

マイクロデータにおける攪乱的手法の有効性に関する比較研究

独立行政法人統計センター

中央大学経済学部

NTTセキュアプラットフォーム研究所

総務省統計局

吉武 透

伊藤 伸介

菊池 亮

阿久津 文香

総務省統計局は、現在、国勢調査、住宅・土地統計調査、全国消費実態調査、労働力調査、就業構造基本調査、及び社会生活基本調査の6調査の匿名データを作成し、統計センターに委託することにより提供を行っている。これらの匿名データには、リコーディングやトップコーディングといった非攪乱的な匿名化技法が施されているが、平成12年及び17年国勢調査の匿名データでは、加えて、スワッピングという攪乱的な匿名化技法も施されている。

本報告では、平成22年国勢調査の調査票情報（個票データ）に対し、スワッピングだけでなく、PRAM (Post Randomization Method) と呼ばれる攪乱的手法をも用いて作成した匿名化技法が異なる匿名化されたマイクロデータ（以下、「匿名化マイクロデータ」と呼称）間の比較・検証を行うことによって得られた攪乱的手法の有効性を報告する。

本研究では、スワッピング及びPRAMにおける情報特性を実証的に明らかにした上で、複数の変数の組み合わせに対してスワッピングやPRAMを適用した場合に、秘匿性の強度と情報量損失の程度がどのような関係になるかを検討した。

先行研究である伊藤他(2017)では、スワッピングが適用された場合、対象となるレコードは他の地域の類似するレコードと入れ替わることから、匿名データの利用者（侵入者、intruder）が個人を特定するのは困難であり、その結果として秘匿性が高まるという仮定に基づいて、実証研究を進めてきた。本研究では、このような先行研究で用いられたスワッピングの仮定を適用するだけでなく、侵入者のシナリオを変更し、特定の個人の情報を持っている侵入者が、その個人の所在地だけでなく、入れ替え先として想定される複数の地域を対象に、当該個人を匿名化マイクロデータのレコードの中から探索するというシナリオについても考察した。このシナリオでは、侵入者が特定の個人を対象とする地域内から「発見できる」割合を秘匿性の指標とした。

また、①ターゲットスワッピングを適用して作成した匿名化マイクロデータ、②ターゲットスワッピングに加えて年齢・国籍にPRAMを施して作成した匿名化マイクロデータを対象に、スワッピング率を変化させた場合の秘匿性と情報量損失の関係について比較・検証を行い、PRAMによって一定程度情報量損失が発生するが、秘匿性の強度に関しては、ターゲットスワッピング2%~3%分程度の効果があることが分かったことを報告する。

参考文献

伊藤伸介・星野なおみ・阿久津文香・菊池亮(2017)「国勢調査の匿名化マイクロデータの作成方法に関する新たな取り組み」『製表技術参考資料』No. 37, 1~27頁