調査報告

国勢調査の「不詳」増加がもたらす統計地図の歪みの可視化

Visualizing the Distortions of Statistical Maps Caused by the Increase in Unreported Cases
in Japan's Population Census

山本 涼子・埴淵 知哉・中谷 友樹・山内 昌和

YAMAMOTO Ryoko, HANIBUCHI Tomoya, NAKAYA Tomoki and YAMAUCHI Masakazu

(2020年10月5日受付 2020年12月1日受理)

本研究のねらいは、国勢調査の「不詳」が増加したシナリオを立て、データから観測される指標の地域差がどのように本来の姿からずれてしまうのかを可視化することである。東京大都市圏の市区町村を対象に、(1) 年齢階級別の比例按分と(2) 属性間の不詳率比に基づく重み付け按分を用いて、2015年国勢調査による未婚率と短期居住率を不詳率の異なるシナリオ別に推計した。その結果、いずれの指標も不詳率が0のシナリオでは都心部で高く周辺部で低い地域差が明瞭であったが、不詳率が1.5倍や2倍に増加したシナリオで推計した地図を見ると、そういった地域差は縮小・消失して見えることが可視化された。このことは、今後さらに「不詳」が増加することで、疑似的な地域差が生じる蓋然性が高まることを意味する。不詳率の高い項目や地域において国勢調査を利用する場合には、結果の慎重な解釈や補正データの利用が求められる。

The purpose of this study was to visualize how regional differences observed from the population census deviate from their “true” forms due to an increase in unreported (missing) cases by considering multiple scenarios in which there was an increase in the number of unreported cases in the census. For municipalities in the Tokyo metropolitan area, we estimated the percentages of unmarried people and those with short-term residence from the 2015 Population Census for different scenarios of unreported rates using 1) proportional distribution by age-group and 2) weighted distribution based on the ratio of unreported rates among attributes. In both indicators, the results showed that there were clear regional differences with high and low percentages in urban centers and suburban areas, respectively, in a scenario where the unreported rate became zero. However, when we examined the maps estimated under scenarios where the unreported rates increased by 1.5- or 2-fold, such regional differences appeared to be reduced or eliminated. This means that further increases in the number of unreported cases will increase the probability of spurious regional differences in the future. Careful interpretation of the results and use of statistically corrected data are required when analyzing the census data for items and regions with high unreported rates.

キーワード: 国勢調査, 「不詳」, 比例按分, 疑似的な地域差

Key words: population census, unreported cases, proportional distribution, spurious regional differences

I はじめに

国・自治体の政策立案や地理学をはじめとする大学・研究機関の研究等で頻繁に利用されるデータとして、国勢調査がある。日本国内常住者に対する唯一の全数調査であり、小地域レベルでの集計結果も利用可能なことから、地域分析に有用なデータとして活用されてきた。また、きわめて高い回収率も国勢調査の統計としての有用性・信頼性を高める一因であった。しかし、近年では調査票の未回収および回答内容の不備等による「不詳」が増加しており、地域分析への影響が懸念されている(阿部 2013; 小池・山内 2014)。国

勢調査に対する懸念は日本だけにとどまらず、対象者の過少把握による精度低下の問題などが海外でも取り上げられている(Martin 2010; O'Hare 2016)。

「不詳」の増加は属性別人口など実数の把握を困難にするほか、失業率や未婚率のような地域特性を表す代表値を実際よりも過小または過大に見積もるリスクを高める。たとえば、「不詳」は若年層や未婚者など特定の属性に多いと見られている(阿部 2013; 山田 2016)。宇南山(2013)は、人口動態統計を利用した補正推計によって、2010年国勢調査の20~39歳の未婚率が過小評価されていることを指摘した。埴淵・山内(2019)も、調査票の未提出が若年層や未婚者、単

身世帯、短期居住者に多いことや、プライバシー意識よりも国勢調査の理解度が強く関連していることを、インターネット調査を基に明らかにした。

むしろ、国勢調査の回収率低下や「不詳」の増加を受けて、回収・回答状況の改善に向けた取組みも進められている。国勢調査の実施主体である総務省統計局は、調査票の郵送回収（2010年）やインターネット回答（2015年）を全国的に導入した。これらの回答方法は、プライバシー保護の観点から回答状況の改善につながる可能性がある。一方で、従来行われてきた国勢調査員による記入状況の確認ができないために、回答内容の不備等を見落としやすくなり、項目によっては「不詳」を増加させることもありうる（小池・山内 2014）。実際、2015年国勢調査では一部の項目で不詳率の低下が見られたものの、大半の項目では上昇が続いており、労働力調査など他の統計と比べても「不詳」は格段に多かったとされる（山田 2016）。つまり現状では「不詳」の増加を止めるには至っておらず、この傾向が続けば国勢調査の統計としての信頼性は徐々に損なわれるおそれがある。

従来の研究では、将来的な「不詳」増加への懸念は一樣に示されているものの、それが具体的にどのような問題を引き起こすのかは十分に議論されていない。たとえば埴淵・村中編（2018）は、不詳率に大きな地域差があることから、国勢調査データに地理的な系統誤差が紛れ込み、疑似的な地域差や地域相関を観測してしまう可能性に言及した。ただし、実際にそれがどこで、どのように表出するのかを描き出すには至っていない。「不詳」増加が進むことで、従来指摘されてきた未婚率等の過小評価や、それらの地理的分布は今後どのように歪められるのだろうか。これを定量的に把握することは、「不詳」問題をわかりやすく可視化し、その深刻さを推し量るためにも重要であろう。

こういった「不詳」増加による影響を計算することは、推計手法の観点からみれば、「不詳」が0になる状態を推計すること、すなわちデータ補正を行うことと同じ枠組内で議論できる。同じ仮定のもとで、「不詳」が増えたり減ったりした状態を計算することにほかならないからである。しかし従来、「不詳」の補正を行った研究はあっても（阿部ほか 2018）、それを「不詳」増加に当てはめて問題を議論したものは管見の限り見

