

## 行政における統計データの利活用の推進に関する研究

- 研究代表者 椿 広計 (大学共同利用機関法人情報・システム研究機構統計数理研究所・名誉教授)
- 研究分担者 南 和宏 (大学共同利用機関法人情報・システム研究機構統計数理研究所・データ科学研究系・教授)
- 研究分担者 岡 檀 (大学共同利用機関法人情報・システム研究機構統計数理研究所・医療健康データ科学研究センター・特任准教授)
- 研究分担者 岡本 基 (大学共同利用機関法人情報・システム研究機構・戦略企画本部・主任URA／特任准教授)
- 研究分担者 高部 勲 (立正大学データサイエンス学部・教授)
- 研究分担者 山下 智志 (大学共同利用機関法人情報・システム研究機構統計数理研究所・教授・副所長・リスク解析戦略研究センター長)
- 研究分担者 山内 慶太 (慶應義塾大学大学院健康マネジメント研究科・教授)
- 研究分担者 渡辺 美智子 (立正大学データサイエンス学部・教授)
- 研究分担者 新井 崇弘 (慶應義塾大学大学院健康マネジメント研究科・大学院生／厚生労働大臣指定法人・一般社団法人いのち支える自殺対策推進センター・分析官)
- 研究分担者 久保田 貴文 (多摩大学経営情報学部・准教授)
- 研究分担者 竹林 由武 (福島県立医科大学医学部・助教)
- 研究協力者 阿部 穂日 (独立行政法人統計センター 統計情報提供課)
- 研究協力者 赤谷 俊彦 (総務省・統計データ利活用センター長)
- 研究協力者 六信 孝則 (株式会社帝国データバンク・企総部企画課 課長補佐)
- 研究協力者 田上 紀代美 (慶應義塾大学大学院健康マネジメント研究科・大学院生)

## 要旨

本研究は、自殺総合対策のEBPM（Evidence Based Policy Making）に資するデータ分析環境の整備を目的に、行政における統計データの利活用推進を加速するための次の3つの活動を実施した。

第1は、自殺総合対策に関係する総務省・厚生労働省の公的統計マイクロデータの政策利用を加速するオンサイト拠点形成である。拠点設置準備並びに既存拠点のデータ分析環境拡充を進め、立正大学データサイエンス学部、慶応義塾大学大学院健康マネジメント研究科における設置を2022年度に実現すべく準備を進めた。また、オンサイト拠点で大規模データ分析に利用可能なデータ解析手法の検討を進め、情報・システム研究機構オンサイト拠点に実装した。

第2は、オンサイト拠点利用を加速する公開可能データ作成や匿名化ツールの開発研究である。官学一般の統計マイクロデータ利活用を容易にする元マイクロデータの構造を保持した疑似マイクロデータの作成を行った。また、マイクロデータの詳細な地域情報公開を可能とする匿名化アルゴリズムを開発し、秘匿の安全性評価も行った。

第3は、自殺総合対策における公的統計マイクロデータの利活用に関する研究を多角的に行った。

全国1,735市区町村の自殺統計から2020年前後の自殺率の変化を推定する指標「自殺率上昇度」を作成し、自殺率上昇の地域差や性差、その背景要因について分析を行い、コロナ禍による日本の自殺率上昇に地域差が生じていたことを明らかにし、その背景には産業構造の差異があり、男女の勤務形態の違いが自殺率の性差を生じさせた可能性が考えられた。また、社会生活基本調査匿名マイクロデータに潜在クラスモデルを適用し、介護者の日常活動時間に関する情報を分析し、介護者が抱える様々な問題を可視化した。特に、介護行動時間と他の生活行動時間との関連から、介護者一般の生活行動パターンの多様性を可視化し、単身男性の介護生活行動に潜む社会的孤立要因の課題を見出した。さらに、自殺対策においてメンタルヘルスサービスの利用を促進・阻害する要因について検討するために、ウェブアンケートを計画し、2022年度に調査を実施する。

## 1. 研究目的

本研究班は、自殺総合対策をはじめとする政策立案において、厚生労働省や総務省等の統計データ、特にマイクロデータと呼ばれる調査票個票情報を一層活用することが、証拠に基づく政策立案（EBPM、Evidence Based Policy Making）にとって極めて有効との問題意識を基に、研究者が分担し、総務省統計局統計データ利活用センターや情報・システム研究機構が運用する「公的統計マイクロデータ研究コンソーシアム」と連携して、次の3つの研究活動を行うことを目的とする。

第1の目的は、機微な個人情報を含む「統計マイクロデータ（公的統計等のために収集されている調査票個票）」を安全な情報環境下で、政策研究に代表されるような公共研究目的ならば、自在に探索的分析できるオンサイト拠点と呼ばれる環境の研究機関における形成を加速することである。この背景には、平成30年4月に総務省統計局が、統計データ利活用センターを設置し、公的統計作成のために収集された全府省の調査票情報（統計マイクロデータ）の利活用を通じた行政課題解決・公益性の強い研究支援を開始したことが挙げられる。特に、令和元年5月に施行された改正統計法により、統計マイクロデータが、研究代表者が所属する情報・システム研究機構、総務省統計データ利活用センター、（独）統計センターと、全国8大学に設置されたオンサイト拠点で総括的かつ探索的に分析する仕組みが生まれたことが挙げられる。自殺総合対策に資する統計マイクロデータ・行政情報などの活用環境整備が必要かつ可能な時機となっているのである。

第2の目的は、統計マイクロデータを分析する研究者等を支援するツールや環境を開発・提供することである。統計マイクロデータが一般には公開できない機微な情報であり、それに基づく分析もオンサイ

