

国勢調査と全国消費実態調査の匿名データの統計量を 家族構成変数で統合する実験結果の紹介

周防 節雄

(公財)統計情報研究開発センター

要旨

国勢調査の匿名データに含まれる世帯員の個人属性を組み合わせ、各世帯の世帯員構成が一目で分かる新発想の「家族構成変数」を発案し、昨年の本研究会で報告した。それとは別に、以前に利用した全国消費実態調査の匿名データにも、ほぼ同様の手法を適用・作成した家族構成変数毎の統計量が一部、手元に残っていた。そこで、今回、国調の匿名データに含まれる唯一の数値変数である「就業時間」について、家族構成変数を使っていくつかの家族形態別に計算した後、両調査から計算した様々な「統計量」を、この家族構成変数をキーとして「統合」する実験を行った。この統合実験の意義としては、異なる調査から得た様々な社会生活に関する統計量を、世帯構造という概念をキーとして統合することで、各世帯構造に特徴的な様々な生活実態を浮かび上がらせることが可能となる。その実験の途中経過を報告する。この手法を使えば、世帯を調査客体とするその他の統計調査データなら、ほとんどの調査データにも適用できるので、多数の異なる統計調査からの解析結果を、家族構成変数という共通の物差しで「統合」して、世帯構造毎の多面的な角度からの生活実態の解析が可能となる。

1. はじめに

1980年代半ばのメインフレーム・コンピュータの時代から、筆者は、公的統計のマイクロデータ(当時は、官庁統計の個票データと呼称していた)の分析を始め、今日に至っている。当時から、貴重な統計調査データを更に効率よく利用したいという発想で、個票データレベルで、異なる統計調査データを「照合」しようとする手法が論じられ、開発されてきた。世帯を調査客体とする公的統計の分野では、完全一致マッチング(Exact Matching)と統計的マッチング(Statistical Matching)という2つの手法が主流であった。

世帯を調査客体とする調査では、ほとんどの場合、各世帯の属性の他に、各世帯を構成する世帯員それぞれの「個人属性」も調査項目にあり、データ構造も、世帯属性と世帯員属性という二層構造になっているのが普通である。こうした二層構造を持つ複数の異なるデータに対して、従来のマッチング手法で果たして意味のあるマッチングが可能なのか、長い間考えを巡らせてきた。

その過程で、「直感的に」家族構成を識別できる変数を開発しようと思い立った。世帯を調査客体とする統計データには、ほとんどの場合、世帯員全員について、4つの個人属性、①続柄、②婚姻状況、③性別、④年齢の情報がある。この4つの変数があれば、それぞれの世帯の家族構成が判別できる。結局、この4つの情報を「一変数に集約」して、かつ、誰が見ても説明なしに家族構成が一目で分かる変数として、「家族構成変数」を開発した。家族構成変数の開発過程および、その仕組みについては周防(2021)で詳細に解説しているので本稿では省略し、付録1に、その概略図を

示すに留める。

本稿では、この変数を共通項として使って、異なる統計調査データから得られる統計量を算出して、家族構成毎にその結果を「統合」する試みを紹介する。従って、ここで言う「統合」とは、従来行われてきた「照合」や「マッチング」とは全く異なる概念である。

2. 世帯構造と各世帯の社会・経済活動の関連性

各世帯の社会・経済活動が、その家族構成と密接に関連していることは議論の余地がない。例えば、8050(9060)問題、少子高齢化、人口減少、ひとり親世帯、貧困世帯、経済格差などをはじめとする懸案事項は、それぞれの世帯の家族構成も大いに影響していると考えるのが自然である。本稿では、(独)統計センターが提供している 2000 年国勢調査(以後、国調)と 1999 年全国消費実態調査(以後、全消)の匿名データから、家族構成変数を利用して、前者から就労時間、後者から収入・支出の状況を算出して、各世帯の家族構成による経済活動の傾向の比較を試みた。

3. 従来からマイクロデータ提供時に組み込まれている世帯分類に関連する変数

国調では、昭和 45 年から「家族類型」という世帯分類が広く使われてきた。現在では、世帯を調査客体とする殆どの統計調査の個票データにも、同様の変数が組み込まれている。因みに、2000 年国調・匿名データでは、表 1 に示す様に、世帯区分を表す変数として、「世帯の種類」、「世帯の家族類型」、「3世代世帯」の3つの変数が含まれている。ただ、各コードに付

表1 2000年国調・匿名データの世帯分類変数

世帯の種類	1	一般世帯
	2	施設等の世帯
世帯の家族類型 (一般世帯)	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
3世代世帯 (一般世帯)	△	対象外
	1	3世代世帯:3世代以上世帯を含む
	0	上記以外
	△	対象外

