公的統計ミクロデータ研究コンソーシアム

NEWSLETTER

Vol.02

OCT.2022

INDEX

特集：公的統計ミクロデータと統計教育Ⅰ
～統計教育とSSDSEの活用の展望～ … 02

イベントレポート …… 05

今後の活動予定 …… 06

活動報告 …… 08

問合せ先 …… 09

座談会

身近な地元の自治体データで若者の統計リテラシーを向上

―SSDSE（教育用標準データセット）で地域の課題解決力を育てる―

独立行政法人統計センター 統計技術・提供部技術研究開発課
統計リテラシー向上担当
田中雅行さん

東京学芸大学 先端教育人材育成推進機構 准教授
山下雅代さん

統計センターはデータ分析のための汎用素材として、SSDSE（Standardized Statistical Data Set for Education：教育用標準データセット）を作成・提供している。日本の統計リテラシーの底上げを目指し、学生を対象としてSSDSEを用いた統計データ分析コンペティションも開催。統計センターの田中雅行さんと、教材開発に携わる東京学芸大学の山下雅代さんに、SSDSEの作成経緯や教育現場での活用方法、統計教育を広めていく上での課題について伺った。

身近な地域のデータセットSSDSEの誕生

――お二人はどのような経緯で、SSDSE（教育用標準データセット）に関わるようになったのですか？

田中雅行さん

田中　「私は大学卒業後、旧総務庁の統計センターに入庁し、国勢調査をはじめとする統計調査の実施・公表など主に公的統計の作成業務全般に従事してきました。また総務省統計研修所に在籍していた頃は、公的統計ミクロデータの利用促進に関する業務にも携わりました。

また、今年3月までの3年間は一橋大学に出向し、学部生・院生向けに『統計調査論』などの講義や、修士学生向けのゼミを担当。現在は統計センターにもどり、SSDSEの提供や利用の促進を通し、統計リテラシーの向上に取り組んでいます。」

山下雅代さん

山下　「大学で品質管理の研究室に入り、研究テーマに選んだのが統計教育でした。QC的問題解決法の知見を小中高の統計教育に適用する研究をしています。博士の学位取得後は、統計センターで4年間、消費動向指数の改善に関わる業務と、SSDSEを活用した中高生向けの教材開発に携わりました。今年4月から東京学芸大学先端教育人材育成推進機構に移り、主にデータサイエンス教育のカリキュラムの検討を行っています。」

――統計センターはどんな組織ですか？また、統計教育の位置づけは？

田中　「統計センターは、総務省統計局の調査をはじめ他省庁から委託された統計調査の集計・製表を担う組織です。近年は『e-Stat（政府統計ポータルサイト）』を含む政府統計共同利用システムの開発や運用をはじめ、統計データの利活用促進も担うようになっています。2009年に全面改正のあった

『統計法』によって二次利用制度が創設されて、政府の統計調査の匿名データの提供やオーダーメード集計を行うようとする公的統計ミクロデータサービスを始めました。そして、2019年施行の法改正によって、調査票情報（政府の統計調査の個票データ）を安全な環境で利用できるオンサイト施設のサービスも担うようになりました。さらに学生や社会人に向けた統計リテラシーの向上のための取り組みも、統計センターの大きな役割の一つとなっています。」

――それが、SSDSEの作成・公開につながったわけですね。

田中　「そうです。SSDSEの目的は、実社会の問題解決力の育成に向け、自由度の高い規定課題を全国で共有することにあります。」

山下　「2017年から政府が進めた『公的統計改革』で、統計が証拠に基づく政策立案（EBPM）を支える基盤と位置づけられたことを受け、総務省は全国各地の自治体職員や国家公務員を対象にEBPMの研修を実施しました。その研修教材として都道府県や市区町村の統計をまとめたデータベースであるSSDS（社会・人口統計体系）を用いたところ、参加者の受けがよかったようで、当時統計センターの理事長だった椿広計先生（現：統計数理研究所所長）が『学習者は自身の生活地域の姿に強い興味を持つ』という仮説を立てられました。そこでSSDSの中から社会問題の解決につながる項目を抽出し、使いやすいデータセットとして提供する『SSDSE』の検討が始まったと聞いています。」

