

国勢調査のマイクロデータから新開発した家族構成変数と、 それを利用した世帯構造の分析事例の紹介

周防 節雄 (公財)統計情報研究開発センター

1. はじめに

ここ 10 数年、国勢調査(以後「国調」)の調査票情報や匿名データ、全国消費実態調査(以後「全消」)の匿名データなどの公的マイクロデータを使って、世帯を対象にしたデータ分析も手がけてきた。8050 問題、少子高齢化、人口減少、ひとり親世帯、貧困世帯、経済格差などの諸問題の対策には、世帯構造を念頭に置いた分析が不可欠となる。国調など、世帯を調査客体とする公的統計では、通常、各世帯員の続柄、性別、年齢、婚姻状態の4つの個人属性が調査項目にあり、エンドユーザーが利用できる調査票情報や匿名データにも変数として含まれている。その他にも、メーカー側で作成された世帯分類に関するいくつかの変数が提供されている場合が多いが、これらは先の4つの個人属性を組み合わせれば、手間はかかるが、エンドユーザー側でも作成できる。エンドユーザーの立場からは、予め提供されたこれらの世帯分類変数は直ぐに利用できて便利だが、それらの変数のコードが実際に何を意味するかは、符号表の日本語記述に頼る他はない。

国調の調査票情報には、世帯の「家族類型」という変数(26 区分)があり、匿名データでは6区分に纏められている。そのどちらにも「夫婦と子供から成る世帯」がある。初めてその調査票情報を利用した時、「エッ」と思う場面に遭遇した。「夫婦と子供」の世帯と言え、①「世帯主夫婦と子」の世帯がすぐに連想されるが、②「世帯主とその両親」の世帯もこのカテゴリーに含まれていた。これは、国調の「家族類型」の概念が、世帯の中で誰が世帯主であっても、世帯の構造は同一になるように設計されているからである。

確かに上記の①、②は共に「夫婦と子供」から成る世帯には違いないが、この両者の世帯は、消費構造、就労状況など、社会生活上の行動パターンは全く異なっているはずなので、エンドユーザー側では、分析の際にはきちんとこの違いを認識する必要がある。この世帯分類の変数と別の変数を組み合わせれば、この両者は明確に区別できるのだが、分析作業過程では、直感的に世帯構造が正確に把握できる自前の変数を使って分析ができたという思いを長年抱いてきた。

数年前に、シンフォニカでは科研プロジェクト¹で「新世帯類型」と称する新機軸の詳細な世帯分類変数(中川雅義他 2016、伊藤彰彦他 2017)を考案して、国調の調査票情報を使って実際に作成し、本研究会でも研究報告を行った。その開発過程で、長年、筆者の頭の片隅にあった別の構想を具体化させるべく、本稿で解説する新構想の「家族構成変数」の設計を試みた。その作成アルゴリズムの開発をするために、乱数を使って、国調の調査票情報に含まれる世帯員の続柄、性別、年齢、婚姻状態の4つの個人属性を発生させて擬似マイクロデータを作成後、SAS で開発した。今回、そのアルゴリズムを、国調の匿名データに適用して、分析を試みた²。

¹ (独)日本学術振興会平成 26 年度学術研究助成基金研究: 基盤研究(C)「新世帯類型の構築と世帯構造の変動に係る長期時系列分析」(課題番号: 26380280、研究期間: 平成 26 年度~28 年度、研究代表者: 伊藤彰彦)

² 本稿は、2019 年の本研究会で行った研究報告の続編である。ただ、この一年の間に試行錯誤を繰り返し、家族構成変数の表

2. 漢字表現の家族構成変数で表現する世帯構造

2000年の国調の調査票情報からは、表1に示すように、続柄は12区分に再編された。年齢は実年齢。家族構成変数は、表2の続柄漢字パターン一覧表に示すように、世帯員を続柄、性別、婚姻状況、年齢によって、各世帯員の家族関係を33個の漢字を使って一文字で表現し、世帯主を左端に配置し、以下、順次残りの世帯員に対応する漢字を右側に並べて作成した文字変数である。この家族構成変数を見れば、その世帯構造が一目で把握できる。これらの続柄漢字パターンの世帯内階層構造は末尾の附録に示している。

なお、表2には、一部、本来の意味とは異なる漢字も含まれている。例えば、単独世帯では世帯員が婚姻中か否かも区別できるように、非婚の男性の単独世帯は、「単」で表現し、非婚の女性の単独世帯には「嬋」を当てている。女偏に「単」の旧字体なので直感的に理解できる。現在婚姻中の単独世帯の場合は、男なら「宿」(宿六からの連想)、女なら「嬢」を当てた。こうしておけば、世帯主が「宿」なら、既婚男性の一人住まいであることが直ぐに判別でき、その配偶者の妻は、どの世帯かは特定できないが、別の世帯³⁾にいることになる。仮に、この妻の世帯に男児が一人居ても、これは見かけ上は母子家庭だが、家計は父母と男児が同一家族であることになり、いわゆる母子家庭ではないが、その判別には、「母」が婚姻中か否かをチェックすればよい。

「子」や「子の配偶者」と同居していない世帯主については、①夫婦同居なら、「夫」と「妻」で、②夫婦同居でない場合は、「亭」(亭主)と「奥」(奥様)で表現した。

世帯主の子供は、息子か娘か、更に、現在婚姻中か独身かの区別ができるように「倅」(婚姻中)、「倅」(倅の略字体:独身)、「娘」(婚姻中)、「好」(独身:「女の子」と読む)として、区別できるようにした。

