデータの利用を推進する上での課題として、マイクロデータを利用した実証分析ができる人材を育成していくことが挙げられる。人材育成のためには、利用において統計法令の制約も受けず、大学等の教育機関における授業や演習などで自由に利用できるマイクロデータが必要と考えられる。このような目的のために作成したのが教育用擬似マイクロデータである。本稿は教育用擬似マイクロデータの作成方法を提示している。教育用擬似マイクロデータの作成においては、調査票情報（個票データ）から高次元の集計表を作成し、集計表の各セルの量的属性値が多変量（対数）正規分布に従うことを仮定し、多変量（対数）正規乱数を生成する方法を採っている。

## キーワード

マイクロデータ、教育用データ、マイクロアグリゲーション、超高次元クロス集計表、多変量対数正規乱数

## 1. はじめに

新統計法<sup>1)</sup>が平成19年5月に公布され、平成21年4月に全面施行された。新統計法において、公的統計は「国民にとって合理的な意思決定を行うための基盤となる重要な情報である」とされ、言うなれば国民の共有財産と位置付けられている。そうした理念の下に、公的統計の利用促進のために、統計データの

二次利用<sup>2)</sup>に関する制度が設けられ、学術研究や高等教育の発展に資する場合に、委託による統計表の作成及び匿名データの提供ができることになっている。

この二次利用制度が開始され、匿名データの提供は、定着しつつあるものの、新統計法に規定されている利用目的の制約、利用環境の制約を受けざるを得ない。そのため、多数の学生を対象とした大学等での講義や統計演習などの利用は、現実問題として困難である。このようなことから、利用において統計法に制約されない統計データの開発が、統計委員会等で議論され、大学の研究者からも要望されていた。こうした背景から、自由に利用できる教育用擬似マイクロデータの開発を計画し、平成16年全国消費実態調査の教育用擬似マイクロデータを統計的な手法を用いて作成し、

\* (独)統計センター

19-1 Wakamatsu-cho Shinjuku Tokyo 162-8686 Japan  
TEL 03-5273-1285

e-mail : kyamaguti@nstac.go.jp

\*\* 明海大学経済学部

1 Akemi Urayasu Chiba 279-8550 Japan

TEL 047-355-5120(内線1419)

\*\*\* (独)統計センター

19-1 Wakamatsu-cho Shinjuku Tokyo 162-8686 Japan  
TEL 03-5273-1188(内線8394)

実際に試行的に提供できるまでに至っている。

本稿では、まず、教育用擬似マイクロデータを作成するに至った背景と目的について述べ、次に、教育用擬似マイクロデータ作成上の基本的な考え方を示し、その基本的な考え方に基づいた平成16年全国消費実態調査を例とする作成方法を述べる。最後に今後の課題を提示する。

## 2. 背景と目的

### 2.1 統計法の改正

新統計法（平成十九年法律第五十三号）は、統計に関する基本法として、旧統計法（昭和二十二年法律第十八号）を全部改正し、統計報告調整法（昭和二十七年法律第百四十八号）の廃止とともに、平成19年5月16日に成立した。この新統計法は、平成19年5月23日に公布され、戦後、統計制度が再建されて以来60年振りの抜本的改革となった。そして、平成21年4月1日に新統計法が全面施行された。

統計法の改正は、「行政のための統計」から「社会の情報基盤としての統計」へ転換し、公的機関が作成する統計が、より体系的・効率的に整備され、国民・事業者にもより使いやすいものとなることを目指しており、①公的統計の体系的かつ効率的な整備及びその有用性の確保を図るため、公的統計の整備に関する基本的な計画の策定、②統計データの利用促進に関する措置、③統計調査の対象者の秘密保護の強化、④統計整備の「司令塔」機能の強化が主な内容となっている。

### 2.2 新統計法の下でのマイクロデータの利用

新統計法においては、先述のとおり、統計データの利用促進がうたわれていて、統計の研究や教育などの公益に資する場合に限り、マイクロデータを利用することが可能になっている。新統計法の下でのマイクロデータの利用については、旧統計制度においても利用され

ていた方法と同じ利用方法である第三十三条による調査票情報の利用、新統計制度における新たな仕組みである第三十四条による委託による統計の作成等（オーダーメイド集計）及び第三十五条・第三十六条による匿名データの作成・提供がある。これらのマイクロデータの利用を二次利用と称している。

新たな仕組みのうち、オーダーメイド集計とは、統計の作成等を希望する者が調査実施者に個別の委託集計を申し出て、その申出を受けて調査実施者が集計し、委託申出者に集計結果を提供する方式のマイクロデータの利用である。委託申出者が直接調査票情報を利用しないので、秘密の保護が確実に保たれる。このように秘密の保護が担保されるので、高度の公益性を満たさなくても、学術研究の発展に資する場合及び高等教育の発展に資する場合の一定程度の公益性が認められる場合に利用が容認されている。なお、公益性は、その利用目的によって判断される。

匿名データの提供は、調査票情報を特定の個人又は法人その他の団体の識別（他の情報との照合による識別を含む。）ができないように加工した匿名データを、一般の利用に供する方式でのマイクロデータの利用である。

匿名データの作成については、調査実施機関が作成することになっており、基幹統計調査に係る匿名データは統計委員会に諮問し、答申を得なければならない。一般統計調査の匿名データについては、統計委員会に諮問し、答申を得る必要はないが、基幹統計調査と同様の匿名化措置を施さなければならないとされている。

匿名データについては、匿名化されることにより、秘密の保護が図られているので、第三十四条と同様に、高度の公益性を満たさなくても、学術研究の発展に資する場合、高等教育の発展に資する場合並びに国際社会における我が国の利益の増進及び国際経済社会の健全な発展に資すると認められる場合の一定

程度の公益性が認められる場合に利用が認められている。

匿名データは匿名化措置を施し、秘密の保護が図られているとしても、調査票情報であることには変りがないので、その管理には十分な注意が必要である。そのため、匿名データの提供におけるガイドライン<sup>3)</sup>では、利用者が申出する際には、管理するための条件が掲げられている。

### 2.3 二次利用の実績

新統計法において、各府省は基本計画の着実な推進が求められることになっており、各府省からの法の施行状況の報告を総務省政策統括官（統計基準担当）において、「統計法施行状況報告」として取りまとめている。

この「統計法施行状況報告」から二次利用の実績をみると、平成21年度に、一般からの申出によって、オーダーメイド集計の結果を提供した件数は、4件となっている。22年度において、集計結果を提供した件数は12件となっている。

