

「全国消費実態調査」匿名データの整理概要
— kuma_do 及び変数対応表の活用方法を中心に —

二木 美苗

1.はじめに

本稿は、神戸大学マイクロアーカイブ研究会において作成した kuma_do (計量分析ソフトウェア用のプログラムファイル) と変数対応表 (Excel ファイル) の説明を中心に、総務省「全国消費実態調査」匿名データ利用時のデータ整理の問題点とその対処方法を記述する。

神戸大学マイクロアーカイブ (KUMA) は、独立行政法人統計センターの西日本で最初のサテライト機関である。平成 22 年から、協定¹に基づき「匿名データ」や「オーダーメイド集計」のサービスを提供し、マイクロデータを用いた学術研究の支援・促進・普及に努めた活動を行っている²。現在、KUMA では、総務省統計局所管の「全国消費実態調査」、「社会基本調査」、「就業構造基本調査」、「住宅・土地統計調査」、「労働力調査」、「国勢調査」の匿名データが利用可能である。本稿で用いた「全国消費実態調査」の匿名データ (1989,1994,1999,2004 年) も KUMA を通してデータの提供を受けている。

「匿名データの提供」は、独立行政法人統計センターにおいて平成 21 年より開始されている³。これは、平成 19 年に統計法が抜本的に改正されたことを受けた取り組みである。この改正によって、学術研究や高等教育の分野における公的統計の活用を図るために、公的統計マイクロデータの利用が広く一般に可能となった。近年多くの学術研究で、マイクロデータを用いた分析の成果が発表されており、研究分野を超えて、個票データを用いた詳細な分析の重要性が高まっている⁴。その一方で、マイクロデータの利用時には、データの整理、加工を分析者自身が行う必要があり、分析に利用できる段階に至るまでに多大な時間を費やすことも多い。また、データの整理、加工には適切な知識が必要であり、アドバイザーが周囲にいない時、データ分析の初学者にとっては、マイクロデータの利点を分析に活かすことが難しい場合がある。

そこで、KUMA では、匿名データの研究利用を促進するために、研究会を通してデータの整理や分析利用における知識を共有する試みを実施している。この取り組みの一環として、本稿で説明する「全国消費実態調査」の匿名データを整理する Stata-dofile(kuma_do) と変数対応表を作成した。本稿の主な目的は、これらのファイルについての詳細を記述し活用を促すことで、匿名データ利用者のデータ整理の簡便化を図ることである。

・ 神戸大学大学院経済学研究科 博士後期課程

¹ 「国立大学法人神戸大学と独立行政法人統計センターとの間における連携協力に関する基本協定書」(平成 21 年 11 月 13 日締結)

² 詳しい活動内容は以下の HP を参照されたい。 <http://www.econ.kobe-u.ac.jp/kuma/index.html> (2014/1/13 閲覧)

³ <http://www.nstac.go.jp/services/archives.html> (2014/1/13 閲覧)

⁴ 公的統計マイクロデータ利用に関する現状については、小林 (2012) を参照されたい。

本稿の内容は以下の通りである。次章で「全国消費実態調査」匿名データの概要を説明し、匿名データ利用時の問題点をまとめる。3章では匿名データ整理用のテキストファイル（kuma_do）と Excel ファイル（変数対応表）の利用方法について記述し、4章では、各ファイルによって整理したデータ分析の応用例として集計値の作成について取り上げる。