けられた日本語の詳細な意味の解説は、脚注のサイト¹にあるが、「世帯の家族類型」の1~6のコードとの対応が複雑過ぎて、実用上の難点があるように思う。例えば、「世帯の家族類型」の「2:夫婦と子供から成る世帯」にある世帯をよく見ると、通常、我々が普通に思い浮かべる世帯とはかけ離れた家族構成も含まれている。この中には、「世帯主夫婦と子供から成る世帯」(いわゆる、子育て世帯)と「世帯主と両親から成る世帯」が混在していることは、確認済みである。このことを認識せずに行った分析結果はどう解釈すればよいのか、疑問が残る²。

4. 世帯構造を表現する家族構成変数に使用される漢字

¹ <https://www.stat.go.jp/data/kokusei/2000/kihon1/00/yougo.html>

² この点については、周防(2021)でも議論している。

家族構成変数の全体の設計は、国調・調査票情報を使い、4つの個人属性、①続柄、②婚姻状況、③性別、④年齢から、世帯内の世帯員間の関係を33個の漢字で表現した(表2)。それを今回は、国調・匿名データにそのまま適用した。

世帯員の各続柄漢字の内、表2に示す様に、調査票情報の符合10~12が、匿名データではコード値が一括りになって10になっており、前者のシステムをそのまま適用した結果、

表記される続柄漢字が「戚」に統一されてしまっているため、結果の解釈には注意が必要になる。これらの漢字の世帯内における階層構造については、周防(2021)で解説済みであるが、今回、一部変更した箇所についてだけここで述べる。

続柄が兄・弟・姉・妹は、開発当初は、世帯員だけではなく、「世帯主」の表現にも使用したが、他の世帯員との関係を考慮すると、若干違和感があったので、世帯主には使用しないことにした。その代わりに、兄弟姉妹の内の誰かが世帯主の世帯では、性別を考慮して、「世帯主」の表記としては、婚姻関係が不明の場合には、単に、「男」・「女」に変更し、今後は、「兄・弟・姉・妹」は「世帯員」としてだけの表記に使用するようにシステム変更をした(付録1)。

5. 家族構成変数を含む家族構成データセットの構造

世帯を調査客体とする統計調査のデータは、データ構造上、「世帯」のデータセットと、「世帯員」のデータセットから成り、両者はユニークな世帯識別情報(世帯 ID 番号)で連結されている。本節では、家族構成変数を含む「家族構成データセット」の構造を解説する(付録2に図解)。

国調の家族構成データセットは、以下に示す4種類の情報から構成されている。

- ① 世帯員情報のデータセットから新規作成された2種類の家族構成変数

kanjiPat(全体版): 全ての世帯員を表示。

shortPat(短縮版): 同じ世帯員表記が複数個ある場合は、一つだけ表示。

- ② 世帯 ID 番号
- ③ 世帯員の人数
- ④ 元の世帯員情報のデータセットから世帯単位にまとめ直した世帯員全員の元々の4つの個人属性(続柄、性別、年齢情報、婚姻状況)

表 2 家族構成変数に使用される 33 個の漢字

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	
単 独 世 帯 以 外	1	世帯主または代表者	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			
	4	子の配偶者	4			
	5	世帯主の父母	5			
	6	世帯主の配偶者の父母	6			
	7	孫	7			
	8	祖父母	8			
	9	兄弟姉妹	9			
	10	他の親族	戚	10	戚	11: 住み込みの雇人
	11	住み込みの雇人	雇			12: その他も含む
	12	その他	他(男) 媒(女)			
単 独 世 帯	1	世帯主: 男・独身	単	表注: 匿名データに、調査票情報用のシステムをそのまま適用しているため、「戚」には11、12も含まれる。		
		世帯主: 男・既婚	宿			
		世帯主: 女・独身	嬋			
		世帯主: 女・既婚	嬬			

⑤ 匿名データ提供時に世帯情報に含まれている変数のうち、世帯分類に関する3つの変数

このデータセットがあれば、世帯又は世帯員に関する他の変数が必要になれば、世帯 ID 番号、世帯内の世帯員 ID 番号をキーとしてマッチングすれば、容易に入手できる。

6. SAS による家族構成変数の使用方法

我々が構築したSASデータベースに組み込まれた家族構成変数を実際にどのようにして利用について解説する。

国調匿名データ(2000年)から作成した家族構成データセットの最初の10オブザベーションから成るデモ用のSASデータセットを作成した(付録2)。このデモデータを使って、SASプログラムによる家族構成変数の基本的な操作方法を付録3に示した。

