

私の政策提言：マイクロデータの提供拡大による地域経済効率化

1 はじめに

今、私たちは息をするかのように様々な情報を日々発信している。インターネット上にはかつてない程の膨大な情報が蓄積されており、調べ学習を図書館で行う人は少数派になりつつあるのかもしれない。だが、驚くべきはデータの量だけではない。それを分析する手法も、データの量に負けず劣らず、急激な進化を遂げている。分野でいえば、統計や計量経済学などである。コンピュータの計算能力向上に起因するデータ分析手法の進化により、今まで卓上の空論とあしらわれてきたものが、確たるエビデンスをもってその正当性を主張できるようになったのである。

私は経済を学び、このような分析に携わる中で、データ分析により地方創生に貢献できるのではないかと考えるようになった。なぜなら、数ある地方創生プロジェクトがある中、地域の特性に合わせた施策は少ないように感じたからだ。地域には、その地域なりの特性がある。その特性をあらわにするツールこそ統計や計量経済学を利用したデータ分析である。今やデータは重要なインフラの一つであることを考えれば、地方自治体が自らの地域についてデータを集め、その個票データを多くの人に公開することが重要だと考える。なぜなら、それにより、学者や企業の積極的データ分析を促し、延いては地域経済効率化という形で地方創生に貢献する期待が大きいからだ。

2 ミクロデータの意義

データを公開することの有用性は近年、さらに認識されるようになり、民間データはともかく、政府や自治体、公的研究機関によって集められた公共性の高いデータは公開されているものが多い。人口、財政、観光、教育、交通や医療など、分野も様々だ。行われた調査は、政府のデータサイトや自治体のホームページを通し還元されている。だが、私たちが手にできるデータのほとんどは集計データである。集計データとは、調査によって得られた個々の情報を、誰かの手によってすでに処理・製表されたデータのことを指す。集計されたデータはそのデータの全体像を把握する際に非常に有用だが、個々の情報が欠落しているため、より細かな情報をそこから手にすることはできない。現在の集計データのみでの公開は、より画期的なイノベーションや政策立案に蓋をしてしまっているのではないか。

例えばある自治体、A市で観光アンケートが取られたとしよう。回答者の年齢、交通手段、滞在中の過ごし方などが質問項目にあり、A市のホームページには、各々パーセント表記で集計結果と報告書が公開されている。これにより、A市への観光客の年齢層が高いことや、

車での来客が多いこと、レジャー・スポーツが主な滞在目的だと分かるかもしれない。しかし、「60歳以上の方になると、電車での訪問が多く、主な滞在目的は温泉である」という事実は、その集計結果からたどりつくことは不可能である。なぜなら、A市のホームページには、「年齢を60歳以上と選択された方が、どの交通手段を選び、どの滞在目的を選択したか」という、よりマイクロなデータが公開されていないからだ。

マイクロデータとは、集計される前の個々の情報を指す。先ほどの観光データでいえば、縦軸に回答者の識別番号を、横軸に質問内容を取り、各々の答えを記入した表データがあるはずだ。これがマイクロデータである。このデータをもとにグラフや表などが新たに作成され、観光調査の報告書といった形で集計されたデータが公開される。では、集計データと異なる、マイクロデータのメリットとは何だろうか。それは、どのような属性の人がどのような選択をするのかという情報を与えてくれることだ。もしA市のマイクロデータ公表により、「60歳以上の方になると、電車での訪問が多く、主な滞在目的は温泉である」という事実が浮き彫りになれば、温泉街のバス一日無料券を販売することが電車で来られた高齢者に受けるかもしれない。このように、より詳細な情報を分析していくことはサービスや政策の妥当性を向上することに貢献する。

3 ミクロデータ提供から始まる地域経済効率化

政府は SSJDA（東京大学社会科学研究所附属社会調査データ・アーカイブ）や独立行政法人統計センターなどの公的研究機関を通して、限定的にはあるが、マイクロデータの提供を行っている。一方、地方自治体が主体となってマイクロデータを開示している地域はほとんどない。私は、地方自治体がマイクロデータを率先して開示すべきだと考える。マイクロデータの提供拡大は民間企業や研究者のデータ分析を加速させ、サービスや地域システムの向上を通して、地域経済の資源配分を効率化させるからだ。また、自治体がデータを公開せずに分析を第三者に発注するよりも、データを多くの人に提供する方が地域経済効率化をコスト面で安く達成できるという利点もある。

