

公的統計ミクロデータ研究コンソーシアム第 29 回運営委員会
議事録

- 1 日時 令和 6 年 2 月 20 日 (火) 16 時 30 分～18 時 00 分
- 2 場所 オンライン会議 (ZOOM ミーティング)
- 3 出席者 (運営委員) 南委員長、伊藤副委員長、伊原委員、岡本委員、小松委員、佐藤委員、高部委員、田中委員、山形委員、山下委員
- 4 概要

(1) 第 27 回議事録の確認

第 27 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し、再確認した。また追加で気がついた点がある場合の連絡を依頼した。

(2) 第 1 号議案：公的ミクロデータのためのチュートリアル・講習会について

資料 4 に基づいて説明し、前回の内容をどのように拡張もしくは変更して行うべきか意見を募った。

(主な意見)

- チュートリアルのアンケート結果は「概ね満足」が多かったが、「前提知識が必要で難しかった」、逆に「非常に専門的な内容を期待したので、その期待に沿うものではなかった」などの厳しい意見もあった。
- 主催側の事前準備が不充分であったことと、本来初級者が来ることを想定して初步的な内容を組んでいたが、蓋を開けると教育者側の方の参加もあり、参加者の実力にかなりの差がつってしまったことが反省点。今後は初級者を対象にした方が良いと思うが、教育側の視点に立っての講習にして、参加して意見をぶつけていただくのが良いかとも思われる。初級者と教育者の両者を相手にしつつ準備が不足したのが厳しい評価に繋がった。
- 演習で使用するデータをどうするか。変数が揃っているようなデータさえあれば、データベースの使い方・R の使い方・レコード対データの集計の仕方は教えられるので、より簡易な形で多くのミクロデータを作り、それを使って、R・dplyr の講習会・データレイアウトの読み方の講習会にした方が初級者も参加しやすいと思う。計量分析まで教えずとも、データのさわりだけにして、かつ不特定多数の方が参加できるようにすると、たくさんの方にご覧いただけるのではないか。
- アンケート時にしっかりと講習の内容を伝えて募集することが重要。リアル開催であれば会場で困っている人を講師が助けることができるが、まずデータに触れてもらう、という事であれば、オンラインで開催して、実際に自分でも実行していく形式にすると参加者を増やすことができてよいのではないか。初級者を対象にした講習のフィードバックがあれば、「これを踏まえて教育してみたら、このような教育モデルができた」というような実践例を教育者側に提供することも考えられる。

- 初級者と上級者が一緒に受講することになってしまったので、2つに分けて開催する工夫が必要かもしれないが、分けることによって参加者が減ってしまう問題がある。
- ハイブリッドも選択肢の1つなのでは。話を聞きたいだけの初級者と、ある程度分かっていて会場でサポートを受けたい人の両方に対応する。教育者向けとデータ利用者向けのどちらにフォーカスするかも難しいと感じた。
- ハイブリッドで手間がかかるようだったら Zoom 講習会を録画して後日公開すれば手間がからず出来るのでは。Zoom の参加者や録画の視聴者に意見を寄せてもらい、その上で教育者向けに第2弾を開催する方法もある。前述の通り、前回欲張りすぎた部分はバッサリ切って、公的ミクロデータに親しんでもらうという意味で、データレイアウトフォームを読み込んでみて集計する、というようなことに主眼を置いた方が良いのでは。そのための一般用ミクロデータのような簡単なものを作って、ラインアップを揃えて出したらどうか。
- 形式だけ揃っているようなデータでも教育の最初の段階では充分かもしれない。
- 色々な可能性がある。一度 Zoom でレクチャーを実施して、それを録画して、あとはオンデマンドで見てもらうなど。ゆくゆくは簡単なチュートリアルのページを整備しても良いのでは。
- 主催者に過度な負担がかかるないよう、地味に長く続けた方が良い。負担が大きいと息切れして2~3年で出来なくなってしまう。マンパワーから見て実現可能かどうか。ハイブリッドは講演者の負担が大きいと思う。
- リアル参加者とオンライン参加者が期待する教育の密度は全然違うと思う。その両方に答えられる講義というのは難しい。リアルで丁寧に講習会を行い、録画して、編集を入れて教材にすると良いと思う。
- 講師があらかじめ決まっているのであれば、受講者が出来るだけ情報を入手して、(受講者自身の)レベル感を調整しなければならない。こちら側が情報を提供しているのであれば、講義レベルの違和感については受講者側の責任ではないか。
- R は世の中で市民権を失っており、Python で天下統一されつつある。R をやり続けると、これから先にユーザーを増やすのは辛いように思う。統計学にどっぷりつかっていると、R が標準言語のような気がしているが、世間では全然そうではない。
- 前回の反省点としては、我々の周知の仕方も曖昧だった部分があり、初級者と上級者が受講者として共存するミスマッチがあった。
- 来年度の開催についてはまだ決まっていないが、講師に負担のない範囲で時期の相談をしたい。
- 最近の高校生の情報の教科書や指導要領も Python が基本的なので、言語は変えていかなくてはと考えている。
- 今後は言語の選択も含めて、データの準備状況を見ながら継続して議論させていただきた

