

全国消費実態調査と国勢調査の照合について

Record Linkage between the Population Census and the National
Survey of Family Income and Expenditure

宇南山 卓
統計研究研修所客員教授
京都大学教授

UNAYAMA Takashi
SRTI Guest Professor
Professor, Kyoto University

令和 4 年 11 月
November 2022

総務省統計研究研修所
Statistical Research and Training Institute (SRTI)
Ministry of Internal Affairs and Communications

受理日：令和4年11月10日

本ペーパーは、総務省統計研究研修所の客員教授である執筆者が、その責任において行った統計研究の成果を取りまとめたものであり、その内容については、統計局または統計研究研修所の見解を表したものではありません。本ペーパーの内容については、執筆者に問い合わせ願いたい（Email: unayama.takashi.4x@kyoto-u.ac.jp）。

本研究では、統計法(平成19年法律第53号)第32条の規定に基づき、国勢調査、家計調査及び全国消費実態調査に係る調査票情報を使用した。

全国消費実態調査と国勢調査の照合について

宇南山 卓

概要

本研究では、国勢調査と全国消費実態調査を個票レベルで照合した新たなデータセットを構築した。日本では、異なる調査間で同一世帯を識別できる識別子は存在していないため、原理的には同一世帯が含まれるはずの調査同士であっても情報を統合して分析することはできなかった。それに対し、本研究では、世帯員の年齢・性別や住居の状況によって世帯の同一性を確認することで調査票情報の照合を試みた。世帯固有ではない属性で照合を可能とするために、国勢調査と全国消費実態調査の対応情報を活用した。照合の結果、おおむね約半数の世帯の照合が可能であった。新たなデータセットでは、全国消費実態調査の詳細な収入・支出・資産の情報に加えて、国勢調査で調査される項目が利用可能となる。その家計分析への応用は今後の課題である。

キーワード：国勢調査、全国消費実態調査、データ照合

Record Linkage between the Population Census and the National Survey of Family Income and Expenditure

UNAYAMA Takashi

Abstract

A new dataset has been constructed by merging the Population Census (PC) and the National Survey of Family Income and Expenditure (NSFIE) using record-linkage technique. In Japanese official statistics, no unique identifier for households is available that is common across surveys, and so we cannot merge information from different surveys even if common households must be included. In this study, however, we try to match survey information by using location as well as demographic information such as the age and gender of household members and characteristics of housing. In order to do so, we exploited the concordance between sampling area units in the PC and the NSFIE, which is not publicly available. As a result, about half of the NSFIE sample households was identified in the PC. With the new dataset, not only the income, expenditure, and wealth information from the NSFIE but also detailed demographic information from the PC are jointly available as one record. Its application to household analysis would be an issue for the future.

Keywords: Population Census, National Survey of Family Income and Expenditure, Record-Linkage

1. はじめに¹

本研究の目的は、国勢調査の調査票情報と全国消費実態調査の調査票情報を照合することで、追加的な個人・世帯の属性を家計収支データに付加し、より高度な家計分析が可能なデータを構築することである。国勢調査は日本に居住する全ての個人・世帯を調査しており、全国消費実態調査の調査世帯の大部分も調査対象となっているはずである。国勢調査・全国消費実態調査の2つの調査への同一世帯の回答を統合することで、より広範な家計の分析が可能となる。

ただし、日本では、国勢調査を含むあらゆる個人・世帯に関する調査において、住所・電話番号などの個人情報、マイナンバーのような識別子など、個人を容易に特定できる情報は調査事項として収集していない。そのため、異なる調査間で同一世帯を識別するための識別子は存在せず、厳密な個人・世帯固有の情報に基づく照合はできない²。そのため、原理的には同一世帯が含まれるはずの調査同士であっても、現実には情報を統合して分析することはできなかった。たとえば、英国では、国勢調査にあたる全数調査と各種標本調査を住所単位で照合することができる(Freeth, White and Martin, 2001)。

本研究では、世帯員の年齢・性別や住居の状況によって世帯の同一性を確認することで、複数の統計の調査票情報を照合する方法によってこの問題に対応した。もちろん年齢や性別などの情報は個人固有の情報ではないが、複数の世帯員の情報や住居の情報を組み合わせれば同一の属性を持つサンプルが存在する可能性はかなり低くなる。たとえば「夫婦子供2人の4人世帯で、夫の年齢が48歳、妻の年齢が47歳、第1子が女で13歳、第2子が男で10歳、80平米の集合住宅の持家に住む世帯」は、日本全体でもそれほど多く存在するわけではない。

さらに、一定の地理的な範囲内に居住する世帯に限定すれば、実質的に世帯を特定する識別子になりうる。そうした地理的な限定を付加するために、ここでは全国消費実態調査の調査区と国勢調査の調査区の対応関係を記録した情報を活用した。全国消費実態調査の調査世帯は、転居などの状況が無視すれば、対応する国勢調査の調査区に存在する世帯の中から抽出されていると考えられる。つまり、調査区の対応関係が分かれば、同一世帯の可能性がある候補を大幅に限定することができるのである。その限られた範囲での世帯であれば、世帯員の年齢・性別や住居の状況が完全に一致する世帯はほとんど存在していないため、住所のような個人情報を使わずとも照合が可能となるのである。

このような方法での全数調査である国勢調査と標本調査である全国消費実態調査との照合は日本で初めての試みとなることから、ここでは構築されたデータを用いて具体的な分析をすることではなく、照合そのものがどのような結果であったかを概観する。

本稿では調査区対応状況の情報が入手できた統計のうち、2010年国勢調査と2009年・2014年全国消費実態調査の照合について報告する。本稿では、時点や照合対象となる調査の違いによって照合できる割合に違いがあるのか、どのような属性の世帯では照合が容易か(もしくは困難か)、照合ができた世帯の属性に偏りはないか、などを示す。また、2015年国勢調査と各回全国消費実態調査との照合、

¹ 本稿は、総務省統計局との共同研究の成果の一部をまとめたものである。研究遂行にあたり、塚田武重、根本和郎、佐藤朋彦(故人)の各氏にさまざまなご協力いただいた。ここに記して感謝したい。

² 米国では、過去の国勢調査を個別レコードレベルで照合し、パネルデータ化する試みが進められている。その照合には名前や年齢などの情報が用いられており、ここでの試みと類似の課題に直面している。詳細については、Abramitzky et al. (2021)を参照。

2010 年国勢調査と 2013-17 年家計調査との照合についても研究を進めているが、その結果については別稿で示すこととする。

全国消費実態調査は、家計調査とならび、収入・支出に関する最も主要な統計である。個人・世帯の属性についても基本的な情報は調査されているが、国勢調査（特に大規模調査）では全国消費実態調査では利用できない個人・世帯の属性についても調査している。個票レベルで各調査の調査対象世帯を照合し情報を統合できれば、調査主体にも客体にも追加的な負担なしに情報の充実が図れる。付加できる情報として有用と考えられるのは、生年月・学歴・詳細な就業状況・従業地・転居の状況などである。学歴に関しては、2019 年に全国消費実態調査が全国家計構造調査に移行し追加的な調査項目となったが、2014 年までの全国消費実態調査では利用できない情報である。同一の学歴の家計の行動がどのように変化したかなどを分析するためにも、時系列的な比較を可能にするデータは重要となる。

