

## 「匿名データを活用した社会課題の解決」のための研究事例の紹介

-令和5年度総務省統計研究研修所の新たな取組み-

長尾 伸一\*

Introducing study cases for solving social issues using anonymized microdata  
-A new challenge carried out by the Statistical Research and Training Institute in 2023 -

NAGAO Shinichi

総務省統計研究研修所では、国・地方公共団体等の職員を対象に、統計作成の中核を担う統計人材の育成及び統計を政策の立案等に活用するための統計リテラシー向上を目的として、従前より統計研修を実施している。近年、公的統計をはじめとした様々なデータを活用したデータ駆動型の意思決定に資する実践的な研修と社会課題の解決に向けた研究を進めることに関心が高まっていることから、令和5年度の本研修において、公的統計の「匿名データ」を活用した研修を初めて実施したので、その概要を紹介する。

キーワード：匿名データ、統計研修、公的統計

The Statistical Research and Training Institute of the Ministry of Internal Affairs and Communications (MIC) has long conducted statistical training for employees of national and local governments, with the aim of developing human resources for the important tasks of conducting surveys and improving statistical literacy for using statistics in policy making. In recent years, there has been growing interest in practical training that contributes to data-driven decision making using various data, including official statistics, and in promoting research aimed at solving social issues. The following is an overview of the research.

Key Words : Anonymized microdata, Statistical training, Public Statistics

---

\* 総務省統計研究研修所管理・研修部

## 1. はじめに

現在、我が国においては、人口減少・少子高齢化、過疎化・東京圏への一極集中、地域産業の空洞化といった様々な課題に直面している。一方で、デジタル・トランスフォーメーション(DX)と呼ばれる、デジタル技術により社会課題を解決する方策を考えることが、国・地方を問わず様々な領域で注目されている<sup>1</sup>。

このような状況下で、総務省統計研究研修所では、国・地方公共団体等の職員を対象に、統計作成の中核を担う統計人材の育成及び統計を政策の立案等に活用するための統計リテラシー向上を目的として、従前より統計研修を実施している。近年、データサイエンスや統計の利活用に関心が高まっていることから、効率的な研修の受講のために、オンライン研修の充実などを実施しているところである。また、公的統計をはじめとした様々なデータを活用することを通じて、社会課題を解決に向けた研究に関心が高まっていることから、実践的なデータ分析を実施するために、「実データ」を利用することで、データ駆動型の意思決定に資する実践的な研修の実施を推進している。

令和5年度の本科研修において、公的統計の「匿名データ」<sup>2</sup>を活用した研修を初めて実施したので、その概要について、紹介する。

## 2. 匿名データの利用について

匿名データとは、行政機関等が行う統計調査によって集められた調査票情報を、特定の個人又は法人その他の団体の識別(他の情報との照合による識別を含む。)ができないように加工したものをいい、平成21年4月から全面施行された統計法第36条により、統計調査を実施する行政機関等が作成した匿名データを、学術研究及び高等教育の発展に資すると認める場合に提供できることになった。独立行政法人統計センターに申請することにより利用することが可能となる<sup>3</sup>。ただし、利用者は、データ漏えいの防止など、匿名データを適正に管理する必要があるなどの条件がある。

前述のとおり、匿名データの利用については、その性格上制約があるものの、データ分析を実施する上で、以下のように、便益があると考えている。

1) 公表データにはないクロス集計表を作成できる

2) 公表される前の大きなデータセットを使える

サンプルサイズの大きなデータの活用が可能となり、分析結果の信頼性が高まる

3) 回帰分析のような統計モデルとの相性がいい

公的統計が提供する公表データは、標本誤差等を考慮して一定の信頼区間に収まるクロス集計表を定義し、公表している。統計利用者にとっては、調査項目を分析のための変数として利用する際に、予め準備された変数の平均値の利用に限られ、定義された変数以外の調査項目を利用できない。言い換えると、公表データを利用する前提で、統計モデルを考えたい際、調査されている全ての項目を変数として分析に加えられるわけではない。匿名データを利用することでほとんどの調査項目を用いて、回帰分析のような統計モデルを作成することが可能となる。

4) 公表前のデータの分布が確認できる

公表前のデータの分布が容易に確認できることから、最適な統計モデルの選択に有効であること、平均からの乖離の大きな外れ値の影響を考慮した分析が可能になることにより、分析結

<sup>1</sup> デジタル田園都市国家構想、内閣官房

<sup>2</sup> 匿名データの利用・提供及びオーダーメイド集計、総務省

<sup>3</sup> 匿名データの利用方法、独立行政法人統計センター

果の信頼性が高まることが期待できる。

以上のことから、データ分析に関わる研修や教育で匿名データを活用することは、社会課題の解決を目的とした研究に有力なツールとなり得ると考えている。

### 3. 本研修で活用した匿名データの種類

令和5年度の本研修で準備した匿名データは、以下のとおりである。

- ①全国消費実態調査(平成21年及び平成26年)
- ②就業構造基本調査(平成24年及び平成29年)
- ③社会生活基本調査(平成23年及び平成28年)