他方、平成21年度に、一般からの申出によって、匿名データを提供した件数は、20件となっている。22年度において、提供した件数は38件である。

このように平成21年度及び22年度のオーダーメイド集計、匿名データの提供の利用実績をみると、大きな期待をもって開始されたにもかかわらず、申出件数は数少ない状況であった。22年度は、21年度に比べて、利用者も、利用件数も着実に増えており、二次利用制度も認知されつつあると思われるが、制度を周知する広報だけでは、利用者の拡大はあまり望めない状況と考えられる。

また、統計委員会、匿名データ部会等では、新統計法の全面施行の準備をする段階で、いわゆるレプリカデータや統計教育・訓練用データの必要性が議論され、このようなデータが要望されていた。二次利用制度の広報活

動を行っている中で、研究者や統計教育の関係者の方からも、このようなデータについての要望が多数あった。政府も二次利用の仕組みを考える際に、課題の一つとしてレプリカデータを取り上げ、「統計データの二次利用促進に関する研究会報告書」に記載しているが、その後、特にレプリカデータに関する具体的な議論はされなかった。

こうしたデータの必要性については、以前からも一部の研究者において指摘されていた。すなわち、日本学術会議学術基盤情報常置委員会（2005）は、リサンプリングとスワッピングによって作成し、研究者に自由に配布できるレプリカデータを提言していた。また、松田(2008)は、①個別の回答者とのリンケージが不可能な、②特別な手続きを必要としない、③大学院生にも自由に使えるレプリカデータの作成を計画していたことに言及している。このレプリカデータの作成については、日本学術振興会の科学研究費補助金による「マイクロ統計データ活用研究会」（研究代表松田芳郎、井出満、森博美）において、平成17年度に匿名化の度合いを高めた教育用の匿名標本データを作成し、大学院生に利用させることが計画されていた。しかし、計画に対する承認が下りず、レプリカデータの作成については断念している。また、実際にマイクロデータ提供を実施している一橋大学経済研究所の経験を踏まえると、実証分析ための教育用データの必要性を指摘することができる（山口(2008)、小林(2011)）。

### 2.4 教育用擬似マイクロデータの開発

想定していたよりも低調であったマイクロデータの利用を促進していくための1つの方策として、マイクロデータを用いた実証分析ができる人材を拡大または育成することが考えられる。未だマイクロデータを利用していない研究者には、まずマイクロデータを使って、データ特性等を理解してもらい、若手の研究者や

学生には、マイクロデータを用いた実証分析の演習等を行ってもらい、そうしたことができる環境を整備することが肝要である。環境を整備することが、実証分析ができる人材を拡大・育成し、その結果として、マイクロデータを用いた実証研究を発展させ、学術研究水準を向上させることになると考えられる。

そのために、未だ利用経験のない研究者、若手の研究者、学生等に実際のマイクロデータを自由に利用させることを実現しなければならない。匿名データの利用には、高等教育目的で利用できることになっているものの、統計法令に定められた利用目的や利用環境などの要件を満たさなければならない制約があり、かならずしも自由に使えるというわけではなく、多くの学生を相手にした大学での統計演習などで利用するには現実的に困難な場合がある。美添(2009)は、匿名化措置を強めた匿名データを一般用、教育用として作成し、それぞれ一般社会人、大学生以上の高等教育を受ける人に自由に使えることを提言している。星野(2010)は匿名データを更に匿名化したマイクロデータはPublic Useとして利用すべきとしている。しかしながら、新統計法に基づく統計制度では、そうした利用は想定されていないので、提言を実現できないのが現状である。

そこで、利用において統計法令に縛られない、何の制約も受けない自由に使える教育用擬似マイクロデータの開発を計画することとした。計画した教育用擬似マイクロデータとは、本来の調査票情報である個票データを集計表としてまとめ、個票データとの関連を断ち切った上で、その集計表に基づいて、マイクロデータの形式を持つ擬似的なデータ（以下「擬似マイクロデータ」と呼称）を作成することとした。作成に当たっては、統計法の第三十三条の規定に基づいて申出を行い、提供を受けた調査票情報（個票データ）を使用している。この擬似マイクロデータは、調査票情

報と異なる。そして、このような擬似マイクロデータを、高等教育等の利用が主たるものとして、教育用擬似マイクロデータ<sup>4)</sup>と称している。

改めて、実証分析の教育の枠組みを想定すると、次のような段階を経ていくことができると考えられる。まず、①統計調査の調査方法、調査票、標本設計、推定方法などを理解する。次に、②報告書等の既存の統計表データを用いた分析による実態把握を行う。その後、③教育用擬似マイクロデータを用いた演習によってマイクロデータの特長、取扱い方、統計的分析手法を習得する。最後に、④匿名データを用いた実証分析する段階に進み、学術研究、学位論文作成を行う。このような教育の枠組みにおいては、教育用擬似マイクロデータは、実証分析のための環境を整備するために必要不可欠なツールであると考えられる。

### 3. 擬似マイクロデータの基本的な考え方

#### 3.1 調査票情報と匿名データ

現在、新統計法で提供されている統計データには、調査票情報、オーダーメイド集計、匿名データの3つがある。どのデータも法令に規定されており、利用する上では制約がある。

新統計法における調査票情報とは、「統計調査によって集められた情報のうち、文書、図画又は電磁的記録に記録されているもの」、すなわち調査客体の調査票ごとのデータであり、「特別の定めがある場合を除き、その行った統計調査の目的以外の目的のために、統計調査に係る調査票情報を自ら利用し、又は提供してはならない」と規定されている。「特別の定めがある場合」以外には利用することはできない。匿名データは、調査票情報を加工したもので、調査票情報の一種と考えられている。統計法施行規則（平成十九年総務省令第百十二号）に定められた要件を満たさなければ利用することはできない。このことは、

匿名化措置を施されていても、自由には利用できず、どれほど匿名化を強めたとしても、調査票情報には変わらないことを意味する。

したがって、自由に利用できるデータとして教育用のマイクロデータを作成するためには、調査票情報から作成することはできないので、別の方法によって作成することが求められた。

### 3.2 教育用擬似マイクロデータ

調査票情報から作成したものは、調査票情報であることを踏まえて、教育用擬似マイクロデータでは、個票データから高次元の集計表を作成し、その高次元の集計表から個票データに近似したマイクロデータを作成するという方法をとっている。集計表から作成するために、個票データでも、匿名データでもない擬似的なマイクロデータと言える。それでいて、この教育用擬似マイクロデータは、実証分析に利用した際に、我が国の実態を反映できるように、つまり個票データの分布にできる限り近似するように工夫して作成する方向で考えた。