田中　「学校教育において公的統計の果たす役割の重要性がより高まるなか、当時の椿理事長主導のもと絶妙なタイミングでスタートを切れたのではないかと思います。」

欠測値のない
すぐに使えるデータセット

――SSDSEにはどのようなデータが入っているのですか？

山下　「2018年にまず『SSDSE-A（市区町村）』の提供が始まりました。縦に全国1741の市区町村が並び、横に人口、出生数など125のデータ項目が並ぶ表形式のデータセットです。

　その後、毎年1つずつ増え、『SSDSE-B（県別推移）』『SSDSE-C（家計消費）』『SSDSE-D（社会生活）』『SSDSE-E（基本素材）』と現在は5種類。エクセル形式とCSV形式で提供しています。欠測値がなく、ダウンロード後すぐに分析に使用できます。

　私が気に入っているのは、今年提供が始まった『SSDSE-E（基本素材）』。都道府県別の多分野データで、県民所得やゴミの量、空き家件数など中高生が仮説を立てる際に知りたがるような項目が多く入っています。」

山下　「利点としては、日本全国のデータがそろっていること、欠測値がなく、収録項目に関する解説が付いていて出典が明らかにされていること、信頼性が高く安心して使えることが挙げられます。誰でも無料でダウンロードできることも特長ですね。」

田中　「SSDSEの作成で留意していることのひとつは、項目の選定です。SSDSEの元になるSSDS（社会・人口統計体系）のデータ項目は都道府県別で2000以上ありますが、その中から授業での利用やデータ分析に適した項目を選ぶようにしています。

　SSDSEの基になるSSDSには統計調査によるもの以外に業務統計のデータなども含まれているので、項目の選定の際には、調査方法やデータの作成方法にも留意した上で選定する必要があります。」

――統計教育の現場では、どのように利用されているのでしょうか。

山下　「京都の中学校3年生の授業で、私が考案した教材を活用してもらったことがあります。公民の授業で『需要と供給』について学んだ後、『京都の食料費は高いか』という導入の問いからスタートするものです。教材では『SSDSE-B（県別推移）』に収録されている2015年の食料費データを活用し、全国で食料費が違う理由を、仮説を立

https://www.nstac.go.jp/use/literacy/ssdse/

SSDSEの一部

てて検証していく流れになっています。『大都市圏で食料費が高いのは人口が多く、需要量が大きいからではないか』などの『需要と供給』に絡んだ仮説を生徒から引き出すことがねらいです。

まず、自分たちの住んでいる京都について身近な近畿地方と比較することで、興味がかき立てられたようです。中には『京都市民は大食いが多いのではないか』などのユニークな仮説も飛び出しますが、その後、全国のデータと比べることで仮説がより一般化されていきます。授業を実践下さった先生からは、『身近な京都市のデータから入ることで印象を強め、需要と供給に対する知識の定着を促すことができた』との感想をいただきました。」

――高校生や大学生らを対象に、SSDSEを使った「統計データ分析コンペティション」も毎年開催されていますね。

山下　「SSDSEを開発する際、椿先生が『データを提供するだけでなく、実践によるベストプラクティスを共有する場を設けよう』と言われ、このコンペティションを企画されました。」

田中　「今年で5回目になります。『高校生の部』と『大学生・一般の部』があり、応募者数は年々増えています。21年は『高校生の部』が55件、『大学生・一般の部』49件で合計104件の応募となり、初めて100を超えました。今年はさらに増える見込みです。」

山下　「私が担当したのは『高校生の部』の1次審査ですが、応募者は偏差値の高い高校が目につきました。分析内容も、多変量解析を用いるなど学習指導要領の内容を超えるレベルの高いものも多かったです。一方で、高校までに習う手法のみを上手に使った論文もありました。」