照合の結果、国勢調査の調査票情報と照合することができた世帯の割合は、2009 年全国消費実態調査の調査世帯のうちの 56%、2014 年調査の 49%となった。原理的には、転居や死亡などが発生していなければ照合ができるはずであることを前提にすれば、想定よりも低い水準となっている。その要因としては、世帯員の年齢や性別だけでは区別できない世帯が複数存在するために発生したものではないと考えられる。各調査区内には、同一の属性を持った世帯が存在する確率は低かったが、同一属性の世帯が国勢調査の調査世帯に存在しないために照合できなかったケースが大部分を占めている。

同一世帯が存在しているのにもかかわらず、対応する世帯が見つからないのは、調査された属性の一部に「測定誤差」が発生しているからと考えられる。ここでの照合の方法は、各統計での世帯属性が正確に記録されていることを前提としているが、その前提は必ずしも満たされない。たとえば、個人が自分の年齢を正確には回答せず 5 歳もしくは 10 歳刻みなどで丸めて回答していると、調査時点が異なる調査での年齢が同一であるかのように見えてしまい、同一人物であることが判別できない。また、住宅の面積についても、全国消費実態調査では実数（平米数）で回答しているのに対し国勢調査では階級値で回答している。こうした質問形式の違いも、測定誤差の原因となる。統計の中身としては問題とならないような軽微な誤差であっても、同一世帯の同定のためには決定的な問題となりうる。2 つの統計を照合するには、一定の誤差の補正手法についての検討が必要である。

本稿の以下の構成は次のとおりである。まず、第 2 節では、2 つの統計を照合する具体的な方法について論じている。第 3 節では、照合の結果について報告している。第 4 節では、構築したデータをさまざまな属性別に観察することで、照合が十分に機能していない原因について考察する。第 5 節はまとめであり、今後の課題についても論じている。

2. 照合の方法

2.1 照合方法の概要

本研究の目的は、国勢調査・全国消費実態調査・家計調査の個票を照合することで、新たなデータを構築することである。「個票の照合」とは、個別調査の結果のレコード（調査票情報）から、同一世帯の調査結果を探し出す作業のことである。別の統計調査の結果を統合することで、同一世帯に関するより多くの情報が利用できるようになる。

異なる統計を照合するには、世帯を識別するための共通の符号なければ困難である³。現実の調査統計において共通の「符号」を利用できるケースは多いが、多くの先行研究では名前と住所を基準に照合がされている(たとえば、Winkler, 2006)。それに対し、日本の統計調査においては、調査客体を識別する符号は実務上の必要に応じて調査ごとにつけられており異なる統計間で共通した符号(たとえばマイナンバーのような個人識別子)はつけられていない。また、個人を容易に特定できるような情報、たとえば住所・電話番号などの個人情報も調査されず記録もされていない。そのため、データとして利用可能な情報では「実存する個人」を特定することはできず、調査横断的に世帯を識別することはできない。

この問題に対応するため、ここでは、全国消費実態調査・家計調査の調査区と国勢調査の調査区を対応付けることで、同一世帯が含まれる可能性のある集合を限定し、その集合内で世帯属性が一致するような世帯を照合するという手順をとる。限定された世帯の集合の中では、世帯属性が完全に一致するレコードはほとんど存在しないため、擬似的に世帯の識別子として利用できるのである。家計調査・全国消費実態調査では国勢調査の結果をサンプリングフレームとして使用しており、直近の国勢調査の近隣する2調査区を「単位区」として設定し対象調査区を決定している。この対応関係を使えば、潜在的に同一世帯が含まれる範囲を高々数百世帯程度に限定することができる。

2つの調査で「同一世帯」とみなすのは、世帯員の年齢・性別、住宅の建て方や広さなどの情報が一致した世帯である。年齢や性別などの情報は個人固有の情報ではないが、複数の世帯員の情報を組み合わせれば同一の属性を持つサンプルが存在する可能性はかなり低くなる。特に、該当する調査区という狭い範囲に限定すれば、同一の属性を持つ世帯はほとんど存在せず、住所のような個人情報を使わずとも照合が可能となる。

この照合方法にはいくつかの制約、限界が存在する。第1に、国勢調査をサンプリングフレームとする標本調査であっても、国勢調査の調査区との対応関係は特定が困難な点である。第2に、各種世帯属性が一致することは、厳密には同一世帯であることの必要条件でも十分条件でもないという点である。以下では、こうした課題への対応について述べ、具体的な照合方法と課題について議論する。

2.2 調査区の照合と必要な情報

国勢調査と全国消費実態調査・家計調査の個票レベルでの照合の最初のステップは、それぞれの調査における調査区同士を対応付けることである。国勢調査で設定される「調査区」は、地理的に固定される「基本単位区」を組み合わせ(もしくは分割して)調査ごとに設定される。基本単位区とは、『平成22年国勢調査ユーザーズガイド』の「調査結果で用いる用語の解説⁴」によれば、

「街区方式による住居表示を実施している地域では、原則として一つの街区を基本単位区の区画としています。それ以外の地域では、街区方式の場合に準じ、道路、河川、鉄道、水路など地理的に明瞭で恒久的な施設等によって分けられた区域を基本単位区の区画としています。基本単位区の区画は、街区方式による住居表示の新たな実施などやむを得ない理由により変更する場合のほかは、固定されています」

³ 医療系の先行研究では、社会保障番号などを使い民間のデータ同士を照合する研究などがある(たとえば、Newman and Brown, 1997)。

⁴ <https://www.stat.go.jp/data/kokusei/2010/users-g/word.html>

とされている。

一方、家計調査・全国消費実態調査では国勢調査の結果をサンプリングフレームとして使用しており、調査対象市町村を決定した上で、調査市町村内から直近の国勢調査の近隣する 2 調査区を「単位区」として設定し、調査対象調査区を決定している。その意味では、国勢調査の調査区とは厳密に一致した地理的な対応関係が存在している。

しかし、抽出された全国消費実態調査・家計調査の単位区には、事後的に 1 から順に符号が振られており、国勢調査の具体的な調査区との対応関係は分からないようになっている。これは、調査対象単位区が地理的にどのエリアと対応するかが明らかになると調査対象世帯の特定につながる可能性が高いためであり、特別に許可を得た個票利用者も含めて、データの利用者に対しては完全に秘匿されている。つまり、全国消費実態調査の個票レベルでは各世帯の単位区が同じか違うかは識別可能だが、その調査単位区がどの国勢調査の調査区と対応するのかは一般に特定できない状態にある。

