

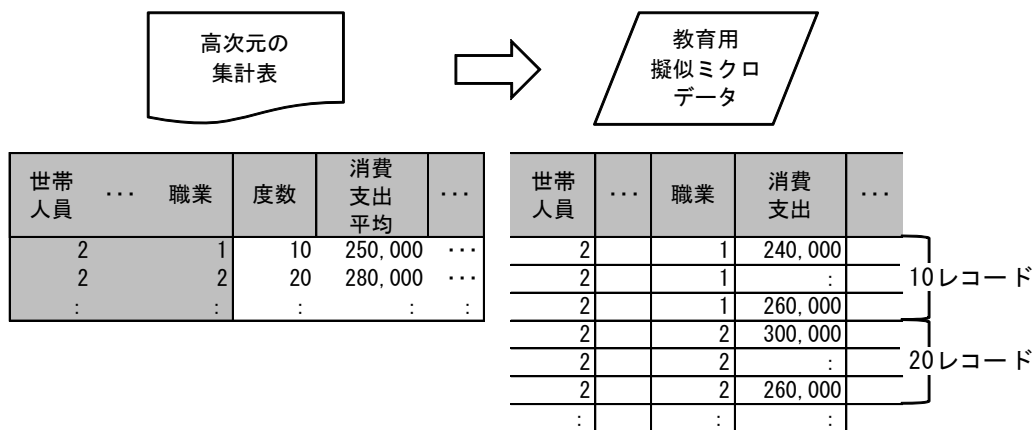
## 教育用擬似マイクロデータの作成方法と試行提供について

秋山 裕美、後藤 武彦、山口 幸三（以上、統計センター）、伊藤 伸介（明海大学）

### 1 はじめに

統計センターでは、公的統計のマイクロデータの利用を推進するために、大学等の高等教育で自由に利用できるデータを想定した『教育用擬似マイクロデータ』の作成方法の研究を行っている。教育用擬似マイクロデータは、個票データ（調査票情報）から集計した高次元の集計表（公表している集計表より詳細な集計表）を基に作成した擬似的なデータで、調査票情報とは異なるが、調査票情報に近似した分布になるように作成している。また、データの仕様は、マイクロデータの形をとっている。第54回(2010年)全国研究大会では、「教育用マイクロデータ作成の試みー政府統計マイクロデータの利用拡大に向けてー」というタイトルで、研究報告を行った。本報告では、昨年度の研究内容を踏まえ、追加検討部分を含めた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について、その特徴を説明する。さらに、現在、予定している試行提供についての紹介を行う。

【教育用擬似マイクロデータの作成イメージ】



### 2 教育用擬似マイクロデータの作成方法

作成方法の基本的な考えは、①個票データから高次元の集計表を作成し、②個票データの量的属性値が多変量（対数）正規分布にしたがうことを仮定し、高次元の集計表を用いて、多変量（対数）正規乱数を生成することである。実際、この方法だけでは、教育用擬似マイクロデータが、個票データに近似した分布にならないため、方法を追加し、次のとおり作成する。

- (1) 調査項目の使用頻度を考慮し、教育用擬似マイクロデータに含める属性を選択する。【属性の選択】
- (2) 個票データを用いて、高次元の集計表（質的属性を分類事項とした度数、量的属性の平均値及び相関係数行列等）を作成する。【集計表の作成】

