

TUTORIAL

政府統計マイクロデータの符号表から SAS 変数のラベルとフォーマットを自動生成する SAS プログラムの作成方法

How to Write a SAS Program to Automatically Create Labels and Formats for SAS Variables from a “Code Table” Provided together with Public Micro Data of Japanese Government

周防 節雄 (兵庫県立大学・名誉教授)

要旨

日本の政府機関が作成したマイクロデータには、メタデータとして「符号表」が付随している。この符号表は、政府機関共通の「政府統計個票データレイアウト標準記法」(平成 18 年)に準拠して作成されており、エクセルファイルでネットからもアクセス・保存が出来る。そこには、変数名、変数の長さ、変数の型(数値/文字)、ラベル、変数コードとその内容等が階層構造で表現されている。このエクセルファイルから SAS 変数のラベルと SAS フォーマットを自動生成し、かつ、SAS 変数も自動的に rename する SAS プログラムについて解説する。本チュートリアルでは、SAS ユーザー総会のデータ分析コンペに使用されている平成 16 年全国消費実態調査の教育用擬似マイクロデータの符号表を例題として用いるので、今後データコンペに参加予定のユーザーには大いに参考になるはずである。

1. はじめに

近年、パソコンの急速な高性能化と、データ分析用の優秀なソフトウェアの整備のおかげで、以前ではメインフレームコンピュータでしか出来なかった公的統計¹のマイクロデータ分析が大学・研究機関の研究室レベルで盛んに行われるようになってきた。公的統計のマイクロデータの利用は、平成 19 年に新「統計法」が制定される以前は、マイクロデータの「目的外使用」と一般的に呼称されていたが、今では「調査票情報の二次利用」²とされている。

マイクロデータの提供の際は、以前は各省庁がそれぞれ独自のファイルレイアウトを使っていた。平成 18 年に政府機関共通の「政府統計個票データレイアウト標準記法」が制定され、現在では、マイクロデータ本体と共に、この標準記法に準拠した「レイアウト表」と「符号表」がエクセルファイルで提供される。符号表(付録 1)に含まれる主な情報としては、項目名、長さ、型(数値型、文字型)、変数名、符号(いわゆるコード値)とその内容(意味)がある。

データ分析の過程では、変数のラベルとフォーマットが不可欠であるが、マイクロデータには変数の数が多いのが通例で、その作成には大変な手間と時間がかかる。そこで、手作業で変数のラベルとフォーマットを作成する代わりに、符号表を入力データとしてこれらを SAS プログラムで自動的に作成

¹ 公的統計とは、国・地方公共団体やその他の公的機関が作成する統計を指し、以前は「政府統計」とか「官庁統計」と呼称していた。公的統計のわかりやすい解説本として、『公的統計の体型と見方』(松井博 2008)がある。

² 統計法(平成十九年五月二十三日法律第五十三号)の第三十二条以下に関連の規定がある。統計法の条文は以下の URL で閲覧出来る。<http://law.e-gov.go.jp/htmldata/H19/H19H0053.html>

すれば、便利だと考えた。本チュートリアルでは、SAS ユーザー総会で 2013 年からデータ分析コンペで使用されている平成 16 年全国消費実態調査の教育用擬似マイクロデータの符号表を例題として用いているが、このプログラムは他の公的統計のマイクロデータの符号表にもそのまま適用できるので、利用者には役に立つプログラムと自負している。

2. 全国消費実態調査の教育用擬似マイクロデータ

2.1 概要

全国消費実態調査の教育用擬似マイクロデータ(以下「マイクロデータ」と呼称する)については「教育用擬似マイクロデータの開発とその利用 ～平成 16 年全国消費実態調査を例として～」と題する文書が以下の URL で閲覧出来るので、詳細についてはそちらを参照して欲しいが、ここでは同文書の「2.2 教育用擬似マイクロデータ」を以下に引用しておく。

<http://www.nstac.go.jp/services/pdf/sankousiryoku2407.pdf>

調査票情報から作成したものは、調査票情報であることを踏まえて、教育用擬似マイクロデータでは、個票データから高次元の集計表を作成し、その高次元の集計表から個票データに近似したマイクロデータを作成するという方法をとっている。集計表から作成するために、個票データでも、匿名データでもない擬似的なマイクロデータと言える。それでいて、この教育用マイクロデータは、実証分析に利用した際に、我が国の実態を反映できるように、つまり個票データの分布にできる限り近似するように工夫して作成する方向で考えた。

このように集計表から作成する教育用擬似マイクロデータは、基本的に、①個票データの分布に近づけるなど、元の個票データに近似したデータであること、②量的属性の相関関係を保つなど、量的属性間の関係が整合的であること、③全国消費実態調査で言えば収入総額と支出総額が合致しているなど、調査特有のデータ構造を保持すること、④標本調査における集計用乗率を考慮すること、⑤データ量は元の個票データに合わせること、の考えの下で作成している。作成例としての全国消費実態調査における考慮点として、質的属性 5 については、集計表の作成における分類項目が該当し、その項目数は限られたものになり、量的属性については、分析上必要と思われる収入項目、支出項目を収録する。

2.2 教育用擬似マイクロデータの SAS データセットへの変換

本チュートリアルの議論に必要なので、提供されるマイクロデータ本体の SAS データセットへの変換方法について簡単に触れておきたい。

提供されたマイクロデータは 197 変数、32,028 レコードから成る CSV ファイルであるが、提供に際しては 7 つの CSV ファイルに分割されている。最初のレコードから直ぐにデータが始まっており、いわゆる項目名とか変数名に相当するメタ情報はない。このような場合、SAS データセットへ変換するには様々な方法があるが、一番簡単にしかも確実に変換するには、**proc import** を使うのが最適である。

当初、7 つの CSV ファイルをそれぞれ先にエクセルファイルに変換してから、**proc import** のマクロを作って、7 つの SAS データセットに取り込んだ後に、その 7 個を set 文で縦に結合しようとした。ところが、**proc import** を使うと、文字変数の長さは、各ファイル内で最大の長さが自動的に定義されるので、データセットによっては同じ名前の変数なのに length が異なるケースが出て、結合しようと