家族構成変数がない場合には、分析作業の度に、先に述べた4つの個人属性の変数値を見比べながら、各世帯員間の血縁関係を割り出す必要があり、世帯員数が多い場合は極めて煩わしい作業になる。一方、家族構成変数があれば、漢字が読める人なら、直にその世帯の家族構成が判別できる。つまり、ひとつの家族構成変数の中に、各世帯の全ての世帯員の4つの個人属性をまとめて詰め込んであり、その漢字文字列で表現された変数値を見るだけで各世帯の家族構成が一目で判る仕掛けになっている。繰り返すが、一変数の中に各世帯の全世帯員の続柄が詰め込まれて、しかも、「変数フォーマット」なしに、直接に変数値を見るだけで、漢字が分かる人なら、その変数値の意味が直ぐに理解できる。従って、分析対象としたい世帯構造の世帯を正確に、かつ、容易に抽出できるのが最大の利点である。

更に、もうひとつの利点がある。家族構成変数の変数値は、更に加工することによって、分析したい世帯構造の新しい変数を簡単に作ることができ、分析作業が能率よく推進できる。

なお、現時点では、家族構成変数には漢字(2バイト文字)を使って各続柄を表現しているが、漢字を使わない外国でも、アルファベット(1バイト文字)の2文字を使って、大文字と小文字の組み合わせによって、各世帯員を表現すれば、漢字を使う場合と全く同じシステムで利用できる。例えば、「父」は、「Fa」(「Father」)、「母」は、「Mo」(「Mother)、「倅」(未婚の息子)は、「Ss」(「Son single)としておけば、「父母倅」の世帯は「FaMoSs」と表現でき、2バイト毎に読み取って行けばよい。

7. 収入・支出と就業時間の分析結果

一般的に、世帯全体の収入金額は、世帯主の性別が要因の一つであり、更に、その収入は、その世帯員全員の就業時間の合計にも影響を受けると考えられる。今回、調査年次の近い全消匿名データ(1999年)と国調匿名データ(2000年)を使って、世帯の収入と就業時間について、家族構成変数別に基本統計量を計算した。分析対象の世帯構成の分布を表3に示す。

全消匿名データ: 収入総額(万円)、実収入(以下、円)、経常収入、

世帯主が男の収入、世帯主が女の収入、支出総額、実支出、消費支出

国調匿名データ: 世帯員全員の平均週間就業時間(数値型変数はこの変数のみ)

各世帯の収入は、その世帯員全員の就業時間の合計にかなり左右されることには議論の余地がない。今回、いくつかの家族形態を取り上げて、調査年次の近い全消と国調のデータを、家族構成変数を利用して「統合」することで、家族形態によって特徴があることが分かった。同様の手法をその他の統計調査の匿名データにも適用することで、ひとつの統計調査だけでは見えない各世帯の社会・経済活動の傾向を推し測ることができる手段を得たことになる。

なお、表3の「家族構成」欄に表示されている続柄は、左から順に、「世帯主」、「その配偶者」、「その未婚の子供」だけであるが、この場合の厳密な意味は、少なくともそれらの続柄が揃っている世帯を指す。この表では、それ以外の続柄の世帯員がいる世帯も含まれているが、もちろん、それ以外の世帯員は含まない世帯だけの分類も容易に出来る。

例えば、更に、厳密に典型的な母子家庭(母一人と未成年の未婚の子一人だけ)のデータが欲しい場合には、家族構成変数 **kanjiPat** の変数値が「母俣」、または、「母好」だけを抽出した後、更に、2番目の世帯員の年齢情報から 18 歳未満だけを絞り込めばよい³。

8. 新しい分析手法の確立を目指す

現在、(独)統計センターから提供されている匿名データには、国調、労働力調査、全消、就業構造基本調査、社会生活基本調査、住宅・土地統計調査がある。そのうち、住宅・土地統計調査を除く残り5つの統計調査データには、各世帯員の続柄情報があり、それぞれ若干の定義が異なるものの、先に述べた家族構成変数を作成することが出来るので、世帯構造による世帯の様々な行動パターンの解析が可能になる⁴。従って、本格的なデータ分析を始める前に、予め家族構成変数を作成しておいて、実際の分析段階では、付録3のSASプログラムに例示しているように、家族

表3 分析対象の世帯の分布

全消・匿名データ1999年

世帯の種類	家族構成	度数	注
① 子が同居しない世帯主夫婦の世帯	妻夫	43,809	左端が世帯主
	夫妻	2,549,950	
② 親(世帯主)が未婚の子と同居する世帯	父子	37,816	左端が世帯主。本表の「子」は「俣」または「好」を指す。
	父母子	4,815,830	
	母子	381,200	
	母父子	35,594	
上記以外の世帯	その他	3,372,465	注: 集計用乗率適用
累計度数		11,236,665	

国調・匿名データ2000年

世帯の種類	家族構成	度数	注
① 子が同居しない世帯主夫婦の世帯	妻夫	740	左端が世帯主
	夫妻	95,476	
② 親(世帯主)が未婚の子と同居する世帯	父子	5,390	左端が世帯主。本表の「子」は「俣」または「好」を指す。
	父母子	173,292	
	母子	26,350	
	母父子	602	
上記以外の世帯	その他	180,154	注: 集計用乗率なし
累計度数		482,004	

