

公的統計の マイクロデータを使って 分析してみませんか？

公的統計データの新たな利用方法
オンサイトのご案内



公表されている集計結果からは
分からないことも…



マイクロデータを使用した
分析によって…



公的統計の調査結果は、公表されている集計表に加え、
マイクロデータを使用することにより、研究者の方々は
更に多様な分析を行うことができます。

マイクロデータの利活用なら **オンサイト**

総務省など国の行政機関で実施した統計調査の結果は、
ホームページ(政府統計の総合窓口 e-Stat)等を通じて広く一般の方にご利用いただいています。
このような調査結果の提供に加え、公益性のある学術研究等にご活用いただくため、
調査対象の秘密の保護を図った上で、世帯単位や事業所単位といった
集計する前の個票形式のデータを提供しています。

この個票形式のデータを
マイクロデータ(調査票情報)と言います。

マイクロデータ(調査票情報)を用いることで、
研究者の方々は、より自由で多様な
分析を行うことが可能となるため、
新たな発見につながることを期待されます。

マイクロデータ(調査票情報)のイメージ

Weight	Y_Income	L_Expendi	Food	Housing	LFW	Furniture	Clothes	Health	Transport	Education
895.2667	3917	201649	47756	16028	9652	6702	8088	726	21546	0
895.2667	6675	166381	34054	7416	26313	17062	6989	7637	20773	0
895.2667	6706	259736	84501	1927	10082	6741	5090	11015	53372	0
895.2667	2790	114511	41664	730	22358	5413	1205	5049	17411	0
895.2667	2577	193505	56981	3779	28747	4812	4243	751	16435	0
895.2667	3452	152109	34924	3418	8131	4164	6970	4247	47698	0
895.2667	3233	136900	49956	203	15429	3659	22843	4365	8684	0
895.2667	9252	192439	68882	2832	23042	2598	5714	2052	37006	0
895.2667	2359	138415	53591	753	13072	5140	1786	5416	11593	0
895.2667	2059	79179	32853	14134	7977	3017	2364	1607	7630	0
895.2667	2324	243835	60528	28118	16392	4652	10759	37060	17666	0
895.2667	4524	241539	104433	5253	40637	7711	13833	3853	43003	0
895.2667	4415	207854	95504	7687	13801	19702	7239	3939	9547	0
895.2667	4162	185110	59798	565	12146	14552	12301	2486	26726	0
895.2667	16647	219935	81572	3704	21164	4944	10282	6573	7103	0
977.1795	6760	176625	43112	5013	11780	1697	3008	4836	12295	848
977.1795	6614	130803	32336	3149	14723	21002	12075	3447	24273	140
977.1795	6849	395294	66755	9524	19962	19777	10598	7398	56370	80
977.1795	6813	284803	86655	13479	15121	22966	1134	3649	84535	551
977.1795	6595	260459	46614	1646	15044	6415	16207	3709	76074	69

*データは擬似データです

オンサイトとは?

オンサイトとは、情報セキュリティが確保された環境で、
許可を受けた研究者がマイクロデータを用いて、独自の集計・分析を行うことができる専用室です。
現在、オンサイト施設及び利用可能な統計調査の拡充に向けて、
関係府省、関係機関等と順次調整しています。