ト拠点という厳格な情報セキュリティ環境下で実施され、分析結果の外部持ち出しと公表も一定の制限がかかるということが背景にある。統計マイクロデータの全情報は、分析者の研究室内で利用できないこと、またオンサイト拠点で分析集計された結果の持ち出しには制限がおかれていることを承知しなければならない。この第2の目的は、統計マイクロデータの構造を保持した疑似的マイクロデータを公開し、研究室などで試行的な分析を可能にすることと、オンサイト拠点で分析した結果に適切な匿名化アルゴリズムを施し、匿名化という観点ではリスクのない分析結果をオンサイト拠点外に持ち出し、公表可能にするものの2つの目標に具体化される。特に、今後自殺総合対策に公的統計マイクロデータと共に利活用が考えられる個人の医療サービスの受診歴に関するマイクロデータには、受診者に関する機密情報が含まれており、研究利用のための2次的利用を行う際、公表する分析結果から機密情報が漏洩しないような防護措置が必要となる。近年、匿名データの代表的な安全性手法として k-匿名化が提案されているが、実際に匿名データを作成するには、外観識別性の高い準識別子情報を定義し、レコード情報の粒度を粗くするための一般化処理方法等、具体的な秘匿処理技術を確立する必要がある。本研究では、公的統計マイクロデータのみならず医療マイクロデータに含まれる情報の統計的性質、外部データとの照合リスク等を考慮して、安全で有用な匿名データの作成手法の確立を目指す。

第3の目的は、統計マイクロデータ等の探索的分析が、自殺総合対策に資する情報を提供し得るというユースケースを示し発信することである。この政策実証研究のために、ヘルスデータサイエンティスト協会運営の中核機関である慶應義塾大学健康マネジメント研究科との共同研究を組織した。なお、自殺総合対策に資するデータサイエンス的研究については、統計マイクロデータに限らず、オープンデータや必要な調査研究などの企画も行っている。

以下では、3つの目的にそって、研究方法、結果、考察・結論、政策提案・提言を記載する。

## 2. 研究方法

### 2-1 統計マイクロデータの政策利用を加速するオンサイト拠点形成

山下、岡本、南、高部はオンサイト拠点における公的統計マイクロデータの利活用を推進するために官学で組織された「公的統計マイクロデータ研究コンソーシアム」の運営メンバー（南が運営委員長）として、総務省統計局統計データ利活用センター、(独)統計センターと連携して、オンサイト拠点設置の全国への呼びかけオンサイト環境設置の方法やそこで行う公共的データ分析の普及啓発行事を企画・実施した。

高部、渡辺、山内、岡は、立正大学データサイエンス学部ならびに慶應義塾大学大学院健康マネジメント研究科におけるオンサイト拠点設置の準備をおこなった。

岡本、椿は、情報・システム研究機構オンサイト拠点における分析環境の整備を行った。

### 2-2 オンサイト拠点利活用を加速する公開可能データ作成や匿名化ツール等の開発研究

高部は、**Synthetic Data**の考え方を基に、自殺対策等に関連する行政記録情報や調査票情報等のデータ利活用に資する公的統計の疑似的なマイクロデータ（一般用マイクロデータ）について、中間的な集計表や回帰モデルの推定結果などを基に作成する方法に関して、実データを基に試作・分析・検討を行った。

このため、企業に関する商用データ（帝国データバンクのデータ）及び世帯に関する匿名データ（社会生活基本調査）を用いて、**Synthetic Data**の考え方に基づく一般用マイクロデータの施策・分析・

検討を行った。具体的には、それぞれのデータに関し、2項・多項ロジットモデル、順序ロジットモデル、重回帰モデル等を用いて、中間的な集計表・モデルの推定結果・残差の情報から、疑似的なデータを生成した。次に、作成した疑似的なデータと元のデータとの差や、それらのデータ及び元のデータを用いた重回帰モデル等の解析結果を比較し、元のデータの構造をどの程度保持しているかについて分析・検討を行った。

南は、代表的な k-匿名化アルゴリズムである Mondrian アルゴリズムを実装し、安全性のパラメータである k 値、データ効用の指標を変数として匿名化データの作成し、データの安全性と有用性に関する実証的評価を行う。また匿名化処理において必要となる情報粒度の調整を実施するための一般化階層の定義を行い、特に外観識別性の高い地域情報に関する一般化階層を具体的に定式化した。

さらに、南は地域情報に含まれる市町村の情報を国土交通省が公開する GPS 位置情報との照合を行うことで、GPS 基準値に基づく地域情報の再帰的な分割を検討し、詳細な地域情報を含む有用性の高い匿名データの実現可能性を検討した。

久保田は、統計マイクロデータ等とリンケージする情報として、テキストデータの活用方法について検討した。特に、SNS (Twitter) のデータ収集技法を整備し、コロナ禍において、漠然とした不安から、顕在化した怒りや症状 (痛みなど) に移行し、さらに希死念慮を抱くようになるようなプロセスの確認を試みた。このため、TwitterAPI に登録し、「漠然とした不安」「怒り」「痛み」「希死念慮」などのキーワード群を設定し、API の制限を超えないように、日々 Twitter におけるつぶやきをクロールし、その日時、テキストの内容、さらには SNS 特有の、リツイートの有無やリプライなどについて収集した。さらに、収集したデータの基礎集計を行うとともに、テキストの内容についても、極性分析でつぶやきのポジティブ・ネガティブについて確認を行った。また、テキストの内容については、トピックモデルによって分類し、その意味について解釈を行った。

### 2-3 自殺総合対策における公的統計マイクロデータの利活用に関する研究

岡、山内、渡辺、椿は、10 年以上にわたり減り続けてきた日本の自殺が 2020 年に入ってから増加に転じ、COVID-19 パンデミックとの関係が指摘されている。失業、経済的困窮、社会的活動の減少、孤立等の変化がどのように自殺増加に影響したのか、実態の把握を試みた。このため、自殺問題に関する先行研究を参照し、対象地域区分、指標の選択、解析手法などを整理した。特に、自殺総合対策に資する公的マイクロデータの活用について検討を行った。

