様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １９ 日現在

機関番号：３２６８７
研究種目：研究活動スタート支援
研究期間：2021～2022
課題番号：２１Ｋ２０１３３

研究課題名（和文）統計データ利活用推進に資する擬似的なミクロデータの作成方法に関する研究

研究課題名（英文）Study for creating pseudo-microdata for promoting the utilization of statistical data

研究代表者
高部　勲（TAKABE, ISAO）

立正大学・データサイエンス学部・教授

研究者番号：３０９０９６１９

交付決定額（研究期間全体）：（直接経費）　　1,000,000円

研究成果の概要（和文）：我が国の法令・統計制度に沿った形で合成データ（Synthetic Data）の手法に基づく疑似的なミクロデータを作成・提供するための方法について検討し、商用データや匿名データ、実際の公的統計ミクロデータを用いて疑似的なミクロデータを作成し、その性質等について分析を行い、当該データを活用した教育・研究を実施するための方法についても検討を行った。これらの研究成果について、学術論文誌に発表するとともに、学会大会、シンポジウム、研究集会等の場で発表・報告を行った。
さらに、作成した疑似的なミクロデータを用いたワークショップを実施し、教育等への活用方法の実践を進めるとともに、今後の課題について整理した。

研究成果の学術的意義や社会的意義
擬似的なミクロデータを作成す際に諸外国で標準的に用いられている合成データ（Synthetic Data）の方法を、我が国の公的統計データに関する法令・制度に沿った形で適用し、実際の公的統計ミクロデータを基に、擬似的なミクロデータを作成・提供するための方法論を提案するとともに、その手順や課題について整理することができた。
また、ワークショップの場における、擬似的なミクロデータを用いた教育・研究の実践を行うことで、こうした擬似的なミクロデータを用いる際の方法や課題について検討し、整理することができた。

研究成果の概要（英文）：I create pseudo-micro data based on the synthetic data method in accordance with the Japanese system, and examine methods for providing these data, using various data to create pseudo-micro data, and examine its properties.
Based on these preparations, permission was obtained to use official statistical microdata. Then, I created pseudo-micro data using the official statistical micro-data and analyzed its properties. I also examined methods for implementing education using this data. The results of these studies were published in academic journals. These research results were reported at conferences, symposiums, research meetings, etc.
A workshop (sponsored by the Official Statistical Microdata Research Consortium) was held using the created pseudo microdata. Then, we considered how to utilize the data in education, and sorted out future issues.

研究分野：統計科学

キーワード：公的統計　経済統計　ミクロデータ　合成データ

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

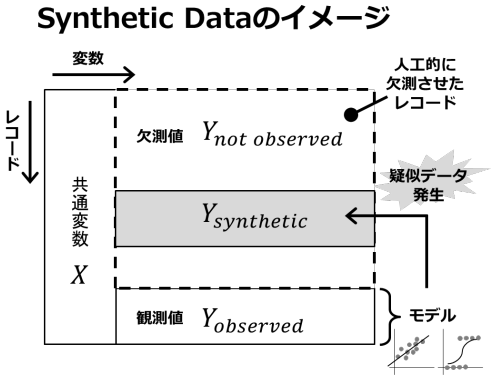
１．研究開始当初の背景

統計調査の調査票を集計する前のレコード単位のデータをミクロデータという。欧米では EBPM の観点から、ミクロデータを活用した多くの実証分析が行われてきた実績がある。我が国においても、2007 年・2018 年の統計法改正により、公的統計ミクロデータの利用要件が緩和されたものの、制度等に対する認知度が低く、利用方法のイメージがしにくいことなどから、ミクロデータの利用は進んでいなかった経緯があり、こうした中で、公的統計ミクロデータについての理解、利用イメージの把握に資する教育用・プログラムテスト用の疑似的なミクロデータの必要性が指摘されていた（坂田編著(2019)「公的統計情報：その利活用と展望」）。

ところで、我が国の統計法・統計制度においては、公的統計ミクロデータの学術的な利用目的は統計の作成に限られており、擬似的なデータであったとしても、レコード単位のデータを最終成果物として作成することは、秘匿性の観点から、制度上、認められていない。

上記のような教育用・プログラムテスト用の擬似的なミクロデータに関し、諸外国では、合成データ（Synthetic Data）と呼ばれるモデルベースの疑似的なデータ作成方法に関する研究が進んでいる（高部・徳富(2020)）。これは一部のレコード・変数を人工的に欠測させ、事前に構築した重回帰モデルやロジットモデルを用いて疑似データを発生させる方法であり（Nowok, Raab and Dibben (2016)、Templ et al.(2017)）、変数間の関係を保持したデータの作成が可能となる。ただし、この方法では、ミクロデータから直接的にレコード単位のデータを作成・提供することになるため、前述のとおり現行の法令・制度上、認められない方法である。そこで制度に合わせた形で、合成データの手法により、擬似的なミクロデータを作成するための工夫が必要となる。