られない。そこで本研究では、「不詳」が将来さらに増加するシナリオを想定し、その場合にデータから算出される地域代表値（具体的には未婚率と短期居住率）が、「不詳」がない場合の本来の値とどのように、そしてどの程度ずれてしまうのかを推計・可視化することを目的とする。

なお、よく利用される補正方法として年齢階級別の比例按分があるが、この方法についても議論の余地がある。この方法では、年齢階級別に属性の割合（たとえば未婚率）を計算し、その値に基づき年齢階級別の「不詳」人口を比例按分した上で全体を合計するという手順をとる。こうすることで、配偶関係「不詳」人口が多い若年層において未婚率が高い、という傾向を考慮した補正がなされるため、年齢を考慮せずに全体を比例按分した場合よりも実際に近い値を得ることができると考えられる。ただし、配偶関係における「不詳」の生じやすさは年齢（およびその他の周辺情報）によって異なるだけでなく、当該の項目そのものの影響を受けている場合もありうる。

そこで本研究では、こういった周辺情報を利用した補正・推計とは異なるアプローチとして、属性間の不詳率比を直接仮定する方法もあわせて検討する。ただし「配偶関係別の配偶関係不詳率」のような属性間の不詳率比を、国勢調査のデータそのものから得ることは論理的にできない。しかし何らかの別のデータを根拠に、属性ごとの不詳率をあらかじめ仮定（たとえば男性は女性に比べて x 倍「不詳」になりやすい、など）することができれば、その条件を満たすように「不詳」人口を重み付けして按分することが可能である。著者らはこの属性間の不詳率比による方法に基づき、不詳率が増加した場合の未婚率を推計し、それが地図上でどのように違った印象を与えるのかを紹介した（埴淵ほか 2020）。ただし、そこでは一つの項目（未婚率）について単一のシナリオに基づく試算結果を例示したにすぎず、方法についても簡潔に述べるにとどまっている。

以上の問題意識を基に本研究では、「不詳」補正後、つまり不詳率が0になった場合と、逆に不詳率が現在よりも高まった場合のそれぞれについて属性の地理的分布（具体的には未婚率および短期居住率の市区町村別分布）を推計し、両者を比較する。それにより、

「不詳」の増加がどのように諸事象の地理的分布を歪めて表現してしまうのかを、視覚的に示すことを試みる。その際に、年齢階級別の比例按分と属性間の不詳率比を利用した二つのアプローチを併用する。これらを通じて「不詳」が増大する国勢調査データの問題点を浮き彫りにするとともに、その適切な分析や解釈に役立つ基礎的知見を得ることが本研究のねらいである。

II 対象とする指標とシナリオ設定

1. 対象とする指標と不詳率の考え方

本研究では、2015年国勢調査のデータを用いて、不詳率が高い東京大都市圏（一都三県）の市区町村を対象に分析を行う。「不詳」はほぼすべての項目において見られるが、その中から本研究では（1）配偶関係の未婚と（2）居住期間の短期間居住を取り上げる。これらはともに、都市地理学や人口地理学で頻繁に利用される項目であり、「不詳」研究においても注目されてきた（阿部 2013; 宇南山 2013; 小池・山内 2014; 埴淵ほか 2018）。ただし不詳率を見ると、両者は対照的な値を示す。配偶関係は、国勢調査の調査項目の中では比較的不詳率が低い。これに対して居住期間は、教育（学歴）や職業と同様に不詳率が高い項目に位置づけられる（埴淵・村中編 2018）。なお、2010年において最も不詳率が高いのは教育（学歴）であったが、2015年国勢調査では調査されていない。

ここで、本研究で使用する指標について簡単に説明しておきたい。表1に各指標と属性および「不詳」と非「不詳」の関係をまとめて整理した。配偶関係のうち、本研究では未婚に注目するため、表ではそれ以外の有配偶・離別・死別はまとめて非未婚として示した。理論上、（真の）未婚率は、未婚人口を総数で除した値 $(a+b)/(a+b+c+d)$ である。しかし、観測された国勢調査のデータには「不詳」が含まれており、その属性ごとの数（表1の a および c ）はわからないため、一般的には「不詳」を除くという処理がなされる。これにより、(i) 観測される未婚率は、非「不詳」の未婚人口 b を、非「不詳」の総数、すなわち総数（15歳以上人口）から配偶関係「不詳」を差し引いた数 $(b+d)$ で除した値 $b/(b+d)$ となる。(ii) 配

表1 指標の整理

属性 \ 「不詳」	「不詳」	非「不詳」
未婚	a	b
非未婚	c	d

偶関係の不詳率は、未婚・非未婚の「不詳」の合計が総数に占める割合が $(a+c)/(a+b+c+d)$ として計算される。本研究は、この (ii) が増加したときに、各地域で (i) がどう変化するかを探るものである。

なお、表1は配偶関係の場合を示しているが、居住期間についても指標は同様である。居住期間が5年未満である割合を短期居住率として、1年未満と1～5年未満人口の和を総数（居住期間）から居住期間「不詳」を差し引いた数で除した値に着目する。