また、全国の 1,735 市区町村の過去 11 年間の自殺統計データを参照し、2020 年前後の自殺率の変化を推定する指標「自殺率上昇度」を独自に作成した。COVID-19 パンデミック後の全国の自殺率上昇の度合いに対し、市区町村毎の自殺率上昇の度合いがどれだけ大きかったか小さかったかを推定する方法を検討し、以下の手順により計算した。人口規模の小さな町村では僅かな発生件数であっても率に過大な変動が生じる可能性があるため、その影響を抑制するために過去 10 年間の自殺率を参照することとした。

- A 1,735 市区町村毎の、COVID-19 前の各市区町村自殺率”偏差値”  
過去 10 年の市区町村毎の自殺率平均値と、過去 10 年全国平均値との差
- B 市区町村毎の、COVID-19 後の各市区町村自殺率”偏差値”  
2020 年の市区町村毎の自殺率と、2020 年全国の自殺率との差

- C 全国の COVID-19 後の自殺率上昇  
過去 10 年の全国自殺率平均値と、2020 年全国自殺率の差
- D 市区町村毎の、COVID-19 後の自殺率上昇度  $= (B-A)/C$

さらに市区町村ごとに 14 種類の産業別住民就業率のデータを連結してパネルデータを構築し、自殺率上昇の地域差や性差、その背景要因について分析を行った。

渡辺、山内、田上は、公的統計マイクロデータの利活用を推進するための分析技術として、観察記録データ、とくに、比較的規模の大きな、所謂、リアルワールドデータに対する因果構造モデル探索の技法として 2020 年度本委託研究で教材化した潜在クラス分析の手法を取り上げ、行政データ、とくに、行政記録および公的統計における調査票データの分析活用の実証研究および行動計量的活用事例の開発を行った。すなわち、(独)統計センターより提供された総務省社会生活基本調査の匿名化データ(調査票 A【生活時間編】272,861 人)を基に潜在クラスモデルを適用し、介護行動時間を軸に他の生活行動時間との関連から、介護者一般の生活行動パターンの多様性を可視化し、とくに単身男性の介護生活行動に潜む社会的孤立要因の課題を探索した。また、新井、山内、渡辺では、介護高負荷による深夜不眠者の同定を目的として、潜在クラスモデル及びヒートマップによる睡眠時間帯の可視化、さらに個別リスク評価を行った。渡辺は、これらの研究結果を発信するために、2021 年 12 月 18～19 日に開催された 35 学会の連合体である横断型基幹科学技術研究団体連合主催第 12 回横幹連合コンファレンス「横幹知で拓くポストコロナ社会」において、企画セッション「行政課題の解決及び国民の生活研究における大規模統計データの利活用」を実施した。

竹林は、メンタルヘルスに不調を抱える人の約 9 割が、専門的なメンタルヘルスサービスを受けていないことが明らかとなっていることから、メンタルヘルスサービスの利用を促進・阻害する要因について検討するために、ウェブアンケートを計画した。本年度は、メンタルヘルスサービスの利活用の阻害、促進要因を検討した研究の文献レビューを実施し、調査項目の選定と整備を行った。調査に適した日本語版の項目が存在しないものについては、原版項目の邦訳をした。具体的にはメンタルヘルスサービスの利用に対する心理的バリアーのリストである Barriers to Treatment Questionnaire (BTQ: Smith et al., 2021)と、多様性の受容性を測定する The Acceptance and Action Questionnaire – Stigma (AAQ-S: Levin et al., 2014)の邦訳を行い、調査項目を整備した。

### 倫理面への配慮

本研究班の用いたマイクロデータの多くは匿名化されたデータの分析であった。ただし、匿名化技法を扱う研究においては、倫理委員会による研究計画の承認を行った上で実施している。また、竹林が実施を計画した調査研究についても倫理委員会審査を受けたが、倫理委員会勧告への対応のため、実施は 2022 年度となった。

## 3. 研究結果

### 3-1 統計マイクロデータの政策利用を加速するオンサイト拠点形成

#### 3-1-1 統計マイクロデータ分析拠点形成・利活用の普及啓発活動

山下、岡本、南、高部、椿は、2021 年 11 月 19 日に「公的統計マイクロデータ研究コンソーシアムシンポジウム 2021」を下記のプログラムで開催した。特に午前中は、多くの研究者に対してオンサイト利

用に関するチュートリアルを行い、公的統計マイクロデータの提供側・利用側の双方からの解説が行われ、今後のオンサイト拠点利活用促進に繋がる実践的な内容を提供できた。午後のセッション後半が統計データの高度利用について、本研究班分担者が研究している疑似マイクロデータ作成やデータの匿名化に関する研究発表に充てられた。結果として、約 100 名の参加者にオンサイト拠点などの利活用についての啓発を行うことができた ([http://jmodc.org/event/sche\\_repo\\_2021sympo.php](http://jmodc.org/event/sche_repo_2021sympo.php))

なお、このチュートリアルセッションについては、オンサイト利用に関する有用な情報を提供する教材として加工し、今後も広く参照提供できるように、下記赤谷俊彦氏（総務省統計局統計データ利活用センター長）と阿部穂日氏（（独）統計センター統計情報提供）、2022 年 3 月コンソーシアムの WEB サイト上で公開した (<http://jmodc.org/videos/>)。

## 【公的統計マイクロデータ研究コンソーシアムシンポジウム 2021】

### I オンサイト利用に関するチュートリアル 司会：伊藤 伸介（中央大学 経済学部）

チュートリアル開会挨拶：南 和宏（統計数理研究所教授）

- ① 「統計データ利活用センターにおけるオンサイト利用推進の取組」  
赤谷 俊彦（総務省統計局・独立行政法人統計センター 統計データ利活用センター長）
- ② 「オンサイト利用による分析結果等の安全性確認における注意と事例」  
阿部 穂日（独立行政法人統計センター 統計情報提供課）
- ③ 「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」  
白川 清美（立正大学 データサイエンス学部教授）