田中　「『大学生・一般の部』は、参加者の傾向としては国公立大学が多く、学部は経済や経営、情報系のほか、最近新設が続くデータサイエンス学部が目立ちます。内容も高度で、中には分析に機械学習を活用した論文もありました。

しかし、あくまでコンペ本来の目的は、次代を担う学生の統計データ利活用マインドと分析技術の涵養を図ることにあります。単に高度な分析手法を競うようなコンペにならないよう、審査では注意を払っています。」

コンペの募集ポスター

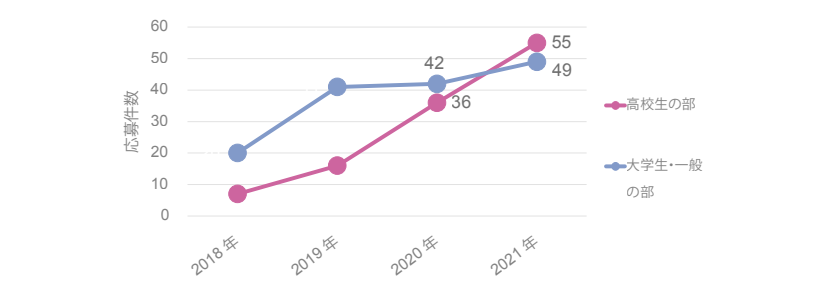
データ利用の促進と学びの事例形成を両輪で

――今後、統計教育を広めていく上で、課題はありますか？

山下　「教材が充実し、統計を学ぶ環境は整ってきました。子どもたちに統計的な探究課題を与えると、興味を持って乗ってきます。その半面、統計や分析の指導に自信を持っている教員ばかりではまだないので、そのサポートが必要だと思います。大学においてはデータサイエンス教育が強化されていますが、統計学の『文法』ともいえる問題解決プロセスの教育により焦点があたるとよいと思っています。」

田中　「SSDSE以外にも、個票データを模して自由に使うことのできる『一般用ミクロデータ』や、最初に挙げた匿名データも提供しています。調査票情報や匿名データは、利用目的や適正管理などを満たしていなければならないといった条件はありますが、ICT技術の進展に応じ、セキュリティなどに配慮しつつ運用ガイドラインの見直しや新たな提供方法などサービスの改善を継続していく必要があると思います。

データ利用を促進するためには、データを用いた分析や考察に関する『学びの事例』を生みだし、共有していく必要があります。統計データ分析コンペ



SSDSEを用いた「統計データ分析コンペティション」応募件数の推移のグラフ

以外にも、そうした事例をもっと増やしていきたいと思います。」

山下　「分析コンペは『問題の原因はここだろう』という仮説の段階で終わります。もしその先に、学生がさまざまな人と一緒に社会実装し、地域の問題を解決したという事例ができれば、『うちの学校でもやってみたい』という声が次々上がってくるのではないでしょうか。そのためには、地域のコーディネーターを中心に仕事で分析の経験を持つ人と学校と自治体職員の四者で協働していければと思います。」

――将来に向けての抱負をお願いします。

山下　「今、私が力を注ぎたいのは小学校のデータサイエンス教育です。子どもの頃から、データを使って身近な生活をより良く改善していける問題解決の力を育てることが大切。統計は日常生活でも役立つ楽しいものです。『面白い！』と思ってもらえるような講義や教材を目指して、研究活動を進めたいと思います。」

田中　「SSDSEに関しては、関係者のこれまでの努力もあり、データの整備とそれを活用した分析コンペが両輪となり、普及を促進してきました。今後も既存のデータの見直しや新たなデータの追加など、ユーザーの使いやすい、時代に合ったデータの提供を心掛け、SSDSEのさらなる普及を図っていきたいと思っています。」