さらに、前職（総務省統計局）においてミクロデータの利活用推進業務に携わった経験や、ミクロデータを活用した実証分析の成果発表の機会、行政データのオープン化に熱心な地方自治体との交流を通じて、ミクロデータに関心を持つ研究者がいる一方で、データの内容や利活用のイメージがつかみにくく、具体的な利用につながりにくい状況があることや、教育・訓練・テスト用データへのニーズが高いこと、それらの機関のデータ利活用推進の取組に本研究の方法論が応用可能なことを認識していた。



このような実践と研究の経験から、ミクロデータに対する理解・関心を深め、その利用推進を図るためには、手続き上の負担なく自由に利用できて、実データに近い分析結果が得られる疑似的なミクロデータの方法論の研究が重要なテーマであると認識し、我が国の法令・制度を踏まえた上での、合成データの手法を活用した疑似的なミクロデータの作成・提供の方法論の研究・検討を行った。

２．研究の目的

前述の課題を踏まえつつ、事前に集計した統計表と推定したモデル（重回帰モデル、ロジットモデル等）の結果を基に、諸外国で研究が行われている合成データ（Synthetic Data）に関するモデルベースの手法を活用し、現行の法令・制度上の制約を満たしつつ元のミクロデータの構造を可能な限り保持した疑似的なミクロデータを作成し、提供を行うための方法を確立することを目的として、研究を行った。

また、擬似的なミクロデータの作成方法の研究だけではなく、合わせて、作成した擬似的なミクロデータを安全に提供するための方法や、それらのデータを用いた教育の方法論を確立することも目的として、研究を行った。

３．研究の方法

上記の目的を達成するために、以下の２点の方法により、研究を進めた。

(1) 疑似的なミクロデータの作成に関する課題、手法の整理・検討：

・疑似的データの作成に関する最新の情報収集、手法の整理、ミクロデータの申請手続きを行うとともに、現行制度を踏まえた疑似的データ作成の課題、方法論について整理した
・整理した課題や方法論を踏まえ、実際の公的統計データを基に、公的統計の匿名データや商用データ、公的統計ミクロデータなどの様々な種類のレコード単位のデータを用いた試算や実証分析や検討を行い、その作成過程や方法論についてまとめるとともに、元のデータとの比較検証、有効性の確認、更なる利用可能性の検討等を行った

(2) 作成した擬似的なミクロデータの利用・提供方法に関する検討：
・それらを実際のワークショップの場において教育に用いて、その結果得られた、擬似的なミクロデータの利用・提供方法に関する課題を整理した
・研究で得られた知見を基に、実際の公的統計ミクロデータを用いて擬似的な公的統計ミクロデータを作成するとともに、作成した擬似的なミクロデータを用いて、学生や教育・研究に携わる方々を対象としたワークショップを行った
・その過程で、このような疑似データを作成し提供する上での課題や、それらのデータを用いて教育等を行う上で検討すべき課題などについて整理し、検討を行った

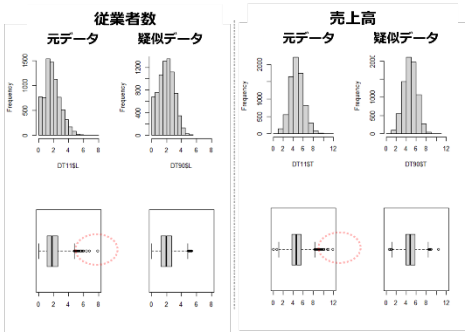
４．研究成果

令和３年度成果

令和３年度は、事前の準備段階として、擬似的なミクロデータの作成に関する最新の情報収集、手法の整理、公的統計ミクロデータのオンサイト利用（総務省統計局オンサイト施設）に関する利用申請を行うとともに、現行制度を踏まえた疑似的データ作成の方法論の研究を進めた。

これと並行して、既に利用が可能となっている公的統計の匿名データ（社会生活基本調査）及び商用データ（帝国データバンクデータ）を用いて、合成データの方法に基づく疑似的なミクロデータを試作し、元のデータとの比較検証を行うとともに、作成方法に関する検討を行った。

擬似的なミクロデータの作成結果



研究の結果、中間的なクロス集計表を適切な粒度で作成し、合わせて回帰モデルの結果を事前に公表することにより、我が国の統計法・統計制度に対応した形で、合成データの手法に基づき、元のデータの構造をある程度保持した擬似的なミクロデータを作成することが可能であるとの結論を得た。

