

公的マイクロデータのプライバシー保護のための秘匿審査

統計数理研究所 南 和宏

1. はじめに

近年、我が国は公的調査票情報の二次的利用を推進し、令和元年5月には改正統計法が施行され、オンサイト利用における調査票情報の学術研究等の利用が可能となり、提供対象の拡大が図られた。この制度により、学術研究を目的とする研究者はオンサイト施設の端末から調査票情報に対する探索的分析を行なうことが可能となる。しかし調査票情報には調査客体の機密情報を含まれるため、利用者が持ち出す分析結果に対して安全性審査を実施し、機密情報の漏洩を防止する必要がある。

2. 持ち出しデータの安全性基準

オンサイト利用における安全性審査は、長年の運用実績のある欧州連合統計局（Eurostat）が編纂した統計開示抑制技術のハンドブック [2] を参照して審査基準を策定している。審査基準は学術論文等に発表する統計データから元の調査票情報の復元を防止するための5つの原則に基づき、それらの原則を各統計量に適用することで具体的な安全性ルールが導き出される。しかし、オンサイト利用では、利用者と審査者の円滑な連携が必要であり、審査者が持ち出しデータの安全性を検証するために必要な説明資料に関する明示的な要件を Eurostat 基準に追加している [3]。

3. 表データの秘匿処理ツールの開発

集計表、度数表等の表データは、データ分析における基本となるデータ形式であるが、行計、列計に関するデータ間の関係性を内包するため、手作業での秘匿処理は困難である。また集計表における占有性ルールの検証には、元の調査票情報の参照が必要となり、利用者が説明資料を準備する負担も大きい。したがって、オンサイト利用での利用を想定し、安全性審査に必要な説明資料の自動生成を行う秘匿処理ツールの開発を R 言語で進めている [4]。

参考文献

1. ミクロデータ利用ポータルサイト 調査票情報の利用. <https://www.e-stat.go.jp/microdata/data-use>
2. Anco Hundepool et al. Handbook on Statistical Disclosure Control Version 1.2.
https://ec.europa.eu/eurostat/cros/content/handbook-sdc_en
3. 南 和宏, 菊池亮. オンサイト利用における持ち出し安全性基準及び審査方法. 統計研究彙報 75, 2019 年 3 月.
4. Kazuhiro Minami and Yutaka Abe. Statistical Disclosure Control for Tabular Data in R. Romanian Statistical Review, No. 4, pp. 67-76, 2017.