実験用マイクロデータ(続柄は12区分)から作成した例を表3に示す。この中に、「女他」があるが、これは「女性(独身)と他人(男性)」から成る世帯を示している。このパターンは、先の新世帯類型による分析結果からも、ほとんどのケースで、両者がほぼ同世代、かつ、両者とも現在独身の、いわゆる同棲世帯であることが判明している。これに対応するために、赤の「他人」の性別が分かるように、「他」(男の他人)、「媒」(女の他人:某女の連想)に2分類した。

3. 家族構成変数の利点

世帯に関して様々な分析処理をする際に、従来のように世帯分類を最初に固定してから始めるのではなく、まず先に家族構成変数を作成しておき、分析過程で、必要に応じて家族構成変数を使って、その都度、分析に適した世帯分類を行えば、より適切な分析が可能となる。なお、家族構

続柄	
1	世帯主または代表者
2	世帯主の配偶者
3	子
4	子の配偶者
5	世帯主の父母
6	世帯主の配偶者の父母
7	孫
8	祖父母
9	兄弟姉妹
10	他の親族
11	住み込みの雇人
12	その他
性別	
1	男
2	女
婚姻状況	
1	未婚(幼児などを含む)
2	配偶者あり
3	死別
4	離別

表1 世帯員属性

現形式の一部を変更した。それと並行して、その間の研究内容を、本研究会の直前に別途、参考文献にある周防(2020a)と周防(2020b)に投稿した。本稿はこれら一連の作業の一つに取り纏めた成果である。

³⁾ 調査票情報と匿名データでは、抽出率がそれぞれ20%、1%なので、必ずしも存在するとは限らない。

成変数には、①各世帯内の全世帯員の続柄を連結した全体版(変数名:kanjiPat)と、②同じ続柄は一つだけにまとめた短縮版(変数名:shortPat)の二通りがあり、用途に応じて使い分けができる。

世帯構造を表現するために、我々の考案した「家族構成変数」のような発想は国内にも海外にもまだない。この変数の生成アルゴリズムは完成しており、国調や全消などの調査票情報にもすぐに適用できる。ただ、この表現法では、漢字を使う中国、台湾、コリアでは直ぐに適用できるが、英字を使う言語圏ではこのままでは使えないが、英字圏の国々でも使えるように実用化したプログラムは既に完成済みである。(ただし、続柄の定義法が国により異なる場合には、若干の修正が必要になる。)

家族構成変数は2バイトで表現される漢字の連結である。従って、半角アルファベット二文字(先頭文字は大文字、二文字目は小文字)を漢字一文字に対応させることで、漢字用に作成したプログラムがそのまま使えるようになる。例えば、「父」は Father

の「Fa」、「母」は Mother の「Mo」、「夫」は Husband の「Hb」、「妻」は Wife の「Wf」のように、いわゆる mnemonic な略称を付けておけば読みやすくなる。「父母倅」(父・母・独身の息子)なら「FaMoSs」(Ss は Son single)の表記になる。漢字表現よりは若干読みづらくなるが、慣れれば問題なく活用できるはずである。つまり、漢字用のプログラムの中で、それぞれの漢字(全角一文字)を、それに対応する英字の半角二文字で置換すれば済むが、手作業で置換作業をするとミスが起こり易いので、置換作業は、別途用意した漢字と英字の対応表を使っ

世帯の種類	続柄漢字パターン	性別	意味	婚姻状況	
単独世帯	1 単	M	单身	非婚姻	
	2 宿			婚姻中	
	3 嬋	F		非婚姻	
	4 嬢			婚姻中	
世帯主	5 男	M	「子」非同居	非婚姻	
	6 女	F		婚姻中	
	7 亭	M	非同居		
	8 奥	F	婚姻状況 問わず		
	9 父	M	「子」同居	婚姻中	
	10 母	F		世帯主夫婦	
	11 夫	M	「子」非同居	婚姻中	
	12 妻	F		婚姻中	
	単独世帯以外の世帯	13 倅	M	息子	非婚姻
		14 倅	M		婚姻中
		15 娘	F	娘	婚姻中
		16 好	F		非婚姻
17 嫁		F	息子の妻	婚姻状況 問わず	
18 婿		M			娘の夫
19 爺		M	世帯主の父		
20 袋		F			世帯主の母
21 舅		M	世帯主の配偶者の父		
22 姑		F			世帯主の配偶者の母
23 翁		M	世帯主の祖父		
24 婆		F			世帯主の祖母
25 孫		MF	世帯主の孫		
26 兄		M	世帯主の兄弟姉妹		
27 弟		M			
28 姉		F			
29 妹		F	他の親族(親戚)		
30 戚	MF				
31 雇	MF	住み込みの雇人			
32 他	M	「その他」男			
33 媒	F		「その他」女		