また、令和5年度の本研修受講生の中で、匿名データを利用して研究レポートを作成した者は10名であった。その受講生が各々活用したデータは下表のとおりである。

表 受講生が利用した匿名データの種類と研究テーマ

| 受講者氏名  | 研究テーマ   | 全国消費実態調査 | 就業構造基本調査 | 社会生活基本調査 |
|--------|---|----------|----------|----------|
| 那須 英和  | 少子化の要因分析とその対策   | H21,H26  | H24,H29  | H23,H28  |
| 櫻井 直也  | 全国消費実態調査の匿名データを用いた高齢者世帯の消費と貯蓄分析                               | H21,H26  |          |          |
| 土橋 卓也  | 自衛隊への就業動向のマクロ分析と採用施策への提言について～防衛力の維持のために～                      |          | H24,H29  |          |
| 中山 拓人  | 非正規雇用から正規雇用への移行要因分析   |          | H24,H29  |          |
| 石井 秀人  | 若年隊員不足解消に向けた若者の離職に関する要因分析<br>～若者の働きやすい自衛隊を目指して～               |          | H29      |          |
| 大西 満斉紀 | 海外情勢不安に起因する生産拠点の国内回帰を背景とする企業の立地意思決定に関する要因分析～成長業種を県内経済に取り込めるか～ |          | H29      |          |
| 川田 佳   | 男女の賃金差の要因分析   |          | H29      |          |
| 林 玲奈   | 働く貧困層（ワーキングプア）の要因分析<br>～労働者の貧困はなぜなくなるのか～                      |          | H29      |          |
| 三村 勇貴  | 働く場所に見る育児時間の決定要因分析  |          | H29      | H28      |
| 吉崎 雅基  | 男性単独世帯における家事時間の規定要因の分析  |          |          | H28      |

総務省が実施する公的統計のうち、3つの統計調査の匿名データを準備したが、全ての調査について利用があった。最も利用者が多かったのが、就業構造基本調査であった。テーマを見ると、自衛隊員の充足に関するテーマが2件あったほか、賃金及び賃金の格差に関するテーマ、労働移動に関するテーマ、企業の立地に関するテーマなど雇用環境をめぐる様々な課題に関心がある者がいたことが窺える。

### 4. 匿名データを活用した受講生の研究レポートの概要

今回、匿名データを利用した受講生の研究レポートについて、以下にそれを紹介する。

(1) 那須英和の研究概要

|                    |  |
|--------------------|--|
| 研究テーマ              | 少子化の要因分析とその対策  |
| 研究の目的              | 本研究は、様々な問題が引き起こされる少子化問題について、男性・女性別にその要因分析を行うことを目的とする。  |
| 分析の方法              | 男性・女性別のほか年収・教育等の個人属性別に婚姻率を計算し、機械学習による予測モデル等を用いて要因を検証した。<br>分析に用いたモデルは、以下のとおり。<br>・クロス集計表の作成(データの可視化)<br>・ベイジアンネットワークによる分類<br>・決定木による分類<br>・アンサンブル学習による予測モデル  |
| 分析結果の概要            | 1) 男性について、分析手法別に考察する。記述統計の手法から年収が高いほど婚姻(既婚)率が高いという結果が得られた。機械学習による分類からは、婚姻率に影響を与える要因として、年収、従業者規模が挙げられ、さらにその背景として、正規・非正規の別、三大都市圏か否か、産業や教育等が影響しているという結果が得られた。<br>2) 女性については、男性のように年収が高いほど婚姻(既婚)率が高いという明確な結果は得られなかった。これは結婚・出産の前後での就業状態の変化の影響を受けている可能性があることが示唆される。また、非正規雇用を除いた場合は、男性と同様に年収が高いと婚姻率が高いという可能性もあるという結論が得られ、一般的に言われるような年収が高いほど結婚しないという仮説は認められない。女性の婚姻率には様々な要因が影響されていると推論できる。<br>以上のことから、男性については、先行研究で言われるように就業状態が結婚に影響を与えているという結論が得られる。一方、女性の婚姻については、就業との関連も見られるものの、それ以外の要因も存在すると考えられる。  |
| 参考文献及びデータ(匿名データ以外) | [1] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 785-794.<br>[2] 内閣府. (2022). 少子化社会対策白書(令和4年版)(大型本, p. 245). 日経印刷<br><a href="https://lead.to/amazon/jp/?op=bt&amp;la=ja&amp;key=4865793291">https://lead.to/amazon/jp/?op=bt&amp;la=ja&amp;key=4865793291</a><br>[3] 和田光平. (2015). 人口統計学の理論と推計への応用(Kindle版). オーム社.<br><a href="https://lead.to/amazon/jp/?op=bt&amp;la=ja&amp;key=B07D7RSZ4B">https://lead.to/amazon/jp/?op=bt&amp;la=ja&amp;key=B07D7RSZ4B</a><br>[4] 坂爪聡子. (2023). 児童手当が結婚と出生に与える影響. 現代社会研究科論集: 京都女子大学大学院現代社会研究科紀要, 017, 41-51. <a href="http://hdl.handle.net/11173/3638">http://hdl.handle.net/11173/3638</a> |

|  |  |
|--|--|
|  | <p>[5] 形, &amp; 智邦. (2014). 雇用形態が男性の結婚に与える影響. 人口学研究, 50(0), 75-89.<br/> <a href="https://doi.org/10.24454/jps.50.0_75">https://doi.org/10.24454/jps.50.0_75</a></p> <p>[6] 情報通信白書_平成30年版(3). (n.d.). Accessed on 11/1/2023. <a href="https://cir.nii.ac.jp/crid/1690013498970166144">https://cir.nii.ac.jp/crid/1690013498970166144</a></p> <p>[7] 日本の将来推計人口(平成29年推計). (n.d.).<br/> <a href="https://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/db_zenkoku2017/s_tables/1-5a.htm">https://www.ipss.go.jp/pp-zenkoku/j/zenkoku2017/db_zenkoku2017/s_tables/1-5a.htm</a>;<br/> Accessed on 11/1/2023.</p> <p>[8] 田辺和俊., &amp; 鈴木孝弘. (2020). 都道府県別の女性未婚率の要因分析 : 自治体の少子化対策の観点から. 厚生指標 = Journal of Health and Welfare Statistics, 67(13), 15-21.<br/> <a href="https://cir.nii.ac.jp/crid/1521136281007242112">https://cir.nii.ac.jp/crid/1521136281007242112</a></p> |
|--|--|