³ ただし、匿名データでは、年齢変数が各歳ではなくて、5歳刻みの場合が多いので、その場合には、20歳未満で処理するのやむを得ない。いずれ調査票情報を使う場合は、各歳年齢を使って厳密に母子家庭を抽出できる。

⁴ 今回使用した全消・匿名データには、世帯員の婚姻情報(配偶関係)の変数がない。今回は、続柄の日本語表現から婚姻情報の変数を作成したが、世帯外の家族に関する設問から作成された変数が別途あるので、その情報から、例えば、世帯主の女性に、別世帯に「子供がいる」とか、「単身赴任や入院中」等で「夫がいる」ことが判別できるケースもある。元の調査票には、配偶関係の設問があるので、調査票情報のデータには、配偶関係が明示された変数は組み込まれていると思われるので、次の研究段階では、この問題は解消されると思われる。なお、調査時点での「世帯主」の定義が厳密には異なり、また、被調査者の「世帯主」の用語の定義が一定でないことが考えられるので、真の「世帯主」が誰なのかは、かなり恣意的にならざるを得ないが、これはやむを得ない。

構成変数を使って世帯を更に分類しながら分析すると、作業能率が上がり、分析結果の解釈も容易になる。

なお、上記の5つの匿名データに対して、家族構成変数を組み込んだ後、それを使って世帯の様々な社会・経済活動の実態を家族構成によって特徴を把握することを目指している。その次の段階では、同じアルゴリズムを調査票情報に適用して、より正確な分析結果を得ることを計画している。その過程で、家族構成変数を介して、異なる統計調査データを「統合」する手法を探りたい。

9. 今後の研究予定

今回開発した家族構成変数は、世帯の各構成員の続柄を直感的に連想しやすい漢字一文字で表現しているので、分析作業に際して、変数コードの取り違えを防ぐことが出来ると同時に、分析者にも分かりやすいだけでなく、分析結果を読むエンドユーザーにも、説明なしに理解してもらえるという利点がある。第2節でも列挙した 8050(9060)問題、少子高齢化、人口減少、ひとり親世帯、貧困世帯、経済格差などの様々な社会問題に対処していく分析ツールとして、大いに有用であると考えられる。

一例として、ひきこもりの世帯員のいる世帯の問題を取り上げると、「高齢の母親と、現在結婚していない無職の中年の息子の二人世帯で、母親が死亡すると、そのまま放置したり、隠蔽したりすることで生存を偽装して、親の年金を搾取し続ける」というニュースを近年、よく目にする。このケースでは、家族構成変数を使うと、二人の内のどちらが世帯主かによって、「男母」または「母伴」のいずれかで表現できる。その世帯の収入がそれまでは高齢の母親の年金だけだとすると、母親が死亡すると残された中年の無職の子は途端に生活費に窮するという背景があると推測される。このような状況の世帯がどのくらい存在するかは、家族構成変数 KanjiPat の値が「男母」または「母伴」の世帯を観察して、付録4-1・付録4-2でも示した様に、全消の収入に関するいくつかの変数や、国調の就労時間の変数を丹念に見ていけば、全体の傾向が浮かび上がってくると予想している。

その他の社会問題についても、家族構成変数を使って、今後も分析作業を継続して行く。それら一連の実験が終了した後の今後の研究計画は、以下のように予定している。

ステップ1 国調や全消の他に、労働力調査、就業構造基本調査、社会生活基本調査の匿名データも活用して就労状況や日常生活での時間の過ごし方などについて、家族構成変数を使って、様々な世帯構造毎に特徴を捉えることができるシステムを開発して、実態の解明をする。

毎月末に実査される労働力調査以外は、実査周期が5年で、実施年次に重複がないので、当面は、2000年国調以外の統計調査データについては、国調のこの実施年に近い年次のデータを使うことにする。

ステップ2 上記の匿名データの代わりに、記憶媒体で提供される調査票情報入手して、ステップ1で開発済みのプログラムを調査票情報用に一部、改訂して、より正確な分析結果を得る。

ステップ3 更に、オンサイトで、国調の100%の調査票情報を使って、ステップ2で完成したプログラムをほぼそのまま適用して、日本全体の世帯構造とその社会・経済活動の全容を解明する。この

段階では、分析用プログラムの開発は、直前のステップ2でほとんど終わっているの、オンライン利用という閉じられた作業空間でも十分に対応できると考えている。

10. 提案

本稿で紹介している家族構成変数を、個票データから一般のエンドユーザーが作成するのは、かなり負担が大きいと思われる。世帯を調査客体とする統計調査を実施する公的機関が、個票データをエンドユーザーに提供する際、関係省庁間で、全て同じ基準で作成された家族構成変数を事前に組み込んで提供できるようになれば、異なる調査間の結果の比較が容易になり、これまで想像もしなかったような社会の現状が見える可能性がある。せひとも、関係機関にご検討して頂きたい。なお、本稿で述べたように、調査によって、「世帯主」の定義が異なるのは、分析する際に障害となるので、是非、各省庁間で統一できれば、よりいっそう正確な分析が可能となるはずである。

