73. 住まいの意向調査における災害公営住宅希望戸数の無回答バイアス補正の検討

- 2016 年熊本地震での熊本県益城町の事例より -

Correcting Non-response Bias in the Demand Estimates for Disaster Public Housing Using Questionnaire Survey

- A Case of Mashiki in the 2016 Kumamoto Earthquake -

佐藤 嘉洋*・円山 琢也**

Yoshihiro SATO* and Takuya MARUYAMA**

Non-response bias, which is caused by non-respondents or missing data in a survey record, can lead to severe problems, and several methods have been proposed to correct the bias. This study aims to explore the correction methods of non-response bias in the demand estimates for disaster public housing using the questionnaire survey in Mashiki following the 2016 Kumamoto earthquake. The survey results indicated that the following households tended to prefer disaster public housing: households in rural areas, with single, with elderly members, with low incomes, and with single parents. Using this data, we corrected the non-response bias caused by missing data for housing preference in this survey. The k-nearest-neighbor imputation with socioeconomic variables and their problem indicators for disaster recovery performed better than hot deck imputation or imputation without these variables.

Keywords: Kumamoto Earthquake, Disaster Public Housing, Non-response Bias, Bias Correction

熊本地震, 災害公営住宅, 無回答バイアス, バイアス補正

1. 研究の背景と目的

大規模災害発生時には、被災者に対して様々な調査が実施される。特に生活再建支援に関する調査は再建に向けて被災者が抱えている課題、悩みを網羅的かつ細やかに把握していくことが望まれる。つまり、このような調査では悉皆性を確保するような調査設計が重要であると言える。しかし、現実には全数調査を達成することは困難であり、少なからず無回答が存在することとなる。

大規模災害時の調査には、復興へ向けた政策決定の基礎データの収集という重要な役割もある。この調査への無回答は「回答を行うほど生活再建には困っていない」という意思表示の可能性もある。しかし、この場合、無回答が完全にランダムに生じている (MCAR: Missing Completely At Random) とは言えない。その場合、回答データだけで被災者の実態を把握することは危険であり、無回答による調査結果の偏り (無回答バイアス) が存在しうるのである。加えて、政策決定に影響する調査項目に対しても、仮に無回答が MCAR で生じているのであれば、無回答の存在は調査結果の不偏推定量に影響しない。しかし、Horikoshi *et al.*¹⁾は、災害時の無回答世帯は回答世帯に比べ様々な問題を抱えている可能性があるとして指摘していることから、無回答は完全にランダムには生じていないと考えられる。この場合、無回答世帯が抱える問題を探求し支援策を検討すると同時に、無回答バイアスを低減させる調査方法・分析論の構築も望まれる。

2016 年熊本地震で最も大きな被害を受けた熊本県益城町では、応急仮設住宅、みなし仮設住宅が約 3,200 戸供給され、多くの住民が不自由な生活を余儀なくされた。同町では被災世帯のために災害公営住宅の建設が計画され、2 度の意向調査を踏まえて整備予定戸数が決定された。予定戸数の決定は 2 回目の意向調査に基づいて行われたが、回

収率向上のため郵送調査無回答世帯に対しては訪問調査が実施された。回収率向上の目的には、被災者の再建意向を可能な限り把握することと共に、災害公営住宅の需要推計誤差を可能な限り少なくしたいというねらいがあった。

佐藤・円山²⁾はこの意向調査での回答世帯の偏りを明らかにし、回答時期によって災害公営住宅の希望割合にバイアスが存在し、郵送調査での無回答は少なくとも MCAR で起きているとは言えないことを示している。本研究では、この両調査手法のデータに着目する。訪問調査で回答を行った世帯を郵送調査の無回答世帯とみなし、これによるバイアスを本研究における「無回答バイアス」と定義する。意向調査で無回答世帯の情報は聞き取れているため、無回答バイアスの存在とその補正法について分析を行うことが可能である。そこで本研究では、

1. 住まいの意向調査からどのような世帯が災害公営住宅を希望しているか、その実態を把握すること
2. 上記実態をもとに、郵送調査で無回答だった世帯の無回答バイアスの補正の比較検討を行うこと

以上の 2 点を目的とする。

本研究の特徴として、佐藤・円山²⁾では行えていなかったバイアス補正法を検討する実証的研究であることが挙げられ、ここに新規性・独自性がある。本研究は、意向調査で得られた調査結果から無回答世帯を定義し、無回答バイアス補正法に関する試算例を検討するものである。郵送調査のみで得られた結果をもとに被災者支援策が決定される例も十分考えられうることから、本研究のアプローチによるバイアス補正法の検討、評価を行う価値は高いと考えられる。

* 学生会員 熊本大学大学院自然科学教育部

(Kumamoto University)