なお、表中の b と d 、および $(a+c)$ は国勢調査の集計表から得ることが可能であるが、 a と c それぞれの値は不明である。国勢調査データの「補正」とは、この $(a+c)$ を、何らかの方法を用いて a と c それぞれに分解する作業を意味する。最もわかりやすい例は、 $a:c$ が $b:d$ と同じ比であると仮定して単純に比例按分するものであろう。これは、「不詳を除く」という処理による未婚率等の計算と同じ仮定に基づくものである。これを年齢階級ごとに繰り返した後で合計する処理が、年齢階級別比例按分である。いずれも、属性ごとの不詳率（たとえば未婚であれば $a/(a+b)$ ）には違いがなく、属性間の不詳率比 r （式(1)）が1:1であると仮定している点で共通している。

$$r = \frac{a}{(a+b)} \bigg/ \frac{c}{(c+d)} \quad (1)$$

他方で、この不詳率が属性別に異なる（たとえば2:1）と考え、それに沿うよう「不詳」人口を重み付けして按分するのが本研究のIVにおいて試みる方法となる。

2. 「不詳」の増加に関するシナリオ設定

本研究では、2015年国勢調査の配偶関係および居住期間の不詳率が0から2倍まで変動するシナリオを立て、未婚率と短期居住率が各市区町村でどう変化するかを推計する。具体的なシナリオは、以下に示し

た4通りである。

- A) 補正シナリオ：不詳率=0%の場合に相当
- B) 現況シナリオ：不詳率=観測値（現況）の場合
- C) 1.5倍シナリオ：不詳率が現況の1.5倍の場合
- D) 2倍シナリオ：不詳率が現況の2.0倍の場合

ここでの不詳率とは、配偶関係および居住期間それぞれの不詳率(ii)のことを指す。不詳率0のシナリオAは、真の値を推計した補正值である一方、現況シナリオ(B)から1.5倍(C)、そして2倍シナリオ(D)へと不詳率が増加するにつれて、観測される未婚率や短期居住率の地域差は、不詳率の地域差に起因する系統的誤差を伴う「疑似的な地域差」(埴淵・村中編2018)を示す蓋然性が高くなると予想される。これらのシナリオに基づいて、以下IIIでは年齢別按分処理によって、IVでは不詳率比を仮定する方法によって未婚率および短期居住率の地域差を分析する。

III 年齢階級別比例按分による分析結果

1. 年齢階級別比例按分による分析結果

「不詳」の補正にしばしば使われている方法として、年齢階級別の比例按分がある。たとえば非「不詳」の未婚率が30%だったとすると、「不詳」もそれと同じであるとみなし、「不詳」人口の30%を未婚、残り70%を非未婚に割り振るのが単純な比例按分である。この按分処理を年齢階級ごとに行い、それを合計すれば、全体をまとめて按分するよりも不詳者の年齢構成を考慮した補正值を算出することが可能となる。なお、年齢以外の項目を利用することも可能であり、たとえば総務省統計局による「市区町村別の年齢・国籍不詳按分後の人口」は単身世帯か否かを、阿部ほか(2018)は移動形態の情報を用いて、年齢や国籍「不詳」人口の按分処理を行っている。

ここでは、まずこの年齢階級別按分処理によって補正シナリオ(A)および「不詳」増加シナリオ(C・D)に基づく分析を行う。まず、年齢階級別に配偶関係「不詳」人口を非「不詳」人口における属性構成比(未婚：非未婚)に基づいて比例按分し、全体を合計した上で未婚率を算出する処理を、東京大都市圏の各市区町村(N=251)において行った。これが補正シナリオである。次に、不詳率が増加するシナリオで

は、年齢階級別の「不詳」人口を1.5倍または2倍し、その増加分を差し引いた非「不詳」人口を全年齢階級について合計することで未婚率を算出した。つまり補正とは逆に、非「不詳」人口を「不詳」人口に比例按分する処理である。居住期間の場合も同様に、短期居住(5年未満)とそれ以外(5年以上)の構成比を年齢階級別に計算し、居住期間「不詳」数に構成比を掛けて按分し、非「不詳」人口に加えて合計することで補正值を算出した。また同じく、不詳率が1.5倍または2倍となるシナリオに基づく推計を行った。

なお、年齢階級別に処理する場合、年齢そのものが「不詳」であるケースをどう扱うかという問題がある。というのも、年齢「不詳」人口については配偶関係が不明であるため、その構成比を基に按分することができないからである。居住期間についても、年齢「不詳」は全数が居住期間「不詳」であるか、非「不詳」があったとしてもごく少数しかない市区町村が多いことから、これを一つの年齢階級として扱い按分すると極端な値をとるおそれがある。そこで本研究では、総務省統計局が公表している「市区町村別の年齢・国籍不詳按分後の人口」を利用し、年齢「不詳」はすべて配偶関係「不詳」・居住期間「不詳」とみなして按分処理する方法を採用した。つまり、まず年齢「不詳」を按分補正し、次に配偶関係・居住期間「不詳」を年齢階級別に按分するという2段階の処理を行った。

2. 未婚率の分析結果

図1と図2は、各市区町村の未婚率をシナリオごとに示したものであり、同じ指標を異なる階級区分を用いて地図化したものである。図1では、階級区分はシナリオAで得られた値による等間隔分類を共通して用いている。同じ基準にすることで、値の絶対的な変化を把握しやすくするためである。各シナリオにおける推計結果の値を用いた階級区分図は図2に示した。こちらの図からは、各シナリオにおける値の相対的な分布の違いを読み取ることができる。ただし、シナリオA～Dで未婚率の最大値と最小値は大きく変化しておらず、その差は0.5%未満であることから、図1と図2にはほとんど違いが見られない結果となった。なお、図には中央値もあわせて記載している。

図1 年齢階級別比例按分による未婚率の補正・推計結果（シナリオAの結果による等間隔分類）

図2 年齢階級別比例按分による未婚率の補正・推計結果（各シナリオの結果による等間隔分類）

いずれの区分方法においても、シナリオAからシナリオDへと不詳率が増加するにつれて特に色の濃い地域が減り中央値に近い地域が増えている。すなわち不詳率増加により地域差が小さく見える様子が表現されているが、その変化は顕著とはいえない。そこで、図3に未婚率のシナリオAとDの差分を示した。これを見ると不詳率増加による過小推計の様子がわかりやすい。一都三県全体で過小推計は見られるものの、都心部での差分が相対的には大きくなっている。「不詳」は若年層で多く、その層では未婚率も高いことから、「不詳」が増えるとそこにより多くの未婚者が含まれることになる。そのため、「不詳」を除いて算出される未婚率は過小推計されやすく、それは特に東京都心部で顕著であることがわかる。