## (2) 櫻井直也の研究概要

|         |   |
|---------|---|
| 研究テーマ   | 全国消費実態調査の匿名データを用いた高齢者世帯の消費と貯蓄分析   |
| 研究の目的   | 本研究は、高齢者世帯の消費支出、所得と介護サービス利用、ライフサイクル投資仮説の妥当性について分析を行い、新たな視点を提供することを目的とする。  |
| 分析の方法   | <ol style="list-style-type: none"> <li>1) 高齢者世帯のうち、2人以上世帯の消費支出額を目的変数とした線形回帰分析</li> <li>2) 高齢者世帯のうち、単身世帯及び2人以上世帯の介護サービス支出額を目的変数とした線形回帰分析</li> <li>3) 有価証券の売買取引の有無を目的変数としたプロビット回帰分析</li> </ol>  |
| 分析結果の概要 | <ol style="list-style-type: none"> <li>1) 高齢者世帯のうち、就業をしている世帯では、公的年金給付だけで生活費を賄うことが難しい状況であるという結論が得られた。就業している世帯のうち、年金だけで生活費を賄っておらず、65歳を越えても働いている世帯の存在があると推察できる。ただし、収入があるために年金の受給を繰下げている可能性もあると考えられる。<br/> 一方、就業していない世帯では就業している世帯に比べて公的年金給付だけで生活を賄っている世帯の割合は高いが、年金以外の収入が少なく、生活費を賄っていない世帯もあることが明らかとなった。</li> <li>2) 高齢者世帯の介護サービス支出については、単身世帯と2人以上世帯では関連性の違いが見られた。2人以上世帯では、収入が多いほど、また、貯蓄現在高が多いほど、介護サービス支出額も多くなるという関係があることが認められた。一方で、単身世帯では、介護サービス支出額と収入・貯蓄現在高のいずれの変数でも相関は見られなかった。</li> <li>3) ライフサイクル投資仮説の妥当性をデータから分析した結果、2009年のデータでは40歳代の世帯の持ち家ダミーの係数が負で有</li> </ol> |

|                        |   |
|------------------------|---|
|                        | <p>意と、有価証券への投資が住宅投資に抑制された可能性が示唆され、50歳代以降の世帯については、経常所得と金融資産の係数が正と推計されていることから、有価証券の売買に積極的になっている可能性があることを示した。この結果からは、世帯の有価証券の売買行動において、2009年のデータについてはライフサイクル投資仮説が妥当している可能性があることを示している。2014年のデータでは、ライフサイクル投資仮説が確認できなかったことから更なる分析が必要だと考えている。</p>  |
| 参考文献及びデータ<br>(匿名データ以外) | <p>[1] 和泉徹彦、齋藤香里、白石憲一、卓涓涓(2016)「全国消費実態調査を用いた消費・貯蓄、貧困、介護サービスの分析」生活経済学研究、43巻 p.65-80</p> <p>[2] 和泉徹彦(2017)「全国消費実態調査に基づく高齢者世帯消費支出の分析」嘉悦大学研究論集 59(2), p.55-67</p> <p>[3] 谷頭子、草薙 仁 (「高齢者世帯の所得格差と食料消費行動」農業経済研究、89巻、p.291-294)</p> <p>[4] S. Bakshi、Zhiwu Chen(1994)「Baby Boom, Population Aging, and Capital Markets」The Journal of Business、Vol.67、No.2 (Apr.,1994) p.165-202</p> <p>[5] 熊本方雄(2015)「人口構造の変化が株価に与える影響」金融調査研究会報告書 Volume 54、p.69-83、2015</p> <p>[6] 【R】ROC 曲線を書く 2020年7月7日 Qiita:<br/>&lt;<a href="https://qiita.com/insilicomab/items/2b4c380fb19f6f656e32">https://qiita.com/insilicomab/items/2b4c380fb19f6f656e32</a>&gt;</p> <p>[7] 統計用語集 統計 WEB:<br/>&lt;<a href="https://bellcurve.jp/statistics/glossary/2167.html">https://bellcurve.jp/statistics/glossary/2167.html</a>&gt;</p> |

### (3) 中山拓人の研究概要

|         |   |
|---------|---|
| 研究テーマ   | 非正規雇用から正規雇用への移行要因分析   |
| 研究の目的   | 本研究は、非正規雇用から正規化する移行要因を明らかにすることで、不本意非正規労働者を減少させることを目的とする。  |
| 分析の方法   | データから、転職者のうち、分析対象を、前職非正規で、かつ離職時期1年以内の者を抽出し、目的変数を現職が正規に就いた者と非正規に就いた者とするプロビット回帰分析の手法を用いて検証した。また、訓練・自己啓発の効果を推測するために傾向スコアマッチングを適用したモデルについても推計した。  |
| 分析結果の概要 | <p>転職によって正規雇用化する者について、訓練・自己啓発の効果を明らかにすることを目的として分析した結果を、訓練・自己啓発が「勤め先が実施したもの」と「自発的に行ったもの」に分けて分析した。また、雇用形態別の訓練・自己啓発の効果を明らかにすることも併せて検証した。</p> <p>1) 前職非正規の者のうち、「訓練・自己啓発の有無」は、現職正規雇用か否かに有意な差が認められた。</p> <p>2) 訓練・自己啓発が、「勤め先が実施したもの)」についても、有意</p> |

|                        |  |
|------------------------|--|
|                        | <p>な差が認められた。</p> <p>3) 訓練・自己啓発が、「自発的に行ったもの」についても、有意な差が認められた。</p> <p>4) 雇用形態別の「訓練・自己啓発の有無」については、パート、アルバイト、契約社員、嘱託、その他では5%有意水準で有意な結果となったが、派遣社員については有意な結果が得られなかった。</p> <p>以上のことから、訓練・自己啓発の効果は、概ね全ての雇用形態においても転職による正規化に寄与すると考えられる。</p>  |
| 参考文献及びデータ<br>(匿名データ以外) | <p>[1] 玄田有史(2008)「前職が非正規社員だった離職者の正社員への移行について」日本労働研究雑誌 No. 580/November 2008</p> <p>[2] Hiromi Hara(2022) “The effect of public-sponsored job training in Japan” Journal of The Japanese and International Economies 64(2022) 101187</p> <p>[3] 樋口美雄、佐藤一磨、石井加代子(2011) 「非正規雇用から正規雇用への転換に能力開発支援は有効か」 KEIO/KYOTO GLOBAL COE DISCUSSION PAPER SERIES DP2011-043</p> <p>[4] 高橋勇介(2021) 「パネルデータから見た非正規雇用の現状と正規雇用への転換」経済ジャーナル第17巻第2号</p> |

