

大学の地域研究を支える調査・データ オープンデータ「地域元気指数調査」のご紹介

多くの大学で、地域の研究や課題解決に取り組む「域学連携」が行われています。大学・公共図書館でも地域資料を収集し、こうした活動を支援しています。一方で、国内の地域を共通の指標で網羅的に調査したデータは少ないのが現状ではないでしょうか。

このフォーラムでは、大学における地域研究やデータを使った学びのトレンド、「地域元気指数調査」の大学における実際の活用事例を紹介するとともに、学習者や研究者の地域研究を図書館はどう支援できるのか、図書館への期待を提示しました。

※このレポートは2022年11月18日に実施したフォーラム内容をRPIがまとめたものです。

「地域元気指数調査」の概要

全国には、気候風土に根ざした暮らし、住人が誇りと愛着を感じる情景など、その土地ならではの個性や特徴もつ多様な地域が形成されています。株式会社アール・ピー・アイは、全国に多様で活気にあふれた地域が存在し続けることこそが、日本の魅力と豊かさを支え、これからの社会の元気や幸福をつくると考え、地域の元気を増幅できる施策や取組を明らかにし、地域の元気や個人の幸福度を高めることを目的に、「地域元気指数調査」を2016年より毎年実施しています。

この調査では、回答者に自分の地域の元気はどのくらいか、10点満点で評価してもらいます。この平均を地域元気指数と呼んでいます。さらに5つの分野を設定し、各分野に紐づいた47の指標について、それぞれ当てはまると感じるかどうか、5段階で評価してもらいます。この47指標を地域の元気度の評価（10点満点）の要因として分析しています。また、回答者自身がどのくらい幸せだと思うか、10点満点で評価してもらいます。この平均を幸せ指数と呼んでいます。



調査名称	地域元気指数調査
調査方法	インターネットによる調査
調査対象	全国20～69歳の男女
調査時期	毎年8月に実施
サンプル数	約100,000件

講師

氏家 清和 氏

筑波大学 生命環境系 准教授



専門は農業経済学、食料消費分析。政策提言にも寄与する、公益性や新規技術をめぐる消費者行動、地域データに基づく環境保全型農業や経済波及効果などの研究に従事。地域元気指数調査を学術研究用データならびに統計教育用データとして、大学・大学院での教育研究に幅広く活用。

地域をめぐるデータ利用可能性の拡大

日本は公式な統計の収集体制が確立されている国のひとつであり、これまでに多くのデータが蓄積されてきましたが、近年までその情報が広く活用される体制になっていませんでした。こうした背景を受け、2007年に統計法が全面改正され、統計の位置づけが「行政のための統計」から「社会の情報基盤としての統計」に転換されました¹。その後、2016年には、官民データ活用推進基本法で国および地方自治体による統計データのオープンデータ化が推進され、2018年には統計法の一部改正により、調査票情報の二次利用範囲の拡大が認められました。さらに2020年には、オープンデータの取りまとめを所管するデジタル庁が設置され、政策だけでなく、民間企業における活用も視野に入れた統計データの利用拡大、オープン化が推進されています²。

デジタル庁によると、オープンデータに取り組む地方公共団体の数は、令和4年6月28日時点で全体の約7割となっており、都道府県や大規模都市を中心にデータ提供が行われています。また、政府統計の総合窓口e-statでは、学術研究等のための非集計デ

ータの提供や、APIによるデータ分析ソフトウェアを介した直接ダウンロードが可能となっています。分析ソフトウェアには、高度な分析に対応した「R」や「QGIS」といった無料ツールが利用できます。こうした、オープンデータの充実、分析ツールの登場を受け、地域をめぐる膨大なデータを、高度な手法を用いて、比較的簡単に分析することが可能になりました。