3. 短期居住率の分析結果

図4は短期居住率を各シナリオについて示したもの

図3 年齢階級別比例按分による未婚率の推計値の差（シナリオD-A）

であり、階級区分はシナリオAで得られた値による等間隔分類に基づく。図5は、同じ分析結果をシナリオごとに階級区分して示したものである。ただし、シナリオA～Cまでは最大値と最小値に変化がないことから、図4と同じ地図になっている。一方、図5のシナ

図4 年齢階級別比例按分による短期居住率の補正・推計結果（シナリオAの結果による等間隔分類）

図5 年齢階級別比例按分による短期居住率の補正・推計結果（各シナリオの結果による等間隔分類）

リオDは全体的に色の濃い地図となっており印象が異なるが、これは推計された最小値が0であり、階級区分値が他のシナリオとは大きく変わったためである。なお、短期居住率が0と推計されるのは、不詳率を2倍にすると100%近い値になるような都市であり、今回の計算方法をそのまま適用すると計算上マイナス値をとることがあるため、その場合には短期居住率を0として処理した。

図には、シナリオAからDにかけて、不詳率が増加するにつれて地域差が小さく見える様子が表現されている。未婚率と比べると、その変化は顕著である。「不詳」は年齢が低い層で多く、その層で短期居住率が高いことから、「不詳」が増えるとその中に含まれる短期居住者が多くなる。そのため、不詳率が高くなるシナリオでは、算出される短期居住率の過小推計が起きていると考えられる。傾向は未婚率の場合と類似

図6 年齢階級別按分処理による短期居住率の推計値の差（シナリオD-A）

しているものの、元の不詳率が全体的に高いことから、この効果が強く結果に反映されたといえるであろう。短期居住率のシナリオAとDの差分をとった地図（図6）は、不詳率増加による過小推計が東京都心部

できわめて大きいことを示している。

このように、年齢階級別に「不詳」の按分処理を行い、補正・推計値を得ることによって、若年層・未婚者・短期居住者そして「不詳」がいずれも多い東京都心部を中心に、不詳率の上昇が各指標の過小推計を招くという結果が示された。年齢は不詳率と強く関連する要因であり、埴淵・山内（2019）は未婚や短期居住と調査票未提出との関連性が年齢によって説明されることを示唆している。これは、按分処理に際して年齢を考慮することの妥当性を示す一つの根拠である。

しかしそれ以外の因子、特に回答カテゴリ自体が「不詳」の発生と相関をもつ場合、この方法では十分とはいえない。たとえば短期居住者が短期居住であることによって「不詳」になりやすい（調査員と顔見知りでないために回答への誘因が働きにくいなど）のであれば、年齢階級別に按分するだけでは不十分であろう。このような観察できる変数の組合せだけでは「不詳」発生を予測できない場合には、別の調査結果と組み合わせた推論を施す必要がある。そこで、以下のIVでは先行研究（埴淵・山内 2019）の調査から得た不詳発生率の推計値を利用して、未婚の不詳率が他の属性の不詳率よりも高いことを直接的に反映させる方法を検討する。

IV 不詳率比を用いた按分による分析結果

1. 属性間の不詳率比を用いた重み付け按分方法

重み付け按分法を用いた分析は、未婚と有配偶など「属性間の不詳率に差がある」という仮定の基に行う。その仮定の根拠となる先行研究（埴淵・山内 2019）では、国勢調査が未提出となりやすい属性の傾向を明らかにしている。アンケート調査による個票データをロジスティック回帰分析により統計学的に検討したところ、未提出と有意な粗オッズ比が示された属性は、未婚、一人暮らし、集合住宅の借家、短期間居住などであった。本研究で対象とする配偶関係については、未提出率は有配偶で低く、未婚の場合に高い値を示した。そこで、推計にあたっては、未婚とそれ以外（有配偶・離別・死別）の属性間の不詳率比（式(1)）が2:1であるとする仮定を置く。これは、先行研究（埴淵・山内 2019）において得られた未提出率が未婚で

14.2%、未婚以外では6.4%（有配偶と離別・死別）であったことを根拠としたものである。

また居住期間については、年数が短いほど未提出率が高くなり、他の属性と大きな差が見られることが報告されている。そこで本研究では、短期居住（5年未満）とそれ以外の属性との不詳率の比を2:1と仮定する。これは埴淵・山内（2019）によって示された未提出率が、居住期間5年未満で17.4%、5年以上（「出生時から」を含む）で7.6%であったことに基づく設定である。むしろ先行研究の不詳率比をそのまま使用することも可能であるが、年齢層やサンプル数の限られた標本調査による値であるため、ここではシナリオ別の分析結果の解釈と項目間の比較の容易さを優先し、いずれも2:1という単純な仮定とした。なお、この不詳率比は配偶関係全体の「不詳」数が大きくなっても一定であり、地域ごとの差もないものとして考える。

この仮定のもと、不詳率比の条件を満たすように「不詳」人口を重み付け按分し、シナリオごとに各市区町村の未婚率・短期居住率を求める。以下はその計算方法を、未婚率を例として式に示したものである。