すると、エラーになり、うまくいかなかった。どうしても、proc import を使いたかったので、まず、DOS コマンドで事前に CSV ファイル 7 個を縦に結合してから、proc import で SAS データセットに変換した。その結果、F1~F197 の変数、32,028 オブザベーションから成るデータセットが作成できた。

このデータセットをそのまま使うのでは、各変数が何を意味しているかわかりにくく、分析過程では大変な不便を感じるし、何よりも重大な変数の取り違えが起こりかねない。そこで、少なくともよく使う変数はそれと分かる変数名 (mnemonic name) に変えて、かつ、変数ラベルも付けておくのが望ましい。この変数名の mnemonic 化、変数ラベルの自動設定及び、変数の format の作成を以下で解説する。

2.3 符号表と SAS データセット化

マイクロデータの提供を受けると、マイクロデータ本体の他に、メタデータとして、「レイアウト表」と「符号表」がそれぞれのシートに保存されたエクセルファイルも提供される。本チュートリアルでは、この符号表から、変数のラベル・フォーマットの作成、及び変数の mnemonic 化をする SAS プログラムを論じているので、まず、符号表の中味に限定して概要を解説する。

提供された「平成 16 年全国消費実態調査符号表」の先頭部分を付録 1 に示す。このエクセルファイルのオリジナル版にはセキュリティ用のパスワードが設定されており、そのままでは SAS プログラムで読めないで、パスワード設定を外したファイルに変換して使う必要がある。そのファイルに対しては、proc import で SAS データセットに取り込む前に加工を施している (付録 2) が、詳細は、3.2 節で述べている。このエクセルファイルをそっくり同じイメージで SAS データセットに取り込んだ結果を表示した

「結果ビューア画面」を付録 3 に示す。

3. 本システムの解説

3.1 全工程のフローチャート

全工程のフローチャートを図 1 に示す。筆者自身が作成した SAS プログラムは、プログラム 1~プログラム 4 の 4 個あるが、プログラム 4 が自動作成した SAS プログラム (auto_format.sas) が 1 個ある。