入退室管理や
監視カメラを備えた
オンサイト室



オンサイトでできること

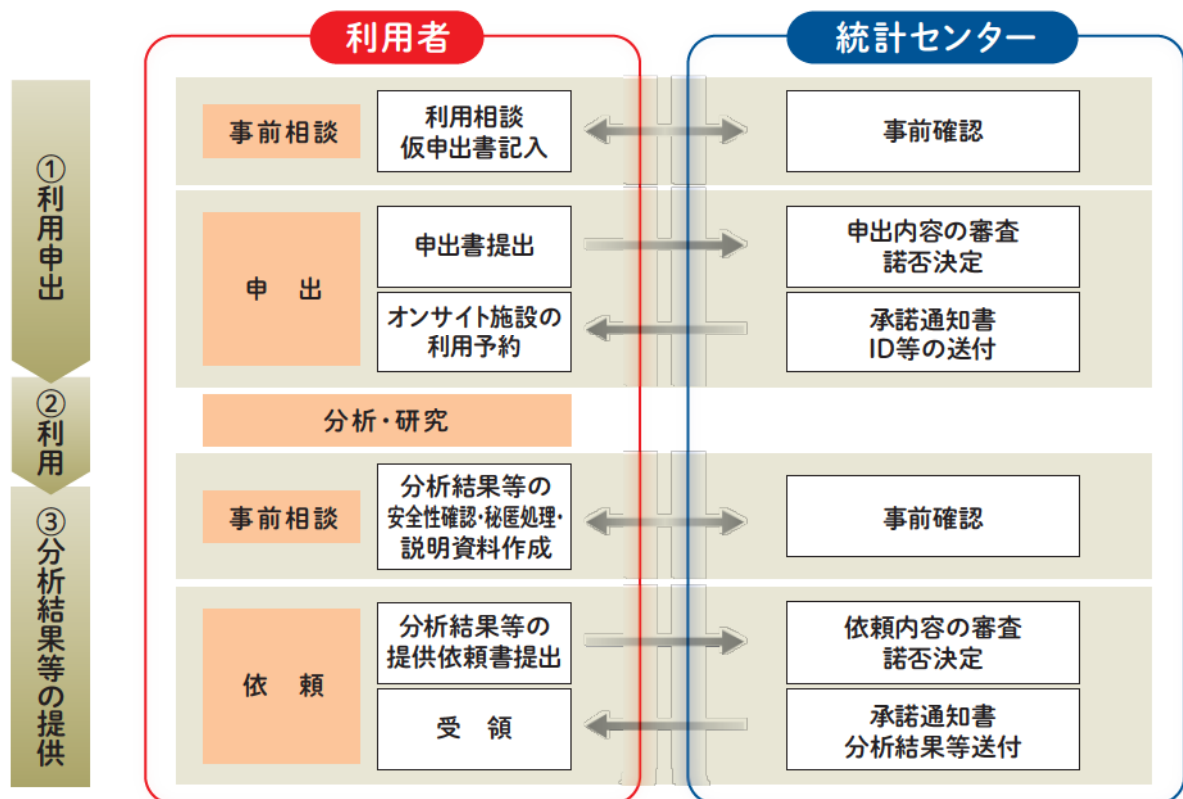
- ・オンサイト内に設置されたPCを用いて、利用申出を行った公的統計のマイクロデータを使った研究分析を行うことができます。
- ・研究者が用意したデータやプログラムも利用することができます。
- ・分析した結果については、セキュリティ保護の観点等から、所定の審査を経た上で提供を受け、利用することができます。

探索的・創造的研究が可能に

利用できる方・条件

- ・行政機関が行う統計の作成等と同等の公益性を有する場合、例えば、行政機関からの委託研究や行政機関との共同研究、行政機関からの公募による補助(科学研究費等)を受けて行う研究などが該当します。
- ・上記以外にも、マイクロデータの学術研究等の利用が可能となりました。例えば、大学等、公益社団法人又は公益財団法人(公益目的事業に限る)が行う研究、大学等に所属する教員が行う研究などが該当します。(この目的で利用する場合は有料となります。)

オンサイト 利用の流れ



統計マイクロデータによる探索的研究



所属 統計数理研究所 筆者氏名 橋 広計 所長

研究背景・マイクロデータ活用の経緯

日本では統計学やデータ解析を専門とした人材が昔から少なかったので、統計学の専門家として、様々な分野の研究者と共同研究を行い、データ解析の支援を行ってきました。これまでも様々なデータに触れてきましたが、中でも公的統計マイクロデータは標本が全国から無作為にサンプリングされていること・回収手続きが厳格で信頼性が高く、サンプルサイズも非常に大きいため、有用性の高いデータとして研究に利用しています。

下記に示す自殺要因に関する研究においてもマイクロデータを活用しました。本研究は公衆衛生・精神衛生を専門とされている先生方が中心で取り組まれていた厚生労働科学研究への参加がきっかけで着手することとなりました。厚生労働省国民生活基礎調査ではK6(気分障害・不安の尺度)を測定しているので、厚労科研チームで竹林由武先生(福島県立医大)らを中心に探索的データ分析を実施しました。その結果、興味深い発見が可能となると分かったため、本格的に研究を進めることとなりました。