未婚率を求めるためには、未婚人口と配偶関係全体の総数および「不詳」数の数値が必要である。配偶関係の各属性人口の観測値 O_{ij} と「不詳」数 U_{ij} ・不詳率 x_{ij} には、以下の関係式が成り立っている。

$$(O_{ij} + U_{ij})x_{ij} = U_{ij} \quad (2)$$

$$U_{\text{all},j} = \sum U_{i,j} \quad (3)$$

ここで、 i は属性を表す添え字で、all: 総数、a: 未婚、b: 未婚以外の属性（有配偶、離別、死別）とする。 j はシナリオを表し、A: シナリオA、B: シナリオB、C: シナリオC、D: シナリオDである。なお、式(2)の $O_{ij} + U_{ij}$ は各属性の総数を示しており、シナリオによらず一定である。

属性間の不詳率比の仮定は次の通りである。

$$x_{a,j} = 2x_{b,j} \quad (4)$$

これら式(2)～(4)を満たす $x_{a,j}$ を求めることにより、未婚人口の「不詳」数 $U_{a,j}$ を得ることができ、未婚率 S_j が式(5)によって計算可能となる。

$$S_j = \frac{O_{a,j}}{O_{all,j} - U_{all,j}} = \frac{O_{a,A} - U_{a,j}}{O_{all,j} - U_{all,j}} \quad (5)$$

シナリオAの未婚人口 $O_{a,A}$ が必要となるため、まずは各属性の観測値 $O_{i,B}$ と配偶関係の「不詳」数 $U_{all,B}$ が既知であるシナリオBにおいて、各属性の不詳率 $x_{i,A}$ を求める。これにより、各属性の「不詳」数 $U_{i,B}$ が計算できるため、シナリオBの観測値と「不詳」数の和がシナリオAの観測値 $O_{i,A}$ となる。

不詳率増加シナリオについては、まず式(3)の左辺である配偶関係の「不詳」数 $U_{all,j}$ を以下の式より求める必要がある。

$$U_{all,j} = (O_{all,B} + U_{all,B})\alpha x_{all,B} = \alpha U_{all,B} \quad (6)$$

ここで、 α は不詳率1.5倍のシナリオCならば1.5、不詳率2倍シナリオDならば2の値をとる。

2. 未婚率の分析結果

各シナリオについて市区町村別の未婚率を推計し、地図化したものが図7である。階級区分はシナリオAで得られた値の等間隔分類による。シナリオAの地図では東京都心部で未婚率が高く、そこから郊外、周辺地域へと低下していく様子が明瞭に見てとれる。一方B～Dへと不詳率が現況から増加していくシナリオ別の結果を見比べると、全体的に色が薄くなり、地域的な傾向を読み取ることが困難になる。各シナリオの結果の値を用いて階級区分した場合の図8でも、不詳率が高くなるにつれて、未婚率の高い地域はばらついていくことがわかる。極端に未婚率の高い地域は減り、中央値に近い地域が広がる結果となった。

全体の傾向としては、未婚率が高いはずの都市部ほど過小推計されることになり、地域差が本来のものよりも小さくなったり、なくなったりしたように見える結果となった。東京都心部などの未婚率が高い地域は

図7 不詳率比を用いた未婚率の補正・推計結果（シナリオAの結果による等間隔分類）

図8 不詳率比を用いた未婚率の補正・推計結果（各シナリオの結果による等間隔分類）

不詳率も高い傾向にあるため、シナリオDでは未婚率が低下しつつ平準化され、全体として都心と周辺部の差が見えにくくなったものと理解できる。これは、「不詳」が増加することで統計地図が本来の姿から乖離していくことを意味する。図9はシナリオAとDの差分をとったものであり、ここで指摘したような未婚率の過小推計が起る市区町村の分布傾向を明確に示している。

この現象は、シナリオA～Dの不詳率増加に伴う各都市の未婚率の推移をグラフ化するとさらにわかりやすい。図10は、横軸にシナリオBを1としたときの各シナリオの不詳率を、縦軸に未婚率をとったグラフである。各都県において配偶関係不詳率が最も高い市区（東京都豊島区、神奈川県横浜市中区、千葉県成田市、埼玉県さいたま市大宮区：図の実線）および最も低い町村（東京都青ヶ島村、神奈川県山北町、千葉県長生村、埼玉県ときがわ町：図の点線）と、一都三県で不詳率が2番目に高い東京都港区を抽出して例示した。

たとえば、不詳率が0の場合（シナリオA）では東京都豊島区のほうが青ヶ島村よりも未婚率が高いはずであったが、不詳率2倍のシナリオDでは逆転して見えている。東京都港区とさいたま市大宮区との間にも同様の現象が起きている。さらに、シナリオAで見ると東京都港区と神奈川県横浜市中区の未婚率にほとんど差はないが、不詳率が2倍になったシナリオでは5%ほどの差が見られる。つまり、（真の）地域差がないはずであるにもかかわらず、「不詳」が増加することで未婚率に地域差があるように（疑似的に）見えてしまう可能性が示されている。

3. 短期居住率の分析結果

続いて、未婚率と同様の分析を短期居住率について行った結果を示す。居住期間の不詳率を0～2倍とした各シナリオについて市区町村別の短期居住率を推計し、地図化したものが図11である。階級区分はシナリオAで得られた値による等間隔分類である。なお、不詳率2倍のシナリオDで不詳率が非常に高い水準に設定される都市においては、短期居住率が負の値で算出されることがある。これは、当該シナリオでは不詳率比2:1という条件を満たすことができないケースがあることを意味しているが、ここではその場合の短

図9 不詳率比を用いた未婚率の推計値の差（シナリオD-A）

図10 不詳率の増加に伴う未婚率の推移（不詳率比を用いた推計値）

期居住率を便宜的に0として処理した。

これまでの結果と同様に、シナリオAの地図では都心部から周辺地域へと短期居住率が低下していく傾向が明瞭である。そして不詳率が高くなるにつれて色が濃い地域が少なくなり、地域的な傾向を読み取ることが難しくなる。各シナリオの結果の値を用いて階級区分し直した図12を見ても、短期居住率の高い地域は不詳率が高いほどばらつくようになり、明瞭な地域差はとらえにくい。不詳率が高くなるシナリオでは、本来あるはずの地域差がなかったり違ったりして見えてしまう様子が可視化されており、特に極端な値をとるシナリオDでは、むしろ東京都心部に短期居住率の低い地域が集中し、ドーナツ化現象のような分布傾向が見てとれる。