2. 「全国消費実態調査」匿名データについて

「全国消費実態調査」は、『国民生活の実態について、家計の収支及び貯蓄・負債・耐久消費財・住宅・宅地などの家計資産を総合的に調査し、消費・所得・資産にかかる水準・構造・分布などを明らかにする』ことを目的とし、昭和34年の第1回調査以来5年ごとに実施されている。調査は、二人以上の世帯(一般世帯)と単身世帯とに分けられて選定され、各年での標本数は一般約55,000世帯、単身世帯約4,000~5,000世帯である。それぞれの調査期間は、一般世帯は9・10・11月の3か月間、単身世帯は10・11月の2か月間である。基本的な調査事項は、調査年によって若干の変更はあるものの、以下の通りにまとめられる。①家計上の収入と支出に関する事項、②品物の購入先に関する事項、③主要耐久消費財等に関する事項、④年間収入及び貯蓄・借入金残高に関する事項、⑤世帯及び世帯員に関する事項、⑥現住居及び現住居以外の住宅・宅地に関する事項である。匿名データには上記6項目のうち「品物の購入先に関する事項」を除いた5項目の内容が一括して含まれている。これらのデータの詳しい内容は、KUMA ホームページ「匿名データの利用」頁の「全国消費実態調査」符号表によって確認できる。各年の符号表のレイアウトフォームに示される匿名データの内容は、①調査票、②耐久財、③用途、④品目(1)(2)、⑤家計資産(1)(2)という分類で示されている。

データ利用時には、調査報告書と統計局ホームページを参照して調査方法や標本設計の方法を十分に理解することが必要である。例えば、各データには集計用乗率が付与されていて、平均値を求める時には乗率を用いて加重する必要がある。また、データは匿名化措置が施してあるため、地域は「3大都市圏」と「その他」の2地域に分けられているのみで都道府県を特定することはできない。更にリサンプリングやトップコーディングによる匿名化処理がなされていること、外れ値補正がなされていない点もデータの特徴として知っておくべきである。その他に、消費支出が「品目」と「用途」の2体系の分類で集計されている点にも注意が必要である。「品目」は世帯が購入した物品を同一商品は同一項目で分類している。しかし、「用途」では世帯が購入した物のうち、世帯員以外の人のための贈答又は接待を目的とした購入物は「交際費」で分類し、その他の物を「品目」で分類している。ここから、「用途」と「品目」データの消費支出が総合計以外では一致しない場合が出てくる。これらのデータ利用時の注意点をよく理解した上で、実際のデータ整理段階へと進む。

データ整理の段階では、以下の3点が大きな問題となる。第一の問題は、元データの容量が非常に大きい点である。上記で説明したように、各年のデータは一般世帯・単身世帯

に分けられており、それぞれのテキストファイル（csv形式）で提供される。しかし、各データに全ての調査項目内容が含まれているため、ファイルの容量は大きく、データの読み込みにかかなりの時間を要する。当然、毎回の作業で全てのデータを読み込むことは効率的ではない。ここで、作業時に素早くデータを取り出せるように、データを適切に分割して保存する必要が生じる。またその一方で、分析に応じて分割データを簡単に合わせる事が可能な設計も求められる。

第二の問題は、データ読み込み時の表示に関する問題である。匿名データの各変数の内容は主に数値化されて保存されているが、一部データは X,Y といった文字列のまま処理されている。そのため、元データを、ソフトウェアを用いて分析するには、それらの文字列データを数値に変換する必要がある。

最後の問題は、データの変数が全て v1,v2,・・・という数字のみで記されている点である。例えば、v1563 という変数が何を示しているかを知るためにはレイアウトフォームを見て確認する必要がある。しかし、提供される符号表にはデータと一致した変数名は記されていない。そのため、各変数の詳細を知るためには、利用者が符号表に一つずつ変数と対応する数字を当て、新たに変数名を設定するという煩雑な作業が必要である。また、数値と変数が、割り当て時のミスによってずれてしまう場合も考えられるので、変数の示す内容と実際のデータの中身が整合的であるかを確認する必要もある。更に、変数の番号は各年で共通でないため、複数年のデータを合わせて用いる場合は、複数年共通の変数名を設定しなければならない。

