

公的統計マイクロデータ研究コンソーシアム第18回運営委員会
議事録

- 1 日時 令和4年2月14日(月) 15時00分～17時00分
- 2 場所 オンライン会議(ZOOM)
- 3 出席者 (運営委員) 南委員長、伊藤副委員長、赤谷委員、稲垣委員、高部委員
中村委員、山下委員、岡本委員

4 概要

(1) 前回の議事録の確認

第16回運営委員会(オンライン)の議事録について、資料3に沿って概略を説明・確認した。

(2) 第1号議案：コンソーシアム会員の認定について

新規の入会申込者について、資料9に沿って審議し、了承した。

(3) 第2号議案：匿名化議事録の一般公開について

現在会員むけにコンソーシアムウェブサイト上で公開している議事録の一般向け公開について、第9回評議会議事録に基づき審議の背景を説明した。また資料15のバックナンバーサイトデザイン(案)と会員向けに公開しているページに基づき一般公開の方法について議論した。

(主な意見)

- 一般公開時には、トップページからリンクを貼り、ウェブサイト内の「お知らせ」の項目でも案内することを考えている。
- 議事録公開の際には、発言者の氏名は公開されるのか。また資料は全て公開するのか。
→ 会員向けと同様に発言者の氏名は公開されない。
→ 資料は一部公開する予定。
- 実際に一般公開する前には、確認をするため事前に連絡をして欲しい。
→ 現在公開している会員向けのページを、そのまま一般向けに公開するという形になるため、確認の上 問題点等があったら、事務局へ連絡して欲しい。
- 運営委員会で議論する内容は、企画セッションの検討内容のように、一般向けに発信する必要がないものと、運営委員会の中で発言者の名前を残した形で記録に残しておく必要があるものがあると思われる。公開用の議事録と名前を残した形で運営委員が共有できる議事録との2種類があると良いのではないか。
→ 現在は、名前ありの議事録と匿名化した公開用の議事録の2つを作成している。
→ 氏名を記載しない匿名化だけでは不十分だと思われるので、公開する内容を検討し、
 - ① 運営委員会用に議事録を作成
 - ② 一般公開用に匿名化し不要な部分は除いた議事録を作成
 - ③ 公開する上で問題がないか運営委員全体に確認をとる
 - ④ 公開 という手続きで進めていく。
- 基本的には第1回からの運営委員会の議事録を全て載せるという認識でよいか。

→ 過去の運営委員会に関しては、一部記録が残っていない回があり、その分は公開できていない。また、評議会の議事録についても公開している。

- 年度内を目安に運営委員が公開する議事録を確認し、その後何かあれば次回の運営委員会で議論することとする。

(4) 第3号議案：公的マイクロデータワークショップ(仮)の開催企画について

活動計画の一つである公的マイクロデータワークショップについて、資料8に沿って企画案を説明し議論した。

(主な意見)

公的マイクロデータに興味はあるが実際の利用に至っていない潜在的なユーザーの掘り出しを目的とし、データの研究利用に興味のある大学院生、研究者を対象とした1日でのワークショップの実施を検討している。こちらのワークショップでは、オンサイトの環境に近いデータ環境を準備し、データ分析のノウハウを提供できるようにしたい。

具体的には、午前中は講義形式としてマイクロデータの生成プロセスなどの「公的マイクロデータの概要」、「主な公的マイクロデータの内容」、オンサイトの利用申請・持出し申請などの「マイクロデータの2次利用の制度について」等を計画している。また午後はR言語などを使い、CSVの読み込みからデータ分析、最終的にはデータの持出しのための秘匿処理までの一連の流れを、分析実習という形でワークショップを行う案を考えている。

