

平成19年就業構造基本調査の標本設計の概要

1 標本抽出方法

標本抽出方法は、第1次抽出単位を平成17年国勢調査調査区(以下「調査区」という。)、第2次抽出単位を住戸とし、それぞれの抽出単位を層化した後に抽出を行う層化2段抽出法である。

第1次抽出では、全国から約3万調査区を抽出し、第2次抽出では、約45万住戸を抽出した。調査標本は、この方法により抽出された住戸に居住する15歳以上の世帯員全員である。

(1) 調査区の抽出(第1次抽出)

第1次抽出単位である調査区の抽出は、以下の手順により行った。

ア 標本調査区は、次のものを除く全国の調査区の中から抽出した。

- (ア) 山岳・森林・原野地帯等のある区域
- (イ) 刑務所・拘置所等のある区域
- (ウ) 自衛隊区域
- (エ) 駐留軍区域
- (オ) 水面調査区

イ 各調査区について、平成17年国勢調査の結果等に基づく特性により、次の7層に分類した。

- 層1:人口が0の調査区
- 層2:世帯数が15以下の調査区
- 層3:学生の寮・寄宿舎のある調査区
- 層4:病院・療養所及び社会施設のある調査区
- 層5:給与住宅のある調査区
- 層6:漁業の就業者の多い調査区
- 層7:上記以外の調査区

ウ 層ごとに、調査区を次の基準により配列した。

- (ア) 都道府県
- (イ) 市区町村コード(標本抽出時)
- (ウ) 平成17年国勢調査の結果等に基づく特性 ※層7のみ
- (エ) 市区町村コード(国勢調査時)
- (オ) 平成17年国勢調査調査区番号

エ ウの配列を基に、層1から層6まではそれぞれ都道府県ごとに全調査区の15歳以上人口を累積し、累積した15歳以上人口に対して確率比例系統抽出により調査区を抽出した。

層7は市区町村ごとに全調査区の15歳以上人口を累積し、累積した15歳以上人口に対して確率比例系統抽出により調査区を抽出した。

(2) 住戸の抽出(第2次抽出)

第2次抽出単位である住戸の抽出は、以下の手順により行った。

ア 標本調査区ごとに、調査区に含まれる全ての住戸を、「居住者無」、「居住者有」の順に配列した。

イ アの配列を基に、標本調査区ごとに、抽出起番号を1とし、抽出間隔を「居住者有」の住戸数を15で除し小数点以下を切り上げた値として、等確率系統抽出法により住戸を抽出した。

ウ 抽出された「居住者有」の住戸数が15未満の場合は、最初に抽出された「居住者有」の住戸の次の住戸以降の配列について、抽出住戸数が15になるまでイと同様に再度抽出を行った。

2 推定方法

結果の推定方法は、比推定の考え方に基づいている。例えば、有業者数は次の式で表される。

$$\text{有業者数} = \frac{\text{線型推定による有業者数}}{\text{線型推定による人口}} \times \text{基準人口}$$

ただし、実務上は、上記の算式を次のように用いている。基準人口は、別途推計した地域、男女、年齢階級、世帯の種類別人口を用いた。

$$\text{有業者数} = \frac{\text{線型推定による有業者数}}{\text{線型推定による人口}} \times \text{基準人口} = \text{線型推定による有業者数} \times \frac{\text{基準人口}}{\text{線型推定による人口}}$$

3 推定値の標本誤差

標本誤差の算出は、副標本方式により行った。

副標本方式による標本誤差の計算は、事後的に6組の副標本を設定し、この6組の副標本ごとに算出された推定値を用いて算出した。

前回答申における「今後の課題」への対応

諮問第13号の答申「全国消費実態調査、社会生活基本調査、就業構造基本調査及び住宅・土地統計調査に係る匿名データの作成について」（平成21年3月9日府統委第22号）における検討事項への対応

検討事項

前回の4統計調査（平成4年、9年及び14年就業構造基本調査を含む。）の匿名データに係る統計委員会の答申では、「複数の匿名データの作成の可能性についての検討」、「平成元年以前の年次拡張及び調査実施後5年基準の緩和」及び「トップコーディング等が行われた変数の平均値等（メタデータ）の整備」の3点が今後検討する事項として答申されている。

対応策

1 「複数の匿名データの作成の可能性についての検討」について

年齢各歳別の匿名データ等の作成の可能性を探るため、複数の匿名データのマッチングによる調査客体の特定の危険性に関する研究を行ったところ、現行匿名データの一部のサンプルを再抽出する方法では、危険性があることが判明した。

今後は、引き続き別の方法による作成の可能性について検討する予定である。

2 「平成元年以前の年次拡張及び調査実施後5年基準の緩和」について

平成元年以前の年次拡張については、長期の時系列分析のニーズがあることは承知しているが、よりニーズの高い直近の匿名データの提供を重視し、継続年追加を優先することで、直近の匿名データのより速やかな提供を開始したいと考えている。

また、調査実施後5年基準の緩和については、調査実施から一定の期間を経過することで、世帯や個人の属性に変化が生じ、個人が特定しにくくなる効果があることや、就業構造基本調査は5年周期であり、5年以上経過することで、次回調査を実施する際に、その次回調査の調査客体が、当該統計調査のうち最新のものに関する匿名データは提供されていないことを知ることにより、匿名データに対して抱き得る不安感を一定程度緩和する効果があると考えられるため、緩和については行わないこととする。

3 「トップコーディング等が行われた変数の平均値等（メタデータ）の整備」について

就業構造基本調査の匿名データは、年齢を85歳以上でトップコーディングしている。多変量解析等に資するため、85歳以上の平均年齢等を提供することとする。

就業構造基本調査に係る匿名データの作成の論点（案）

平成28年3月22日

1. 匿名性及び有用性の確保

平成19年就業構造基本調査（以下、「19年調査」）の匿名データ作成においては、過去に答申を得た平成4年就業構造基本調査、平成9年就業構造基本調査、平成14年就業構造基本調査（以下、「14年調査」）の匿名データ作成手法を用いつつも、社会情勢の変化や他調査の作成方法等を勘案し、以下の事項について、19年調査の作成手法を変更するもしくは変更しないこととしているが、匿名性及び有用性が確保されているか。

(1) 世帯員の年齢の匿名化措置

19年調査における世帯員の年齢の匿名化措置は14年調査から変更せず、15～84歳は5歳階級、85歳以上はトップコーディングとしているが、調査の利用目的に照らして適当か。

(2) 19年調査において変更された項目の匿名化措置

①新規調査項目の提供方法は、適当か。

②調査事項の選択肢が変更された項目は原則としてそのまま提供することとしているが、適当か。

2. 前回答申における「今後の課題」への対応

前回答申^(注)において「今後の課題」とされた以下の事項への対応は適当か。

(1) 複数の匿名データの作成

「匿名化措置を課す情報及びその程度が異なる複数の匿名データの作成の可能性について検討する」についての対応。

(2) 匿名データの提供時期の短縮化

「調査実施後5年以上経過したものを提供するという基準を緩和することについて検討する」についての対応。

(3) トップコーディング等が行われた変数

「トップコーディング等が行われた変数についても多変量解析に十分利用できるよう、当該変数の平均値等をメタデータとして整備する」についての対応。

(注) 諮問第13号の答申は、就業構造基本調査のほか、全国消費実態調査、社会生活基本調査、住宅・土地統計調査を含めた統計調査の匿名データ作成に係る答申