

## 「消費生活に関するパネル調査」における 残存サンプルの配偶率について

坂口 尚文

(公益財団法人 家計経済研究所 次席研究員)

### 1. はじめに

本稿は、「消費生活に関するパネル調査（以下、JPSC）」において、対象者の脱落が、残存しているサンプルの配偶構成にどのような影響を与えているか検証する。脱落により配偶率に歪みが生じている場合は、その配偶率の値が意味をなさないばかりではなく、各時点の労働力率や出生率など多くの指標にも影響を与える。JPSCでは、ほとんどの調査項目が対象者の配偶状態の影響を強く受けていることが予想されるため、残存サンプルの配偶率に歪みがないか検証することは、調査の信頼性を担保する上で重要な作業となる。

JPSCは、20歳代後半の女性を対象にした追跡調査である。若年女性を調査対象にしている理由の一つとして、追跡期間中に結婚の発生を効率的に観測するねらいがある。一方で結婚は、対象者が調査から脱落する要因ともなっている。そのことは、対象者の拒否理由からも明らかにされてきた。JPSCでは対象者の調査拒否理由を調査員の聞き取りにより収集しており、結婚は拒否理由の上位に位置している（村上 2003; 坂本 2006）。結婚、出産により、調査に協力する時間的、心理的余裕が失われることが調査拒否へとつながっていると予想される。

ただ個々の分析において、JPSCのデータを使用する際の問題と関心は、対象者の調査拒否理由にあるというよりも、対象者が脱落することによって、抽出したすべての対象者について結婚の有無が観測できていないこと、さらに、そのことに

起因するサンプルの歪みにある。調査の拒否理由は、結婚を契機にした脱落が一定数存在することを物語っているが、すべての無配偶の対象者が結婚を契機に脱落するわけではない。脱落した無配偶者の中には一定期間を経て結婚、出産する対象者も多いことが予想され、拒否理由だけではそのような潜在的な結婚、出産層の情報は得られない。また、対象者は結婚、出産をしていたとしても、それらを拒否の理由に挙げるとは限らず、そもそも4割前後の対象者については、拒否の明確な理由さえ分かっていない。結婚については住所移転による追跡不能や、調査に対して夫の理解が得られないなど、間接的な要因で調査から脱落するケースも考えられる。結婚、出産の発生と拒否理由は1対1に対応していない。このような点から、調査の拒否理由は結婚、出産の逸失情報を捕捉する手段としては不十分なものであり、情報を捕捉するには別のアプローチを考えなければならない。

配偶率の歪みを検証する際の困難は、脱落が発生した後のサンプルの配偶状態を類推しなければならない点にある。パネル調査の脱落を扱った研究の多くでは、対象者が脱落するプロセスに強い仮定をおき、観測できるデータ——脱落していない対象者の情報、および脱落した対象者については脱落する前の情報が得られる——のみを用いて、脱落後の情報を推計している。本稿でも、改善の策として、このような強い仮定をおいたモデルを用いて、残存している対象と脱落した対象の配偶率に相違があるかを検証する。ただ、JPSCには、その調査設計上、「国勢調査」等の大規模

な人口統計調査を補助情報として活用できる余地がある。JPSCは年齢と配偶構成を「国勢調査」の結果に基づいて標本抽出を行っている。配偶率は対象者の加齢に伴い経年変化する。脱落による磨耗が生じなかった場合を想定すれば、抽出サンプルの配偶率は「国勢調査」の結果とシンクロナイズしていたことが予想される。JPSCでは抽出の年齢幅を5歳程度に限定していること、また対象を女性のみに行っていることで、サンプルが特定の人口集団であることが明確に認識できる。その集団の配偶率も自然な形で算出でき、対象を限定していることで母集団との誤差も小さく抑えることができる。このような追加的な情報の存在は、モデルにしいた仮定の妥当性、および仮定の制約を減じることを可能にする。

本稿では残存サンプルのバイアスを検証するモデルに、「国勢調査」の情報を取り込む方策を考える。近年、回収率の低下や費用調達の困難など、社会調査をとりまく環境は厳しくなっている。既存の大規模統計の活用可能性を探ることは、追加的な調査コストを発生させることなしに当該調査の信頼性を検証する手段として有用と考えられる。

