

令和 7 年国勢調査有識者会議企画WG（第 1 回）

# 国勢調査の不詳補完に係る課題と対応の方向性

令和 5 年 7 月  
総務省統計局

## 国勢調査における不詳データの増加に対する集計上の対応

国勢調査における不詳データの増加に伴う影響に対応するため、従前、以下の補完処理を実施

### 平成27年国勢調査 あん分による補完

- 人口等基本集計において、あん分により、年齢・国籍不詳を補完

### 令和2年国勢調査 不詳補完値の公表

- 人口等基本集計で合理的な推計が可能な部分について、年齢及び国籍を個票レベルで補完した上、残る不詳について、年齢、国籍及び配偶関係をあん分により補完
- その他の集計区分※についても、あん分により主要項目の不詳を補完

**〔前処理〕 世帯単位で基本的項目がすべて不詳のデータに対し、個票レベルで年齢及び国籍（日本人・外国人の別）を補完**

- **二人以上の世帯**

小地域、男女・世帯人員構成及び住宅の建て方別に、ホットデック法により年齢、国籍を補完

- **単身世帯**

小地域、男女別にコールドデック法により行政記録情報（在留外国人登録データ）から年齢及び国籍を補完。その上で、民営賃貸共同住宅の年齢不詳は推計人口をベンチマークとして、市区町村、男女別に確率的に補完。

※ **その他の集計区分**；就業状態等基本集計（労働力状態、産業、職業及び従業上の地位）、従業地・通学地集計（労働力状態及び従業地・通学地）及び人口移動集計（年齢5歳階級及び5年前の常住地）

## 令和2年国勢調査の不詳補完値に係る課題と対応の方向性

### 令和2年国勢調査の不詳補完の効果

- 調査困難な層とみられる若年層、外国人及び未婚者の不詳が減少するなど結果精度上の観点で一定の効果がみられた。

### 不詳補完値の課題

- あん分を基本とする手法であり、対応可能な統計表は一部に限られる。また、小地域集計にも対応していない。
- あん分は、集計区分ごとに独立して行うため、集計区分間の整合性が担保できない。
- 統計表レベルでの補完が中心のため、対応する個別データがなく、二次利用できない。



### 対応の方向性

- これらの諸課題に対応する方法としては、不詳データを個別データレベルで補完する方法がある。
- 令和7年国勢調査の集計・公表に向けて、補完処理に関する国際的な動向を確認した上、我が国の国勢調査における不詳補完の改善に向けた検討を行う。

## 国際的な動向① 国連による原則及び勧告

- 国連統計部によると、未加工のデータファイルに含まれる様々な種類のエラーのうち重大なエラーについては全て訂正し、重大でないエラーについても可能な限り訂正されなければならないとされている。
- 一方、**大規模調査においては、手作業によるデータ訂正はほぼ実施不可能であり、訂正の諸条件は自動のエラー検査とインプテーション用のプログラムで規定される。**

### 人口・住宅センサスに関する原則及び勧告（国連統計部） [抜粋]（仮訳）

- ✓ 3.188. 未加工のデータファイルには、様々な種類のエラーが含まれており、回答者が発生させたものもあれば、回答者の回答を誤解した調査員が引き起こしたものもある。さらに、データ処理作業、コーディングやデータ入力、あるいは転記の過程でもミスが発生する。作業上の観点から、このようなエラーには2つのタイプがある：(a) それ以降の処理を妨げる可能性のあるもの（重大なエラー）、(b) それ以降の処理作業の論理的な流れを妨げずに、センサス結果に歪みをもたらすもの（重大でないエラー）。前者のエラーは全て、後者のエラーは可能な限り多く訂正されなければならない。（以下略）
- ✓ 3.189. **大規模な調査の場合、手作業による訂正は経済的にほぼ実行不可能であるため、自動のエラー検査とインプテーション用に特別に設計されたコンピュータ・プログラムにより、人や世帯又はその他の人々や世帯に関する情報に基づき条件が規定される。（以下略）**

参考資料) United Nations Statistics Division (2017)

“Principles and Recommendations for Population and Housing Censuses Revision 3”