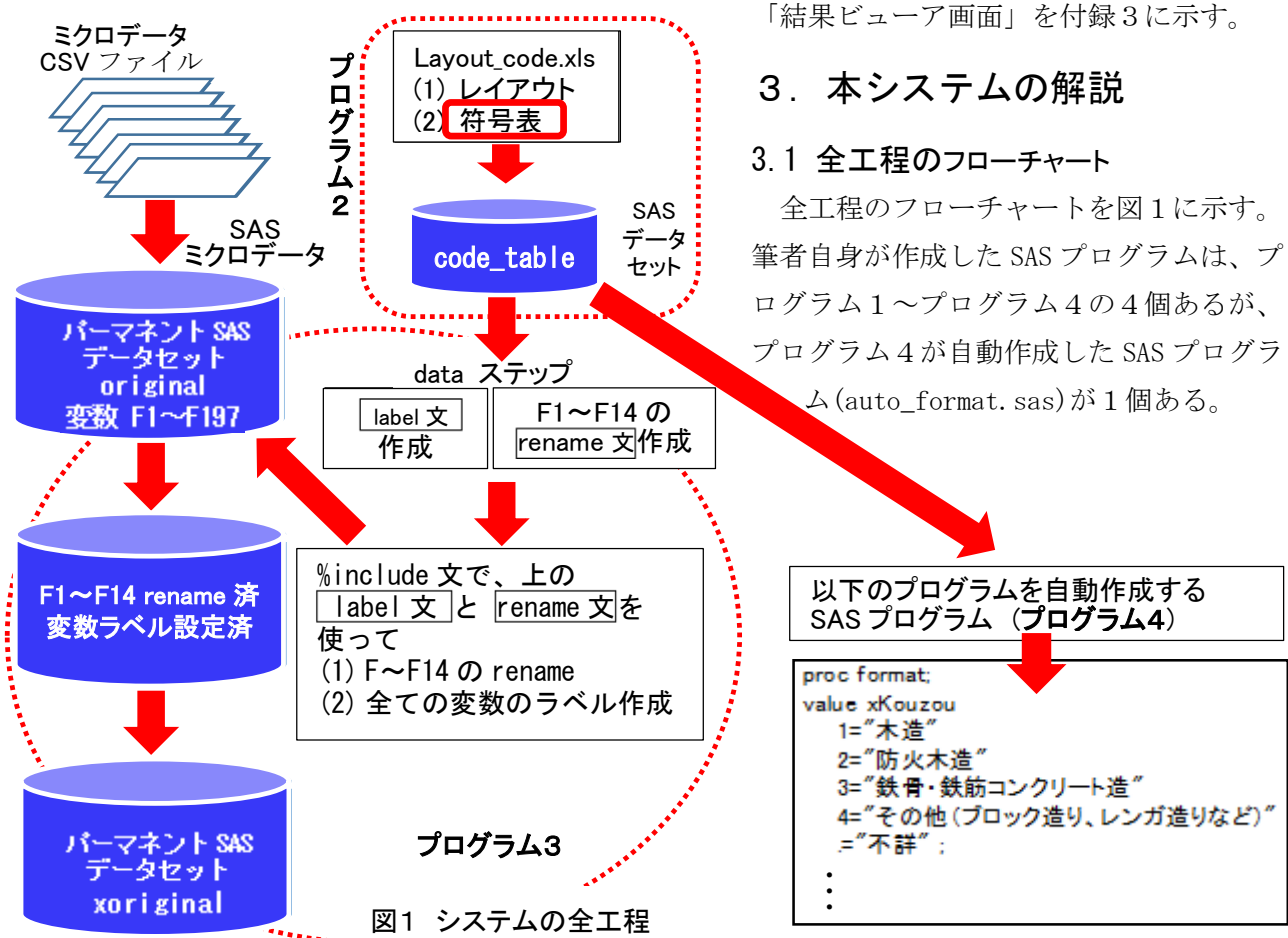


図1 システムの全工程

3.2 符号表を SAS データセットに変換

プログラム1でプログラムの動作環境を設定する。利用者は自分のパソコン環境に応じて、ここで設定しておけば、以下に示す一連のプログラムで再度設定は不要である。

エクセル形式で統計センターから提供される符号表(付録1)の先頭部には符号表に関する情報や、各列の項目名があるが、proc importを使う際には、余分なレコードなので削除して、その代わりに、項目名に相当する SAS 変数名を入力しておく(付録2)。プログラム2を実行すると、このエクセルファイルがパーマネント SAS データセット code_table(付録3)に変換される。これを使って、次節以降の処理を行う。

```
/* path_define.sas */ *以下の①～⑤を指定して下さい。;  
*① 符号表(Excel形式)が保存されているパスの設定;  
%let folder = G:\全消全消擬似マイクロデータ;  
*② SAS データセット(マイクロデータ本体)の保存先 folder 名の指定;  
* ★この folder は①で指定した folder の直ぐ下に作成しておく★;  
%let ds_folder = zensho_permanent;  
*③ SAS データセット(マイクロデータ本体)のファイル名の指定;  
* ★このファイルは②で指定した folder に保存しておく★;  
%let dsname = original;  
*④ マイクロデータ符号表(エクセルファイル)のファイル名の指定;  
* ★このファイルは①で指定した folder に保存しておくこと★;  
%let CodeTable = 統計センター_layout_code(鍵なし).xls;  
*⑤ マイクロデータ符号表(エクセルファイル)のシート名の指定;  
%let sheet = sheet1;  
*⑥ 統計調査名(日本語/英語)の指定; *途中結果表示の際に title 文で使用;  
%let SurveyName = 平成16年全国消費実態調査 擬似マイクロデータ;  
  
libname xyz "&folder%\&ds folder"; *この命令はそのまま;
```

プログラム1 プログラム実行環境の設定

```
/* giji_micro_zensho_codetable_import.sas */ *符号表を SAS データセットに変換;  
proc import out=xyz.code_table  
datafile="&folder%\&CodeTable" dbms=excel replace; getnames=yes; sheet="&sheet";  
run;  
proc print data=xyz.code_table(obs=56); title "&SurveyName.符号表"; run;
```

プログラム2 符号表を SAS データセットに変換

3.3 階層構造の符号表

符号表から変換した SAS データセット code_table(付録3)は、いわゆる、階層構造型である。つまり、複数のオブザベーションでひとまとまりの情報を構成しており、SAS で処理するには若干の熟練度が必要となる。この SAS データセットをよく見ると、以下の規則性が読み取れる。

- (1) 一つの変数に対する情報は、一行または複数行から成る。
- (2) 変数の区切り目のオブザベーションの SAS 変数 VcodeContents の値は「項目の区切り」
- (3) 変数が数値型の場合は、変数 Vtype の値は「1」で、一つのオブザベーションから成る。
- (4) 変数が文字型の場合は、変数 vtype の値は「欠損値」(半角空白)で複数のオブザベーションから成る。変数フォーマットが必要なのはこのタイプである。
- (5) 変数フォーマット作成に必要な変数は、Vcode(コードの値)と VcodeContents(コードの意味)のペアである。
- (6) 変数ラベル作成には、変数 variable と変数 Vlabel があればできる。変数 variable に変数名のローマ字表記のあるオブザベーションだけを対象にすればよい。

3.4 変数名の mnemonic 化と変数のラベル作成をする SAS プログラム: label_rename. sas

前節の(1)から(6)を考慮しながら、階層構造の SAS データセット code_table を読み取り、変数名の mnemonic 化と変数のラベル作成をするプログラムをプログラム 3 (次ページ)に示す。

このプログラムでは二つのテキストファイルを一旦 SAS の外に出力して、それらを再びこのプログラムの中で `%include` 文によって、命令文の一部として使っている。一つは、変数の rename に使用する右のテキストファイル F1toF14.txt、もう一つは、変数ラベルの作成のためのテキストファイル label.txt (付録 4)である。この二つのテキストファイルを見ると、誰でも進んでやりたくない手作業なので、自動化できることは作業の大幅な効率化になる。

このプログラムの各データステップで作成されるデータセットは、「結果ビューア」の①～⑤に表示しているので、プログラムの流れを理解し易い。

```
F1=SetaiKubun
F2=SetaiJinin
F3=ShuugyouJinin
F4=Kouzou
F5=Tatekata
F6=Shoyuu
F7=S1_Sex
F8=S1_Age
F9=S1_Shuugyou
F10=S1_KigyokuKubun
F11=S1_KigyokuKibo
F12=S1_Sangyou
F13=S1_Shokugyou
F14=Weight
```

ファイル名：
F1toF14.txt

3.5 変数 format を作成する SAS プログラム: format. sas

分析の過程や出力結果を見る場合、変数のコードのままでは大変読みにくいので、変数フォーマットを使用すべきである。変数の数が多いマイクロデータの場合、この変数フォーマットを作成するのに手間暇がかかる。符号表の SAS データセットから、これを自動的に作成する SAS プログラム format. sas をプログラム 4 に示す。このプログラムを実行すると、SAS プログラム `auto_format. sas` (付録 5) が自動作成される。

これを使用する際は、`%include` 文で対象の SAS プログラムの中に埋め込めば良い。簡単な使用例を付録 6 に示す。

なお、プログラム 1～4 は http://mighty.gk.u-hyogo.ac.jp/confidential/Tutorial_Suoh2014.zip からダウンロード出来る。

4 まとめ

日本の政府機関が作成したマイクロデータの提供を受けると、エクセル形式の「符号表」がメタデータとして付随してくる。変数名を分かり易い名前に rename をする、変数ラベルを付ける、変数フォーマットを定義するといった作業は、大変手間と時間がかかる。この作業は通常は手作業で行われるが、既に電子ファイル化された今回、符号表を SAS プログラムの入力データとして読み込み、こうした作業を全自動で行える SAS プログラムを開発した。今回は、平成 16 年全国消費実態調査の擬似マイクロデータの符号表を使用例として取り上げたが、他の政府統計の場合でも、ユーザーのパソコン環境の設定をする SAS プログラムを最初に実行するだけで、後のプログラムは変更なしに簡単に使えるように設計した。