## 2. 問題の所在

JPSCは特定の出生コーホートを追跡している調査である。出生コーホートはデモグラフィックな集団であるため、その集団の属性値として婚姻率や配偶率などの人口統計が自然な形で算出できる。対象者の脱落はこれらの人口統計に影響を与える。脱落が結婚に関して引き起こす問題でいえば、(a) 結婚という事象そのものにかかわる問題と (b) 残存しているサンプルについて、その配偶状態の構成に歪みが生じる問題とがある。人口統計的な観点からみると、(a) は脱落が引き起こす人口動態への影響、(b) は配偶率に人口動態への影響である。本稿では (b) について検証していくが、以下には両者について簡単に触れておく。

(a) について、脱落が発生する第一義的な影響は、結婚というイベントの観測数が減少することである。観測数の減少は推定値の誤差を大きくす

るため、結婚に焦点をあてた分析を行う際には大きな問題となる。もとより、調査期間中に結婚を経験する対象はサブサンプルであるため、その大きさは全対象に比べて限定されている。JPSCでは、すべてのコーホートの事象を合算しても、新規婚姻数が100件を超える調査年はほとんどない。このことを考慮すれば、脱落1件あたりの情報損失量は大きいといえる。結婚で脱落する層、残存している無配偶層になんらかの偏りが発生している場合は、結果として観測される結婚の情報にもバイアスが発生する。坂本（2006）では、JPSCにおいてそのバイアスの存在が指摘されている。

(b) については、結婚に伴う脱落が系統的に発生していれば、残存サンプルの配偶率は母集団に比べて低めに出る可能性がある。新たに結婚した人数を有配偶の対象として、過小にしか積み増さないためである。結婚は毎年発生する事象であるため効果は累積し、母集団と残存サンプルにおける配偶率の乖離は経年で大きくなる。その場合、配偶率、および配偶状態に強い影響を受ける事象については、異なる調査時点の集計結果を単純には比較できないことになる。

ただし、配偶状態に変化がない層でも恒常的に脱落は発生しており、その絶対数は結婚に伴い脱落する層より多い。また、結婚した人たちもすべてが脱落しているわけではない。結婚に伴い一定数の脱落が発生することは不可避であったとしても、脱落している層の配偶構成と残存している層の配偶構成が同じであれば、バイアスの発生は最小限に抑えられる。なお、配偶率は変化ではなく水準を示す指標であるため、残存しているサンプルの配偶率に対するカウンターパートは、各時点で脱落した層の配偶率ではなく、各時点におけるそれまでに脱落した対象全体の配偶率である。

このような人口静態面での問題は、生涯未婚率が高まっている今日では、より深刻な問題とも考えられる。未婚率が低い状況であれば、残存している層、脱落していった層にかかわらず、最終的にはほとんどの対象者が無配偶から有配偶へ移行する。また、結婚適齢期の規範が存在していた状況では、短い期間で配偶状態の変化が落ち着くこ

とが期待できた。未婚率の上昇はこれらの楽観的予想を成り立たせづらくさせ、JPSCでは個々人の配偶状態、およびそれらを集計化した配偶率は、時間で変化し、かつ調査時点に強く依存する変数となっている。さらに言えば、離婚率の上昇は有配偶から無配偶への可逆性を生じさせ、事態をより深刻なものにさせている。調査に残っている層と脱落していった層に偏りがなく、常にモニタリングする必要性が生じている。

残存したサンプルについて、配偶状態以外の属性にバイアスが生じている可能性はあるが、今回は、配偶状態の構成のみに着目する。

### 3. アプローチ

本稿では2期間の配偶状態、およびその遷移に焦点を合わせる。関心のある確率変数は、対象者の2期間の配偶状態と脱落の有無である。よって、それら3つの状態空間の間の関係を考える。確率変数は、以下のように設定する。 $W_{it}$ は対象者*i*の第*t*期の配偶状態を示し、有配偶には1、無配偶には0を割り当てる。 $Z_i$ は第2期の観測を示すインディケータであり、2期目に対象者*i*が残存している場合は1、脱落している場合は0を割り当てる。対象を集計し、同時確率は、

$$P(W_1=w_1, W_2=w_2, Z=z) \quad \text{for } w_1, w_2, z \in \{0,1\}$$
の形で表せる。それぞれ2値の値をとる確率変数なので、値の組み合わせに応じて8つのセルを持つ分割表が構成される。このうち $Z=1$ となる4つのセルの確率は残存サンプルのデータから直接計算できる。ただ、脱落した $Z=0$ については、 $P(Z=0)$ 、 $P(W_1=w_1, Z=0)$ までしか情報がわからず、4つのセルの確率を識別するには何らかの制約を設定する必要がある。