イベントレポート

新学習指導要領とSTEAM教育が創り出す「社会に活かす統計の考え方」シンポジウム

2022年6月16日（木）に、統計数理研究所オープンハウスの連携イベントとして「新学習指導要領とSTEAM教育が創り出す『社会に活かす統計の考え方』シンポジウム」がオンラインで開催されました。STEAMとは「Science」「Technology」「Engineering」「Art」「Mathematics」の5分野の頭文字を取ったものであり、教科横断の重要性を指摘するものです。本シンポジウムは、STEAM教育とその方法論のひとつである数理的・統計的問題解決プロセスによる「社会に活かす統計の考え方」をテーマとしたものになっています。

オープニングでは、統計数理研究所の椿広計所長によるオリエンテーションが行われ、問題解決プロセス教育の歴史とその考え方、現在の高等学校・大学教育においてどのようなことが起きているのかなどについて解説しました。続けて、電気通信大学の鈴木和幸氏から「問題解決プロセスとAI・データサイエンス」、東京学芸大学の西村圭一氏から「統計教育の社会実装に向けて」、独立行政法人統計センターの槙田直木氏から「社会に活かす公的統計」、東京学芸大学の山下雅代氏から「問題解決すごろくと身近な例」の演題で講演があり、問題解決プロセス教育とSTEAM教育の考え方や教育・産業における実践事例、教育利用可能な統計データなどについて紹介がありました。最後にパネル討論が行われ、数理・データサイエンス的問題解決力教育の目的と方法、初中等教育における問題解決力教育を巡る問題と解決の方向性などについて意見が交わされました。

なお、本イベントの模様はアーカイブ配信されており、下記ウェブサイトよりご覧いただけます。

https://www.ism.ac.jp/openhouse/2022/pre_event.html

問題解決すごろく

シンポジウムイメージ

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

2022年11月25日（金）に「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」をオンラインで開催いたします。

本シンポジウムは、公的統計ミクロデータを学術研究に利用する大学・研究所、総務省統計局、統計センターなどの関係機関が連携して毎年開催しているシンポジウムです。

午前の部では「オンサイト施設運営に関するチュートリアル」と題して、総務省統計局や統計センターによるオンサイト施設開設のノウハウや運営支援の取組、またオンサイト施設の設置担当者による体験談などのチュートリアル講演を実施します。またオンサイト施設及びオンサイト利用全般に関する Q&A の時間も設ける予定です。

また午後の部では「わが国のデータインフラ整備、データベース化」「統計をめぐる諸課題に関する研究」と題して、２つのセッションを実施し、さまざまな角度からの報告を予定しています。

本シンポジウムは事前申し込み制（無料）のシンポジウムとなっており、申し込み方法は10月初旬に公的統計ミクロデータ研究コンソーシアムのウェブサイト（http://jmodc.org）にてお知らせいたします。ぜひご参加ください。

日　　時　2022年11月25日（金）10時～16時40分（予定）
開催方式　オンライン開催（ZOOM ウェビナー）

プログラム（予定）

【オンサイト施設運営に関するチュートリアル】

- 「オンサイト施設の設置・運営支援の取組」
  田中 久睦（総務省統計局・統計データ利活用センター）
- 「オンサイト開設の手続きのノウハウ」
  伊藤 直人（（独）統計センター）
- 「公的ミクロデータの利活用に関する事例紹介」
  寒河江 雅彦（金沢大学）
  原田 魁成（金沢大学）
- 「オンサイト施設の設置、運用、将来展望
  ー施設管理者と利用者の視点からー」
  中村 健太（神戸大学）
  勇上 和史（神戸大学）
- 「オンサイト施設・オンサイト利用に関する Q&A」