<全体の構成について>

- ミクロデータの2次的利用の制度とマイクロデータの生成プロセスで1つのセクションを置き、その次に2次的利用の手続き(申請)関係について話をするという流れで行うと良いのではないか。
 - 大学院生などを対象とする場合には制度の話は固い印象があるので、利用申請の話の中で参考までに制度の話をするという程度が良いのではないか。
 - 大学院生を主な対象とするならば、2次的利用の申出者になり得ないので、制度というより具体的なデータの扱い方に重点を置いた方が良いと思う。
- ワークショップを今後継続的に行っていく場合、まずは入門的なことを実施し、それを聞いた参加者が次のワークショップで応用的なことを行う流れになるかと思う。応用的な話の中では、統計表をみただけでは分からないような調査票の癖のようなものを話したり、各先生が行っているデータ分析の加工や工夫を話したりというのでも検討に値すると思う。
 - 対象やレベルも色々あるので、まずはベーシックなものを行い、その後発展させていくというのは、今後の活動として十分あり得ると思う。
 - 調査票の癖を話すとは？
 - 利用者が使い始めた時に疑問に思う点（「欠損値に対してどのように対応しているのか」や、「データセットの項目と調査票情報の項目が対応していない」など）を担当部局の方が話をする、またどの時期によって調査したかによって回答にも癖があるのでそれについて話してもらうなど、やり方は色々あると思う。

- とある勉強会の中では、どのような変数やレンジになっているかというデータの中身に対して参加者は興味があるようだった。初回のワークショップでは、過去の事例を説明するというよりも、まずは「そもそも公的統計とは何か」という概要とデータの中身について話し、興味を持った参加者がいれば統計センターの公開している研究成果の案内をするという流れにすると良いのではないか。
 - 別のセミナーでは、「公的マイクロデータは他のデータと違い、簡単には利用できない」という意見があった。そういった方々を対象とするならば、データの特徴をまずは説明し、その後一般用マイクロデータや疑似データで理解してもらうという流れというの
は良いと思う。

<大学院生をワークショップの対象とする場合の検討>

- 博士課程の大学院生は二次的利用の申出者になることはできないのか。
 - ケースバイケースだと思う。
 - 通常の大学院生が研究で使いたいというだけでは公的マイクロデータの利用は難しいが、大学にリサーチアシスタント等で雇用されていて研究者協力者として名前が載っていれば利用が可能な場合もある。ケースバイケース。
 - 現行法では、研究分担者と研究協力者は区別をしないと考えるか。
 - 都度判断になると思う。
 - 磁気媒体かオンサイト利用かというのは分けているか。
 - 磁気媒体かオンサイト施設かという利用形態は利用者要件とは関係なく、寧ろ具体的な集計様式をどこまで出す、事前審査型か持出し審査型かといった事が関わってくる。
- 大学院生の人でもどのような人がデータを使えるかというような Q&A を講義の中で行うと良いと思う。
 - 研究内容やプロジェクトか、申請している調査の種類、担当部局がどう判断するのかなど、その時々によって状況がかなり変わるため、ワークショップで解説するのは難しいのではないか。
 - 事例の紹介くらいになると思う。
- 大学院生を対象とする場合「大学院生が自分の研究を行うための分析手法を学ぶ」または「授業で一般用マイクロデータや疑似データを使う場合、どのように授業を組み立てれば疑似データを使って効果的に授業ができるというようなモデルケースを提示する」という2つの切り口の考えられるのではないか。
- 大学院生を対象とする場合、学んだ内容を発揮する場所が限られてくるため、対象を再度検討する必要がある。個票データを扱える若手研究者や教育に活用したい大学教員を対象とするのが良いか。
 - ワークショップの対象者に加え、これを学ぶと何に役立つのかという目的を明確にする必要があると考える。若手研究者や大学教員の場合、午前中の分析事例が重要になってくる。

<分析事例の紹介について>

- 分析事例の紹介については、研究課題と利用データを結びつける事例があれば、潜在的

な需要が掘り起こせるのではないか。

- 事例紹介についてはまずは分野を特化する必要があるのではないか。また分析した方に話をしてもらおうか、それとも統計局などの調査内容に詳しい方にデータの癖を話してもらおうのかによって全体の構成が変わってくると思う。やはり目的をどこに設定するかが重要。
- miripo 内に載っている研究に対し、どのような調査票情報を使っているのか、全体の傾向を提示し、それを受けて一つの調査を深く掘り下げると良いのではないか。
→ 活用事例まで追っていないので難しいと考える。

