

## 匿名データの作成方法の変更点

### <平成19年調査の匿名データの作成方法>

平成16年調査の匿名データの作成方法(平成23年4月答申)を用いつつ、社会情勢の変化や他調査の作成方法等を勘案し、匿名データを作成するものである。



主な変更点	
① 年齢のトップコーディング (答申の「今後の課題」への対応)	個人の年齢階級の上限について、当該年の人口構成を踏まえ、「90歳以上」に変更
② 1%閾値基準に基づく上限値(グルーピング)の変更	家計支出額、普段の活動ができなかった日数のトップコーディングの変更 ※
③ 平成19年調査で把握された項目の取り扱い	<ul style="list-style-type: none"> <li>・「すぐに仕事につけるか否か」を提供</li> <li>・「こころの状態」について、閾値に満たない項目を統合のうえ提供</li> </ul>
④ 平成16年調査で提供を見送った項目への対応	<ul style="list-style-type: none"> <li>・「手助け、見守りを要する者の状況」を提供</li> <li>・「自覚症状」、「通院傷病」の項目を提供</li> </ul>

注:

- (1) 家計支出額は、家計支出総額、子への仕送り額及び育児費用でトップコーディング値が変更となった。
- (2) 「日常生活影響の就床日数」は、「普段の活動ができなかった日数」となり、設問が選択肢から実日数記入となったため、閾値基準に基づき作成した。

## 平成19年国民生活基礎調査に係る匿名データの作成方法（案）

## 1 基本的な考え方

平成16年国民生活基礎調査（以下、「平成16年調査」という。）に係る匿名データの作成方法に準拠した秘匿措置を講じて作成、提供する。

## 2 作成する匿名データの構成概要

平成16年調査に係る匿名データの種類と同様に、以下の2種類の匿名データを作成する。

匿名データの種類	匿名データの構成 (接続する調査票)	リサンプリング率 (提供レコード件数 ／調査本体のサンプルサイズ)	拡大乗数	想定される 利用ニーズ
A	世帯票、健康票	約2割 (約4万 / 約23万世帯)	全国一律の拡大乗数を再付与する。	人口、社会統計分野での世帯数の推計分析等を中心とした利用を想定
B	世帯票、健康票、 所得票、貯蓄票	約2割 (約5千 / 約2万4千世帯)	拡大乗数は付与しない。	世帯の所得及び貯蓄に関する分析等を中心とした利用を想定

平成19年国民生活基礎調査  
匿名データのチェックリスト

匿名データを作成する統計データの名称および年次

統計調査名：国民生活基礎調査

調査年：平成19年

1 地理的情報

- (1) 提供するファイルにはどのレベルの地理的情報が含まれていますか。匿名化のために地理的情報を加工していますか。

①地理情報のレベル : 全国  
②地理情報の加工の有無 : 有 無

【平成16年調査と変更なし】

項目：地域ブロック、県・指定都市番号、市郡、地区番号・単位区番号

方法：当該符号を削除し、全国一区分とするとともに、世帯単位にランダムに並び替えを行う。

- (2) 直接的な地理的情報以外に地理的情報が明らかになるような情報がありますか。

地理情報以外の地理的情報 : 有 無

- (3) 地域分析用に詳細な地理的情報を提供していますか。提供している場合、どのレベルの地理情報が含まれていますか。

地域分析用の地理情報提供の有無 : 有 無

- (4) ある特定の種類の施設であることが明らかになることはありますか。

特定の種類の施設の情報 : 有 無

2 世帯の識別情報

- (1) 世帯の識別情報として考えられるデータ項目を挙げてください。

【平成19年調査変更項目は太字下線】

世帯人員数、住居の室数、住居の床面積、同居していない者の状況、**家計支出総額**、  
親への仕送り額、**子への仕送り額**、**育児費用**、医療費世帯総額、総所得、課税等の状

況、掛金、貯蓄現在高、貯蓄減少額、借入金額

- (2) それぞれの識別情報について、どのような匿名化措置を行っていますか。また、母集団に対する割合はどのようになっていますか。

【平成 19 年調査変更項目は太字下線】

項目：世帯人員数

方法：世帯人員が 8 人以上いる世帯を削除。

項目：住居の室数

方法：10 室以上をトップコーディング。

項目：住居の床面積

方法：19.9 平方メートル以下をボトムコーディング、300 平方メートル以上をトップコーディング。

項目：同居していない者の状況

方法：「老人福祉施設入所者」、「社会福祉施設入所者」、「長期入院者」を「入院・入所」に統合する。

項目：家計支出総額

方法：単独世帯の 120 万円以上をトップコーディング、2 人以上世帯の 200 万円以上 をトップコーディング。(平成 16 年は 300 万円以上をトップコーディング)

項目：親への仕送り額

方法：6 万円以上をトップコーディング。

項目：子への仕送り額

方法：16 万円以上をトップコーディング。(平成 16 年は 20 万円以上をトップコーディング)

項目：育児費用

方法：7 万円以上をトップコーディング。(平成 16 年は 6 万円以上をトップコーディング)