大学でのデータ教育

データを扱う基礎的な知見を持った人材育成を目指す日本政府による「AI戦略」（2022年）を受け、大学でもデータ教育が広がっています。2017年に滋賀大学へのデータサイエンス学部設置を皮切りとして、多くの大学でデータサイエンス関連分野の教育組織の設立が相次いでいます。また、2023年度に新設予定の18学部のうち5学部がデータサイエンス系となっています。さらに、2024年度以降に実施する全ての国立大学の入試において、大学入試共通テストでの「情報I」が原則必須化し、大学入学後のカリキュラムでもデータ教育が重視されています。

データサイエンスには、プログラミングなどによるデータ処理についての知識、データの特徴の要約やデータに基づいた事実の推測についての知識だけでなく、記述的内容も含む、分析対象に関する深い知識も学習していく必要があります。こうしたいわゆる文系的な知識は「価値創造・領域（ドメイン）知識」とされ、「データ処理・コンピュータ科学」、「データ分析・統計学」とあわせてデータサイエンスの3要素とされています³。データサイエンスには、文系（価値創造）と理系（データ処理・データ分析）を融合した学習課程が必要となります。例えば、報告者が授業担当をしている筑波大学生命環境学群生物資源学類では、1年次、2年次で基礎的な統計学やデータ分析の知識、データの処理方法を学んだ後、3年次からは各学生が関心のある分野でさらに実践的・発展的なデータ分析を身に付け、4年次で特定の研究課題を対象として分析手法の総まとめを行っています。

教育研究におけるオープンデータの活用の必要性

大学の教育研究にオープンデータを活用する理由として、多彩なオープンデータがあり、教育研究用途に幅広いデータを利用することが可能である点、

学生の多様な興味関心に近い情報をオープンデータから取得し、それらを分析することで、信頼性が高く有意義な研究が実施できる点があります。加えて、大学の教育研究資金は減少を続けており、社会調査や経済実験への費用、データ分析ソフトウェアの購入等を考えると、支給される教育研究費では十分に対応できないこともあります。こうした様々な理由から、オープンデータを教育研究に活用しています。わたくしどもの研究室でオープンデータを利用した研究では、ある地域をキーに様々なデータをつなげて、分析をより面白いものに行っています。

地域元気指数調査・統計データを使った研究

地域元気指数調査は、2016年から継続的に行われている10万人を対象とした大規模調査で、主観的幸福（Subjective Well-being）など学術的バックグラウンドを備えつつ、多岐にわたる調査項目が効率よくデザインされています。これまで重視されてこなかった、地域住民の意識（主観的評価データ）を詳細に知ることができ、さらに時系列的な変化も追うことが可能です。市町村別に集計が可能であることから、様々な地域データと組み合わせることで多角的に分析できます。

主観データである、地域住民の意識は回答者本人にたずねてみないとわかりません。それとは対照的に、客観データは外から観察可能な行動データ（例えば、支払い金額など）です。経済的指標と個人の幸福感がからなずしも一致しないように、客観データと主観データには乖離があるのではないかの議論から、近年、主観データをしっかり分析することの重要性が高まっています。主観とは個人の思いであり分析対象として安定したデータになりうるのかという疑念もありますが、これまでの研究によれば、適切な処理をすればおおよその正しさを得られることがわかっています^{4,5}。よく知られているものでは、全国企業短期経済観測調査（日本銀行）やBetter Life Index（OECD）、満足度・生活の質に関する調査（内閣府）なども主観データと考えられます。主観データの一つである地域元気指数調査は、調査規模がかなり大きく（10万サンプル）、地域に対する評価がわかり、2016年からの継続的調査であることが特徴です。ここからは、実際に地域元気指数調査を使った研究内容を紹介します。

地域元気指数調査の研究への活用

今野 蒼都 氏

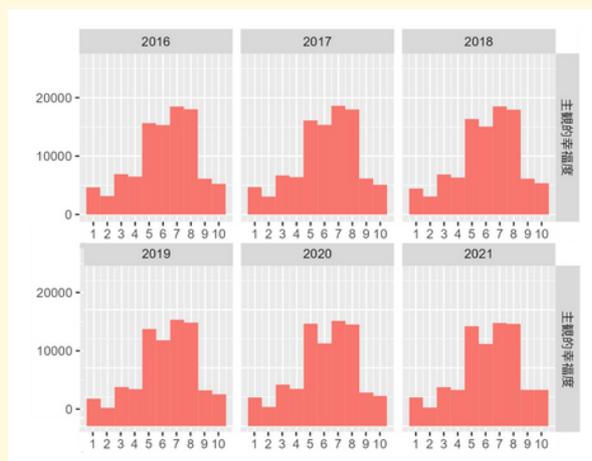
筑波大学 理工情報生命学術院/生命地球科学研究群/生物資源科学学位プログラム 博士前期課程2年