<使用するデータについて>

- 使用するデータとしては、公開済みの全国消費実態調査（全消）の一般用マイクロデータで十分であって、あえて疑似データを作り直す必要はないかと思う。
→ 現在作成途中の疑似データを使う事もできる。
- 「このような統計調査が存在している」「そのマイクロデータが使えるようになっている」「符号表などでこのような調査事項になっている」というような、マイクロデータの持ち方が分かるだけでも参加者にとっては良い経験になるのではないかと考える。
- 一般用マイクロデータを使うことによって、リアルなマイクロデータを使う時の追体験のようなものはできるか。
→ 追体験くらいならばできるかと思う。
- データクリーニングの部分と解析の方法とを分けて行うのも良い。クリーニングの部分はシンプルな例で特徴を教える形になると思う。実際にワークショップを行うのは、総務省の職員というよりも実際にクリーニングの経験がある研究者などが良いと思う。
- 現在作成しているという疑似データを使用する場合は、ワークショップに間に合いそうか。
→ ギリギリになりそうだが、恐らく間に合うと思う。ただし利用にあたっては各方面へ調整が必要。間に合わない場合は一般用マイクロデータで行う形になると思う。
→ この疑似データは教育目的と理解して良いか。
→ 教育目的に加えて、研究者向けにはプログラムのテストおよびデータのイメージをつかんでもらうことを目的としている。
- ゆくゆくはその疑似データを、コンソーシアムのホームページへアップし、研究者が気軽に使ってもらえるようになると良い。
→ ワークショップで使うことは問題ないと思うが、実際にホームページで公開する場合には色々な立場があるので、疑似データをそのまま公開するのはやはり問題があると思う。
→ (公開されている)全消の一般用マイクロデータは疑似データではないのか。
→ 全消の一般用マイクロデータは疑似データではあるが、元となる集計表を総務省統計局側が公開している。作成前の集計表がオープンになっており、そこの手続きがきちん

とできているならば疑似データだったとしても公開が可能となる。

- 既に公開されているデータを元に作った疑似データは、とりあえずは問題がないが、どこが公開するかというのは議論が必要になってくる。
- 制度論というよりも官側としてどう考えるかにもよるかと思う。整理をして検討が必要かと思う。
- 次回の運営委員会で報告して欲しい。

<ワークショップ開催時期について>

- 時期的について7月下旬ということだったが、官側の人事異動もあるので考慮した方が良い。
 - 9月の連合大会との関係も重要となる。
 - 統数研のリスクセンターでは、7月上旬に金融シンポ、8月の下旬にリスクシンポを開催予定。事務局担当者が被るので配慮が必要だと思う。
 - 事務局で個別に関係者と調整する。
- ワークショップの時期は、11月のシンポジウムの前日に実施するというのはどうか。例年のシンポジウムの午前中はチュートリアルを行っているので、そこをワークショップにするというのはどうか。
 - やはり開催を5月にするのは厳しい。午前中のチュートリアルを発展させた形として11月のシンポジウムと組み合わせてワークショップを開催するというのは有望な案だと思う。
 - 11月ならば先に取り上げた疑似データも完成されている頃ではないか。
 - 開催時期については、メールなどで適宜相談していく。

(5) 意見交換

- ✓ NewsLetter 発行とチュートリアル動画公開（報告）

NewsLetter 発行とチュートリアル動画公開について報告があった。NewsLetter 創刊号は2月末から3月上旬にPDF発行予定。シンポジウムのチュートリアル動画は統数研のYouTubeにアップし、コンソーシアムのホームページを介して公開予定。