い。ちなみにオンラインでは Python は使えるのか？R を使う人とどちらが多いのか？

→Python は使える。

→ログを確認したところ、R→45%・Python→20%・Stata→35%での使用率で、R が多い。

(3) 第 2 号議案：来年度の統計関連学会連合大会の企画セッション案について

資料 5 に沿って説明し、意見を求めた。

オーガナイザー、座長を決定し、それぞれの講演候補者へ依頼をすることで、了承された。

(主な意見)

- 「調査票情報の提供の動向」に関しては、開催時までにアップデートした状況で説明できると思う。
- 研究のアウトラインになってしまふかも知れないが、異なる組織が整備しているオンラインシステム間のデータベース互換性が無い中での相互利用について話すことができるかも。と考えている
- 「わが国の公的統計における匿名データの作成可能性に関する実証研究 仮」の講演について、今取り組んでいる匿名化に関する共同研究の報告をこのタイミングで出来ればと思っている。
- 「海外における公的大規模データのリンクエージの動向 仮」の講演については海外調査を踏まえた内容だが、ここでいう「リンクエージ」は研究者側でなく統計作成部局で、公的機関側がリンクエージをどのようにしているかの現状報告できればと思っている。
- 概ねこの内容で企画を進める。

(4) 報告事項 1：第 15 回評議会報告

資料 6 に沿って説明した。評議会では「統計データ利活用センターと本コンソーシアムとの関係性がわかりづらい」との指摘があり、本件について意見を求めた。

(主な意見)

- 評議会では、「利活用センターが主にオンライン施設側の推進をしていて、コンソーシアム側は利用者側の推進や利用ノウハウなどの啓蒙活動をしている、また共有部分もある」と回答した。
- 利活用センターというよりは、総務省として連携している意味合いが強いと感じている。
- 数理社会学会や東大の社研と連携を進めているが、①「他の分野ともいろいろ連携できる可能性があるのではないか」という意見や、②「どのようなデータを使うことができ、それが自分の希望とする政策立案研究に使えるのか」ということが分かりにくく、自分の研究のニーズとデータのマッチングが難しい」という意見もいただいた。

→ ①については、社会学系以外のへ展開が可能かどうかは、今後の議論になると考へて いる。

→ ②の解決策については、オンサイトを利用するためにはデータを使う必然性が問わ れる、など利用要件がかなり厳しいが、登録制によるデータの内容を事前に見ることができ るような機会やイベントがあれば、非常に助けになるのではないか、という意見をいた だいた。そのような意見を伺って、どういうデータが入っているのかを見られるよう、デー タ見学会があっても良いと思った。

- 評議会では、コンソーシアムウェブサイトについて、単なる情報発信に留まらず、実際にど ういう人に見られているのか、リファラログ等を流動解析するべきとの有益な意見をいた だいたので、対応したい。シンポジウムは現状オンライン開催だが、今後は開催形態も多様 になると思われるので、録画のオンデマンド公開を検討すべき、との意見をいただいた。ま た、必要とするデータが公的統計で公開されているのか・非公開なのか・現状非公開なの か、その基準がどこにあるのか、という質問もいただいた。