このように集計表から作成する教育用擬似マイクロデータは、基本的に、①個票データの分布に近づけるなど、元の個票データに近似したデータであること、②量的属性の相関関係を保つなど、量的属性間の関係が整合的であること、③全国消費実態調査で言えば収入総額と支出総額が合致しているなど、調査特有のデータ構造を保持すること、④標本調査における集計用乗率を考慮すること、⑤データ量は元の個票データに合わせることを考えの下で作成している。作成例としての全国消費実態調査における考慮点として、質的属性については、集計表の作成における分類事項が該当し、その項目数は限られたものになり、量的属性については、分析上必要と思われる収入項目、支出項目を収録する<sup>5)</sup>。

このように教育用擬似マイクロデータの作成に当たっての基本的な考え方は、以上のとお

りである。これを具体的にどのような統計的な手法を用いて作成したのかは、平成16年全国消費実態調査を例として、次節で述べる。

### 3.3 高次元の集計表とマイクロアグリゲーション

教育用擬似マイクロデータでは、個票データから高次元の集計表を作成し、その高次元の集計表から個票データに近似した擬似マイクロデータを作成するという方法をとっている。この教育用擬似マイクロデータの基になる高次元の集計表の基本的な考え方を述べることにする。

統計作成機関が統計調査を実施し、その調査結果を結果表（又は統計表）として、報告書等で公表する。公表される結果表は、一般的に基本的と考えられる調査事項をクロスさせた集計表であり、低次元の集計表として作成されている。高次元の集計表を作成することは少なく、その公表時期に注目されている問題に対応するために、まれに高次元の集計表を作成することはありえる。

教育用擬似マイクロデータの作成においては、擬似マイクロデータとして収録する分類事項については、その調査事項をすべて使って高次元の集計表の作成をする必要がある。クロスさせなかった調査事項については、データとして収録することはできない。したがって、擬似マイクロデータに収録すべき事項は、基本的な調査事項に限定してもその数は多くならざるを得ない。利用する者にとっては、基本的な調査事項は最低限必要と考えられるからである。平成16年全国消費実態調査による擬似マイクロデータでは14項目を選定して、収録することにしたが、この14項目をクロスさせた集計表を作成し、それが擬似マイクロデータ作成の基になる集計表である。

擬似マイクロデータを作成する方法としては、高次元の集計表から作成する外に、調査票の調査事項ごとに確率分布を作成し、その確率

分布から擬似マイクロデータを作成する方法なども考えられる。集計表から作成するという考えは、特別なものではなく、これまでも1つの方法として指摘されている<sup>6)</sup>。統計制度において、集計表は調査票情報から作成されるが、作成された集計表は調査票情報ではない。そうでなければ報告書等に掲載される結果表は調査票情報になり、公表できないことになる。今回、集計表から作成することとしたのは、調査票情報と集計表とが切り離され、別のものと理解されているからである。実際に、統計法（平成19年法律第53号）第三十三条の規定に基づいて、総務省に全国消費実態調査に係る調査票情報の提供を申し、集計表からの教育用擬似マイクロデータの作成及び利用者への提供が承認されている。さらに、政策統括官（統計基準担当）からも、集計表から擬似的な個票のデータを生成する方法に関して問題なしとの見解を得ている。

ところで、教育用擬似マイクロデータの基になる高次元の集計表の作成については、マイクロデータに対する匿名化技法の1つであるマイクロアグリゲーション（Microaggregation）が方法的に展開されたものと考えられることのできる（Bethlehem *et al.* (1990), Höhne (2003)）。マイクロアグリゲーションとは、マイクロデータ（個票データ）を  $k$  個（ $k$  は閾値（threshold））のレコードを有する同質的なレコード群にグループ化した上で、そのレコードにおける個々の属性値を平均値等の代表値に置き換えることであって（伊藤 (2008)<sup>7)</sup>、ヨーロッパ諸国を中心に、事業所・企業系のマイクロデータにおける匿名化技法として、マイクロアグリゲーションに関する調査研究が進められてきた<sup>8)</sup>。

マイクロアグリゲーションは、一般に、マイクロデータに含まれる量的属性に対して適用される。それに対して、質的属性については、対象となる質的属性のおのおのにおいて同一の属性値を有するレコードをグループ化した

場合、それらの属性値をグループの代表値への置き換えとみなせば、質的属性値に関するレコードのグループ化もマイクロアグリゲーションの一形態として位置付けることが可能である。その場合、マイクロアグリゲートデータは、特定のグループ内で同一の質的属性値群とそれに対応する量的属性の平均値を含むレコード群と考えられる。このようなマイクロアグリゲートデータは、質的属性値群と量的属性の平均値群から構成された個票データに準じたデータとみなすことができるが、各レコードが持つ属性値群は、あくまで集計値として位置付けられている。

一方、グループ化の対象となる質的属性を分類事項とした集計表を作成することが可能であるが、この集計表におけるある特定のセルの度数とマイクロアグリゲートデータにおいて対応するグループ内のレコード数は一致する。このことは、集計表が高次元になるにしたがって、グループ化において使用する質的属性の数が増えることを意味している。こうした議論を展開することによって、「個別データが有するすべての属性群を集計事項の対象とした上で作成される  $n$  次元の多重クロス集計表」である「超高次元クロス集計表」を考えることができ（伊藤 (2008)）、その集計表に含まれるセルと対応関係を持ったレコード群から構成されるマイクロアグリゲートデータが理論的に設定可能となる。なお、超高次元クロス集計表は、一次元から  $n$  次元までのあらゆる次元のクロス集計表を包含している。このことは、超高次元クロス集計表の枠組において、擬似マイクロデータ作成の基になる高次元の集計表を設定するための様々な次元のクロス集計表が作成可能であることを意味している。

他方、マイクロアグリゲーションでは、個票データに含まれるレコードを閾値  $k$  のレコード群にグループ化した上で、グループ内のレコードにおける個々の属性値を平均値等の代

表値に置き換える。この場合、対象となる属性群について同一の属性値を有するレコード群（以下「同質属性値レコード群」と呼称）に存在するレコード数は、同じ属性群を分類事項として作成した超高次元クロス集計表におけるセルの度数と対応関係にある。したがって、同質属性値レコード群内に含まれるレコード数の下限が決まれば、超高次元クロス集計表に含まれるセルの度数に関する閾値が確定する。閾値  $k$  を設定した場合、超高次元クロス集計表の分類事項となる属性群から、属性の組合せを適当に選択することによって、超高次元クロス集計表に含まれるすべてのセルが  $0$  以外の  $k$  未満の数にならないように集計表を構成することができる。また、超高次元クロス集計表において閾値  $k$  未満のセルが存在する場合、そのセルに該当するレコードの属性群に対して不詳による処理を行うことによって、閾値  $k$  以上のレコードを持つ同質属性値レコード群へのグループ化を行うことも可能である。