今回の研究では、家計調査・全国消費実態調査の単位区と各調査の調査市町村を決定する時点での最新の国勢調査の調査区との対応関係（以下では、調査区対応表と呼ぶ）を用いて、全国消費実態調査・家計調査の調査区と国勢調査の調査区とを対応づけた。調査時点での直近の国勢調査とは、全国消費実態調査では約 4 年前の時点もので、2009 年全国消費実態調査であれば 2005 年国勢調査、2014 年全国消費実態調査であれば 2010 年国勢調査が該当する。参考までに、家計調査では、調査区選定時点で最新の結果を用いるため、国勢調査の結果が公表されるごとにサンプルフレームとなる調査が入れ替えられており、たとえば 2008 年から 2012 年に抽出される家計調査の調査区は 2005 年国勢調査を使って選定され、2013 年から 2017 年に抽出される調査区は 2010 年国勢調査を使って選定されている。

ただし、最新の国勢調査と全国消費実態調査であれば、調査区対応表を使ってダイレクトに調査区同士の関係が対応付けられるが、それ以外の国勢調査との対応関係を知るには、各回の国勢調査の調査区同士の対応関係をつけるというステップが必要となる。国勢調査では、調査回ごとに調査区コードを付与しており、地理的に同一の地域が必ずしも同じ調査区コードとなっているわけではない。本研究では、各回の国勢調査における調査区の対応関係（以下、単位区変遷表と呼ぶ。）を活用した。

この情報によって、全国消費実態調査の調査区は「最新の国勢調査」の調査区との対応関係を通じて、間接的に他の国勢調査の調査区とも照合できる。調査時点での「最新」以外の国勢調査と対応させることは、特に全国消費実態調査において重要な点となる。西暦の末尾が 4・9 の年に調査される全国消費実態調査から見れば、直近の国勢調査とは調査実施の 4 年前のものとなる。4 年間の間に転居や世帯員の死亡・転出などがあると、同一世帯の照合は困難となる。調査時点間の時間的距離を考えれば、むしろ調査実施 1 年後に実施される国勢調査のほうが近く、より多くの世帯を照合できる可能性が高い。

ここでは「最新の国勢調査」と全国消費実態調査の照合として、2014 年全国消費実態調査と 2010 年国勢調査、国勢調査同士の対応関係も使った照合として 2009 年全国消費実態調査と 2010 年国勢調査の照合を試みる。ここでのケースで照合方法が十分に確立できれば、2014 年全国消費実態調査と 2015 年国勢調査、2013-17 年家計調査と 2015 年国勢調査など他の年次にも応用が可能である。

2.3 調査区の照合手順

ここでは、まず 2014 年全国消費実態調査と 2010 年国勢調査の調査区を例にして、具体的な照合手順を示す。このケースでは、2014 年の全国消費実態調査のための「調査区対応表」に記録されているのは、

最新の（4年前の）国勢調査である2010年国勢調査の調査区との関係である。全国消費実態調査の単位区は市町村コード（都道府県番号を含む）と単位区符号で特定ができる。一方、国勢調査の調査区は（市町村コードに加えて）主番号・後置番号・単位番号で特定される。「調査区対応表」には市町村コード・単位区符号ごとに、対応する国勢調査の調査区の主番号・後置番号・単位番号が書かれている。

全国消費実態調査の調査単位区は、原則として2つの国勢調査区と対応しており、さらに国勢調査の調査区は1または複数の基本単位区に対応するため、結果として1つの全国消費実態調査単位区に2から4つ程度の国勢調査基本単位区に対応する⁵。一部の調査区では国勢調査の調査区を分割して全国消費実態調査の調査単位区が設定されているケースがある。たとえば、大規模宅地の開発などで、調査区の人口が大きく増加したケースなどである。こうした場合、本来国勢調査の調査区の一部だけが単位区に対応することになるが、調査区対応表ではどの基本単位区と対応するのかが判明しないため、分割対象となる国勢調査の調査区に対応する基本単位区をすべて対応させた。その意味では、実際には抽出対象ではない基本単位区にも対応関係がつけられている可能性がある。

この基本のケースに対し、2009年の全国消費実態調査と2010年の国勢調査の調査区を対応付ける手順は次のようなものになる。まず、2009年の全国消費実態調査のための「調査区対応表」から2009年時点での「最新」の国勢調査（すなわち2005年国勢調査）の調査区との関係をリストアップする。すでに述べたように、この時点で、全国消費実態調査の調査単位区ごとに2-4程度の国勢調査の調査区がリストアップされていることになる。

次に、対応付けられた2005年国勢調査の調査区ごとに、該当する2010年国勢調査の調査区を「単位区変遷表」を用いて異なる時点の国勢調査の調査区同士を対応付ける。国勢調査の調査区は一つ又は複数の基本単位区を組み合わせで設定されるが、人口変動などに応じてその組み合わせは変更される。世帯数が多くなれば基本単位区を分割して新たな基本単位区が設定される場合もあり、国勢調査各調査の基本単位区は一方が他方を完全に覆うような単純な包含関係になっていない。ここでは「調査区対応表」でリストアップされた2005年国勢調査の単位区を完全にカバーするように新しい国勢調査の単位区を選択した。この手順によって、本来は対応する基本単位区が漏れることを防ぐことができるが、全国消費実態調査の単位区に対応する国勢調査の範囲はやや広めになっている。ただし、この作業によっても、1つの国勢調査の基本単位区が複数の全国消費実態調査の調査単位区に対応することはなかった。

2.4 世帯の照合について

国勢調査は全数調査であり、原理的には国内外の人口移動がなければ、全国消費実態調査の調査世帯は必ず国勢調査の調査対象にもなっているはずである。転居などの事由がない限りにおいて、全国消費実態調査の調査対象世帯は、全国消費実態調査の当該調査単位区と対応するとされた国勢調査の調査区に存在する世帯のいずれかに対応するはずである。日本の世帯調査においては、住所・電話番号・マイナンバーなど個人を容易に特定できる情報は収集していないため、全国消費実態調査と国勢調査の世帯同士を照合するには両統計で共通して調査されている項目で照合する必要がある。

両統計で調査される共通の事項として、住居にかかる事項と、世帯員の属性にかかる事項が利用できる。ここでも、具体的に照合キーとするために、変数を整合的になるように修正して使用した。

⁵ 調査区の拡張があった場合には、拡張した調査区についても記載されている。

- 住居にかかる項目
 - ・住宅の建て方（一戸建て・集合住宅など）
 - ・共同住宅の階数（2014年全国消費実態調査のみ）
 - ・住宅の広さ（国勢調査は階級値）
 - ・住宅の所有形態（持家か賃貸かなど）
- 各世帯員にかかる項目
 - ・世帯主との続柄
 - ・性別
 - ・年齢（国勢調査では生年月）
 - ・就学状況（在学する学校の種類）
 - ・就業状況（職業・雇用形態・産業など）