以上の問題は、「全国消費実態調査」匿名データの利用者に対して等しく生じる問題である。そのため、データ整理作業のシステムを広く公開し、作業過程のプログラムを共有することで、分析の第一段階を効率的に進めることが可能である。KUMA では、匿名データの利用促進を図るために、上記作業を簡単に実行できる Stata（統計分析ソフト）のプログラムファイル（以下、dofile と呼ぶ）と Excel の変数対応表を作成し、データ利用者に提供している。

3. kuma_do 及び変数対応表

今回 KUMA で提供する「全国消費実態調査」の匿名データ整理用ファイル（kuma_do、変数対応表）は全て、統計分析ソフト Stata での分析を想定して作成している⁵。Stata ではデータベースの作成から推計・分析結果のとりまとめまでを一貫して作業することが可能で、複雑で大容量になることの多いマイクロデータの分析に適している⁶。また、Stata の命令（コマンド）を作業順に書いたテキスト形式のプログラムファイル（dofile）を作成することで、一度行った作業を何度も再現することが可能である。即ち、同じ匿名データを利用する場合、整理用プログラムの dofile を残すことで、利用者全員が匿名データの整理

⁵ 他のソフトを使用する場合は、kuma_do(テキストファイル)を使用ソフトに合わせた形へ変換する必要がある。

⁶ Stata の利用方法については、松浦（2010）筒井他（2011）を参照されたい。

を同じ手順で実行することができる。

3.1. kuma_do

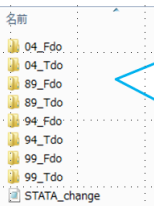
kuma_do は、KUMA 研究会において作成した Stata データ整理用 dofile を一つのフォルダにまとめたものである。kuma_do には、1989 年、1994 年、1999 年、2004 年「全国消費実態調査」匿名データの一般世帯 (F)・単身世帯 (T) のデータを整理するために必要な dofile フォルダ (89_Fdo~04_Tdo) と元データの読み込みを指示する dofile (STATA_change) が含まれている。(図 1 参照)

また、各年のフォルダの中には、データの名前付け用 dofile をまとめた rename フォルダがあり、その他に 8 つの分割データ (調査、世帯、耐久財、用途、品目 1、品目 2、資産 1、資産 2) を作成するために必要な dofile と、それらを一括して実行する dofile(make_all) が含まれている。(図 2 参照)

図 1

kuma_do内容

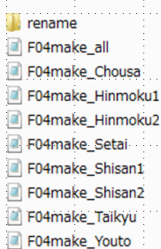
- ▶ kuma_do : フォルダを開くと以下のフォルダとdofileが表示される。



- ▶ STATA_change (dofile)
オリジナルデータをSTATA形式に変換する。
- ▶ フォルダ : [89_Fdo] - [04_Tdo]
数字は各年を示し (2004年は04)、Fは世帯データ、Tは単身データを示す。

図 2

- ▶ 各フォルダの内容 (04_Fdoフォルダの場合)



- ▶ renameフォルダ
各データの名前付けをするdofileをまとめたもの (後で説明)
 - ▶ F04make_all
以下の全てのdoファイルをまわし、8個のデータを作成する。
 - ▶ F04make_Chousa - 調査データを作成
 - ▶ F04make_Setai - 世帯データを作成
 - ▶ F04make-Taikyū - 耐久財データを作成
 - ▶ F04make_Youto - 用途データを作成
 - ▶ F04make_Hinmoku1 - 品目1データを作成
 - ▶ F04make_Hinmoku2 - 品目2データを作成
 - ▶ F04make_Shisan1 - 資産1データを作成
 - ▶ F04make_Shisan2 - 資産2データを作成
- * 1989年は資産データがないため、doの数は2個少ない