** 正会員 熊本大学くまもと水循環・減災研究教育センター

(Kumamoto University)

2. 無回答バイアスとその補正法

(1) 先行研究のレビューと本研究の位置づけ

社会調査方法論の研究対象として、Tourangeau は標本誤差から非標本誤差に関心が移ってきていると主張している⁴⁾。無回答も非標本誤差の一つであり、それによって結果に誤差(バイアス)が起こりうる事が先行研究でも明らかにされている。例えば宮野⁵⁾が標本誤差と非標本誤差の両方に着目して投票率に関する誤差の分析を行い、特に無回答バイアスの影響が大きいことを示した。このように無回答バイアスが結果に与える影響の精査は重要な課題である。

社会調査における無回答とは、調査票自体が欠落するユニット無回答と、調査票の一部設問の回答が欠損するアイテム無回答に分けられるが、藤原ら⁶⁾も指摘しているようにアイテム無回答と比べてユニット無回答に対するバイアス補正は困難を伴うことが多い。

アイテム無回答が存在する調査データに対しては、数理的に無回答バイアスを補正する様々な手法が提案されている。例として、石井・野崎⁷⁾は家計パネル調査でのサンプル脱落の問題に対して Cross-Sectional/Longitudinal ウェイトを用いてバイアス補正を行っており、交通工学分野で上述の藤原ら⁸⁾は交通日誌データをもとに EM アルゴリズムによるバイアス補正を試み、アイテム無回答の場合に良好な補正結果を得られたものなどが挙げられる。このように、補正の試みに関する多くの研究例があるが、特に大規模災害時の調査においての研究例は少なく、知見の蓄積は十分とは言えない。

以上を踏まえ、本研究では、1.でも述べたように、益城町での意向調査の目的の一つである災害公営住宅の希望戸数を研究対象とし、アイテム無回答に着目する。住まいの意向についての回答が欠落もしくは不明である場合とし、それ以外の情報は他の調査から利用可能とする。つまり、本研究で着目する無回答バイアスは、欠測データによるバイアス(欠測バイアス)と読み替えることができる。調査対象世帯を仮設住宅居住世帯とし、入居情報のほか大規模災害時に多く実施される様々な調査を横断的に結合すれば、この状況に近づけることは実務上可能であると考えられる。

(2) 欠測データの補完法

欠測を含むデータにおいて推定量を求めるとき、欠測データの取り扱いが問題となる。最も簡便な方法は欠測データを削除して分析を行うことであり、これはリストワイズ法と呼ばれる。同手法は欠測により推計誤差を生みやすいたことが知られている。この問題に対し、欠測項目にある値を代入し、擬似完全データを作成して分析を行うデータ補完法(Imputation)が古くから用いられている。本研究では、以下の2つの補完手法の比較検討を行う。

(a) ホットデック法による補完

代入法を用いた補完では、大きく単一代入法と多重代入法⁹⁾に分けられるが、「分析の目的が合計値(平均値)を記述的に求めるだけであれば確定的単一代入法で十分である」

とされている。そこで本研究では、確定的単一代入法として多くの実績があるホットデック法(Hot-Deck Imputation)を用いて分析を行う。ホットデック法とは欠測を含むデータ(recipient)に対し、欠損していない情報が最も類似する標本(donor)をデータセットから選定し、その標本が持つ値を欠損値に代入するノンパラメトリックな代入法である。

(b) k-近傍法を応用した補完

k-近傍法(k-Nearest Neighbor Algorithm)とは機械学習の1種で、特徴空間Sにおいてテストデータにもっとも距離の近いk個の学習データを抽出し、それらに最も多くみられるクラスを付与する教師あり学習の分類器である。先に述べたホットデック法が単一の類似標本からの補完であるのに対し、k-近傍法は複数の類似標本からの補完であることが特徴として挙げられる。通常、k-近傍法は特徴空間Sを構成する要素に欠測がないことが求められるが、k-近傍法を応用したデータ欠測補完手法(kNN Imputation)が開発されており¹⁰⁾、本研究ではこれを使用する。本手法では Gower 距離を拡張した距離関数を用いて近傍を判別しており、連続変数、カテゴリカル変数、順序変数の補完に対応していることが特徴として挙げられる。

3. 意向調査概要

表-1に益城町が実施した「第2回今後のお住まいに関するアンケート調査」(以下、意向調査)の概要を示す。調査対象は在宅等も含む半壊以上の全世帯となっているが、本研究での対象は応急仮設およびみなし仮設住宅とする。なお、本研究で用いた意向調査データは最終締切後に回収されたサンプルまで含んでおり、益城町の各種委員会に提出された資料とはサンプルサイズが異なることに留意されたい。意向調査での主な質問項目は表-2に整理している。