主観的評価のデータの扱い方

はじめに、主観的評価データを学術的に信頼性のあるデータとして取り扱うための集計方法について説明します。

地域元気指数調査では、全国の20～69歳の男女を対象に、幸福度・元気度をそれぞれ10段階で評価した主観的評価データを収集しています。各調査年の幸福度の分布を見ると、5～8点と回答する人が多いようです（図1）。

図1 幸福度のヒストグラム

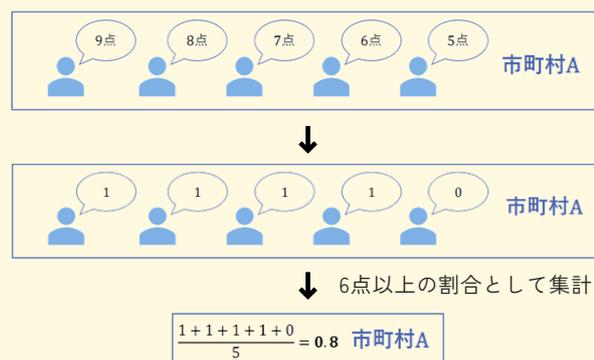


幸福度は主観的評価であるため、回答者ごとに評価の尺度が異なると考えられ、「同じ幸福感＝同じ得点」とは限りません。つまり、同じ程度の幸福感を感じていても、人によって5点と回答する人もいれば、7点と回答する人もいるため、このまま比較や定量的な分析に使うのは問題があるということです。これを個人間比較の問題と呼びます*。

個人間比較の問題に対し先行研究では、グループ（地域元気指数調査の場合は市町村）単位で集計した集計値のほうが安定するという提案や**、一定値（閾値）以上の回答かどうか（0/1）割合として集計するという手法が提案さ

れており***、主観的評価でもデータの扱い方によって個人間比較の問題を軽減することができるとされています。先行研究に倣い、地域元気指数調査の幸福度・元気度にも一定値（閾値）を設定し、市町村ごとの集計値として扱うこととします。具体的には、ある市町村において回答者の幸福度が6点以上かどうかで、幸福か（1）そうでないか（0）に分けて集計を行い（閾値を6点にする）、回答者における6点以上の回答をした人の割合を集計します（図2）。これにより、同じ幸福感を感じていても、6点と答えるか5点と答えるかの差異のみに問題を限定することができます。

図2 集計方法のイメージ



次に、適切な閾値の値を検証するため、幸福度の分布を、5点以上の割合（閾値を5点にする）、6点以上（閾値を6点にする）、7点以上（閾値を7点にする）とした場合の、それぞれの平均と標準偏差を見てみます（表1）。

表1

年	5以上割合		6以上割合		7以上割合	
	平均	標準偏差	平均	標準偏差	平均	標準偏差
2016	0.7796	0.1496	0.6168	0.1688	0.4653	0.1699
2017	0.7817	0.1433	0.6170	0.1627	0.4597	0.1673
2018	0.7846	0.1429	0.6123	0.1663	0.4590	0.1639
2019	0.7769	0.1434	0.6027	0.1728	0.4577	0.1747
2020	0.7744	0.1311	0.5824	0.1656	0.4422	0.1637
2021	0.7841	0.1381	0.6040	0.1679	0.4676	0.1690
全年	0.7802	0.1415	0.6058	0.1678	0.4586	0.1683