本稿では、Hirano et al. (1998, 2001) のアプローチに準拠して問題の解決を図る。Hirano et al. では、上記の同時確率を下記の確率でパラメトライズしておいている。

$$q_{wz} = P(W_2=1 \mid W_1=w, Z=z),$$

$$r_{wz} = P(W_1=w, Z=z).$$

$q$ は第1期の状態、および脱落の有無で条件づけ

ているため、この再パラメトライズは配偶状態の変化を描写する上でも有用である。例えば、 $q_{01}$ と $q_{00}$ を比較することは、脱落の有無で結婚する確率が異なるかを比較することになる。上記の制約については、Hirano et al.はある特定の性質をもつ関数 $g$ で、各期の配偶状態とそれらで条件付けた脱落に関する確率をリンクさせている<sup>1)</sup>。

$$P(Z=1 \mid W_1=w_1, W_2=w_2)$$

$$= g(a_0 + a_1 \cdot w_1 + a_2 \cdot w_2). \quad (1)$$

(1)式で先見的に $a_2=0$ とおいた場合、2期目に脱落しているかどうかは第1期の配偶状態だけに依存するモデル、いわゆるMissing At Random (MAR)の仮定をおいたモデルと同値になる<sup>2)</sup>。MARの場合は2期目の配偶状態は脱落の有無に依存しないため、 $q_{00}=q_{10}$ 、 $q_{10}=q_{11}$ と脱落の有無にかかわらず、遷移確率は同じものとなる。

また(1)式で $a_1=0$ とした場合は、第2期の配偶状態だけに依存するモデルとなる。脱落分析の文脈では、脱落の発生が脱落時の情報にも依存している状況をnon-ignorableな状況というが、 $a_1=0$ の場合はそのクラスのうちHausman and Wise (1979)で提示されたモデル(HWモデル)と同値となる。

HWについても、観測されたデータだけから計算により算出できる。 $a_1 \neq 0$ 、 $a_2 \neq 0$ の場合も、non-ignorableなケースである。Hirano et al.は第1期と第2期の情報をseparableに足し合わせていることから、このモデルをAdditive Non-ignorable model (ANモデル)と呼んでいる。ただ、ANモデルの場合は、すべての確率を識別するためには、第2期のサンプル全体の配偶率、 $P(W_2=1)$ の周辺確率の情報がさらに必要となる。Hirano et al.ではリフレッシュメント・サンプル<sup>3)</sup>の情報を母集団の情報とみなし、その値を $P(W_2=1)$ として外挿することで問題の解決を図っている。本稿では周辺確率を「国勢調査」から算出した配偶率を用いて外挿することにする。

### 4. 推計結果

JPSCには4つの追跡コーホートがあり、今回はそれら4つのコーホートについて、それぞれ推計を



図表-1 各コーホートの配偶状態

	コーホートA	コーホートB	コーホートC	コーホートD
対象数	1,500	500	836	636
初回調査時の年齢	24～34歳	24～27歳	24～29歳	24～28歳
初回調査時の配偶率	66.8%	40.2%	42.0%	34.3%
残存率（全体）	83.7%	73.2%	70.3%	75.9%
残存率（初回有配偶）	86.3%	77.6%	74.6%	82.1%
残存率（初回無配偶）	78.3%	70.2%	67.2%	72.7%
配偶状態の変化（100%）				
有配偶→有配偶	56.0%	29.4%	28.7%	26.4%
有配偶→無配偶	1.7%	1.8%	2.6%	1.7%
無配偶→有配偶	9.3%	15.6%	12.3%	15.4%
無配偶→無配偶	16.7%	26.4%	26.7%	32.4%
有配偶→脱落	9.1%	9.0%	10.6%	6.1%
無配偶→脱落	7.2%	17.8%	19.0%	17.9%