国勢調査において「住宅の建て方」は、一戸建て・長屋建（テラスハウスを含む）・共同住宅・その他にカテゴリされ、共同住宅はさらに別途「建物全体の階数」が調査されている。それに対し、たとえば2009年全国消費実態調査では、一戸建て・長屋建（テラスハウスを含む）・共同住宅（1・2階建）・共同住宅（3-5階建）・共同住宅（6-10階建）・共同住宅（11階建以上）と複合的に調査している。2014年全国消費実態調査では、「建て方」は一戸建・共同住宅のみの分類になっているが、国勢調査と同様に共同住宅は「何階建」であるかも調査されている。そこで、2009年については国勢調査の「住宅の建て方」を「建物全体の階数」を用いて全国消費実態調査と整合的になるように修正し、最大限の共通情報を照合のためのキーとして利用した。

住宅の面積については、全国消費実態調査では「住居の延べ床面積」を実数で調査しているのに対し、国勢調査では10平米の幅を持つ階級値で調査している。両調査で比較可能とするため、全国消費実態調査の数値を国勢調査の分類に合わせて階級値とした。住宅の所有形態については、両統計共に、持家・公営賃貸住宅・民営の賃貸住宅・給与住宅・都市再生機構等の賃貸住宅・間借り・寮寄宿舎に分類されている。

世帯員に関する情報については、年齢と性別を基本的な照合キーとする。ただし、国勢調査では生年月が調査されるのに対し、全国消費実態調査では満年齢が調査されている。そこで、国勢調査を全国消費実態調査の実施時点での満年齢を計算することで、両統計を比較可能とした。就学状態や就業状態は時点によって変化する可能性があるため、できる限り照合キーとしては使用しない。

2.5 論理的な照合可能性

一見すると、これらの照合キーが全て一致することが「同一の世帯」であることの必要条件であるように見える。しかし、これらの項目が全て一致することは、世帯の同一性の観点からすると必要条件でも十分条件でもない。つまり、照合キーが完全に一致したとしても異なる世帯である可能性があり、逆に一致しない項目があっても同一世帯である可能性がある。

たとえば、公営住宅などが調査区となった場合、住宅の状況が完全に一致する世帯（間取りの同じ部屋に住む世帯）は多く存在する。さらに、公営住宅などでは無職の高齢夫婦・高齢単身世帯などの類型が少なくない。そのため、住宅の特性と世帯員の属性全てが完全に一致する世帯が同一調査区に複数存

在しても不思議ではない。こうしたケースでは、上記の照合キーの一致だけでは同一世帯かどうかの判断はできない。

逆に、照合キーが完全には一致しなくとも「同一世帯」の可能性もある。調査区という地理的な範囲に基づき調査を実施している以上、住宅については、住宅の建て方・住宅の面積・住宅の所有形態が一致することは、基本的に同一世帯であることの必要条件である。一方で、世帯員については、照合キーの完全な一致は必ずしも必要条件とならない。典型的には、新生児の誕生や世帯員の転入、世帯員の転出・死亡などの異動が想定できる。全ての照合キーの一致を前提とすれば、こうした世帯属性の変化によって、同一世帯であっても照合することができない。

世帯員の異動を許容した場合、どのような状態を「同一の世帯」と考えるかは概念的な問題となる。たとえば、四人家族の長子が一人暮らしをするために転出した場合であれば「世帯員の異動のあった世帯」とみなすことは自然である。しかし、たとえば、高齢単身世帯に子供世帯が転入してきたような場合、転入前から世帯に居住するのは高齢一人親のみであり、このケースを「同一世帯の世帯員の異動」とみなすべきか議論の余地がある。

ここでの目的は、国勢調査で収集された情報を全国消費実態調査に付加することであり、どの世帯員の情報が必要であるかによって照合の基準を考える必要がある。そこで、この課題に対応して、世帯員については「コアな世帯員」を設定し、その世帯員の年齢・性別を照合キーとした。コアな世帯員として、世帯員のうち 15 歳未満（照合する調査のうち一方の時点で誕生していない年齢を除く）および 30 歳以上 70 歳未満の世帯員全員の年齢・性別が一致することを「同一世帯」であることの必要条件とした。この年齢設定は、15 歳以上 29 歳未満および 70 歳以上は、転出・転入および死亡の可能性が他の年齢層よりも高いことを反映している。

コアな世帯員のみを照合キーとすれば、たとえば子供が大学入学に伴い転出した、70 歳以上の親が子供の世帯に転入してきた、70 歳以上の世帯員が死亡した、などのケースがあっても同一世帯として照合できる。逆に、70 歳以上の両親の家に子供家族が転入してきたようなケースは「別の世帯」とみなされることになる。妻の就業や子供の就学についての分析をするのであれば、こうした現役世代を中心として「コアな世帯員」を設定することを正当化できる。

もちろん、こうした照合キーの作成方法では、70 歳以上の世帯員だけで構成される世帯では年齢・性別の情報が使われないことになる。しかし、これは高齢世帯の照合ができないことを意味しない。たとえば、調査区内に 70 歳以上の世帯員のみで構成される世帯が他になければ同一世帯であると判断できる。また、下で述べるように、コアな世帯員の定義を適宜入れ替えての照合もしており、高齢世帯の照合も一定程度達成できている。

2.6 照合の具体的な方法

この「コアな世帯員」の情報を用いた具体的な実際の照合手順は、次のようなものである。

1. 家計調査・全国消費実態調査および国勢調査の対応する調査区内のすべての世帯について、上記の住宅の属性（住宅の建て方・建物全体の階数・住居の延べ床面積）と世帯員に関する項目（15 歳未満および 30 歳以上 70 歳未満の世帯員の年齢・性別）に基づく照合キーを作成する。
2. この照合キーが、対応する家計調査・全国消費実態調査および国勢調査それぞれの調査区におい

て 1 世帯だけが該当するかどうかをチェックし、照合キーに該当する世帯が複数存在する場合には、当該照合キーに該当する世帯を照合対象から除外する。このステップにより、各調査の調査内での「世帯 ID」として機能することになる。

3. 該当する世帯が全国消費実態調査・国勢調査それぞれで 1 世帯ずつになった場合には、他に一致する世帯がないという意味で「照合」と判定する。

もちろん、世帯属性は論理的には世帯固有のものではなく、一般的な状況でこうした照合キーを用いてマッチングをすれば異なる世帯を同一と取り違える可能性は高い。しかし、ここではすでに 2~4 つの国勢調査の基本単位区に照合対象を限定しており、数百世帯程度まで候補世帯は絞られている。調査区内においては同一照合キーをもつ世帯は限られており、十分に機能するキーとなる。

ステップ 2 について、家計調査・全国消費実態調査内および国勢調査の各調査区内に同一の照合キーをもつ世帯が複数存在すれば、利用している情報だけでは世帯を特定できないため照合は不可能であり、必須のステップである。ただし、家計調査・全国消費実態調査の調査区ごとの調査対象世帯数は 12 世帯程度であるため、1 照合キーに対応するのが 1 世帯という状況は容易に達成できるが、数百の調査区内の世帯全てを含む国勢調査で 1 照合キー 1 世帯となるのは限定されたケースのみであり、照合の大きな障害となる。