しかし、根本的な問題として、マイクロデータ提供が本当に分析を盛んにするのかという疑問がある。答えがノーならこの論文の理論が土台から崩れるわけだが、答えはイエスである。主要経済ジャーナル¹の中で、30年前には40%にも満たなかったデータを利用した論文が2013年には、70%を超える割合で掲載された（Hamermesh 2013）。これは、データ分析手法の進化に加え、利用可能なデータの拡大に起因する。また、経済分野で研究をしている一人として、私は未だ日本においては、マイクロデータは供給不足であると感じる。興味のあるテーマが見つかったとしても、データの入手ができないため、あきらめるといったことがあま

¹ American Economic Review, Journal of Political Economy, Quarterly Journal of Economics の経済三大トップジャーナル。

りにも多い。地方創生は、その最たる例だ。経団連の民間企業に行った調査²によれば、最もニーズの高い公共データの保有機関は地方公共団体で 87%であり、研究者だけでなく、事業者も自治体の詳細なデータを求めている。しかし、地方自治体が提供しているデータは集計データが多くマイクロデータが少ないため、どのような政策や事業が成功し、その成功要因は何かを調べようとしても困難であった。これから自治体が率先して、自らのマイクロデータを公開すれば、多くの研究者や事業を営む人がここぞとばかりに分析を始めるだろう。そして、その恩恵はその地域が受けることになる。

4 地方による自発的なマイクロデータ提供に向けて

地方自治体が主体となって、自らの地域に関するマイクロデータを提供するのが望ましい。地方創生という観点でみれば、地方の情報が重要度を増すからだ。そして、提供モデルを一言で表すのであれば、「地方自治体が収集したそのままのデータを開示すること」である。これこそマイクロデータであり、集計データにはない可能性を秘めている。ここからは、この取り組みの導入プランとより具体的なデータ提供モデルについて述べていく。

4.1 導入プラン

総務省の調査³によれば、マイクロデータに関わらず、データを公開する上で地域が優先している課題の第1位は「具体的な利用イメージやニーズの明確化」であり、第2位は「提供側の効果・メリットの具体化」が挙げられている。データの公開に対する認識が希薄であることが実施のインセンティブを下げていることが窺える。第3位には費用、第4位には個人情報等の取扱いに関する制度的整理が挙げられた。

この結果を見ると、マイクロデータ公開がよい結果を生むと説明されたところで中々一歩を踏み出せないのが正直なところだろう。そこでまずすべきことは先例をつくることだ。先例の存在が、挙げられた課題を解決に向かわせる。方法としては政府がマイクロデータの有用性を説明した上で、データ開示を積極的に行いたいと意思表示をした自治体からランダムでいくつかを選ぶ。その自治体に補助金を出してマイクロデータ提供を支援する形が望ましい。ランダムに選ぶことで、サンプリングにバイアスがかからず得られた効果の普遍性が説得力を増す。データ開示に関して問題が発生した場合は、適宜修正を加え、先例を作っていく。これによって「具体的な利用イメージやニーズの明確化」や「提供側の効果・メリットの具体化」の課題は解決されるだろう。費用については、新しい調査を始めるわけではなく、システムの整備をするだけであるため、多くはかからないだろう。このような導入プランを

² 経団連「公共データの産業利用に関する調査結果」（平成25年3月19日）。

³ 総務省「地域におけるICT利活用の現状等に関する調査結果」（平成26年）。

とれば、データの提供拡大の効果が視覚化され、多くの自治体が自発的にマイクロデータを提供ようになるはずだ。次節では、残された課題である個人情報の取扱いに留意した具体的なデータ提供モデルを述べる。