このようにして得られた成果について、以下のような学会・研究集会において報告を行った。

・「2021 年度統計関連学会連合大会」における企画セッション「公的統計ミクロデータにおけるさらなる利活用をめぐって」（2021 年９月）
・「経済統計学会 2021 年度（第 65 回）全国研究大会」（2021 年９月）
・「第 12 回横幹連合コンファレンス」（2021 年 12 月）
・「大規模データの公開におけるプライバシー保護の理論と応用」研究集会（2021 年 12 月）

また、公的統計に関する官学の共同研究の場である以下の研究集会等において報告を行い、学術研究者のみならず、政府統計・公的統計の作成・提供の実務を担っている行政官・有識者とともに本研究内容について議論し、そこでの質疑等を経て、その後の擬似的なミクロデータの作成方法等の改善につながる示唆を得た。

・「公的統計ミクロデータ研究コンソーシアムシンポジウム 2021」（公的統計ミクロデータ研究コンソーシアム主催）（2021 年 11 月）
・「2021 年度官民オープンデータ利活用の動向及び人材育成の取組に関する研究集会」（独立行政法人統計センター主催）

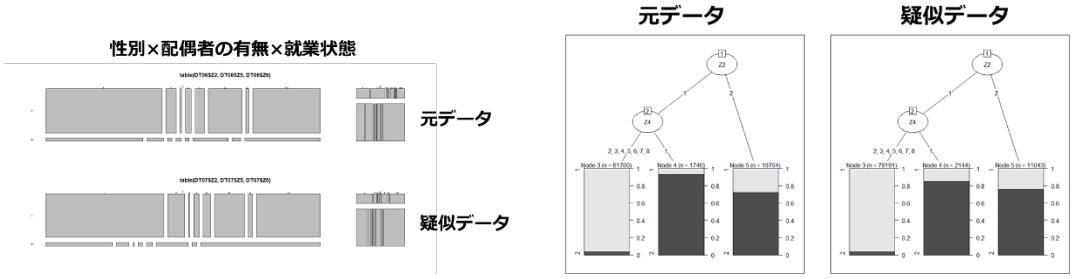
そして、上記のような研究を通じて得られた成果について、学術論文としてまとめ、「統計研究彙報」、「データサイエンス研究」などの学術誌において発表を行った。

令和４年度成果

令和４年度は、前年度においてまとめた課題や取組、方法論を踏まえ、実際の公的統計ミクロデータ（平成 26 年全国消費実態調査）の二次的利用を行い（総務省オンサイト施設）、合成データの手法を用いて擬似的なミクロデータを作成し、元のデータとの比較検証、有効性の確認、更なる利用可能性の検討等を行った。

実際の公的統計ミクロデータ（全国消費実態調査）を用いた分析・検討の結果，元データの構造をある程度保存したデータの作成は可能であり，その分布等は，元データと疑似データで，大

きくは異ならないことが示された。また、元のデータと疑似的なミクロデータを対象に、重回帰分析や決定木分析、主成分分析などの各種の多変量解析などを行い、それらの分析結果が元のデータと疑似的なミクロデータとで、それほど大きく異ならないことを確認し、元のデータの秘匿性を配慮した上で、現行の法令・制度に沿った形で、教育やプログラムテスト用のデータとして有用なデータを作成することが可能であるとの結論を得た。



このようにして得られた成果について、以下のような学会・研究集会において報告を行った。

- 「2022 年度統計関連学会連合大会」における企画セッション「公的統計ミクロデータ利活用の現状と課題」（2022 年 9 月）
- 「経済統計学会 2022 年度（第 66 回）全国研究大会」（2022 年 9 月）
- 「公的統計ミクロデータ研究コンソーシアムシンポジウム 2022」（公的統計ミクロデータ研究コンソーシアム主催）（2022 年 11 月）

また、上記の研究の成果として作成した公的統計ミクロデータに基づく擬似的なミクロデータを用いて、実際の教育等の場で利用が可能かということについて、実践を試みた。具体的には、ミクロデータに関心のある学生や、公的統計データに関連する教育・研究に携わる者を対象として、「公的統計ミクロデータのためのチュートリアル・講習会」（2023 年 1 月及び 3 月、於：統計数理研究所）「公的統計ミクロデータのためのチュートリアル・講習会」（対面演習形式）を実施し、公的統計ミクロデータを基に作成した擬似的なデータを用いてオンサイト施設に類似するデータ環境を用意し、公的統計ミクロデータの利用方法や主な分析方法などについて講習を行った（公的統計ミクロデータ研究コンソーシアム NEWSLETTER vol.03 (2023)）。その結果、元のデータの安全性に配慮した上で作成した擬似的なミクロデータを、教育研究用に利用することが可能であり、効果的であるとの結論を得た。

これらの成果について、「ESTRELA」などの雑誌に成果を発表した。

＜引用文献＞

[1] 高部勲，徳冨智哉（2020）「公的統計ミクロデータ等に基づく Synthetic Data の作成及び分析の試み」、『ESTRELA』、321、19-24、統計情報研究開発センター

[2] Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke creation of synthetic data in R. J Stat Softw, 74(11), 1-26.

