

## 第9回評価分科会 議事概要

1 日 時 令和3年3月26日（金）10:00～11:26

2 場 所 総務省第二庁舎6階特別会議室

3 出席者

【委員】

椿 広計（分科会会長）、岩下 真理（分科会長代理）

【臨時委員】

久我 尚子、山本 渉、美添 泰人

【審議協力者】

厚生労働省政策統括官付参事官付統計企画調整室室長補佐、農林水産省大臣官房統計部企画管理官補佐（統計調整班担当）、経済産業省大臣官房調査統計グループ統計企画室参事官補佐

【事務局（総務省）】

統計委員会担当室：萩野室長、栗原次長、鈴木次長、福田補佐、増成補佐

4 議 事

- （1）欠測値の補完に係る主な方法等について
- （2）その他

5 議事概要

（1）欠測値の補完に係る主な方法等について

公的統計の作成における欠測値の補完に係る主な方法等について議論を行った。

事務局から、資料に基づき説明が行われた。

審議結果を踏まえて事務局にて資料を修正の上、各委員の確認を経て、最終的に審議結果報告書においてまとめていくこととなった。

主な質疑は以下のとおり。

- ・資料を提供された各府省統計の担当者が、相談など問い合わせができる窓口のようなものはあるのか。
- 今回の資料であれば事務局でもお答えできるし、総務省統計研究研修所も統計作成の支援を行っている。
- ・参考1はよい。それぞれの補完方法が用いられている事例も示すと各府省の参考になると思う。

- 補完方法ごとに活用事例を示している「参考1」も併せて各府省に提供する。
- ・欠測値への対応方法について、時代の変遷に伴い見直しなどは行われるのか。
- 周期調査などで5年ごとの周期的な見直しの際に合わせてブラッシュアップするなどの形で行っているところもあるようだ。
- ・今回まとめた資料を、研修などにもいかしてもらえるとよい。
  - ・今回はこれでよいが、将来的にはアメリカなど外国の不完全データに関する報告書など理論や実例について研究していただきたい。
  - ・欠測値は発生させないことや減らす努力が一番大切で、やむを得ず欠測が出た場合に補完を行うという考え方が最後の18ページに出てくるが、一番重要なことなので、冒頭にも記載して欲しい。3ページに「・・発生しないように対処することが必要であるが」とあるが、ここは「必要である」で切って1つの「○」としてまとめた方がよい。
  - ・補完技術を紹介する時には、まず可能な限り回収率を上げる努力をすべきことを言及することが望ましく、3ページ一つ目の○の「統計調査においては、調査項目の一部について未記入である場合や・・」という記述の前に「まず回収率の向上の努力や、不完全な回答の客体の確認の努力を重ねた上でも」といった記述を追加しておいた方がよいのではないか。
  - ・3ページ目に「可能な限り補完を行うことにより統計表及び平均値などの結果の有用性を確保する」とあるが、マイクロデータとして提供する際の結果の有効性についても書いてもいいのではないか。
  - ・マイクロデータについて、欠測値であったかどうかという情報を付与するかどうかは、研究者側からは必要だと思うが、それが一朝一夕にできるかどうかは大問題であり、論点として残るものである。
  - ・どういう欠測値補完を行うかについて透明性みたいなものも本当はあった方がいいかと思う。手法は進化していき、いろいろな補完技術ができてくると思うが、それはそれとして、推計精度を上げるためには欠測になったデータをできるだけ偏りなく、精度よく予測推定することが必要であるという原理原則を入れておいていただければよいのではないか。
  - ・第1段階としてはいいものをまとめていただいたが、欠測値に対して統計のプロフェッショナルが考える印象と、一般の人が考える印象とは大分違うと思うので、この技術がバイアスなどのない統計を作るために重要であることを周知徹底していただけるとよい。
  - ・17ページに「まず、演繹的な補完や、過去の結果から経年で安定的なものであれば利用を検討」とあるが、この箇所は他府省のデータ・調査結果等の活用についても明記した方がよい。
  - ・8ページ目「比率補完」に「相関関係が高い項目を利用」とあるが、比率補完なので比例関係が高い項目とするべきではないか。
  - ・平均値補完で平均を推定とあるが、「平均」「平均値」の用語を統一すべきかどうか

検討して欲しい。

- ・用語について、「調査客体」と「標本」と「ユニット」と「単位」という用語が使われているが、サンプル全体としての「標本」を指す場合と、レコードとしての「標本」を指す場合の2つの使い方があり、レコードは「客体」としてもいいかもしれない。また、「アイテム」と「項目」と「変数」という用語が使われているが、調査票に関しては「項目」、データに関しては「変数」といった使い分けがあってもよい。
- ・4ページや、6ページに記載されている「完全にランダム」という言葉の意味が理解できるよう、19ページに参考で付けている「完全にランダムな欠測」についての説明を4ページの下に小さく書くとよい。5ページにある、「主にランダムな欠測」という言葉のところで「主に」は無くてもよい。「ランダムな欠測」についても19ページに付けている説明を5ページの下に書くとよい。
- ・10ページのホットデックのところにある傾向スコアはカッコ書きで「傾向スコアマッチング」という名称を付け、14ページで傾向スコアに言及されている箇所は「傾向スコアによる推定」のようにするなど違いが分かるようにした方がよい。
- ・全体の構成については、補完の方法の説明より補完の手順についての説明を前に持ってきた方が、先に全体像が見えて分かり易いし、15ページの補完を行うための主な方法の一覧表も個別の補完方法の説明の前に持ってきた方が分かり易いと思う。
- ・資料の目次には、大項目だけではなく、()の項目まで明記した方が資料として有効ではないか。

## (2) その他

- ・次回の第10回評価分科会は、4月15日(木)16時から、第6特別会議室で開催する予定である旨、事務局から案内された。

以上