### II 公的統計マイクロデータ研究コンソーシアムシンポジウム

シンポジウム開会挨拶：藤井 良一（情報・システム研究機構長）

#### i 「公的統計と統計教育」 司会：岡本 基（統計数理研究所）

- ① 「政府における統計人材の確保・育成に向けた取組」  
稲垣 好展（総務省 統計局 調査企画課）
- ② 「データサイエンス教育と公的マイクロデータに期待する役割」  
渡辺 美智子（立正大学 データサイエンス学部）
- ③ 「統計エキスパート人材育成プロジェクトの推進」  
千野 雅人（統計数理研究所）
- ④ 「SSDSE を中心とした統計教育への取組み」  
山下 雅代（独立行政法人統計センター）

#### ii 「統計データの高度利用に関する研究」 司会：山下 智志（統計数理研究所）

- ① 「Synthetic Data の考え方に基づく疑似的なマイクロデータ作成の可能性」  
高部 勲（立正大学 データサイエンス学部）
- ② 「公的統計データの匿名化に関する海外の動向とわが国における課題」  
伊藤 伸介（中央大学 経済学部）
- ③ 「公的マイクロデータに対する k-匿名化加工の検討」  
南 和宏（統計数理研究所）

閉会挨拶: 椿 広計 (統計数理研究所長)

また、公的統計マイクロデータ研究コンソーシアムでは、公的統計マイクロデータの研究利活用を促進するために2022年3月ニュースレターを創刊した。創刊号では、椿、山下、南、岡本が、座談会「公的マイクロデータ二次利用の道のりと活用促進への期待」を行い、統計マイクロデータの研究利活用が可能になったこれまでの経緯と今後の展望を示し、活動の重要性をアピールした。

### 3-1-2 オンサイト拠点形成準備

高部、渡辺は立正大学データサイエンス学部に、山内、岡は慶應義塾大学大学院健康マネジメント研究科にオンサイト拠点形成の準備を行った。オンサイト拠点形成の要件となるサイネットへの接続が、サイネットのバージョンアップで今年度難しかったこともあるが、いずれも2022年度には設置が予定されている。オンサイト拠点形成は、(公財)統計情報研究開発センターが大学に対する設置経費支援を2020年度から開始したために、現在地方大学などでの設置が加速しつつある。なお立正大学は2021年度のオンサイト拠点形成支援経費を獲得した。

### 3-1-3 既存のオンサイト拠点の分析ツール拡充

情報・システム研究機構オンサイトなどで、2020年度にオンサイト拠点でマイクロデータ分析に有用として教材開発を実施した、潜在クラス分析を利用可能にする、あるいはオンサイト拠点で利用可能とするメッシュ統計データなどの購入実装を進めた。

## 3-2 オンサイト拠点利活用を加速する公開可能データ作成や匿名化ツールの開発研究

### 3-2-1 疑似的なマイクロデータの性能

高部は、2-2で策定した疑似的なマイクロデータの統計分析の結果、周辺分布の統計表、回帰モデルの結果、残差の分布に関する情報を事前に作成・公開することにより、実際のマイクロデータの構造をある程度保持した疑似データを作成できることを確認した。ただし、一部の変数では、相関係数の符号が逆になるなどの結果もあり、より大量の公的統計マイクロデータを利用して、更なる分析を行う必要がある。

### 3-2-2 匿名化ツールと匿名化データの提供

南は、準識別子情報として、年齢、性別、地域情報を想定し、国立がん研究センターより提供を受けたがん登録情報のk-匿名化データを作成し、元データのレコードの識別リスクの評価を行った。評価実験の結果、若年層、高年層のレコード数が極端に少なく、年齢区分に関するトップ(ボトム)コーディングの必要性が判明した。また地域によるサンプル数のばらつきは大きいため、画一的な地域情報の一般化処理では情報損失が大きく、本研究で提案した手法による地域情報のGPS座標値に基づく隣接性を考慮した分割手法の有効性が確認できた。

### 3-2-3 言語情報データの収集分析のプロセスに関する研究

久保田の用いた方法によって、希死念慮のツイートについては、頻度の高い時間帯・月などを確認できた。また、感情分析の結果、月や曜日による違いは無かったものの、時間帯による傾向を確認す

ることができた。また、期間を限定してトピックモデル（重要語）の結果より、3つの特徴的なトピックに分類することは実施したが、本来の意味での希死念慮のトピックは検出されなかった。

図1に2021年5月から10月に収集した全 Tweet データの頻度を左から月ごと、曜日ごと、時間帯ごとに示す。これより、月としては5月および8月が他の月よりも多く、曜日としては日曜日と水曜日が多いことがわかる。また、時間帯としては朝の時間帯は5時台から8時台にいくにつれて増加し、9時台になると減少し、そののちは時間帯とともに増加し、夜の9時・10時頃が最も高くなっている。