SAS ユーザー総会のチュートリアルでは、これら一連の SAS プログラムの作成方法と解説を行う。階層構造型のファイルを SAS プログラムで処理するトレーニングにもなるはずである。これを通じて、SAS システムの持つ強力な「データハンドリング機能」を改めて認識する機械になれば、極めて有意義だと信じる。なお、本システムの利用方法は「利用の手引き」(付録 2 の後)を参照されたい。

参考文献

松井博(2008)『公的統計の体系と見方』日本評論社

```

/* label_rename.sas */ *教育用擬似マイクロデータの変数ラベル作成と変数の rename;

filename out1 "&folder¥label.txt"; *変数ラベル定義用テキスト(最後に自動削除);
filename out2 "&folder¥F1toF14.txt"; *変数 F1~F14 の rename 用テキスト(最後に自動削除);

* オリジナル変数 F15~F197 に変数ラベルを付するための LABEL 文の中味を外部ファイルに出力;
data label_define; keep var_label variable Vlabel;
  file out1; *LABEL 文の中味を外部ファイルに出力;
  length var_label $ 100;
  length No 8; *数値変数;
  set xyz.code_table; ← 付録3
  if variable="" then delete;

  if substr(variable, 1, 5)="Youto"
    then do; No=substr(variable, 6, 3);
      var_label=compress('v' || No) || '=' || compress(Vlabel || '');
    end;
  else var_label=compress(variable) || '=' || compress(Vlabel || '');
  put var_label;
run;
proc print; title "label"; var var_label variable Vlabel; run; ← 結果ビューア ①

data renameF1toF14; keep statement variable;
  file out2;
  *変数 F1~F14 を「符号表」にある変数名に変える rename 文の中味をテキストファイルに出力;
  array F F1-F14;
  set label_define;
  if _N_ <= 14 then do; statement=compress("F" || _N_ || "=" || variable);
    put statement;
  end;
  else stop;
run;

proc print; title "renameF1toF14"; run; ← 結果ビューア ②

data x; *変数 F1~F14 だけを rename をする;
  set xyz.&dsname;
  rename
  %include "&folder¥program¥出力テキストファイル¥F1toF14.txt";
  ;
proc print data=x(obs=40); title "変数 F1~F14 の rename"; run; ← 結果ビューア ③

data x; drop i F15-F197; * オリジナル変数 F15~F197 を V1~V183 に rename する;
  set x;
  array V {183};
  array F {183} F15-F197;
  do i=1 to 183; V[i]=F[i]; end;
  label
  %include "&folder¥label.txt";
run;
                                     結果ビューア ④
                                     ↓
proc print data=x(obs=20);          title "&SurveyName. (先頭 40 件:変数名表示)"; run;
proc print data=x(obs=20) label; title "&SurveyName. (先頭 40 件:変数ラベル表示)"; run;
                                     結果ビューア ⑤ ↑
data xyz.x&dsname; set x; run; *マイクロデータのパーマネントデータセット完成;

x "del &folder¥label.txt" ; x "echo label.txt を削除しました。";
x "del &folder¥F1toF14.txt"; x "echo F1toF14.txt を削除しました。";

```

プログラム3 変数名の mnemonic 化と変数ラベルの自動作成 SAS プログラム label_rename.sas

label

結果ビューア ①

OBS	var_label	variable	Vlabel
1	SetaiKubun="世帯区分"	SetaiKubun	世帯区分
2	SetaiJinin="世帯人員"	SetaiJinin	世帯人員
3	ShuugyouJinin="有業人員"	ShuugyouJinin	有業人員
4	Kouzou="住居の構造"	Kouzou	住居の構造
5	Tatekata="住居の建て方"	Tatekata	住居の建て方
6	Shoyuu="住居の所有関係"	Shoyuu	住居の所有関係
7	S1_Sex="性別"	S1_Sex	性別
8	S1_Age="年齢5歳階級"	S1_Age	年齢5歳階級
9	S1_Shuugyou="就業・非就業の別"	S1_Shuugyou	就業・非就業の別
10	S1_Kigyoubun="企業区分"	S1_Kigyoubun	企業区分
11	S1_Kigyokibo="企業規模"	S1_Kigyokibo	企業規模
12	S1_Sangyou="産業符号"	S1_Sangyou	産業符号
13	S1_Shokugyou="職業符号"	S1_Shokugyou	職業符号
14	Weight="集計用乗率"	Weight	集計用乗率
15	v1="年間収入"	Youto001	年間収入
16	v2="収入総額"	Youto002	収入総額
17	v3="実収入"	Youto003	実収入
18	v4="経常収入"	Youto004	経常収入
19	v5="勤め先収入"	Youto005	勤め先収入

中略

187	v173="個人・企業年金保険掛金"	Youto173	個人・企業年金保険掛金
188	v174="他の保険掛金"	Youto174	他の保険掛金
189	v175="有価証券購入"	Youto175	有価証券購入
190	v176="株式購入"	Youto176	株式購入
191	v177="他の有価証券購入"	Youto177	他の有価証券購入
192	v178="土地家屋借金返済"	Youto178	土地家屋借金返済
193	v179="他の借金返済"	Youto179	他の借金返済
194	v180="分割払・一括払購入借入金返済"	Youto180	分割払・一括払購入借入金返済
195	v181="財産購入"	Youto181	財産購入
196	v182="その他"	Youto182	その他
197	v183="繰越金"	Youto183	繰越金

結果ビューア ②

renameF1toF14

OBS	variable	statement
1	SetaiKubun	F1=SetaiKubun
2	SetaiJinin	F2=SetaiJinin
3	ShuugyouJinin	F3=ShuugyouJinin
4	Kouzou	F4=Kouzou
5	Tatekata	F5=Tatekata
6	Shoyuu	F6=Shoyuu
7	S1_Sex	F7=S1_Sex
8	S1_Age	F8=S1_Age
9	S1_Shuugyou	F9=S1_Shuugyou
10	S1_Kigyoubun	F10=S1_Kigyoubun
11	S1_Kigyokibo	F11=S1_Kigyokibo
12	S1_Sangyou	F12=S1_Sangyou
13	S1_Shokugyou	F13=S1_Shokugyou
14	Weight	F14=Weight