【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】

挨拶：藤井 良一（情報・システム研究機構）

わが国のデータインフラ整備、データベース化

- 「公的統計におけるデータ整備」
  長藤 洋明（総務省統計局）
- 「データ共有におけるメタデータの役割：日本学術振興会の人文学・社会科学総合データカタログ（JDCat）を事例として」
  前田 幸男（日本学術振興会研究員・東京大学）
- 「東大社研 SSJDA におけるデータキュレーションとメタデータ整備の実際」
  三輪 哲（東京大学）

統計をめぐる諸課題に関する研究

- 「経済のグローバル化・デジタル化の統計的把握」
  萩野 覚（総務省統計委員会担当室）
- 「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」
  高部 勲（立正大学）
- 「事業所生命表の試算」
  菅 幹雄（法政大学）

閉会挨拶：椿 広計（統計数理研究所）

※講演者や講演題目は変更になる場合もございます。ご了承ください。

公的統計ミクロデータ研究コンソーシアムワークショップの開催について

企画責任者ご挨拶

立正大学データサイエンス学部
高部 勲 教授

立正大学データサイエンス学部の高部と申します。公的統計ミクロデータ研究コンソーシアムの運営委員を務めております。2021年4月から立正大学に着任しておりますが、大学の教員となる前には、公務員として、総務省統計局において、統計調査の実施などに関する業務に携わっておりました。

具体的には、統計調査を実施する際の調査票・調査用品の作成や予算の要求、体制の整備などの調査の企画立案・実施に関する業務のほか、公的統計の作成・公表、企業に関するデータベースの整備などの、正確なデータを作ることに関する、公的統計のメーカーとしての様々な業務に携わってきました。一方で、新たな経済指標の開発や公的統計データの周知とその利活用の推進などの、公的統計データのユーザーとして関わる業務にも携わってきました。こうした業務を行っていく中で、公的統計ミクロデータを使った分析・研究の重要性や、公的統計ミクロデータの持つ魅力・可能性を実感してきました。

公的統計ミクロデータの概要

昨今、多くの分野で、ビッグデータを始めとする様々なデータの利活用が進んでいるところですが、こうしたデータは、対象範囲や項目などに偏り（データバイアス）が含まれることもあり、分析に用いるデータに偏りがないかといった観点からの慎重な確認は、人間による判断を要する重要な作業となっています。このようなデータのバイアスに関する問題に対して、データが作成される現場の各段階において、厳重な品質管理の下、バイアスの発生を抑えるという努力を地道に行ってきたのが公的統計です。公的統計では、調査票や調査用品等の設計段階での入念なチェック、調査員による確認、プログラムによるチェック、地方自治体や国の職員による審査が行われており、調査の企画設計から公表に至るまでの各段階で、バイアスを取り除くための様々なチェックが行われ、多くの労力を費やして作成された品質の高い公的統計ミクロデータは、母集団の特性を把握する際に非常に有用なデータとなっています。

ワークショップ概要

公的統計ミクロデータは、利用件数が増加してきておりますが、まだこれらのデータに馴染みの無い方も多くいらっしゃると思います。今回企画しているワークショップでは、公的統計ミクロデータを基に作成した疑似的なデータを活用して、その利用方法や主な分析方法などについて御紹介することを予定しています。また合わせて、実際のミクロデータのデータレイアウトの詳細や、分析を行う前の段階の、データの成型などの前処理の方法などについても紹介していく予定です。ワークショップを通じて、公的統計ミクロデータの持つ魅力や重要性をお伝えしたいと考えております。一人でも多くの方々に公的統計ミクロデータの利活用に関心を持っていただければ幸いです。

公的統計ミクロデータ研究コンソーシアムワークショップは、2023年3月に対面実習を予定しております。詳細が決まりましたら、コンソーシアムウェブサイト（**http://jmodc.org/**）でお知らせいたします。

Report 1 公的統計ミクロデータ研究コンソーシアム　アンケート　実施

本コンソーシアムでは、公的統計ミクロデータを利用した研究に関心のある研究者を対象に、コンソーシアム会員向け・一般向けの２回に分け、公的ミクロデータに対する利用意向、利用状況に関するアンケートを実施しました。その集計結果をコンソーシアムのウェブサイトで公開しました。