→ 統計によってオンサイトへの登録やミクロデータの公開がされていない理由につい ては、単純に作業上の問題が多いと思われる。省庁によってはマンパワーの不足などで、基幹 統計より後回しにされたり、またデータを作成してもユーザーがいなさそうな統計もある、 といった事情もあるので、出来る範囲で優先順位をつけて対処しているのが現状である。 もしデータを利用したい統計が有るのであれば、「こういうデータが欲しい」などの声が上 がってくると優先順位を上げてくれるものもあるのではないか、と思われる。

→ 公開・非公開を厳格な基準で決めているのではなく、純粹にマンパワーやプライオリ ティの問題であると良くわかった。要望を上げればプライオリティが上がる可能性もある ことも承知した。

→ 制度の改善が進んできており、公開済みのデータも 50→95 統計と倍近くに着実に増 えていることを補足する。

(5) 報告事項 2：調査票情報提供に関するガイドライン改訂の周知について

資料 7 に沿って説明し、意見を求めた。

(主な意見)

- 「調査票情報の提供に関するガイドライン」の改定をコンソーシアム会員に周知してもら いたい、という依頼があり、会員向けに情報展開した。ただ、今回の改正ポイントが一目で はわかりづらいので、どこかに明記されているのであれば伺いたい。今後より会員に分か りやすく改定内容を伝えるには、コンソーシアムのニュースレターに掲載し、説明記事も 入れたらよいかと思っている。

→ 今回の改訂は皆様にとって良いことのはずなので、広めていただけるのは非常にあり がたい。(改訂内容の説明については) 遅くとも 9 月 (統計関連学会連合大会) まで待って

いただければその時に説明できるかと。

→ ガイドラインの後半に別紙で審査表が出ているが、他省庁もこちらの審査表に基づいて審査する、という事が今回の改正の大きなポイントになるのか？ 審査をいかに短くするか、という事が主眼にあると思われるが、統計法 33 条に基づき調査票情報を磁気媒体提供という形で申請した場合、利用者側にとっては手続き的に楽になるかどうか、が一番の関心事だと思われる。今回のガイドライン改正によってどのように楽になるか、ということについて、コンソーシアムから具体的な情報提供をすることは非常に有益と思われる。

→ 規制改革実施計画として、データを二次的利用で使えるようにするために何をすれば良いかや、処理を何日以内にしなくてはならないなどが上がっている。様式を統一してローカルルールを出来るだけなくしていくことも計画の中に入っている、それが実現化したものがチェックシート審査表などで、これもひとつの大きな特徴である。今回の改善をどのような表現で打ち出していくのかはまだ把握していないが、対外的に宣伝するのは多分年度明け以降になるのではないか。miripo にもいずれ掲載されると思われる。

(6) 報告事項 3 : NewsLetter 第 5 号の発行と動画公開について

資料 8 を基に、3 月末の公開に向けて、掲載内容を説明した。

- オンサイト利用担当者へのインタビューということで、統計センターの阿部穂日氏と千葉亮太氏に、利用申請・持ち出し申請の業務内容、苦労されている点を伺った。
- 昨年 12 月の岡山大・佐井至道先生と金沢大・星野伸明先生主催の「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」イベント報告。
- 活動報告として、昨年のコンソーシアムシンポジウム・評議会・チュートリアル動画公開・ウェブサイトの改修等について掲載する予定。
- 昨年のコンソーシアムシンポジウムにおける東京学芸大・山下雅代先生の SSDSE の利活用例の講演のチュートリアル動画を 3 月末に公開する予定で準備を進めているところ。

(7) 意見交換：数理社会学会・東京大学社会科学研究所との連携について

資料 9 を基に、数理社会学会、特に東大社研との共同ワークショップ案について意見交換した。本件は継続審議とする。

以上