注：非婚姻＝未婚、死別、離別

表2 続柄漢字パターン一覧表

家族構成変数	意味
奥姉姉	婚姻中の女性(奥さん)+姉二人 (「亭」主が別世帯にいる)
男爺袋婆姉	男性(独身)+両親(親爺とお袋)+祖母+姉
父母娘孫	両親+娘(婚姻中)+孫 (娘婿が別世帯にいる)
父母娘婿孫	両親+娘夫婦+孫
母父倅好	母(世帯主)+父(世帯主の配偶者)+息子(独身)+娘(独身)
女他	女性(独身)+男性(「他」人)

表3 家族構成変数の例

て、別の SAS プログラムが自動変換にできるように設計されている。

4. エンドユーザーに提供されるマイクロデータの物理的構造

国調の調査票情報や匿名データが利用者に提供される際は、最近では固定長の CSV ファイルで提供されている。CSV ファイルなら固定長でなくても良いはずだが、メインフレーム時代からの名残と思われる。このファイルの物理的構造は、1 レコードが各世帯員に対応しており、世帯主のレコードが先頭に来る。各レコードには、先頭部にある世帯毎のユニークな識別情報に続いて世帯属性の変数が並び、その右側に一世帯員の個人属性が続く。つまり、各世帯員のレコードの先頭部には、同じ世帯属性の繰り返しがあり、その右側に各世帯員の個人属性が続いているので、世帯毎のユニークな識別情報を頼りに同一世帯の世帯員の識別ができるが、各レコードには、繰り返される世帯情報と、世帯員情報が混在しており、世帯と世帯員というレベルの異なる二種類のデータで一つのレコードが構成されている。従って、今回の家族構成変数の作成には、同一世帯を構成する全世帯員のレコードから個人属性を集めて総合的に決めつけをすることになる。家族構成変数の生成には、連続する同一世帯の全世帯員のレコードから、4つの変数「続柄」、「性別」、「婚姻状況」、「年齢」を組合せて、各世帯員に漢字一文字を割り当てる。兄、弟、姉、妹の判別には年齢比較が必要となる。

5. 国勢調査の調査票情報と匿名データの符号の違い

本稿では、国調の匿名データ(2000年、2005年)を使用しているが、今回は、調査票情報用のシステムをそのまま匿名データに適用した。その結果、調査票情報の「10:他の親族」(戚)、「11:住み込みの雇人」(雇)、「12:その他」(他・媒)が匿名データでは、「10:その他」に一括りになっているために、システム上では、「10:戚」と表示されている。参考までに、国調の調査票情報と匿名データの符号の対応表を表4に示す。

国勢調査:調査票情報			国勢調査:匿名データ				
符号	続柄	家族構成変数用漢字	符号	続柄	家族構成変数用漢字		
1	世帯主または代表者	男女亭奥夫妻父母	1	調査票情報と同じ	調査票情報と同じ		
2	世帯主の配偶者		2				
3	子	3	3			3	3
4	子の配偶者	4	4			4	4
5	世帯主の父母	5	5			5	5
6	世帯主の配偶者の父母	6	6			6	6
7	孫	7	7			7	7
8	祖父母	8	8			8	8
9	兄弟姉妹	9	9			9	9
10	他の親族	戚	10	その他	戚 11:住み込みの雇人 12:その他 も含む		
11	住み込みの雇人	雇					
12	その他	他(男) 媒(女)					

表注：調査票情報用のシステムをそのまま匿名データに適用しているため、「戚」には11、12も含まれる。

表4 国勢調査の調査票情報と匿名データの符号の違い

その結果、匿名データでは、近年増加傾向にある、いわゆる同棲や事実婚がデータからは、読み取りにくくなっている。匿名データが1%抽出であることを考慮すると、今回の匿名データで出現する続柄「10:戚」は、全て調査票情報の「12:その他」と見なしてもよいと思われる。

6. 匿名データから作成した家族構成変数の利用

国調・匿名データ(2000年、2005年)には、「世帯の家族類型」という変数があり、6種類の世帯形態に分類されている(表5)。2000年データでこの「世帯の家族類型」と、我々の提案する家族構成変数(短縮版)のクロス表を作成した。この短縮版は、全部で956パターンあったが、その内の36パターンが家族類型の2に該当した。表6に上位10位までの世帯数を示す。なお、この表の元表によると、「一般世帯」462,355世帯の内、「世帯の家族類型」(表5)の「2:夫婦と子供から成る世帯」は148,452世帯あった。

「夫婦と子供から成る世帯」と言えば、直感的には、表6の①～③の家族構成(子供は独身)を連想するのが普通である。⑤と⑥でも、両親と子供(息子または娘)が同居しているが、この子供は「現在婚姻中」であり、その配偶者は何らかの理由で別の世帯にいることになり、先の①～③の世帯とは、世帯員の年齢層や世帯単位の経済・社会的活動の点で大いに異なる。

更に、表6を見ると、上記の5つのパターンの他にも、④「男爺袋」(世帯主である男性[現在独身]+その両親)、⑦「女爺袋」(世帯主である女性[現在独身]+その両親)、⑧「男爺袋妹」(世帯主である男性[現在独身]+その両親+世帯主の妹)も含まれている。確かにこれらも「夫婦と子供から成る世帯」ではあるが、世帯主が親か子供では、経済活動や社会的活動のパターンは全く異なり、同一の世帯類型に分類するのはかなり無理があると言わざるを得ず、国調のマイクロデータに含まれる「世帯の家族類型」を使った分析結果の合理的な解釈が難しくなる。