## 国際的な動向② 国連によるデータエディティングに関する整理

- 国連統計部による人口及び住宅センサスのデータエディティング（※1）に関するハンドブックによると、
  - 多くのインピュテーション・システムにおいて、複数のインピュテーションの方法を活用
  - 演繹的インピュテーション（※2）を可能な限り適用した後、1つ又はそれ以上の手順を適用するのが一般的
  - 現在最もよく使われているのは、シーケンシャル・ホットデック法（※3、4）とフェレギ-ホルト法（※5）
  - フェレギ-ホルト法を活用している国の1つは**新インピュテーション法（NIM）**（※6）に変更しており、その他のいくつかの国もNIMを検討

- （※1） **人口及び住宅センサスのエディティング**；データ収集・捕捉が行われる間あるいはその後に、データレコード内やレコード間の誤りを検出し、個々の項目を調整する方法
- （※2） **演繹的インピュテーション（deductive imputation）**；欠測が生じている標本において、観測されている項目間の関係から、欠測している項目の値を論理的に定めることができる場合、その値により補完する方法
- （※3） **ホットデック法**；データセット内の欠測値又は矛盾値を、同一のデータセット内のデータから一定の方法で選択された値でインピュテーションする方法。インピュテーションするデータの選択方法としては、①ランダムに選択する方法や、②インピュテーションを必要とするデータに類似性があるデータの中から、ドナーをランダムに選択する方法などがある。
- （※4） **シーケンシャル・ホットデック法**；上記②と同様に、類似性があるデータセットについて、当該データセットのレコードの並び順に置き換え候補を検討する方法（例えば住所順に並べたデータセットから地理的に類似したドナーを検討）
- （※5） **フェレギ-ホルト法（Fellegi-Holt edit and imputation method）**；全てのエディティングを同時に検討し、各不合格ユニットに対して、インピュテーションする変数の最少セットと数値の許容範囲を決定し、インピュテーションする方法
- （※6） **新インピュテーション法（New Imputation Methodology）**；全てのエディティングを同時に検討し、利用可能なドナー候補を先に選定し、それを不合格レコードへインピュテーションする方法

（参考資料） United Nations Statistics Division（2019）“Handbook on Population and Housing Census Editing Revision 2”

## 国際的な動向③ 新インピュテーション法 (NIM)

### 国連ハンドブック [抜粋] (仮訳)

- ✓ 395. カナダ統計局は、フェレギ-ホルト法を発展させ、1976年から1991年までのカナダの人口センサスに使用した。1996年のカナダの人口センサスでは、この手法が改良され、**新インピュテーション法 (NIM)** と呼ばれるようになった。
- ✓ NIMは、「大規模な (EditとImputationの) 問題に対して、数値変数と質的変数を最小の変化で同時にインピュテーションする」ことを初めて可能にした (Bankier, Houle and Luc, n.d.)。
- ✓ **NIMは、最初にCANEDITに実装され、その後Canadian Census Edit and Imputation System (CANCEIS)に引き継がれた** (Bankier 2005, Chen 2007)。CANCEISは2001年からカナダ統計局で使用されており、電子調査票の妥当性エディットや重複レコードのチェックなど、データの収集や取り込み時における一定のエディットが施された後のデータ処理に使用される。

(参考資料) United Nations Statistics Division (2019)  
“Handbook on Population and Housing Census Editing Revision 2”

## (参考) CANCEISと一般的なE&Iシステムの違い

|           | CANCEIS   | 一般的なE&Iシステム   |
|-----------|---|---|
| <b>原理</b> | <p><b>NIM（最近隣法）</b></p> <p>最初に距離関数等を用いて最も類似したドナーを複数見つけ出し、どのドナーを使って補完するかどうか検討する。</p> | <p><b>Fellegi and Holt</b></p> <p>最小限の変更要件を満たすために不合格ユニット上の補完すべき変数を最初に特定する。どの変数を補完するか決めた後にドナー候補が検討される。</p> |
| <b>弱点</b> | <p>ドナーとして利用できるデータが不十分な場合（データが少数の場合）、補完が困難</p>                                       | <p>最初に補完する変数を特定するため、補完に適したドナーを見つけることが困難</p>   |

「補完箇所の特定」と「ドナー候補検索」という処理の順序を入れ替えることで、CANCEISのNIMはより優れたデータ駆動を実行することができ、計算上の大きな利得を得ている。

一般的な  
E&Iシステム



参考資料：CANCEIS USER'S GUIDE Version 5.4, Statistics CANADA(2020)  
Handbook on Population and Housing Census Editing Revision 2, UNSD(2019)