項目：医療費世帯総額

方法：単独世帯の 6 万円以上をトップコーディング、2 人以上世帯の 18 万円以上をトップコーディング。

項目：総所得

方法：単独世帯の 1100 万円以上をトップコーディング、2 人以上世帯の 2200 万円以上をトップコーディング。

項目：課税等の状況

方法：単独世帯の 250 万円以上をトップコーディング、2 人以上世帯の 490 万円以上をトップコーディング。

項目：掛金

方法：単独世帯の 40 万円以上をトップコーディング、2 人以上世帯の 80 万円以上をトップコーディング。

項目：貯蓄現在高

方法：単独世帯の 6300 万円以上をトップコーディング、2 人以上世帯の 9000 万円以上をトップコーディング。

項目：貯蓄減少額

方法：単独世帯の 800 万円以上をトップコーディング、2 人以上世帯の 1300 万円以上をトップコーディング。

項目：借入金額

方法：単独世帯の 2400 万円以上をトップコーディング、2 人以上世帯の 4000 万円以上をトップコーディング。

(3) 世帯単位のデータを提供することに対応して特別な匿名化措置を行っていますか。

特別な措置： 行っている 行っていない

【平成 16 年調査と変更なし】

項目：父子世帯

方法：削除

項目：年齢差の大きい夫婦のいる世帯

方法：夫が妻より 20 歳以上、妻が夫より 15 歳以上の夫婦のいる世帯を削除。

項目：年齢差の大きい又は小さい親子のいる世帯

方法：父親と末子の年齢差が 55 歳以上又は母親と末子の年齢差が 50 歳以上の世帯を削除。父親と長子の年齢差が 15 歳以下又は母親と長子の年齢差が 10 歳以下の世帯を削除。

項目：同一年齢階級に4人以上の人員がいる世帯  
方法：削除

### 3 個人の識別情報

(1) 個人の識別情報として考えられるデータ項目を挙げてください。

【平成19年調査変更項目は太字下線】

年齢、公的年金・恩給の受給状況、手助けや見守りの要否、要介護認定の有無、仕事の有無、別居の子の数、職業分類番号、就業期間、1週間の就業時間、1日の片道通勤時間、自立期間、主な介護者の続柄、最も気になる傷病、普段の活動ができなかった日数、悩みやストレスの相談状況、最も気になる悩みやストレスの相談状況、この状態

(2) それぞれの識別情報について、どのような匿名化措置を行っていますか。また、母集団に対する割合はどのようになっていますか。

【平成19年調査変更項目は太字下線】

項目：年齢

方法：5歳階級で提供。90歳以上をトップコーディング。（平成16年は85歳以上をトップコーディング）

項目：公的年金・恩給の受給状況

方法：「福祉年金」、「恩給」を「その他」に統合。

項目：手助けや見守りの要否

方法：手助けや見守りを要する者が2人以上いる世帯を削除。

項目：要介護認定の有無

方法：要介護認定を受けている者が2人以上いる世帯を削除。

項目：仕事の有無

方法：「主に通学で仕事あり」、「家族・通学以外のことが主で仕事あり」を「その他」に統合。

項目：別居の子の数

方法：4人以上をトップコーディング。

項目：職業分類番号

方法：「農業作業者」「林業作業者」「漁業作業者」を「農林漁業従事者」に統合。

項目：就業期間

方法：50年以上をトップコーディング。

項目：1週間の就業時間

方法：80時間以上をトップコーディング。

項目：1日の片道通勤時間

方法：90分以上をトップコーディング。

項目：自立期間

方法：「1月未満」、「1月～3月未満」、「3月～6月未満」、「6月～1年未満」を統合。  
(新規提供項目)

項目：主な介護者の続柄

方法：「その他の親族」を「その他」に統合。(新規提供項目)

項目：最も気になる傷病

方法：各傷病を対応する上位区分に統合。

項目：普段の活動ができなかった日数

方法：25日以上をトップコーディング。(平成16年は4日以上をトップコーディング)

項目：悩みやストレスの相談状況

方法：「民間の相談機関」、「テレビ等」及び「01～07以外」を統合。

項目：最も気になる悩みやストレスの相談状況

方法：「民間の相談機関」、「テレビ等」及び「01～07以外」を統合。

項目：こころの状態

方法：「いつも」「たいてい」の回答個数が3個以上を統合。(新規提供項目)

#### 4 誤差（ノイズ）

匿名化措置として、誤差を付加する方法を採っていますか。誤差を付加する方法を採っている場合には、その方法を記載してください。

誤差の付加： 採用している 採用していない

## 5 リサンプリング

匿名化措置として、リサンプリングをしていますか。リサンプリングをしている場合には、その抽出方法と抽出率を記載してください。

リサンプリング：  行っている  行っていない

※「行っている」場合

【平成 16 年調査と変更なし】

全国一律の拡大乗数とするため、以下の方法でリサンプリングを行う。

抽出方法：

国勢調査区抽出（第一段）：都道府県・指定都市の国勢調査区数に比例するよう、都道府県・指定都市別に国勢調査区をリサンプリング。

世帯抽出（第二段）：第一段でリサンプリングされた国勢調査区から世帯をリサンプリング。

抽出率：全体の約 2 割

## 6 外部の情報

(1) 個人・世帯を特定できる可能性のある外部の情報は存在しますか。

外部情報：  有  無

(2) 母集団情報として利用している情報は何か。

国勢調査の調査区名簿

行政記録から作成した名簿（行政記録の名称： \_\_\_\_\_）

その他（具体的に記載 \_\_\_\_\_）

## 7 その他

(1) データの一連番号、データの並び順について、何らかの匿名化措置を施していますか。

実施している  実施していない

【平成 16 年調査と変更なし】

データの並び順は、世帯単位に乱数によりランダムに並び替える。ランダムな並びにした後、データの世帯番号を付与する。

(2) サンプリング情報によって、地理的情報以外に特定の地域や集団であることが明らかになる可能性はありますか。

ない。

(3) 提供時期と調査時点とはどの程度の期間が開いていますか。

5 年以上。



(4) そのほか、データを匿名化するに当たり、措置していることがありますか。

ない。