4.2 ミクロデータ提供モデル

自治体に期待される場所は以下の2点のみである。原則、誰でもマイクロデータの取得は可能だが、研究目的の場合、報告義務を設ける。学生や教授は研究成果の報告で、民間企業はサービスという形で地域経済の効率化に貢献してもらうわけだ。ただ、分析経験が浅い人の研究については研究自体に欠陥がある可能性もあるため、定期的にデータサイエンティストなどデータ分析に精通した人を加え、成果報告された研究の正当性について審議する必要があるかもしれない。

4.2.1 データの収集

今年5月30日より改正個人情報保護法が全面施行され、氏名だけでなく、単体で個人の特定が可能なマイナンバーや指紋情報、旅券番号なども個人情報と識別されるようになった。よって、この個人情報に分類されないレベルで、かつより詳しい情報を集めることがデータを収集する上でのポイントになってくる。また、研究目的もしくは事業目的で利用されることについて収集時に同意をとる必要がある。一つ確認しておくべきは、回収率である。個人に対しての調査の場合、回収率が低いことも少なくない。これが個人情報に対する意識からくるものなら仕方ないが、特定のセレクションバイアス⁴による場合は問題だ。例えば、ある地域の健康状態の調査を行ったときに、健常者の回答は集まったが、病気の方の回答は中々集まらなかったという場合だ。これだと収集したデータに偏りが生じ、地域の代表的データとして取り扱うことは危険である。もし、データの回収率が低く、そこに何かしらのバイアスがあると考えられる場合、回収率を上げる努力をする必要がある。例えば、その自治体内で使える商品券を付与するなどだ。地域経済にも貢献し、回収率も高まるだろう。

4.2.2 データの提供

収集時点で配慮されているはずだが、もう一度データの開示前に、それら情報によって個人が特定される可能性がないかに留意すべきだ。間接的な個人特定を防ぐため、調査の回答を実数値で記入させていたものを、いくつかの階級からの選択式にするという工夫があるかもしれない。データ提供は、条件付きという形をとる。研究目的、事業目的の2つの目的において提供を許可し、研究目的においてはデータ利用から得られた成果の報告を義務化する。事業目的に報告義務がないのは、それにより競争上不利になることを防ぐためだ。法規定通り、自治体は取得経緯の確認、授受した情報の項目や提供年月日などを記録し、目的

⁴ セレクションバイアスとはサンプリング過程において、ある種の作為が生じ、正しい統計結果でなくなることを。

を逸脱した場合や記入情報に虚偽があった場合の法的処置についても合意をとる。形式に関しては、集計データは従来通りの PDF などによいが、マイクロデータは CSV や Excel といった二次利用ができる状態で提供することが求められる。

5 おわりに

2010 年、ハイチを地震が襲った直後、衛星画像を公開されると、世界中のこれを助けたいとする人が速やかに避難所を加えた詳細な地図を作り上げた。またアフリカ諸国では、政府が家計のマイクロデータを公開すれば、貧困削減を願う開発経済学者によって多くの研究が行われ、アフリカはその恩恵に預かっている。世界の多くの国でマイクロデータの提供が始まっており、その有用性も実証されているのだ。話が少し大きくなりすぎたかもしれないが、地方創生も同じである。地方消滅と叫ばれて、地方をどうにかしたいと思っている人はたくさんいる。マイクロデータの提供は人々にその方法を与えることになるはずだ。

参考文献

・一般社団法人 日本経済団体連合会「公共データの産業利用に関する調査結果」
2013 年 3 月

・北村行伸（2009）「マイクロ計量経済学入門」日本評論社

・個人情報保護委員会「個人情報保護法の基本」

https://www.ppc.go.jp/files/pdf/28_setsumeikai_siryoku.pdf

・「進化する経済学の実証分析」『経済セミナー増刊』2016 年 9 月、日本評論社

・総務省「情報通信白書」2014 年度版、2017 年度版

・総務省ホームページ「統計法」

http://www.soumu.go.jp/toukei_toukatsu/index/seido/houbun2n.htm

・内閣府経済社会総合研究所「マイクロデータを活用した政策研究の推進」『Economic & Social Research』2015 年夏号 No.9

・Hamermesh, D. S. (2013) "Six Decades of Top Economics Publishing. Who and How?" *Journal of Economics Liter*