## (4) 大西満斉紀の研究概要

|         |  |
|---------|--|
| 研究テーマ   | 海外情勢不安に起因する生産拠点の国内回帰を背景とする企業の立地意思決定に関する要因分析<br>～成長業種を県内経済に取り込めるか～  |
| 研究の目的   | 本研究は、企業が生産拠点についてどのような立地条件を重視しているのかを把握することを目的とする。   |
| 分析の方法   | <p>1) 工場立地動向調査の統計表「立地地点選定理由別選択件数(新設のみ)(業種別)」を8年分集計し、産業別、理由別のクロス集計表を作成し、特化係数を計算する。</p> <p>2) 2016年の経済センサス-活動調査と、国土交通省提供の地理国土数値情報を組み合わせ、岡山県の事業所の分布と物流に関連する地物との位置関係を視覚的に分析する。加えて、就業構造基本調査の匿名データから企業に勤める雇用者からみた産業別転居者の把握のために新たなクロス集計表を作成し、その特徴を分析した。</p>   |
| 分析結果の概要 | <p>1) 工場立地動向調査から理由別に件数を合計し、パレート図にしたものを分析すると、1位から順に「本社・他の自社工場への近接性」、「市場への近接性」、「関連企業への近接性」、「地価」、「その他」、「工業団地である」、「人材・労働力の確保」、「国・地方自治体の助成」であった。一方、物流に関する項目は予想外に低かった。</p> <p>また、岡山県について、同様の分析をすると、工業団地と助成についての項目が高く、人材・労働力についての項目は低い。</p> <p>2) 地域メッシュ統計から、製造業の事業所と交通機関との接合部および物流拠点の距離について分析すると、県南部においては接続部</p> |

|                                |  |
|--------------------------------|--|
|                                | <p>および物流拠点の比較的近い地点に製造業集中メッシュが多く位置しているように見える。一方で、接続部および物流拠点がなくとも集中メッシュが存在した。県北部においては、接続部および物流拠点の比較的近い地点に集中メッシュが多く位置するという傾向がみられた。この結果をみると、物流という条件は立地に関係があると考えられる。</p> <p>また、就業構造基本調査のデータから住居を転居した理由と産業別特徴をみると、製造業の誘致による雇用創出効果をもたらす社会増は、他産業のよりも高い可能性が示唆される。</p>   |
| <p>参考文献及びデータ<br/>(匿名データ以外)</p> | <p>[1] 第一生命経済研究所経済調査部 研究理事 佐久間啓：「Market Side Mirror 製造業の国内回帰に大きな期待、と不安～国内回帰というより体内直接投資の拡大という視点が重要～」, 2022年10月31日金融市場レポート</p> <p>[2] 帝国データバンク情報統括部(2023年1月27日)：「特別企画：国内回帰・国産回帰に関する企業の動向調査」</p> <p>[3] 工場立地動向調査、「立地地点選定理由別選択件数 【新設のみ】 (業種別)」, 「立地地点選定理由別選択件数 【新設のみ】 (都道府県別)」, 2022～2015年, 経済産業省 地域経済産業グループ 地域産業基盤整備課, <a href="https://www.e-stat.go.jp/stat-search/files?page=1&amp;layout=datalist&amp;toukei=00550510&amp;tstat=000001024255&amp;cycle=7&amp;tclass1=000001078716&amp;tclass2val=0">https://www.e-stat.go.jp/stat-search/files?page=1&amp;layout=datalist&amp;toukei=00550510&amp;tstat=000001024255&amp;cycle=7&amp;tclass1=000001078716&amp;tclass2val=0</a></p> <p>[4] 工場立地動向調査の調査概要, <a href="https://www.meti.go.jp/statistics/tii/ritti/gaiyo.html">https://www.meti.go.jp/statistics/tii/ritti/gaiyo.html</a></p> <p>[5] 工場立地動向調査の調査計画, <a href="https://www.e-stat.go.jp/surveyplan/p00550510001">https://www.e-stat.go.jp/surveyplan/p00550510001</a></p> <p>[6] 平成28年経済センサス-活動調査-世界測地系(500mメッシュ), 「産業(大分類)別事業所数及び従業者数」, 2016年, 総務省統計局 統計調査部 経済統計課経済センサス室, <a href="https://www.e-stat.go.jp/gis/statmap-search?page=1&amp;type=1&amp;toukeiCode=00200553&amp;toukeiYear=2016&amp;aggregateUnit=H&amp;serveyId=H002005112016&amp;statsId=T000918">https://www.e-stat.go.jp/gis/statmap-search?page=1&amp;type=1&amp;toukeiCode=00200553&amp;toukeiYear=2016&amp;aggregateUnit=H&amp;serveyId=H002005112016&amp;statsId=T000918</a></p> <p>[7] 国土数値情報、「高速道路時系列データ」(2022年), 「物流拠点」(2013年), 「工業用地データ」(2009年), 国土交通省 他</p> |



