

諮問第 76 号の答申（案）

国民生活基礎調査に係る匿名データの作成について

本委員会は、厚生労働省が作成を予定している平成 10 年国民生活基礎調査（以下「10 年調査」という。）及び平成 22 年国民生活基礎調査（以下「22 年調査」という。）に係る匿名データの作成方法の計画について審議した結果、下記の結論を得たので答申する。

記

1 計画の適否

本計画については、これにより作成される匿名データにおいて、10 年調査及び 22 年調査の回答者の匿名性及び学術研究等における有用性がおおむね確保されるものと認められることから、適当である。

ただし、以下の「2 理由等」で指摘した事項については、修正が必要である。

2 理由等

(1) しきい値基準に基づく上限値・下限値の設定

本計画では、平成 19 年国民生活基礎調査（以下「19 年調査」という。）の匿名データ作成方法同様、しきい値に基づき上限値・下限値を設定し、それを上回る・下回る場合に上限値以上・下限値以下でまとめる措置（以下「トップコーディング」、「ボトムコーディング」という。）で匿名データを作成することとしている。

このしきい値に基づく 10 年調査の豊数、22 年調査の家計支出額の匿名化措置については、地域情報が付与されない全国区分での提供であり匿名性が確保されることから、適当である。

(2) 22 年調査において変更した匿名化措置

ア 「こころの状態」のデータ提供方法の変更

本計画では、健康票の調査項目である「こころの状態」について、出現頻度の低い項目を統合するなどの加工措置はせず、調査票の区分で提供することとしている。

これについては、有用性が高まることから、適当である。

イ 「主な介護者」のデータ提供

本計画では、世帯票の「手助けや見守りを要する者の状況」における「主な介護者」のうち「その他の親族」を「その他」に統合して提供することとしている。

これについては、両者を統合すると親族か親族以外かを区別できなくなり「主な介護者」に関する情報の有用性が低下すること、また、外観識別性が低い項目であり、匿名性を確保するために両者を統合する必要性は低いことから、調査票の区分と同様「その他の親族」と「その他」を別々の区分として提供することが適当であると指摘する。

(3) 22年調査において新たに把握された項目の匿名化措置

本計画では、22年調査において新たに把握された項目について必要に応じ匿名化の措置をして提供することとしている。これらのうち、世帯票の「同居していない者の人数」、健康票における検診や人間ドックに関する「医療機関への受診勧奨」、「医療機関への受診状況」、「がん検診受診状況」のデータ提供については、有用性が高まることから、適当である。

ただし、世帯票の教育に関する項目である「在卒の状況」と「(在卒の)学校の種類」を提供することに関して、データ数が少ない「在学したことがない」を「不詳」に統合して提供することについては、「在学したことがない」を「無回答」と同じ扱いをすることになり適当ではない。この情報の有用性を高めながら匿名性を確保するために、「在卒の状況」の「在学したことがない」は便宜上「卒業」に含めることとし、その場合の「(在卒の)学校の種類」は「小学・中学」とし、その区分の名称を「小学・中学以下」に変更して措置すべきと指摘する。ただし、この措置は公表している表章区分と異なることから、利用者への周知を図る必要がある。

(4) 前回答申（諮問第54号、平成25年9月27日答申）における「今後の課題」への対応

ア 地域情報の付与等

国民生活基礎調査の匿名データに関しては、世帯票、健康票のデータから作成した匿名データAと世帯票、健康票に所得票、貯蓄票を加えたデータから作成した匿名データBという2種類のデータを提供している。これらのデータは、調査対象から一定の割合で再抽出する措置（以下、「リサンプリング」という。）をして作成されている。前回の答申で「今後の課題」とされた、これらの匿名データに地域情報を付与することについては、リサンプリング後の標本規模が小さいことから今回は見送ることとしている。

さらに、地域情報を付与した新たな匿名データを作成することについても、提供可能な情報が限定的にならざるをえないことから今回は見送ることとしている。

また、世帯員単位でリサンプリングすることについても、今回は見送ることとしている。

これらについては、同調査が調査対象数の少ない集落抽出であることを考慮すればやむを得ないと理解できるが、調査方法に依拠する面があることから、引き続き、後述「3 今後の課題」で示した方向で検討する必要がある。

イ 所得票の内訳情報の提供

本計画では、所得等の情報について、総額についてのみトップコーディングして提供することとしている。これに対して、所得等の内訳情報も有用性が高いことからトップコーディング等の匿名化措置をして提供することが検討されたが、総額と内訳の情報に整合性が取れないこと、匿名性が十分に確保することができないことが明らかになった。

このような検討の結果、所得等は総額のみをトップコーディングして提供することとし、その内訳情報の提供については、今後、より精緻な匿名化手法に関する慎重な研究・検討が必要であることから、引き続き、後述「3 今後の課題」で示した方向で検討する必要がある。

ウ 匿名データの提供時期と作成対象年次の拡大

国民生活基礎調査の匿名データについては、調査実施からおおむね5年後の提供を計画している。この提供時期は、同調査が集落抽出であること及び同調査を親標本とする複数の後続調査が実施されていることを考慮すれば、匿名性の確保に留意されており、適当である。

また、同調査の匿名データの対象年次の拡大については、昨年の19年調査提供開始に続き、本年に10年調査及び22年調査を提供開始予定であるなど年次追加が着実に進んでおり、適当である。

3 今後の課題

(1) 地域情報の付与及び再抽出の単位

国民生活基礎調査に係る匿名データ作成においては、匿名性を確保するため、回答者である世帯の特定につながる可能性が高い地域情報を削除し、地域区分を「全国」のみとする厳格な匿名化措置を講じることとしている。これは、調査対象数から考えればやむを得ないものの、地域情報は、公衆衛生や疫学分野の研究において世帯員情報と同様に有用性が高いことから、今後、同調査の調査方法が変更される際には、匿名性の確保を十分に図りつつ、匿名データの利用者のニーズを踏まえてリサンプリング等の匿名化措置を検討することによって何らかの地域情報を付与することの妥当性と可能性について検討する必要がある。

(2) 所得票の内訳情報の提供

本計画においては、所得票に含まれる情報について世帯の総所得、課税等の状況及び掛金に限定して提供することとしている。しかしながら、近年、社会保障や所得格差等に関する研究の重要性が増しており、その分析には所得等に関する内訳情報の必要性が指摘されている。一方、匿名性を十分に確保した内訳情報のデータ作成方法は、確立されておらず、より精緻な匿名化手法に関する慎重な研究・検討が必要となっている。

このため、今後、所得等の内訳情報の提供に向け、匿名性と有用性の確保の観点から、トップコーディング以外の適用も含めて匿名化措置を検討する必要がある。