政府の施策では、よく「モデル世帯」と称して、「夫婦と子供(現在結婚していない)二人」世帯を取り上げるが、国調・匿名データから我々が作成した家族構成変数(全体版)を使えば、父が世帯主の「父母好好、父母好倅、父母倅好、父母倅倅」と、母が世帯主の「母父好好、母父好倅、母父倅好、母父倅倅」の世帯をカウントすれば簡単に求められる。匿名データではこのモデル世帯は全世帯の約12.9%しか存在せず、施策用のモデル世帯としては適切とは言い難い。

7. 家族構成変数による世帯の変遷例

今、仮に、ひとり暮らしで、それぞれ独身の男女が結婚するケースを取り上げ、家族構成変数を

1:夫婦のみの世帯
2:夫婦と子供から成る世帯
3:男親又は女親と子供から成る世帯
4:その他の親族世帯
5:非親族世帯
6:単独世帯

表5 国調・匿名データの家族類型

① 父母倅	54,578
② 父母好	47,131
③ 父母倅好	45,122
④ 男爺袋	509
⑤ 父母倅	314
⑥ 父母娘	223
⑦ 女爺袋	74
⑧ 男爺袋妹	50
⑨ 父母倅戚	49
⑩ 父母好戚	47
11位～36位省略	
①位～⑩位の合計	148,452

表6 夫婦と子供から成る世帯

使い、図1に沿って架空の世帯構造の変遷を辿ってみよう⁴。

ステージ① 独身の男性「単」と独身の女性「嬢」の2世帯がある。

ステージ② この両者が結婚してふたりだけの世帯「夫妻」になる。

ステージ③ 長女が生まれて、「父母」になる。

ステージ④ 更に、長男が生まれて、「父母好倅」になる。

ステージ⑤ 長女が結婚し、別世帯になり、「父母倅」と「夫妻」の2世帯に分かれる。

ステージ⑥ 長男が結婚し、両親と同居のままで「父母倅嫁」になる。一方、長女の世帯では、長男が誕生し「父母倅」になる。

ステージ⑦ 実家に同居する長男夫婦に赤ん坊(世帯主の孫)が生まれ、「父母倅嫁孫」になる。一方、長女の世帯では、夫が死亡して、母子家庭「母倅」になる。

ステージ⑧ 実家の父が転勤・異動したので、実家は「母倅嫁孫」、単身赴任の世帯主は「宿」に分離。

ステージ⑨ 孫が結婚・独立の結果、実家は「母倅嫁」、孫は「夫妻」の2世帯に分離。

ステージ⑩ 長男夫婦が転勤・異動のため、実家を出たので、母親が一人暮らしの実家は「嬢」、長男夫婦は「夫妻」の2世帯に分離。

ステージ⑪ 単身赴任していた父が定年退職して、実家に戻ってきたので、実家は「夫妻」となる。

ステージ⑫ 実家の世帯主である「夫」が死亡して、実家の世帯は「嬢」となる。

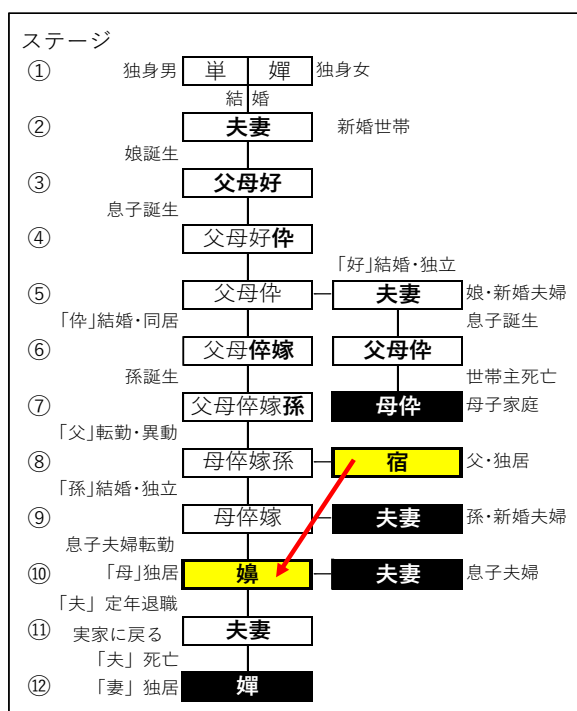


図1 世帯の変遷例

ステージ①で結婚してできた新婚世帯が、最終ステージ⑫では、図1の黒塗りで示した4つの世帯(「母倅」、「夫妻」、「夫妻」、「嬢」)に分離したことになる。これらの世帯は、配偶者との離婚・死別、再婚等を経て、更に複雑に分割・合体を繰り返しながら変貌していく。また、ステージ⑫の「嬢」は、いずれ消滅する。

8. 世帯主は誰か

各世帯の社会・経済活動は、世帯主が誰であるかによって大きく左右される。匿名データを使って、家族構成変数の中で使われる世帯主用の漢字「父母亭奥夫妻男女単嬢宿嬢」の出現件数を、2000年と2005年で比較した⁵(表7)。