分割された各データは、調査年 (v3) とレコード一連番号(v4)を用いることで、分析に応じて自由に組み合わせることが可能である。また、それぞれのデータに含まれる主な内容は、①世帯情報の基本事項 (Chousa)、②世帯員に関する情報 (Setai)、③各耐久財の所有総数、取得時期 (Taikyū)、④世帯情報、世帯収入、世帯支出 (Youto)、⑤購入方法が現金による消費支出の内訳 (Hinmoku1)、⑥購入方法が月賦掛買いによる消費支出の内訳 (Hinmoku2)、⑦総資産である家計資産の数量と資産額 (Shisan1)、⑧純資産である家計資産の数量と資産額 (Shisan2) である。元データをこのように分割することで、まずデータの容量が大きいうという最初の問題を解決することができる。

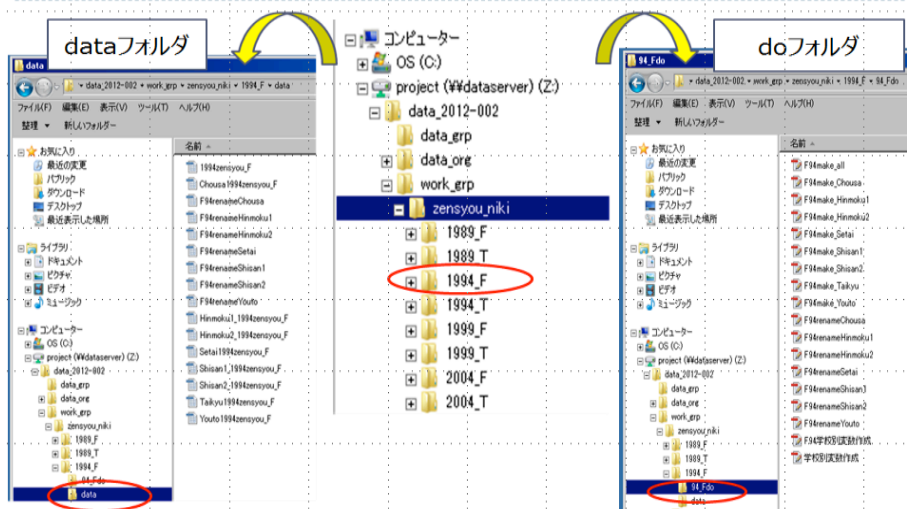
次に二番目の問題であるデータの文字列表記を数値に直す作業について説明する。以下は 2004 年世帯データ作成用 dofile の一部抜粋である。

```
*make setai data*
clear
set more off
cd Z:\data_2012-002\work_grp\zensyou_niki\2004_F\data ①
use 2004zensyou_F.dta
*setaiin*
forvalues num=1/6{
    local com=54+(\num'-1)*14
    *change Tuzukigara*
    replace v`com'="10" if v`com'=="X"
    replace v`com'="11" if v`com'=="Y" } ②
}
destring v39-v136,replace ③
keep v3 v4 v39-v136 ④
save Setai2004zensyou_F.dta,replace
```

下線部①は、使用するフォルダを指定するコマンドである。利用者は、この部分を各自が使用するデータのファイルフォルダに合わせて書き換えることによって、kuma_do の全ての dofile が実行可能となる。kuma_do のデータ設定はオリジナルデータと別に work_grp というフォルダを作成し、その下に zensyou_niki、2004_F、data という順でデータの階層を設定している。そしてこの data フォルダの中に、作成した各年の Stata データが保存される。図 4 は、パソコン画面表示を用いてデータの階層設定を示している。

図 4

データの階層の構造



下線部②の 2 列は、X,Y という文字列でデータ入力されている値をそれぞれ 10, 11 という値に変換する作業を命令している。(各 X,Y が何を指すかは、分析時に符号表で確認する必要がある。) その後、下線部③の `destring` コマンドで変数全てを数値に置き換えている。

また、下線部④では、データを分割する際に残す変数を指定している。この時、`v3,v4` は調査年とレコード一連番号を表し、どの分割データにおいても含まれるように設定している。この 2 つの変数を用いることで、分割データを分析に応じて自由に組み合わせ、データを編集することが可能である。以上から、分割データを作成する `dofile` を実行することで、第 2 の問題点も自動的に解決できる。