(1) 意向調査における回答群の定義

表-1のように、意向調査は2017年7月31日を回答期限として実施されたが、回収率向上のため回答期限を同年8月10日まで延長した。さらに、8月10日以降は無回答世帯へ訪問し、可能な限り調査票回収に努めた。ここで、調査票の回収日は記録されていたため、回収日に基づいた分類が可能となる。分類結果を表-3に示す。当初の締め切りである7月31日までに回収された調査票を「期間内回答」、8月1日から8月10日に回収された調査票を「遅延回答1」、8月11日以降に回収された調査票を「遅延回答2」と分類した¹⁾。遅延回答1は調査票の放置などによる回答の遅れ、遅延回答2は主に訪問調査での回答と見なすことができる。ただし、放置後の回答が遅延回答2に影響していることが考えられるが、両者を厳密に区別することは困難であるため、本研究では回収日に基づいて回答群の整理を行った。

(2) 意向調査における無回答バイアス

意向調査に真に無回答だった世帯の住まいの意向は得られていないため、意向調査全体の無回答バイアスは把握で

表-1 第2回 今後のお住まいに関するアンケート調査概要⁽¹⁾

調査主体	益城町
目 的	熊本地震からの住まい・集落の復旧に向けた基本方針等の検討や住宅自力再建支援策の検討、災害公営住宅等の供給等に向け、被災世帯の住まいの再建方法や予定時期を伺い、今後の復旧・復興に向けた取り組みへ繋げる
対 象	熊本地震により益城町内で被災された世帯のうち、半壊以上の世帯
調査期間	2017年7月3日から7月31日
調査方式	郵送配布/回収形式 ※仮設住宅の期限内無回答世帯には個別訪問の上、一部聞き取り調査形式にて実施
対象世帯数	7,284票（応急仮設住宅：1,463票、みなし仮設住宅：1,455票、在宅等：4,366票）
回収数	5,210票（応急仮設住宅：1,364票、みなし仮設住宅：1,174票、在宅等：2,672票）
回収率	71.5%（応急仮設住宅：93.2%、みなし仮設住宅：80.7%、在宅等：61.2%）

表-2 意向調査の調査項目（全世帯共通のみ）

家族構成	未就学児/小・中学生/15～19歳/20～59歳/60～64歳/65～74歳/75歳以上 それぞれの区分の人数
世帯主	・就業状況（就業中/求職中/無職/年金/無職/収入なし）/生活保護 ・年齢
世帯年収	なし～1,000万円以上まで50～300万円単位の12肢
被災住家	・所有形態（持家/借家） ・被災状況（罹災判定） ・応急修理制度の利用状況
住まいの再建に関する意向（単一回答）	ア. 震災前と同じ住所に自宅を再建したい イ. 震災前に住んでいた住宅を修理して、住み続けたい ウ. 震災前とは別の住所（災害公営住宅を除く）に移転したい エ. 新たに整備される災害公営住宅に入居したい
住宅再建、住宅の確保に向けて、現時点で困っていること（表-6参照）	

きない。そこで本研究では、前節に示した回答分類をもとに、期限内回答と遅延回答1を回答世帯、遅延回答2を無回答世帯と定義する。これにより、災害公営住宅を希望した世帯の割合に関して、郵送調査では把握することが難しい無回答バイアスをシミュレーションすることが可能となる。

意向調査データをもとに、災害公営住宅希望戸数の推計を行った結果を表-4に示す。推計では、最も簡便な推計方法であり、一般に採用されることの多い「回答と無回答世帯の希望割合は同一である」との仮定を置いたリストワイズ法による拡大推計法を用いる。表中の(E)はこの手法で推計された無回答世帯中の希望世帯数で、(F)は聞き取り調査で実際に得られた希望世帯数であるが、応急仮設、みなし仮設ともに希望世帯数は過大に推計される結果となった。この需要の差が本研究で定義した無回答バイアスである。

4. 意向調査における災害公営住宅希望世帯の基礎分析

本章では、補正法に利用するデータを探索することを目的に、意向調査の基礎分析を行う。なお、本章では意向調査での実態を明らかにするため、回収全世帯での分析を行っているが、無回答世帯を除いた場合でも傾向に大きく差がないことを確認している。

表-3 意向調査の時期別回収数（応急仮設、みなし仮設のみ）

（数値上段の単位：世帯）

回収日	～7月31日	～8月10日	それ以降	合計
名称	期間内回答	遅延回答1	遅延回答2	
応急仮設住宅	939 (68.8%)	190 (13.9%)	235 (17.2%)	1,364 (100.0%)
みなし仮設住宅	853 (72.7%)	135 (11.5%)	186 (15.8%)	1,174 (100.0%)