## 国際的な動向④ 主要国の人口センサスにおけるインピュテーションの方法

### G7各国のインピュテーションの方法

- 我が国を除くG7各国の人口センサスにおけるインピュテーションの方法は以下のとおり。  
(各国とも欠測値に対するインピュテーションを集計プロセスに実装済み)

- **カナダ、イギリス及びドイツ** . . . . . **CANCEIS**
- アメリカ . . . . . ホットデック法
- フランス . . . . . シーケンシャル・ホットデック法
- イタリア . . . . . DIESIS

※ CANCEIS 及び DIESISが新インピュテーション法（NIM）に該当する。  
なお、G7以外ではニュージーランドもCANCEISを利用している。

## 国際的な動向⑤ CANCEISの採用理由等

○ CANCEISの採用理由として、大規模データを効率的に処理可能な点をメリットとしてあげる国が多い。

|       | カナダ  | イギリス   | ドイツ   | ニュージーランド   |
|-------|--|--|---|--|
| 利用状況  | <ul style="list-style-type: none"> <li>▶1996年より利用</li> <li>▶人口センサス等で利用</li> </ul>  | <ul style="list-style-type: none"> <li>▶2011年より利用</li> <li>▶人口センサス等で利用</li> </ul>                                  | <ul style="list-style-type: none"> <li>▶2011年住宅センサスより利用</li> <li>▶人口センサス等で利用</li> </ul>   | <ul style="list-style-type: none"> <li>▶15年以上前から利用</li> <li>▶人口センサス等で利用</li> </ul>   |
| 採用理由  | <ul style="list-style-type: none"> <li>▶大規模データのEdit&amp;Imputationを効率的に処理可能なシステム</li> </ul>  | <ul style="list-style-type: none"> <li>▶無料である上、大規模データを理論に基づいて効率的に処理可能</li> <li>▶国連ハンドブックでも例示されているシステム</li> </ul>   | <ul style="list-style-type: none"> <li>▶CANCEISの大きな利点はEdit&amp;Imputationの同時実行及び効率的なアルゴリズムによる処理速度の速さ</li> </ul>   | <ul style="list-style-type: none"> <li>▶CANCEISの補定処理はFellegi-Holt法よりもはるかにデータ駆動性が優れている。</li> <li>▶CANCEISはカナダ統計局による公式なサポートがあり、継続的な改善がある。</li> </ul> |
| 今後の展望 | <ul style="list-style-type: none"> <li>▶人口センサスのため、引き続きCANCEISを利用していく。</li> <li>▶CANCEISは今後もエラーが見つければその都度改善することで性能を向上させ、新機能も導入していくだろう。</li> </ul> | <ul style="list-style-type: none"> <li>▶いつの人口センサスにおいても、CANCEISの利用は調査計画や評価の場面における意思決定過程の一部であるため、今後の利用は不明</li> </ul> | <ul style="list-style-type: none"> <li>▶今後は様々な方法を利用していくだろう。</li> <li>▶インピュテーションの方法の選択は、調査データに依存するだろう。</li> <li>▶現状の方法やシステムを維持した上で、適切なツールが開発されれば、それを取り入れていくだろう。</li> </ul> | <ul style="list-style-type: none"> <li>▶CANCEISを標準的なツールの一つとして今後も使用し続ける可能性が高いが、多重代入法に関する研究も行っている。</li> </ul>  |

※ 上表は、インピュテーションにCANCEISを利用している国に対し、メールによる聞き取り調査を実施した結果に基づく（2022年時点）。

## 令和7年国勢調査における不詳補完の方向性

- 我が国を除く主要国では、欠測値に対するインプテーションを集計プロセスに実装済み。
- 一方、我が国の令和2年国勢調査の不詳補完は一部の統計表レベルに限定した補完であり、集計体系間での整合性を担保できず、あん分可能な項目も限定される。さらに個別データを補完していないため、二次利用結果と不詳補完値は整合しない。したがって、この方法の拡張は困難。
- 調査環境の悪化に伴う調査票未回収や未回答の増加に伴う不詳の増加は構造的な課題であり、統計の有用性の維持・向上の観点からは、不詳補完の方法を精緻化し、適用範囲を拡充することが有効と考えられる。



- 国際的には、多変量エディット法によるインプテーションが主流であり、また、主要国で共通に用いられている方法のうち、新インプテーション法（CANCEISなど）が大規模データの処理に優れているとされている。



- 我が国においても、CANCEISなどの新インプテーション法を活用した不詳補完を導入する可能性について検討してはどうか。