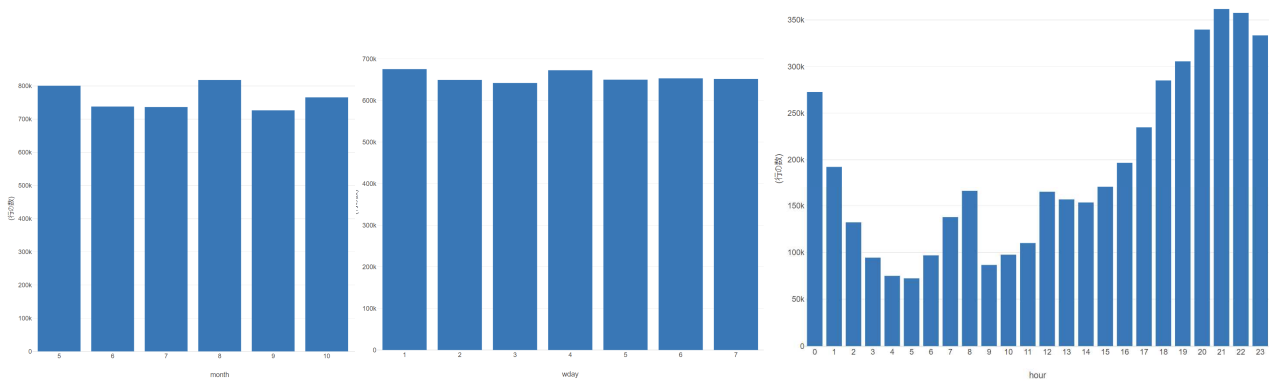


図1 Tweetの頻度の棒グラフ（左から月ごと、曜日ごと、時間帯ごと）

図2に同期間のデータのうち、10件以上リツイートがあるデータの感情分析の結果のうちの、時間帯ごとのスコアの箱ひげ図を示す。これより、先に示した、頻度として増加した朝の時間帯（5時台から8時台にかけて）については、ツイートはネガティブな内容に推移していることがわかる。なお、月ごと曜日ごとの感情分析のスコアについては大きな変化は見受けられなかった。

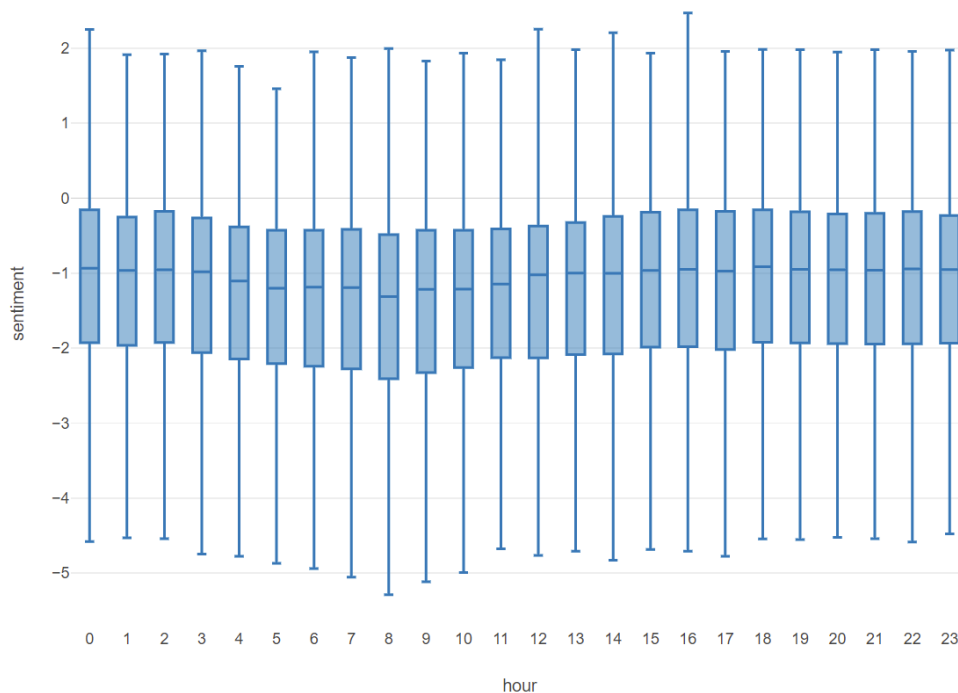


図2 Tweetの感情分析のスコアの箱ひげ図（時間帯ごと）



図3に5月1日の Tweet だけをピックアップして、その内容をトピックモデルにより分析した結果のうち、重要語の棒グラフを示す。ここでは、重要語だけでは判断できなかったため、個々の Tweet を確認したところ、トピック1はリンクやTV番組のタイトルなど「情報提供」の内容、トピック2は恥ずかしい気持ちやソーシャルゲームに課金をしたことに対する反省など「ネガティブな気持ちの高ぶり」の内容、トピック3は〇〇してから死にたいというような「ポジティブな気持ちの高ぶり」の内容を含むことが判明した。なお、これらのトピックについては、1日だけの内容のため、さらなるデータを用いた分析が必要である。

