

## 前回の議論を踏まえた論点整理（擬似マイクロデータ）

平成 25 年 3 月 26 日

総務省政策統括官室（統計基準担当）

前回（平成 24 年 12 月 19 日）の研究会での主な意見と今後の方向性について、以下のとおり整理を行った。

（第三者への再配布）

- 擬似マイクロデータは本物のデータではないため、再配布しても構わないという考え方もあるが、第三者から見た場合に調査票情報が漏えいしたと誤解を生じさせるおそれがあることや、試行提供段階であり擬似マイクロデータの使われ方を把握しておきたいことから、現時点では再配布は認めない取扱い。（（独）統計センター統計技術研究課 坂下課長）
- ドイツの例と対比させると、日本の擬似マイクロデータも PUF（パブリックユースファイル）として利用の汎用性を高めても良いのではないか。（縣委員）
- 擬似マイクロデータは擬似的（＝リアリスティック）なものであり、リアル（＝実在する）データではないが、あたかもリアルデータであると誤解を生む点については社会に与える影響が大きい。擬似マイクロデータ提供の前提として、利用者側・提供者側の双方で利用上の注意が徹底される必要。（樁委員）
- 擬似マイクロデータがリアルデータとして誤用されることを防止する観点、利用者の声をフィードバックさせて更に改善する観点から、現在（独）統計センターで行われているような登録方式が望ましいのではないか。

（擬似マイクロデータ作成のメリット）

- 一般的に匿名化措置が困難と考えられている企業情報や年齢各歳別などの詳細データについては、特定されるリスクが小さくなっていると考えられる。（樁委員）
- 匿名データでは作成できない複数の統計データをリンケージさせたデータも作成できる可能性がある。（樁委員）
- 樁委員御指摘の可能性について、（独）統計センター等において更に研究を進める必要があるのではないか。

（擬似マイクロデータの作成目的）

- 教育目的以上に研究目的でも本格的な研究の準備段階で活用できる可能性がある。（樁委員）
- このようなデータを開発した目的は、訓練用データとしてフォーマットも含めてリアルデータと同一なものを作成すること、授業に用いる教育用データを作成することの 2 つ。教育用であれば、変数やデータ数をコンパクトにしたものでも良いのではないか。また、本当の目的として研究用テストデータやプログラム送付集計型のテストデータと

しての使用が想定され、この場合データは詳細な方が良い。このような棲み分けができれば良いが、初めから全変数の分布の再現は目指さない方が良い。(安田委員)

- 匿名データを基にした作成方法に注力することが重要。匿名データを利用する前に、匿名データとレイアウトを同様にした擬似マイクロデータを作成することは大変有意義であり、現実的ではないか。匿名データの利用前にレプリカデータで事前検証を行うことで、匿名データの正確な利用につながる。(玄田委員)
- 匿名データを基にした擬似マイクロデータを作成したことがあるが、目的は教育だった。利用方法の可能性を広く考えた場合、方向性が若干異なってくるのではないか。(総務省統計研修所 小林調査官)
- 調査票情報を基に作成すべきか、匿名データを基に作成すべきかという点も引き続きの論点。(廣松座長)
- 擬似的なデータであるため、実証研究より、マイクロデータの取扱い・分析手法の習得等の教育目的での活用に適している。一方で、プログラム送付集計型のテストデータや複数の統計データをリンケージさせたデータなど、様々な活用の可能性への期待も大きい。今後の方向性として、教育用擬似マイクロデータと研究用擬似マイクロデータとを分けて議論する必要があるのではないか。

(擬似マイクロデータの名称)

- 擬似マイクロデータという名称は混乱を招くおそれがあり、反対。リアルなデータではないということを認識してもらう必要がある。強いて言うなら「パイロット版データ」か。(玄田委員)
- 名称については、当面、今の「擬似マイクロデータ」を仮称として用いることとするが、利用目的等に応じたより適当な名称を検討する必要があるのではないか。

(事業所・企業対象の統計調査の擬似マイクロデータ)

- できれば作成する方向で研究を進めてほしい。(樫委員)
- (独)統計センターとしては、次に質的属性が多い「就業構造基本調査」の擬似マイクロデータを開発し、その次に事業所・企業系の調査に関する擬似マイクロデータの開発を目指したい、としている。

(その他：廣松座長の取りまとめ)

- 内容については、調査票情報とは切り離されたものであり、現行制度上問題ないということで、研究会として合意が取れたものとする。その上で、今後どのようなものを作成すべきか、この研究会等で意見を聴いて提言したい。
- 既に多くの利用者があるようだが、今後も潜在的な利用者を含めて利用者ニーズを把握する必要。
- (独)統計センター以外の者が作成する場合についても今後検討する必要。