最後に残った問題は、データの変数名を複数年で共通の適切な名前に変換する作業である。この複雑な作業に対しては、`dofile` とともに変数対応表という Excel ファイルを作成した。次節でこのファイルの活用方法を詳しく説明する。

3.2 変数対応表

Excel ファイルによる変数対応表とは、「全国消費実態調査」各年の変数と符号表を対応させ一覧とした表である。先に問題点として挙げたように匿名データの各年の元変数は、`v1,v2,...` で表示されるが、年ごとに変数と対応する数値が異なっている。また、各変数の数値が何を指すかは、各年の符号表で 1 つずつ確認し、その上で各自が設定する名前をつける必要がある。この問題に対し、変数対応表を利用すると、簡単に元変数を設定したい名前に変換できる。そして、各年度の変数名を共通にすることで複数年のデータを合わせる (`merge`) ことも可能である。以下では、具体的に変数対応表の活用方法を説明していく。

図 5 は、品目に関する変数対応表の一部抜粋である。左側の品目項目は、データの符

号表のレイアウトフォームを階層ごとに可視化して整理している。そして、code という列には項目分類表（全国消費実態調査報告書に掲載）のコード番号を対応させている。また、各年の列には、対応する変数の数値(v1~)を入力し、数値のない年は斜線を記して整理している。例えば、パンの支出に対する変数として、2004年は、v678、1999年は、v676、1994年は、v775、1989年は、v1012 が対応していることが示されている。

図 5

		code	2004	1999	1994	1989	大分類	中分類	小分類	
品目	購入方法	hl	v673	v671	v768	v1005				
	消費支出	hl	Lexpend	v674	v672	v769	v1006			
	食料	hl	food	v675	v673	v770	v1007	1	0	
		穀類	hl	cireal	v676	v674	v771	v1008	0	1
		米類	hl	102	v677	v675	v772	v1009	0	0
		うるち米	hl	102_1			v773	v1010	0	0
		他の米	hl	109			v774	v1011	0	0
		パン	hl	120	v678	v676	v775	v1012	0	0
		めん類	hl	130	v679	v677	v776	v1013	0	0
		他の穀類	hl	140	v680	v678	v777	v1014	0	0
		魚介類	hl	fish	v681	v679	v778	v1015	0	1
		生鮮魚介	hl	170	v682	v680	v779	v1016	0	0
		塩干魚介	hl	195	v683	v681	v780	v1017	0	0
		魚肉練製品	hl	203	v684	v682	v781	v1018	0	0
		他の魚介加工品	hl	213	v685	v683	v782	v1019	0	0
		肉類	hl	meat	v686	v684	v783	v1020	0	1
		生鮮肉	hl	freshmeat	v687	v685	v784	v1021	0	0
		牛肉	hl	220	v688		v785	v1022	0	0
		豚肉	hl	221	v689		v786	v1023	0	0
		鶏肉	hl	222	v690		v787	v1024	0	0
		合いびき肉	hl	22X	v691		v788	v1025	0	0
		他の生鮮肉	hl	224	v692		v789	v1026	0	0
		加工肉	hl	meatproducts	v693	v686	v790	v1027	0	0
		ハム・ソーセージ(89 ハム)	hl	225	v694	v687	v791	v1028	0	0
		ソーセージ	hl	226				v1029	0	0
		他の加工肉	hl	229	v695	v688	v792	v1030	0	0

また、大・中・小分類の欄には0と1の数値を入れている。この分類フラグは、Excel ツールバーのフィルター機能を用いて、分類ごとの一覧表を作成する際に利用できる。図 6 に示す手順にしたがってフィルターをかけることで、図 7 のように大分類（大分類フラグが1の列）のみの表が作成可能である。「全国消費実態調査」には、多くの項目に関するデータが含まれているため、分類フラグは、関心のある項目にデータを絞りたい時に有効に機能する。各自が分析に必要な分類フラグを変数対応表に新たに加えることで、データ整理の効率化がよりいっそう進むと考えられる。