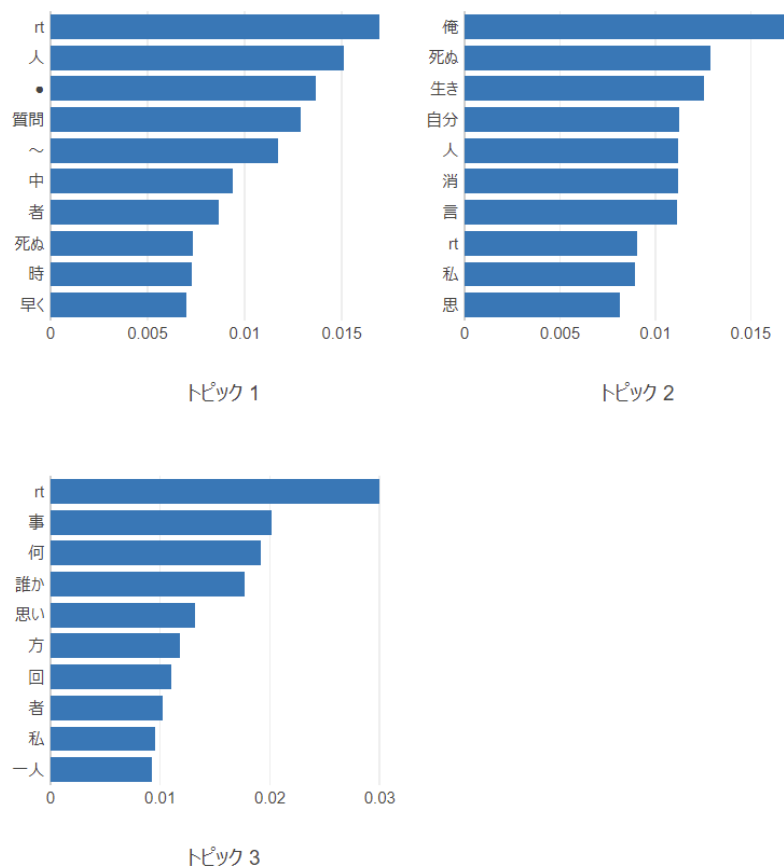


図3 5月1日の Tweet の内容のトピックモデルにおける重要語の棒グラフ

### 3-3 自殺総合対策における公的統計マイクロデータの利活用に関する研究

#### 3-3-1 COVID-19 パンデミック下での自殺率上昇の地域差や性差、その背景要因について分析

岡、山内、渡辺、椿は、自殺総合対策研究における公的マイクロデータの活用について、データの種類やそれぞれの特徴、有効性を挙げ、先行事例を参照して活用の利点を整理した。自殺統計と産業構造指標を組み合わせた分析の結果、2020年の市区町村の自殺率上昇は内需型サービス業への就業率と有意な関係があり、失業や休業の増加が自殺リスクを高めている可能性が示唆された。中でも、宿泊業・飲食サービス業においては、女性の自殺率上昇度は男性よりもはるかに大きいことが明らかとなった(図4)。静岡県を取り上げて自殺率上昇度の分布を地図上に描出したところ、女性の自殺率上昇度が特に高かった市区町村は、宿泊業・飲食サービス業が盛んな伊豆半島に集中していた。本研究は、第12回横幹連合コンファレンスで発表した。発表者の岡はこの発表によって、第11回木村賞を受賞した。

[製造業就業率上位100の市区町村]

[宿泊業・飲食サービス業上位100の市区町村]

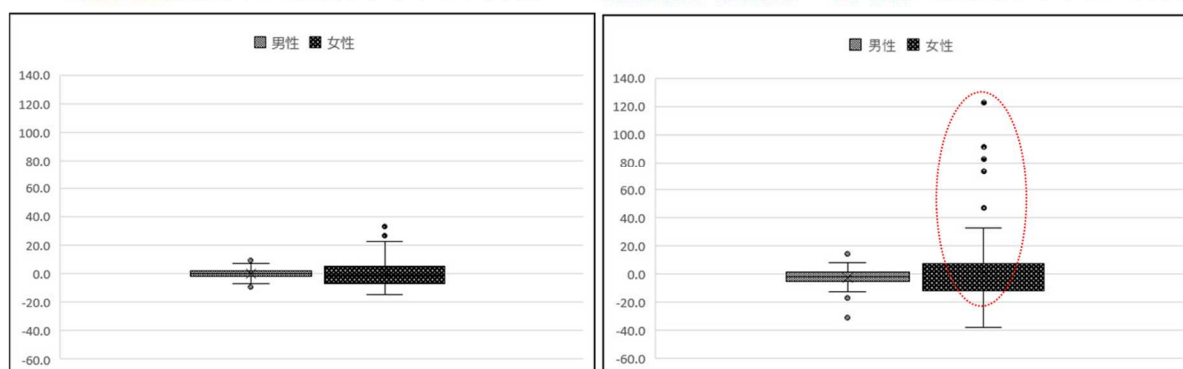


図4 製造業と宿泊業・飲食サービス業住民就業率；上位100市区町村での自殺率上昇度の分布

### 3-3-2 総務省社会生活基本調査匿名化データの分析

渡辺、山内、田上は、社会生活基本調査の匿名マイクロデータを基に潜在クラスモデルを適用し、介護時間を主に他の行動時間との関連から、介護者の生活行動パターンを明らかにし、とくに単身男性の介護生活行動に潜む社会的孤立要因の課題を見出した。特に、新井、山内、渡辺は介護負荷の高いグループでは、0:00 から 6:00 にかけて、睡眠行動者の割合が他のグループよりも低下していることがわかった。さらに、帰属確率に基づく個別リスク評価によって深夜不眠者を同定した。これらの研究は、第12回横幹連合コンファレンスで発表した。

竹林は、先行研究のレビューから、メンタルヘルスサービスの利用阻害要因としては、性別、年齢、社会経済的地位などの人口統計学的要因、精神疾患に対するスティグマに基づく多数の心理的障壁が存在することを明らかにした。一方、メンタルヘルス・リテラシーや多様性の受容は、メンタルヘルスサービスの利用を促進する可能性が示唆された。これらの要因の複雑な関係について実証的な検討を行うために、次年度に調査を実施することとし、福島県立医大の倫理委員会の審査を受けた。

## 4. 考察・結論

自殺総合政策に資する公的統計マイクロデータを探索的に分析することが可能なオンサイト拠点については、その制度的成立から3年を経て、官学共同のコンソーシアムにおける啓発もあり全国の研究者に対する認知が進むとともに、大学への設置を支援する補助金制度も確立された。一方、オンサイト拠点はマイクロデータを自在に分析できるとはいえ、3つの問題が明らかになりつつある。

第1は、これまでの集計データに基づく実証分析とは全く異なる大きさのデータを扱える社会科学・公衆衛生学・医学系研究者や行政関係者が、まだまだ日本には多くないことである。このため、簡単な機械学習や潜在クラスモデルを公的統計マイクロデータで扱える教材の開発は、2020年度のJSCP委託研究で開発したが、潜在クラスモデル等の有用性と必要性を多くの研究者が認識する必要がある。

第2は、データ分析環境としてのオンサイト拠点が、自身の研究室におけるデータ分析に比して、研究者にとって依然として窮屈なものであることである。本研究班では、研究室で利用可能な疑似マイクロデータの開発を進めているが、この種のオープンデータの作成技法の研究は研究者が行えるが、実際のオープンデータの作成は、マイクロデータ作成部局が定常的に行える体制の整備も必要ではないかと考える。