平均：市町村ごとの幸福な人の割合の平均

標準偏差：市町村ごとのばらつき大きさ

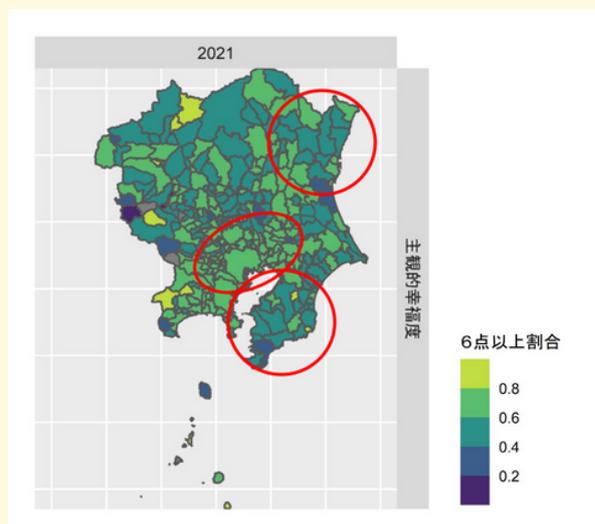
5点以上の割合（閾値を5点にする）の場合は、全年の標準偏差が小さく、市町村ごとのばらつきを集約しすぎていることがわかります。また、コロナ禍前後での変化が表れている2019年から2020年の平均を見ると、6点以上の割合（閾値を6点にする）の場合が変化が大きく出ていることが分かります。このことから、コロナ禍前後の幸福度の変化を検証するため、閾値を6と設定して分析することとします。このように集計方法を工夫することで、主観データであっても、学術的に信頼性のあるデータとして分析に使うことができるのです。

地域の幸福度に影響を与える要因の分析

ここからは住民の幸福度に影響する地域の特徴について、地域元気指数調査を使って進めている研究の一部を途中経過として紹介します。

幸福度を6点以上と回答した人の割合で、関東地方の市町村を塗り分けたのが図3です。紫色に近いほど幸福度が低く、黄緑色に近いほど幸福度が高い地域となっています。2021年では、23区を中心に東京都市圏エリアで幸福度が高く、千葉県南部や茨城県北部などは幸福度が低いことから、市町村によって幸福度のあり方がかなり異なっていることがわかります。

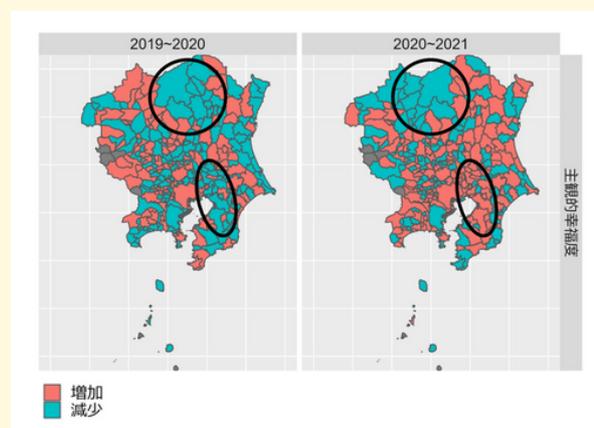
図3 幸福度の地理的分布（関東地方）



地域元気指数調査と行政区画ポリゴン（国土数値情報ダウンロードサービス）より作成

また、2019年から2020年と、2020年から2021年での幸福度の変化を示したのが図4で、赤色が幸福度が増加した地域、青色が減少した地域です。これを見ると、コロナ禍における幸福度の変化が地域によって異なることがわかります。例えば、栃木県北部では2019年から2021年の3年間を通して幸福度が低下しています。一方、茨城県・千葉県首都圏郊外エリアは2019年から2020年にかけては幸福度が低下したものの、2020年から2021年にかけては上昇しています。