*数値下段は回収数に対する割合、住宅区分は調査回答にしたがう

表-4 災害公営住宅需要推計結果⁽²⁾

	応急仮設	みなし仮設
回答世帯数 (A)	1,118 世帯	980 世帯
うち災害公営住宅希望世帯 (B)	473 世帯	220 世帯
希望割合 (C) = (B)/(A)	42.3%	22.4%
無回答世帯数 (D)	233 世帯	186 世帯
推計世帯数 (E) = (C)*(D)	99 世帯	42 世帯
実際の希望世帯数 (F)	64 世帯	37 世帯
実際の希望割合 (G) = (F) / (D)	27.4%	19.9%
希望割合の差 (G) - (C)	-14.9%	-2.5%
推計誤差 (E)-(F)	+35 世帯	+5 世帯

表-5 意向調査の基本統計量

（単位：世帯、カッコ内は対象世帯に対する希望割合、不明は除く）

	応急仮設	みなし仮設
	災害公営住宅 希望世帯(39.7%)	災害公営住宅 希望世帯(22.0%)
熊本地震前の居住校区		
飯野校区	54 (28.3%)	— (13.7%)
広安校区	224 (45.3%)	152 (24.8%)
木山校区	153 (38.8%)	71 (19.8%)
福田校区	47 (45.2%)	— (20.0%)
津森校区	43 (33.9%)	— (14.0%)
世帯主の就業状態		
就業中	192 (34.1%)	123 (20.1%)
求職中、無職（年金含む）	322 (44.1%)	111 (26.0%)
世帯主の年代		
30代以下	44 (46.8%)	30 (20.3%)
40代	37 (35.6%)	31 (20.1%)
50代	74 (44.3%)	41 (19.6%)
60代	154 (41.3%)	83 (26.7%)
70代	121 (37.9%)	42 (22.6%)
80代以上	105 (37.5%)	30 (19.2%)
自宅の被災状況		
全壊	441 (40.9%)	194 (22.6%)
大規模半壊	42 (33.1%)	31 (20.5%)
半壊	47 (37.6%)	30 (20.4%)
2016年の世帯年収		
0～250万円未満	380 (52.9%)	155 (37.6%)
250～400万円未満	93 (30.1%)	68 (21.5%)
400万円以上	34 (14.8%)	24 (8.3%)
被災前の住居区分		
持ち家（共同名義含む）	331 (30.7%)	109 (13.4%)
借家・社宅	195 (81.3%)	145 (43.4%)
世帯人数		
1人（単身世帯）	211 (61.5%)	90 (33.0%)
2人	165 (38.7%)	86 (23.4%)
3人	85 (31.1%)	38 (15.3%)
4人	36 (22.6%)	25 (18.2%)
5人以上	29 (23.0%)	— (8.3%)
世帯類型		
高齢者のみの世帯	203 (49.5%)	78 (30.1%)
後期高齢者のみの世帯	100 (51.3%)	32 (27.4%)
ひとり親世帯	— (74.1%)	— (35.1%)
子どもがいる世帯	67 (30.7%)	40 (16.7%)

* 個人が特定される恐れがあるため、10世帯以下の項目は割合のみ表示とした。

(1) 基本統計量

表-5に属性別の基本統計量を示す。熊本地震前に居住していた校区別では、飯野校区、津森校区が他の校区に比べて災害公営住宅の希望割合が低い。広安、木山、福田の3校区合計と比較して検定を行うと、応急仮設で1%水準の有意差が認められた。飯野、津森の2校区は自宅の居住年数が長い世帯が多い¹⁴⁾。渡邊らは同町の応急仮設住宅聞き取り調査から、自宅再建意向の要因の一つとして住み慣れた土地への愛着を挙げている。意向調査では居住年数を知ることができないが、居住年数が長い地区であることが、希望割合が低くなった要因の一つであると推察される。

世帯主または世帯の経済状況に関連する分析として、世帯主の年代別では大きな差が見られず、世帯主の年代は災害公営住宅の希望に影響を与えにくいと考えられる。次に、世帯主の就業状態では、就業中世帯の希望割合が低い。2016年の世帯年収別では、年収が少なくなるほど希望割合が高くなる。本稿で詳細に紹介することは控えるが、調査票での12区分で見ても同様の傾向が示されている。

自宅に関連する分析として、地震前の住居区分では、持ち家と比較して借家世帯の希望割合が高く、特に応急仮設は81.3%と高い結果となった。借家世帯を所有する土地がない世帯とみなせば、賃貸住宅への入居とほぼ同様の形態かつ低廉な賃料で供給される災害公営住宅への選好が高まったと推察される。被災状況別では、全壊が最も希望割合が高いが、他の選択肢と大きな差は見られない。