[3] Templ, M., Meindl, B., Kowarik, A., & Dupriez, O. (2017). Simulation of synthetic complex data: The R package simPop. Journal of Statistical Software, 79(10), 1-38.

[4] 公的統計ミクロデータ研究コンソーシアム NEWSLETTER vol.03 (2023),Report 3「公的統計ミクロデータのためのチュートリアル・講習会（対面演習形式回）開催」

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　2件）

１．著者名 高部勲	４．巻 79
２．論文標題 Synthetic Dataの考え方に基づく疑似的なミクロデータの作成方法の検討	５．発行年 2022年
３．雑誌名 統計研究彙報	６．最初と最後の頁 111-130
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 高部勲	４．巻 1
２．論文標題 公的統計ミクロデータの利活用推進に資するSynthetic Dataの作成方法の検討	５．発行年 2022年
３．雑誌名 データサイエンス研究	６．最初と最後の頁 3-18
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 高部勲	４．巻 349
２．論文標題 公的統計ミクロデータ研究コンソーシアムの活動	５．発行年 2023年
３．雑誌名 ESTRELA	６．最初と最後の頁 2-7
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計11件（うち招待講演　5件／うち国際学会　0件）

１．発表者名 高部勲
２．発表標題 政府の公的統計ミクロデータを活用した研究事例：多様なデータによる新たな価値創造
３．学会等名 九州大学マス・フォア・インダストリ研究所(IMI)コロキウム（招待講演）
４．発表年 2022年

１．発表者名
高部勲

２．発表標題
Synthetic Dataの考え方に基づく疑似的な公的統計ミクロデータ作成・提供方法の検討

３．学会等名
科研（A）「公的統計ミクロデータを活用したEBPM支援研究プラットフォームの構築」、革新的自殺研究推進プログラム委託研究「行政における統計データの利活用の推進に関する研究」合同研究集会（招待講演）

４．発表年
2022年

１．発表者名
高部勲

２．発表標題
公的統計に関する疑似ミクロデータの提供可能性

３．学会等名
第12回横幹連合コンファレンス

４．発表年
2021年

１．発表者名
高部勲

２．発表標題
Synthetic Dataの考え方に基づく疑似ミクロデータ作成の可能性

３．学会等名
公的統計ミクロデータ研究コンソーシアムシンポジウム2021（招待講演）

４．発表年
2021年

１．発表者名
高部勲

２．発表標題
公的統計ミクロデータの利活用推進に資する疑似データ活用の可能性

３．学会等名
2021年度統計関連学会連合大会

４．発表年
2021年

１．発表者名
高部勲

２．発表標題
公的統計ミクロデータの利活用推進に資する疑似データ活用の可能性：現行制度に即した疑似データの作成・提供方法の検討

３．学会等名
経済統計学会2021年(第65回)全国研究大会

４．発表年
2021年

１．発表者名
高部勲

２．発表標題
公的統計ミクロデータに基づく疑似的なミクロデータの作成・提供方法に関する研究

３．学会等名
経済統計学会2022年(第66回)全国研究大会

４．発表年
2022年

１．発表者名
高部勲

２．発表標題
秘匿性・安全性を考慮した統計的マッチングの手法による複数データの結合

３．学会等名
2022年度統計関連学会連合大会

４．発表年
2022年

１．発表者名
高部勲

２．発表標題
公的統計ミクロデータに基づく疑似的なミクロデータの作成・提供・利活用方法の検討

３．学会等名
公的統計ミクロデータ研究コンソーシアムシンポジウム2022（招待講演）

４．発表年
2022年

１．発表者名 高部勲
２．発表標題 公的統計ミクロデータに基づく疑似的なデータの作成及び活用方法について
３．学会等名 研究集会「大規模データの公開におけるプライバシー保護の理論と応用」（招待講演）
４．発表年 2022年

１．発表者名 高部勲
２．発表標題 公的統計ミクロデータの更なる利活用に向けた取組について
３．学会等名 科研（A）「公的統計ミクロデータを活用したEBPM支援研究プラットフォームの構築」、革新的自殺研究推進プログラム委託研究合同研究集会
４．発表年 2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関