

国勢調査における不詳補完値の試算 — CANCEIS による全国データの補完を中心に —

北原 昌嗣*
會澤 豪大†
羽田 大起‡

Trial Estimation of the Imputed Values in Population Census - Focusing on National Results Imputed by CANCEIS -

KITAHARA Masatsugu
AIZAWA Takehiro
HANEDA Hiroki

2020年国勢調査では、結果利用者の利便性を図るために参考表として、あん分等による不詳補完値を算出し公表しているが、結果表レベルでのあん分による不詳補完になっているため、粒度に限界があることや集計区分間の整合性が取れないなどの課題がある。そこで、不詳を補完するためのソフトウェアとして、カナダ統計局の開発した補完システム「CANCEIS」を利用し、不詳補完値を試算した。CANCEISによる不詳補完は、個票レベルで実行するため、これらの課題を打開できる補完方法であり、開発国であるカナダ以外の人口センサスでも利用されている。

本稿では、全都道府県において、CANCEISによる不詳補完値を試算し、その結果を合算することで全国結果を作成した。そして、その試算した全国結果を用いて、あん分等による不詳補完値との比較や、CANCEISによる補完数を推定した不詳率と比較することで、妥当性を検証した。そのほか、他の公的統計の結果数値も参考にして、CANCEISによる不詳補完値の妥当性を検証している。

キーワード：国勢調査、補完方法

For the 2020 Population Census, the imputed values based on proportional distribution, etc. were calculated and published as reference tables to improve the convenience of users. However, imputed values based on proportional distribution, etc. have issues such as limited granularity and a lack of consistency between result tables because they are calculated at the results table level. Therefore, we used CANCEIS, an imputation system developed by Statistics Canada, to calculate imputed values. CANCEIS imputation is performed at the individual record level, making it an imputation method that can overcome these issues. It is also used in the Population Census in countries other than Canada, which developed CANCEIS.

In this paper, we calculated the imputed values by using CANCEIS for all prefectures and totaled the results up to make national results. We then used these calculated national results to verify their validity by comparing them with the imputed values based on proportional distribution, etc. and by comparing the imputed numbers by using CANCEIS for each aggregation item with the estimated “unknown” rate. In addition, we also refer to the results of other official statistics to verify the validity of the imputed values by using CANCEIS.

Key words: Population Census, Imputation method

* 総務省統計局統計調査部国勢統計課 Email: mkitahara@soumu.go.jp

† 独立行政法人統計センター統計編成部人口統計編成課 Email: taizawa@nstac.go.jp

‡ 独立行政法人統計センター統計編成部人口統計編成課 Email: hhaneda2@nstac.go.jp

1 はじめに¹

国勢調査では、結果利用者の利便性向上のために、2015年から不詳を除いた国勢調査のデータ構成比を利用した「あん分」により、不詳を補完した結果（不詳補完値）を参考表として提供している。「あん分」による補完方法は、国際的にも認知された一般的な不詳を補完する方法であるが、一方で、結果表レベルでのあん分による集計となるため、集計区分間の整合性が取れないなどの課題がある。その解決策として、主要国で実施されているソフトウェアによる個票レベルの補完が挙げられ、本稿では、カナダ統計局が開発した補完システムであるCANCEIS(CANadian Census Edit and Imputation System)を利用することで、不詳補完値を試算し、その妥当性について検証した。北原(2025)において試算したCANCEISによる補完結果は、CANCEISを実行するための要でもあるモジュールを都道府県別に作成する必要があることや、演算時間の確保が困難だったことから、主に愛知県のみを掲載していたが、本稿では全都道府県において、CANCEISによる補完結果を作成し、それを合算することで全国結果を試算した。

CANCEISについて簡単に説明すると、CANCEISは最近隣法(NIM)²を基本原理とした国勢調査(人口センサス)に特化した補完システムであり、大容量データを効率的に処理するアルゴリズムを持つため処理速度が速く、開発国のカナダだけでなく、イギリスやドイツにおける人口センサスでも利用されており、信頼性の高さにおいて十分に実績のあるシステムである³。

今般、都道府県別にモジュールを作成し、CANCEISを各都道府県で実行することにより、全国のCANCEISによる不詳補完値を試算し、その妥当性⁴について、あん分等による不詳補完値や他の公的統計とも比較することで検証した。また、各集計事項において、推定した不詳率とCANCEISによる補完数を比較することで、CANCEISがその集計事項に補完を実行した意味をマクロ的な観点から考察した。

2 CANCEISによる不詳補完値⁵の検証

2020年国勢調査結果を用いて、全ての集計区分⁶でCANCEISによる不詳補完値を全都道府県の国勢調査データを利用して試算⁷した結果について、2.1 CANCEIS補完値の人口及び構成比による検証、2.2 各集計事項におけるCANCEIS補完値の検証を実施した。

2.1 CANCEIS補完値の人口及び構成比による検証

本項では、原数値⁸、あん分等による不詳補完値(以下、あん分等補完値)及びCANCEISによる不詳補完値(以下、CANCEIS補完値)の全国結果における人口や構成比⁹を比較することで、

¹ 本稿の内容は個人的見解も含まれており、筆者の所属する機関の公式見解を示すものではない。

² 国連統計部の発刊している「人口・住宅センサスのエディティングに関するハンドブック (Handbook on Population and Housing Census Editing Revision 2)」では、CANCEISに使われている補完方法は、単純な最近隣法(NIM: Nearest-neighbor Imputation Methodology)ではなく、フェレギーホルツ法(Fellegi-Holt Method)の最小変化量の概念も取り入れた新インピュテーション法(New Imputation Method)であると評している。

³ CANCEISの詳細な解説については、北原・寺垣内(2023)、北原(2025)を参照されたい。

⁴ 妥当性の定義は、真値が不明である以上、定義付けることが困難であるが、各種統計や行政記録情報の傾向、推定値を元に顕著な乖離が認められるかどうかを妥当性の判断材料としている。

⁵ 本節のCANCEISによる不詳補完値は、総務省統計局「国勢調査」の調査票情報を用いて、CANCEIS version5.4により独自に算出し、集計したものである。

⁶ 全ての集計区分とは、人口等基本集計、就業状態等基本集計、従業地・通学地集計及び人口移動集計のことである。抽出詳細集計は、抽出による推計値であるため除いている。また、試算に使用した調査項目は、簡易調査項目に限定している(教育や利用交通手段を補完対象から除いている。)

⁷ CANCEISによる不詳補完手順については、北原(2025)を参照されたい。

⁸ 本稿の「原数値」は、国勢調査の「公表値」と同義である。

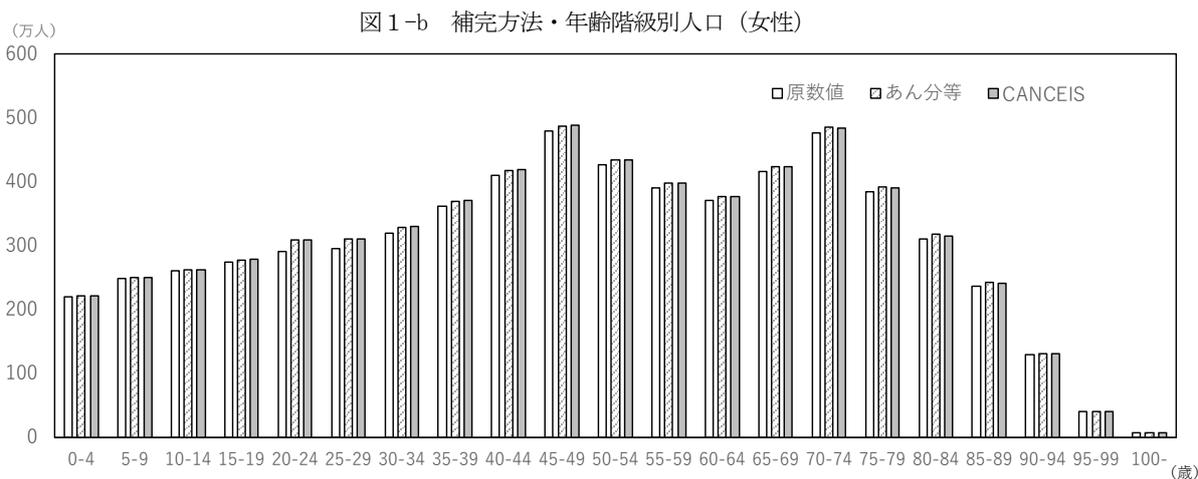
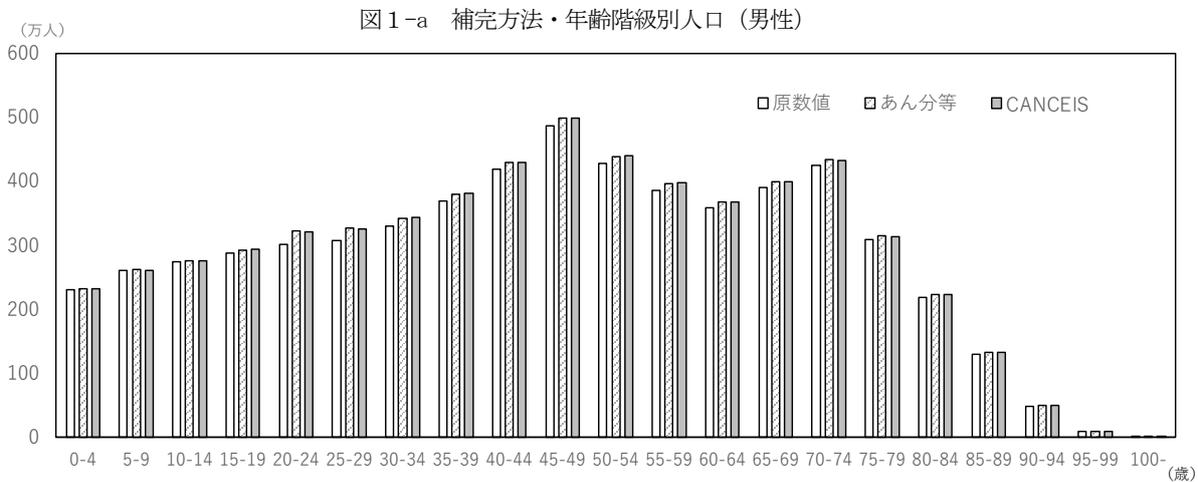
⁹ 「原数値」及び「CANCEIS補完値」の構成比については、分母である「総数」から「不詳」を除いて算出している。「CANCEIS補完値」においては、CANCEISによって補完ができなかった「不詳」が極めて少ないものの残存している。「あん分等補完値」においては、「不詳」は存在していない。

CANCEIS 補完値の妥当性を検証した。妥当性の検証には、あん分等補完値の構成比を参考指標としている。

また、構成比よりも順位に基づく検証の方が妥当であると考えられる「国籍」、「5年前の常住地」及び「従業地・通学地」については、スピアマン(Spearman)の順位相関係数¹⁰を見ることで補完方法間の単調性(順位の類似性)を評価した。

2.1.1 年齢(人口等基本集計)

図1-a及び図1-bは、補完方法・男女・年齢階級別人口のグラフであり、「原数値」、「あん分等補完値(以下、グラフや表上では「あん分等」)」及び「CANCEIS 補完値(以下、グラフや表上では「CANCEIS」)」を比較している。このグラフを見ると、男女ともに各年齢階級で、「原数値」と比較して、「あん分等」と「CANCEIS」が近接した数値になっている。また、表1-a及び表1-bは、補完方法・年齢階級別人口の構成比であり、「原数値」「あん分等」「CANCEIS」は、ほぼ等しい値になっている。



¹⁰ 順位付けに利用した令和2年国勢調査の国名や市区町村別結果等は、Q-Qプロット(Quantile-Quantile Plot)により確認したところノンパラメトリック(非正規分布)であったことから、スピアマンの順位相関係数 r_s により評価した。