ご回答いただきました貴重なご意見は、本コンソーシアムの活動への反映の他、関係省庁にも共有し、今後のオンサイト利用施設の整備等にも活用予定です。

調査方法

●実施期間：

会員向け：2020年10月６日（火）～11月１日（日）

一般向け：2021年５月11日（火）～６月11日（金）

●実施対象：

会員向け：公的統計ミクロデータ研究コンソーシアム個人会員（73名）

一般向け：公的統計ミクロデータを利用した研究に関心のある研究者

●アンケート実施方法

Google フォーム回答によるオンラインアンケート

●実施広報

コンソーシアム公式ウェブサイトからの告知

会員向けメーリングリストによる依頼

関連学会へのメーリングリストへの案内送付

●質問内容

公的ミクロデータの利用状況

オンサイト利用施設の利用意向

オンサイト利用の経験

コンソーシアムへの要望

●回答数

会員向け：24名（回答率：約33%）

一般向け：26名

調査結果

アンケート結果サマリー：http://jmodc.org/questionnaire/gaiyou.html

会員向けアンケート結果詳細：http://jmodc.org/questionnaire/member.html

一般向けアンケート結果詳細：http://jmodc.org/questionnaire/nonmember.html

Report 2 統計関連学会連合大会企画セッション

本年度の統計関連学会連合大会が、2022年９月４日（日）～８日（木）に成蹊大学にてハイブリット形式で開催されました。本コンソーシアムでは「公的統計ミクロデータ利活用の現状と課題」と題した企画セッションを実施し、各報告とも更なる公的統計ミクロデータの利活用につながる活発な議論が展開されました。

https://confit.atlas.jp/guide/event/jfssa2022/session/5A06-10/tables?QbWVQfTQhi

統計関連学会連合大会企画セッション

「公的統計ミクロデータ利活用の現状と課題」

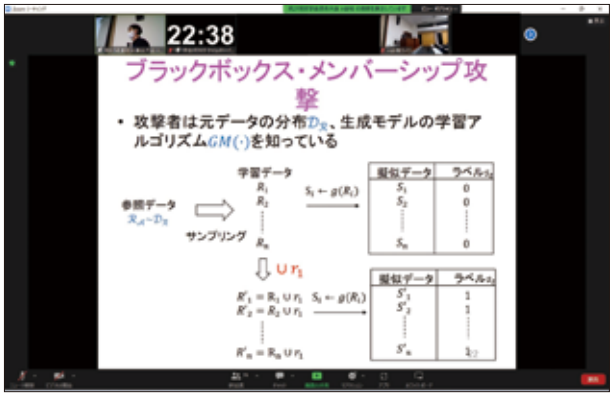
日　　時　2022年９月８日（木）13:00～15:00

参加人数　会場（成蹊大学）：20名　オンライン：83名

プログラム

座長　山下智志（統計数理研究所）

- 「公的統計における擬似データのプライバシー保護と安全性評価」
  南 和宏（統計数理研究所）
- 「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」
  高部 勲（立正大学）
- 「賃金所得の統計的分布を考慮した消費構造の計量分析―公的統計のミクロデータを用いて―」
  伊藤 伸介（中央大学）
  出島 敬久（上智大学）
  村田 磨理（統計情報研究開発センター）
- 「公的統計匿名データを利用した統計的因果推論」
  高橋 将宜（長崎大学）
- 「厚生労働省におけるEBPMとデータ利活用の取組について」
  戸田 淳仁（厚生労働省）



問合せ先

ADDRESS

〒190-8562　東京都立川市緑町10-3
大学共同利用機関法人　情報・システム研究機構
データサイエンス共同利用基盤施設　社会データ構造化センター内
公的統計ミクロデータ研究コンソーシアム　事務局

MAIL　office@jmodc.org
URL　http://jmodc.org/