このように、マイクロアグリゲーションの枠組みにおいて超高次元クロス集計表に基づく個票データに準じたデータの作成を方法的に位置付けることが可能になる。このことは、教育用擬似マイクロデータの作成方法においてマイクロアグリゲーションが方法的な基礎を成していることを示すものである。

#### 4. 教育用擬似マイクロデータの作成方法

本節では、平成16年全国消費実態調査を例に、教育用擬似マイクロデータの作成方法を述べることにしたい。先述のとおり、教育用擬似マイクロデータの基本的な考え方は、集計表から個票データとは異なる擬似マイクロデータを作成することにある。集計表は、一般に表頭及び表側に用いられる分類事項と集計量として表される集計事項（度数等）から構成されるが、それは、擬似マイクロデータにおいてはそれぞれ質的属性と量的属性に対応する。

そこで、教育用擬似マイクロデータの作成においては、最初に、擬似マイクロデータに含める質的属性と量的属性を選択する。質的属性と量的属性の選択については、集計表の分類事項によって分割された擬似マイクロデータのレコード群内における度数1又は2の割合が考慮される。次に、集計表の度数1又は2のセルに該当するレコードが出現しないような集計表を作成した上で、作成した集計表のセルごとに多変量正規乱数を発生させることによって量的属性値を作成する。さらに、量的属性については、乱数によって作成した量的属性値から合計値及び内訳値を作成し、収入と支出のバランス調整を行う。最後に、教育用擬似マイクロデータにおける集計用乗率を付与する。以下では、教育用擬似マイクロデータの具体的な作成手順を述べる。

##### 4.1 量的属性と質的属性の選択

教育用擬似マイクロデータ作成における第1の手順は、平成16年全国消費実態調査に含まれるすべての属性の中から擬似マイクロデータに含める量的属性と質的属性を選び出すことである。教育用擬似マイクロデータは集計表に基づいて作成されることから、量的属性と質的属性の選択は、集計表における分類事項と集計事項の探索的な設定ととらえられる。分類事項の選択によっては、公表されている結果表として存在しない高次元の集計表を作成することができる。

教育用擬似マイクロデータの作成における質的属性は、男女別、年齢、就業・非就業の別といった世帯主に関する調査事項と世帯区分、世帯人員階級といった世帯に関する調査事項に区分されている。本稿では、これらの質的属性の中から、12属性、13属性、14属性、16属性と18属性の5つのパターンに関する高次元の集計表が想定されている。

ところで、高次元の集計表においては、度数1又は2のセルが出現する可能性がある。

度数1に該当するレコードは、個票データと1対1で対応することから、それは個票データとみなされるおそれがある。また、セルに度数1又は2が存在する場合、秘匿性の観点から、公表されている結果表においては、そのセルに「X」等の秘匿処理が施されていることが少なくない。このことから、度数1又は2については、集計表内のセルに出現しないような処理を施す必要がある<sup>9)</sup>。

集計表の分類事項の選択によって、集計表に出現する度数1又は2となるセルの数は異なる。表1は、検討した質的属性の数及び度数1、2と3以上に該当するレコード数及びセル数である。表1のように、集計表における分類事項として用いられる質的属性が多くなるにつれて、集計表に含まれるセルの総数が増大するから<sup>10)</sup>、度数1又は2が出現するセル数も多くなる。そこで、教育用擬似マイクロデータの作成においては、度数1又は2の出現数を考慮した上で、擬似マイクロデータに含める質的属性の数及び種類を選択した。さらに、質的属性の選択においては、①世帯主及び世帯に関する基本的な調査事項であること、②旧統計法の目的外使用（第十五条第二項）で提供している属性や結果表で用いられている属性の中で使用頻度が高いことを考慮した。その結果、教育用擬似マイクロデータに含める質的属性として次の14属性を選別した。一方、量的属性については、本稿で用いている作成方法において提供することが可能な184属性を選択している。

世帯主事項(6)：男女別，年齢5歳階級，就

業・非就業の別，企業区分，企業規模，産業分類

世帯事項(8)：世帯区分，世帯人員階級，就業人員階級，住居の建て方，住居の構造，入居時期・入居年

したがって、教育用擬似マイクロデータで提供する属性は、個票データに含まれる全属性から選び出した、質的属性14属性、量的属性184属性、及び集計用乗率の全199属性である。

#### 4.2 度数1又は2に該当するレコードの処理

教育用擬似マイクロデータの作成における第2の手順は、擬似マイクロデータ用の高次元の集計表のセルにおいて度数1又は2に該当する個票データのレコードに対して、度数3以上のセルとなるような処理を施すことである。

後述するように、教育用擬似マイクロデータの作成において多変量正規乱数を発生させる必要があることから、高次元の集計表の集計事項として算出されるのは、度数だけでなく、平均、分散・共分散である。なお、平均、分散・共分散は、集計用乗率を乗じることによって計算された重み付きの統計量である。この重み付きの平均、分散・共分散については、度数のように集計表上で作成することができないために、元の個票データから作成する必要がある。そのため、集計表の度数1又は2に該当するレコードにおいて、質的属性の一部の値を不詳に置き換えている。このような度数1又は2に該当するレコードの処理によって、多変量正規乱数の生成で用いる集計

表1 検討した質的属性の数と度数1、2と3以上に該当するレコード数及びセル数

	12属性	13属性	14属性	16属性	18属性
度数1	4,612	13,583	22,583	26,549	46,255
度数2	2,954	4,084	5,806	6,918	3,526
度数3以上	47,490	37,387	26,667	21,589	5,275
セル数	9,505	18,538	28,481	32,897	49,084

表の作成が可能になる。

不詳扱いにする質的属性については、あらかじめ質的属性の有用性を考慮した上で不詳を適用する属性の優先順位を決め、それに従って、該当する質的属性値に対して不詳の付与を行った。なお、世帯事項の世帯区分、世帯人員階級、就業人員階級、世帯主事項の男女別については、基本的な事項であるため、不詳となる属性から除外した。さらに不詳に関する処理を行っても、度数3以上にならないデータは削除した。不詳を適用する属性の優先順位は次のとおりとなった。

入居時期・入居年>住居の構造>住宅の所有関係>住居の建て方>職業符号>産業符号>企業規模>企業区分>就業・非就業の別>年齢5歳階級

度数1又は2の処理に関する具体的な方法は、次のような手順で行われる。最初に、分類事項14項目のクロス表を作成し、クロス表の度数1又は2になるセルに該当するレコードについて、不詳を適用する属性の優先順位に従って、入居時期・入居年の項目を不詳に置き換える。次に、その不詳処理をしたレコードを使って、再度、クロス表を作成し、その再集計したクロス表の度数1又は2になるセルに該当するレコードの住居の構造の項目を不詳に置き換える。クロス表に度数1又は2のセルがなくなるまで、このように不詳処理を繰り返す。