行った。コーホートは、出生順にコーホートA（1959～1969年生）、コーホートB（1970～1973年生）、コーホートC（1974～1979年生）、コーホートD（1980～1984年生）と呼称する<sup>4)</sup>。配偶状態は初回調査年度と4年後の値を用いた<sup>5)</sup>。外挿する「国勢調査」データは、各コーホートとも調査開始から4年目の年に最も近い実施年の「国勢調査」を用いた。コーホートAは1995年、Bは2000年、Cは2005年、Dは2010年のデータを使用している。「国勢調査」からは、各コーホートの4年後の年齢幅に対応する配偶率を算出した。比較するデータ間には調査時点のずれがあるため、配偶率を計算するにあたっては「国勢調査」の年齢別の人口構成比が各コーホートの初回抽出時の年齢構成比に沿うよう割り戻している。

図表-1には、各コーホートの基本情報と初回調査時から4年後の配偶状態の変化、および脱落の状況を示した。初回調査時の配偶率は最も早生のコーホートAが66.8%と高く、後生のコーホートであるDが最も低い34.3%となっている。コーホートAは、他のコーホートに比べて20歳代の未婚率が低い世代であることに加え、初回調査時に30歳代前半の層をサンプルに含めたことが大きく影響している。コーホートAは、サンプルの年齢構成が他のコーホートと大きく異なっていることに注意を要する。コーホートBとコーホートCはそれぞれ40%強の配偶率である。Cの方がBに比

べて後生のコーホートではあるが、CではBより28歳と29歳の2歳ほど年齢幅が広い。

初回時に抽出した対象者のうち、4年後も調査に参加していた人の割合を示す残存率は、コーホートAでは83.7%と、その他後生のコーホートよりも比較的高い値を示している。最も残存率が低いコーホートはCで、その値は70.3%である。図表-1において、その下2行の残存率は、初回時の配偶状態別に算出した残存率である。どのコー

ホートでも初回時に有配偶の方が残存率は高く、無配偶の方が残存率は低くなっている。裏を返せば、無配偶の対象者の方が脱落しやすいことを意味している。いずれのコーホートでも、両者の間には7～10%程度の開きがある。

配偶状態の変化は、初回調査時の状態から4年後にどの状態に変化したかを示したものである。脱落を含めた6つの経路について、それぞれの割合を示している。コーホートAでは、「有配偶→有配偶」が過半数を占める。他のコーホートでは配偶率の低さを反映して、無配偶を起点にする状態遷移の割合が大きい。特に後生コーホートほど「無配偶→無配偶」の相対的なウエイトが増している。配偶変化で特に留意すべきは、いずれのコーホートとも「無配偶→有配偶」の割合も相対的な位置を占めていることである。結婚に伴う脱落が含まれる「無配偶→脱落」と比べても大きな開きはない。つまり、結婚を経験する、ほとんどすべての対象者が脱落しているわけではない。

図表-2は、前節の $r_{wz} = P(W_1=w, Z=z)$ と $q_{wz} = P(W_2=1 | W_1=w, Z=z)$ を計算したものである。例えば、 $r_{00}$ は初回調査時に無配偶で4年後には脱落していた人の割合となる。また、 $q_{01}$ 、 $q_{11}$ は4年後にも脱落していない対象者について、初回調査時の配偶状態で条件づけた遷移確率である。 $q_{01}$ は初回調査時に無配偶だった対象者が、4年後に有

