

# 質的変数に関わる擬似マイクロデータについて

独立行政法人統計センター 滝澤 有美

## 1 はじめに

平成 21 年の新統計法(平成 19 年法律第 53 号)全面施行に伴い、総務省統計局は公的統計の二次的利用拡大に向けて、匿名データの提供及びオーダーメイド集計を開始した。一方、研究者、教育関係者等から統計教育・訓練用データの要望が高まり、統計委員会においても統計教育・訓練用データの必要性に関する指摘があったため、当センターにて大学等の教育機関における授業や演習で利用可能な「擬似マイクロデータ」に関する研究を行うこととなった。

擬似マイクロデータは、調査票情報から集計した高次元のクロス集計表の確率分布に基づき、乱数を発生させて作成したマイクロデータ形式のデータであり、調査票情報とは異なるが、公表結果表に近い数値の再現を目的としたものである。

当センターでは、公的統計の二次的利用を推進するための研究活動の一環として、平成 16 年全国消費実態調査データを用いて「擬似マイクロデータ」の作成方法に関する研究を進めてきた。この研究で作成した擬似マイクロデータは、平成 23 年 8 月より当センターにて試行提供を行っている。

先行研究では支出・収入という量的変数を対象としていたが、本報告では、新たに質的変数を対象とした擬似マイクロデータ作成方法の研究結果を報告する。

## 2 作成方法

質的変数のデータは量的変数と異なり、平均値や相関係数に基づく乱数から作成することが困難であるため、乱数発生方法には新たな工夫が必要となる。

本研究では、平成 14 年就業構造基本調査の「有業者」のデータを用いて、以下の方法で擬似マイクロデータを作成した。

- (1) 使用する変数を、多くの公表結果表で用いる基本事項の変数(以下「基本変数」と、これ以外の「加工対象変数」とに分類する。
- (2) 全変数を用いてクロス集計表を作成する。
- (3) (2)のクロス集計表を「基本変数」でグルーピングする。
- (4) 「基本変数」のグループごとに「加工対象変数」の出現確率を求める。
- (5) (4)の出現確率に基づき一様乱数を発生させ、「加工対象変数」の出現確率を変化させる。

この作成方法には以下の利点があると考えている。

- (1) 公表結果表で用いる基本事項の分布を再現可能であるので、統計教育・訓練用のデータとして適当と思われる。
- (2) データ内の変数の関連性をよりよく再現できる。

本報告では、作成方法及び作成した擬似マイクロデータを公表結果表の数値と比較検証した結果について報告する。

## 主要参考文献

秋山裕美ほか(2012)「教育用擬似マイクロデータの開発とその利用～平成 16 年全国消費実態調査を例として～」(『製表技術参考資料』No. 16)