変数F1~F14のrename

結果ビューア ③

OBS	SetaiKubun	SetaiJinin	ShuugyouJinin	Kouzou	Tatekata	Shoyuu	S1_Sex	S1_Age	S1_Shuugyou	S1_Kigyoubun	S1_Kigyokibo	S1_Sangyou	S1_Shokugyou	Weight	F15	F16	F17	F18	F19	F20
1	1	2	1	1	1	1	1	2	1	1	3	W	1	19	##	##	##	##	##	0
2	1	2	1	1	1	1	1	2	1	1	3	W	1	15	##	##	##	##	##	0
3	1	2	1	1	1	1	1	2	1	1	3	W	1	15	##	##	##	##	##	0

結果ビューア ④

平成16年全国消費実態調査 擬似マイクロデータ(先頭40件:変数名表示)

OBS	Setaikubun	Setaijinin	Shuugyoujinin	Kouzou	Tatekata	Shoyuu	S1-Sex	S1-Age	S1-Shuugyou	S1-Kigyokubun	S1-Kigyokibo	S1-Sangyou	S1-Shokugyou	Weight	V1	V2	V3	V4	V5	V6	V7	V8	. . . 中略 . . .							
	1	2	1	1	1	1	1	2	1	1	3	W	1	19	##	##	##	##	##	0	0	0	V177	V178	V179	V180	V181	V182	V183	
1	1	2	1	1	1	1	1	2	1	1	3	W	1	19	##	##	##	##	##	0	0	0	0	##	0	##	0	##	0	##
2	1	2	1	1	1	1	1	2	1	1	3	W	1	15	##	##	##	##	##	0	0	0	0	##	0	##	0	##	0	##
3	1	2	1	1	1	1	1	2	1	1	3	W	1	15	##	##	##	##	##	0	0	0	0	##	0	##	0	##	0	##

...以下略...

結果ビューア ⑤

平成16年全国消費実態調査 擬似マイクロデータ(先頭40件:変数ラベル表示)

OBS	世帯区分	世帯人員	有業人員	住居の構造	住居の建て方	住居の所有関係	性別	年齢5歳階級	就業・非就業の別	企業区分	企業規模	産業符号	職業符号	集計用乗率	年間収入	収入総額	実収入	経常収入	勤め先収入	事業・内職収入	農林漁業収入	家賃収入	. . . 中略 . . .							
	1	2	1	1	1	1	1	2	1	1	3	W	1	19	##	##	##	##	##	0	0	0	0	0	0	0	0	0	0	
1	1	2	1	1	1	1	1	2	1	1	3	W	1	19	##	##	##	##	##	0	0	0	0	0	0	##	0	##	0	##
2	1	2	1	1	1	1	1	2	1	1	3	W	1	15	##	##	##	##	##	0	0	0	0	0	0	##	0	##	0	##
3	1	2	1	1	1	1	1	2	1	1	3	W	1	15	##	##	##	##	##	0	0	0	0	0	0	##	0	##	0	##

...以下略...

注 結果ビューアの④と⑤は対応している。

補足 システムのバージョンアップについて

プログラム2を少し書き換えれば、利用の手引き(付録2の下を参照)の「プログラムの実行手順(1)」を省略できることに気づいたが、本稿の原稿提出締め切り直前だったために、差し替え作業が間に合わなかった。当初、今回の教育用擬似マイクロデータだけに使うつもりで本システムを開発したが、汎用性を考慮すれば手順(1)はない方がはるかによい。

新バージョンでは、事前にユーザー側で、符号表のエクセルファイルのヘッダー行を「全て」削除し、中味のデータだけにしておく。これを proc import で読み込むと、変数名が自動的に F1, F2, ... となる。その後に、data ステップを一つ挿入して、この変数 F1, F2, ... を付録3にある変数名に rename するコードを挿入するだけで済む。チュートリアルの際は、この新バージョンを使って解説する。

なお、この新プログラム2は、3.5 節に記載の URL からダウンロードできるようにする。


```

/* format.sas */ *教育用擬似マイクロデータ用のproc format 作成; options nocenter;

filename out1 "&folder%auto_format.sas";

data format; keep formatName Vcode VcodeContents;
  set xyz.code_table; *符号表のデータセットを使用;
  retain sw formatName;

  if Vlabel="," then sw=0;
  if variable NE " " AND Vtype=" " /*文字型変数*/
    then do; sw=1; formatName=variable; output; return; end;
  if sw=1 then output;
run;

proc print; title "format 対象変数(1)"; var formatName Vcode VcodeContents; run;

*★===== dataset 'format' に対して、「V」と「△」の処理を追加する=====★;
data format; set format;
  if index(Vcode,"V") > 0 then Vcode=".";
  if (formatName="S1_KigyokuKibo" OR formatName="S1_KigyokuKubun") AND Vcode="△"
    then Vcode="99";
  else Vcode=kcompress(Vcode,"△");
run;
proc print; title "V converted to period"; run;
*★===== dataset 'format' に対する「V」と「△」の追加処理終了=====★;

proc sort data=format out=format1; by formatName; run;

data format1; set format1; by formatName;
  if first.formatName then statement2='value ' || compress('x' || formatName);
  statement=compress(Vcode || '=' || VcodeContents || '');
run;

proc print;
  title "format 対象変数(2)";
  var formatName Vcode VcodeContents statement statement2;
run;

data _null_; file out1;
  set format1 end=final; by formatName;
  if _n_=1 then put "proc format:";
  if first.formatName then put statement2;
  if last.formatName then put " " statement ";" /;
  else put " " statement;
  if final then put "run:";
run;

```

プログラム 4 SAS フォーマットの自動作成プログラム:format.sas

付録1 平成16年全国消費実態調査擬似マイクロデータ符号表 (オリジナル先頭部)

政府統計コード	実施時期	平成16年	作成日	訂正日
統計調査名	全国消費実態	集計区分	2人以上勤労世帯	ファイル名
調査票名	擬似マイクロデータ	(備考・補足事項)		(照会先等)
コード体系	Shift_JIS	「△」はブランク(空白)を示す。		
レコード長	2974			