国勢調査で「1 照合キー 1 世帯」の状況になりやすくするため、より多くの世帯員の情報を用いることが重要である。ここではコアな世帯員のみを利用しており、全世帯員の情報を使っていないが、コアな世帯員以外の世帯員の情報を使うことにはメリットとデメリットが存在する。できるだけ多くの世帯員の情報を用いれば、より世帯の特徴が明確になるため、同一の照合キーを持つ世帯が複数存在する可能性を下げることができる。これは、コアな世帯員のみを使うと同一世帯であるにもかかわらず（複数の世帯が同一の照合キーを持つため）「世帯の同一性を完全には確認できない」という誤りが発生していることを意味する。一方で、より多くの世帯員情報を用いれば、世帯員の異動などが発生する可能性が高くなり、実際には同一世帯であっても（国勢調査でより多くの世帯で 1 照合キー 1 世帯の状況が達成できても）「同一世帯を異なる世帯とみなす」誤りが発生しやすくなる⁶。

どの世帯員情報を用いるかは 2 つの誤りのバランスで決定する。ここでも、このバランスを検討するために、上記の基本パターンでの照合ができなかった世帯を対象に、コアな世帯員の範囲を拡大・縮小して照合できるケースがどの程度増加するかも観察する。また、各調査で提供される情報に一定のエラーが含まれる可能性を考慮して、一部の変数に若干のずれを許容した。

3. 照合の結果について

3.1 照合キーの重複状況

第 2 節で示した方法で、実際に照合をした結果をまとめたのが表 1 である。ここでは、時点の変化などの影響を避けるため、2009 年および 2014 年の全国消費実態調査と 2010 年の国勢調査を照合したケ

⁶ 2 つのデータにおける同一サンプルのレコードの識別についての古典的な研究として Fellegi and Sunter (1969)が知られている。

ースを見る。照合対象となるのは、2009 年の全国消費実態調査の調査世帯 52,787 世帯および 2014 年調査の 51,919 世帯である。

表 1 で「国勢調査」として示されているのは、全国消費実態調査で抽出された各調査単位区に該当する国勢調査の調査区内の世帯数であり、合計の項を見ると 2009 年は 557,638 世帯、2014 年では 633,226 世帯であった。これは、全国消費実態調査の 1 調査世帯当たり 11 世帯程度が対応していることになる。全国消費実態調査では単位区ごとに 12 世帯が抽出されていることから、対応する国勢調査の調査区内に存在する世帯数は 130 世帯程度という計算となる（逆に言えば、この国勢調査の 130 世帯から全国消費実態調査の 12 世帯を探すことになる）。これは、国勢調査の調査区がおおむね 50 世帯になっていること、全国消費実態調査の単位区は国勢調査の 2 調査区で構成されることと整合的な結果である。

表 1 基本世帯属性による照合キーの重複状況

	2009		2014	
	全国消費実態調査	国勢調査	全国消費実態調査	国勢調査
1 世帯	51,437 (97.4%)	466,060 (84%)	50,256 (96.8%)	523,753 (83%)
2 世帯	1,116 (2.1%)	32,052 (6%)	1,388 (2.7%)	35,512 (6%)
3 世帯	193 (0.4%)	15,948 (3%)	232 (0.4%)	19,995 (3%)
4 世帯	41 (0.0%)	10,328 (2%)	28 (0.1%)	14,676 (2%)
5 世帯以上	0 (0.0%)	33,250 (6%)	15 (0.0%)	39,290 (6%)
合計	52,787 (100%)	557,638 (100%)	51,919 (100%)	633,226 (100%)

表 1 の各行はそれぞれの調査において、上で述べた基本照合キーの重複状況をまとめたものである。すでに述べたように、全国消費実態調査では調査区ごとに 12 世帯しか存在しないため、ほとんどの調査区で 1 照合キー1 世帯の状況が達成できている。すなわち、全国消費実態調査程度のサンプル数であれば、住居の状況とコアな世帯員の情報だけで、ほぼユニークに世帯を特定できる状況となる。実際、調査世帯のうちユニークに識別できた世帯（同一の照合キーに対応する世帯が 1 世帯であるケース）は、2009 年調査・2014 年調査それぞれで全世帯 52,787 世帯、51,919 世帯のうちの約 97 パーセントである。一方、国勢調査では、基本的な属性だけでユニークに識別できるケースは全体の 8 割程度であり、調査世帯内の全世帯をユニークに識別するには、追加的な情報が必要な状況である。

この状況を所与として、第 1 ステージとしては、表 1 の全国消費実態調査側で照合キーが「1 世帯」だけに該当している世帯だけを対象として照合作業をする。全国消費実態調査の各調査世帯の照合キーと同一の照合キーを持つ国勢調査の世帯をカウントし、その数が 1 である場合に照合できたケースとみなす。参考のため、国勢調査側では照合キーでユニークに識別された世帯（約 8 割）の世帯に限定せず、対応関係を観察した。

3.2 基本パターンでの照合結果

この第1ステージでの照合結果をまとめたのが表2である。両統計で照合ができた世帯（すなわち基本的な属性が完全に一致する世帯のペアが1組しかいなかったケース）は2009年で16,387世帯、2014年で12,969世帯であった。これは、全国消費実態調査で照合キーによってユニークに識別できていた世帯（表1の1行目の全国消費実態調査の列に相当する世帯）の32%および26%に相当する。

この表2の結果は、使用した世帯情報では単一の世帯を識別するのに不十分であるという可能性を否定するものである。全国消費実態調査の1つの世帯と同じ照合キーを持つ世帯が「2世帯」存在するケースが1,231世帯、1,315世帯、さらに「3世帯」のケースでも618世帯、813世帯にとどまっており、照合ができないケースの中では少数派である。

それに対し、国勢調査と照合できなかった世帯（対応する世帯が1世帯ではないケース）の大部分は、「国勢調査に対応する世帯が存在しないケース」であった。大部分の世帯は、転居等を経験していないと考えられるにもかかわらず、全国消費実態調査の多くの世帯では、全く同じ世帯属性を持っている国勢調査の調査世帯が存在していない状況なのである。

表2 基本世帯属性による照合結果

	2009 全国消費実態調査	2014 全国消費実態調査
該当世帯なし	32,174 (63%)	33,452 (67%)
1世帯	16,387 (32%)	12,969 (26%)
2世帯	1,231 (2%)	1,315 (3%)
3世帯	618 (1%)	813 (2%)
4世帯以上	1,027 (2%)	1,707 (3%)
照合キーが重複していない世帯	51,437 (100%)	50,256 (100%)

3.3 照合キーの拡張

照合できた世帯が全体の約3割という比率は、原理的には大部分の世帯が照合できるはずであることを前提とすれば、極めて低いと言わざるを得ない。照合できなかった世帯の大部分が「該当世帯なし」であることは、世帯の同一性を「完全には確認できない誤り」よりも、むしろ世帯情報を追加しすぎたために「同一世帯を異なる世帯とみなす誤り」が発生していると考えられる。以下では、照合できない理由を想定しながら照合キーを一部変更して、さらなる照合を試みる。その結果をまとめたものが表3である。