図 6

＜分類フラグの活用法＞

ツールバーのデータを選択し、整理したい列をクリックして指定した後にフィルターをクリック、赤い囲みで示す印が表示される。

1999	1994	1989	大分類	中分類	小分類
71	v768	v1005			
72	v769	v1006			
73	v770	v1007	1	0	
74	v771	v1008	0	1	
75	v772	v1009	0	0	
	v773	v1010	0	0	

赤枠内の印をクリックすると右のようなBOXが表示される。
今回は0, 1で分類を分けているため、1の欄のみにチェックを入れてOKをクリックする。

図 7

h1	food	v675	v673	v770	v1007	1
h1	housing	v783	v774	v872	v1110	1
h1	fuel_light	v794	v785	v883	v1119	1
h1	furniture	v803	v794	v893	v1128	1
h1	cloth_footw	v850	v842	v941	v1177	1
h1	medicalcare	v905	v896	v995	v1231	1
h1	transp_com	v921	v911	v1009	v1244	1
h1	education	v956	v945	v1041	v1277	1
h1	recreation	v974	v963	v1058	v1294	1
h1	other_living_e	v1037	v1028	v1118	v1355	1

大分類のみのデータを表示した場合

次に変数対応表を用いた名付け dofile (rename-do) の作成について説明する。作業の第一段階として、変数対応表を用いて貼り付け用のデータを作成する。図 8 は、貼り付け用データ作成に関する変数対応表の例として世帯データの一部を抜粋している。

図 8

	名前	2004	1999	1994	1989		2004	1999
世帯員番号	Setaiinn1	v39	v43	v40	v36		rename v39 Setaiinn1	rename v43 Setaiinn1
続柄	S1_tudukigara	v40	v44	v41	v37		rename v40 S1_tudukigara	rename v44 S1_tudukigara
性別	S1_sex	v41	v45	v42	v38		rename v41 S1_sex	rename v45 S1_sex
年齢5歳階級	S1_age	v42	v46	v43	v39		rename v42 S1_age	rename v46 S1_age
年齢各歳(15歳未満のみ)	S1_u15	v43	v47	v44	v40		rename v43 S1_u15	rename v47 S1_u15
就業・非就業の別	S1_work	v44	v48	v45	v41		rename v44 S1_work	rename v48 S1_work
企業区分	S1_csort	v45	v49	v46			rename v45 S1_csort	rename v49 S1_csort
企業規模	S1_csize	v46	v50	v47	v42		rename v46 S1_csize	rename v50 S1_csize
勤務形態	S1_wform		v51	v48	v43			rename v51 S1_wform
本業以外の勤め先	S1_secondw				v44			

図 8 の左側部分の表では、名前の列の赤枠内に Setaiinn1 という名前が入力されている。そして右側部分の 2004, 1999 の列には rename v39(43) Setaiinn1 という文字が 1 枠の中に収まっている。これらの列には Excel の CONCATENATE という関数を用いている。この関数は、指定した複数のセルに入力されている文字列をつなげた文字列で返す関数である。例えば 1 行目であれば、同じ行の 2004 (1999) 年の列の数値 v39 (v43) と名前列の文字 Setaiinn1 を rename という文字の後に続けて入れるように指定している。この関数を用いると、赤枠内の名前を変換するだけで右の青枠内の部分も一斉に変換することができる。即ちこの変換機能によって、kuma_do で設定した名前を簡単に各自が設定する名前に変換することが可能である。