行番号	項目名	階層	位置	バイト数	配置	型	種別	変数名	符号	符号内容	備考
1	世帯区分		1	1				SetaiKubun	1	勤労	
2									2	勤労以外	
3									3	無職	
4			1	2	1				.	項目の区切り	
5	世帯票		1								
6	世帯人員		2	3	2	2	1	SetaiJinin	△2~	2人~	
7			2	5	1				.	項目の区切り	
8	有業人員		2	6	2	2	1	ShuugyouJin	△1~	1人~	
9									VV	不詳	
10			2	8	1				.	項目の区切り	

以下略(残りの内容は付録3と同じ)

付録2 平成16年全国消費実態調査擬似マイクロデータ符号表 (統計センター_layout_code(鍵なし).xls 先頭部)

lineNO	Vlabel	strata	position	bytes	xhaichi	Vtype	xshubetsu	variable	Vcode	VcodeContents
1	世帯区分	1	1	1				SetaiKubun	1	勤労
2									2	勤労以外
3									3	無職
4		1	2	1					.	項目の区切り
5	世帯票	1								
6	世帯人員	2	3	2	2	1		SetaiJinin	△2~	2人~
7		2	5	1					.	項目の区切り
8	有業人員	2	6	2	2	1		ShuugyouJin	△1~	1人~
9									VV	不詳
10		2	8	1					.	項目の区切り

利用の手引き

最初に SAS プログラム `path_define.sas` で以下の①~⑥を指定して下さい。

- ① 符号表(Excel 形式)が保存されているパスの設定
`%let folder = G:\全消¥全消擬似マイクロデータ;`
- ② SAS データセット(マイクロデータ本体)の保存先 folder 名の指定
 ★この folder は①で指定した folder の直ぐ下に作成しておく★
`%let ds_folder = zensho_permanent;`
- ③ SAS データセット(マイクロデータ本体)のファイル名の指定
 ★このファイルは②で指定した folder に保存しておく★
`%let dsname = original;`
- ④ マイクロデータ符号表(エクセルファイル)のファイル名の指定
 ★このファイルは①で指定した folder に保存しておくこと★
`%let CodeTable = 統計センター_layout_code(鍵なし).xls;`
- ⑤ マイクロデータ符号表(エクセルファイル)のシート名の指定
`%let sheet = sheet1;`
- ⑥ 統計調査名(日本語/英語)の指定:
`%let SurveyName = 平成16年全国消費実態調査 擬似マイクロデータ;`

ファイル配置例 `G:\全消¥全消擬似マイクロデータ¥統計センター_layout_code(鍵なし).xls`

`G:\全消¥全消擬似マイクロデータ¥zensho_permanent¥original.sas7bdat`

②

③

プログラムの実行手順

- (1) 3.2 節の太字下線部で述べる前処理を符号表に対して行う。
(付録2の一行目参照)
この変数名が SAS プログラムで使われる。

注
補足(結果ビューア⑤の下)を参照)

- (2) 本文のプログラム①~④までを順番に実行する。

付録3 平成16年全国消費実態調査擬似マイクロデータ符号表
(SAS データセット code_table の冒頭部)

平成16年全国消費実態調査 擬似マイクロデータ符号表											
OBS	lineNO	Vlabel	strata	position	bytes	xhaichi	Vtype	xshubetsu	variable	Vcode	VcodeContents
1	1	世帯区分	1	1	1	.	.	.	SetaiKubun	1	1 勤労
2	2			2	2 勤労以外
3	3			3	3 無職
4	4		1	2	1	項目の区切り
5	5	世帯票	1			
6	6	世帯人員	2	3	2	2	1		SetaiJinin	Δ2~	2人~
7	7		2	5	1	項目の区切り
8	8	有業人員	2	6	2	2	1		ShuugyouJinin	Δ1~	1人~
9	9			VV	不詳
10	10		2	8	1	項目の区切り
11	11	現住居等に関する事項	2			
12	12	住居の構造	3	9	1	.	.	.	Kouzou	1	1 木造
13	13			2	2 防火木造
14	14			3	3 鉄骨・鉄筋コンクリート造
15	15			4	4 その他（ブロック造り、レンガ造りなど）
16	16			V	不詳
17	17		3	10	1	項目の区切り
18	18	住居の建て方	3	11	1	.	.	.	Tatekata	1	1 一戸建
19	19			2	2 長屋建
20	20			3	3 共同住宅（1・2階建）
21	21			4	4 共同住宅（3～5階建）
22	22			5	5 共同住宅（6～10階建）
23	23			6	6 共同住宅（11階建以上）
24	24			7	7 その他
25	25			V	不詳
26	26		3	12	1	項目の区切り
27	27	住居の所有関係	3	13	1	.	.	.	Shoyuu	1	1 持ち家（世帯員名義）
28	28			2	2 持ち家（その他名義）
29	29			3	3 民営賃貸住宅（設備専用）
30	30			4	4 民営賃貸住宅（設備共用）
31	31			5	5 市区町村営賃貸住宅
32	32			6	6 都市再生機構・公社等賃貸住宅
33	33			7	7 社宅・公務員住宅（借上げ含む）
34	34			8	8 借間
35	35			9	9 寮・寄宿舎
36	36			V	不詳
37	37		2	14	1	項目の区切り
38	38	世帯員に関する事項	2			
39	39	世帯主	3			
40	40	性別	4	15	1	.	.	.	S1_Sex	1	1 男
41	41			2	2 女
42	42		4	16	1	項目の区切り
43	43	年齢5歳階級	4	17	2	.	.	.	S1_Age	Δ1	2 4歳未満
44	44			Δ2	2 5～29歳
45	45			Δ3	3 0～34歳
46	46			Δ4	3 5～39歳
47	47			Δ5	4 0～44歳
48	48			Δ6	4 5～49歳
49	49			Δ7	5 0～54歳
50	50			Δ8	5 5～59歳
51	51			Δ9	6 0～64歳
52	52			10	6 5～69歳
53	53			11	7 0～74歳
54	54			12	7 5歳以上
55	55			VV	不詳
56	56		4	19	1	項目の区切り