まず、照合ができない理由としては少数派であるが、同一の世帯属性を持つ世帯が複数存在するために同一性が完全には確認できないケースへの対応を考える。表2での照合では、全国消費実態調査側は照合キーでユニークに識別できた世帯のみを用いているが、国勢調査側では全ての世帯を用いた。表1で示されたように国勢調査では約2割の世帯が同一の照合キーを持っており、全国消費実態調査側で2009年に5.6%、2014年に7.6%の世帯が照合できない理由となっている。

このケースについては、追加的な世帯員の情報の利用によって、照合キーの重複を可能な限り排除することで照合できるケースが増加する可能性がある。そこで、表2で照合ができた世帯を除外した上で、コアな世帯員以外の情報も活用した照合キーを作り直して、さらなる照合を試みた。

具体的には、15歳から29歳の世帯員の情報も照合キーに加えることで「70歳未満の全世帯員の情報」を使うケース、「90歳未満の全世帯員の情報」（70歳から89歳までの情報も照合キーに加える）ケース、70歳未満の世帯員の就業状態に関する情報を用いるケース、をそれぞれ考慮した。その結果が表3の行A、B、Cである。この変更で影響が大きかったのは、70歳から89歳の情報を追加したケースBである。高齢者のみの世帯でも年齢・性別の情報が活用されるようになり、2009年で1,275世帯、2014年で1,892世帯が追加的に照合できた。

次に、「国勢調査に対応する世帯が存在しないケース」への対応について検討する。表2で示されたように、照合できないケースの大部分はこちらのケースと考えられる。論理的には、全国消費実態調査の調査対象世帯は国勢調査でも調査対象となっているはずであり、国勢調査のデータに含まれないのであれば、次のような可能性が考えられる。

1. 転居により実際に対応する世帯が調査区内に存在しない
2. 全国消費実態調査の調査世帯が国勢調査には協力しなかった
3. 照合キーに関する世帯属性が変更になった
4. 全国消費実態調査もしくは国勢調査に測定誤差がある

このうち、第1のケースについては、調査区を対応づけて国勢調査内で同一世帯を探すという方針そのものが有効ではないことになり、手元の情報では対処はほぼ不可能である。しかし、転居が主たる原因であるとすれば、4年の間隔のある2014年調査でより多くの該当世帯なしのケースが発生するはずである。表2の「該当世帯なし」の割合が2009年では62.6%であるのに対し、2014年では66.6%となっている。2009年と2010年では1年しか離れていないのに対し、2014年は4年差となっている。これは、年1%程度該当世帯が減る計算であり、転居率を考慮すればおおむね妥当な計算になる。言い換えれば、第1の理由で照合できない世帯の割合は高々数パーセントであり、照合できた世帯の割合が低いことの主たる原因とは考えられない。

第2の、世帯が国勢調査に協力していないケースについては、さらに蓋然性は低い。近年では国勢調査であっても一部の調査項目で多くの「不詳」が発生するなどの問題が知られている⁷。世帯員の属性などで不詳が発生したり、そもそも回答を拒否したりしていれば、当然に全国消費実態調査との照合はできない。しかし、相対的に調査負担の大きな全国消費実態調査に協力をした世帯という条件つきであれば、国勢調査に不適切な回答をする世帯の割合はほぼゼロと考えられる。その意味で、ここで観察されたような大きな割合の「該当世帯なし」の現象を説明することはできない。

⁷たとえば、宇南山(2013,2022)を参照。

表3 拡張した照合結果

	2009年		2014年	
	世帯数	比率	世帯数	比率
基本照合キー	16,387	31.0%	12,969	25.0%
A. 15-29歳を含む	323	1%	187	0%
B. 70-89歳を含む	1,275	2%	1,892	4%
C. 職業を含む	101	0%	68	0%
D. 30-69歳のみ	711	1%	1,156	2%
E. 建物の階数を除く	--	--	357	%
F. 住宅の面積の誤差	8,271	1%	6,843	13%
G. 世帯員の年齢誤差	2,386	4%	1,963	4%
合計	29,454	56%	25,435	49%
照合できず	23,333	44%	26,484	51%
全世帯	52,787	100%	51,919	100%

第3のケースは、コアな世帯員の情報をすべて照合キーとしたことが、同一世帯の識別をするには条件が厳しすぎた可能性を示唆する。世帯員の転入や転出、婚姻・離婚・死亡などの異動があれば、コアな世帯員の情報全てが一致するとは限らないのである。そこで、むしろ一致を要求する世帯員の情報を削減することの検討が必要である。

この課題に対応するために、むしろ子供と若年者および70歳以上の高齢者を除外して、30歳から69歳までの世帯員のみを性別・年齢を照合キーとしたケースも検討した。それが、表3のケースDである。さらに、住居の「建物全体の階数」と「居住階」についても、照合キーから除外したものがケースEである。これは、建物の階数の数え方が、地下を含むかどうかなどで変化する可能性に対応するものである。しかし、このように「同一世帯」の判定条件を緩和しても、新たに照合できる世帯は2%程度であり、大きな影響をもたらさない。

3.4 測定誤差の考察

こうした条件の緩和にもよらず、世帯属性が一致する世帯が見つからないことから考えられるのが、第4のケースである「測定誤差」の可能性である。すなわち、同一の世帯が国勢調査と全国消費実態調査で異なる情報を回答している可能性である。測定誤差のケースには、全国消費実態調査の情報が誤っているケース、国勢調査の情報が誤っているケース、両方が誤っているケースが考えられるが、ここではそれらのケースの区別をすることは考えない。

測定誤差の第1として、住居の床面積について注目する。全国消費実態調査では「住居の床面積」の実数（平方メートル）を記入することが求められているのに対し、国勢調査では階級値で調査されてい

る。ここでは、両者を比較可能なように全国消費実態調査の数値を階級化しているが、階級値を直接聞いた場合とでは回答が異なる可能性がある。たとえば、住居の床面積が69平方メートルであった場合、実数を回答する際に四捨五入によって「70平方メートル」と回答する可能性は十分にある。一方で、階級値で答える際には「60～70平方メートル未満」を回答することは自然である。この場合、原理的には全国消費実態調査の結果を「70～80平方メートル未満」の階級値に換算することで、両者が一致しなくなる。この問題に対応するため、住居の床面積の階級値を前後1階級まで誤差を認めて照合したものが、ケースFである。このような誤差を許容することの影響は大きく、新たに15%程度の世帯を照合することができた。

さらに、世帯員の年齢についても誤差を認めたものがケースGである。年齢については、国勢調査では生年月を調査しているのに対し、全国消費実態調査では満年齢を調査している。両者を照合するために、全国消費実態調査の実施された時点での満年齢を計算した。しかし、誕生日前後で回答する年齢がずれる可能性があり、誤差が生じうる。さらに、満年齢の分布は5歳ごとにピークを持つことが知られており、丸め誤差の可能性もある。しかし、年齢は世帯を特定する非常に重要な情報であるため、あまり柔軟に誤差を認めてしまうと異なる世帯を同一世帯とみなす誤りが多く発生する可能性がある。