⁴ 以下の解説にある「」内は、家族構成変数の中で用いられる漢字。

⁵ 本稿で利用する匿名データは1%抽出で、集計用乗率の変数はない。従って、本稿に掲載した全ての表の数字の100倍程度

	世帯主	2000年		2005年		世帯主	2000年		2005年		性別
		実数	%	実数	%		実数	%	実数	%	
単 独 世 帯	男性	78,857	16.4	86,629	17.1	宿	7,651	1.6	8,124	1.6	男
		71,206	14.8	78,505	15.5	単	71,206	14.8	78,505	15.5	
	女性	69,834	14.5	80,859	15.9	嬢	3,046	0.6	3,679	0.7	女
		66,788	13.9	77,180	15.2	嬢	66,788	13.9	77,180	15.2	
	小計	148,691	30.8	167,488	33.0						
単 独 世 帯 以 外 の 世 帯	父	194,461	40.3	186,989	36.8	子と同居中の父親(婚姻状況問わず)					
	母	29,499	6.1	33,845	6.7	子と同居中の母親(婚姻状況問わず)					
	亭	467	0.1	494	0.1	子と非同居、現在結婚中(妻と別居)の亭主					
	奥	176	0.0	184	0.0	子と非同居、現在結婚中(夫と別居)の奥さん					
	夫	95,476	19.8	103,866	20.4	子と非同居、現在結婚中(妻と同居)の夫					
	妻	740	0.2	937	0.2	子と非同居、現在結婚中(夫と同居)の妻					
	男	8,356	1.7	9,594	1.9	子と非同居、現在結婚していない男					
	女	4,138	0.9	4,594	0.9	子と非同居、現在結婚していない女					
	小計	333,313	69.2	340,503	67.0	注 「子と非同居」とは、 子が「別居、死別、最初からいない」のいずれか					
	世帯総数	482,004	100.0	507,991	100.0	(国勢調査・匿名データ)					

表7 世帯主は誰か？

世帯総数では2005年の方が26,000世帯ほど多いので、全世帯に占める割合(%)で比較している。

①世帯主では、やはり「父」が4割前後を占めるが、2005年では、「父」が4ポイントほど減少し、逆に「母」が0.6ポイント増加した。

②世帯主夫婦(亭・奥)が別居している世帯の数は650世帯前後と少なく、変動は殆どない。

③子と同居していないか、または、子のいない世帯主夫婦(「夫」・「妻」)では、夫が世帯主の割合が若干増加した。

④現在結婚していない男性が世帯主「男」の世帯は、0.2ポイント増加した。一方、現在結婚していない女性が世帯主「女」の世帯は、ほぼ横這いである。

⑤単独世帯は、男女全体で2.2ポイント増加した。男女別では、それぞれ、それぞれ0.7ポイント、1.4ポイント増加した。

⑥単独世帯の独身者の割合は、男性(「単」)・女性(「嬢」)、それぞれ0.7ポイント、1.3ポイント増加した。一方、既婚者の割合は、男性(「宿」)は変わらず、女性(「嬢」)は0.1ポイント増加した。

9. 世帯内の夫婦の分布

ひとつの世帯の中にどのような種類の夫婦が含まれているのかを調べるのは、世帯構造を知る上で重要な情報のひとつである。ここ10数年来、筆者は世帯内の夫婦や家族構成について、関心を持ってきた。そのきっかけは、30年以上継続してきた日本の芸術家の方々に対するアンケート調査である。多岐に亘るその調査票項目の中に、調査対象者である芸術家本人のご家族の芸術活動状況があった。調査をする側の意図は、そもそも大人になって人並みの生活していく上で、将来を予めきちんと見通せない職業である芸術家を選んだ理由を知りたかったからである。そこで立てた仮説は、芸術家本人が幼少期から、同居する親・兄弟・祖父母等から芸術的な何らかの影響

の値が、実際の推定件数と思われる。

や教育活動があったのではないかと、ということであった。

その中で、夫婦が結婚前に、互いに芸術的な影響を受けあった可能性もあるのではないかと、という仮説から、以前に国調の調査票情報を使って、夫婦を同定し、職業小分類コードを使って分析を試みたことがある。統計局から提供を受けた個票データのファイル構造は、今と殆ど変わっていない。その元ファイルから分析用のデータベースを構築した後、世帯内の夫婦の同定作業を行ったが、その時は、かなり複雑なアルゴリズムになった。人間の目でファイルを見ていけば、その同定作業自体は難しいことではないが、膨大なデータ件数では目視の作業は論外である。

そんな中、数年前から、ひとつの変数の中に、一世帯分の全世帯員の血縁関係を、一見して分かる形で詰め込めないか思案していた。その結果、辿り着いた変数が今回解説している家族構成変数である。

近年の核家族化の流れの中で、同一世帯に世代の異なる夫婦が同居することは次第に少なくなっていると思われるので、本節では、世代の異なる夫婦の分布を分析してみた。

本システムでは、同一世帯に同居する夫婦には、以下の7種類がある⁶。

①「父母」、②「夫妻」、③「倅嫁」、④「娘婿」、⑤「爺袋」、⑥「舅姑」、⑦「翁婆」

表8にこれらの夫婦を含む世帯について、国調の匿名データから求めた件数を示す。抽出率が1%のデータなので、ざっと百倍すれば実数に近い数字が得られると思われる。

この内、③と④については、同一世帯に複数個、または複数組合まれることがあり得るが、そのケースは今日では極めて希になってきている。例えば、家族構成変数が、「母倅嫁嫁」の場合、二人の嫁のどちらが「倅」の嫁なのかを決定するには、ふたりの嫁の婚姻状況を見れば「概ね」判定できる。例えば、ひとりの「嫁」が「婚姻中」、もうひとりの「嫁」が「死別」であれば、前者が倅の嫁と判定できる。しかし、二番目の嫁も「婚姻中」(例えば、その夫が単身赴任中で別居中)なら、どちらの組合せが夫婦か断定は出来ない。