### 4.3 高次元の集計表の作成

第3の手順は、度数1又は2を処理した個票データを用いて、分類事項を質的属性、集計事項を度数、基本的な量的属性の平均及び分散・共分散の高次元の集計表を作成することである。本稿では、量的属性の共分散を計算するために、相関係数行列を計測する。なお、相関係数行列の算出においては、選択された14の質的属性の中から、量的属性の値が0にならないように質的属性を限定した上

で、質的属性によってグループ化されたレコードごとに相関係数行列を算出した。

集計表の作成方法は、以下の①～③の手順で行われる。

- ① 世帯区分別にグループ化した上で、量的属性値が0となるレコードが除外されたレコード群について相関係数行列を計算する。
- ② 質的属性ごとにさらにグループ化した上で度数を計測するだけでなく、量的属性ごとに属性値が0となるレコードを除外したレコード群について、平均、標準偏差を計算する。
- ③ 質的属性ごとにグループ化されたレコード群別の分散・共分散に関しては、②で求めたレコード群別の標準偏差に、①で求めたそのレコード群に対応する相関係数行列を乗じることによって求められる。

なお、平成16年全国消費実態調査の量的属性値は、年間収入や消費支出などの収支金額で、必ずしも正規分布に従わないため、量的属性値が対数正規分布に従うことを仮定した。このため、作成する集計表も、度数1又は2を処理した個票データの量的属性値を対数変換し、各統計量を計算している。

### 4.4 多変量正規乱数の生成

第4の手順は、乱数を発生させることによって擬似マイクロデータを生成することである。乱数の発生方法については、最初に、①乱数を発生させずレコード群内の平均値を当てはめる方法（以下「平均法」という）、②単変量正規乱数法、③2変量正規乱数法の3つの方法が、年間収入と消費支出の2つの量的属性を用いた実験によって検討された。

#### ① 平均法

集計表の質的属性別の平均値を用いて、マイクロデータ形式でセル内の度数分のデータを作成する。

## ② 単変量正規乱数法

個票データの量的属性について、セル内における各属性の値が属性ごとに正規分布に従うことを仮定し、各属性の平均及び標準偏差を用いて、個票データのばらつきを加味した正規乱数を生成する。

## ③ 2変量正規乱数法

個票データの量的属性のなかの2属性が関連性を持ちながら、正規分布（2変量正規分布）に従うことを仮定し、各属性の平均及び2属性間の相関係数を用いて、個票データにおける量的属性の分散を加味した正規乱数を生成する。

上記の3つの方法を用いて作成した擬似マイクロデータにおける年間収入と消費支出の相関係数が、表2に示されている。表2を見ると、平均法における度数分布は個票データと比較して大きく異なることが確認できるが、2変量正規乱数法のように、2つの属性の関連性を考慮した形で乱数を発生させた場合には、擬似マイクロデータにおける相関係数が個票データのそれに近似していくことがわかった。

本実験では、量的属性として年間収入と消費支出の2属性で行ってきたが、教育用擬似マイクロデータにおいては、数多くの量的属性を含むデータセットの作成が指向されていることから、本稿では、多変量の属性を用いた

作成方法を検討した<sup>11)</sup>。

## ④ 多変量正規乱数法

個票データの量的属性について、質的属性ごとにグループ化されたレコードにおける属性値が多変量正規分布に従うことを仮定し、各属性の平均値、分散と属性間の共分散を用いて、個票データのばらつきを加味した多変量正規乱数を生成する。

対象となる量的属性については属性値に0が含まれるために、対数変換を行うことができない。そこで、当該量的属性値の0の有無にかかわらず、量的属性値に1を加えてから対数変換し、教育用擬似マイクロデータ用に作成した集計表を基に生成した乱数を実数に戻してから1を引く処理を行った。

このように、多変量正規乱数法を用いて作成することによって、数多くの量的属性が擬似マイクロデータに設定可能になるだけでなく、①平均法、②単変量正規乱数法、及び③2変量正規乱数法と比較して、個票データにより近い分布特性を持つ擬似マイクロデータの作成が期待されるが、量的属性によっては、擬似マイクロデータにおける標準偏差が個票データのそれよりも大きくなる属性が存在する。その理由として、擬似マイクロデータの作成において、個票データの対数を取り、多変量対数正規乱数を生成した後、指数変換によって実数値に戻すような処理を行っているために、

表2 年間収入と消費支出の相関係数

個票データ			①平均法		
	年間収入	消費支出		年間収入	消費支出
年間収入	1.00		年間収入	1.00	
消費支出	0.41	1.00	消費支出	0.68	1.00
②単変量正規乱数			③2変量正規乱数		
	年間収入	消費支出		年間収入	消費支出
年間収入	1.00		年間収入	1.00	
消費支出	0.22	1.00	消費支出	0.44	1.00

右に裾が長い分布となる可能性があることが考えられる。そこで、多変量正規分布の標準偏差に基づいて乱数の生成可能な区間を設定し、その閾値を超えないように乱数を生成する方法を採用した。具体的には、乱数の生成可能な区間として、 $4\sigma$ 、 $3\sigma$ 、 $2\sigma$ 、 $1.5\sigma$ 、 $2\sigma$ （住居のみ $1.5\sigma$ ）の5つのケースを比較・検討した。その結果、 $2\sigma$ （住居のみ $1.5\sigma$ ）に基づいて作成された擬似マイクロデータが個票データに最も近似していることが明らかになったことから、教育用擬似マイクロデータの作成においては、乱数の生成可能な区間として、 $2\sigma$ （住居のみ $1.5\sigma$ ）を採用した（表3）。住居の標準偏差が他の収支項目によりも過大なのは、データの中に持家世帯と借家世帯が混在し、持家世帯では住居の主な支出金額である家賃はほとんど支出しておらず、一方借家世帯は世帯割合が低く、家賃は比較的高い支出金額になっているためである。

#### 4.5 パターン別集計表を用いた量的属性値0の付与

先述の多変量正規乱数の生成によって作成した擬似マイクロデータには、個票データには含まれる量的属性値の0が存在しない。そこで、教育用擬似マイクロデータ作成における第5の手順として、個票データに近似するように、擬似マイクロデータに量的属性値0を新たに付与した。具体的には、度数1又は2を処理した個票データを用いて、質的属性ごとにグループ化した上で、そのグループ内で量的属性値が0か0以外のパターン別の集計表を作成し、そのパターンに従って、該当する量的属性に0を付与した。具体的な処理の方法の例は以下のとおりである。