SASデータセット CODE_TABLE: 527オブザベーション、11変数

付録4 全国消費実態調査擬似マイクロデータ
(SAS 変数ラベルの定義用テキストファイル: label.txt)

SetaiKubun="世帯区分"	v53="牛乳"	v119="健康保持用摂取品"
SetaiJinin="世帯人員"	v54="乳製品"	v120="保健医療用品・器具"
ShuugyouJinin="有業人員"	v55="卵"	v121="保健医療サービス"
Kouzou="住居の構造"	v56="野菜・海藻"	v122="交通・通信"
Tatekata="住居の建て方"	v57="生鮮野菜"	v123="交通"
Shoyuu="住居の所有関係"	v58="乾物・海藻"	v124="自動車等関係費"
S1_Sex="性別"	v59="大豆加工品"	v125="自動車等購入"
S1_Age="年齢5歳階級"	v60="他の野菜・海藻加工品"	v126="自転車購入"
S1_Shuugyou="就業・非就業の別"	v61="果物"	v127="自動車等維持"
S1_Kigyokubun="企業区分"	v62="生鮮果物"	v128="通信"
S1_Kigyokibo="企業規模"	v63="果物加工品"	v129="教育"
S1_Sangyou="産業符号"	v64="油脂・調味料"	v130="授業料等"
S1_Shokugyou="職業符号"	v65="油脂"	v131="教科書・学習参考教材"
Weight="集計用乗率"	v66="調味料"	v132="補習教育"
v1="年間収入"	v67="菓子類"	v133="教養娯楽"
v2="収入総額"	v68="調理食品"	v134="教養娯楽用耐久財"
v3="実収入"	v69="主食的調理食品"	v135="教養娯楽用品"
v4="経常収入"	v70="他の調理食品"	v136="書籍・他の印刷物"
v5="勤め先収入"	v71="飲料"	v137="教養娯楽サービス"
v6="事業・内職収入"	v72="茶類"	v138="宿泊料"
v7="農林漁業収入"	v73="コーヒー・ココア"	v139="バック旅行費"
v8="家賃収入"	v74="他の飲料"	v140="月謝類"
v9="他の事業収入"	v75="酒類"	v141="他の教養娯楽サービス"
v10="内職収入"	v76="外食"	v142="その他の消費支出"
v11="本業以外の勤め先・事業・内職収入"	v77="一般外食"	v143="諸雑費"
v12="他の経常収入"	v78="学校給食"	v144="理美容サービス"
v13="財産収入"	v79="住居"	v145="理美容用品"
v14="社会保障給付"	v80="家賃地代"	v146="身の回り用品"
v15="公的年金給付"	v81="設備修繕・維持"	v147="たばこ"
v16="他の社会保障給付"	v82="設備材料"	v148="その他の諸雑費"
v17="仕送り金"	v83="工事その他のサービス"	v149="こづかい(使途不明)"
v18="特別収入"	v84="光熱・水道"	v150="交際費"
v19="受贈金"	v85="電気代"	v151="食料"
v20="その他"	v86="ガス代"	v152="家具・家事用品"
v21="実収入以外の収入"	v87="他の光熱"	v153="被服及び履物"
v22="預貯金引出"	v88="上下水道料"	v154="教養娯楽"
v23="保険取金"	v89="家具・家事用品"	v155="他の物品サービス"
v24="個人・企業年金保険取金"	v90="家庭用耐久財"	v156="贈与金"
v25="他の保険取金"	v91="家事用耐久財"	v157="他の交際費"
v26="有価証券売却"	v92="冷暖房用器具"	v158="仕送り金"
v27="株式売却"	v93="一般家具"	v159="非消費支出"
v28="他の有価証券売却"	v94="室内装備・装飾品"	v160="直接税"
v29="土地家屋借入金"	v95="寝具類"	v161="勤労所得税"
v30="他の借入金"	v96="家事雑貨"	v162="個人住民税"
v31="分割払・一括払購入借入金"	v97="家事用消耗品"	v163="他の税"
v32="財産売却"	v98="家事サービス"	v164="社会保険料"
v33="その他"	v99="被服及び履物"	v165="公的年金保険料"
v34="繰入金"	v100="和服"	v166="健康保険料"
v35="支出総額"	v101="洋服"	v167="介護保険料"
v36="実支出"	v102="男子用洋服"	v168="他の社会保険料"
v37="消費支出"	v103="婦人用洋服"	v169="他の非消費支出"
v38="食料"	v104="子供用洋服"	v170="実支出以外の支出"
v39="穀類"	v105="シャツ・セーター類"	v171="預貯金"
v40="米"	v106="男子用シャツ・セーター類"	v172="保険掛金"
v41="パン"	v107="婦人用シャツ・セーター類"	v173="個人・企業年金保険掛金"
v42="めん類"	v108="子供用シャツ・セーター類"	v174="他の保険掛金"
v43="他の穀類"	v109="下着類"	v175="有価証券購入"
v44="魚介類"	v110="男子用下着類"	v176="株式購入"
v45="生鮮魚介"	v111="婦人用下着類"	v177="他の有価証券購入"
v46="塩干魚介"	v112="子供用下着類"	v178="土地家屋借入金返済"
v47="魚肉練製品"	v113="生地・糸類"	v179="他の借入金返済"
v48="他の魚介加工品"	v114="他の被服"	v180="分割払・一括払購入借入金返済"
v49="肉類"	v115="履物類"	v181="財産購入"
v50="生鮮肉"	v116="被服関連サービス"	v182="その他"
v51="加工肉"	v117="保健医療"	v183="繰越金"
v52="乳卵類"	v118="医薬品"	