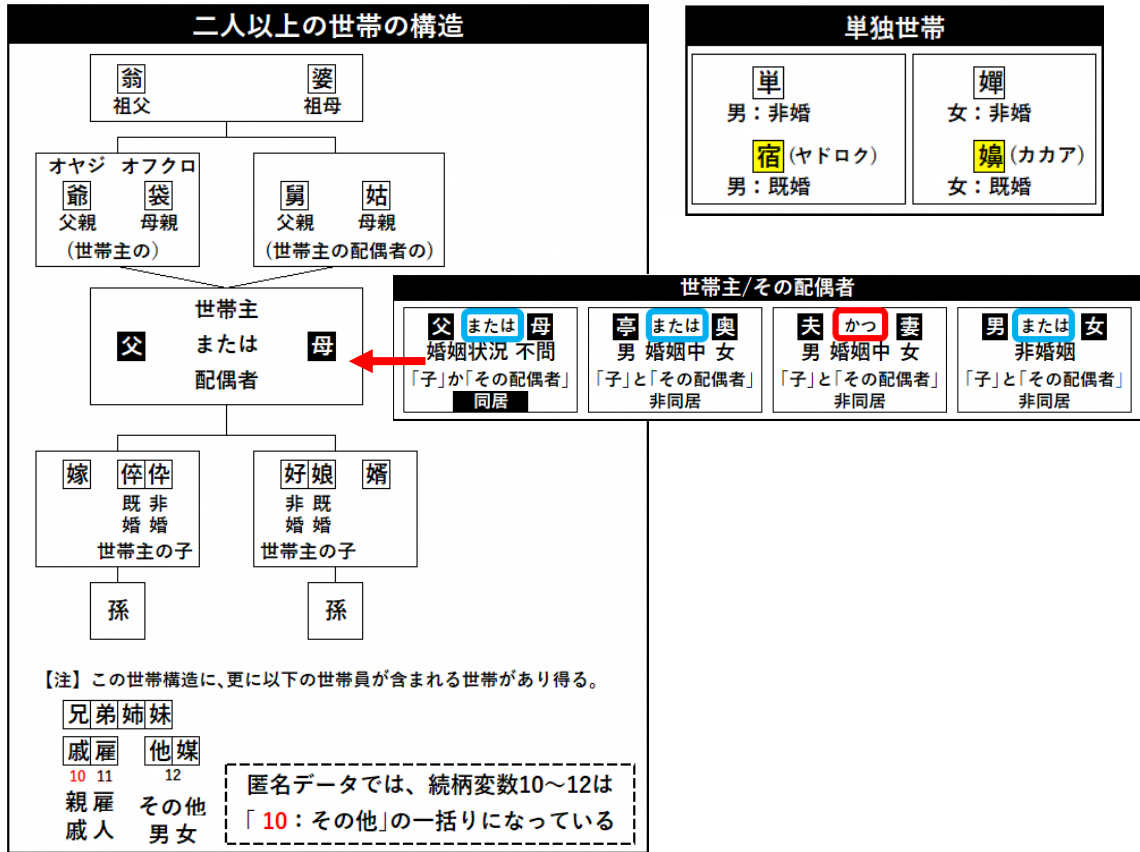
参考文献

- ① 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』、統計情報研究開発センター、2017年
- ② 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
- ③ 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
- ④ 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
- ⑤ 周防節雄(2021)「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」、『官民オープン利活用の動向及び人材育成の取り組み(2020年度)報告要旨集』、独立行政法人 統計センター、pp.155-167、2021年3月
- ⑥ 周防節雄(2022)「国勢調査と全国消費実態調査の匿名データの統計量を家族構成変数で統合する試み」、『ESTRELA 4月号』、(公財)統計情報研究開発センター、2022年4月刊行予定

謝辞

本研究の遂行に際しては、国勢調査と全国消費実態調査の匿名データを使用したが、その際、SAS インステチュート ジャパン(株)から使用料の補助を頂いた。また、両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。

付録1 続柄漢字パターンの世帯内階層構造図(改訂版)



家族構成変数の例：国勢調査調査票情報データ

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)
嬢	女性(既婚)の単独世帯
単	男性(独身)の単独世帯