こういった見え方の変化の仕組みも、未婚率の場合と同様である。つまり、居住期間不詳率が増加するにしたがって、短期居住率が高く不詳率も高い都心部の短期居住率が大幅に過小推計されることが理由と考え

図11 不詳率比を用いた短期居住率の補正・推計結果（シナリオAの結果による等間隔分類）

図12 不詳率比を用いた短期居住率の補正・推計結果（各シナリオの結果による等間隔分類）

られる。シナリオ間の差分を描いた図13は、このような東京都市部における短期居住率の過小推計をはっきりと示している。ただし、居住期間の不詳率が未婚率よりも全体的に高く、またその地域差も大きいため、未婚率の推計結果と比較すると都市部の過小推計はより強く表れることになったと理解される。

さらに、不詳率増加に伴う短期居住率の推移をグラフで見てみたい。図14は、横軸にシナリオBを1としたときの各シナリオの不詳率を、縦軸に短期居住率をとったグラフである。各都県の居住期間不詳率が最高の市区（東京都港区、神奈川県横浜市中区、千葉県市川市、埼玉県和光市：図の実線）および最低の町村（東京都御蔵島村、神奈川県山北町、千葉県睦沢町、埼玉県小鹿野町：図の点線）と、一都三県で不詳率が2番目に大きい東京都千代田区の値を例示した。

各都県の不詳率が最も低い町村（図の点線）は、2015年の居住期間不詳率が0%であるため、シナリオ

図13 不詳率比を用いた短期居住率の推計値の差（シナリオD-A）

によって短期居住率は変化しない。また不詳率が17～25%程度である埼玉県・千葉県・神奈川県の不詳率一位の市区は、不詳率が増加するにつれて緩やかに短期居住率が低下していく。一方、東京都港区や千代田区の短期居住率は、不詳率が増加するにつれて大幅に低下する様子が描かれている。この2区は、2015年国

図14 不詳率の増加に伴う短期居住率の推移（不詳率比を用いた推計値）

勢調査の居住期間不詳率がそれぞれ48%・43%とかなり高いため、シナリオDともなると、居住期間の不詳率は80～100%近い設定になってしまう。これにより、不詳率比を用いた重み付け按分では、計算上、非「不詳」人口に短期居住者が1人も残らない状態となるため、短期居住率0となる（計算上負の値をとる場合は0として処理）。こういったケースでの不詳率比（2:1）の仮定には議論の余地があるものの、いずれにしてもここでの結果は、「不詳」の増加が結果を大きく変えてしまう可能性を示している。

V 年齢階級別比例按分と重み付け按分の比較

本研究では、「不詳」を一定の条件や仮定のもとで按分して非「不詳」人口に足し合わせる補正と、逆に不詳率が上昇したシナリオを立てて同様に値を計算する推計を行った。その方法として、一つは年齢階級別に比例按分を繰り返す方法、もう一つは属性間の不詳率比を仮定して重み付け按分を行う方法を用いた。このうち前者は「不詳」の補正法として広く利用されてきたものであり、年齢が「不詳」および属性（未婚等）の分布と強く関連していることを利用した方法である。一方、後者は属性そのもの（たとえば配偶関係の未婚）が「不詳」発生と関連することを考慮した異なるアプローチによる手法である。いずれも集計データから表計算ソフト等で計算可能であることから、「不詳」の補正や特定条件下におけるデータの歪みを把握するための簡便な手法と考えることができる。そ

図15 不詳率の増加に伴う推移の比較（東京都豊島区）

ここで最後にこれら二つの方法を比較して、それぞれの利点と問題点を検討したい。

IIIおよびIVにおける分析を通じて、「不詳」が増えることで未婚率および短期居住率の地域差が見えにくくなるという傾向は、いずれの手法によっても確認された。しかし、具体的な数値に基づいて詳しく比較すると、両者には違いが見られる。たとえば、二つの方法による差が大きい東京都豊島区について、各シナリオの不詳率が増加するにつれて未婚率・短期居住率がどのように推移するかを示すと、図15のようになる。未婚率・短期居住率共に不詳率比に基づく重み付け按分のほうが、年齢階級別按分よりも各指標の変化が大きくなっていることが読みとれる。

他の地域でもこの傾向は同様であった。図16は、年齢階級別比例按分（①）と不詳率比による重み付け按分（②）のそれぞれで得られた指標の値の差分（②－①）を示している。ここでは、補正シナリオ（A）と2倍シナリオ（D）についてのみ結果を地図化した。まずはシナリオAの結果に着目する。ここでの値が正であることから、不詳率比に基づく重み付け按分の方が年齢階級別按分よりも各指標をより大きい値に補正していることがわかる。周辺部では大きな差はほとんどないが、各都県の都市部で差が見られ、未婚率では最大で2.4%、短期居住率では10.9%の差がある。他方、シナリオDについては負の値が得られていることから、不詳率比に基づく按分の方が未婚率・短期居住率をより低く推計したことがわかる。シナリオAと同様に、周辺部よりも都市部で差が大きく、未婚率では最大で－3.5%、短期居住率では－12.4%の差が生じている。

図16 年齢階級別比例按分 (①) と重み付け按分 (②) の値の差 (②-①)