第3は、公的統計マイクロデータとしてオンサイト拠点で利用可能なデータとして、厚生労働省人口動

態統計調査、総務省社会生活基本調査、総務省労働力調査などが着々として増えてきているが、K6を有し、K6を調査に加えた根拠を総務省統計審議会（当時）答申時に自殺対策に資するとされた、厚生労働省国民生活基礎調査が、依然としてオンサイト拠点で利用できない状況が続いていることである。政府方針として、基幹統計調査のマイクロデータはオンサイト拠点で利用可能にするということが掲げられている以上、速やかな提供が必要である。なお、国民生活基礎調査匿名化データは提供されており、竹林、久保田、椿は、K6を対象とした探索的リスク分析を厚生労働科研費の支援を受けて実施したが、当該データには地域情報が含まれておらず、リスクの地域性を分析するには、マイクロデータが必要である。

本研究班では、公的統計データを自殺統計データとマクロ的に結合した結果、コロナ禍による日本の自殺率上昇は全国一律に起きたわけではなく、より上昇した地域とそうでない地域の格差が生じていたことを明らかにした。地域差の背景には産業構造の差異があり、また、男女の勤務形態の違いが自殺率上昇の性差を生じさせている可能性が考えられた。介護負荷の高いグループにおいては、平日や土日によって睡眠行動者の割合が断片的であった。睡眠の連続性が失われている可能性が示唆された。この分析については、さらに労働力統計などのマイクロデータによって、失業の地域性情報などを追加すれば、さらに詳細な自殺リスクに対するエビデンスを得られるものとする。今後、公的統計マイクロデータのオンサイト拠点、医療情報のオンサイト拠点、自殺統計マイクロデータのオンサイト拠点の間でセキュアな分析環境を作成し探索的な分析を加速させることで、より詳細な地域自殺実態とリスク要因を浮かび上がらせる可能性がある。

## 5. 政策提案・提言

### 5-1 統計データの活用環境整備に関する提言

- ①公的統計マイクロデータとしてオンサイト拠点で分析可能なマイクロデータについては、自殺総合対策に資するデータがまだ完全に提供されているわけではなく、特に厚生労働省国民生活基礎調査のマイクロデータのオンサイト拠点への提供は優先順位の高いものであり、早期の実現を要望する。
- ②自殺統計マイクロデータと公的統計マイクロデータ（人口動態統計等）が同時に分析可能なオンサイト拠点環境の整備を行うことが、自殺統計マイクロデータに基づく政策分析の精度改善にも繋がり必要である。
- ③統計マイクロデータや行政情報のリンケージや分析を支援可能な人材の系統的育成に高等教育機関は取り組み、政策研究に一定規模の人材が投入される必要がある。

### 5-2 統計データ活用に基づく自殺総合政策研究の加速に関する提言

- ①コロナ禍による日本の自殺率上昇は全国一律に起きたわけではなく、より上昇した地域とそうでない地域の格差、そして性差が生じていたことに鑑みて、対策の実施における優先順位の設定と、背景要因によって異なる支援内容を策定することが重要である。この政策研究のエビデンスを向上させるためにも地域産業連関構造や労働力統計情報との迅速な結合環境の整備が必要となる。
- ②介護休暇やレスパイトケアによって、介護者は睡眠時間や睡眠の質を改善できる可能性があり、公的統計マイクロデータを用いた本研究結果は、介護負荷の高い介護者に対する十分かつ質の高い睡眠時間確保の必要性（例えば、夜間対応型介護訪問介護の充実など）を提示している。この提言を導いた総務省社会生活基本調査マイクロデータには調査対象者の生活時間帯に応じた行動種類が詳細

な項目で記録されており、中でも 15 分刻みでの時間帯行動情報は、個人の社会的関係性や生活リズムを分析する上で貴重であり、総務省社会生活基本調査を自殺総合対策などの政策研究に利活用すべきである。特に、コロナ下に行われた令和 3 年社会生活基本調査と前回の平成 28 年社会生活基本調査の比較は、コロナ下で日本人の生活がどのように変化し、どのようなリスクが新たに生じているかを調べ、必要な政策提案を行うために極めて重要と考える。

## 6. 成果外部への発表

### (1) 学会誌・雑誌等における論文一覧（国際誌 2 件、国内誌 3 件）

- 1) Shiori N, Hiraku K, Satoshi S, Eiji N, Hiroyuki Y, Nao I, Yoshiki M, Alexander T S, Paul A H, Keita Y, Michiko W, Hiroaki M, Shun K. Assessment of coding-based frailty algorithms for long-term outcome prediction among older people in community settings: a cohort study from the Shizuoka Kokuho Database. *Age and Ageing* 2022; 51(3): 1-9
- 2) Seiji H, Michiko W, Yoshimitsu A. Determinant analysis and developing evaluation indicators of grade of execution score of double axel jump in figure skating. *Journal of sports sciences* 2022;40(4): 470-481
- 3) 岡檀, 久保田貴文, 椿広計, 山内慶太. 日本における COVID-19 パンデミック後の自殺率上昇の地域差及び性差に関する分析—全国市区町村の産業構造に着目して—. *統計数理* 2022; 70: (掲載予定)
- 4) 高部勲. 公的統計マイ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. *データサイエンス研究(立正大学データサイエンス研究所紀要)* 2022; 1:3-18 (掲載予定)
- 5) 高部. Synthetic Data の考え方に基づく疑似的なマイクロデータの作成方法の検討. *統計研究彙報* 2022; 79: 111-130