- (3) 高次元の集計表に、度数 1 がある場合、個票データと 1 対 1 で対応することから、度数 1 に該当するレコードが個票データとみなされる可能性がある。また、度数 2 のセルは、公表している集計表で秘匿されている。そこで、度数 1 及び度数 2 に該当するレコードについては、質的属性値が類似であるレコードを 3 レコード以上にグループ化し、グループ内で同一でない質的属性値を不詳に置き換えることで、高次元の集計表のセルが、度数 3 以上になるようにする。【度数 1 及び度数 2 の取り扱い】
- (4) 高次元の集計表を用いて、セルごとに、多変量（対数）正規乱数を生成する。【多変量正規乱数の生成】
- (5) 量的属性値が 0 か 0 以外かのパターン別集計表を作成し、そのパターンにしたがって、量的属性値に 0 を付与する。【量的属性値 0 の作成】
- (6) 各セルの集計用乗率の平均値を、集計用乗率として付与する。【集計用乗率の付与】
- (7) 量的属性値を合算し、量的属性の合計値を算出する。【加法性に基づいた量的属性値の合計値の算出】
- (8) 支出総額に対する収入総額の比率を算出し、収入に関する各量的属性値にその比率をかけることで収入と支出のバランスを調整する。【収支バランスの調整】
- (9) 量的属性値の内訳について、合計値に対する比率（構成比）を計算し、その比率を用いて按分することにより作成する。【量的属性値の内訳の作成】

### 3 教育用擬似マイクロデータの試行提供

上記の方法を用いて作成した教育用擬似マイクロデータの試行提供を行い、利用者から意見等を募集する予定である。利用者の意見を参考にしながら、さらなる検討を行いたいと考えている。なお、試行提供する教育用擬似マイクロデータは、以下のとおりである。詳細については、独立行政法人統計センターホームページ：<http://www.nstac.go.jp/>に掲載する予定である。

- ・平成 16 年全国消費実態調査 二人以上の勤労者世帯
- ・レコード数：約 32,000 レコード
- ・データ形式：CSV 形式

【教育用擬似マイクロデータのイメージ】質的属性 14 属性、量的属性 187 属性（予定）

世帯情報		世帯員情報			集計用乗率	年間収入	収入		支出総額			
世帯区分	...	世帯主		...			収入総額	...	支出総額	...	消費支出	...
	...	...	年齢	...								
1	...	...	50	...	27	5000	790000	...	790000	...	280000	...

#### 参考文献

山口 幸三、秋山 裕美、後藤 武彦、伊藤 伸介（2010）、「教育用マイクロデータ作成の試みー政府統計マイクロデータの利用拡大に向けてー」、経済統計学会 第 54 回(2010 年)全国研究大会 報告要旨集、52～53 頁

## 厚生労働省で提供するオーダーメイド集計及び匿名データについて

小貫 正子、小松 史子、久住 和弘（厚生労働省 大臣官房統計情報部）

### 【 概要 】

従前、調査実施府省以外による調査票情報の利用には行政目的や科学研究費補助金による学術研究等が要件とされていたが、平成 21 年 4 月に全部改正された統計法では「行政のための統計から社会の情報基盤としての統計へ」と考え方が変わるとともに、近年の高度情報通信技術の発達や学術研究の高度化等を踏まえ、被調査者の秘密を確保した上で学術研究、高等教育の発展に資するオーダーメイド集計、匿名データに関する規定が新たに設けられた。

本発表ではその制度の概要、サービスの内容、留意点等について説明、報告することにより、多くの方に利用いただき、学術研究、高等教育の発展、ひいては社会の発展につなげていただきたい。

（詳細は <http://www.mhlw.go.jp/toukei/itaku/>）

### 1 匿名データの作成・提供（貸与）

統計調査で得られた調査票データについて、被調査者が識別できないように秘匿措置を施した上で、匿名データを作成し、提供（貸与）する。

匿名データは、統計法で適正管理義務、承諾された目的以外での利用、第三者提供等の禁止と罰則が規定されている（匿名データは Public Use File ではない。）。

厚生労働省では、平成 16 年国民生活基礎調査の匿名データの作成準備を進め、内閣府統計委員会においてその秘匿措置及び有用性の妥当性について審議いただき、本年 4 月 22 日に答申され、現在、提供開始に向けて準備を進めているところ。

#### （1）提供する匿名データの種類

本調査は、世帯票を中心に複数種類の調査票で構成されているため、利用者における利便性を踏まえ、単一ではなく複数の調査票を組み合わせた以下の 2 種類の匿名データを提供する。