## (5) 川田佳の研究概要

|                        |   |
|------------------------|---|
| 研究テーマ                  | 男女の賃金差の要因分析   |
| 研究の目的                  | 本研究は、男女の賃金格差の要因を把握することを目的とする。   |
| 分析の方法                  | <p>データから得られる年間所得について、対数変換し、変数の正規性を確認した上で、線形回帰モデルの目的変数とした。また、変数増減法により説明変数を適切に選択し、以下の統計モデルから女性の賃金を決定する要因を分析した。</p> <ol style="list-style-type: none"> <li>1) 相関分析</li> <li>2) 決定木分析</li> <li>3) 重回帰分析</li> <li>4) 残差分析</li> </ol> <p>モデルの推計にあたり、事前にヒストグラムによる外れ値の確認と除外の検討等データクレンジングを実施し、事後に残差分析、交互作用分析の手法を利用することでモデルの妥当性を検証した。</p>   |
| 分析結果の概要                | <p>先行研究から、男女の賃金格差は、職業別の労働者数の偏り、具体的には、管理的職業に就く女性が少ない等の要因から生じているという仮説のもとで分析を実施したが、年間所得の決定においては、職業よりも「性別」、「非正規雇用」、「継続就業期間」、「高卒か否か」及び「従業者規模」の5つの要因があるとの結論が得られた。また、交互作用分析の結果からは、「女性×高卒×非正規雇用」の3つの条件が揃うと交互作用に年間所得に対する有意な影響があることが確認できた。このことから、女性に限定された影響ではなく、男性・女性に共通して、非正規雇用や高等学校卒業の変数が年間所得に影響しているという結論が得られた。</p> <p>以上のことから、男女の賃金格差の要因を改善するために3点の改善策を提案する。(1) 非正規雇用者を中心に所定内給与の上昇を促す、(2) 教育水準を大卒まで引上げる、(3) 資格の取得の促進による正規雇用への転換を促すことによって男女の賃金格差が改善されることが期待できると考える。</p> |
| 参考文献及びデータ<br>(匿名データ以外) | <ol style="list-style-type: none"> <li>[1] 「男女の雇用格差と賃金格差」(2005), 阿部 正浩</li> <li>[2] 「職種を考慮した男女間賃金格差の分析」(2009), 佐藤浩一</li> <li>[3] 「働き方の男女不平等」(2017), 山口一男</li> <li>[4] 「日本における男女間賃金格差と職業選択」(2010), 野崎 華世</li> <li>[5] 永田靖・棟近雅彦共著: 「多変量解析法入門」(株)サイエンス社、2019年出版</li> </ol>   |

(6) 林玲奈の研究概要

|                        |  |
|------------------------|--|
| 研究テーマ                  | 働く貧困層（ワーキングプア）の要因分析<br>～労働者の貧困はなぜなくなるのか～   |
| 研究の目的                  | 本研究は、就業者の貧困（以下「ワーキングプア」という）を捉えるために、ワーキングプアの要因および属性を分析することを目的とする。   |
| 分析の方法                  | 目的変数を「ワーキングプアかどうか」に設定し、決定木分析の手法により、適切な説明変数を決定する。さらに、適切な説明変数から以下の3モデルから、ワーキングプアの要因と属性を分析する。推計モデルは、<br><ul style="list-style-type: none"> <li>・ロジスティック回帰分析</li> <li>・ニューラルネット</li> <li>・ランダムフォレスト</li> </ul> とし、モデルの予測精度を検証するために、元データから学習用データとテストデータを作成し、予測精度を確認する。  |
| 分析結果の概要                | 1) ロジスティック回帰分析の結果から、性別では、女性は男性よりもワーキングプアに陥りやすい。配偶者の有無では、未婚を基準とすると、配偶者と離別・死別している方がワーキングプアに陥りやすい。従業上の地位（非正規かどうか）では非正規雇用の方がワーキングプアに陥りやすい。従業者規模では勤め先の従業者数が少ないほどワーキングプアに陥りやすい。訓練・自己啓発の有無では訓練・自己啓発をした人を基準にしたときに、していない人の方がワーキングプアに陥りやすい。職業では、管理的職業従事者を基準としたときに、販売従事者、サービス職業従事者、保安従事者等がワーキングプアに陥りやすい。母子世帯かどうかの項目では、父子世帯・母子世帯のいずれにも該当しない世帯を基準としたときに、母子世帯はワーキングプアに陥りやすい。一般・単身世帯の別では、単身世帯はワーキングプアに陥りやすいという結論が得られた。<br>2) 予測モデルにより、ロジスティック回帰分析、ニューラルネット、ランダムフォレストの予測結果を検証すると、以下のように、ロジスティクス回帰の予測モデルの正答率は90.4%、ランダムフォレストの予測モデルは90.5%、ニューラルネットワークの予測モデル89.7%、いずれのモデルも9割近くの正答率となった。<br>以上のように、ワーキングプアの要因を分析したが、要因は1つだけではなく、複数の要因が関連している可能性が高いと考える。3つのモデルで分析したが、どの分析結果においても、一番大きく寄与しているのは非正規雇用という変数であり、ワーキングプアを解消するためには、不安定就業について、社会保障の拡充、賃金・給料の増額等の改善策が有効であると考えられる。 |
| 参考文献及びデータ<br>(匿名データ以外) | [1] 「就業構造基本調査」を用いたワーキングプアの規定因の検討, 村上雅俊 「統計学」第109号 2015年9月 <murakami.pdf (toukeigaku.sakura.ne.jp)>   |

|  |   |
|--|---|
|  | <p>[2] 都道府県別の貧困率の、ワーキングプア率、子どもの貧困率、捕捉率の検討, 戸室健作 山形大学人文学部研究年報第 13 号 (2016. 3) &lt;山形大学学術機関リポジトリ (nii.ac.jp)&gt;</p> <p>[3] 令和 4 年国民生活基礎調査 厚生労働省</p> <p>[4] OECD 統計 OECD.stat</p> |
|--|---|