世帯の構成に関する分析として、世帯人数別では単身世帯の希望割合が最も高く、世帯人数が多くなるほど希望割合が低くなる傾向が見られる。ただし、この中には様々な世帯の形が混在しているため、「高齢者(65歳以上)のみ世帯」「後期高齢者のみ世帯」「子どもがいる世帯(19歳以下の構成員と複数人の成人構成員がいる世帯)」「ひとり親世帯(20～64歳の構成員が1人かつ19歳以下の構成員が存在する世帯)」の4類型を定義して分析を行った。これらのうち希望割合がもっとも高かったのはひとり親世帯であり、特に応急仮設は74.1%と顕著である。対象世帯数が少ないため全世帯と比較して統計的に有意な差は認められなかったが、ひとり親世帯が災害公営住宅を希望する傾向にあると考えられる。次に多かったのは単身世帯で、応急仮設で61.5%、みなし仮設で33.0%と全世帯平均より高い。また、高齢者のみ世帯、後期高齢者のみ世帯ともに希望割合が高いが、両者の間に特筆すべき差はみられず、高齢世帯であることが災害公営住宅の希望に影響を与えていると考えられる。

(2) 意向調査時点での悩み

意向調査では、前節で紹介した基礎属性、住まいの再建方法の意向とともに、「住宅再建、住宅の確保に向けて、現時点で困っていること」を伺っている。表-6に選択肢の一覧を示すが、住宅の再建、確保に向けた費用に関すること、周辺環境等に関すること、住宅再建に向けた情報提供や各種支援に関することの17肢およびその他困っていること

表-6 「自宅の再建に向けて困っていること」の選択肢

住宅の再建、確保に向けた費用に関すること	
1. 住宅の再建（建設）にかかる費用の準備が難しい	
2. 住宅の補修にかかる費用の準備が難しい	
3. よう壁の修繕・補修費用の準備が難しい	
4. 震災時に住んでいた住宅の債務（ローン）の支払いが難しい	
5. 宅地復旧費用の準備が難しい	
6. 賃貸アパート等の家賃・敷金等の支払いが難しい	
7. 融資が受けられない	
周辺環境等に関すること	
8. 自宅の敷地が安全かどうかわからない（震災前と同じ住所に自宅を再建してよいかわからない）	
9. 敷地から道路（公道）に接していないことや、再建とあわせて道路（公道）の拡幅が必要なこと等が理由で、住宅の再建が難しい	
10. 公的事業（土地区画整理や県道拡幅等）の区域内であるため、住宅の再建が難しい	
11. 道路（私道）の復旧が進んでいない、復旧費用の準備が難しい	
住宅再建に向けた情報提供や各種支援に関すること	
12. 新たな生活に向けた準備費用（家財・引越等）が不足している	
13. 住宅の再建、確保に向けた支援情報等が得にくい	
14. 住宅の再建、確保に向けてどこで相談したら良いかわからない	
15. 業者がいない、どうやって探せば良いかわからない	
16. 賃貸アパート等を借りる際の保証人がいない	
17. 物件が見つからない	
その他困っていること	
18. その他（自由回答）	

表-7 住まいの意向別 悩みの選択率

	分類	選択肢	住まいの意向	
			災害公営住宅(N=794)	その他(N=1723)
災害公営希望世帯が少ない	自宅の再建・維持費用	1	32.0%	48.2% ***
		3	4.2%	13.3% ***
		4	1.6%	4.2% ***
	自宅敷地への不安	8	18.1%	30.6% ***
		10	4.4%	7.3% ***
	情報・支援不足	13	11.6%	20.8% ***
		14	10.2%	13.1% **
		18	16.6%	24.3% ***
	その他(有意差なし)	2	2.5%	3.8%
		5	8.1%	8.9%
		9	4.4%	4.5%
		12	22.2%	25.4%
		15	3.9%	5.3%
災害公営希望世帯が多い	賃貸住宅	6	21.8% ***	5.2%
		16	8.2% ***	0.7%
		17	10.3% ***	4.3%
	資金調達	7	14.5% ***	8.1%
	その他(有意差なし)	11	5.7%	5.5%

***1%水準有意, **5%水準有意

として自由回答欄を設け、複数回答方式としている。表-7には、災害公営住宅希望世帯とそれ以外の意向を持つ世帯で選択肢ごとの選択率を示した。なお、本分析では応急仮設とみなし仮設に分けた場合、一部選択肢で対象世帯数が極端に少なくなるため、応急仮設とみなし仮設を合わせて示しているが、それぞれの区分で分析を行っても概ね傾向に差は見られないことを確認している。