大きさ n の組標本 $(x_i, y_i) (i = 1, \dots, n; n \geq 2)$ があるとき、昇順に並べてある x 標本の数値の順位は i であり、それに対応する y 標本の数値の順位を Y_i とすると、スピアマンの順位相関 r_s は次式となる。

$$r_s = 1 - \frac{6 \sum_{i=1}^n (i - Y_i)^2}{n(n^2 - 1)} \quad \text{※同順位については、当該順位の総和を個数で除した順位に変換}$$

表 1-a 補完方法・年齢階級別人口の構成比（男性）

年齢階級	構成比（%）			構成比の差（%ポイント）	
	①原数値	②不詳補完値	③CANCEIS	③-①	③-②
0-4	3.9	3.8	3.8	▲0.1	0.0
5-9	4.4	4.3	4.3	▲0.1	0.0
10-14	4.6	4.5	4.5	▲0.1	0.0
15-19	4.8	4.8	4.8	0.0	0.0
20-24	5.1	5.3	5.2	0.1	▲0.1
25-29	5.1	5.3	5.3	0.2	0.0
30-34	5.5	5.6	5.6	0.1	0.0
35-39	6.2	6.2	6.2	0.0	0.0
40-44	7.0	7.0	7.0	0.0	0.0
45-49	8.1	8.1	8.1	0.0	0.0
50-54	7.2	7.2	7.2	0.0	0.0
55-59	6.5	6.5	6.5	0.0	0.0
60-64	6.0	6.0	6.0	0.0	0.0
65-69	6.5	6.5	6.5	0.0	0.0
70-74	7.1	7.1	7.1	0.0	0.0
75-79	5.2	5.1	5.1	▲0.1	0.0
80-84	3.7	3.6	3.6	▲0.1	0.0
85-89	2.2	2.2	2.2	0.0	0.0
90-94	0.8	0.8	0.8	0.0	0.0
95-99	0.2	0.2	0.2	0.0	0.0
100-	0.0	0.0	0.0	0.0	0.0

表 1-b 補完方法・年齢階級別人口の構成比（女性）

年齢階級	構成比（%）			構成比の差（%ポイント）	
	①原数値	②不詳補完値	③CANCEIS	③-①	③-②
0-4	3.5	3.4	3.4	▲0.1	0.0
5-9	3.9	3.8	3.9	0.0	0.1
10-14	4.1	4.0	4.0	▲0.1	0.0
15-19	4.3	4.3	4.3	0.0	0.0
20-24	4.6	4.8	4.8	0.2	0.0
25-29	4.7	4.8	4.8	0.1	0.0
30-34	5.0	5.1	5.1	0.1	0.0
35-39	5.7	5.7	5.7	0.0	0.0
40-44	6.5	6.4	6.5	0.0	0.1
45-49	7.5	7.5	7.5	0.0	0.0
50-54	6.7	6.7	6.7	0.0	0.0
55-59	6.1	6.1	6.1	0.0	0.0
60-64	5.8	5.8	5.8	0.0	0.0
65-69	6.6	6.5	6.5	▲0.1	0.0
70-74	7.5	7.5	7.5	0.0	0.0
75-79	6.0	6.0	6.0	0.0	0.0
80-84	4.9	4.9	4.9	0.0	0.0
85-89	3.7	3.7	3.7	0.0	0.0
90-94	2.0	2.0	2.0	0.0	0.0
95-99	0.6	0.6	0.6	0.0	0.0
100-	0.1	0.1	0.1	0.0	0.0

2.1.2 続柄（人口等基本集計）

表 2 は、一般世帯における補完方法・世帯主との続柄別人口及び構成比である。人数については、続柄の不詳が補完されている分だけ「原数値」よりも「CANCEIS」の方が多くなっているが、構成比で見れば、ほぼ等しい値になっている。

「増加数」の構成比を見ると、続柄「不詳」のうちの約半数（50.3%）が「子」に補完、約3分の1（31.8%）が「配偶者」に補完されている。これら2つの集計事項で「不詳」の8割強を占めており、この傾向は、補完の構造から見て妥当と考えられる。単独（単身）世帯の続柄不詳は、CANCEIS 補完前に、既に「世帯主」に補完されているため、続柄「不詳」は必然的に2人以上の世帯に限定される。したがって、人数の多い、「配偶者」と「子」の補完が多くなるのは自然なことであり、世帯主の「配偶者」は1世帯あたり最大1人に対して、「子」は複数ありえるため、子>配偶者>世帯主の父母>孫・・・の順に補完されたと考えられる。また、平成27年国勢調査事後調査（以下、事後調査）の結果によると、「本体調査で把握されなかった」世帯員は、子>配偶者>世帯主の父母>孫・・・の順になっており、CANCEIS の補完結果とも一致する¹¹。

表 2 補完方法・世帯主との続柄別人口及び構成比（一般世帯）

世帯主との続柄	人数（人）			構成比（%）			構成比の差(%ポイント) CANCEIS - 原数値
	原数値	CANCEIS	増加数	原数値	CANCEIS	増加数	
配偶者	27,662,817	27,761,075	98,258	41.2	41.2	31.8	0.0
子	31,990,501	32,146,054	155,553	47.6	47.7	50.3	▲0.1
子の配偶者	901,647	909,298	7,651	1.3	1.3	2.5	0.0
世帯主の父母	2,595,771	2,612,114	16,343	3.9	3.9	5.3	0.0
世帯主の配偶者の父母	523,555	525,488	1,933	0.8	0.8	0.6	0.0
孫	2,016,335	2,031,353	15,018	3.0	3.0	4.9	0.0
祖父母	37,643	37,953	310	0.1	0.1	0.1	0.0
兄弟姉妹	705,702	712,390	6,688	1.1	1.1	2.2	0.0
他の親族	188,116	190,343	2,227	0.3	0.3	0.7	0.0
住み込みの雇人	12,992	13,148	156	0.0	0.0	0.1	0.0
その他	513,243	518,324	5,081	0.8	0.8	1.6	0.0
不詳	309,724	506	-	-	-	-	-

注1) 2020年国勢調査で、参考値として公表している「あん分等」による不詳補完値では、続柄の場合、区分数が多いことなどから不詳補完値を作成していない。

注2) 一般世帯の単独世帯を除く続柄を表章している。また、「世帯主」はCANCEISによる補完の前後で数値が変化しないため表章から除いている。（∵「世帯主」は一意に定まるため、「不詳」が存在しない。）

¹¹ 平成27年事後調査結果を見ると、配偶者の次に「本体調査で把握されなかった」世帯員が多いのは、「その他」であるが、「その他」には、表2には含まれていない施設等世帯の寄与があるため除いて評価をしている。

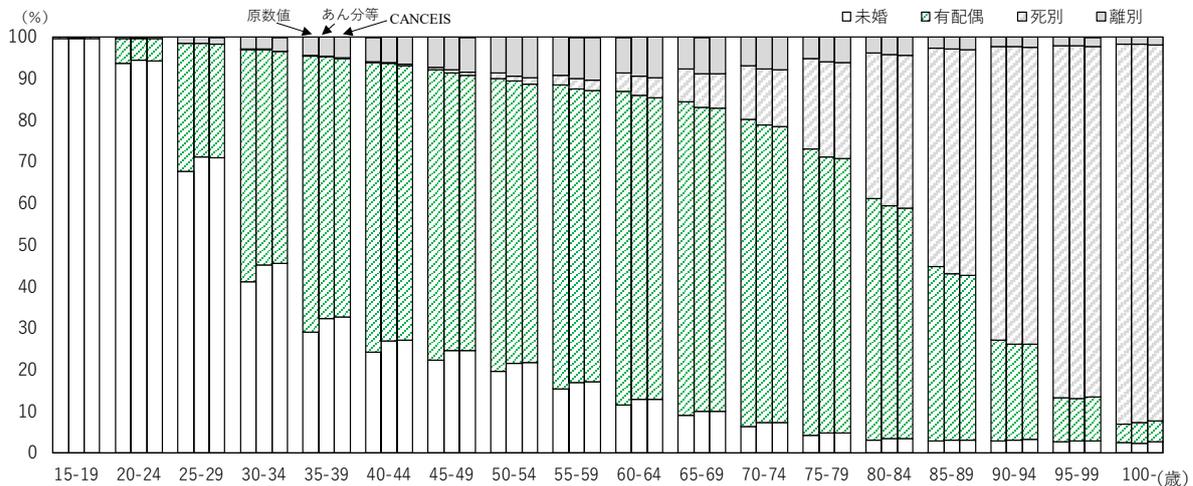
2.1.3 配偶関係（人口等基本集計）

図2は、補完方法・配偶関係別人口の構成比のグラフである。配偶関係においても、「CANCEIS」は「原数値」と比較して、「あん分等」に近接した構成比になっている。また、「原数値」と比較すると「あん分等」及び「CANCEIS」では、「未婚」の割合が増加、「有配偶」の割合が減少している。

事後調査では、配偶関係別に見ると、「未婚」、「死別」及び「離別」は、「有配偶」より「本体調査で把握されなかった」世帯員の割合が高いと分析している。そして、阿部（2013）は、「不詳」人口の未婚率のほうが、非「不詳」人口のそれよりも高いと考えられることから、「不詳」を除く人口から算出された未婚率は、過少推計の可能性があることに言及しており、宇南山（2013）は人口動態統計を利用した補正推計によって、2010年国勢調査の「20～39歳」の未婚率が過少評価されていることを指摘している。これは、図2の両補完値における「未婚」の割合が、「25～39歳」で顕著に増加していることとも一致する。

以上のことから、不詳は未婚者に多い可能性があるため、CANCEIS 補完値において、「未婚」の割合が増加することは妥当であると考えられる。

図2 補完方法・配偶関係別構成比



2.1.4 国籍（人口等基本集計）

2020年国勢調査において国籍で使用されている国名は195か国（日本を除く。）であり、国名が多いことから、あん分により不詳補完値を算出することが難しく、あん分等補完値に国名別の国籍は存在していない。一方、CANCEIS 補完値の場合、個票レベルでの補完になることから国名を補完した不詳補完値の作成が可能¹²であるが、データの並び上、少数派の国籍が大量に補完される事象が起きてしまった場合に、歪んだ結果になってしまうことが国籍補完におけるCANCEIS 補完値の懸念点として挙げられる¹³。国籍においては、全国でどの国籍が何人いるかを行政記録情報（在留外国人統計）により把握することができるため、特に少数派の国名が大量に増加するような補完結果が算出された場合には、国勢調査と行政記録情報を比較することで、結果に疑念を持たれる可能性がある。

表3-a及び表4-aは、補完方法・国名別人口ランキング（男女別）であるが、人数の大きい順に、在留外国人統計>CANCEIS>原数値となっている。CANCEIS による補完を実施しても、在

¹² 日本人・外国人別の国籍は、あん分等補完値にも存在する。

¹³ CANCEIS は不詳のあるユニット（レシピエント）からデータの並びに近い順に、不詳のない整合性の取れたユニットをドナー候補として選定してくるため、同じ国籍の外国人データが近隣に密集している場合、補完によって偏りが生じる可能性がある。

留外国人統計の人口が大きくなる主な要因は、CANCEISでは「無国籍・国名不詳」が約13万件あるためである。「無国籍・国名不詳」は「無国籍」と「国名不詳」に分割することができないため、CANCEISは、「国名不詳」を不詳補完の対象にできず、国籍が「無国籍・国名不詳」である（「不詳」ではない）ドナーとして扱っている。一方の在留外国人統計では、「無国籍・国名不詳」が700件ほどで国勢調査と比較して少ない。また、調査方法や調査時点の違いもあることから、在留外国人統計よりもCANCEISの方が小さい値になっているが、順位においては、各補完方法でほぼ同じ順位となった。「原数値」「CANCEIS」「在留外国人統計」におけるスピアマンの順位相関係数を見ると、ほぼ1になっていることから、順位相関という観点では「原数値」「CANCEIS」「在留外国人統計」は極めて類似した結果であると言える（表3-b、表4-b）。そして、CANCEISによる補完結果からは、懸念していた少数派の国名の大量補完は生じていないことが確認できた。