注：左端が世帯主

付録2 家族構成変数 kanjiPat と shortPat を含む SAS データセット

変数名 obs	家族構成変数		世帯 ID番号	世帯員 数	続柄	性別	年齢階層	婚姻状況	世帯主 歳階層	世帯の 種類	世帯の 家族類型	3世代 世帯
	全体版 kanjiPat	短縮版 shortPat										
			HHidno	HHmNo	tsuzuki	sex	age	marry	headAge	HHKind	familyType	gene3
			1 2 3 4 5 6 7 8	1 2 3 4 5 6 7 8	1 2 3 4 5 6 7 8	1 2 3 4 5 6 7 8	1 2 3 4 5 6 7 8	1 2 3 4 5 6 7 8				
1	夫妻孫	夫妻孫	1	3	1 2 7	2 2	15 15 5	2 2 1	15	1		0
2	父母伴侶	父母伴侶	2	4	1 2 3 3	1 2 1 1	7 7 2 1	2 2 1 1	7	1		0
3	嬬	嬬	3	10	1 10	2	18	3	18	2		
4	単	単	4	1	1	1	6	1	6	1		0
5	嬬	嬬	5	1	1	2	6	1	6	1		0
6	単	単	6	1	1	1	5	1	5	1		0
7	父母伴侶 父母孫	父母伴侶	7	4	1 2 3 8	1 2 1 2	11 11 6 18	2 2 1 3	11	1		0
8	父母孫 父母孫	父母孫	8	5	1 2 3 4 7	1 2 2 1 1	14 13 9 10 4	2 2 2 2 1	14	1		0
9	母好伴	母好伴	9	3	1 3 3	2 2 1	9 4 4	3 1 1	9	1		0
10	夫妻	夫妻	10	2	1 2	1 2	16 16	2 2	16	1		0

1～8の数字は、家族構成変数kanjiPatの漢字の左からの並びに対応している。

国調2000年・匿名データから作成(最初の10オブザベーションだけ表示)

【注】3番目の世帯の世帯員数が「.」(欠損値)は、一般世帯の「対象外」を意味する。
 年齢階層が「18:85歳以上」なので、施設・病院等に入居と推測できる。
 分析の内容によっては、このようなオブザベーションを外してもよい。

付録3 家族構成変数を含む SAS データセットの使用方法

```

/* 国調匿名2000年kanjiPatの使い方例デモプログラム.sas */ options nocenter;
libname koku2000 "E:\国調匿名\匿名sasDS";

*★ 国調匿名データには、集計用乗率の変数はないので、事前にweight=1とした;
data temp; set koku2000.kokutokumei2000sorted_kanjipat; by HHidNo;
*setするデータセットは、各世帯員がobservationのSASデータセットで、
  事前に家族構成変数kanjiPatとshortPatが組み込まれている;
retain totaltime 0; retain HHage;
keep HHidNo kanjiPat shortPat totaltime HHage weight /*var29*/;
label totaltime="全世帯員就業時間計";
label HHage="世帯主年齢階層";
if first.HHidNo then HHage=age;
if first.HHidNo then totaltime=0;
totaltime=sum(totaltime,var29);
if last.HHidNo;
run;

*★デモ用SASデータセットの作成★;
data DEMOdata; set; keep HHidNo kanjiPat shortPat totaltime; if _N_<=10; run;

proc print data=DEMOdata label; title "国調匿名2000年:DEMOdata"; run;

*★目的：父母俣/父母好を含む世帯(子)は結婚していない)を抽出したい★;
*☆step1:父母を含む世帯を抽出☆;
data fmc; set DEMOdata; targetHH=ksubstr(kanjiPat,1,2);
  if targetHH="父母"; drop targetHH; run;
proc print data=fmc label; title "★step1：父母を含む世帯を抽出★"; run;

*☆step2: 更に、「俣」または「好」を含む世帯を抽出☆;
data fmc; set; p1=kindex(kanjiPat,"俣"); p2=kindex(kanjiPat,"好");
  HHtype="fmc"; if p1+p2 > 0; drop p1 p2; run;

proc print data=fmc label;
title "★step2：父母俣/父母好：両親(父親が世帯主)と未婚の子を含む世帯を抽出"; run;

```

プログラム実行結果

国調匿名2000年:DEMOdata

Obs	世帯一連番号	kanjiPat	shortPat	全世帯員就業時間計
1	1	夫妻孫	夫妻孫	50
2	2	父母俣俣	父母俣	54
3	3	嬪	嬪	0
4	4	単	単	50
5	5	嬪	嬪	0
6	6	単	単	50
7	7	父母俣俣	父母俣俣	128
8	8	父母娘婿	父母娘婿孫	200
9	9	母好俣	母好俣	56
10	10	夫妻	夫妻	0

★step1: 父母を含む世帯を抽出★

Obs	世帯一連番号	kanjiPat	shortPat	全世帯員就業時間計
1	2	父母俣俣	父母俣	54
2	7	父母俣俣	父母俣俣	128
3	8	父母娘婿	父母娘婿孫	200