研究概要

活用したデータと手法等

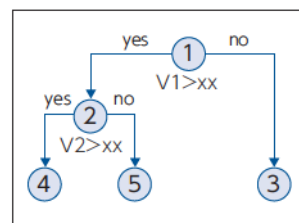
まず、国民生活基礎調査のマイクロデータを対象に、1980年代後半に開発された第2世代人工知能(分類樹)による分析を竹林由武先生が中心に行い、介護者など自殺ハイリスク群で自殺リスクを高めている要因を抽出しました。

次に、岡檀先生(統計数理研究所)が社会生活基本調査のマイクロデータと教育用標準データセット(総務省統計局:都道府県・市区町村のすがた(社会・人口統計体系)の市町村別データから作成)を用いて、個人要因と1,741自治体レベルで集計した環境要因を合わせて分類樹分析をおこない、(自殺リスクを高める要因である)介護に携わる時間を増加させる要因を分類樹形式で抽出しました。

研究結果

国民生活基礎調査マイクロデータを対象とした分析では、精神疾患者・介護者・無職者を自殺ハイリスク群としました。いずれの群においても、最も自殺リスクを高めている要因(予測要因)は主観的なストレス状態の自覚であることが分かりました。その他にも世帯種類、貯蓄高、総所得、職業分類等が自殺リスクに大きな影響を与えることを明らかとしました。また、精神疾患のリスクを高める大きな要因として脳卒中既往や低い貯蓄高等が大きいかを確認しました。

次に社会生活基本調査マイクロデータを用いた分析により、個人要因よりも環境要因、市区町村の特性よりも都道府県が、介護に携わる時間に強い影響を与えている可能性が示唆されました。



再起分割分析により、分類樹形式で予測要因を抽出

マイクロデータ活用の有用性など

上記のように、公的統計マイクロデータは、対象が広範で解析可能な変数の種類もサンプルサイズも非常に大きいデータです。一定のサンプルサイズを必要とする人工知能など探索的分析も可能な有用性の高いデータなのです。社会生活基本調査、国民生活基礎調査などのマイクロデータを用いないとできない地域研究は他にも多々存在すると考えています。一方で、自治体別特徴の集計分析を行うと、小さな自治体ではサンプルサイズが小さくなり現状では公開不可能となるケースもありました。より地域政策分析に利活用しやすい標本設計についても検討が進むことを期待しております。

経済学等の実証研究では既に公的統計マイクロデータの認知度が高い一方、環境、医療等認知度を高める余力のある分野も数多くあります。公的統計マイクロデータの認知度向上、データサイエンスの専門家育成推進、オンサイト施設で利用できる統計マイクロデータの種類・量の増加等に伴い、様々な分野で統計マイクロデータの利活用が進んでいくのではないのでしょうか。

利活用事例

① Takebayashi, Kubota and Tsubaki(2016) [Risk profiles for severe mental health problem: classification and regression tree analysis] Proc. COMPSTAT 2016, pp.291-302.

② 岡、岡本、久保田、竹林、谷道、橋(2019)『社会生活基本調査マイクロデータを利用した介護高負担要因の探索的解析』「官民オープンデータ活用の動向及び人材育成の取組」研究会

観光学におけるマイクロデータ利活用



所属 和歌山大学観光学部 筆者氏名 大井 達雄 教授

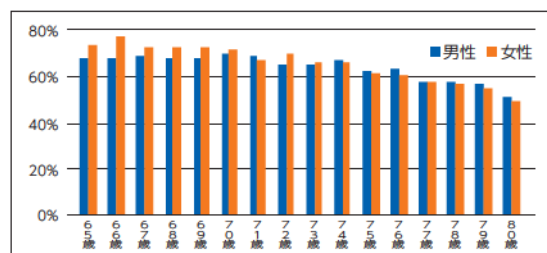
観光学と公的統計マイクロデータ

観光学は、観光に関する諸事象を研究する学際的、かつ比較的新しい学問分野です。その範囲は、社会学、文化人類学、都市計画、心理学、経済学や経営学などに及び、学会ではさまざまな研究分野から活発な議論が行われています。しかし、このような学問的背景の異なった研究者が議論するため、その内容がかみ合わないこともしばしばみられます。そのため、広範囲な研究分野を統合し、研究分野としての地位を高めるためには、データに基づいた実証分析がますます重要となります。一方で観光統計は観光庁を中心に整備が進められていますが、その蓄積は十分ではありません。このような状況の中でデータを使用した分析を行うために、「社会生活基本調査」に代表されるような公的マイクロデータの活用が期待されています。

