

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

周防 節雄

安井 浩子

(公財) 統計情報研究開発センター

共同研究集会「官民オープンデータ利活用の動向及び人材育成の取組」

@統計数理研究所

2019年11月15日

概要

- 使用マイクロデータ → **国勢調査**(2000, 2005年)・**匿名データ**
- 入手データ(CSVファイル) → **自動変換** → SASデータベース化
- テーマ → **世帯の分類法** / **世帯員構成の表現法**
- **家族類型**(統計局) → **新世帯類型**(シンフォニカ)



家族構成変数 = 各世帯員を漢字一文字

それらの漢字を横一列に一つの変数に詰め込む

例：両親・息子(未婚)・娘(未婚) = 「**父母**倅**好**」

両親・息子・嫁・世帯主の母 = 「**父母**倅**嫁**婆」

付録1 - ①

1	2	3	4	5	6	7	8		
00200521,0000190004,200010,28,100,0000000001	0000000001	01	最初の世帯 (3名)	2000年 国勢調査・ 匿名データの CSVファイル (固定長)
00200521,0000190004,200010,28,100,0000000002	0000000001	02	...	省	2番目の世帯 (4名)	
00200521,0000190004,200010,28,100,0000000003	0000000001	03	略		
00200521,0000190004,200010,11,999,0000000004	0000000002	01		
00200521,0000190004,200010,11,999,0000000005	0000000002	02		
00200521,0000190004,200010,11,999,0000000006	0000000002	03		
00200521,0000190004,200010,11,999,0000000007	0000000002	04		

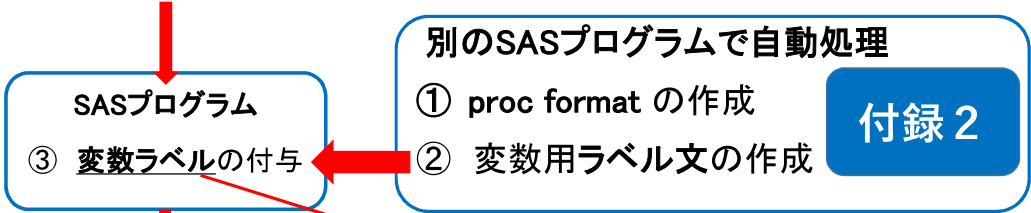
SASプログラム
proc import

SASデータセット

	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8
1	200521	190004	200010	28	100	1	1	1
2	200521	190004	200010	28	100	2	1	2
3	200521	190004	200010	28	100	3	1	3
4	200521	190004	200010	11	999	4	2	1
5	200521	190004	200010	11	999	5	2	2
6	200521	190004	200010	11	999	6	2	3
7	200521	190004	200010	11	999	7	2	4

残りの変数省略

付録1 - ②



	政府統計コード	管理コード	調査年月(西暦)	都道府県	市区町村	レコード一連番号	世帯一連番号	世帯員番号
1	200521	190004	200010	28	100	1	1	1
2	200521	190004	200010	28	100	2	1	2
3	200521	190004	200010	28	100	3	1	3
4	200521	190004	200010	11	999	4	2	1
5	200521	190004	200010	11	999	5	2	2
6	200521	190004	200010	11	999	6	2	3
7	200521	190004	200010	11	999	7	2	4

(必要に応じて
変数名のRENAME)

CSVファイルの
SASデータセットへの変換終了

付録2

符号表
(2000年
国勢調査・
匿名データ)

行 #	項目名	省略	項目番号	省略	対象	符号	符号内容
13	世帯一連番号	...	7	...		000000001~	000000001~
14				
17	世帯員番号	...	8	...		01~08	世帯内番号
18	世帯の種類	...	9	...		1	一般世帯
19			2	施設等の世帯
21	住宅の建て方	...	10	...		1	一戸建
22			2	共同住宅
23			3	その他(長屋建を含む)
24			△	対象外
92		*		
93	世帯主との続き柄	...	19	...		01	世帯主または代表者
94			02	世帯主の配偶者
95			03	子
96			04	子の配偶者
97			05	世帯主の父母
98			06	世帯主の配偶者の父母
99			07	孫
100			08	祖父母
101			09	兄弟姉妹
102			△	その他 **

* 住宅に住む一般世帯
** その他(他の親族, 住み込みの雇人を含む)