日本において外国人は少数派であり、CANCEISによる外国人への補完の弱さが見られることや、事後調査により、国籍が「外国」である世帯員は、国籍が「日本」である世帯員と比べて「本体調査で把握されなかった」世帯員の割合が高いと分析されていることから、推定が可能な単身の外国人については、行政記録情報である在留外国人統計データを利用して、部分的な補完をCANCEIS実行前に実施した上で、CANCEISによる補完を実施している¹⁴。

表 3-a 補完方法・国名別人口ランキング（男性）

【男性】 国名	人数（人）			順位		
	原数値	CANCEIS	(参考)在留 外国人統計	原数値	CANCEIS	(参考)在留 外国人統計
中国	290,994	345,858	377,681	1	1	1
ベトナム	174,095	214,406	238,540	2	2	2
韓国, 朝鮮	172,515	201,774	215,207	3	3	3
ブラジル	96,506	105,154	114,468	4	4	4
フィリピン	67,887	76,677	85,627	5	5	5
ネパール	38,461	48,689	54,849	6	6	6
アメリカ	32,829	36,564	38,375	7	8	8
インドネシア	32,405	37,682	44,969	8	7	7
ペルー	21,392	22,936	25,279	9	9	10
インド	18,366	22,033	27,260	10	10	9
スリランカ	13,121	15,767	20,186	11	11	11
タイ	11,424	12,739	15,358	12	13	13
ミャンマー	11,243	13,074	15,562	13	12	12
イギリス	10,422	11,339	13,240	14	14	15
パキスタン	8,536	9,533	13,475	15	15	14
バングラデシュ	7,845	8,855	11,509	16	16	16
フランス	7,141	8,139	9,469	17	17	17
カナダ	5,900	6,370	7,443	18	18	19
オーストラリア	5,433	5,904	7,368	19	20	20
カンボジア	5,407	6,113	8,209	20	19	18
⋮	⋮	⋮	⋮	⋮	⋮	⋮

表 3-b スピアマンの順位相関係数（国名・男性）

比較対象	スピアマンの 順位相関係数 (国名・男性)
原数値×CANCEIS	0.99982
原数値×在留外国人統計	0.99627
CANCEIS×在留外国人統計	0.99645

表 4-a 補完方法・国名別人口ランキング（女性）

【女性】 国名	人数（人）			順位		
	原数値	CANCEIS	(参考)在留 外国人統計	原数値	CANCEIS	(参考)在留 外国人統計
中国	376,481	436,078	469,083	1	1	1
韓国, 朝鮮	202,078	232,696	247,947	2	2	2
フィリピン	162,464	182,906	196,396	3	3	3
ベトナム	146,710	168,556	181,875	4	4	4
ブラジル	83,508	89,791	96,710	5	5	5
タイ	31,278	34,667	37,986	6	7	7
ネパール	28,864	35,819	40,518	7	6	6
ペルー	19,642	20,933	23,116	8	8	8
インドネシア	16,742	18,357	21,115	9	9	9
アメリカ	15,046	16,651	18,839	10	10	10
ミャンマー	13,673	15,441	17,741	11	11	11
インド	9,549	10,675	12,504	12	12	12
カンボジア	5,472	5,920	7,447	13	14	14
スリランカ	5,460	5,981	7,549	14	13	13
ロシア	4,840	5,196	6,162	15	16	16
モンゴル	4,814	5,352	6,681	16	15	15
バングラデシュ	4,029	4,260	5,326	17	17	17
マレーシア	3,942	4,184	4,877	18	18	18
パキスタン	3,435	3,546	4,821	19	19	19
イギリス	3,168	3,382	4,538	20	20	20
⋮	⋮	⋮	⋮	⋮	⋮	⋮

表 4-b スピアマンの順位相関係数（国名・女性）

比較対象	スピアマンの 順位相関係数 (国名・女性)
原数値×CANCEIS	0.99989
原数値×在留外国人統計	0.99649
CANCEIS×在留外国人統計	0.99647

¹⁴ 詳細については、北原(2025)を参照されたい。

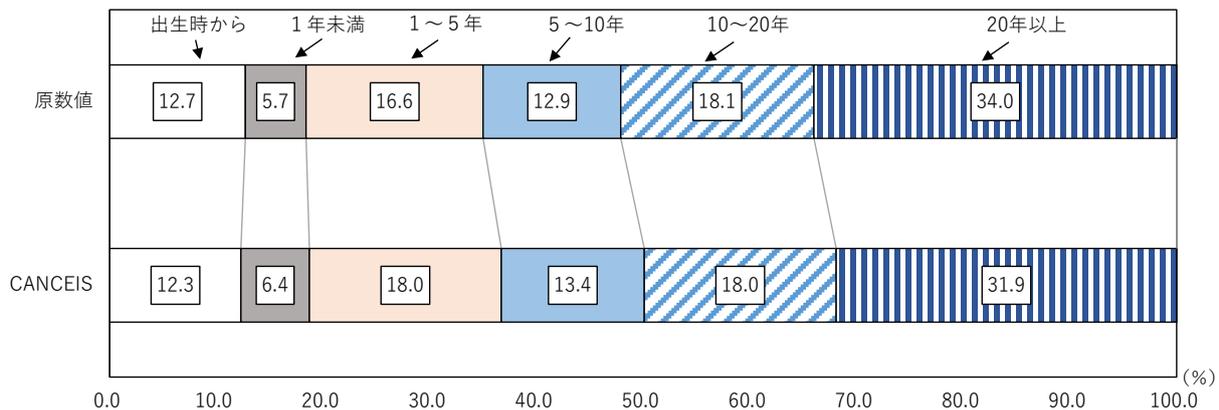
2.1.5 居住期間（人口等基本集計）

表5は、補完方法・居住期間別人口であり、居住期間「1から5年」が最も増加しており、次に「10から20年」が増加している。図3は、補完方法・居住期間別人口の構成比を表しており、居住期間が「出生時から」及び「20年以上」については、「原数値」よりも「CANCEIS」の方の構成比が減少していることから、これら集計事項では、CANCEISによる補完が他の集計事項より少なかったと言える。逆に、「1から5年」などの集計事項では、構成比は増加しており、CANCEISによる補完が他の集計事項より多かったと言える。すなわち、CANCEISによる補完結果からは、居住期間が長い人ほど調査に回答している可能性が高いことが読み取れる。

表5 補完方法・居住期間別人口 (人)

居住期間	原数値	CANCEIS	増加数
出生時から	14,214,580	15,538,948	1,324,368
1年未満	6,388,862	8,018,425	1,629,563
1から5年	18,477,107	22,670,106	4,192,999
5から10年	14,394,320	16,924,137	2,529,817
10から20年	20,139,057	22,706,627	2,567,570
20年以上	37,913,458	40,285,435	2,371,977
不詳	14,618,715	2,421	-
総数	126,146,099	126,146,099	-

図3 補完方法・居住期間別構成比



2.1.6 5年前の常住地（前住地）人口（人口移動集計）

表6は、補完方法・前住地別人口であり、「あん分等」と「CANCEIS」が類似した数値となっている。図4は、補完方法・前住地別人口の構成比であり、「あん分等」や「CANCEIS」では、「原数値」と比較して、前住地が「現住所」の構成比が減少、「他市区町村」の構成比が増加している。すなわち補完結果からは、前住地が「現住所」である人に「不詳」が少なく、前住地が「他市区町村」である人に「不詳」が多いと読み取ることができる。

表7-a及び表8-aは、補完方法・前住地別人口ランキング（他市区町村及び他都道府県）であるが、「あん分等」「CANCEIS」が類似した数値になっている。表7-b及び表8-bは、スピアマンの順位相関係数であるが、他市区町村及び他都道府県の両者ともに1に近い数値になっており、順位相関という観点からは、「原数値」「あん分等」「CANCEIS」は極めて類似した結果になったと言える。

表6 補完方法・前住地別人口 (人)

前住地	原数値	あん分等	CANCEIS
総数（常住者）	126,146,099	126,146,099	126,146,099
現住所	87,851,473	97,474,410	96,774,209
自市区町村	10,032,027	11,595,423	12,021,918
他市区町村	12,597,830	16,040,973	16,337,517
国外	769,382	1,035,293	1,010,052
前住地区分不詳	14,895,387	0	2,403

図4 補充方法・前住地別構成比

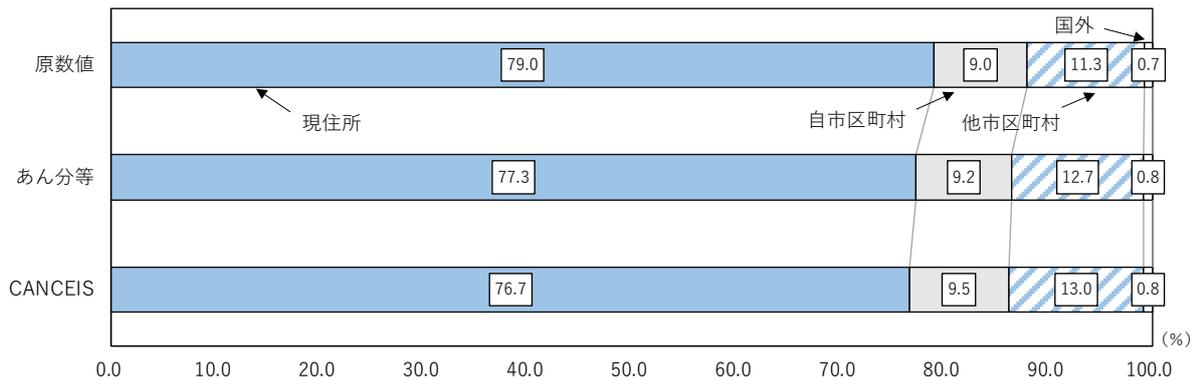


表7-a 補充方法・前住地別人口ランキング (他市区町村)

前住地 (他市区町村)	人数 (人)			順位		
	原数値	あん分等	CANCEIS	原数値	あん分等	CANCEIS
東京都世田谷区	121,108	167,193	168,698	1	1	1
東京都大田区	89,409	121,485	122,106	2	3	3
東京都練馬区	89,224	121,813	122,785	3	2	2
東京都杉並区	82,932	115,743	116,824	4	4	4
東京都江戸川区	73,449	98,598	99,729	5	5	5
東京都板橋区	67,961	93,033	94,404	6	6	6
千葉県船橋市	63,534	84,321	84,469	7	8	8
東京都江東区	62,760	85,317	86,062	8	7	7
千葉県市川市	61,351	82,106	82,128	9	9	10
東京都足立区	61,066	81,723	82,982	10	10	9
東京都新宿区	56,495	80,474	81,363	11	11	11
東京都品川区	55,433	76,579	77,573	12	12	12
鹿児島県鹿児島市	54,527	66,605	67,925	13	18	18
兵庫県西宮市	54,282	71,470	72,240	14	15	15
埼玉県川口市	53,193	70,637	71,507	15	16	16
東京都中野区	53,134	75,289	76,102	16	13	13
東京都八王子市	51,663	72,958	73,414	17	14	14
神奈川県横浜市港北区	51,161	68,339	69,042	18	17	17
宮城県仙台市青葉区	50,031	66,469	65,160	19	19	22
大阪府吹田市	48,738	64,741	65,169	20	22	21
⋮	⋮	⋮	⋮	⋮	⋮	⋮

表7-b スピアマンの順位相関係数 (前住地・他市区町村)

比較対象	スピアマンの 順位相関係数 (前住地・他市区町村)
原数値×あん分等	0.99653
原数値×CANCEIS	0.99658
あん分等×CANCEIS	0.99985

表8-a 補充方法・前住地別人口ランキング (他都道府県)

