

令和元年 11 月 27 日

国勢調査の匿名データの年齢の中位数の求め方（案）

平成 12 年国勢調査の匿名データを作成する際に、「年齢」については、調査票情報の「各歳」から「0～84 歳を 5 歳階級でグルーピング及び 85 歳以上をトップコーディング」の匿名化の処理を行った上で提供をしている。

また、匿名データの有用性に関する情報の 1 つとして、調査結果で公表されており、比較可能な「中位数」の提供を予定している。匿名データの年齢から中位数を以下のように度数分布から中位数を求める方法で算出を検討している。

年齢	下限値	級間隔	人数	累積人数
0～4歳	0	5	1,356	1,356
5～9歳	5	5	1,516	2,872
10～14歳	10	5	1,689	4,561
15～19歳	15	5	1,771	6,332
20～24歳	20	5	1,870	8,202
25～29歳	25	5	2,193	10,395
30～34歳	30	5	1,909	12,304
35～39歳	35	5	1,913	② 14,217
40～44歳	⑤ 40	④ 5	③ 1,893	16,110
45～49歳	45	5	2,286	18,396
50～54歳	50	5	2,456	20,852
55～59歳	55	5	2,034	22,886
60～64歳	60	5	1,710	24,596
65～69歳	65	5	1,489	26,085
70～74歳	70	5	1,305	27,390
75～79歳	75	5	920	28,310
80～84歳	80	5	585	28,895
85歳以上	85	—	458	29,353

(+ 2) → ① 14,676.5

1. 年齢が小さい階級から順に、累積人数を求める。また、階級の下限值と級間隔の列を追加する。

2. 中位数は、「個人を年齢順に並べたとき、その中央で 2 等分する境界点にある年齢」を指すので、ここでは年齢の階級が小さい方から、29,353 人の半分である、14,676.5 人目の値を求めることとなる。

この境界点がどの階級にあるのかをみると、35～39 歳の累積人数が 14,217 人、40～44 歳の累積人数が 16,110 人であることから、40～44 歳のどこかにあることとなる。

3. 40～44 歳のうち、境界点がどの位置にあるのかを計算する。

度数分布から中位数を求める場合、階級内に値が均等に分布していることを前提としているため、階級の下限を 0、上限を 1 とした際の境界点の位置は、上記の表の (①—②) ÷ ③ で求めることが可能である。

これにより、境界点の位置は、

$$(14,676.5 - 14,217) \div (1,893) = 0.24$$

となる。

ところで、この区間の級間隔は④のとおり 5 歳であることから、先ほどの 0.24 に 5 を乗じる。更にこの区間の下限値は⑤のとおり 40 歳なので、これを加えることより、41.2 歳が匿名データの年齢中位数となる。