災害公営住宅を希望する世帯の選択率が少なかったものとして、「自宅の再建」「よう壁の修繕・補修」「震災前の住宅の債務」といった自宅の再建・維持費用、「自宅の敷地が

安全かわからない」という不安、「公的事業（土地区画整理事業等）の区域内である」という外的要因、「住宅の再建、確保に向けた支援が得にくい」「どこで相談したら良いかわからない」といった情報・支援の不足に関する課題が見られた。いずれも自宅の再建を視野に入れたときに直面する課題と言える。一方、災害公営住宅希望世帯の選択率が高くなったものは、「賃貸アパート等の家賃・敷金等の支払いが難しい」「物件が見つからない」などの賃貸住宅に関する悩みのほか、「融資が受けられない」という自宅再建を断念せざるを得ない大きな悩みが見られた。自宅再建に係る費用の悩みの中で、本選択肢のみ災害公営住宅希望世帯の方が選択率が高くなっており、災害公営住宅の希望に影響を与える特徴的な悩みと言える。

5. 欠測データ補完手法を用いたバイアス補正の検討

本章では、4.で明らかとなった災害公営住宅の希望に係る要因を変数として、欠測データ補完法によりバイアス補正を行う。ここで、本研究では(a)意向調査における社会経済変数/悩みが利用可能な状況に加え、比較対象として(b)意向調査における社会経済変数/悩みが利用不可の状況の2つを想定する。(a)の状況は、災害時における他の調査データと連携することで、無回答世帯においても入居情報だけではなく社会経済変数や悩みに関する情報が手に入る状況、(b)の状況は連携を行わず、もっとも基礎的な情報である仮設住宅の入居世帯の属性しか手に入らない状況下におけるバイアス補正の試みである。

(1) 補完に用いる変数選択とデータ欠測状況

前節での分析から、(a)の場合の補完に用いる変数を「地震前の居住地区」「単身世帯」「ひとり親世帯」「高齢者のみ世帯」「地震前の自宅の形態」「世帯主の就業状況」「世帯年収」に4.(2)で有意差が見られた悩みの6分類を加えた13変数、(b)では上記のうち「地震前の居住地区」から「地震前の自宅の形態」のみの5変数とした。なお、自由回答の有無に関しては、様々な課題が内包されているため、本稿では変数から除外した。自由回答の詳細な分析と新たな変数の導入については今後の検討課題としたい。

表-8に、これら変数の意向調査におけるデータ欠測状況を示す。みなし仮設において世帯主の就業状況、世帯年収の項目に欠測が多いが、その他は概ね10%未満の欠測率である。なお、6つの悩みについて、この設問は任意回答であるため一つも選択していない場合は「悩みがない」と「欠測」の区別がつかない。本研究では、選択されてないものに関してはすべて「悩みがない」と取り扱うため、悩みの6変数については欠測がない状態にある。

(2) k-近傍法で用いる近傍データ数の選定

k-近傍法を用いる場合、誤差を最小とするためには適切なkを選ぶことが重要である。そこで、通常のk-近傍法で用いられる手法にならない、交差検証法により近傍データ数の

表-8 補完に使用するデータセットの欠測率

(単位：世帯、カッコ内は対象レコード数に対する割合)

	応急仮設	みなし仮設
地震前の居住地区	33 (2.4%)	38 (3.3%)
地震前の自宅の形態	32 (2.4%)	16 (1.4%)
世帯主の就業状況	45 (3.3%)	128 (11.0%)
世帯年収	94 (7.0%)	148 (12.7%)
世帯類型(単身/高齢者/ひとり親)	24 (1.8%)	20 (1.7%)

図-1(a) 交差検証における誤判定率(a)

図-1(b) 交差検証における誤判定率(b)

選定を行った。具体的な手順は以下のとおりである。

1. 意向調査データをランダムに並べ替える
2. データを m 分割し、上から i ($i = 1, 2, \dots, m$) 番目の意向情報を欠損させる仮想データセットを作成する
3. それぞれのデータを補完し、誤判定率の平均を求める

なお、無回答者の割合は応急仮設で17.2%、みなし仮設で15.9%であり、欠損率を実態に近づけるため $m = 6$ (6-fold Cross Validation)とした。また、k-近傍法では使用する近傍データ数が多くなるほど判定の曖昧さが増すため、本研究ではkを1から9までの奇数の5段階で評価した。(a)社会経済変数/悩みが利用可能である場合、(b)社会経済変数/悩みが利用不可の場合の結果をそれぞれ図-1(a), (b)に示す。近傍データ数を多く取ると判定の基準が曖昧となり、誤判定率が高くなる傾向にある。住宅区分別では、応急仮設に比べみなし仮設の誤判定率がやや低い。これは表-4でも示したように、みなし仮設居住世帯の災害公営住宅希望割合が少なく、近傍データと無回答データが類似しやすかったことが理由として挙げられる。また、(a)と比較して(b)の場合で誤判定率が高い傾向にあることも見て取れる。(b)では補完に使用する変数が不足し、特徴を捉えにくかったことを表していると考えられる。(a), (b)ともに応急仮設は $k = 5$ のとき、みなし仮設は $k = 3$ のとき誤判定率が最小となったため、これらの値を近傍データ数として採用する。