前住地 (他都道府県)	人数 (人)			順位		
	原数値	あん分等	CANCEIS	原数値	あん分等	CANCEIS
東京都	1,695,712	2,336,280	2,360,727	1	1	1
神奈川県	1,000,250	1,310,288	1,334,539	2	2	2
大阪府	871,498	1,166,126	1,191,040	3	3	3
愛知県	777,056	967,808	985,295	4	4	4
埼玉県	712,578	932,728	937,669	5	5	5
福岡県	616,905	765,800	787,148	6	7	7
千葉県	606,885	798,933	805,585	7	6	6
北海道	601,360	753,256	748,560	8	8	8
兵庫県	503,550	649,398	668,249	9	9	9
静岡県	341,751	408,362	418,687	10	10	10
広島県	287,758	351,436	363,978	11	12	12
宮城県	272,711	345,087	341,855	12	13	13
京都府	266,857	356,946	365,585	13	11	11
茨城県	250,291	323,784	324,008	14	14	14
熊本県	201,227	237,545	244,953	15	15	15
長野県	184,367	220,323	226,223	16	17	17
新潟県	180,707	221,875	228,606	17	16	16
岡山県	172,288	211,304	220,127	18	18	18
岐阜県	167,166	203,297	209,453	19	19	19
鹿児島県	161,028	189,151	195,473	20	24	22
⋮	⋮	⋮	⋮	⋮	⋮	⋮

表8-b スピアマンの順位相関係数 (前住地・他都道府県)

比較対象	スピアマンの 順位相関係数 (前住地・他都道府県)
原数値×あん分等	0.99792
原数値×CANCEIS	0.99884
あん分等×CANCEIS	0.99954

2.1.7 労働力状態 (就業状態等基本集計)

表9は、補充方法・労働力状態別人口及び構成比であるが、人数については、数値の大きい順に「CANCEIS」>「あん分等」>「原数値」となっている。また、「原数値」と比較して、「あん分等」と「CANCEIS」が類似した数値となっている。「あん分等」より、「CANCEIS」の人数が多くなっている要因は、「CANCEIS」では、年齢補充も実行されており、それにより15歳以上人口

が「あん分等」よりも増加、それを含めて労働力状態を補完したためである。

また、補完方法・労働力状態別人口の構成比であるが、「原数値」と比較して「あん分等」と「CANCEIS」が類似した構成比となっている。構成比の差は、「CANCEIS－あん分等」において、「主に仕事」が最も大きく、「CANCEIS」では、多くの不詳が「主に仕事」に補完された。

さらに、全国結果を算出したことで、我が国における就業及び不就業の状態を明らかにすることを目的とした労働力調査¹⁵とも比較ができるようになったが、国勢調査と調査方法が異なるため単純な比較ができないことから参考としている。例えば、労働力調査は、サンプル調査であり、調査票の設計も異なることや、不詳は不詳として推計している。「失業者」を見ると、「CANCEIS」の数値が労働力調査よりも大きくなっているが、それは労働力調査では、不詳に「失業者」がいること、そして調査票の設計において、設問「月末1週間に仕事をしたかどうかの別」における選択肢「仕事を探していた」に続きの設問があること等が影響していると考えられる。

表9 補完方法・労働力状態別人口及び構成比

労働力状態	人数(万人)				構成比(%)				構成比の差(%ポイント)		
	原数値	あん分等	CANCEIS	【参考】 労働力調査	原数値	あん分等	CANCEIS	【参考】 労働力調査	CANCEIS －原数値	CANCEIS －あん分等	【参考】 CANCEIS －労働力調査
労働力人口											
主に仕事	4,795	5,468	5,640	5,515	49.7	50.5	50.8	49.7	1.1	0.3	1.1
家事のほか仕事	711	780	789	834	7.4	7.2	7.1	7.5	▲0.3	▲0.1	▲0.4
通学のかたわら仕事	111	130	140	175	1.1	1.2	1.3	1.6	0.2	0.1	▲0.3
休業者	148	169	170	199	1.5	1.6	1.5	1.8	0.0	▲0.1	▲0.3
失業者	231	265	278	211	2.4	2.5	2.5	1.9	0.1	0.0	0.6
非労働人口											
家事	1,314	1,440	1,470	1,322	13.6	13.3	13.2	11.9	▲0.4	▲0.1	1.3
通学	528	597	612	606	5.5	5.5	5.5	5.5	0.0	0.0	0.0
その他	1,818	1,976	2,013	2,244	18.8	18.3	18.1	20.2	▲0.7	▲0.2	▲2.1
不詳 [※]	1,170	0	0	7	-	-	-	-	-	-	-

※労働力調査は就業状態不詳

2.1.8 従業上の地位

表10は、補完方法・従業上の地位別人口及び構成比であるが、「原数値」と比較して、「あん分等」と「CANCEIS」が類似した数値になっている。「あん分等」より「CANCEIS」の人数が多くなっている要因は、労働力状態の補完と同様に、年齢補完により15歳以上人口が増加したことによるものである。

構成比については、「あん分等」「CANCEIS」ともに「原数値」と類似した結果になった。構成比の差を見ると、大きな差ではないものの、「CANCEIS」は、「原数値」や「あん分等」と比較して、「正規の職員・従業員」などが増加、「パート・アルバイト・その他」などが減少した。

表10 補完方法・従業上の地位別人口及び構成比

従業上の地位	人数(万人)				構成比(%)				構成比の差(%ポイント)		
	原数値	あん分等	CANCEIS	【参考】 労働力調査	原数値	あん分等	CANCEIS	【参考】 労働力調査	CANCEIS －原数値	CANCEIS －あん分等	【参考】 CANCEIS －労働力調査
総数	5,764	6,547	6,738	6,722	-	-	-	-	-	-	-
正規の職員・従業員	3,080	3,599	3,748	3,553	54.5	55.0	55.6	53.1	1.1	0.6	2.5
労働者派遣事業所の派遣社員	152	178	187	144	2.7	2.7	2.8	2.2	0.1	0.1	0.6
パート・アルバイト・その他	1,462	1,677	1,708	1,943	25.9	25.6	25.3	29.1	▲0.6	▲0.3	▲3.8
役員	313	362	363	354	5.5	5.5	5.4	5.3	▲0.1	▲0.1	0.1
雇人のある業主	106	120	121	129	1.9	1.8	1.8	1.9	▲0.1	0.0	▲0.1
雇人のない業主	372	427	428	413	6.6	6.5	6.3	6.2	▲0.3	▲0.2	0.1
家族従業者	161	174	174	154	2.8	2.7	2.6	2.3	▲0.2	▲0.1	0.3
家庭内職者	8	9	9	8	0.1	0.1	0.1	0.1	0.0	0.0	0.0
不詳	110	0	0	34	-	-	-	-	-	-	-

¹⁵ 労働力調査では、都道府県別結果をモデル推計値により公表しているものの、項目が限定（労働力人口、就業者、完全失業者、非労働力人口、完全失業率）されていることや、月次の結果公表ではなく、都道府県別四半期平均及び年平均の結果公表であるため、全国結果でなければ国勢調査との比較が困難である。（国勢調査の就業状態は9月の月末一週間を調査しているため労働力調査の9月結果との比較が可能）

2.1.9 産業

表 11 は、補充方法・産業大分類別人口及び構成比であるが、「原数値」と比較して、「あん分等」と「CANCEIS」が類似した数値となっている。「あん分等」よりも「CANCEIS」の人数が多い要因は、他の就業状態等基本集計の調査項目同様に、年齢補充により労働力人口が増加した影響である。構成比については、「原数値」「あん分等」「CANCEIS」ともに類似した結果となった。構成比の差を見ると、「CANCEIS」は、「原数値」や「あん分等」と比較して、「情報通信業」などが増加した。

表 11 補充方法・産業大分類別人口及び構成比

産業大分類	人数(万人)				構成比(%)				構成比の差(%ポイント)		
	原数値	あん分等	CANCEIS	【参考】 労働力調査	原数値	あん分等	CANCEIS	【参考】 労働力調査	CANCEIS -原数値	CANCEIS -あん分等	【参考】 CANCEIS -労働力調査
農業	177	192	195	195	3.2	2.9	2.9	2.9	▲0.3	0.0	0.0
林業	6	6	7	5	0.1	0.1	0.1	0.1	0.0	0.0	0.0
漁業	13	14	14	14	0.2	0.2	0.2	0.2	0.0	0.0	0.0
鉱業、採石業、砂利採取業	2	2	2	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
建設業	418	486	493	506	7.5	7.4	7.3	7.6	▲0.2	▲0.1	▲0.3
製造業	906	1,044	1,073	1,035	16.2	15.9	15.9	15.6	▲0.3	0.0	0.3
電気・ガス・熱供給・水道業	28	32	33	35	0.5	0.5	0.5	0.5	0.0	0.0	0.0
情報通信業	196	252	273	240	3.5	3.8	4.0	3.6	0.5	0.2	0.4
運輸業、郵便業	312	368	377	342	5.6	5.6	5.6	5.2	0.0	0.0	0.4
卸売業、小売業	881	1,032	1,060	1,068	15.8	15.8	15.7	16.1	▲0.1	▲0.1	▲0.4
金融業、保険業	136	162	166	188	2.4	2.5	2.5	2.8	0.1	0.0	▲0.3
不動産業、物品賃貸業	125	152	155	145	2.2	2.3	2.3	2.2	0.1	0.0	0.1
学術研究、専門・技術サービス業	210	256	267	245	3.8	3.9	4.0	3.7	0.2	0.1	0.3
宿泊業、飲食サービス業	310	367	380	404	5.5	5.6	5.6	6.1	0.1	0.0	▲0.5
生活関連サービス業、娯楽業	198	231	240	246	3.5	3.5	3.6	3.7	0.1	0.1	▲0.1
教育、学習支援業	283	331	342	337	5.1	5.1	5.1	5.1	0.0	0.0	0.0
医療、福祉	763	882	910	872	13.7	13.5	13.5	13.2	▲0.2	0.0	0.3
複合サービス事業	44	49	50	51	0.8	0.7	0.7	0.8	▲0.1	0.0	▲0.1
サービス業(他に分類されないもの)	380	454	463	448	6.8	6.9	6.9	6.8	▲0.1	0.0	0.0
公務(他に分類されるものを除く)	203	234	239	246	3.6	3.6	3.5	3.7	▲0.1	▲0.1	▲0.2
分類不能の産業	174	0	0	100	-	-	-	-	-	-	-

2.1.10 職業

表 12 は、補充方法・職業大分類別人口及び構成比であるが、人数については、「原数値」と比較して、「あん分等」と「CANCEIS」が類似した数値となっている。「あん分等」よりも「CANCEIS」の人数が多い要因は、他の就業状態等基本集計の調査項目同様に、年齢補充により15歳以上人口が増加した影響である。構成比については、「原数値」「あん分等」「CANCEIS」ともに類似した結果となった。構成比の差を見ると、「CANCEIS」は、「原数値」や「あん分等」と比較して、「専門的・技術的職業従事者」が増加した。

表 12 補充方法・職業大分類別人口及び構成比

職業大分類	人数(万人)				構成比(%)				構成比の差(%ポイント)		
	原数値	あん分等	CANCEIS	【参考】 労働力調査	原数値	あん分等	CANCEIS	【参考】 労働力調査	CANCEIS -原数値	CANCEIS -あん分等	【参考】 CANCEIS -労働力調査
管理的職業従事者	117	136	135	139	2.1	2.1	2.0	2.1	▲0.1	▲0.1	▲0.1
専門的・技術的職業従事者	1,028	1,225	1,280	1,218	18.3	18.7	19.0	18.4	0.7	0.3	0.6
事務従事者	1,167	1,372	1,413	1,370	20.8	21.0	21.0	20.6	0.2	0.0	0.4
販売従事者	686	811	836	863	12.3	12.4	12.4	13.0	0.1	0.0	▲0.6
サービス業従事者	680	794	818	836	12.1	12.1	12.1	12.6	0.0	0.0	▲0.5
保安職業従事者	110	128	128	135	2.0	2.0	1.9	2.0	▲0.1	▲0.1	▲0.1
農林漁業従事者	190	206	209	212	3.4	3.1	3.1	3.2	▲0.3	0.0	▲0.1
生産工程従事者	764	874	900	862	13.6	13.4	13.4	13.0	▲0.2	0.0	0.4
輸送・機械運転従事者	195	226	229	213	3.5	3.5	3.4	3.2	▲0.1	▲0.1	0.2
建設・採掘従事者	246	286	289	302	4.4	4.4	4.3	4.6	▲0.1	▲0.1	▲0.3
運搬・清掃・包装等	419	488	500	487	7.5	7.5	7.4	7.3	▲0.1	▲0.1	0.1
分類不能の職業	163	0	0	87	-	-	-	-	-	-	-