★step2: 父母俣/父母好: 両親(父親が世帯主)と未婚の子を含む世帯を抽出

Obs	世帯一連番号	kanjiPat	shortPat	全世帯員就業時間計	HHtype
1	2	父母俣俣	父母俣	54	fmc
2	7	父母俣俣	父母俣俣	128	fmc

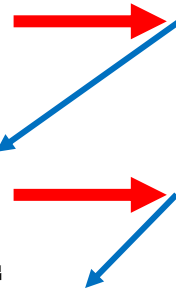
左のデータセットから、変数「kanjiPat」の先頭部に「父母」の漢字文字列がある世帯だけを抽出する方法

SAS関数 KSUBSTRを使う

そのデータセットから、変数「kanjiPat」の中に「俣」(未婚の息子)又は、「好」(未婚の娘)の漢字がある世帯だけを抽出する方法

SAS関数 KINDEXを使う

変数kanjiPatに、2番目の引数の文字が左から数えて、何文字目にあったかを値として返す。その文字がなければ、0を返す。



付録4-1 子が同居していない世帯主夫婦の世帯と、
結婚していない「子」と親(世帯主)が同居する世帯の収支分布

全消・匿名データ 1999 年
収入

var405 年間収入(単位:万円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	648	445	60	2,500	526	347	805	458
夫妻	13,893	626	384	1	2,500	519	370	781	411
父子	204	690	377	60	2,500	628	452	871	419
父母子	25,841	828	407	4	2,500	750	543	1,019	476
母子	1,989	542	376	2	2,500	450	252	741	489
母父子	198	761	426	130	2,391	670	430	1,030	600
その他	6,165	400	300	1	2,500	325	200	500	300
全世帯	48,522	643	413	1	2,500	556	350	849	499

var406 収入総額(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	694,668	586,879	0	6,180,222	626,246	344,508	935,561	591,053
夫妻	13,893	676,562	706,919	0	25,026,145	627,965	328,798	922,929	594,131
父子	204	619,923	515,831	0	4,707,618	656,089	128,822	944,018	815,196
父母子	25,841	899,719	805,539	0	44,670,521	883,819	578,063	1,207,138	629,075
母子	1,989	654,541	664,651	0	16,566,782	550,748	328,041	848,820	520,779
母父子	198	914,605	853,151	0	6,852,533	730,607	478,158	1,112,170	634,012
その他	6,165	477,835	389,691	0	7,262,307	432,433	258,601	616,280	357,679
全世帯	48,522	712,446	699,095	0	44,670,521	645,008	332,244	991,912	659,668

var407 実収入(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	318,161	259,667	0	1,374,525	287,801	136,610	422,030	285,420
夫妻	13,893	293,775	285,836	0	15,859,979	247,004	97,650	438,586	340,936
父子	204	322,698	276,775	0	1,305,667	308,147	1,667	524,546	522,880
父母子	25,841	458,193	351,914	0	22,324,786	454,500	289,743	627,419	337,676
母子	1,989	281,552	223,148	0	1,484,515	238,278	129,438	404,560	275,122
母父子	198	441,122	380,415	0	2,823,911	388,305	230,050	568,912	338,862
その他	6,165	230,018	186,942	0	3,529,207	215,000	105,983	313,860	207,877
全世帯	48,522	345,351	308,221	0	22,324,786	309,741	136,167	506,516	370,350

全消・匿名データ 1999 年：収入(続く)

var408 経常収入(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	301,731	240,874	0	1,157,537	261,435	129,213	406,672	277,459
夫妻	13,893	283,547	252,308	0	3,524,000	235,650	88,500	428,856	340,356
父子	204	316,309	270,147	0	1,120,909	308,147	0	514,210	514,210
父母子	25,841	445,984	302,737	0	3,088,000	446,847	282,183	615,342	333,159
母子	1,989	273,416	218,560	0	1,482,115	226,840	122,423	397,436	275,013
母父子	198	427,429	376,841	0	2,822,731	377,925	213,753	560,142	346,388
その他	6,165	224,559	183,154	0	3,529,207	209,675	100,000	307,844	207,844
全世帯	48,522	335,753	275,748	0	3,529,207	302,493	129,912	497,426	367,513

var411 世帯主が男の収入(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	0	0	0	0	0	0	0	0
夫妻	13,893	158,253	224,943	0	1,958,200	0	0	314,500	314,500
父子	204	201,186	239,880	0	982,790	100,000	0	357,044	357,044
父母子	25,841	335,941	238,420	0	3,088,000	365,207	157,000	486,940	329,940
母子	1,989	0	0	0	0	0	0	0	0
母父子	198	391	3,741	0	46,133	0	0	0	0
その他	6,165	112,791	178,244	0	3,529,207	0	0	240,000	240,000
全世帯	48,522	214,421	240,007	0	3,529,207	171,812	0	395,958	395,958

var414 世帯主が女の収入(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	191,251	207,476	0	1,144,367	153,945	0	255,144	255,144
夫妻	13,893	0	0	0	0	0	0	0	0
父子	204	0	0	0	0	0	0	0	0
父母子	25,841	5	365	0	29,480	0	0	0	0
母子	1,989	107,767	144,913	0	1,253,400	63,883	0	166,833	166,833
母父子	198	219,451	247,199	0	2,002,731	176,484	83,863	291,000	207,137
その他	6,165	47,849	110,009	0	855,488	0	0	0	0
全世帯	48,522	19,460	75,522	0	2,002,731	0	0	0	0