## (7) 三村勇貴の研究概要

|                        |  |
|------------------------|--|
| 研究テーマ                  | 働く場所に見る育児時間の決定要因分析   |
| 研究の目的                  | 本研究は、育児休業の取得有無が育児時間に影響を与えているか勤める企業の属性との関連で検証することを目的とする。  |
| 分析の方法                  | 就業構造基本調査のデータから、育児時間を目的変数として、回帰分析を行う。育児時間は 6 区分の階級値となっているため、量的変数として変数変換することとする。説明変数の選択による 4 つのモデルと、変数間の交互作用を考慮した 1 モデル、合計 5 つの回帰モデルを作成し、それぞれについて結果を検証する。  |
| 分析結果の概要                | 5 つの回帰モデルから育児時間を検証すると、勤めている産業により、係数が統計的に有意な差があることが確認できた。以上の結果から、育児休業の取得は育児時間を増加させるが、その効果の大きさは勤め先の属性(産業や従業者規模)に影響される。この特性は、特化係数で分析しても同様の結果が得られた。育児時間の増加のためには、その属性ごとの重点的な対応を検討する必要があると考えられる。<br>以上のことから、育児時間の増加につながる育児休業の取得者数の増加のためには、産業ごとに「育児休業制度の整備促進」と「育児休業の取得促進」といった 2 段階に分離し、それぞれの進捗を踏まえた対応を検討することを提案する。  |
| 参考文献及びデータ<br>(匿名データ以外) | <p>[1] 竹沢 純子 (2023)「就業構造基本調査の個票データを用いた父親の育児休業取得に関する分析」(厚生労働行政推進調査事業費補助金(地球規模保健課題解決推進のための行政施策に関する研究事業))</p> <p>[2] 「日中韓における少子高齢化の実態と対応に関する研究」令和 4 年度 総括・分担研究報告書) 三菱UFJリサーチ&amp;コンサルティング (2017)「仕事と家庭の両立に関する実態把握のための調査研究事業報告書(平成 29 年度)」</p> <p>[3] 大島敬士・佐藤朋彦 (2021)「就業構造基本調査の個票データを用いた出産前後の女性の就業継続に関する要因分析」(『統計研究彙報』総務省統計研究研修所)</p> <p>[4] 内閣府 (2023)「令和 5 年度男女共同参画白書」</p> <p>[5] 内閣府 (2022)「令和 4 年少子化社会対策白書」</p> <p>[6] 厚生労働省 (2023)「令和 5 年度厚生労働白書」</p> <p>[7] OECD “Balancing paid work unpaid work and leisure(2021)</p> <p>[8] 厚生労働省 平成 29 年度雇用均等基本調査</p> |

(8) 吉崎雅基の研究概要

|                        |  |
|------------------------|--|
| 研究テーマ                  | 男性単独世帯における家事時間の規定要因の分析   |
| 研究の目的                  | 本研究は、男性単独世帯における家事時間を決定する要因や属性を明らかにすることを目的とする。  |
| 分析の方法                  | <p>社会生活基本調査のデータから以下のように男性単独世帯の家事時間を分析する。</p> <ul style="list-style-type: none"> <li>・男性と女性の家事時間に関して、新たなクロス集計表を作成し、男性と女性の差について統計的仮説検定の方法で検証する。</li> <li>・トービット・モデルの打ち切り回帰モデルによる分析</li> </ul>  |
| 分析結果の概要                | <p>1) 家事時間に関して、男性と女性を比較すると、予め用意した6区分の「仕事の時間」でも、全ての時間区分において、女性の家事時間が長いことが、統計的に有意な差として確認できた。また、「ふだんの日」「仕事の日」の比較においても女性の方が家事時間が長いことから、仕事に関わりなく女性の方が家事時間が長いという結果が得られた。</p> <p>2) ふだんの日トービット分析の結果をみると、仕事時間が長いほど家事時間が短い傾向にある。年齢については、年齢が高いほど家事時間が長い傾向にある。三大都市圏に居住している方が、三大都市圏以外に居住しているよりも、家事時間が短い傾向にある。休みの日のトービット分析の結果をみると、ふだんの日家事時間が長いほど家事時間が長い傾向にあり、ふだん家事をする人は、休みの日も家事をする時間が長いことが確認できた。このことは、ふだんの日家事時間が短い人は休みの日に埋め合わせるということではなく、家事の習慣や経験が家事時間の長さに影響があることが示唆される。ふだん家事をしている人が休みの日も家事をするという傾向は、結婚前から家事をしている人は結婚後も家事をするという傾向が存在する可能性を示唆している。</p>   |
| 参考文献及びデータ<br>(匿名データ以外) | <p>[1] 外務省 (n. d.). SDG グローバル指標 (SDG Indicators) 5:ジェンダー平等を実現しよう<br/><a href="https://www.mofa.go.jp/mofaj/gaiko/oda/sdgs/statistics/goal5.html">https://www.mofa.go.jp/mofaj/gaiko/oda/sdgs/statistics/goal5.html</a> (2023年10月30日取得).</p> <p>[2] 平井太規 (2019). 休日における夫の家事・育児動向の時系列的推移—「社会生活基本調査」政府統計匿名データの分析から— Discussion Paper Series A No.691, Institute of Economic Research, Hitotsubashi University.</p> <p>[3] 平井太規 (2022). 祖父母による家事・育児動向の基礎的分析—3世代世帯家族を対象に— 立教大学コミュニティ福祉研究所紀要, 10, 115-127.</p> <p>[4] 平田道憲・貴志倫子 (1999). 夫婦の家事労働時間に与える家事労働に関する態度の影響 日本家政学会誌, 50(9), 915-924.</p> <p>[5] 平田道憲 (2007). 共働き世帯と非共働き世帯の夫妻のワーク時間の時系列変化—家族類型からみた分析— 広島大学大学院教</p> |