夫婦形態	2000年	2005年
①父母	187,508	179,472
②夫妻	96,216	104,803
③倅嫁	14,664	12,950
④娘婿	3,016	2,847
⑤爺袋	5,598	4,566
⑥舅姑	921	862
⑦翁婆	24	20

(世帯数)

表8 夫婦の形態

10. プログラミングの解説

本節では、SAS 言語を使って、家族構成変数の利用法について、若干の解説を行う。

前節の①～⑦の夫婦形態を含む世帯を検索してみよう。

家族構成変数(全体版)kanjiPat に対して、以下に示すように、KINDEXC 関数を使えば、「倅」かつ「嫁」を含む世帯を簡単に特定することができる。この例では、家族構成変数 kanjiPat(長さは最大漢字8文字)には母倅嫁嫁□□□□が格納されているとする。(□は半角スペース二つ分)

```
x1=KINDEXC(kanjiPat, "倅");  
x2=KINDEXC(kanjiPat, "嫁");  
if x1 > 0 AND x2 > 0 then output fufu3;
```

最初の二つの命令文を実行すると、2変数 x1 と x2 の値は、それぞれ 2 と 3 となり、漢字

⁶ 「①父母」・「②夫妻」は、それぞれ、子と同居・非同居の世帯主夫婦。

の「倅」と「嫁」が母倅嫁嫁□□□□(kanjiPat の値)の左から数えて2文字目と3文字目にあることが分かる。この場合は、当該漢字がそれぞれ少なくともひとつあればよいので、同じ漢字が複数個あるかどうかはチェックする必要がない。因みに、当該漢字が見つからなければ KINDEXC 関数には値として0が返される。この二つの数値は世帯内でユニークな世帯員 id 番号に対応している。

上記の例では、直後の if 文では条件式が満足されたので、当該世帯の中に探している夫婦形態が見つかったことになり、当該オブザベーションの世帯 id 番号、および、世帯内ユニーク id 番号(この場合、2と3)が、事前に指定された3番目のデータセット(fufu3)に、保存される。同様の処理を残りの夫婦形態にも適用すれば、7つのデータセット fufu1~fufu7 が出来上がる。この7つのデータセットに含まれるそれぞれのオブザベーションの件数が、表8である。

筆者が構築した国調・匿名データの SAS データベースは、世帯情報データセットと世帯員情報データセットから成る。前者には、ユニークな世帯 id 番号が付与され、世帯情報が格納されている。後者には、ユニークな世帯 id 番号の他に、世帯内でユニークな世帯員 id 番号と共に、当該世帯員の個人情報および、同一世帯内の全世帯員の4つの基本属性(続柄、性別、婚姻状況、年齢)が保存されている。先に計算した変数 x1 と x2 は、世帯内ユニーク id 番号に対応しているので、世帯と世帯員のユニーク id 番号を使えば、上のふたつのデータセットにアクセスできる。

世代の異なる複数の夫婦が同居する世帯は、fufu1~fufu7 の複数のデータセットにそれぞれ含まれるので、この7つのデータセットを縦にマージした後、世帯 id 番号でソートして同じ世帯 id 番号のオブザベーションを連続に並べれば、同一世帯にどの夫婦形態が何組あるかが容易に分かる。

なお、変数 kanjiPat を構成する各漢字一文字を、各世帯員の世帯内続柄の変数として、世帯員情報データセットにも埋め込んでおくと、更に便利である。

11. 8050 問題の現状

世帯の形態は、現実社会において、様々な格差を生む要因のひとつと考えられる。そのひとつに「8050 問題」がある。80 代の親が 50 代の子供の生活を経済的に支えるという問題。その背景には子どもの「ひきこもり」がある。ひきこもりは当初、1980 年代~90 年代には若者の問題と考えられていた。しかし、約 30 年の年月が経った今、当時の若者が 40~50 歳代になっても定職に就かず、70~80 歳代の高齢になった親の家に独身のまま同居し続けて、経済的援助を受けながら、その家に籠もってしまう、いわゆるパラサイト・シングルとなる。

本稿ではこの問題の本質を論じる意図はなく、2000 年国調の匿名データから作成した家族構成変数を使って、そのような引きこもり世帯がどれくらいの割合で存在するのか、推定を試みた。

8050 問題を抱えている可能性のある世帯は、家族構成変数を使えば、以下の様に2種類に表現できる。子供の年齢は 40 歳以上とした。

- (1) 世帯主が「父」または「母」の場合：
 - ① 未婚の息子「倅」・未婚の娘「好」が世帯員
 - ② かつ、子供の年齢が 40 歳以上
- (2) 世帯主が「男」(息子)又は「女」(娘)の場合：
 - ① 親が世帯員(「爺」・「袋」)
 - ② かつ、子供(世帯主)の年齢が 40 歳以上

家族構成変数を使いながら段階を追って、世帯を絞っていき、該当する可能性のある世帯を絞り込んで行った経過を表9に示す。最終的に、表9の黒塗りのセルにあるように、(1)のケースが9,335世帯、(2)が3,403世帯あった。