- ① 度数1又は2を処理した個票データを用いて、分類事項について質的属性は世帯区分、量的属性は実収入及び繰越金（実収入と繰越金については0の値と0

表3 擬似マイクロデータの乱数生成可能区間別標準偏差

収支項目	個票データ	擬似マイクロデータ					
		処理なし	$4\sigma$	$3\sigma$	$2\sigma$	$2\sigma$ (住居 $1.5\sigma$ )	$1.5\sigma$
年間収入	355.49	362.89	362.89	361.16	327.20	329.66	311.86
実収入	313,118.37	457,774.91	457,774.91	455,730.60	405,187.48	321,234.75	225,193.81
実収入以外の収入	334,179.08	436,428.68	436,428.68	433,544.09	295,466.99	258,169.35	229,319.51
繰越金	85,271.65	143,984.99	143,984.99	136,181.10	90,979.43	91,537.30	63,621.91
食料	29,777.53	29,862.35	29,862.35	29,809.28	27,372.66	27,439.17	26,137.28
住居	53,558.37	581,041.68	581,041.68	580,704.40	361,910.19	52,120.42	44,956.43
光熱・水道	8,041.10	8,096.77	8,096.77	8,069.63	7,652.85	7,680.78	7,094.38
家事・家具用品	16,653.53	28,433.90	28,433.90	28,331.69	13,992.71	13,683.43	9,653.25
被服及び履物	19,391.95	27,703.69	27,703.69	27,624.41	20,481.71	38,291.78	14,776.43
保健医療	19,691.49	33,156.22	33,156.22	31,743.35	21,693.65	21,361.73	14,029.66
交通・通信	87,206.49	103,957.31	103,957.31	88,654.91	75,994.82	77,918.22	49,487.79
教育	49,867.14	112,888.83	112,888.83	105,143.03	54,158.48	62,989.77	46,688.69
教養娯楽	31,476.04	52,690.20	52,690.20	52,671.81	29,103.34	28,384.30	24,240.14
その他の消費支出	100,500.49	137,450.39	137,450.39	137,030.53	100,410.46	100,366.66	78,976.20
非消費支出	55,573.93	100,813.17	100,813.17	100,697.20	64,279.44	62,686.33	50,718.26
実収入以外の支出	405,185.34	421,922.06	421,922.06	420,826.23	396,401.89	378,661.31	294,557.16
繰越金	94,957.51	161,278.30	161,278.30	155,061.72	110,699.37	111,416.33	69,925.80

以外の値の2区分), 集計事項については度数とした集計表を作成する。なお, 実収入と繰越金の組み合わせに関しては, 世帯区分別に, (1)2属性とも0の値の場合, (2)2属性のうちどちらか一方が0の値の場合, (3)2属性とも0以外の値の場合の3つのパターンが考えられる。

- ② ①の集計表において, ある特定の世帯区分(例えば世帯区分が「1 勤労世帯」とする)のグループ内で, 実収入は0以外の値, 繰越金が0の値に該当するパターンの度数は10とする。多変量正規乱数の生成によって作成したレコードの中で, 世帯区分が1のグループに該当するのは70レコードとすると, そのうち10レコードの繰越金の値に0を付与する。なお, 何番目のレコードの繰越金に0の値を付与するかに関しては, 乱数を1~70の範囲で10回発生させ(乱数の値が重複した場合, 再度発生させる), 乱数の値に該当するレコードの繰越金に0の値を付与する。

#### 4.6 加法性と収支バランスの調整

第6の手順は, レコードレベルで加法性を保ち, 収支バランスの調整を行うことである。一般に, 統計調査の調査事項には, 総計と内訳が存在し, その加法性は保たれている。例えば, 全国消費実態調査においては, 消費支出の金額と十大費目分類<sup>12)</sup>の金額の総計は一致している<sup>13)</sup>。このような加法性に基づき, レコードごとに, 量的属性の合計について対応する量的属性を合算し, 作成した。

一方, 全国消費実態調査では, 収入(総額)と支出(総額)の値が一致するように収支のバランスが図られている。ゆえに, 教育用擬似マイクロデータにおいても, 収入と支出のバランスを調整した。具体的には, レコードレベルで支出総額に対する収入総額の割合を算出し, その割合を当該レコードの収入に関す

る各量的属性値に乗じる。

さらに, 量的属性における内訳の値を求めるために, ①度数1又は2を処理した個票データを用いて, 質的属性別に多変量正規乱数により作成した各量的属性値を1としたその内訳の構成比を集計し, ②集計した結果に基づき, 収支バランスの調整を行った擬似マイクロデータを用いて, 質的属性によってグループ化されたレコード群ごとに, 量的属性値にその内訳の構成比を乗じることによって, 内訳の値を付与する<sup>14)</sup>。

#### 4.7 集計用乗率の付与

第7の手順は, レコードに教育用擬似マイクロデータの集計用乗率を付与することである。標本調査では, 結果数値の推定を行うため, 個票データに集計用乗率が付与されている(推定は当該項目の値に集計用乗率を乗じる)。よって, 擬似マイクロデータでも, ①度数1又は2を処理した個票データを用いて, 質的属性によってグループ化されたレコード群について集計用乗率の平均値を求め, ②多変量正規乱数の生成によって作成した擬似マイクロデータの対応するレコード群ごとに集計用乗率の平均値を付与した。

#### 5. おわりに

本稿では, 最初に, 擬似マイクロデータの作成に関する基本的な考え方を論じた上で, 全国消費実態調査の個票データを用いた教育用擬似マイクロデータの作成を行った。なお, 本稿では論じていないが, このように作成した教育用擬似マイクロデータ<sup>15)</sup>の分布特性については, 個票データの分布と比較してより大きな散らばりを持つ場合もあるが, 全般的には個票データのそれにほぼ近似していることがわかっており, 高等教育における授業・演習用として教育用擬似マイクロデータを用いることは有用である。

擬似マイクロデータの作成に関する特徴は,

高次元の集計表に基づいて個票データに近似した擬似的なデータを作成することにある。教育用擬似マイクロデータの作成については、マイクロアグリゲーションの枠組みにおいて方法的に位置付けられるが、この枠組みから導かれた超高次元クロス集計表に基づく個票データに準じたデータの作成・提供は、公的統計のマイクロデータにおける新たな作成可能性を提示するようと思われる。なぜなら、個票データに準じたデータは集計表を基に作成されることから、擬似的なデータの側面を持つものの、個票データと同様の属性群を有していると考えられるからである。このような超高次元クロス集計表に基づいた擬似マイクロデータの作成は、公的統計の二次利用の観点から見ても新たな展開方向を示すものと言える。