(3) 補完手法による誤判定率の比較

ホットデック法および(2)で選定した近傍データ数を用いた k -近傍法によって無回答データの補完を行い、住まいの意向の誤判定率の比較を行う。結果を表-9に示すが、いずれの場合もホットデック法よりも k -近傍法で誤判定率が低く、良好な結果を示していることがわかる。住宅区分別では、みなし仮設の方が低い結果となったが、これは(2)でも述べたようにみなし仮設居住世帯の災害公営住宅希望割合が少ないことに起因すると考えられる。

(4) データ補完による無回答バイアス補正結果

前節で補完された補完データで推計された希望戸数が把握可能であるため、これと実際の希望戸数との差を比較し、バイアス補正結果を比較する。表-10に推計結果を示す。

(a)の社会経済変数/悩みが利用可能な状況下では、応急仮設において k -近傍法で10世帯の過大推計とリストワイズ法に比べ推計誤差が71.4%減少した。回答世帯と無回答世帯の特徴に差がある場合、リストワイズ法ではその差を無視することになるため、バイアスを生み出しやすい。しかし、世帯属性や悩みなどの特徴を捉え、近い世帯の情報を欠測部分に補完することでバイアスが低減できることを示している。さらに、災害公営住宅の希望には観測されていない様々な要素が関係していると考えられるため、近傍とする世帯の不確実性を表現できる k -近傍法の方がより良好な結果になったと考えられる。一方、(b)の社会経済変数/悩みが利用不可な状況下では、応急仮設住宅でリストワイズ法と大きな差はなく、みなし仮設で過小推計に転じるなど補正精度が悪い。補完結果を確認すると、誤判定を行ったデータのうち74.4%が災害公営住宅を希望しない世帯と誤って補完された。(b)では世帯を特徴づける情報が不足しており、(2)でも述べたが、数が多い災害公営住宅を希望しない世帯と判定する傾向がより強くなったことが考えられる。

6. おわりに

本研究では、熊本県益城町における、熊本地震被害からの住まいの再建に関する意向調査をもとに、災害公営住宅希望世帯の実態把握と、希望戸数に関する無回答バイアスの補正について検討を行った。具体的な成果を示す。

- 1) 住まいの再建に係る意向調査から、震災前の住宅が借家、世帯年収が低い、世帯主が非就業、震災前の住居形態が借家、単身世帯や高齢者のみ世帯の世帯に災害公営住宅の希望が高い傾向にある。
- 2) 災害公営住宅の希望に関して、郵送調査の回答世帯と無回答世帯に差がある場合、リストワイズ法では推計誤差が大きく、欠測データ補完手法を用いたバイアス補正が有効であること、複数の補完手法を検討し、ホットデック法と比較して k -近傍法が有効であることが分かった。加えて、社会経済変数および世帯が抱える悩みを特徴に導入することで、より良好なバイアス補正が行えることを示した。

表-9 無回答データ誤判定率の比較

	ホットデック法	k -近傍法
(a) 社会経済変数/悩みが利用可能		
応急仮設	29.6%	25.8% (k=5)
みなし仮設	28.5%	22.0% (k=3)
(b) 社会経済変数/悩みが利用不可		
応急仮設	31.6%	30.9% (k=5)
みなし仮設	21.5%	20.9% (k=3)

表-10 データ補完による無回答バイアス補正結果

(単位：世帯、カッコ内は実際の希望戸数との差)

		応急仮設	みなし仮設
実際の希望戸数		537	257
リストワイズ法		572 (+35)	262 (+5)
(a) 社会経済変数/悩みが利用可能	k -近傍法	547 (+10)	258 (+1)
	ホットデック法	550 (+13)	278 (+21)
(b) 社会経済変数/悩みが利用不可	k -近傍法	563 (+26)	238 (-19)
	ホットデック法	559 (+22)	237 (-20)

* k -近傍法において、応急仮設：k=5、みなし仮設：k=3で補正

本研究の結果では、社会経済変数や各世帯が抱く悩みが考慮できない場合にバイアス補正を行うと推計誤差が大きくなった。このことが示しているように、安易な補正処理は誤った結果を導くことになるため避けるべきである。しかし、2)のように多くの情報から影響を与える変数を選び出した場合では比較的良好な補正結果も得られている。大規模災害時における社会調査を横断的に利用することで、被災世帯の特徴をどのように捉え、バイアスが存在する場合はその特徴を使ってどのように補正すれば良好な結果が得られるのか、更なる知見の蓄積を行う必要がある。