|  |  |
|--|--|
|  | <p>育学研究科紀要, 56, 297-302.</p> <p>[6] 堀内かおる・天野寛子・伊藤純 (1997). 家事労働観と生活時間から見る夫妻のジェンダー関係—1995年世田谷区在住雇用労働者夫妻の調査から— 日本家政学会誌, 48(10), 851-864.</p> <p>[7] 稲葉昭英 (1998). どんな男性が家事・育児をするのか?—社会階層と男性の家事・育児参加— 現代日本社会階層調査研究資料集: 1995年SSM調査報告書, 4, 467-508.</p> <p>[8] 工藤寧子 (2015). 夫婦の家事分担に関する文献レビュー 東北女子大学・東北女子短期大学紀要, 54, 58-64.</p> <p>[9] 栗原由紀子 (2021). 子育て世代における生活様式の長期トレンドの変容—地域効果を考慮したマルチレベルモデルを用いて— 中央大学経済研究所年報, 53(2), 259-287.</p> <p>他</p> |
|--|--|

## 5. おわりに

令和5年度本科研修受講生の10名の研究レポートに匿名データが利用された。いずれも分析の方法については、多変量解析の手法を活用している。また、統計研究研修所での講義のカリキュラムの対象となっていない機械学習を用いた分析手法を利用している者もあり、近年、高度なデータ分析の手法の理解が広がっていることが窺えた。

2016年に「数理及びデータサイエンス教育の強化に関する懇談会」<sup>4</sup>として、文部科学省が大学における統計・データサイエンス教育強化の必要性や方策を取りまとめた。この後、データサイエンス教育が広く普及していることは、周知のことと思うが、このような状況が本研修所の本科研修の研究レポートの作成にも良い影響を与える結果となったと感じた。

また、受講生のうちの何人かは、分析結果に加えて、もう少し踏み込んだ個人の意見として、施策の提言としての改善策を記述していたので、分析結果の概要の項目に改善策の提案としてまとめている。

今年度の本科研修については、修了者全体が20名となっており、そのうちの半分の10名に匿名データを利用してもらえたことは幸いだと感じている。統計研究研修所としても初めての取り組みであったことから、新しい挑戦への高揚感とともに、不安やリスクがあった。また、事務的な準備が必ずしも十分でなかったこともあり、利用者の方に、迷惑や不便な思いをさせたことについては、率直にお詫び申し上げたい。

今年度の実績を踏まえ、統計研究研修所としても公的統計の匿名データの利活用推進及び社会課題の解決に向けたデータ分析研修の改善について、さらに尽力していきたいと考えている。

<sup>4</sup> 数理及びデータサイエンス教育の教科に関する懇談会(2016), 文部科学省

資料 匿名データを利用した令和5年度本科研修受講生一覧

| 氏名     | 所属機関(令和5年9月現在)      |
|--------|---------------------|
| 石井 秀人  | 防衛省海上幕僚監部防衛部        |
| 大西 満斉紀 | 岡山県総合政策局            |
| 川田 佳   | 総務省統計局統計調査部         |
| 櫻井 直也  | 独立行政法人統計センター情報システム部 |
| 土橋 卓也  | 防衛省航空自衛隊航空教育集団司令部   |
| 中山 拓人  | 総務省統計研究研修所研究部       |
| 那須 英和  | 独立行政法人統計センター情報システム部 |
| 林 玲奈   | 独立行政法人統計センター統計編成部   |
| 三村 勇貴  | 総務省政策統括官(統計制度担当)付   |
| 吉崎 雅基  | 独立行政法人統計センター統計編成部   |

参考資料1 匿名データ利用の留意事項・利用要件<sup>5</sup>

匿名データの利用は、以下の目的のものに限られる。また、下記のとおり利用要件が設けられている。

- 1) 学術研究の発展に資すると認められる統計の作成等（学術研究目的）
- 2) 教育の発展に資すると認められる統計の作成等（教育目的）
- 3) 国際社会における我が国の利益の増進及び国際経済社会の健全な発展に資すると認められる統計の作成等（国際比較統計利活用事業目的）
- 4) デジタル社会形成基本法（令和3年法律第35号）第38条第2項第13号に規定する特定公共分野に関する統計の作成等であって、国民経済の健全な発展又は国民生活の向上に寄与すると認められる統計の作成等（デジタル社会形成統計利活用事業目的）

本サービスを利用した場合、氏名、所属や研究成果・教育・事業の内容について提供申出書等に基づき、マイクロデータ利用ポータルサイト【miripo】で公表を行う。

所定の手数料を納付する必要がある。

匿名データの提供を受けた者には当該データの適正管理義務が課され、利用終了後は統計センターへ返却する必要がある。

匿名データを初めて利用する場合、本人確認のために統計センターまたは統計センター同様、匿名データの手続きの窓口業務を行っている機関（一橋大学、神戸大学、情報システム研究機構）へ来訪し、本人確認及び利用条件の説明を受ける必要がある。

法令に違反した場合の罰則のほか、利用条件（利用規約等）に反した場合は、統計センター及びすべての行政機関等による公的統計のデータ利用に係るサービスの提供禁止措置が科される。