2000年国勢調査・匿名データ：全世帯数 482,004世帯					
①「父」または「母」が世帯主の世帯 223,960世帯			②「男」または「女」が世帯主の世帯 12,494世帯		
「伴」・「好」（世帯主の「子」：未婚・死別・離別）が			「爺」・「袋」（世帯主の親）が		
いる世帯			いる世帯		
205,842世帯			6,869世帯		
いない世帯			いない世帯		
18,118世帯			5,625世帯		
「伴」・「好」40歳以上		40歳未満	世帯主 40歳以上		40歳未満
12,173世帯		193,689世帯	4,882世帯		1,987世帯
未婚		死別・離別	未婚		それ以外
9,335世帯		2,838世帯	3,403世帯		1,479世帯

表9 パラサイト・シングルがいる可能性のある世帯

ただ、これだけでは絞り込みでは不十分で、このままでは、定職に就いて、それなりの収入があり、親と同居している独身の中高年齢者まで含まれてしまう。当然、仕事をどれ程しているのかも考慮する必要がある。国調では収入額は調べないが、2000年の調査では、「9月24日～30日までの一週間に仕事をした時間」の記入欄がある⁷。匿名データにもその仕事時間の実数の変数はあるが、その他に、10区分や14区分などにまとめた就業時間の変数もある。今回は、10区分の労働時間の変数を使い「労働力状態」とのクロス表(表10、表11)を、「子」の男女別に作成した。

1日8時間、週5日勤務で週40時間になるので、表10と表11に、就業時間が週40時間の所で線を引いてみた。

更に、教育歴、雇用形態など、匿名データには他にも分析に役に立つ情報が含まれているので、そうした情報も利用すれば、このクロス表から、更なる踏み込んだ分析作業も可能となる。

①「父母伴好」がいる世帯

平均一週間の就業時間(10区分)	男「伴」：世帯主の親と同居する未婚の息子							女「好」：世帯主の親と同居する未婚の娘						
	労働力状態							労働力状態						
	1:就業者	2:仕事を探していた(完全失業者)	3:家事	4:通学	5:その他	9:不詳	合計	1:就業者	2:仕事を探していた(完全失業者)	3:家事	4:通学	5:その他	9:不詳	合計
非就業者	0	702	102	22	725	50	1,601	0	189	726	13	363	46	1,337
1:休業者	52	0	0	0	0	0	52	21	0	0	0	0	0	21
2: 1～4時間	5	0	0	0	0	0	5	13	0	0	0	0	0	13
3: 5～9時間	44	0	0	0	0	0	44	41	0	0	0	0	0	41
4:10～14時間	58	0	0	0	0	0	58	55	0	0	0	0	0	55
5:15～29時間	143	0	0	0	0	0	143	184	0	0	0	0	0	184
6:30～34時間	195	0	0	0	0	0	195	146	0	0	0	0	0	146
7:35～39時間	183	0	0	0	0	0	183	232	0	0	0	0	0	232
8:40～48時間	2,134	0	0	0	0	0	2,134	1,146	0	0	0	0	0	1,146
9:49～59時間	871	0	0	0	0	0	871	269	0	0	0	0	0	269
10:60時間以上	450	0	0	0	0	0	450	142	0	0	0	0	0	142
不詳	10	0	0	0	0	0	10	3	0	0	0	0	0	3
合計	4,145	702	102	22	725	50	5,746	2,252	189	726	13	363	46	3,589

2000年国勢調査 匿名データ 男女計 9,335

表10 親(世帯主)と同居する40歳以上の未婚の子供がパラサイト・シングルの可能性のある世帯

⁷ 余談になるが、本研究会の報告準備をしていた時が国調の実査期間中だったので、2020年の調査票を見直したら、既に「仕事時間」数の記入欄はなくなっていた。

②「男女爺袋」がいる世帯

平均一週間の 就業時間 (10区分)	「男」：親と同居する未婚の息子(世帯主)							「女」：親と同居する未婚の娘(世帯主)						
	労働力状態							労働力状態						
	1:就業者	2:仕事を 探していた (完全失業者)	3:家事	4:通学	5:その他	9:不詳	合計	1:就業者	2:仕事を 探していた (完全失業者)	3:家事	4:通学	5:その他	9:不詳	合計
非就業者	0	213	41	2	206	6	468	0	26	236	0	92	4	358
1:休業者	53	0	0	0	0	0	53	7	0	0	0	0	0	7
2: 1～ 4時間	6	0	0	0	0	0	6	3	0	0	0	0	0	3
3: 5～ 9時間	11	0	0	0	0	0	11	4	0	0	0	0	0	4
4:10～14時間	17	0	0	0	0	0	17	7	0	0	0	0	0	7
5:15～29時間	71	0	0	0	0	0	71	29	0	0	0	0	0	29
6:30～34時間	93	0	0	0	0	0	93	24	0	0	0	0	0	24
7:35～39時間	106	0	0	0	0	0	106	34	0	0	0	0	0	34
8:40～48時間	1,141	0	0	0	0	0	1,141	158	0	0	0	0	0	158
9:49～59時間	453	0	0	0	0	0	453	41	0	0	0	0	0	41
10:60時間以上	291	0	0	0	0	0	291	23	0	0	0	0	0	23
不詳	5	0	0	0	0	0	5	0	0	0	0	0	0	0
合計	2,247	213	41	2	206	6	2,715	330	26	236	0	92	4	688

