

## I.5 データセットの重み付けについて

この節では、JGSS-2016 データセットに附属する変数【WEIGHT】の作成方法、【WEIGHT】を用いることによる推定結果への影響度を概観し、データセット利用者が【WEIGHT】を使用するか否かを判断するための情報を提供する。また、【WEIGHT】の作成方法の改定について付記する。

文中、アルファベットを【 】で括った文字は、データセットにおける変数名を意味する。

### 1. JGSS-2016 の推定方法

JGSS の標本設計では、全国の市町村を地域ブロック及び市郡規模により 24 の層に分け、その層の人口規模（住民基本台帳登録者数）に比例した数だけ国勢調査の基本単位区を抽出し、更に各単位区からそれぞれ約 15 人の調査対象を無作為に抽出している（「I.3 調査の方法」参照）。このため、推定の過程が単純化され、原理的には調査対象ごとの値を単に加算するだけで、結果の構成比を推定できる<sup>1)</sup>。

一方、官公庁が標本調査により実施する指定統計調査などでは、結果の推定において標本の分布を直近の統計調査などから得られた人口の年齢分布などに合わせる推定法が用いられることがある。現実の標本調査においては、偶然又は実務上の制約（不在や拒否）のため推定結果に偏りが生じることがある。特に、最近、統計調査における回収率の低下が顕著で、かつ、それが若年層など特定の層で高いことが問題となっている。このようなことから、少しでも推定の精度を高めるため、既存の補助情報を活用しようという試みである。

諸外国の社会調査や類似の調査を見ても、カナダやイタリアなどいくつかの国では国勢調査や労働力調査の結果を補助情報として活用する推定法が用いられている。また、米国の GSS では、このような推定法は採用していないものの、結果の分布が全数調査などと比べ大きな差のないことを確認するなど、標本の分布に偏りが無いことを確かめた上で比較・分析するという視点は重視されている<sup>2)</sup>。

補助情報を用いる推定法では、乗率＝「ウエイト」を計算してそれを用いて集計・分析することになるが、その作成方法は必ずしも一つに定まるものではない。また、補助情報に依存しない変数もあるので、場合によってはこの推定法を適用することで、かえって推定値が現状から離れてしまう可能性もある。ウエイトを用いるか否かの選択は、分析の目的に応じて、データセット利用者が判断してほしい。

注 1) ただし、住民基本台帳ベースの人口に基づく推計値となる。

2) 清水誠, 2001, 「JGSS 第 2 回予備調査 データセットのウエイトについて」 SSJ Data Archive Research Paper Series 14, 東京大学社会科学研究所附属日本社会研究情報センター

## 2 変数【WEIGHT】の作成方法

JGSS-2016 データセットには、推定用ウエイトとして変数【WEIGHT】が付されている。これは、調査票1枚が代表する人数を、下記の式で男女、年齢階級別の14区分ごとに計算したものである。基準人口としては、「2015年10月1日現在推計人口」（総務省）の年齢、男女別全国日本人人口を用いている。

$$\text{ウエイト} = \text{「基準人口」} / \text{「JGSS-2016 回答者数」}$$

表1 男女、年齢階級別ウエイト計算表-JGSS-2016

		人口 (千人)	回答者数	ウエイト
男	25-29	3,211	67	47,925
	30-34	3,653	82	44,549
	35-39	4,192	83	50,506
	40-44	4,922	102	48,255
	45-49	4,366	108	40,426
女	25-29	3,083	57	54,088
	30-34	3,531	78	45,269
	35-39	4046	114	35,491
	40-44	4764	130	36,646
	45-49	4254	147	28,939

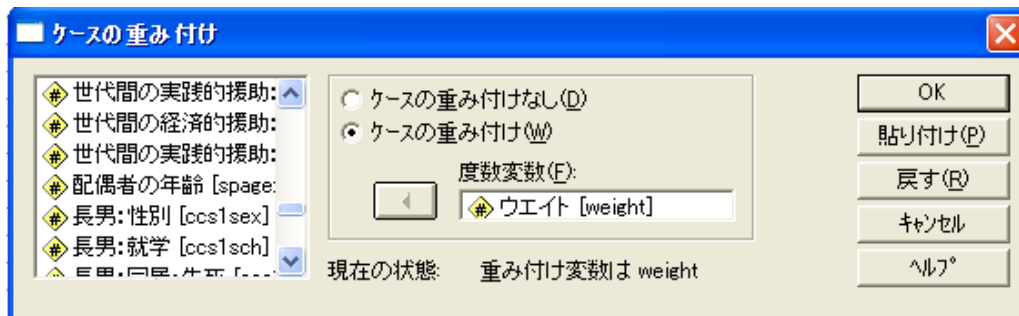
算出されたウエイトの値を男女別で比較すると、25-29歳と30-34歳では女性が高く、35歳以上の階級では男性が高い。年齢で比較すると、女性では若年層で高いが、男性では35-39歳で高い。ウエイトの最大値は25-29歳女性の54,088、最小値は45-49歳女性の28,939であり、最大値は最小値の約1.9倍となっており、このウエイトによる補正の効果がうかがわれる。

このようなウエイト値の差は、調査票の回収率の差によって生じたものである。すなわち、回収率は若年層で低く、年齢が上がるにつれて徐々に上昇するが、高齢になると再度低下していることを意味している。