### (2) 学会・シンポジウム等における口頭・ポスター発表（国際学会等 2 件、国内学会等 29 件）

- 1) Shunichi N, Michiko W, Yuko O. Hierarchical Topic Model for Tensor Data and Extraction of Weekly and Daily Patterns from Activity Monitor Records. *Pacific-Asia Conference on Knowledge Discovery and Data Mining. Virtual. 2022*
- 2) Hajime O, Kazuhiro M, Hideitsu H. One-bit Submission for Locally Private Quasi-MLE: Its Asymptotic Normality and Limitations. *The 25th International Conference on Artificial Intelligence and Statistics (AISTATS). Virtual. 2022*
- 3) 岡檀. 未来を生き抜く力、見つけたい—日本で“最も”自殺の少ない町の調査から. *日本ブリーフサイコセラピー学会. オンライン. 2021*
- 4) 岡檀. COVID-19 のパンデミックによる自殺率上昇の地域差および性差の検討,市町村の産業構造に着目して. *第 45 回日本自殺予防学会 日本自殺予防公開シンポジウム. オンライン. 2021*
- 5) 南和宏. 公的統計における擬似データの構築可能性. *2021 年度統計関連学会連合大会 .オンライン. 2021*
- 6) 南和宏, 阿部穂日. 表データの最適セル秘匿処理に対するマッチング攻撃とその防御手法の検討. *2021 年度統計関連学会連合大会 .オンライン. 2021*
- 7) 高部勲. 「公的統計マイクロデータにおけるさらなる利活用をめぐる」、公的統計マイ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. *2021 年度統計関連学会連合大会 .オンライン. 2021*

- 8) 高部勲. 公的統計マイ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. 経済統計学会 第 65 回全国研究大会. オンライン. 2021
- 9) 南和宏. 差分プライベートな合成データの作成の動向. コンピュータセキュリティシンポジウム. オンライン. 2021
- 10) 岡檀. 日本における COVID-19 パンデミック後の自殺率上昇の地域差および性差, 全国市区町村の産業構造に着目した分析. 新型コロナウイルス (COVID-19) の世界的流行下における自殺予防・自死遺族支援のための学際的・共同研究集会. 統計数理研究所. オンライン. 2021
- 11) 岡檀. コロナ感染拡大後の自殺率上昇の地域差と性差—全国市区町村の産業構造に着目して—, 統計数理研究所公開シンポジウム「新型コロナウイルス関連データを解析する」. オンライン. 2021
- 12) 南和宏. 公的マイクロデータに対する k-匿名化加工の検討. 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」. オンライン. 2021
- 13) 岡檀. COVID-19 感染拡大による生活変化の把握と対策に資する質的/量的混合アプローチ. 九州大学 IMI マス・フォア・インダストリ研究所 研究集会. 2021
- 14) 岡檀, 椿広計, 山内慶太. COVID-19 感染拡大による生活変化の把握と対策に資する質的/量的混合アプローチ. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 15) 山内慶太. 自殺総合対策研究における公的統計データ活用の意義. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 16) 田上紀代美, 新井崇弘, 山内慶太, 渡辺美智子. 社会生活基本調査匿名データを用いた社会的孤立要因の検討—潜在クラス分析による単身男性の介護生活行動に潜む課題抽出. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 17) 新井崇弘, 山内慶太, 渡辺美智子. 「介護・看病疲れ」による自殺と行動時間情報の統合的研究-公的統計マイクロデータ利用による社会課題解決への橋渡し-. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 18) 赤谷俊彦. 地方の EBPM 推進に資する統計データ利活用センターの取組について. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 19) 六信孝則. 地域状況や行動の可視化に向けた RESAS、V-RESAS の紹介とその活用. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 20) 高部勲. 公的統計に関する疑似マイクロデータの提供可能性. 第 12 回横幹連合コンファレンス. オンライン. 2021
- 21) 岡檀, 久保田貴文, 椿広計, 山内慶太. コロナ禍の自殺率上昇の地域差および性差に関する分析, 市区町村の産業構造に着目して. 第 80 回日本公衆衛生学会. 京王プラザホテル/東京大学. 2021
- 22) 椿広計. 公的統計周辺の 2 つの技術的課題: 欠測値補完と母集団復元並びに残差のブートストラッピングに基づく仮想データ作成の数理的性能評価について. 令和 3 年度革新的自殺研究推進プログラム「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A)「公的統計マイ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」合同研究集会. オンライン. 2022
- 23) 山内慶太. 自殺・社会的孤立研究から見た公的統計マイクロデータ活用の意義と課題. 令和 3 年度革新的自殺研究推進プログラム「行政における統計データの利活用の推進に関する研究」, 科

- 学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 24) 高部勲. Synthetic Data の考え方に基づく疑似的な公的統計マイクロデータ作成・提供方法の検討. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 25) 山下智志. 法人マイクロ統計データの特徴と信用リスクモデル. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 26) 久保田貴文. コロナ禍における精神的不安や希死念慮に関連する Tweet データのテキスト分析. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 27) 渡辺美智子. 大規模観察データの分類技法ー潜在クラスモデルの行動計量的活用例ー. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 28) 新井崇弘. 介護者の自殺念慮と睡眠時間との関係性: 「社会生活基本調査」匿名データの利活用. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 29) 岡檀. COVID-19 感染拡大による生活変化の把握と対策に資する質的/量的研究混合アプローチ. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 30) 竹林由武. メンタルヘルスサービスの利活用状況とその阻害、促進要因. 令和 3 年度革新的自殺研究推進プログラム 「行政における統計データの利活用の推進に関する研究」, 科学研究費補助金基盤研究 (A) 「公的統計マイ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」合同研究集会. オンライン. 2022
- 31) 久保田貴文. コロナ禍における精神的不安や希死念慮に関連する Tweet データのテキスト分析. 第 106 回行動計量シンポジウム. オンライン. 2022
- (3) その他外部発表等
- 1) 岡檀. 生き心地のよいコミュニティのために社会教育ができることー日本で“最も”自殺が少ない町からの気づきー. 東京都社会教育指導員研修. 2021
- 2) 岡檀. ウィズコロナ社会でやわらかな思考を育もうー自殺希少地域と自殺多発地域の子どもコホートスタディの結果から. 生活協同組合連合会研修会. 2021

7. 引用文献・参考文献

なし

8. 特記事項

(1) 健康被害情報 なし

(2) 知的財産権の出願・登録の状況 なし