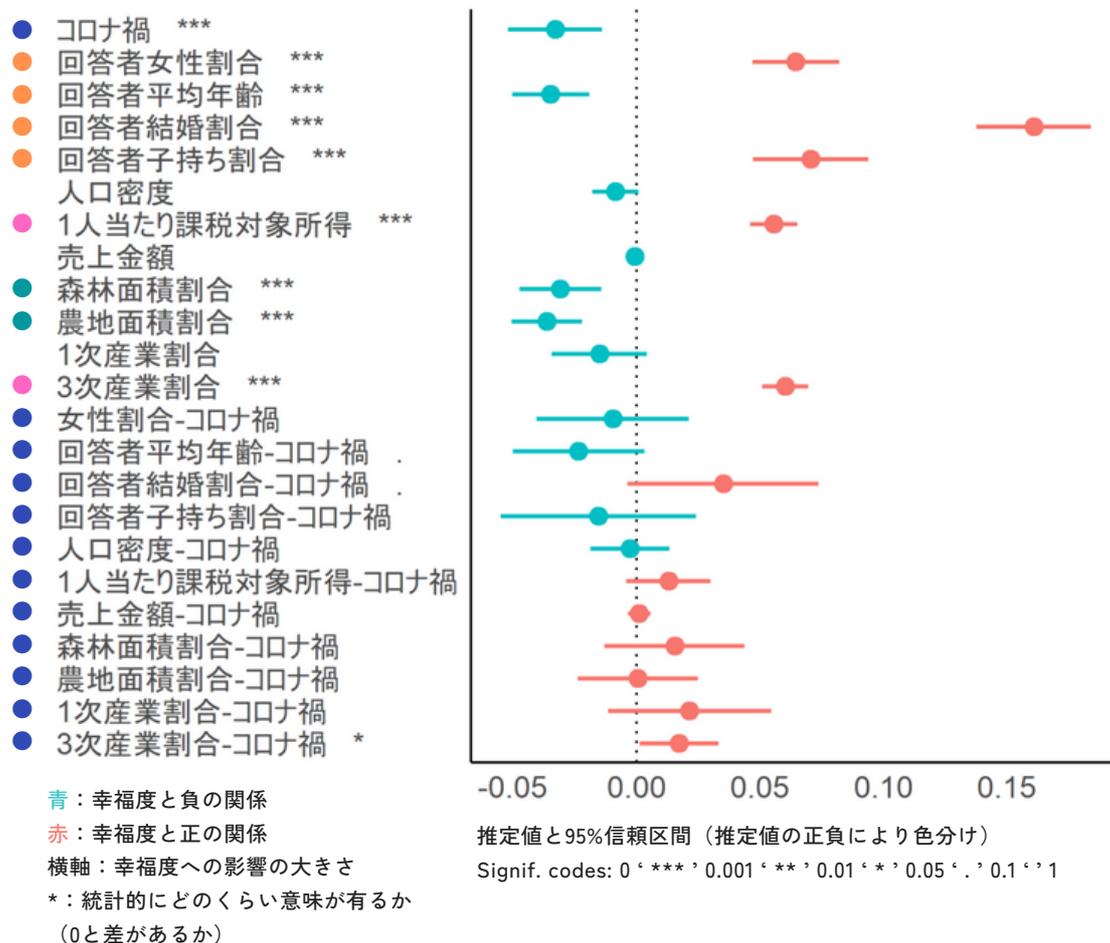
図4 幸福度の変化の地理的分布（関東地方）



地域元気指数調査と行政区画ポリゴン（国土数値情報ダウンロードサービス）より作成

これらの分析から、市町村ごとに幸福度のあり方が異なること、コロナ禍による幸福度の変化が地域によって異なることがわかります。地域元気指数調査では、回答者の居住市町村が把握されているため、公的統計による各市町村のデータと紐づけることによって、幸福度と地域の特徴との関係を分析できます。ここでは、地域の幸福度（6点以上の回答割合）に対して、地域の特徴が与える影響の推定結果（途中経過）の一部を図5で紹介しています。地域の特徴によってその地域の幸福度が異なることと、コロナ禍での幸福度の変化は、個人の特徴や地域の特徴により異なることがわかってきました。より多くの地域特性を考慮することで、今後さらに多角的な議論が可能になると考えています。