そこで、ここでは限定的に世帯員1名だけが前後1歳の誤差があった場合のみ許容することとした。たとえば、複数の世帯員の年齢が誤っていたり、5歳刻み・10歳刻みと大きな誤差が生じていたりするケース（たとえば「33歳」の世帯員を「35歳」もしくは「30歳」と回答するようなケース）は含めていない。世帯員の年齢に誤差を認めても、新たに照合できる世帯は4%程度である。

こうした状況をすべて加味すると、照合できた世帯の割合は約半数になる。基本の照合キーを用いたケースに比べると25%の世帯が追加で照合できた。しかし、約半数の世帯が照合できない状況は、何らかの測定誤差を示唆するものと考えられる。しかし、世帯の提供した情報を大規模に変更して照合することは不適切と考えられる。そのため、今回は最終的に照合できた世帯(2009年が29,454世帯、2014年が25,435世帯)を照合できた世帯とする。

4. 照合された世帯の属性

4.1 世帯人員別の結果

基本の照合キーに加え、上記のAからGまでの拡張した方法で照合された世帯の特徴を明らかにするために、照合された世帯とされなかった世帯の世帯属性を比較する。

まず、世帯人員別に照合結果を比較したものが表4である。2009年については、単身世帯と5名以上の世帯で照合できた比率が低めになっている。これは、照合キーの拡張について検討した事項を考慮すれば自然な結果である。単身世帯では、照合キーとなる変数が限定的であるため、そもそも国勢調査では各世帯をユニークに識別できていない可能性がある。たとえば、同じ間取りのアパートに同じ年齢・性別の個人が単独で居住していれば、同一世帯を同定することは不可能なのである。逆に、5名以上の世帯人員がいるケースでは、世帯員の転入や転出の可能性が高く、年齢の測定誤差なども発生しやすくなるため、照合が困難であると考えられる。

一方で、2014年では全般として照合確率が低くなっている。これは調査どうしの間隔が4年になることで、世帯属性が変化する可能性が高まった結果と考えられる。おおむね世帯員数が多くなるほど照合確率が低くなっていることと整合的な結果である。

このように世帯人員で、若干の照合確率の差が存在しているが、その差は必ずしも大きくなく、照合されたデータにおいて世帯人員別の分布を大きく歪ませるようなものではなく、分析する上で大きな支障とはならないと考えられる。

表 4 世帯人員別の照合結果

	2009 年			2014 年		
	世帯数	照合された世帯	比率	世帯数	照合された世帯	比率
単身世帯	3,959	2,029	51%	4,298	2,182	51%
世帯人員 2 名	18,844	10,720	57%	20,523	10,718	52%
世帯人員 3 名	12,241	6,946	57%	11,772	5,476	47%
世帯人員 4 名	11,366	6,504	57%	10,299	4,925	48%
世帯人員 5 名以上	6,377	3,255	51%	5,027	2,134	42%
合計	52,787	29,454	56%	51,919	25,435	49%

4.2 住宅の所有形態別の結果

ここでの照合では、世帯の居住する住宅が重要な情報として使われていた。照合キーの一部として、住宅の建て方・広さ・所有状況が用いられていたからである。そのため、照合ができた世帯の住居の状況は、全体の状況と比べて偏る可能性が存在する。その偏りの大きさを確認したのが、表 5 である。

全国消費実態調査では、約 8 割が持家世帯であるが、その持家世帯は相対的に照合確率が高くなっている。この理由として考えられるのは、第 1 に転居の確率である。ここでの照合方法では調査区をベースとしているため、転居をした場合には照合は不可能である。一般に、持家世帯の方が賃貸住宅等に居住する世帯よりも転居する確率は低くなると考えられ、照合確率は高くなると予想できる。上でも述べたように、2014 年の照合確率が 2009 年よりも低くなっているのは、この要因によるものと考えられたが、その性質は住宅の所有形態によっても確認できるのである。

一方で、2010 年の国勢調査の 1 年前である 2009 年の持家世帯と民営借家において 18%もの照合確率に差が存在していることは、転居だけでは説明できない。こちらについては、住宅に関する情報（正確な面積や構造）に対する理解が持家世帯の方が高いことも寄与していると考えられる。

表5 住宅の所有形態別の照合結果

	2009年			2014年		
	世帯数	照合された世帯	比率	世帯数	照合された世帯	比率
持家	42,787	25,177	59%	43,169	22,677	53%
民営借家	5,977	2,422	41%	5,072	1,551	31%
公営住宅	1,984	1,023	52%	1,952	785	40%
公団・公社住宅	709	364	51%	444	218	49%
社宅・公務員宿舎	1,254	466	37%	999	201	20%
借間	50	2	4%	263	3	1%
寮・寄宿舍	26	0	0%	20	0	0%
合計	52,787	29,454	56%	51,919	25,435	49%

4.3 世帯主の職業別の結果

ここでの照合では、時点によって変化する世帯員の情報は照合キーとしては利用しなかった。世帯員の就学状況・就業状況が全国消費実態調査と国勢調査の両方で調査されているが、年齢や性別と異なり、こうした変数は調査時点によって変化する可能性があるため、照合には使えないのである。逆に、照合キーとして用いていないのであれば、基本的には、照合確率と就業状況などは相関を持たないはずである。それを確認しているのが表6である。

この表は、全国消費実態調査の時点における世帯主の職業別に照合確率を示したものである。全国消費実態調査における世帯主の職業としては、いわゆるホワイトカラー労働者である「民間職員」が約2割で最多となっており、ブルーカラー労働者である「常用労務作業員」がほぼ同じ程度存在している。それ以外にも、公務員である「官公職員」や自営業者である「商人および職人」や「個人経営者」などさまざまな職業が存在しており、それぞれごとの照合確率を計算したものである。

計算された照合確率は、職業には依存せずおおむね一定である。これは、照合結果に特定の偏りが存在しないことを示唆しており、照合方法の妥当性を示すものと考えられる。差は小さいが、自営業者において照合確率がやや低くなっている。これは、住宅に関する事項のうち、事業に係る部分の切り分けで測定誤差が生じている可能性を示唆している。

表 6 世帯主の職業別の照合結果

	2009 年			2014 年		
	世帯数	照合された世帯	比率	世帯数	照合された世帯	比率
常用労務作業	10,720	5,502	51%	10,776	4,729	44%
臨時及び日々雇労務者	793	409	52%	912	382	42%
民間職員	12,512	7,421	59%	11,065	5,378	49%
官公職員	5,329	3,214	57%	4,704	2,303	36%
商人および職人	4,548	2,201	48%	3,907	1,700	44%
個人経営者	293	149	51%	256	103	40%
農林漁業従事者	1,680	902	54%	1,474	666	45%
法人経営者	1,596	896	56%	1,662	872	52%
自由業者	747	423	57%	787	368	47%
その他	85	39	46%	7	3	43%
無職	14,484	8,298	57%	16,369	8,931	55%
合計	52,787	29,454	56%	51,919	25,435	49%