一方、本稿では、公的な統計調査の1つである全国消費実態調査の擬似マイクロデータの作成について論じたが、我が国で、公的統計におけるこのような擬似的なマイクロデータを開発したのは、これが最初であり、マイクロデータを用いた実証分析の教育のために意義ある取組みと考えている。今後、この取組みが研

究者・教育者に理解されて、実証分析の教育上の有効なツールとして確立されることを望むとともに、教育上のツールとして、実証分析の研究者の育成に貢献でき、公的統計の二次利用を促進させ、さらには我が国の社会科学分野における研究の水準が向上することを期待するものである。

教育用擬似マイクロデータについては、試行提供という形で、既に研究者や学生に提供してきている。これらの利用者の意見等や実際の作成段階で明らかになった改善点など、擬似マイクロデータの作成における残された課題については検討していきたい。

また、教育用擬似マイクロデータとして、消費行動のための擬似マイクロデータの次は、就業行動と就業意識を明らかにできる就業構造基本調査の擬似マイクロデータの作成を考えている。全国消費実態調査と異なり、就業構造基本調査では、レコードに含まれる量的属性は少なく、その多くが質的属性である。作成方法についても全国消費実態調査の教育用擬似マイクロデータの作成方法をそのまま適用するのは難しいことから、作成方法を再検討する必要があると考えている。

## 注

- 1) 本稿では、統計法（昭和二十二年法律第十八号）を旧統計法、統計法（平成十九年法律第五十三号）を新統計法と称する。
- 2) 新統計法の第三十二条による調査票情報（マイクロデータ）の利用を二次利用という。すなわち統計調査計画時点での統計表作成以外の作表は、当初の目的外とみなされる。新統計法第三十三条、委託による統計の作成等並びに匿名データの作成及び提供を二次的利用としているが、本稿では二次利用で表す。
- 3) 総務省政策統括官（統計基準担当）は、第三十三条による調査票情報の提供、第三十四条によるオーダーメイド集計、第三十六条による匿名データの提供における各府省共通のガイドラインを設定している。
- 4) 教育用擬似マイクロデータの名称について、開発当初においては擬似データと称していたが、外部への提供を考えた場合に、「擬似」よりも主たる目的である「教育」を冠し、自由に使えるデータと言う意味を込めて、教育用マイクロデータとした。その後、匿名データ等の調査票情報でないことを強調するために、「擬似」を追加し、教育用擬似マイクロデータとした。
- 5) 本稿での質的属性は集計表を作成する場合の分類事項、例えば世帯主の男女別、年齢階級別であり、量的属性は集計表の表章事項、全国消費実態調査では収入と支出の項目が該当する。
- 6) 寺崎（2000）、松田（1999）、美添（2009）を参照のこと。

- 7) ミクロデータに適用される匿名化技法は、非攪乱的な (non-perturbative) 手法と攪乱的な (perturbative) 手法に大別される (Willenborg and de Waal (2001))。非攪乱的な手法には、リコーディング (global recoding, local recoding), データの削除 (local suppression), トップ (ボトム)・コーディング等が存在する。一方、攪乱の手法に関しては、ノイズの付加 (additive noise), ラウンディング (rounding) 等が含まれる。なお、マイクロアグリゲーションは、攪乱的な匿名化技法の1つとして位置付けられている。詳細については、例えばWillenborg and de Waal (2001) を参照されたい。
- 8) Thorogood (1999) によれば、ヨーロッパ諸国の企業におけるイノベーションの活動状況を調査した Community Innovation Survey (1994) においては、匿名化技法の1つとしてマイクロアグリゲーションが適用されていることが知られている。
- 9) 後述のとおり、教育用擬似マイクロデータ用の集計表を作成した後に、セルごとに多変量正規乱数が生成される。セルに含まれる度数が1又は2になる場合、多変量正規乱数の生成のために用いられる相関係数行列が計算できない。このような多変量正規乱数の発生の観点からも、度数1又は2が集計表内のセルに出現しないような処理が必要である。
- 10) 厳密には、質的属性の数だけではなく、質的属性の選択肢の数にも影響される。
- 11) 年間収入又は消費支出を量的属性間における基本的な相関関係として設定し、それ以外の量的属性については年間収入及び消費支出との関係性の中に逐次的に位置づける方法もあるが、年間収入又は消費支出以外の属性間の関連性も考慮するのが望ましいと判断し、多変量で行うこととした。
- 12) 十大費目分類は、食料、住居、光熱・水道、家具・家事用品、被服及び履物、保健医療、交通・通信、教育、教養娯楽及びその他の消費支出の10分類である。
- 13) 統計調査によっては、内訳の項目としては存在しないが、その値を合計には含める場合もある。
- 14) 当初は、セル内のデータが、全て同じ量的属性値に対する内訳の構成比とならないように、構成比の集計表を消費支出の分位階級別で作成し、分位階級に該当するデータについては、該当分位階級の構成比を用いることとしていた。その後、開発上の都合からこうした処置を施していないため、実際のセル内のデータは画一的な構成比となっている。
- 15) 作成した教育用擬似マイクロデータは、二人以上の勤労者世帯のみであり、勤労者以外の世帯、単身世帯が含まれていないことに留意する必要がある。本来、対象とすべきであったが、開発上の都合から含めていない。

## 参考文献

- [1] 伊藤伸介 (2008) 「マイクロアグリゲーションに関する研究動向」『製表技術参考資料』No. 10, 3~31頁。
- [2] 小林良行 (2011) 「匿名データの教育用目的利用に関する一考察」『統計学』第100号, 100~105頁
- [3] 総務省政策統括官 (統計基準担当) (2008) 「統計データの二次利用促進に関する研究会報告書」
- [4] 総務省政策統括官 (統計基準担当) (2010) 「平成21年度 統計法施行状況報告」
- [5] 総務省政策統括官 (統計基準担当) (2011) 「平成22年度 統計法施行状況報告」
- [6] 総務省統計局 (2009) 「平成16年全国消費実態調査報告書」
- [7] 寺崎康博 (2000) 「リスト形式による集計表とパターン化変数」松田芳郎・伴 金美・美添泰人 (編著) 『講座マイクロ統計分析ーマイクロ統計の集計解析と技法』日本評論社, 111~122頁
- [8] 統計委員会 (2009) 「第20回統計委員会議事録」
- [9] 統計委員会匿名データ部会 (2009) 「第1回匿名データ部会議事概要」
- [10] 日本学術会議 (2005) 「政府統計・世帯調査等の一次データ (含む個票データ) の体系的保存と活用・公開方策について」学術基盤情報常置委員会報告
- [11] 星野伸明 (2010) 「公的統計マイクロデータ提供制度の課題」『日本統計学会誌』第40巻第1号, 23~45頁
- [12] 松田芳郎 (1999) 『マイクロ統計が描く社会経済像』日本評論社
- [13] 松田芳郎 (2008) 「日本におけるマイクロデータ政府統計活用の新しい夜明け」『統計』第59巻第

