

**第 16 回 ビッグデータ等の利活用推進に関する産官学協議のための連携会議
(第 7 回統計技術・データソースの多様化等検討会 同時開催)
議事概要**

1 日 時 令和 4 年 2 月 3 日 (木) 10:00～12:00

2 場 所 総務省第二庁舎 6 階特別会議室 (Web 会議併用)

3 出席者

- ・ 構成員 高橋座長、和泉構成員、大西構成員、庄司構成員、田原構成員、
水野構成員
- ・ 審議協力者 横浜市立大学データサイエンス学部 佐藤教授
- ・ 事務局 総務省統計改革実行推進室

4 議 題

- (1) 人流データを活用した宿泊旅行統計調査延べ宿泊者数の推計について
- (2) ビッグデータ連携会議の取りまとめについて

5 配付資料

- 資料 1 人流データを活用した宿泊旅行統計調査延べ宿泊者数の推計
- 資料 2-1 各府省におけるビッグデータ等の活用の状況 (事務局におけるヒアリング結果) について
- 資料 2-2 「ビッグデータ関係会議」における取りまとめの修正の方向性について

- 参考資料 第 15 回ビッグデータ等の利活用推進に関する産官学協議のための連携会議
(第 6 回統計技術・データソースの多様化等検討会 同時開催) 議事概要

6 議事概要

- (1) 人流データを活用した宿泊旅行統計調査延べ宿泊者数の推計について

概要は以下の通り。

総務省統計改革実行推進室松井参事官補佐から、資料 1 の説明が行われた。

主な質問・意見は次のとおり

- 人流データが宿泊旅行統計調査の値とかなり連動しており、推計値が算出できるということがよく分かった。今は京都市だけのデータを使って算出していると思うが、他の地域も含めて、時間方向だけでなく空間方向まで分析を広げることも考えられるのではないかと。解析自体もアウトオブサンプルのものとインサンプルのものが混じっていて、なかなか評価がしづらいが、かなりの精度で結果が出ているのではないかと思う。また、同じような解析でテレワーク率のようなものが出せるのではないかと。これまでもヒアリング調査などで、東京都のテレワーク率は何%、といった数値がいくつか出ていると思うので、それとここでの手法を重ね合わせて情報を見える化すると、ビッグデータ活用の理解が進むと思う。
 - ▶ 現在は京都府、京都市、神奈川県で分析しており、ある程度フォローができるが、対象を広げた場合、場所によってデータの偏りの程度が異なると思われるので、そこをどう押さえていくかがこれからの大きな課題と認識している。テレワーク率については、厚生労働省が人流データを用いて分析を試みているという話も聞いている。

- 速報性を高めることにも意味があるが、日次や週次でも把握できると思うので、時系列での粒度、地域的な粒度を上げるといった方向でも付加価値がつけられる。また、神奈川県が宿泊旅行統計調査2次速報の8月の値が突出している点について、スマホを持たない子供や外国人は、このデータではどれくらい捉えられているのか。また、地域によってスマホの所有者に偏りがあると思うが、それをどのように補正しているのか。
 - ▶ 日次、週次の数値については興味深いと思っているが、費用面の問題もあり現在は月次データを使用している。現在使用しているデータには、年齢や男女別の区分はないが、可能であればそのようなデータも検証したいと考えている。また、このデータでは国勢調査の日本人人口で推計、人数合わせを行っている。スマホの所有者についても、その人の住所地のデータではなく補正等を行っていない。

- 常住地人口について、国勢調査の500mメッシュ人口を使っているのが目についた。国勢調査は5年に1度であり、特に観光地のような都市部の場合、ちょっとした要因で500mメッシュ内の人口が変動することもあるかと思う。毎月自治体が公表している住民基本台帳のデータを活用してはどうか。
 - ▶ おそらくメッシュ単位の情報にはなっていないため、メッシュ単位の人口を考えると少し加工が必要になると思うが、標準システムで自動的・統一的に数値が出ると利用する側も活用しやすい。

- 自治体ごとに様式やタイミングなどが微妙に異なると、データの収集に時間と労力が必要となり、せっかくのオープンデータの活用が遅れている原因となっている。各自治体、省庁がメッシュ、単位、様式等を統一化することで、国勢調査より早く、情報を一元的に入手できるようになれば、活用が進むのではないか。
- 以前似たような解析を行った時は、宿泊施設のポリゴンを使って、500mメッシュの中でその施設がどの程度を占めているのか、そのメッシュは全体の宿泊旅行統計と連動するのかを吟味して、必要なメッシュのみを購入した。そのようにすると予算削減ができるのではないか。
- 説明変数、被説明変数とも自己相関や季節自己相関が見られるうえ、両系列間には系列相関も見られるように思われる。こうした変数について、水準のまま回帰分析を行っているようだが、推計において何らかの調整をしているのか。
 - 区分的回帰分析の部分については、尤度比検定を使った再現性のある方法である。区間ごとの回帰分析の尤度比を使った式を作り、区間と回帰係数を同時に決定している。確かに、尤度を計算する時にIIDの仮定がどうしても必要になるが、IIDの仮定を置かないというのも難しいと思っている。
- 季節調整をかけた後の前月比を利用して確認すると、季節相関や自己相関の問題がなくなるので、そういった方法もあるのではないか。
 - 回帰分析の方法を改良するところで検討していきたい。

(2) ビッグデータ連携会議の取りまとめについて

概要は以下の通り。

- 事務局 前原参事官補佐から資料2-1、資料2-2の説明が行われた。

主な質問・意見は次のとおり

- BigData-STATSダッシュボード(β版)では更新を終了した指標もあるということで、インセンティブや購入費用、指標の継続に当たっても課題があるのではないか。資料では特段の課題等はないといった記載が多いが、ヒアリングをした感触はどうか。
 - 既に公的統計の中に組み込まれて実施している調査に関しては、定期的に、例えば毎月公表していく必要があるため、比較的順調に進めているところが多く、公表スケジュールに堪え得る範囲でビッグデータを活用しているという印象もある。BigData-STATSダッシュボード(β版)は、指標を試行的に公開しているものであり、毎年度必要に応じて実施していると聞いている。こういった試行的な取組も有効かと思う。

- 今回の資料 2 - 1 でオルタナティブデータとしてJCB消費NOWの紹介があるが、こういった公的統計に活用されていないオルタナティブデータについて、連携会議の取りまとめの中ではどう位置付けられるか。
 - ▶ 公的統計へのビッグデータの活用は大きな目的の一つであるが、ビッグデータの使い方は様々であり、こういったオルタナティブデータについても取りまとめの中に記載していきたい。

- ビッグデータの活用方法、分析手法などをどこまで公開するかといった観点も重要。
 - ▶ 「取りまとめ」の中の公表方法にかかる記載のところ等で、記載ぶりを検討したい。

- POSデータの活用などもまだまだ品目が限られている。難しいと思うが、今後の目標などが少しでも記載できると良いのではないか。また、現行の統計を作成しながら、更に新たな手法を導入するというのはなかなか難しい面もある。省庁を超えた情報共有をこの連携会議で今後もやっていくことが有効ではないか。
 - ▶ ヒアリングにおいても、現行の統計調査を実施しながらビッグデータの活用を進めるのはなかなか難しいということを感じた。本日の議題（1）で紹介した人流データの検証などの取組を進め、共有していくことで、ビッグデータを使いやすい環境を整備していきたい。

以上