2.1.11 従業地・通学地(従通地)別人口(従業地・通学地集計)

表 13 は、補充方法・従通地別人口であり、「原数値」と比較して、「あん分等」と「CANCEIS」が類似した数値となっている。図 5 は、補充方法・従通地別人口の構成比であり、「原数値」「あん分等」「CANCEIS」ともに類似した結果となっている。「CANCEIS」では、「原数値」及び「あん分等」と比較して、従通地が「現住所」及び「自市区町村」の構成比が減少、「他市区町村」の構成比が増加する結果となった。すなわち補充結果からは、従通地が「現住所」及び「自市区町村」である不詳は少なく、従通地が「他市区町村」である不詳は多いということが読み取れる。

従通地が他市区町村及び他都道府県である人口のランキングにおけるスピアマンの順位相関係

数を見てみると、ともに1に近い数値となっており、順位相関という観点から、「原数値」「あん分等」「CANCEIS」は極めて類似した結果になっていると言える。(表14-b、表15-b)

表13 補完方法・従通地別人口

従通地	原数値	あん分等	CANCEIS
総数	126,146,099	126,146,099	126,146,099
現住所	5,766,887	6,577,533	6,680,922
自市区町村	35,049,555	40,545,630	41,378,266
他市区町村	28,072,270	33,051,253	34,228,094
従通地区分不詳	15,831,580	0	1,705

図5 補完方法・従通地別構成比

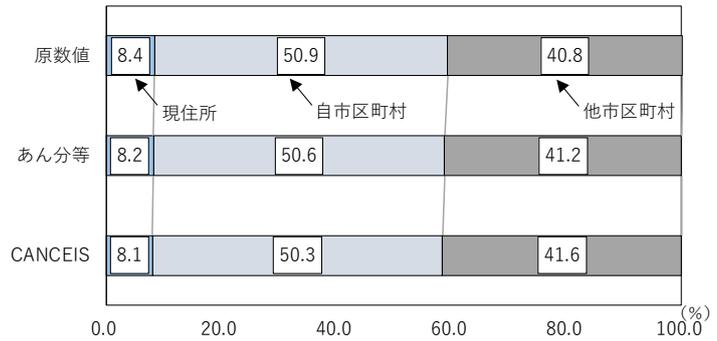


表14-a 補完方法・従通地別人口ランキング

従通地 (他市区町村)	人数(人)			順位		
	原数値	あん分等	CANCEIS	原数値	あん分等	CANCEIS
東京都千代田区	850,924	1,123,372	1,182,036	1	1	1
東京都港区	759,827	1,000,940	1,063,820	2	2	2
東京都新宿区	517,476	683,015	722,489	3	3	3
東京都中央区	513,030	667,851	707,294	4	4	4
大阪府大阪市中央区	367,067	466,325	491,208	5	6	6
東京都渋谷区	363,883	482,973	517,902	6	5	5
大阪府大阪市北区	311,551	395,024	417,400	7	7	7
東京都品川区	286,241	373,314	393,572	8	8	8
東京都江東区	260,182	340,828	359,311	9	9	9
愛知県名古屋市中区	219,646	259,168	267,606	10	10	11
東京都豊島区	195,287	255,347	267,660	11	11	10
福岡県福岡市博多区	192,221	232,122	239,680	12	14	13
東京都文京区	188,585	251,134	259,553	13	12	12
北海道札幌市中央区	180,582	215,963	224,980	14	16	16
東京都世田谷区	178,827	232,548	239,224	15	13	14
東京都大田区	172,370	222,731	233,033	16	15	15
兵庫県神戸市中央区	162,091	197,748	206,029	17	17	18
東京都台東区	151,748	196,156	207,486	18	18	17
福岡県福岡市中央区	139,862	169,411	174,963	19	19	19
宮城県仙台市青葉区	139,072	160,287	164,048	20	21	21
⋮	⋮	⋮	⋮	⋮	⋮	⋮

表14-b スピアマンの順位相関係数
(従通地・他市区町村)

比較対象	スピアマンの 順位相関係数 (従通地・他市区町村)
原数値×あん分等	0.99648
原数値×CANCEIS	0.99942
あん分等×CANCEIS	0.99937

表15-a 補完方法・従通地別人口ランキング

従通地 (他都道府県)	人数(人)			順位		
	原数値	あん分等	CANCEIS	原数値	あん分等	CANCEIS
東京都	6,192,357	8,084,897	8,504,302	1	1	1
大阪府	2,636,914	3,324,594	3,485,281	2	2	2
愛知県	2,016,097	2,343,351	2,405,954	3	4	4
神奈川県	1,950,821	2,386,386	2,480,884	4	3	3
埼玉県	1,410,966	1,682,898	1,736,329	5	5	5
福岡県	1,179,859	1,391,059	1,433,615	6	6	6
千葉県	1,135,720	1,358,067	1,404,599	7	7	7
兵庫県	1,027,194	1,225,934	1,268,769	8	8	8
北海道	762,412	890,743	918,547	9	9	9
静岡県	732,031	787,323	799,509	10	11	11
京都府	641,789	797,686	830,583	11	10	10
茨城県	567,752	640,674	656,807	12	12	12
宮城県	523,620	592,456	606,480	13	13	13
広島県	503,616	570,683	581,264	14	14	14
群馬県	383,625	425,454	432,414	15	15	15
岐阜県	368,554	407,955	416,317	16	16	16
新潟県	346,273	373,896	378,477	17	19	19
熊本県	337,991	375,094	382,046	18	17	18
栃木県	334,449	375,079	382,672	19	18	17
長野県	322,744	348,059	353,132	20	21	21
⋮	⋮	⋮	⋮	⋮	⋮	⋮

表15-b スピアマンの順位相関係数
(従通地・他都道府県)

比較対象	スピアマンの 順位相関係数 (従通地・他都道府県)
原数値×あん分等	0.99838
原数値×CANCEIS	0.99792
あん分等×CANCEIS	0.99965

2.2 各集計事項における CANCEIS 補完値の検証

本項では、2020年国勢調査結果を使って、調査項目ごとに集計事項¹⁶*m*の年齢階級別不詳率を推定し、CANCEIS及びあん分等による集計事項*m*の年齢階級別補完数との比較を試みた¹⁷。

¹⁶ 本稿では、調査項目に対する回答選択肢を集計事項としており、例えば、調査項目が「配偶関係」の場合、集計事項は、「未婚」「有配偶」「死別」「離別」「配偶関係不詳」である(未回答の場合は、「不詳」となる。)

¹⁷ あん分等補完値については、補完方法の限界から補完値のない集計事項が存在しているため、一部の調査項目に限定して比較を行っている。

集計事項 m の年齢階級別不詳率 r_n^m については、集計事項別の不詳を調査結果から把握することができないため、当該年齢階級別集計事項 m の全集計事項（「不詳」を除く）に対する比率を当該調査項目 i の不詳数に乗じることにより求めた（式(1)(2)）。CANCEIS 又はあん分等による年齢階級・集計事項別補完数については、式(3)(4)により算出している。

$$P(A_n^m|F^i) = P(F^i) \times \frac{P(A_n^m|\neg F^i)}{\sum_m P(A_n^m|\neg F^i)} \quad \dots (1)$$

$$r_n^m(\%) = \frac{P(A_n^m|F^i)}{\sum_n P(A_n^m|F^i)} \times 100 \quad \dots (2)$$

$$CI_n^m = C_n^m - O_n^m \quad \dots (3)$$

$$DI_n^m = D_n^m - O_n^m \quad \dots (4)$$

A_n^m : 調査項目 i のうちの集計事項 m (不詳の集計事項を除く) における年齢階級が n である事象 F^i : 調査項目 i が不詳である事象 $P(F^i)$: 調査項目 i が不詳である事象の人口 $P(A_n^m F^i)$: 調査項目 i が不詳である事象のうちの集計事項が m で年齢階級が n の人口 $P(A_n^m \neg F^i)$: 調査項目 i が不詳でない事象のうちの集計事項が m で年齢階級が n の人口 r_n^m : 集計事項 m の年齢階級別不詳率
$CI_n^m (DI_n^m)$: 集計事項が m で年齢階級が n である CANCEIS (あん分等)による補完数 $C_n^m (D_n^m)$: 集計事項が m で年齢階級が n である CANCEIS (あん分等)による不詳補完数 O_n^m : 集計事項が m で年齢階級が n である原数値

また、都道府県別に CANCEIS 補完値と原数値において、年齢階級別に各集計事項の構成比の差を式(7) (10)のとおり算出し、どの集計事項に重点的に CANCEIS による補完がなされたかを確認した。都道府県別に見ると、動きの傾向にほとんど差異は見られなかったため、本稿では、構成比の差が大きかった都道府県を5つ選定し、他統計とも比較をすることで、CANCEIS による補完値の妥当性を検証した。ここでいう構成比とは、調査項目である「配偶関係」を例に出すと、「不詳」を除いた集計事項は、「未婚」「有配偶」「離別」「死別」であり、年齢階級別にこれらの集計事項の総数を分母に、各集計事項を分子にしたものである。この構成比を年齢階級別に CANCEIS 補完値と原数値でそれぞれ算出し、その差を取った。結果を見る観点としては、ある集計事項・年齢階級における構成比の差（CANCEIS－原数値）がプラスになっている場合、その区分において、主に精緻なドナー選択（不整合・外れ値の条件を厳密に設定したこと）などの結果により、CANCEIS による補完が重点的になされたことを意味しており、差の大きさは補完の程度（強さ）を示している。なお、当該集計事項において、関連する他の公的統計の結果が存在する場合は、その結果も参考にして分析している。

○各年齢階級×集計事項に占める人口の構成比の差（図7、図10、図11、図13）

$$CR_{nm}(\%) = \frac{P(A_n^m|\neg F^i)}{\sum_{n,m} P(A_n^m|\neg F^i)} \times 100 \quad \dots (5)$$

$$OR_{nm}(\%) = \frac{P(A_n^m|F^i)}{\sum_{n,m} P(A_n^m|F^i)} \times 100 \quad \dots (6)$$

$$D_{nm}(\%ポイント) = CR_{nm} - OR_{nm} \quad \dots (7)$$

CR_{nm} : CANCEIS 補完値における年齢階級 n ×集計事項 m に占める人口の構成比 OR_{nm} : 原数値における年齢階級 n ×集計事項 m に占める人口の構成比 D_{nm} : CANCEIS 補完値と原数値の年齢階級 n ×集計事項 m に占める人口の構成比の差
--

○各年齢階級における集計事項に占める人口の構成比の差（図8）

$$CR_m(\%) = \frac{P(A_n^m|\neg F^i)}{\sum_m P(A_n^m|\neg F^i)} \times 100 \quad \dots (8)$$

