

SPSS による全国消費実態調査「擬似マイクロデータ」の利用例

2013. 6. 29 (7. 2改)

明海大学経済学部/東京大学社会科学研究所 伊藤 伸介
統計センター/東京大学社会科学研究所 榎田 直木

公的統計の匿名データは、一般に、相当ボリュームの大きな CSV 形式ファイルである（数十万レコード、数十～百超メガバイト）。そんな匿名データを、統計パッケージソフトに取り込むに当たって、事前に演習として「擬似マイクロデータ」を扱うことが役に立つかもしれない。なお、このデータの作成及び試行提供（無償）は、（独）統計センターである。

擬似マイクロデータ（平成 16 年全国消費実態調査（総務省統計局実施）に基づく）

<http://www.nstac.go.jp/services/giji-microdata.html>

統計センターに所定のフォームにより申出てファイル解凍用パスワードを入手すること。

全消擬似マイクロは、将来的に申出を行おうとしている匿名データの種別を問わず、データのインポート、データセットに対するラベルの付与といった演習用に気軽に利用可能である（全国消費実態調査以外にも、就業構造基本調査や社会生活基本調査（ともに総務省統計局）、国民生活基礎調査（厚生労働省）など）。

なお、統計パッケージソフトとして、ここでは SPSS を念頭に記しているが、Stata や SAS、R の場合であっても、考え方は同様である。

1. データのインポート

全消擬似マイクロは、大規模データと簡易データの 2 種類ある。このうち、簡易データ（CSV ファイル又は Excel ファイル）は、統計センターウェブサイトからダウンロードして、パスワードを用いて解凍後、SPSS にそのまま 1 つのデータセットとしてインポートすればよい。

一方、大規模データは、5000 ケース（レコード）ずつに分割された 7 つの CSV ファイルを同ウェブサイトからダウンロードして、それぞれ解凍後、7 つのファイルを連結して 1 つのデータセットを完成させる必要がある。

連結には、以下の方法などが考えられる。

- 7 つの CSV ファイルをまず 1 つにつないで*から、SPSS の上でインポートする。
※ 例えば <http://www.itmedia.co.jp/help/tips/windows/w0066.html> を参照
- SPSS で CSV ファイルを 1 つずつインポートしてそれぞれデータセット化し、しかる後に SPSS の上で 1 つにつなぐ。

参考： 全消匿名データは、全消擬似マイクロ（簡易データ）と同様に、提供段階で受け取るものがそのまま 1 つのデータセットである（連結作業は要しない）。

2. データセットに対するラベルの付与

SPSS を用いて全消擬似ミクロの CSV ファイルをインポートすると、初期状態のデータセットでは、その変数名は v1, v2, …と機械的な形で設定される。また、このままの状態では、変数の値が何を表しているのか人間には理解できない形式となっている。

そこで、全消擬似ミクロのデータレイアウト及び符号表を参照しながら、データセットに対して変数ラベルを付与したり、値ラベル（例えば年齢階級について、1 ならば「24 歳未満」、2 ならば「25～29 歳」、…）を付与したりするとよい。

この際、この [SPSS シンタックス](#) を参考にして変数名、ラベルを付与するとよい。

備考： 全消匿名データの場合も、匿名データのデータレイアウト及び符号表を参照しながら、変数ラベルや値ラベルを付与するとよい。

3. データ構造の概要

全消擬似ミクロを扱うには、平成 16 年全国消費実態調査の調査報告書やウェブサイト（調査の概要、調査票等）の他に、データレイアウト及び符号表、全消擬似ミクロ付録の利用の手引などを、十分参照すること。

- (1) 全消擬似ミクロは、平成 16 年全国消費実態調査（二人以上の勤労者世帯）を対象に、作成されている。レコード数は約 32,000

備考： 全国消費実態調査の調査対象は、もともとは約 5 万世帯

- (2) 全消擬似ミクロでは、世帯員の属性を用いた分析は、世帯主についてのみ可能。

備考： 全消匿名データには、世帯主以外の世帯員の属性もある。

- (3) 全消擬似ミクロには、収入・支出に関する調査事項がある。

備考： 全消匿名データには、資産に関する調査項目も（貯蓄現在高、負債現在高 等）

- (4) 全消擬似ミクロの各レコードには、母集団推計のための乗率が設定されている。

擬似ミクロを SPSS に取り込んだら、まず、乗率を用いた単純集計を行い、その集計が正しいかどうか、擬似ミクロ付録の基礎集計表と照合すること。なお、その集計値は、総務省統計局による公表値と一致しない。

備考： 一般に、公的統計の匿名データにも各レコードに乗率が設定されている。全消匿名データも、匿名化ゆえ、それを集計しても総務省統計局による公表値と一致しないが、その匿名化は内閣府統計委員会の意見を聞いた上で行われているものであり、匿名データを用いた分析結果は、学術研究（論文作成など）には有用とされている。

なお、擬似ミクロは、学術研究というよりも演習などに向けたものである。

- (5) 全消擬似マイクロでは、基とするデータに含まれる外れ値の除外処理を行っている。なお、全消擬似マイクロを用いて分布を取ると、場合によっては一見外れ値のように見えるものが出てくる場合がある。これは、全消擬似マイクロの作成において、量的属性を対象として多変量正規乱数を発生した結果生じたものである。

[参考] 授業における全消擬似マイクロの利用

全消擬似マイクロは、大学における実証分析の演習にも適したデータであり、多くの変数を含むことから、クロス集計表の作成やモデル分析の練習だけでなく、これを用いて学生にレポートを書かせることも可能である。

- 例 基本統計量の計測
散佈図等による分布特性の把握
クロス表の作成
相関係数の計測
回帰分析
等