### 3 SPSSにおいて【WEIGHT】により重み付けをする方法

JGSS のデータセットは、SSJ データアーカイブから SPSS 形式で提供されている。SPSS の以下の操作により、各ケースが【WEIGHT】で重み付けされる。

- (1) メニューバーから「データ (D)」→「ケースの重み付け (W)」を選択
- (2) 「ケースの重み付け」ダイアログから「ケースの重み付け (W)」を選択、「度数変数 (F)」として【WEIGHT】を指定したのちに「OK」
- (3) 通常どおりに集計・解析



### 4 推定結果へのウエイトの影響

ウエイトを用いた場合と用いない場合で JGSS-2016 の推定結果がどのように異なるか、属性別構成比で比較すると、表 2 のとおりである。表には参考として、JGSS-2016 と同じところに実施された回答者数の大きな調査の結果も掲載している。

ウエイトを用いた集計結果の主な特徴は、次のとおりである。

- (1) 男女、年齢階級区分については、ウエイト作成の際に補助情報として用いているので、ウエイトを用いた集計結果は参考値と一致する（母集団と一致する）。ウエイトを用いない場合、男性及び若年層の構成比が母集団に比べて小さめに推計されることが分かる。
- (2) 配偶関係の「有配偶」の構成比は、ウエイトを用いない場合、比較対象とした労働力調査を上回っている。ウエイトを用いた場合もまだ差が見られるものの、その差は縮小している。この結果には調査方法の違いが影響している可能性もあるので一概に言えないが、ウエイトを用いることにより若年層の比重が高まり、母集団により近い推定値が得られていることが分かる。
- (3) 就業状態の構成比は、ウエイトによる補正の効果が概して小さい。いずれの場合も、ウエイトを用いた集計結果とウエイトを用いない修正結果がほとんど変化しない。性別や年齢に偏りなく、様々な就業状態の対象者が回答していることを示唆する。
- (4) ちなみに、ウエイト作成に用いた男女、年齢階級別で違いが大きいと考えられる意識項目、例えば、生活の程度に対する意識について集計を行ってみると、ウエイトを用いた場合と用いない場合で構成比の推定値にはほとんど差が生じていない。

表2 推計方法（単純集計、ウエイト付集計）による構成比の比較（％）－JGSS-2016

	単純集計 <sup>1)</sup>	ウエイト付集計 <sup>2)</sup>	参考値	
男女【SEXA】				
男	45.7	50.8	50.8	2015年10月現在 推計人口 <sup>3)</sup> (総務省)
女	54.3	49.2	49.2	
年齢【AGEB】				
25～29歳	12.8	15.7	15.7	2015年10月現在 推計人口 <sup>3)</sup> (総務省)
30～34歳	16.5	18.0	18.0	
35～39歳	20.4	20.6	20.6	
40～44歳	24.0	24.2	24.2	
45～49歳	26.3	21.5	21.5	
配偶関係【DOMARRY】				
有配偶	68.6	66.9	62.6	2016年1月 労働力調査 <sup>4)</sup> (総務省)
離死別	4.4	4.1	4.8	
未婚	27.0	29.0	32.6	
就業状態 <sup>5)</sup> 【XJOB1WK】， 【TP5UNEMP】				
従業	85.0	85.7	81.2	2016年1月 労働力調査 <sup>6)</sup> (総務省)
休業	0.9	1.0	1.9	
失業	1.4	1.5	2.9	
家事	10.7	9.9	11.2	
学生	0.5	0.5	0.5	
その他	1.3	1.4	2.3	
生活の程度【OP5LEVK】				
上	1.2	1.3	1.0	2015年6月 国民生活に関する 世論調査 <sup>7)</sup> (内閣府)
中の上	14.1	14.2	14.6	
中の中	46.0	45.5	57.1	
中の下	31.2	31.2	22.6	
下	7.5	7.8	4.8	

- 1) 単純集計は、ウエイトを用いないで集計した結果であり、無回答者は集計対象から除いている。
- 2) ウエイト付集計は、ウエイトで重み付けを行って集計した結果であり、無回答者は集計対象から除いている。
- 3) 推計人口の集計は、25～49歳の日本人についてである。
- 4) 労働力調査の集計は、25～49歳の者についてであり、日本人以外の者も含む。配偶関係が不詳の者は集計対象から除いている。
- 5) 就業状態の単純集計、ウエイト付集計は、従業 (XJOB1WK=1)、休業 (XJOB1WK=2)、失業 (TP5UNEMP=1)、家事 (TP5UNEMP=4)、学生 (TP5UNEMP=3)、その他 (TP5UNEMP=2, 5, 6, 7, 9) と集計している。
- 6) 労働力調査の集計は4)と同様である。集計の定義は、従業 (就業者中の従業者)、休業 (就業者中の休業者)、失業 (完全失業者)、家事 (非労働力人口中の家事)、学生 (非労働力人口中の通学)、その他 (非労働力人口中のその他) であり、就業状態不詳の者は集計対象から除いている。
- 7) 国民生活に関する世論調査の集計は、20～49歳の日本人についてであり、「わからない」「無回答」は集計対象から除いている。また、質問文 (お宅の生活の程度は、世間一般からみて、どうですか。この中から1つお答えください) や調査方法 (面接) が、JGSSとは異なる。

孟 哲男