【注】

上記の「世帯主が男の収入」は、「世帯主が男の世帯の収入」を意味しているため、家族構成が「母父子」の世帯では、この変数は非該当のはずなので、斜線を引いておいた。
匿名データ化の処理の際に誤って紛れ込んだのかもしれない。
「世帯主が女の収入」の場合も同様。なお、家族構成が「その他」と「全世帯」の場合でも、同様の矛盾が含まれているケースがあるので、斜線を引いておいた。

全消・匿名データ 1999 年：支出

var457 支出総額(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	791,775	636,234	64,764	6,180,222	643,493	428,534	950,434	521,900
夫妻	13,893	739,841	664,543	34,210	25,026,145	642,338	405,765	930,472	524,707
父子	204	705,087	485,839	55,419	4,707,618	668,373	320,000	961,998	641,998
父母子	25,841	961,929	750,703	33,844	44,670,521	888,070	605,160	1,209,704	604,544
母子	1,989	685,375	644,545	51,000	16,566,782	564,408	349,725	852,948	503,222
母父子	198	946,159	823,981	69,740	6,852,533	730,607	478,603	1,112,170	633,567
その他	6,165	512,039	367,602	21,238	7,262,307	445,181	287,649	621,675	334,026
全世帯	48,522	765,545	661,228	21,238	44,670,521	658,111	392,325	996,622	604,297

全消・匿名データ 1999 年：支出（続き）

var458 実支出(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	359,882	273,824	64,764	2,061,634	292,361	199,153	415,092	215,939
夫妻	13,893	343,639	228,925	20,826	4,017,930	290,953	204,454	413,244	208,790
父子	204	323,736	231,002	55,419	2,718,752	284,424	198,150	392,536	194,386
父母子	25,841	437,332	236,199	29,204	4,396,320	387,066	285,320	527,220	241,900
母子	1,989	307,586	215,725	32,041	2,567,200	254,376	171,363	375,685	204,322
母父子	198	379,791	222,361	69,740	1,235,913	313,684	222,912	464,572	241,660
その他	6,165	236,556	177,029	15,810	3,179,466	200,166	133,520	283,693	150,173
全世帯	48,522	350,543	233,716	15,810	4,396,320	298,818	199,800	438,660	238,861

var459 消費支出(単位:円)									
家族構成	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	232	316,525	256,753	51,746	1,965,475	252,952	178,639	335,856	157,217
夫妻	13,893	298,710	203,403	20,226	4,017,930	251,249	182,139	350,987	168,849
父子	204	272,148	220,407	55,419	2,718,752	233,340	165,104	315,158	150,054
父母子	25,841	360,133	203,339	29,204	4,291,040	314,128	236,230	426,700	190,470
母子	1,989	278,022	200,696	32,041	2,344,592	226,591	156,229	339,922	183,693
母父子	198	308,113	167,194	69,740	1,039,086	266,815	194,072	372,273	178,201
その他	6,165	203,946	162,946	15,810	3,179,466	167,111	115,664	238,981	123,317
全世帯	48,522	295,901	203,166	15,810	4,291,040	251,681	170,658	361,654	190,996

付録4-2 子が同居していない世帯主夫婦の世帯と、 結婚していない「子」と親(世帯主)が同居する世帯の就業時間の分布

国調・匿名データ 2000 年 週間就業時間(唯一の数値データ)

分析変数：VAR29 平均週間就業時間										
家族構成	Obs 数	有効件数	平均	標準偏差	最小値	最大値	中央値	下側四分位点	上側四分位点	四分位範囲
妻夫	740	740	54	41	0	195	52	11	88	77
夫妻	95,476	95,476	43	39	0	326	40	0	71	71
父子	5,390	5,390	66	43	0	310	60	40	95	55
父母子	173,292	173,292	82	44	0	438	74	50	104	54
母子	26,350	26,350	49	39	0	330	42	24	73	49
母父子	602	602	77	44	0	368	74	48	98	50
その他	180,154	180,154	33	39	0	416	32	0	49	49
全世帯	482,004	482,004	54	47	0	438	48	7	80	73

注: このデータには、集計用乗率の変数はない。

付録5 子が同居していない世帯主夫婦の世帯と、
結婚していない「子」と親(世帯主)が同居する世帯の
収入・支出の箱ひげ図：全消・匿名データ 1999 年