12号, 2~9頁

- [14] 山口幸三 (2008) 「政府統計の個票利用と統計法改正」『経済研究』 Vol. 59, No. 2, 139~152頁
- [15] 美添泰人 (2009) 「統計の有効活用に関する展望と課題」『ESTRELA』 No. 181, 9~17頁
- [16] Bethlehem, J.G., Keller, W.J. and Pannekoek, J. (1990) Disclosure Control of Microdata, *Journal of the American Statistical Association*, Vol. 85, No. 409, pp.38-45.
- [17] Höhne, J. (2003) SAFE – A Method for Statistical Disclosure Limitation of Microdata, Joint ECE/Eurostat Work Session on Statistical Data Confidentiality (Luxembourg, 7-9 April 2003), <http://www.unece.org/fileadmin/DAM/stats/documents/2003/04/confidentiality/wp.37.s.e.pdf>
- [18] Thorogood D. (1999) “Protecting the Confidentiality of Eurostat Statistical Outputs”, *Netherlands Official Statistics*, Volume 14, Spring, pp.30-33.
- [19] Willenborg, L. and de Waal, T. (2001) *Elements of Statistical Disclosure Control*, Springer, New York.

## Generating Pseudo Microdata for Educational Use in Japan

Kozo YAMAGUCHI\*, Shinsuke ITO\*\* and Hiromi AKIYAMA\*\*\*

### Summary

One of the key tasks in promoting the use of official microdata is to nurture researchers who are able to conduct empirical research based on such data. The education and training of these researchers requires the free availability of unrestricted microdata for classes and seminars in higher educational institutions. Pseudo microdata for educational use can fulfill this role. This paper suggests an approach for creating pseudo microdata for educational use. Pseudo microdata is generated using multidimensional tabulation on individual data from official statistics. Random numbers are generated based on the assumption that the values of quantitative attributes contained in each cell of the hyper multidimensional tabulation are based on multivariate lognormal distribution.

### Key Words

Microdata, Data for educational use, Microaggregation, Hyper multidimensional tabulation, Multivariate lognormal random numbers

---

\* National Statistics Center

\*\* Faculty of Economics, Meikai University

\*\*\* National Statistics Center

## 編集委員会からのお知らせ

山口秋義（編集委員長）

機関誌『統計学』の編集・発行について

1. 常時、投稿を受け付けます。
2. 各号ごとに投稿の締め切りを設けます。その期日までに受け付けた原稿でも、査読の進捗如何によつては、その号に掲載されないことがあります。
3. 投稿に際しては、2012年9月の総会において改正された「投稿規程」、「執筆要綱」、「査読要領」をご熟読願います。
4. 原稿は編集委員長に宛ててお送り願います。
5. 原稿はPDF形式のファイルとして提出してください。また紙媒体での提出も旧規程に準拠して受け付けます。紙媒体の送付先も編集委員長としてください。
6. 原則としてすべての投稿原稿が査読の対象となります。
7. 今後の締め切りは次のとおりです。  
A：「論文」・「研究ノート」；B：その他
  - (1) 第105号（2013年9月30日発行予定）  
A：2013年7月31日；B：2013年8月31日
  - (2) 第106号（2014年3月31日発行予定）  
検討中（学会HPなどでお知らせします）
8. 次年度（2013年4月－2014年3月）編集委員会メンバー（敬称略）は次のとおりです。  
金子治平（関西，委員長，原稿送付先），西村善博（九州，副委員長），山田 満（関東），橋本貴彦（関西），栗原由紀子（関東）

以上

---

### 編集後記

研究成果をご投稿いただいた会員諸氏に御礼申し上げます。また製版と発送の作業を昭和情報プロセス株式会社様と音羽リスマチック株式会社様にお世話になりました。この場をお借りして御礼申し上げます。本号では山口秋義（編集委員長）が責任編集を務め、前田修也（東北支部編集委員）が発行業務を担当しました。

（山口秋義 記）

---

## 執筆者紹介 (掲載順)

山口 幸三 ((独)統計センター)

伊藤 伸介 (明海大学経済学部)

秋山 裕美 ((独)統計センター)

櫻本 健 (松山大学経済学部)

IRINA ELISEEVA (Dr. Professor, Member-in-correspondence of Russian Academy of Sciences)

## 支部名

## 事務局

北海道	062-8605	札幌市豊平区旭町 4-1-40 北海学園大学経済学部 (011-841-1161)	水野谷武志
東北	986-8580	石巻市南境新水戸 1 石巻専修大学経営学部 (0225-22-7711)	深川通寛
関東	192-0393	八王子市東中野 742-1 中央大学経済学部 (042-674-3424)	芳賀寛
関西	525-8577	草津市野路東 1-1-1 立命館大学経営学部 (077-561-4631)	田中力
九州	870-1192	大分市大字旦野原 700 大分大学経済学部 (097-554-7706)	西村善博

## 編集委員

水野谷武志 (北海道)

前田修也 (東北)

岡部純一 (関東)

良永康平 (関西) [副]

山口秋義 (九州) [長]

## 統計学 No.104

---

2013年3月31日 発行	発行所	経済統計学会 〒194-0298 東京都町田市相原町4342 法政大学日本統計研究所内 TEL 042(783)2325 FAX 042(783)2332 <a href="http://www.jsest.jp/">http://www.jsest.jp/</a>
	発行人	代表者 森 博 美
	発売所	株式会社 産業統計研究社 〒162-0801 東京都新宿区山吹町15番地 TEL 03(5206)7605 FAX 03(5206)7601 E-mail: <a href="mailto:sangyoutoukei@sight.ne.jp">sangyoutoukei@sight.ne.jp</a> 代表者 品 川 宗 典

---

# STATISTICS

---

No. 104

2013 March

---

## Articles

Generating Pseudo Microdata for Educational Use in Japan  
..... Kozo YAMAGUCHI, Shinsuke ITO, Hiromi AKIYAMA (1)

Analysis of IO-based Annual Supply and Use Tables for the Development of QNA  
..... Takeshi SAKURAMOTO (16)

## Note

Introduction of the Theory of Correlation into Russia and E. Slutsky  
..... IRINA ELISEEVA (41)

## Activities of the Society

Activities in the Branches of the *Society* ..... (52)

Bylaws of the Society, Regulation of the Editorial Committee, Prospects for the Contribution  
to the Statistics ..... (57)

---

JAPAN SOCIETY OF ECONOMIC STATISTICS

---