2000年国勢調査 匿名データ 男女計 3,403

表11 親と同居する40歳以上の未婚の子(世帯主)が
パラサイト・シングルの可能性のある世帯

12. 今後の展望

今般、家族構成変数を開発した狙いは、世帯の分類の際に、分析用の分類変数を事前に用意するのではなく、誰にでも瞬時に、直感的に目で見えて理解できる形で世帯構造を表象する「変数」を準備しておき、分析段階で、各自の分析目的に応じて、自前の世帯分類を容易にできるようにするためである。

本稿では詳細には触れなかったテーマのひとつに、近年増加傾向にある同棲や事実婚があるが、家族構成変数と年齢情報があれば、調査票情報から、ほぼ確実にその実態を明らかにできる⁸。

更に、次のステップとして、国調・調査票情報を用いて家族構成変数を作成し、各世帯員の生涯の変化、例えば、結婚、出産、離婚、死別、再婚、同棲、進学、卒業、就職等の人生のイベントが発生する確率情報を入手、或いは、想定して、各世帯員の5年毎の生活環境の変化をシミュレーションし、各世帯の世帯構造の変遷を追っていき、数十年後の世帯構造の変遷の状況を観察しながら、長期的人口減少傾向を可視化したい。

この家族構成変数を使えば、世帯構成の分布が直感的に把握できる。ただ、国調だけでは経済・労働・余暇活動・住居環境などの関連情報が十分でないが、全消、就業構造基本調査、労働力調査、社会生活基本調査、住宅・土地統計調査等の調査票情報からは入手可能である。幸い、これらのマイクロデータには、国調の調査票情報とほぼ同様の世帯員の基本的な4つの個人属性があり、家族構成変数を作成できるので、世帯構造別による社会・経済活動等の詳細な観察が可能となる。

マイクロデータ分析の研究分野では、これまで、異なる統計調査のマイクロデータに対して、様々な手法を使って個票同士のマッチングが行われてきたが、我々は視点を変えた「マッチング」手法を目指している。つまり、異なる統計調査のマイクロデータから世帯構成変数をそれぞれ作成し、その家

⁸ 母子家庭については、2019年の本研究会で報告している(周防・安井 2020)。

族構成変数を介して、世帯構成毎に、異なる統計調査から、経済・労働・余暇活動・住居環境などの情報を収集・結合することで、世帯の実態を様々な側面から観察することを目指している。

参考文献

- ① 周防節雄(2016)「国勢調査マイクロデータを使った新しい世帯分類方式「新世帯類型」の設計と開発」、『公的統計のマイクロデータ等を用いた研究の新展開(平成28年度)報告要旨集』、統計数理研究所、pp249-267 (https://www.nstac.go.jp/services/pdf/161125_5-3-2.pdf)
- ② 中川雅義、米澤香、新井郁子(2016)「国勢調査マイクロデータから新開発した新世帯類型による世帯構造の長期時系列分析の結果」、『公的統計のマイクロデータ等を用いた研究の新展開(平成28年度)報告要旨集』、統計数理研究所、pp217-248 (https://www.nstac.go.jp/services/pdf/161125_5-2-2.pdf)
- ③ (公財)統計情報研究開発センター(2016)『新たな個人消費を創造する新世帯類型別世帯数・世帯人員』、Sinfonica 研究叢書、2016年3月
- ④ 伊藤彰彦・中川雅義・周防節雄・米澤香・安井浩子・新井郁子(2017)『新世帯類型の構築と世帯構造の変動に係る長期時系列分析の結果』、(公財)統計情報研究開発センター、2017年
- ⑤ 周防節雄・安井浩子(2020)「国勢調査の続柄情報に婚姻状況・性別・年齢を加味して世帯員構成を直感的に表現できる変数の開発」、『官民オープン利活用の動向及び人材育成の取り組み(2019年度)報告要旨集』、独立行政法人 統計センター、pp.81-91、2020年3月
- ⑥ 周防節雄(2020a)「世帯を調査客体とする公的マイクロデータの世帯構造を直感的に把握できる「家族構成変数」の開発と利用(1)」、『ESTRELA 9月号』、pp.25-30、(公財)統計情報研究開発センター、2020年9月
- ⑦ 周防節雄(2020b)「世帯を調査客体とする公的マイクロデータの世帯構造を直感的に把握できる「家族構成変数」の開発と利用(2)」、『ESTRELA 11月号』、pp.25-31、(公財)統計情報研究開発センター、2020年11月

謝辞

本研究の遂行には国勢調査の匿名データを使用したが、その際、(株)SAS インスティテュートジャパンからデータ使用料の補助を賜った。また、国勢調査の匿名データの利用に際しては、独立行政法人統計センターから助言や便宜を頂いて頂いた。本稿は、財団法人統計情報研究開発センターの同意を得て、同センター発行の月刊誌『エストレーラ』の2020年9月号、及び2020年11月号に掲載された拙稿をひとつに取り纏めて本稿を完成することが出来た。ここに記して、関係各所に対して謝意を表したい。

附録 国勢調査の調査票情報の続柄を基にした世帯員の階層構成