付録5 全国消費実態調査擬似マイクロデータ(SAS 変数用フォーマット)

auto_format. sas

```

proc format;
value xKouzou
  1="木造"
  2="防火木造"
  3="鉄骨・鉄筋コンクリート造"
  4="その他(ブロック造り、レンガ造りなど)"
  .="不詳";

value xS1_Age
  1="24歳未満"
  2="25～29歳"
  3="30～34歳"
  4="35～39歳"
  5="40～44歳"
  6="45～49歳"
  7="50～54歳"
  8="55～59歳"
  9="60～64歳"
  10="65～69歳"
  11="70～74歳"
  12="75歳以上"
  .="不詳";

value xS1_KigyokuKibo
  1="1～4人"
  2="5～29人"
  3="30～499人"
  4="500～999人"
  5="1000人以上"
  99="非就業又は官公"
  .="不詳";

value xS1_KigyokuKubun
  1="民営"
  2="自営"
  3="官公"
  99="非就業"
  .="不詳";

value xS1_Sangyou
  1="農業"
  2="林業"
  3="漁業"
  4="鉱業"
  5="建設業"
  6="製造業"
  7="電気・ガス・熱供給・水道業"
  8="情報通信業"
  9="運輸業"
  10="卸売・小売業"
  11="金融・保険業"
  12="不動産業"
  13="飲食店・宿泊業"
  14="医療・福祉"
  15="教育・学習支援業"
  16="複合サービス事業"
  17="サービス業(他に分類されないもの)"
  18="公務(他に分類されないもの)"
  19="その他(非就業を含む)"
  .="不詳";

value xS1_Sex
  1="男"
  2="女";

value xS1_Shokugyou
  1="常用労務作業者"
  2="臨時及び日々雇労務作業者"
  3="民間職員"
  4="官公職員1"
  5="官公職員2"
  6="商人及び職人"
  7="個人経営者"
  8="農林漁業従事者"
  9="法人経営者"
  10="自由業者"
  11="その他"
  12="無職"
  .="不詳";

value xS1_Shugyou
  1="就業"
  2="うちパート"
  3="非就業"
  4="うち仕事を探している"
  .="不詳";

value xSetaiKubun
  1="勤労"
  2="勤労以外"
  3="無職";

value xShoyuu
  1="持ち家(世帯員名義)"
  2="持ち家(その他名義)"
  3="民営賃貸住宅(設備専用)"
  4="民営賃貸住宅(設備共用)"
  5="県市区町村営賃貸住宅"
  6="都市再生機構・公社等賃貸住宅"
  7="社宅・公務員住宅(借上げ含む)"
  8="借間"
  9="寮・寄宿舎"
  .="不詳";

value xTatekata
  1="一戸建"
  2="長屋建"
  3="共同住宅(1・2階建)"
  4="共同住宅(3～5階建)"
  5="共同住宅(6～10階建)"
  6="共同住宅(11階建以上)"
  7="その他"
  .="不詳";

run;

```

付録6 SAS 変数用フォーマットプログラム auto_format. sas の使用例

```

/* freq. sas */ options nocenter;
*擬似マイクロデータ用 SAS フォーマット設定プログラム「auto_format. sas」の使用例;
%let microdata=xoriginal; * ユーザーが付けたマイクロデータ本体(SAS data set)のファイル名;
libname micro "G:¥全消¥全消擬似マイクロデータ¥zensho_permanent"; * ユーザーの環境に設定;
%include "G:¥全消¥全消擬似マイクロデータ¥auto_format. sas"; *proc format プログラム実行;

title "擬似マイクロデータ 度数分布表";

%macro freq(variable);
proc freq data=micro.xoriginal; tables &variable / missing;
    format &variable x&variable.;
run;
%mend;

%freq(SetaiKubun)
%freq(Kouzou)
%freq(Tatekata)
%freq(Shoyuu)
%freq(S1_Sex)
%freq(S1_Age)
%freq(S1_Shuyugyo)
%freq(S1_Kigyokubun)
%freq(S1_Kigyokibo)
%freq(S1_Sangyo)
%freq(S1_Shokugyo)

```

このプログラムは平成 16 年全国消費実態調査の擬似マイクロデータに含まれる全ての分類変数の度数分布表を出力する。%include 文によって、実行時に全ての変数のフォーマットを定義する auto_format. sas が実行される。その結果、マクロ freq 内の format 文が有効になり、出力されるクロス表の表側には、変数コードではなくて、下に示すように、format 値が表示される。

なお、表頭には変数名の代わりに変数ラベルが表示されている。

実行結果のうち、2 番目から 5 番目までの 4 つの表「結果ビューア」画面を以下に示す。

擬似マイクロデータ 度数分布表

FREQ プロシジャ

住居の構造				
Kouzou	度数	パーセント	累積 度数	累積 パーセント
不詳	5584	17.43	5584	17.43
木造	14064	43.91	19648	61.35
防火木造	3626	11.32	23274	72.67
鉄骨・鉄筋コンクリート造	8739	27.29	32013	99.95
その他(ブロック造り、レンガ造りなど)	15	0.06	32028	100.00

擬似マイクロデータ 度数分布表

FREQ プロシジャ

住居の所有関係				
Shoyuu	度数	パーセント	累積 度数	累積 パーセント
不詳	1446	4.51	1446	4.51
持ち家(世帯員名義)	22936	71.61	24382	76.13
持ち家(その他名義)	785	2.45	25167	78.58
民営賃貸住宅(設備専用)	3952	12.34	29119	90.92
民営賃貸住宅(設備共用)	3	0.01	29122	90.93
京市区町村営賃貸住宅	1287	4.02	30409	94.95
都市再生機構・公社等賃貸住宅	325	1.01	30734	95.96
社宅・公務員住宅(借上げ含む)	1282	4.00	32016	99.96
借間	12	0.04	32028	100.00

擬似マイクロデータ 度数分布表

FREQ プロシジャ

住居の建て方				
Tatekata	度数	パーセント	累積 度数	累積 パーセント
不詳	5665	17.69	5665	17.69
一戸建	21623	67.51	27288	85.20
長屋建	87	0.27	27375	85.47
共同住宅(1~2階建)	1055	3.29	28430	88.77
共同住宅(3~5階建)	2679	8.36	31109	97.13
共同住宅(6~10階建)	641	2.00	31750	99.13
共同住宅(11階建以上)	278	0.87	32028	100.00

擬似マイクロデータ 度数分布表

FREQ プロシジャ

性別				
S1_Sex	度数	パーセント	累積 度数	累積 パーセント
.	1	0.00	1	0.00
男	29381	91.74	29382	91.74
女	2646	8.26	32028	100.00