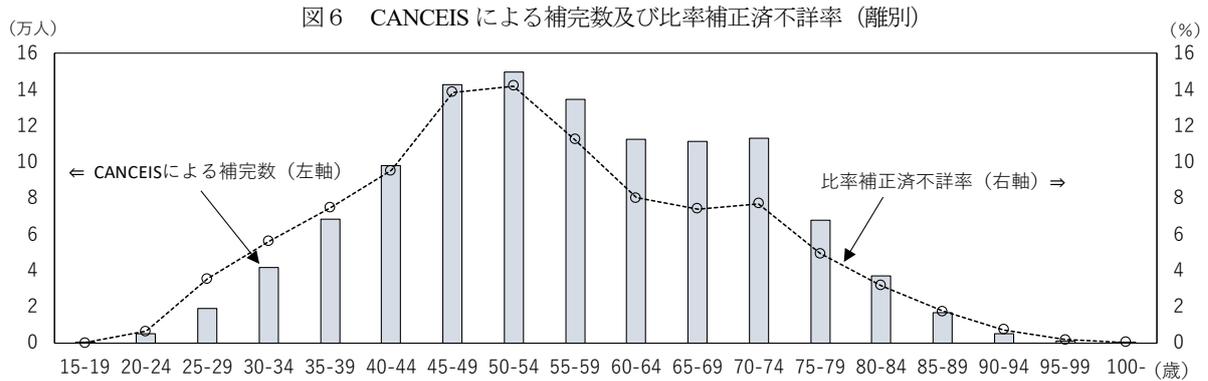
$$OR_m(\%) = \frac{P(A_n^m|F^i)}{\sum_m P(A_n^m|F^i)} \times 100 \quad \dots (9)$$

$$D_m(\%ポイント) = CR_m - OR_m \quad \dots (10)$$

CR_m : CANCEIS 補完値における各年齢階級の集計事項に占める人口の構成比 OR_m : 原数値における各年齢階級の集計事項に占める人口の構成比 D_m : CANCEIS 補完値と原数値の各年齢階級の集計事項に占める人口の構成比の差
--

2.2.1 配偶関係（人口等基本集計）

集計区分「人口等基本集計」においては、一般世帯の配偶関係における集計事項「離別」を選定し検証を行った。全国結果における CANCEIS による「離別」の補完数と比率補正済の「離別」不詳率の比較は図6のとおりである。比率補正済「離別」不詳率の動きは CANCEIS による補完数とほぼ同じ動きであるが、50～79歳の間では CANCEIS による補完数の動きが若干強くでている。当該年齢層は、近年増加していると言われている、いわゆる熟年離婚が起きる年代であり、「離別」の件数が多くなる年代であると推測できる。

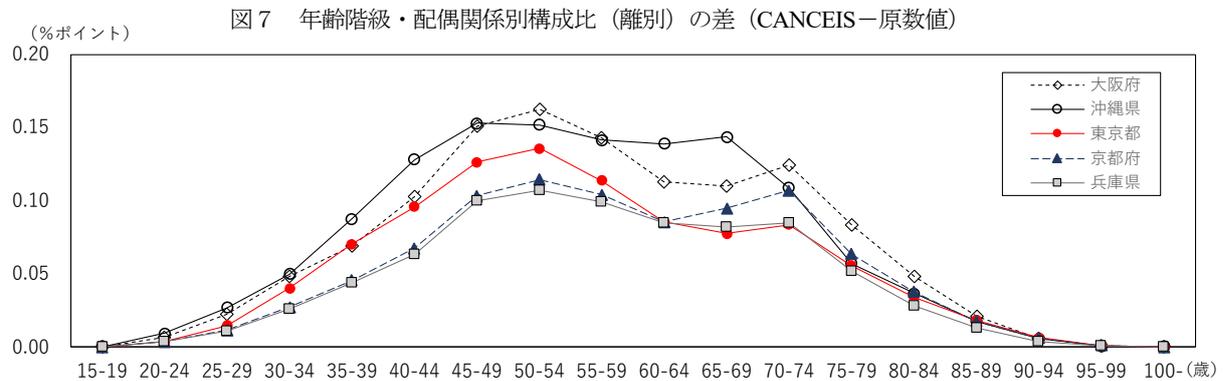


全都道府県において、年齢階級別配偶関係（離別）の構成比の差を見ると、「原数値」よりも「CANCEIS」の方が、構成比が高くなることわかる（図7）。構成比の差は、最大でも0.16%ポイントであり、あまり大きくはないが、これにより「離別」に関して言えば、CANCEIS による距離関数の判別結果として、「離別」となるデータが他の集計事項と比較して若干多く配分されたことがわかる。また、「大阪府」「沖縄県」は特に「離別」となる傾向が強いが、令和2年人口動態調査の都道府県別離婚率に関する結果（表16）を見ると、離婚率1位は「沖縄県」、5位は「大阪府」となっており、人口動態調査の結果においても、上位にランキングしている。都道府県別離婚率においては、各都道府県で大きな差が見られず、また、補完結果の動きも小さいため、CANCEIS 補完値と人口動態調査結果の間に明らかな類似性が見られなかったと考えられる。

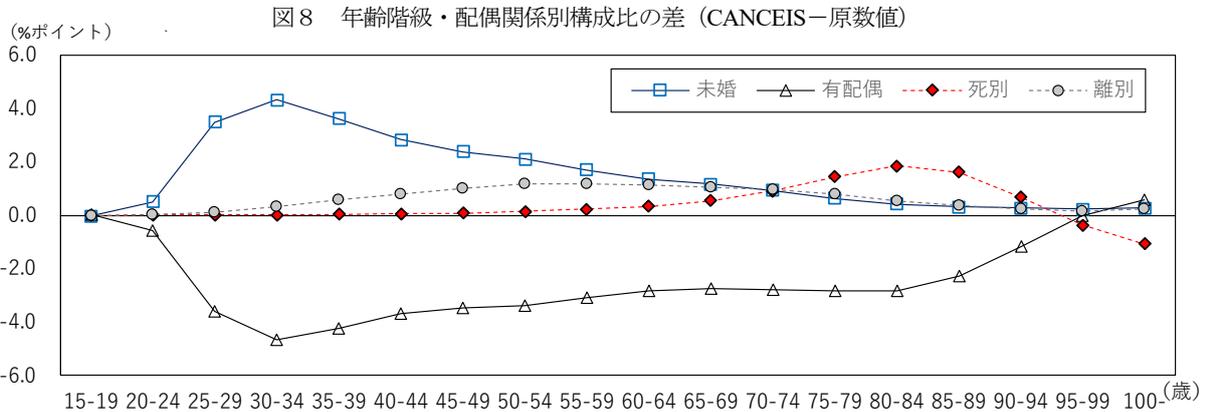
表16 都道府県別離婚率ランキング

順位	都道府県	離婚率 (%) (人口千対)
1	沖縄県	2.36
2	宮崎県	1.79
3	福岡県	1.77
4	北海道	1.75
5	大阪府	1.73
6	大分県	1.70
7	和歌山県	1.67
7	高知県	1.67
9	福島県	1.63
9	山梨県	1.63

引用元：令和2年人口動態調査（厚生労働省）

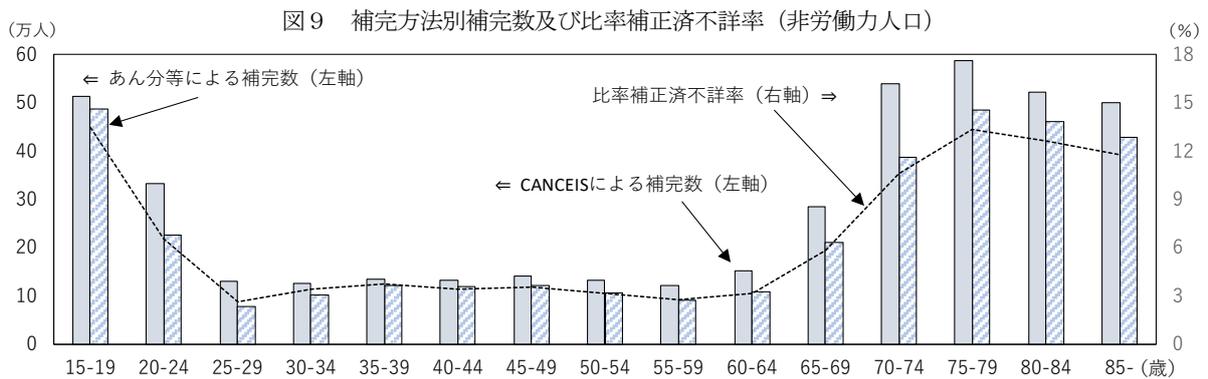


一方、図8の年齢階級・配偶関係別構成比の差を見ると、配偶関係「不詳」の多くが、若年層では「未婚」に補完されており、高齢層では「死別」に補完されていることがわかる。また、全年代（100歳以上除く）を通して、「有配偶」への補完は少ないことがわかる。この原因として考えられるのが、2020年国勢調査結果より、①配偶関係「不詳」は単独世帯の世帯員が多くを占めている（約79%）、②単独世帯の約53%が「未婚」であるということである。こういった国勢調査データの特性から、配偶関係「不詳」の多くが「未婚」に補完されたと推測される。



2.2.2 非労働力人口（就業状態等基本集計）

集計区分「就業状態等基本集計」においては、集計事項「非労働力人口」を選定し検証を行った。全国結果におけるCANCEIS及びあん分等による「非労働力人口」の補完数と比率補正済「非労働力人口」不詳率の比較は図9のとおりである。これを見ると、あん分等による補完数の動きが比率補正済不詳率の動きに近いことがわかる。あん分等補完値は、「不詳」を除いた集計事項の比率を用いて作成しているため、このような類似した結果となる。一方のCANCEISによる補完数であるが、比率補正済不詳率の動きと傾向としては同じであるが、若年層や高齢層において、若干強くでていることがわかる。この層は非労働力人口の多い層であるため、補完数も多くなると推測できる。また、補完数が「あん分等」よりも「CANCEIS」に多い要因は、年齢不詳が「CANCEIS」では補完されているためであり、その寄与の分だけ補完数が多い。



全都道府県において、年齢階級別労働力状態の構成比（非労働力人口）の差を見たところ「15～19歳」及び「55～59歳」以上において、「CANCEIS」の方が「原数値」よりも構成比が低くなる傾向が見られた（図10）。これは、CANCEISによる距離関数の判別結果として、「15～19歳」及び「55～59歳」以上では、「非労働力人口」になり得る不詳が少なかったということの意味する。これは、図10が年齢階級別労働力状態における「非労働力人口」の構成比の差なので、①当

該年齢階級以外の不詳が多い可能性がある、②当該年齢階級以外の不詳の多くが「非労働力人口」ではなく、「労働力人口」に補完されている可能性があることが考えられるため、「相対的に見て、非労働力人口になり得る不詳が少なかった」ということである。反対に、「20～24歳」及び「25～29歳」では、「CANCEIS」の方が、構成比が高くなっており、これは「20～24歳」及び「25～29歳」では、「非労働力人口」になり得る不詳が多かったことを意味する。

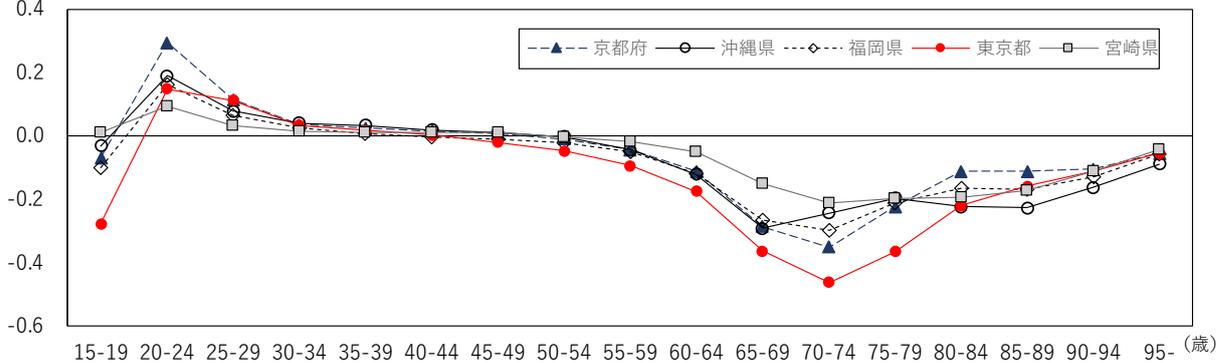
令和2年度学校基本調査より、都道府県別人口に対する学生（大学・大学院生）比率を計算すると、全都道府県のなかで、「京都府」の割合が最も高く、次に「東京都」の割合が高かった（表17）。こういったことから、図10において「京都府」や「東京都」の「20～24歳」の構成比の差が大きくなっていると考えられる。一方で、他の道府県においては、「福岡県」が上位にランキングされているものの、明らかな類似性は見られなかった。