- ✓ 統計関連学会連合大会企画セッションについて

今年度コンソーシアムで企画セッションを開催したが、同様に来年度も開催したいと考えている。そのテーマについて議論したい。

- アジアのマイクロデータの企画セッションを実施しているが来年はどうか。国際マイクロとの兼ね合いもあると考える。
 - まだ決まっていないが、国際マイクロ側としては企画が組めれば良いと考えている。
- もしワークショップが連合大会よりも前に開催されるのであれば、ワークショップでデータの特徴などの入門的な内容をおこない、連合大会ではそれを踏まえてマイクロデータを使う上での注意点や加工や工夫をテーマということも考えられる。また疑似データの話を行う場合であれば、疑似データの加工の方法論や、実際に疑似データを公開するよ

うな方向性があるならば、利活用センターなどにその進捗などを話してもらおうという方法もある。

- オンサイトの利活用推進という側面を考えた場合、リンケージというのもテーマになりうると考える。
- 公的マイクロデータを使った研究発表というよりも、実際に公的マイクロデータを使ってみて苦労した点や分析上の工夫、調査票情報の癖やその対処方法など、利用に関するノウハウなどを盛り込むと良いのではないか。
 - 基本的には事象分析の報告になるかと思う。その事象分析のプロセスの中に、分析上の工夫などを話すといった内容が、企画セッションとして集まれば良いと考える。
- 昨年の企画セッションの延長線上の報告というのもあると考える。
- マッチングをした時の話なども報告できるかと思う。
 - マッチングを研究されている方は他にもいるのか。
 - マッチングそのものを研究されている方はいないが、関連した研究をされている方は4-5人いる。
 - マッチングは重要なテーマだと思うので、登壇者があつまるといったところならそこにフォーカスするのも良いかと思う。また相談させて欲しい。
- 現在全消のデータを扱って分析しているが、それにプラスして他のデータとの掛け合わせという内容について報告することも可能。
- 現在研究を進めているデータの構造化についての提言などについても、テーマとして入れたいと考えている。
- 報告のメンバーは運営委員と考えて良いか。
 - 決定はしていないが、基本的にはここにいる運営委員会で報告を行うことを考えている。ただしいつも同じメンバーになってしまうので、例えばマッチング等であまり登壇されていない方をお願いするのも良いかと思う。
 - 可能性としては、運営委員数名からの報告と、それに加え 1-2 本の報告内容があれば、セッションとして成立すると考えて良いか。
 - そのように考えている。残りの 1-2 本については、例えばマッチングや何か興味深い分析をされている方などがいれば、セッションが組めるかと考えている。
 - 厚生労働省でマッチングしたデータの分析をされている方がいる。恐らく日本統計学会の会員だと思うので、候補の一人になるのではないか。
 - 詳細は別途相談したい。

✓ データサイエンス系の学部との連携について

前回の評議会で、データサイエンス系のコンソーシアムとの連携について提言があった。運営委員会としてどのような可能性があるか議論していきたい。

- 「数理・データサイエンス教育強化拠点コンソーシアム」が念頭におかれていると思う。こちらはデータサイエンスの標準カリキュラムや教材の作成といったことがメインのコ

ンソーシアムなので、教育の題材の一つとして公的マイクロデータを使ってもらおうというのは、一つの方向性だと考える。

→ 「数理・データサイエンス～」は6大学連携が名前を変えたもので、統計というよりも情報が中心のコンソーシアム。少し我々とカラーが違う印象を持っている。まとまってはいるが動きとしては、各大学別の個別となっているようだ。この中でどの大学と組むか特定して動いた方が良く考える。

→ AI 戦略やイノベーションが強いイメージ。

→ 本格的に提携を行うようであれば、東大の駒木文保先生、元統数研所長の北川源四郎先生などがキーパーソンになるかと思う。統数研の統計エキスパート人材育成と比較され、あまり関りが無い感じではある。

→ 今すぐ一緒に何かをやるというのは難しいと思うので、まずはキーパーソンとなりそうな先生にコンタクトを取って話を伺い、様子を見るという形で進めていく。本件は継続審議とする。

✓ 次回の運営委員会の開催について

次回は5月をめどに開催を予定していたが、ワークショップや企画セッションについては、その前にメールなどで適宜相談していく。