マイクロデータ利活用事例

本研究では、「社会生活基本調査」の旅行・行楽に関する回答結果を用いて、高齢者の旅行・行楽行動について実証分析を行っています。具体的には、1年間で旅行(宿泊を伴う観光)または行楽(日帰り観光)をしたことがあるかどうかの行動率を分析しています。65歳以上の高齢者を対象にすると、当然のことながら加齢に伴い、行動率は下落します。しかし男女別の行動率でみると、65～69歳では女性の方が高く、70～74歳ではほぼ同じ水準となり、75歳以上となると男性の方が高くなります。つまり女性のほうが落ち込みが大きくなることが報告書からわかります。そこで5歳間隔ではなく1歳ずつで行動率を集計しました。グラフから70歳までは女性のほうが行動率が高いですが、74歳を過ぎると男性の行動率のほうが高くなる傾向にあることがわかりました。今後は他の調査とのデータ連携も視野に入れつつ、家族世帯、単身世帯による差についても分析することを予定しています。

■男女・年齢階層別の旅行・行楽の行動率(2016年)



70歳までは女性のほうが高いが、74歳を過ぎると男性のほうが高くなり、旅行・行楽の行動率が逆転する傾向にある。

マイクロデータを活用される方々への期待

マイクロデータを用いた研究では、集計後のデータからは見えてこない、細やかなデータの揺らぎや小さな傾向を発見できることがメリットとして挙げられます。一方でデータ量が膨大な為、それらに埋もれてしまうことがあります。私自身、オンライン施設でマイクロデータを使った分析を進めていると、ついつい時間が経つのを忘れて研究に没頭してしまいます。回帰分析はもちろんのこと、クロス集計だけで解釈は色々できますし、良くも悪くも非常に自由度が高いと思います。そのため、こういった統計分析にはある程度のスキルが必要となります。しかし、実証分析に慣れていない研究者の方であってもこれらのデータの山に埋もれてもらいたいと考えています。そういった経験をもとにデータ解析の経験を積んでいただき、観光学において新たな発見が生まれることを期待しています。

観光分野におけるマイクロデータ利活用促進に対する期待

観光学に限っていうと、そもそも公的統計マイクロデータを利活用できることを知らない研究者の方もいると考えています。一方、人文科学においても、最近では純粋な理論研究だけではなく、データ利活用の機運は高まってきています。その為、まずはマイクロデータをオンライン施設で利用できることを認知してもらう必要があると思います。加えて、マイクロデータを利用するためには利用申請手続きを経る必要がありますが、申請してから審査が終わるまでの期間が長いこともあります。利用申請手続きの簡素化についても併せて進めることで、より観光分野の研究として使いやすくなると期待しております。

利活用事例

大井 達雄
経済統計学会第63回全国研究大会報告(2019年9月6日)
社会生活基本調査を使用した観光行動の地域分析—高齢者の観光行動に注目して—

研究キーワード

観光学、旅行、行楽、行動率、高齢者、シニア、男性、女性

外生的ショックに対する 雇用調整・賃金調整の実証研究



所属 東京大学大学院経済学研究科 筆者氏名 川口 大司 教授

研究背景・マイクロデータ活用の経緯

労働経済学・教育経済学における実証研究、特に人的資本形成過程の分析とその経済格差への影響に関する実証分析を多く行っています。その中で、日本における正規労働者・非正規労働者間の雇用調整に関する非対称性(景気や業績の悪化時、正規労働者ではなく非正規労働者の雇用を減らす傾向)の存在とその原因を明らかにするため、外生的ショックとして為替レートの変動を利用し、外生的ショックに対して各企業がどのような雇用調整・賃金調整を行うのか分析することにしました。その際、個社の雇用調整・賃金調整推移や、為替レート変動によって受ける影響の大きさ(本研究では輸出入比率を利用)を特定できる企業活動基本調査・賃金構造基本調査のマイクロデータを利用することにしました。