4.4 都道府県別の結果

最後に、都道府県別の照合確率についても確認をしておく。都道府県別に差が出るとすれば、転居の確率である。首都圏や近畿圏では、相対的に賃貸住宅に居住する若年世帯が多く、転居をする確率が高いと考えられるため、照合確率が低くなると考えられる。一方で、地方部では、3世代同居など世帯人員の多い世帯の割合が相対的に高くなると考えられる。すでに見たように、世帯人員が5名を超えるような世帯では照合確率が低くなる。こうした状況を考慮すると、事前的には都道府県別に大きな差があるかは明らかではない。

表7は、実際に都道府県別に照合確率を計算したものである。この表では、都市部でやや照合確率が低くなっていることが確認できるが、全体としては大きな差は存在していない。2009年の岡山県などいくつかの県では特異的に低い照合確率となっているが、偶発的な要因である可能性が高い。ただし、沖縄県については2009年と2014年ともに他の都道府県と比べて低い水準になっている。この結果について、原因は明らかでないため、更なる分析が必要である。

表7 都道府県別の照合結果

	2009年			2014年		
	世帯数	照合された世帯	比率	世帯数	照合された世帯	比率
北海道	2,187	1,227	56%	2,098	977	47%
青森県	746	457	61%	690	346	50%
岩手県	729	423	58%	734	330	45%
宮城県	771	442	57%	738	382	52%
秋田県	703	405	58%	735	370	50%
山形県	722	437	61%	723	366	51%
福島県	926	540	58%	846	392	46%
茨城県	1,388	854	62%	1,384	728	53%
栃木県	879	563	64%	835	436	52%
群馬県	938	553	59%	845	445	53%
埼玉県	2,739	1,657	60%	2,771	1,518	55%
千葉県	2,567	1,417	55%	2,654	1,403	53%
東京都	2,347	1,289	55%	2,397	1,137	47%
神奈川県	2,658	1,381	52%	2,551	1,339	52%
新潟県	957	571	60%	940	475	51%
富山県	743	452	61%	725	400	55%
石川県	739	425	58%	733	327	45%
福井県	718	365	51%	737	388	53%
山梨県	713	379	53%	715	354	50%
長野県	1,015	598	59%	973	467	48%
岐阜県	972	563	58%	932	455	49%
静岡県	1,467	854	58%	1,556	817	53%
愛知県	2,406	1,416	59%	2,441	1,263	52%
三重県	880	472	54%	904	468	52%
滋賀県	738	434	59%	728	381	52%
京都府	819	419	51%	759	369	49%
大阪府	2,802	1,497	53%	2,718	1,307	48%
兵庫県	2,140	1,293	60%	2,061	982	48%
奈良県	723	399	55%	696	336	48%
和歌山県	719	389	54%	713	319	45%
鳥取県	703	360	51%	727	343	47%
島根県	706	396	56%	733	314	43%
岡山県	845	295	35%	772	373	48%
広島県	1,171	654	56%	1,049	520	50%
山口県	873	498	57%	787	392	50%
徳島県	695	350	50%	733	341	47%
香川県	736	403	55%	751	355	47%
愛媛県	764	441	58%	732	347	47%
高知県	715	345	48%	707	306	43%
福岡県	1,621	900	56%	1,568	752	48%
佐賀県	684	389	57%	736	370	50%

長崎県	725	372	51%	698	306	44%
熊本県	659	341	52%	701	208	30%
大分県	733	409	56%	734	332	45%
宮崎県	734	379	52%	729	345	47%
鹿児島県	829	450	54%	738	335	45%
沖縄県	743	301	41%	692	219	32%
合計	52,787	29,454	56%	51,919	25,435	49%

5. まとめ

本稿では、同一世帯における国勢調査と全国消費実態調査に対する回答を統合できるよう、調査票レベルでのデータの照合を試みた。各調査では個人を識別できる情報を収集していないため、世帯員の年齢や性別などの属性が同じ世帯を「同一」とみなす方法で照合した。原理的には重複がありえる世帯属性を照合キーとできるよう、調査区の情報を用いて、潜在的な照合対象世帯を限定した。

この方法によって、全国消費実態調査の調査対象世帯のうち約半数が照合できることが分かった。2人以上の世帯や持家世帯など、転居の可能性が低い世帯類型ほど照合できる可能性が高かった。また、全国消費実態調査の実施時点が、国勢調査の1年前のケースと4年後のケースでは、1年前の方が照合できる可能性が高かった。今後は、同様の照合方法を家計調査や他の年次の全国消費実態調査・国勢調査にも適用し、ここでの結果がどの程度頑健なものであるかを確認することを目指す。

また、照合ができたデータを用いた分析を進めることも、次の重要なステップである。国勢調査の情報が付加されたことで、世帯員の生年月・学歴・転居の状況などが観察可能となっている。たとえば、生年月の情報を使えば、「早生まれ」の個人の収入や支出を分析することができる。早生まれについては、Bedardv and Dhuey (2006)や川口(2007)などで分析されてきたが、家計収支の詳細な情報とセットで観察できるデータは存在しておらず、ここで構築したデータによって新たな発見が期待できる。また、2019年に全国消費実態調査に代わり新たに調査されるようになった全国家計構造調査では、家計収支と共に学歴も調査されるようになった。ここでのデータを活用することで、過去の学歴別の家計収支の状況と比較することができる。ただし、こうした分析は照合方法を確立した上での課題であり、今後の課題である。

参考文献

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. 2021. "Automated Linking of Historical Data." *Journal of Economic Literature*, 59 (3): 865-918.

Bedard, Kelly, and Elizabeth Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics*, 121 (4): 1437-72.

Fellegi, Ivan P. and Alan B. Sunter. 1969 "A Theory for Record Linkage," *Journal of the American Statistical Association*, 64(328): 1183-1210. (DOI: 10.1080/01621459.1969.10501049)

Freeth, Stephanie, Amanda White, and Jean Martin (2001) "USING MATCHED CENSUS-SURVEY RECORDS TO EVALUATE THE QUALITY OF SURVEY DATA," *Proceedings of Statistics Canada Symposium 2001 Achieving Data Quality in a Statistical Agency: A Methodological Perspective*. (<https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2001001/session12/6270-eng.pdf?st=NdGh2XYw>)

Newman, Thomas B. and Brown, Andrew N. (1997) "Use of Commercial Record Linkage Software and Vital Statistics to Identify Patient Deaths" *Journal of the American Medical Informatics Association*, vol. 4(3) pp. 233-237.

Winkler, William E. (2006) " Overview of Record Linkage and Current Research Directions" US Census Working Paper *Statistical Research Reports and Studies* RSS2006-02.

宇南山卓 (2013) 「仕事と結婚の両立可能性と保育所：2010 年国勢調査による検証」 RIETI Discussion Paper Series 13-J-039

宇南山卓 (2022) 「国勢調査」『日本労働研究雑誌』 No.741. pp. 6-9.

川口大司・森啓明 (2007) 「誕生日と学業成績・最終学歴」『日本労働研究雑誌』 No.569. pp. 29-42.