表17 都道府県別学生数比率ランキング

順位	都道府県	学生数比率(%)
1	京都府	6.31
2	東京都	5.40
3	石川県	2.83
4	大阪府	2.79
5	愛知県	2.57
6	宮城県	2.50
7	滋賀県	2.44
8	福岡県	2.40
9	兵庫県	2.32
10	岡山県	2.31

引用元：令和2年度学校基本調査：
都道府県別学生数を用いて独自に算出（分母に用いた都道府県別人口は令和2年国勢調査の値）

図10 都道府県・年齢階級別労働力状態における構成比（非労働力人口）の差（CANCEIS－原数値）（%ポイント）



2.2.3 昼間人口（従業地・通学地集計）¹⁸

集計区分「従業地・通学地集計」においては、集計事項「昼間人口」を選定し検証を行った。全都道府県において、年齢階級別昼間人口の構成比の差を見たところ、「15歳未満」「55～59歳」以上では、「CANCEIS」の方が、「原数値」より構成比が低くなっており、非労働力人口の構成比が高い年齢階級では、CANCEISによる補完が少なくなっていることがわかる。反対に、「20～24歳」から「45～49歳」では、「CANCEIS」の方が「原数値」より構成比が高い。特に「東京都」の「20～24歳」から「40～44歳」において、その傾向が強く出ていることがわかる（図11）。令和3年経済センサス-活動調査の結果からもわかるように「東京都」は事業所数が最も多い都道府県である（表18）。働き盛り世代における、事業所数の多い都道府県の「昼間人口」への補完が目撃の結果となっている（都道府県別事業所数ランキングにおいて「京都府」は12位）。

表18 都道府県別事業所数ランキング

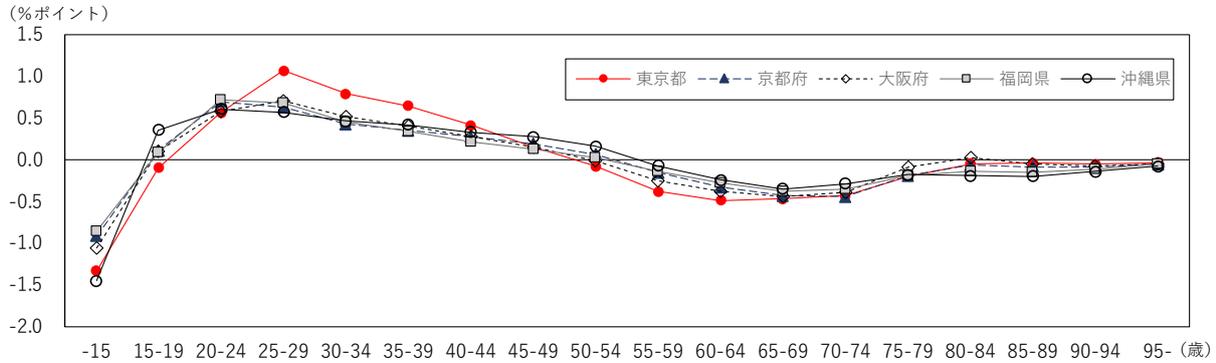
順位	都道府県	事業所数 [※]	従業者数（男女計）
1	東京都	810,248	10,093,781
2	大阪府	474,300	4,764,797
3	神奈川県	343,684	3,769,519
4	愛知県	343,572	4,056,894
5	埼玉県	266,768	2,789,750
6	北海道	247,542	2,408,665
7	福岡県	244,292	2,462,637
8	兵庫県	237,747	2,397,358
9	千葉県	214,005	2,332,923
10	静岡県	176,504	1,852,997

※事業内容等不詳を含む。
引用元：令和3年経済センサス-活動調査

¹⁸ 「昼間人口」の場合、地域別に見ることで意味を成すため、全国結果におけるCANCEISによる補完数と比率補正済不詳率の比較は行っていない。

一方、「沖縄県」は別の要因で構成比の差に特徴が出ていると考えられ、図 11 の「15～19 歳」において、「沖縄県」は他の都道府県と比べて構成比の差に特徴が出ているが、これは、従通地「不詳」のほとんどが自県に補完されたことによるものである。「沖縄県」は島であるため、他県への移動が容易ではないことに関係していると推測できる。

図 11 都道府県・年齢階級別昼間人口における構成比の差 (CANCEIS—原数値)



2.2.4 転入 (人口移動集計)

集計区分「人口移動集計」においては、集計事項「転入」を選定し検証を行った。全国結果における CANCEIS 及びあん分等による「転入」の補完数と比率補正済「転入」不詳率の比較は図 12 のとおりである。「非労働力人口」と同様に、あん分等による補完数の動きが比率補正済不詳率の動きに近いことがわかる。一方の CANCEIS による補完数であるが、比率補正済不詳率の動きの傾向としては、ほぼ同じであるが、「20～24 歳」においては、顕著な違いが見られる。

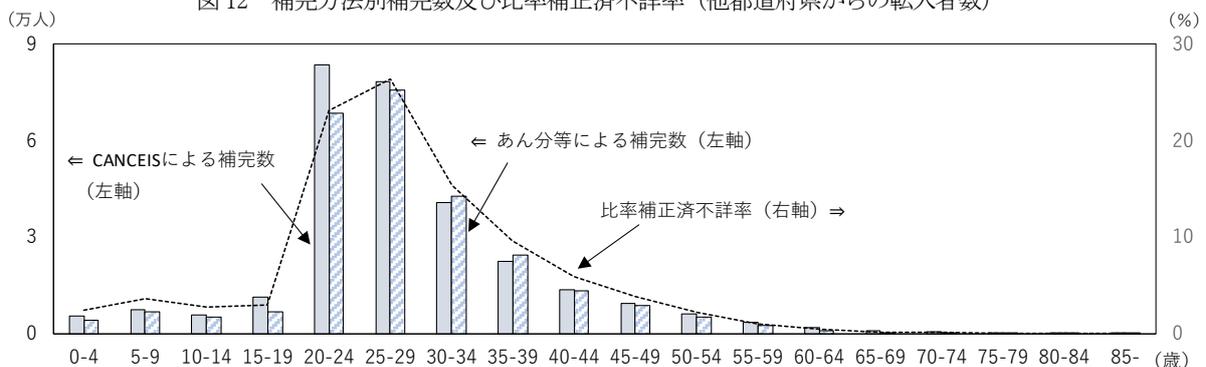
2020 年住民基本台帳人口移動報告 (年報) の他都道府県からの転入者数を見ると、「20～24 歳」は、「25～29 歳」よりも転入者数が多くなっており、これは、CANCEIS による補完数と同じ傾向になっている (表 19)。転入者数の多い年齢階級に、CANCEIS ではより多くの補完がされており、その傾向は住民基本台帳人口移動報告とも一致している。また、図 12 より高齢層では、転入者数がほとんど補完されておらず、実際に転入者数はほとんどいないことも住民基本台帳人口移動報告と一致している。

表 19 年齢階級別転入者数

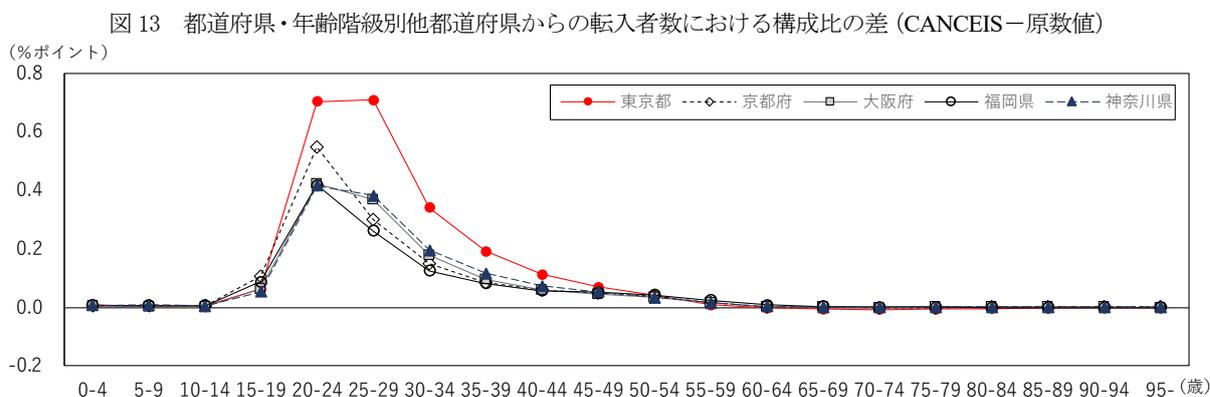
年齢階級	他都道府県からの転入者数	割合 (%)
総数	2,463,992	100.0
0-4	123,096	5.0
5-9	61,733	2.5
10-14	31,590	1.3
15-19	143,449	5.8
20-24	566,969	23.0
25-29	482,199	19.6
30-34	301,685	12.2
35-39	198,506	8.1
40-44	137,764	5.6
45-49	115,961	4.7
50-54	87,787	3.6
55-59	62,073	2.5
60-64	42,380	1.7
65-69	30,180	1.2
70-74	23,970	1.0
75-79	16,987	0.7
80-84	15,287	0.6
85-89	13,650	0.6
90-	8,721	0.4

引用元：2020 年住民基本台帳人口移動報告 (年報)

図 12 補完方法別補完数及び比率補正済不詳率 (他都道府県からの転入者数)



全都道府県において、年齢階級別移動人口（転入）の構成比の差を見ると、「20～24歳」または「25～29歳」で極大になっている（図13）。これは、地域によらず、若年層においては、「転入」に補完される「不詳」が多いことを意味している。差が大きい県は、「東京都」「京都府」「大阪府」「神奈川県」及び「福岡県」となっており、大都市が存在する都府県でその傾向が強くなっている。また、住民基本台帳人口移動報告より、「20～29歳」の転入者数は全体の約43%を占めており、実際に若年層において転入者数が多いことがわかる。

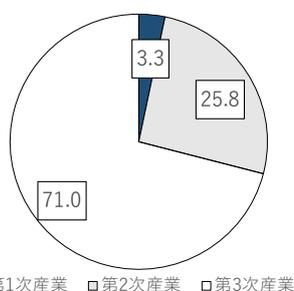


2.2.5 居住期間（人口移動集計）

2020年国勢調査結果より、産業が判明している居住期間「不詳」の内訳を見ると、第3次産業（サービス業等）は71.0%、第2次産業（製造業等）は25.8%、第1次産業（農林漁業等）は3.3%となっており、第3次産業において、居住期間が「不詳」となっている割合が高いことがわかる（図14）。

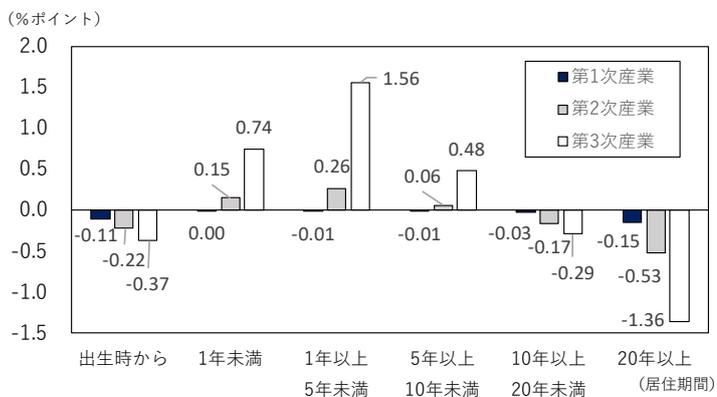
図15は、産業・居住期間別構成比の差であるが、「第3次産業」、「第2次産業」では、居住期間が「1年未満」「1年以上5年未満」及び「5年以上10年未満」において、「CANCEIS」の構成比が原数値よりも高くなっており、CANCEISによる補完が多くなされたことがわかる。一方、それ以外の集計事項では、「原数値」よりも低くなっている。「第1次産業」は、どの居住期間においても、「CANCEIS」の構成比が「原数値」よりも低くなっている。「第3次産業」の居住期間「不詳」の割合が高いことから、「第3次産業」の居住期間「不詳」に補完が多くなされるのは自然なことであり、居住期間が短い集計事項への補完が多いことは、不詳が比較的多い、居住期間の短い単独世帯の若者への補完が影響していると考えられる¹⁹。