研究概要

活用した データと 手法等

まず、企業活動基本調査のマイクロデータを用いて個社の正規労働者・非正規労働者別の雇用調整推移を算出し、同様に算出した個社の輸出入比率・R&D費率・利益率、産業ダミー変数、為替レート(国際決済銀行のデータ利用)等を説明変数、企業の業績・正規労働者雇用者数・非正規労働者雇用数を目的変数として、回帰分析を行いました。

その後、産業・規模別に集計した賃金構造基本調査のマイクロデータと統合し、為替レート変動に伴う産業・規模別の正規労働者・非正規労働者賃金調整推移を分析しました。

研究結果

企業活動基本調査・賃金構造基本調査のマイクロデータを用いた分析から、外生的ショックに対して、正規労働者についてはほぼ雇用調整は行われずに賞与を中心とする賃金調整により対応している一方、非正規労働者については賃金調整は行われずに雇用調整により対応していることが明らかになりました。影響を受けるグループを特定することは、政策立案においても重要な示唆を与えます。

本研究は、輸出入比率等の企業の異質性に着目し、時間を通じた為替レートの変動が企業の業績や、正規労働者・非正規労働者別に雇用に与える影響を分析した点が特徴的と考えます。

景気悪化時の推移	雇用数	賃金
正規	→	↘
非正規	↘	→

景気悪化時、正規労働者は賃金調整、非正規労働者は雇用調整が生じる

マイクロデータ活用の有用性など

公的統計は回収率が高く、厳格な手続きにより情報も正確である点が強みとして挙げられます。企業活動基本調査が(多くの民間データと異なり)未上場企業も対象としており、回答義務もあることから分かるように、カバレッジの広さも長所であり、民間データでは代替しにくいと考えます。また、「薄い」関係性を捉えるためには一定以上のサンプルサイズが必要となるため、サンプル数の非常に多い公的統計はこの点でも利便性が高いと言えます。更に、長期的に同じ枠組みで調査しているため時系列での分析がしやすい点も研究者にとってメリットだと感じます。

今後、オンサイト施設で利用可能な公的統計の幅や時系列の深さが増え、統計マイクロデータ間のマッチングもしやすくなるような仕組み(例えば分析プログラムをシェアできる仕組み等)が整備されると、より一層統計マイクロデータの利用が促進されるのではないのでしょうか。

利活用事例

Yokoyama, Higa and Kawaguchi(2019)
[Adjustments of Regular and Non-Regular Workers to Exogenous Shocks: Evidence from Exchange-Rate Fluctuation] Industrial and Labor Relations Review

研究キーワード

Exchange Rate; Permanent Shocks, Temporary Shocks

労働経済学とマイクロデータの関係性



所属 神戸大学大学院経済学研究科 筆者氏名 勇上 和史 准教授

研究背景・マイクロデータ活用の経緯

労働経済学の分野では、賃金と雇用が2大テーマとして捉えられています。その中で、私は知識・技能の異なる労働者が自らの時間を切り売りして働く中で生じる賃金格差・雇用格差の構造解明をテーマとしています。従来、企業間の賃金格差については企業規模の違い、労働組合の有無、産業・業種の違いが説明要因として用いられてきましたが、各企業の社齢(創業してから何年経ったか)により、企業の存続確率や人材を集められる力等が異なり、結果として賃金格差が生じるのではないかと、という問題意識を持ちました。

政府統計は厳密な抽出手続と高い回収率を保持しつつ、継続的な調査を行っており、日本の労働市場の問題そのものである賃金格差について議論するにあたり有用性の高いデータです。労働経済学の分野ではマイクロデータの利活用が世界的に進んでいることもあり、本研究においても政府統計のマイクロデータを利用することにしました。