ここから、今回の条件では不詳率比 (2:1) に基づく按分補正・推計のほうが、年齢階級別比例按分よりも結果に与える影響が大きかったことがわかる。前者のほうが都心部の未婚率や短期居住率を大幅に過小推計する傾向があったのに対し、後者ではそれほど大きな値の変化は見られず、周辺地域を含めて全体を緩やかに過小推計していた。両手法の妥当性について、どちらが真の値に近いのかは基準がなく判断できないものの、属性 (未婚や短期居住) 自体の影響を含むさまざまな理由による「不詳」発生が増加すれば、年齢のみを考慮した按分は相対的にみればより限定的なものとなる可能性は否定できない。

ちなみに未婚率の年齢階級別比に基づく按分補正・推計では、年齢「不詳」をすべて配偶関係「不詳」として按分処理したため、ベースとなる不詳率が不詳率比に基づく場合よりも年齢階級別比例按分のほうで若干高い設定になる自治体がある。そういった自治体では年齢階級別比例按分のほうが推計・補正による値の変化の余地が大きくなるはずである。にもかかわらず、結果として不詳率比に基づく按分補正・推計のほうが未婚率を大きく変化させたということは、年齢階級別比例按分よりも不詳率比に基づく按分のほうが補正・推計の結果に与える影響が大きいという知見と矛盾はないと解釈できる。

なお、年齢階級別按分による未婚率においては、不詳率が高いシナリオのほうが未婚率がわずかに大きく推計される例があった。たとえば、青梅市の未婚率はシナリオAが29.4%、シナリオBが29.7%であり、横浜市中区ではシナリオA～Dにかけて33.1%、33.3%、

33.4%、33.5%と微増する。不詳率比に基づく按分では理論上、未婚率が過大に推計されることがないため、これらの都市も「不詳」が増加するにつれて未婚率は低下する結果となっている。一方、年齢階級別按分では、有配偶の割合の高い年齢階級で「不詳」人口が多いなど、地域の人口構成と「不詳」の分布によっては過大推計となるケースもあることが確認できた。

一方で不詳率比を用いた場合、不詳率が高くなりすぎると、属性間の不詳率比の仮定 (今回は2:1) を満たすことができず、正しく計算できない点に限界がある。同様の限界は年齢階級別比例按分の場合にも生じるものの、今回の分析では不詳率による按分のほうがこれに該当するケースが多く見られた。未婚率を例にとると、未婚の不詳率はつねに非未婚の2倍である必要があるため、不詳率が高まるにつれて観測される (= 非「不詳」人口の) 未婚者が減少していき、0を下回るところで計算不能になる (今回は便宜的にこのケースを0として処理)。今回の場合、不詳率がまだまだ低い水準にある配偶関係ではこの問題は生じなかったが、居住期間については不詳率がすでに著しく高い市区町村があり、それが2倍となるシナリオでは計算できない地域が散見された。

しかし実際には、短期居住者のすべてが「不詳」となり、短期居住率が0になるということはやや考えにくい。この計算で置いている不詳率の属性間の比率が「つねに一定」という仮定はその意味で単純化しすぎており、実際には不詳率の高低によって (したがって地域ごとに) 不詳率比も変動することが予想される。また各地域の不詳率も全体の増減に対して単純に比例

するわけではないと思われる。したがって、全体の不詳率が一律に増減した場合のみを想定することもやや非現実的であり、不詳率が大幅に上昇した状態のシミュレーションには、それを適切に表現できる仮定を置く必要があるといえよう。

さらに、今回用いた「2:1」という不詳率比の仮定は暫定的なものであり、より現実に近い値を探る試みも必要になる。本研究では、過去のインターネット調査から得られた調査票未提出に関する属性間の比を利用したが、国勢調査では未提出がそのまま「不詳」となるわけではなく、調査員による近隣への聞き取りや行政情報による補記がなされる場合がある。他方で、調査票を提出しても特定の項目のみ未回答である場合には「不詳」となる。本来であればこれらすべてを考慮した上で属性間の不詳率比を得ることが望ましいものの、現状ではこれらの補記や項目未回答に関する情報はほとんど入手できないため、何らかの間接的な情報を組み合わせて実態に迫ることを考えることが現実的に可能な取組みとなろう。

最後に地域単位について考察すると、年齢階級別比例按分は、年齢別に各項目のデータが必要になるため、それが入手可能な市区町村単位までに利用範囲が限られ、小地域レベルにそのまま適用することはできない。それに対して不詳率比を用いた計算にはこのような制約がないため、単純な仮定に基づく計算ではあるものの、小地域単位での補正値を求めることは可能である。ただし、小地域単位では項目や地域によって極端に大きな不詳率をとりやすいため、上述のように、計算不能となるケースが多く出ることも予想される。実際にどこまで有用なものとなるのかは、対象とする項目・地域によって個々に判断する必要があると思われるため、さまざまな項目・地域の事例を蓄積していく必要がある。

VI おわりに

本研究では、近年増加傾向にある国勢調査の「不詳」問題に対して、データの分析や解釈に役立つ基礎情報を得ることを目的に、不詳率が将来高まることを想定したシナリオに基づく分析を行った。具体的には、一都三県の市区町村を対象に、配偶関係の不詳率

が増加した場合の未婚率の変化と、居住期間の不詳率が増加した場合の短期居住率の変化を例として、将来的な「不詳」増加が各種指標にどう影響し、どのような見かけ上の地域差を生じさせるのかを可視化した。

まず、従来しばしば利用されてきた年齢階級別の比例按分を、「不詳」の補正だけではなく不詳率が上昇するシナリオのシミュレーションに適用する方法を試みた。計算の結果、「不詳」が増加することで都心部を中心に各指標の過小推計が起り、それに伴い地域差が不明瞭化することが示された。続いて先行研究の知見を基に、未婚および短期居住の不詳率が他の属性の2倍であるという仮定を置き、それを満たすような補正値・推計値をシナリオごとに求めた。その結果、対象とした指標（未婚率・短期居住率）は、補正値で見ると都心部で高く周辺部にいくにつれて低くなる地域差が見られたが、「不詳」が1.5倍や2倍に増加したシナリオで推計した値の地図を見ると、そういった地域差は縮小ないしは消失して見えることが明瞭に可視化された。