図14 産業別居住期間「不詳」の割合（%）



引用元：2020年国勢調査

図15 産業・居住期間別構成比の差（CANCEIS－原数値）



注) 図15の構成比の差は、年齢階級別ではなく、産業×居住期間のクロス集計表における全体の構成比の差となっている。

¹⁹ 埴淵・山内(2019)では、若年層や単独世帯ほど調査票が未提出となる傾向が強いとの見解を示している。

3 おわりに

本稿では、全都道府県における CANCEIS 補完値を試算し、それを合算することで全国の CANCEIS 補完値を算出した。そして、主に全国の CANCEIS 補完値の人口や構成比を、あん分等補完値と比較し、妥当性を検証した。また、各集計事項における CANCEIS による補完数を、推定した不詳率と比較することで妥当性を検証、さらに他の公的統計に国勢調査の集計事項と同類の結果が存在する場合は、CANCEIS 補完値と比較することで妥当性を検証した。

人口については、都道府県別結果だけでなく、全国結果においても、CANCEIS 補完値とあん分等補完値は、年齢補完の影響が見られる調査項目を除いて、どの集計区分でも両者が類似した結果になることを確認した。構成比については、年齢補完の影響の有無に関わらず、両者の構成比は類似していた。他の公的統計との比較においては、都道府県別に大きな特徴が見られるものについては、補完数との間に関連が見られたが、そうではないものについては、特筆すべき関連は見られなかった。

各集計事項における CANCEIS による年齢階級別の補完数については、年齢階級別に推定した各集計事項の不詳率と動きが概ね一致する結果となった。これは、あん分等補完値と CANCEIS 補完値が類似した結果となる誘因の一つでもある。あん分等補完値は、原数値における集計事項の正データ²⁰の比率を用いて不詳をあん分しており、不詳の中の属性も正データと同じ比率を持っていると仮定した補完方法により算出されている。この補完方法の機構から読み取れることは、正データにおける当該属性が多ければ多いほど、その属性における不詳の配分率も高くなるということで、要するに、CANCEIS 補完値もあん分等補完値も、アプローチの仕方は異なるが、不詳が多いと推定される集計事項に多く補完されており、不詳が少ないと推定される集計事項には少なく補完されているのである。これは当然のことのように思うかもしれないが、CANCEIS の場合、莫大なユニット数を持った個票データに対する最近隣法による補完から始めているわけで、そこからの積み上げでマクロレベルの補完値を再現しており、いわば何通りもの組み合わせが考えられる中で、あん分等補完値と結果が類似するという事実が、CANCEIS 補完値の確度の高さを裏付けている。多くの国が信頼できる人口センサスの補完システムとして CANCEIS を採用している理由が、これまでの検証でより理解できた。

「あん分等による補完数」と「比率補正済不詳率」が算出できる集計事項のグラフ（図9及び図12）について見ると、「あん分等による補完数」と「比率補正済不詳率」が同じ動きであるのに対し、「CANCEIS による補完数」では一部の区分において動きが強く出ているのを確認できる。例えば、「転入」のグラフである図12の場合、20～24歳が25～29歳よりも強く出ている。この動きは転入状況の行政記録情報とも一致している。これは、不詳になる属性が、正データの比率に依存するとは限らないためであり、一概に正データの属性比率が不詳データにおける属性比率と一致するとは言えないことを如実に表した結果となっている。国勢調査のデータの場合、単独世帯や外国人に不詳が多いなど、ランダムに不詳があるわけではなく、偏りが存在しているため、この結果は国勢調査データが MNAR (Missing Not at Random) であるということをサポートしている。また、CANCEIS 補完におけるドナー候補の選定について、ドナーとレシピエントの距離は最短だったとしても、補完した結果、不整合データや外れ値が生成されてしまう場合は、当該補完にそのドナーが使われないように、DLT (Decision Logic Table) による制御を実施しているため、CANCEIS 補完が単純な最近傍マッチング代入法ではないこともこの結果が証明している。

集計区分を横断して CANCEIS 補完値を見てみると、どういう属性の人物が補完されたのかを推測することができる。例えば、表10の補完方法・従業上の地位別人口の構成比の差（CANCEIS－原数値）を見ると、「正規の職員・従業員」が最も多く増加し、同様に表11の補完方法・産業

²⁰ 「不詳」を除いたデータ

大分類別人口の構成比の差（CANCEIS—原数値）では、「情報通信業」、表 12 の補完方法・職業大分類別人口の構成比の差（CANCEIS—原数値）では、「専門的・技術的職業従事者」が最も多く増加している。「情報通信業に従事する正規の職員・従業員である専門的・技術的職業従事者」が CANCEIS による補完処理で最も多く増加したという説明ができる。「情報通信業に従事する正規の職員・従業員である専門的・技術的職業従事者」といえばシステムエンジニア（SE）を連想するが、確かに仕事が忙しいイメージがあり、調査に回答するのが困難な職業の一つであると推論できる。

また、小地域集計における CANCEIS 補完値の妥当性の検証については、紙幅の都合上、割愛しているが、年齢の小地域集計において、ある特定の年齢が大幅に増加している調査区が数は少ないもののいくつか存在することがわかった。この調査区に存在している世帯は一般世帯であり、マンションやアパートのある区域で、学生寮等同じ年代の者が集団で生活しているような地域ではなく、現実的に特定の年齢が集中する事象は考えづらい地域であった。そのため、CANCEIS がどういった経緯で補完処理を実行したのか分析をしてみると、それは同一ユニットをドナーとして複数回利用したことによって起きた事象であることが判明した。CANCEIS は、デフォルトでは 100 回まで、同一ユニットをドナーとして利用できる仕様になっている。一方で、不詳のある複数のユニット（レシピエント）が、ある地域に集中しており、その周囲にはドナー候補となるユニットが 1 つしかない場合、その 1 つのドナーが複数回使われることで、それが結果的にバイアスとなるリスクをはらんでいる。そのリスクを回避する手法として、同一ドナーの最大利用回数を制限する手法が挙げられる。これは、カナダ統計局でも実行されている手法であり、実際にドナーの最大利用回数を制限することで、集中して補完されていた年齢の増加がなくなり、一般的に考えて人数が増加するであろう年齢（不詳が多いと推定される年齢層）へと均された補完結果になることがわかった。CANCEIS は、1 つのレシピエントに対して、距離の近い不詳のないユニットをデフォルトでは 10 個、ドナー候補としてストックすることができ、このストックされたドナー候補のプールからランダムに 1 つのユニットを選択し、レシピエントへ補完を実行する。そのため、大抵の場合は、最大利用回数を設定したことで利用できなくなったドナー候補以外にも、ドナー候補は存在しているわけで、同一ドナーの最大利用回数を制限したとしても、距離が同質の他のドナー候補が補完に利用されるため、マクロ的に見れば、集計事項の数値への影響は極小となる。一つのドナーが何度も使われることにより、小地域単位で起きるバイアスの可能性を持った事象については、同一ドナーの最大利用回数を制限することで、ある程度解消することができ、小地域集計の妥当性を高めることができることがわかった。

まとめとして、これまでの CANCEIS 補完における研究で知り得た、その性能を遺憾なく発揮するための重要なポイントは次のとおりである。

- | |
|---|
| <ul style="list-style-type: none"> ① ドナーが不足していると考えられる集計事項には、行政記録情報等を用いて、CANCEIS 実行前に部分的な事前補完を実施 ② 類似したユニットを近傍に集めるためのインプットデータのグルーピングやソート処理
(ex. 世帯人員、世帯の種類) ③ インプットデータにおける続柄等のユニット（世帯）内のソート処理 ④ CANCEIS 補完による不整合データや外れ値の生成を未然に防ぐための適切な DLT の作成 ⑤ 同一ドナーにおけるドナーとしての最大利用回数の適切な設定 |
|---|

一方で、CANCEIS による不詳補完では、ドナーの見つからないレシピエントがある場合、不詳を補完することができないという課題が残っており、例えば、国勢調査において少数派である外国人世帯については、それが多人数世帯ともなると、ドナーを見つけ出すことが困難となり、補完できずに不詳として残ってしまうケースがある。こういったケースはユニット数としては極め

て少ないものの、補完結果から不詳を完全に無くしたい場合は、CANCEISとは別の補完方法を講じて、不詳を補完しなければならないだろう。一つの案として、CANCEIS補完後に別の補完方法によって、残りの不詳を補完することが考えられるが、それは今後の研究課題としたい。

謝辞等

本研究の遂行に当たり、総務省統計局統計調査部国勢統計課の中村英昭氏及び宮下佳孝氏の両氏からは大変有益なコメントをいただいた。また、外部有識者の菅幹雄教授（法政大学）、加藤久和教授（明治大学）及び西郷浩教授（早稲田大学）の皆様方からも大変貴重なご意見をいただいた。そして、独立行政法人統計センター統計編成部人口統計編成課の野崎政志氏及び佐々木大軌氏の両氏からは、CANCEISを実行する上での困難な演算について、技術的にもご協力いただいた。

さらに、本稿の執筆に当たっては、2名の匿名査読者から大変有益なコメントをいただき、内容を改善することができた。

この場を借りて、ご協力いただいた皆様方へ感謝の意を表したい。

参考文献

- [1] 阿部隆 (2013) 「国勢調査結果の「不詳数」に関わる諸問題」, 『統計』 2013年12月号, 2013年12月
- [2] 宇南山卓 (2013) 「仕事と結婚の両立可能性と保育所: 2010年国勢調査による検証」, RIETI Discussion Paper Series 13-J-039, <https://www.rieti.go.jp/publications/dp/13j039.pdf>, 2013年5月
- [3] 北原昌嗣・寺垣内雅子 (2023) 「諸外国の国勢調査におけるインピュテーション方法」, 『統計研究彙報』第80号, 2023年3月, pp137-162
- [4] 北原昌嗣 (2025) 「国勢調査の補完方法に関する一考察」, 『統計研究彙報』第82号, 2025年3月, pp41-62
- [5] 清水邦夫 (2020) 「相関係数」, 統計スポットライト・シリーズ4, 2020年6月
- [6] 総務省統計局 (2017) 「平成27年国勢調査事後調査確報結果について」, 令和2年国勢調査有識者会議(第3回)資料4, <https://www.stat.go.jp/info/kenkyu/kokusei/yusiki32/pdf/03sy0401.pdf>, 2017年12月
- [7] 総務省統計局 (2021) 「不詳補完結果(参考表)に関する解説及び参考資料(令和2年及び平成27年国勢調査に関する不詳補完結果(参考表)について)」, <https://www.stat.go.jp/data/kokusei/2020/kekka.html>, 2021年11月
- [8] 内閣府経済社会総合研究所(ESRI)景気統計部 (2017) 「欠測値補完に関する調査研究報告書」, 2017年3月
- [9] 埴淵知哉・山内昌和 (2019) 「国勢調査「不詳」発生に関連要因—インターネット調査を用いた未提出者の分析—」, 『E-journal GEO Vol.14(1)』, 2019年1月, pp14-29
- [10] Office for National Statistics (ONS) (2022) “Item editing and imputation process for Census 2021, England and Wales”
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/itemeditingandimputationprocessforcensus2021englandandwales>
- [11] Statistics Canada (2020), “CANCEIS USER’S GUIDE Version 5.4”
- [12] United Nations Statistics Division (2019) “Handbook on Population and Housing Census Editing Revision 2”
https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev2e.pdf