研究概要

活用したデータと手法等

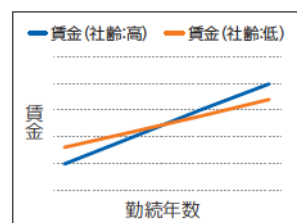
社齢の差が賃金構造に与える影響について分析するため、本研究では賃金構造基本統計調査と事業所・企業統計調査の2004年・2005年・2008年マイクロデータを利用しました(分析にあたって、共通して割り振られている事業所番号を用いてマッチングを行っています)。

まず、①目的変数に賃金、説明変数に社齢、企業に係るその他変数(従業員数、産業・所在地ダミー変数等)、従業員に係る変数(年齢、勤続年数、教育水準ダミー変数等)を取った重回帰分析を行い、社齢に係る係数(社齢が高くなるにつれて、賃金がどの程度変化するか)を確認しました。次に、②社齢と勤続年数の交差項を加えて重回帰分析を行うことで、社齢が勤続年数と賃金との関係性に与える影響を分析しました。

研究結果

上記①の分析により、社齢に係る係数が負、すなわち社齢が高いほど賃金が低下する傾向があることを明らかにしました。次に②の分析により、社齢が高いほど、勤続年数の増加に伴う賃金の増加率が大きい(賃金の年功度合が高い)ことが示されました。更に、製造業と非製造業に分けて分析することで、上記傾向が非製造業においてより顕著に見受けられることを明らかにしました。本研究のように複数の政府統計のマイクロデータをマッチングすることにより、より幅広い観点での分析が可能となります。これにより従来はできなかった研究が可能となり、研究の幅が大きく広がりました。

■勤続年数と賃金の関係性



社齢が高いほど、勤続年数-賃金カーブの勾配が急になる

マイクロデータ活用の有用性など

マイクロデータと違い、集計データでは(集計単位ごとの)平均値しか分からず、集計単位内部の変数の変動が分からないため、変数間の関係を捉えることが出来ません。応用マイクロ経済学の分野ではマイクロレベルでの予測が重要となるので、マイクロデータを利用することで、研究の幅を大きく広げることができます。統計法改正やオンサイト施設の導入により、従前よりも多くの人が、比較的自由にマイクロデータを利用できるようになったため、今後マイクロデータ利活用が促進され、応用マイクロ経済学における研究がより進んでいくのではないかと考えています。また、以前は一部の学生でなければマイクロデータを利用できない仕組みでしたが、法改正により幅広い学生のマイクロデータ活用が可能となったことは博士課程の学生を中心とした学生の研究環境改善に繋がっています。

一方、マイクロデータの利活用にあたって研究者は政府統計の仕組み(母集団リストの決定方法、抽出の仕組み等)を事前に良く把握しておく必要があります。今後、こうした仕組みが研究者の中で周知されていくことで、マイクロデータの利活用がより一層促進されるのではないのでしょうか。

利活用事例

Okajima, H., Yugami, Morimoto, Okajima, S., Nakamura (forthcoming)
[Firm age and wage determination: Evidence from matched employer-employee data in Japan] Applied Economics Letters

研究キーワード

Wage differentials, Wage profile, Firm age, Matched employer-employee data

■ オンサイト施設一覧 (令和2年3月現在)



マイクロデータ利用 ポータルサイト

令和元年5月1日の改正統計法施行に伴い、調査票情報のオンサイト利用を含む、公的統計マイクロデータ利用の情報を集約したサイトとして、新たに「**マイクロデータ利用ポータルサイト(miripo)**」を開設しました。

miripoには

- ・オンサイト利用に必要な手続き、各種申請書類
- ・オンサイトで利用可能なデータ一覧
- ・利用実績

などの情報を掲載しています。

オンサイト利用の詳細な手続きを確認したい場合は、下記URLをご覧ください。

〈HPアドレス〉 <https://www.e-stat.go.jp/microdata/>

〈オンサイト利用についての相談はこちら〉 onsite@nstac.go.jp

※相談の対応は、統計データ利活用センター又は統計センター統計情報提供課で行います。



統計データ利活用センター



〒640-8203 和歌山県和歌山市東蔵前丁3-17 南海和歌山市駅ビル5階
TEL 073-425-0205(代表) E-mail data-rikatsuyou@nstac.go.jp
ホームページ <https://www.stat.go.jp/rikatsuyou/index.html>