図 9 は、右の 2004 の列をコピーして、dofile に貼り付けた rename-do の一部である。図内の青線でくくっている部分が、Excel で作成した貼り付けデータである。各 rename_do の貼り付け部分を各自が設定したい名前に変換することで、個人仕様の名付け dofile が作成できる。

図 9

<rename-doの作成>

```

▶ clear
▶ set more off
▶ cd ¥¥DATASERVER¥user¥niki_2012-002¥デスクトップzensyou_niki¥2004_F¥data
▶ use Setai2004zensyou_F.dta

▶ *rename Setai2004F*
▶ rename v3 year
▶ rename v4 id
▶ rename v39 Setaiinn1
▶ rename v40 S1_tudukigara
▶ rename v41 S1_sex
▶ rename v42 S1_age

```

この部分にExcelで作成した新しい名前を張り付ける。
→kuma_doから個人仕様のdofileの作成が可能

(ただし、dofileにおいてloopを使って変数を作成したい場合には、v1,v2の形のままの方が作業しやすい場合もある。)

最後に 1994 年耐久データに関する注意点を記す。耐久データは、基本的には品目と変数が 1 対 1 に対応している。例えば、2004 年 v137 は全て和ダンスを示す番号である 1 をとり、続く質問の回答は、全世帯において和ダンスの所持に関する回答である。ところが、1994 年の耐久データでは同一変数の中に様々な品目が含まれている。例えば、v147 には、ある世帯は和ダンスを示す 1 が入っているが、またある世帯ではエアコンを示す 27 という数値が入っている。そのため、世帯ごとの番号から v147 の品目名を特定していかなければならない。このようなデータの設定から、図 10 に示すように 1994 年の耐久データについての変数対応表は、変数の数字ではなく、○印（品目設定あり）と品目番号を記入している。そのため、上記で説明した貼り付けデータの作成は 1994 年耐久財に限っては、変数対応表で行うことができない。この点には、十分に注意が必要である。

1994 年耐久財に関する名前付け dofile は、図 11 に示す通りの内容で作成している。具体的には、繰り返しの loop をかけることでそれぞれの番号と変数名を 1 つずつ対応させ、その後で各番号の耐久財に対応した名前付けを実行している。

図 10

1994年に品目設定がある場合は○印と品目番号を記入

	名前	2004	1999	1994	1989		
取得時期を調査した品目(1)	品目番号	Taikyu1	v137	v148	○1	v141	*1989は所有時期が年で設定
和だんす(取り付けを除く)	所有総数	T1_ownN	v138	v149	○	v142	
	過去1年	T1_u1	v139	v150	○	v149	
	過去1~5年	T1_u5	v140	v151	x	v145	v146 v147 v148
	過去5年超	T1_ov5	v141	v152	x	v143	v144

図 11

▶ F94renameTaikyu_do内容

```

forvalues num =1/60 {
  ** 番号指定 **
  local bangou=147+( `num'-1)*4
  local own=`bangou'+1
  local new=`bangou'+2
  local kaikae=`bangou'+3

  replace v`new`=0
  replace v`kaikae`=0

  ** rename 和だんす (取り付けを除く)
  replace Taikyu1 =1 if v`bangou'== 1
  replace T1_ownN =v`own' if v`bangou'== 1
  replace T1_u1 = v`new'+v`kaikae' if v`bangou'== 1

```

Loopをかけて、それぞれの番号と変数名を対応させた名づけを指示している。

<変数を数値で残す利点>
 dofileの作成時に、変数をv1.v2という数値にしておくことloopを使ったプログラムを作成することができる。

4. kuma_do データによる集計値の作成

最後に本稿のしめくくりとして、kuma_do で作成したデータの簡単な分析応用例を取り上げる。「全国消費実態調査」の匿名データ借入時には「参考 1 主な項目についての匿名データによる集計結果（一例）及び公表値」と「参考 2 全国消費実態調査匿名データの集計方法の一例について」という参考資料が提供される。参考 1 には匿名データを用いた主な項目の集計値と公表値を比較した表が掲載され、参考 2 には具体的な項目番号と計算式が記されている。この表に示される集計値を kuma_do データで再現する。