図5 推定結果（途中経過）の一部



推定結果の一部

回答者属性の影響は先行研究と整合的

- 女性は高い点数をつける傾向がある
- 結婚していると幸福度が高い

経済的な豊かさ・消費生活の充実等は、幸福度を有意に高める

- 1人当たり課税対象所得
- 3次産業割合

自然に関する特徴は、幸福度に対して負の関係がある

- 森林
- 農地
- 住民は僻地度のように解釈している可能性がある

コロナ禍において、幸福度が有意に低下

- 結婚していると幸福度が高い傾向
- 年齢が高いと幸福度が低い傾向
- 3次産業割合が高い地域は幸福度が高い傾向

* Ravallion, M. (2016) The economics of poverty: History, measurement, and policy. Oxford University Press (柳原透監訳 (2018) 『貧困の経済学・上』 『日本評論社』)

** Krueger and Schkade(2008) "The reliability of subjective well-being measures" Journal of Public Economics, 92, 1833-1845

*** SechelC(2021) "The share of satisfied individuals: A headcount measure of aggregate subjective well-being" Journal of Economic Behavior and Organization, 186, pp.373-394

人・データ・地域社会

政府がデータサイエンスに力を入れ、大学でもデータ教育の拡充が進められている背景には、既存産業への貢献に対する期待があります。ビジネスの諸活動において蓄積されたデータを活用し生産性を上げることが求められている一方で、データを適切に取り扱い分析できる人材が不足しているのが現状です。

また、大量のデータが社会に提供されるようになった今、データの活用には公共的な観点からもインパクトがあると感じています。

いま、OSINT（Open Source INTelligence）と呼ばれるネット上の公開情報を使った調査や分析が広がっています。そのきっかけは、2014年にオランダを発ったマレーシア航空の飛行機が対空ミサイルで撃墜されたマレーシア航空17便撃墜事件でした。調査集団ベリングキャットがSNSや衛星画像といったオープンデータを分析し、親ロシア派の関与を指摘したことが、ロシアの関与を断定する有力な証拠の一つとなり、OSINTが注目される契機となったのです⁶。現在も続いているウクライナ戦争の情勢分析や、新型コロナウイルス感染状況の評価でも、一般市民によるオープンデータの解析が力を発揮しています。このように、オープンデータや無料ツールを利用することで、さまざまな事実の確認を、誰もが行えるようになっていきます。

データがオープン化され誰でもアクセスできるようになったことで情報の民主化がおきているだけでなく、無料の分析ツールや教材によって、データ分析の民主化も進んでいます。データ関連科学は、こうしたデータやツールを活用し、意思決定のための情報を得る「自由人にふさわしい学芸」として位置づけられるのではないのでしょうか。

図書館への期待

図書館はいまでも、そしてこれからも情報の集積地として存在しています。知の媒体は、口伝から紙・書籍、そしてデジタルデータへと技術の進歩に伴い発展していきますが、図書館には、媒体の種類によらず、知の集積地としての役割を担ってほしいと思います。また、地域をめぐるデータには様々なものがありますが、地域によってオープンデータの利用可能性や所在がさまざまに異なる、データベースの使い方、アクセス方法、データ仕様など

も必ずしも標準化されていないなどの課題があり、使いやすいカタログが無いのが現状です。今後も、研究者が望ましいデータにたどり着けるように、また、市民が適切な判断を行えるように図書館が支援することを期待しています。

- 1.内閣府（2007）『第8回FISIM検討委員会資料』
<https://www.esri.cao.go.jp/jp/sna/seibi/kaigi/shiryou/pdf/fisim/070730/shiryou4.pdf>
- 2.大澤（2018）「統計改革と統計法等の改正」調査と立法 No. 403
- 3.竹村（2018）『データサイエンス入門』岩波書店
- 4.岡（2006）「労働経済学における主観的データの活用」『日本労働研究雑誌』No.551
- 5.桑原（2015）「日本はなぜ、主観的幸福度が低いのか」『DIO：data information opinions』
- 6.毎日新聞取材班（2022）『OSINT新時代』毎日新聞出版