このような疑似的な地域差が生じるのは、未婚率や短期居住率が高い都心部では不詳率も高い傾向にあるためと理解できる。不詳率が上昇するシナリオでは、都心部の未婚率・短期居住率が低下することで全体が平準化され、都心と周辺部の差が見えにくくなることは、いずれの手法によってもシナリオごとの指標の推移から確認された。特に不詳率が高い居住期間の短期居住率においては、都心部での「不詳」増加に伴う指標の過小推計が顕著であった。したがって、近年上昇を続けてきた国勢調査の不詳率がさらに高くなった場合、今回シナリオに基づく分析結果として提示したような各種の疑似的な地域差が現実のものとなってしまうことが懸念される。現状でも、高い不詳率をもつ項目や地域を対象とした分析を行う際には、本研究で例示したような過小（または過大）な評価が引き起こされている可能性に留意しつつ、補正値を利用したり結果を慎重に解釈したりすることが必要であろう。

また、従来頻繁に使用されてきた年齢階級別の比例按分法と、属性間の不詳率比を仮定して重み付け按分する方法を比較したところ、今回の仮定では後者のほうが変化の絶対値が大きいなど、結果に違いがあることを確認した。方法の妥当性を単純に比較することは

困難であるものの、両者に違いがあることは、「不詳」発生の仕組みをより詳しく解明し、それに応じた補正・推計方法を検討する必要性があることを示唆する。また、不詳率比による方法は市区町村よりも詳細な、年齢階級別の値が表章されない地域単位にも適用可能である一方、不詳率が極端に高い場合には計算不能となる地域が生じることや、「属性間の不詳率比が一定」という仮定が必ずしも現実的ではない場合があることなどの課題も指摘された。

本稿執筆中の2020年8月時点では、2020年国勢調査の実施に向けて準備が進められている段階にあるが、新型コロナウイルス感染症の流行が続く中、調査員の成り手不足などを理由に、回収や公表時期の変更が決まっている。現時点では調査そのものは実施予定とされているものの、調査依頼の際に対面による説明を行わないことや、インターネット回答と郵送回答を推奨することなど、これまでとは実施方法が異なる点も少なくない。こういった諸種の変更は、おそらく2020年国勢調査における「不詳」をこれまでの国勢調査以上に増加させる方向に作用し、また、方法面の変更が「不詳」の発生傾向（たとえば今回利用した属性間の不詳率比）にも影響することが予想される。国勢調査のデータとしての信頼性を維持するためにも、この問題に対する基礎研究は今後さらに重要な課題となるであろう。

謝 辞

本研究は、JSPS 科研費基盤研究 (A) 「地理的マル

〈著者略歴〉

山本 涼子 (やまもと りょうこ)

特定非営利活動法人traNeco代表。博士 (環境学)。主な論文: 交通社会資本プロジェクトのLCAにおける評価範囲設定が結果の不確実性に及ぼす影響の分析枠組 (日本LCA学会誌7(4), 2011年)

埴淵 知哉 (はにぶち ともや)

東北大学大学院環境科学研究科准教授。専門社会調査士。博士 (文学)。主な著書: 『地域と統計——〈調査困難時代〉のインターネット調査』 (共編, 2018年, ナカニシヤ出版)

中谷 友樹 (なかや ともき)

東北大学大学院環境科学研究科教授。博士 (理学)。主な著書: *The Atlas of Health Inequalities in Japan* (共編, 2019年, Springer)

山内 昌和 (やまうち まさかず)

早稲田大学教育・総合科学学術院准教授。博士 (学術)。主な著書『地域社会の将来人口——地域人口推計の基礎から応用まで』 (共編, 2020年, 東京大学出版会)

チレベル現象の解明に向けた基盤的データの構築」 (JP 17H00947) の助成を受けたものである。

文 献

- 阿部 隆 2013. 国勢調査結果の「不詳数」に係わる諸問題. 統計 64(12): 51–54.
- 阿部 隆・磯田 弦・澁木智之 2018. 東日本大震災後の東北地方の人口移動パターンの変化——仙台一極集中現象を中心として. 東北地理学会編『東日本大震災と地理学』157–173. 笹氣出版.
- 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.
- 小池司朗・山内昌和 2014. 2010年の国勢調査における「不詳」の発生状況——5年前の居住地を中心に. 人口問題研究 70: 325–338.
- 埴淵知哉・中谷友樹・村中亮夫・花岡和聖 2018. 国勢調査小地域集計データにおける「不詳」分布の地理的特徴. 地理学評論 91A: 97–113.
- 埴淵知哉・村中亮夫編 2018. 『地域と統計——〈調査困難時代〉のインターネット調査』ナカニシヤ出版.
- 埴淵知哉・山内昌和 2019. 国勢調査「不詳」発生の関連要因——インターネット調査を用いた未提出者の分析. *E-journal GEO* 14: 14–29.
- 埴淵知哉・山本涼子・中谷友樹・山内昌和 2020. 「不詳」の増加は統計地図をどう歪めるのか?——開始100年を迎える国勢調査の困難. 地理 785: 56–71.
- 山田 茂 2016. 2015年国勢調査結果の精度について——抽出速報集計を利用した暫定的考察. 國土館大學政経論叢 178: 1–35.
- Martin, D. 2010. Understanding the social geography of census undercount. *Environment and Planning A* 42: 2753–2770.
- O'Hare, W. P. 2016. Who lives in hard-to-count neighborhoods. *International Journal of Social Science Studies* 4(4): 43–55.