SASプログラム用format

```
proc format /*2000年国調匿名データ*/
;
*世帯の種類;
value x9f
1="1:一般世帯"
2="2:施設等の世帯"
;
*住宅の建て方;
value x10f
1="1:一戸建"
2="2:共同住宅"
3="3:その他(長屋建を含む)"
.;
*世帯主との続き柄;
value x19f
1="1:世帯主または代表者"
2="2:世帯主の配偶者"
3="3:子"
4="4:子の配偶者"
5="5:世帯主の父母"
6="6:世帯主の配偶者の父母"
7="7:孫"
8="8:祖父母"
9="9:兄弟姉妹"
10="10:その他(他の親族住み込み..."
;
```

SASプログラム用 LABEL文

```
VAR7="世帯一連番号"
VAR8="世帯員番号"
VAR9="世帯の種類"
VAR10="住宅の建て方"
...
VAR19="世帯主との続き柄"
...
```

SASプログラムで
自動作成

なぜ自動変換で SAS データベース化？

・調査票情報、匿名データ

① **変数、レコードの数**が多い。

e.g. 2005年国調・調査票情報(20%抽出)：約**1600万件**、132変数

2005年国調・匿名データ(1%抽出)：約**124万件**、33変数

② 一度に利用する **調査回次**が多い、時には5回次分以上

③ 調査年毎に **CSV ファイルのデザイン、変数のコード値**が異なる

④ 膨大な個数の **変数フォーマット**や**変数ラベル文**の作成は、
一組のプログラムだけで、**全ての回次分を一括処理**したい。

⑤ **データベース化作業の効率化で節約できた時間**を、
分析作業や論文作成に回せる。

なぜ、家族構成変数？

・国勢調査「家族類型」

一例：「夫婦と子供から成る世帯」

①「世帯主夫婦と子供から成る世帯」

↑↓ **社会通念上異なる性格の世帯？**

②「世帯主と両親から成る世帯」

③「世帯主と両親、および世帯主の兄弟姉妹から成る世帯」



もっと直感的に理解できる世帯の表現法？

国勢調査(2000年以降)