- ① 世帯票及び健康票のデータセット（世帯数等の推計を想定）
- ② 世帯票、健康票、所得票及び貯蓄票のデータセット（平均所得額等の推計を想定）

#### （2）適用する匿名化技法

リサンプリング、直接的識別情報の削除、特徴的レコードの裾切りによる削除、上限（下限）値をトップ（ボトム）コーディングにより統合、分類をリコーディングにより粗くする等。

### 2 オーダーメイド集計

統計調査で得られた調査票データを活用し、委託に応じた統計表を作成し、提供する。

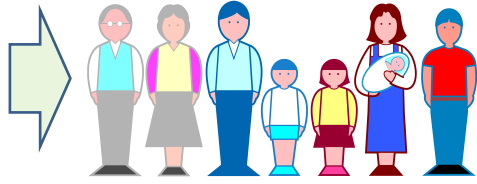
- 提供を行っている調査（対象年次等は WEB ページに掲載）  
人口動態調査（出生票、死亡票）、医療施設静態調査、患者調査、  
毎月勤労統計調査（特別調査票）、賃金構造基本統計調査（個人票）

### 【 利用対象者、手続等 】

- ① 学術研究、学位論文、大学院生や学部学生対象の実習での利用が可能
- ② 事前相談 ⇒ 申出書の提出 ⇒ 依頼書の提出（手数料納付） ⇒ 提供 ⇒ 利用 ⇒ 成果公表 ⇒ 返却（匿名データのみ） ⇒ 利用実績報告書の提出

# 統計調査の調査票情報等の提供について

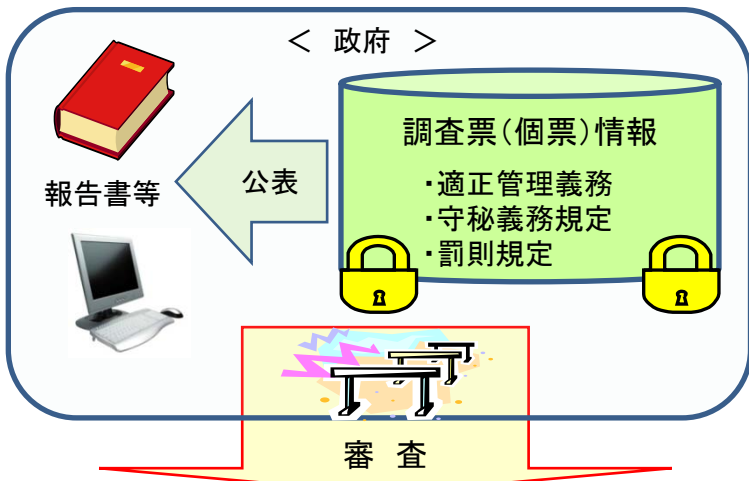
基幹統計調査には、  
報告義務と罰則規定



- ・ 個人情報保護意識の高まり
- ・ 調査環境の悪化等

○ 回収率は年々低下

信頼・協力



個票を利用して、  
報告書に掲載されて  
いない分析（学術研  
究等）が必要。

利用  
申出



＜ 二次的利用の提供  
に関するポイント ＞



【 旧法下 】

- 学術研究
  - ・ 研究論文
  - ・ 学位論文 等
- 高等教育
  - ・ 講義、実習 等

- 高度な公益性
  - ・ 行政目的
  - ・ 科学研究費補助金

- 営利目的
  - ・ 商業利用
  - ・ 商品開発

＜調査票情報の  
提供＞

法改正による二次的利用の  
提供の拡大

【 新法下 】

- ＜オーダーメイド集計＞  
(法第34条)
- ＜匿名データ＞  
(法第35,36条)

(法第33条)  
(変更なし)

(変更なし)

適正管理義務、目的外利用・  
第三者提供の禁止、罰則等

- ・ 万が一、秘匿性が破られると、国民の政府統計に対する信頼が損なわれ、回収率の低下、更には「根拠に基づく政策」の推進に影響が生じる。

- ・ 国民の信頼を維持できる確実な秘匿措置、適切な利用・管理

社会の発展に寄与することを通じて国民に還元

制度に対する国民の理解の広がり