具体的には、意向調査では世帯に関する情報はほとんどが選択肢形式となっており、カテゴリカル変数として扱うものが多かった。例えば世帯構成員に関する詳細な情報(年齢、性別、就業状態等)が利用可能な場合は、傾向スコアを利用したIPW推定法などが有力な補正法として考えられる。機械学習におけるアンサンブル学習の概念も取り入れながら複数の手法の試行を今後の研究課題としたい。

本稿で行った補正は災害公営住宅希望の意向という選好の無回答バイアス補正であるが、佃ら¹²⁾が指摘しているように意向から実際の需要戸数を把握することは難しい。しかし、益城町のように意向調査をもとに災害公営住宅の建設予定戸数を決定する場合、補正効果はたかだか30世帯ほどであっても、災害公営住宅の整備費用は1戸あたり数千万円であることを考慮すると、この誤差低減の意義は大きいと言える。

謝辞：本研究は益城町と熊本大学との共同研究「熊本地震の復興期における効率的な調査手法の構築」の一環として行われたものです。益城町企画財政課、復興整備課、生活再建支援課ほか職員の皆様には多大なる協力と助言を頂きました。この場を借りて御礼申し上げます。また、適切な

修正意見を頂いた匿名の査読者にも感謝致します。ただし、本稿に含みうる誤りの責は筆者のみにあります。また、JSPS 科研費 JP18H01561 の研究助成を一部受けています。

補注

- (1) 回収日とは役場担当職員が郵便物を受領した日であり、回答者が発送した日とは若干のタイムラグがある。
- (2) 益城町における災害公営住宅整備戸数は「平成 28 年熊本地震により住宅が滅失し、自力では住宅再建・住宅確保が困難な世帯に対し、当面 680 戸を目標として災害公営住宅を供給する」¹³⁾とあり、第 2 回郵送調査で回答された希望世帯数とは一致しない。

参考文献

- 1) Horikoshi, N., Iwasa, H., Yasumura, S., Maeda, M.: The characteristics of non-respondents and respondents of a mental health survey among evacuees in a disaster: The Fukushima Health Management Survey, *Fukushima J Med Sci*, 63(3), pp.152-159, 2017.12.
- 2) 佐藤嘉洋, 円山琢也: 複数の手法を導入した大規模災害時調査の無回答バイアス: 熊本地震益城町での実証分析, 第 58 回土木計画学研究発表会, 2018.11.
- 3) 佐藤嘉洋, 円山琢也: 郵送調査未回答世帯の実態と調査回答行動分析: 益城町仮設住宅居住者を対象として, 土木学会論文集 D3, Vol.74, No.5, pp.209-218, 2018.
- 4) Tourangeau, R.: Cognitive aspects of survey measurement and mismeasurement, *International Journal of Public Opinion Research*, Vol. 15, Issue 1, pp. 3-7, 2003.
- 5) 宮野勝: 誤答効果と非回答バイアス—投票率を例として—, 理論と方法, 1 巻 1 号, pp.101-114, 1986.
- 6) 石井加代子, 野崎華世: 「慶應義塾家計パネル調査(KHPS)」と「日本家計パネル調査(JHPS)」における Cross-sectional/Longitudinal ウェイトおよびパネル統合ウェイトの作成, 三田商学研究, 57(4), pp.123-145, 2014.
- 7) 藤原章正, 杉恵頼寧, 原田慎也: 交通日誌データにおける無回答バイアスの修正方法, 土木計画学研究・論文集, Vol.16, pp.121-128, 1999.
- 8) Rubin, D.B.: Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. Proceedings of the Survey Research Methods Section, American Statistical Association, pp.20-34, 1978.
- 9) 高橋将宜, 渡辺美智子: 欠測データ処理 R による単一代入法と多重代入法, 共立出版, p.24, 2017.
- 10) Kowarik, A. and Templ, M.: Imputation with the R package VIM. *Journal of Statistical Software*, 74(7):1-16, 2016.
- 11) 渡邊萌, 佐藤嘉洋, 円山琢也: 熊本地震における益城町仮設住宅入居者の居住地選択意向分析, 都市計画論文集, Vol. 52, No.3, pp.1094-1100, 2017.10.
- 12) 佃悠, 山野辺賢治, 小野田 泰明: 災害公営住宅入居登録者の登録までの住宅再建意向変化とその要因, 日本建築学会計画系論文集, 82 巻 731 号, pp.1-9, 2017.
- 13) 益城町復興整備課: 益城町における災害公営住宅の供給方針について, 2017.10.