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

年齢：一歳刻み



家族構成変数に使用される漢字

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

注1：非婚姻中＝未婚、死別、離別
注2：単独世帯のみ、補助情報を世帯漢字パターンに付加する

図2 続柄漢字パターン表現一覧

付録3 続柄漢字パターンの 世帯内階層構造図

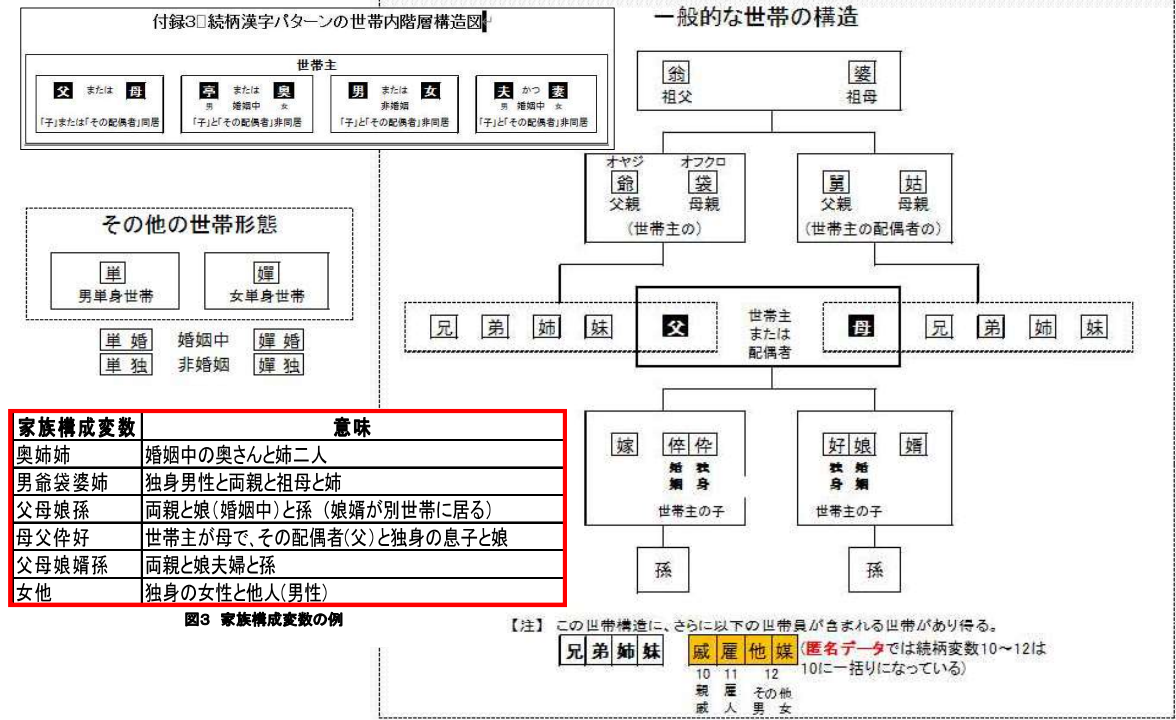
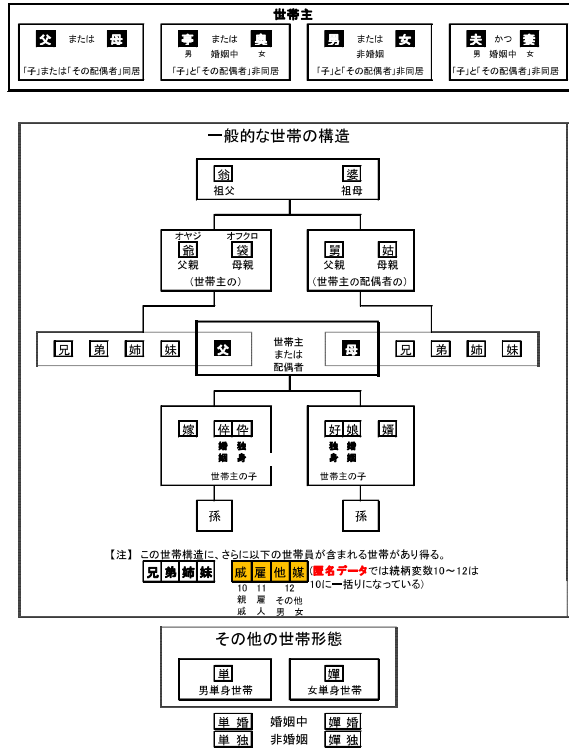


図3 家族構成変数の例(実験用データ)

家族構成変数	意味
奥姉姉	婚姻中の奥さんと姉二人
男爺袋婆姉	独身男性と両親と祖母と姉
父母娘孫	両親と娘(婚姻中)と孫(娘婿が別世帯に居る)
母父倅好	世帯主が母で、その配偶者(父)と独身の息子と娘
父母娘婿孫	両親と娘夫婦と孫
女他	独身の女性と他人(男性)

家族構成変数を利用して国調・匿名データに含まれる変数の定義の確認例

・家族構成変数

- ①世帯内の全世帯員の続柄を羅列:**全体版**(変数名: kanjiPat)
- ②同じ続柄は一つだけにまとめた**短縮版**(変数名: shortPat)

shortPat X 世帯の家族類型

shortPat	2:夫婦と子供から成る世帯
合計	148,452
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

欠損値 (19,649件)

国調・匿名データ (2000年)

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

図4 世帯の家族類型

shortPat X 世帯の家族類型

	shortPat	2:夫婦と子供から成る世帯
	合計	148,452
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

今後の研究計画

- ① 複数回次の**国調**と**全消**の**調査票情報**の利用申請を行い、両調査の調査票情報のデータベースを構築し、**家族構成変数を生成**して、データベースに組み込む。
- ② **国調・調査票情報**から作成した**家族構成変数**を用いて、**世帯構造の時系列的変遷**を分析して、これから確実に少子高齢化が進展する中で、世帯構造の側面から変遷を分析することによって、**家族構成変数の有用性**を示す。
- ③ **全消・調査票情報**から作成した**家族構成変数**も使い、**世帯構造と家計収支**の関係を分析したい。
- ④ **家族構成変数を介して、国調と全消**のデータを**関連づけて**全国規模で**世帯形態と収支状況の関係**を分析したい。

謝辞

本研究の遂行に国勢調査の匿名データを使用した。その際、(株)SASインスティテュート ジャパンから使用料の援助を賜った。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便宜を図って頂いた。

ここに記して謝意に替えたい。