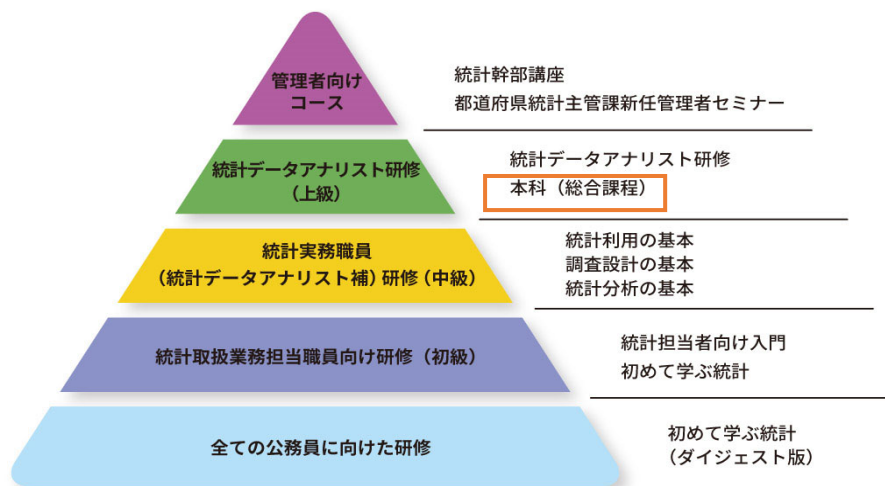
---

<sup>5</sup> 匿名データ利用ポータルサイト 独立行政法人統計センター

参考資料2 総務省統計研究研修所の研修体系<sup>6</sup>

■業務レベル別研修

統計の知識を体系的・段階的に習得



【本科(総合研修)の主要科目】

| 基礎   | 統計調査論                                 | 経済統計分析  |
|--|---------------------------------------|---|
| 統計法規<br>統計解析ソフト(R)の基本<br>統計グラフ作成演習<br>e-statの使い方 | 統計調査<br>意識調査<br>標本調査法<br>統計分類<br>国際統計 | 経済分析基礎理論<br>国民・県民経済計算<br>産業連関分析<br>労働・雇用分析<br>消費・物価分析<br>生産流通分析<br>経済・金融統計の見方 |
| 統計学  | 人口・社会統計分析                             | 研究  |
| 記述統計<br>推測統計<br>時系列解析<br>多変量解析                   | 人口分析<br>地域分析<br>GIS(地理情報システム)<br>社会分析 | 発表のための資料作成方法<br>調査企画(グループ演習)<br>個人研究<br>自主研究                                  |

■分野別研修

上記の業務レベル別研修以外に、「データ利活用研修」(7講座)、「統計作成実務コース」(3講座)、「人口・経済統計コース」(3講座)、「地域分析コース」(2講座)、「マイクロデータコース」(3講座)、「都道府県職員・教育関係者向けコース」(4講座)を分野別研修として実施している。

<sup>6</sup> 令和5年度統計研修のご案内 総務省統計研究研修所



## 【参考文献】

- [1] 総務省 匿名データの利用・提供及びオーダーメイド集計.  
<<https://www.stat.go.jp/info/tokumei/index.html>>
- [2] 内閣官房 デジタル田園都市国家構想.  
<<https://www.cas.go.jp/jp/seisaku/digitaldenen/about/index.html>>
- [3] 独立行政法人統計センター 匿名データの利用方法.  
<<https://www.nstac.go.jp/use/archives/anonymity/>>
- [4] 文部科学省 数理及びデータサイエンス教育の強化に関する懇談会(2016).  
<[https://www.mext.go.jp/b\\_menu/shingi/chousa/koutou/080/index.htm](https://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/index.htm)>
- [5] 独立行政法人統計センター ミクロデータ利用ポータルサイト, 匿名データの利用実績 提供状況一覧 (匿名データ) .<<https://www.e-stat.go.jp/microdata/anonymity>>
- [6] 総務省統計研究研修所 令和5年度統計研修のご案内.  
<<https://www.stat.go.jp/training/1kenshu/pdf/2023panf.pdf>>
- [7] 小林良行(2011), 匿名データの教育目的利用に関する一考察, 『統計学』第100号, 100-105頁.
- [8] 星野伸明(2010), 「公的統計マイクロデータ提供制度の課題」『日本統計学会誌』, 第40巻, 第1号, 23-45頁, 日本統計学会.
- [9] 松田芳郎(2008), 「日本におけるマイクロ政府統計活用の新しい夜明け」, 『統計』2008年12月号, 2-9頁, 日本統計協会.
- [10] 伊藤伸介(2022), ミクロデータの匿名化と統計情報の秘匿可能性について, 経済学論纂(中央大学)第63巻第1,2合併号.
- [11] 伊藤伸介(2016), 諸外国における公的統計マイクロデータの提供の現状とわが国の課題, 中央大学経済研究所年報 第48号(2016) 233-249頁.
- [12] 森博美(2004)「マイクロデータの利用特性と統計利用論」法政大学日本統計研究所『研究所報』No.32, 9-38頁.
- [13] 山口幸三・伊藤伸介・秋山裕美(2013)「教育用擬似マイクロデータの作成—平成16年全国消費実態調査を例として—」『統計学』104号, 1-15頁.
- [14] 山口幸三(2008)「政府統計の個票利用と統計法改正」『経済研究』第59巻第2号, 139-152頁.
- [15] 白川清美, 武藤杏里, 田中雅行, 合田智一, 千葉亮太(2022), 公的統計匿名データを利用したデータサイエンス講義のための取組み, データサイエンス研究創刊号, 2022年3月 63~77頁.
- [16] 統計情報研究開発センター編(2015), 公的統計のマイクロデータ利用ガイド: 社会生活基本調査の匿名データを用いた分析を例として, 統計情報研究開発センター.
- [17] ミクロデータ利用ポータルサイト, 匿名データの利用実績 提供状況一覧 (匿名データ) .
- [18] Thorogood D.(1999) “Protecting the Confidentiality of Eurostat Statistical Outputs”, Netherlands Official Statistics, Volume 14, Spring, pp.30-33.
- [19] Bethlehem, J. G., Keller, W. J. and Pannekoek, J.(1990) “Disclosure Control of Microdata”, Journal of the American Statistical Association, Vol.85, No.409, pp.38-45.
- [20] 欧州統計局 (Eurostat) .  
<<http://ec.europa.eu/eurostat/web/microdata>>