ここで、集計値を再現する推計用 dofile は、Shuukei_do というファイル名で設定しており、kuma_do と共に匿名データ利用者に提供している。フォルダの中には各年の集計値の再現を実行する dofile (Shuukei 1989 F&T~Shuukei2004 F&T) が入っている。以下に 2004 年の集計値作成 dofile の内容（一部抜粋）とその推計結果を示す。

** 平成 16 年(2004 年) 匿名データ集計結果表 **

** 普通世帯 **

```
use F04renameChousa.dta
merge 1:1 year id using F04renameYouto.dta
drop _merge
merge 1:1 year id using F04renameTaikyu.dta
drop _merge
merge 1:1 year id using F04renameShisan2.dta
drop _merge
```

①

*** 2 人以上世帯 世帯数分布(抽出率調整) ***

** 世帯数分布 **

* 全世帯:bunpu *

```
gen bunpuS=(y2 /y1)/10
egen bunpuS1 = total(bunpuS)
sum bunpuS1
```

②

Variable	Obs	Mean	Std. Dev.	Min	Max
bunpuS1	43861	873231.9	0	873231.9	873231.9

< 中略 >

*** 自動車保有台数 : y13 ***

* 全世帯 *

svyset [pweight = bunpu1]

svy: mean y13

} ③

	Linearized			
	Mean	Std. Err.	[95% Conf. Interval]	
y13	1.438362	.0053028	1.427968	1.448755

まず①では、分割データを **merge** コマンドで合併し、分析に必要なデータを作成している。そして②は、参考資料 2 に記されている平均値等の計算式を実行するコマンドである。また、③は、計算式で作成したウェイトを用いて自動車保有台数の平均値を推計している。②③のコマンドの下の各表は、それぞれのコマンドを実行した時に **Stata Results** 画面上に表示される結果表である。この表の **Mean** で示される数値が、匿名データを用いて計算した集計値を示している。

図 12 は、2004 年参考 1 資料の抜粋であるが、表に記載される全世帯列の世帯数分布 873,232 と自動車保有台数 1,438 の値がそれぞれ、上記の推計結果表の数値と一致していることが確かめられる。即ち、推計値の再現 **dofile** を用いることで、データ内容を確認するという分析の第一歩が開始できる。

図 12

1. 二人以上の世帯

集計項目	匿名データ (乗率を合算した場合)		
	全世帯	勤労者世帯	備考
○世帯数分布 (抽出率調整)	873,232	512,312	全コードの (乗率/調査月数) の和
①自動車保有台数 (1000世帯当たり)	1,438	1,524	
②消費支出 (円)	318,867	338,552	
③年間収入 (千円)	6,819	7,376	
④貯蓄現在高 (千円)	15,360	12,146	
世帯数分布 (抽出率調整)	791,438	467,199	集計対象の (乗率/調査月数) の和
⑤負債現在高 (千円)	5,311	6,672	
世帯数分布 (抽出率調整)	829,441	488,396	集計対象の (乗率/調査月数) の和

参考文献

小林良行(2012) 「公的統計マイクロデータ提供の現状と展望- 一橋大学での取り組みをもとに」『日本統計学会誌』 第 41 巻, 第 2 号, pp.401~420

筒井淳也 水落正明 秋吉美都 坂本和靖 平井裕久 福田亘孝(2011)『Stata で計量経済学入門 第 2 版』ミネルヴァ書房.

松浦寿幸(2010)『Stata によるデータ分析入門』東京図書.

*匿名データの整理及び本稿の作成におきまして、神戸大学勇上和史准教授、財務総合政策研究所宇南山卓総括主任研究官、山形大学窪田康平講師に多くのご助言をいただきました。ここに記して感謝申し上げます。