図表-2 残存の有無、配偶状態による同時確率・条件付き確率

	4年後・脱落	4年後・残存	$q_{w1}$
コーホートA			
初回・無配偶	$r_{00}=0.072$ (0.007)	$r_{01}=0.216$ (0.011)	$q_{01}=0.359$ (0.024)
初回・有配偶	$r_{10}=0.091$ (0.007)	$r_{11}=0.576$ (0.013)	$q_{11}=0.971$ (0.006)
コーホートB			
初回・無配偶	$r_{00}=0.178$ (0.017)	$r_{01}=0.420$ (0.022)	$q_{01}=0.371$ (0.033)
初回・有配偶	$r_{10}=0.090$ (0.013)	$r_{11}=0.312$ (0.021)	$q_{11}=0.942$ (0.019)
コーホートC			
初回・無配偶	$r_{00}=0.190$ (0.014)	$r_{01}=0.390$ (0.017)	$q_{01}=0.316$ (0.026)
初回・有配偶	$r_{10}=0.106$ (0.011)	$r_{11}=0.313$ (0.016)	$q_{11}=0.916$ (0.017)
コーホートD			
初回・無配偶	$r_{00}=0.179$ (0.015)	$r_{01}=0.478$ (0.020)	$q_{01}=0.322$ (0.027)
初回・有配偶	$r_{10}=0.061$ (0.010)	$r_{11}=0.281$ (0.018)	$q_{11}=0.939$ (0.018)

( )内は標準誤差

配偶になっている確率であり、 $q_{11}$ は初回調査時に有配偶だった対象者が両期とも有配偶である確率である。1からそれぞれの値を引いたものは4年後に無配偶の状態にある確率となる。これらの数値は図表-1の値の見方を変えただけのものであるので再度個々の値への言及はしないが、 $q_{11}$ についてはコーホートAに比べて他のコーホートの値が低くなっており、後生世代での離婚率の高まりがJPSCのデータにも反映されていることがうかがえる。図表-2のすべてのセルの値は、図表-1の結果を再集計することで計算できる。

図表-3は、脱落した層で予想される配偶状態の変化を、各モデルに基づいて推計した結果である。ANモデルでは前節の(1)式にロジットリンクを仮定した。 $q_{00}$ は脱落した層で予想される「無配偶→有配偶」の確率であり、 $q_{10}$ は「有配偶→有配偶」の確率である。 $q_{10}$ と $q_{00}$ について、それぞれを第1期の配偶状態に依存する仮定をおいたMARモデルと4年後の第2期の状態に依存するHWモデルで比較すると、いずれのコーホートでもMARモデルの方の値が大きく、HWの方が小さくなっている。MARモデルの $q_{00}$ の値は、脱落がない場合の遷移確率と同値である。対して、4年後の配偶

率を基準にしたHWのモデルでは、 $q_{01} > q_{00}$ となり、脱落していない層より脱落した層の方で「無配偶→有配偶」の遷移確率が低いことになる。つまり、HWモデルの仮定では、無配偶の脱落層は脱落していない層に比べて、4年後も無配偶の状態にある傾向が示唆されている。

一方、有配偶については脱落した層の方で「有配偶→有配偶」の遷移確率が低く、離婚による脱落傾向の存在を示唆している。また、「国勢調査」の配偶率を外挿したANモデルでは、MARとほぼ同じ結果を示している。図表-1の「無配偶→脱落」にこの $q_{00}$ をかけた値が、対象全体に占める脱落した「無配偶→有配偶」の推計値になる。決して無視できる値ではないものの、いずれのモデルの $q_{00}$ を用いたとしても、観測できた結婚経験層の割合、すなわち図表-1の「無配偶→有配偶」の値に比べて小さな値となっている。

各コーホートの下段にある配偶率は4年後の配偶率の推計値(=  $P(W_2=1)$ )を示したものである。周辺確率であるため、前節の表記に従えば、 $P(W_2=1) = \sum_{wz} r_{wz} q_{wz}$ と書け、MARとHWの両モデルについては、図表-2、図表-3の値を用いて直接計算できる。ANモデルの配偶率は「国勢調査」から算出した同年齢層の配偶率である。この値を母集団の配偶率と考えると、他のモデルで推計した配偶率の値を評価できる。どのモデル、およびどのコーホートの配偶率とも、それが母集団の値と等しいとする仮説を標準的な有意水準では棄却できない。つまり、MAR、HWいずれの仮定をおいたとしても、推計される配偶率にバイアスが発生しているとは言えない。コンプライト・データ(初年度と4年後の両方で観察される対象に限定したもの)の配偶率、あるいは今回の分析では同義な4年後に残存していた対象だけのデータを用いた配偶率でもこの仮説は棄却できない。

配偶率の値で評価する限り、仮定が妥当でないと判断されるモデルはないため、いずれかのモデルを他のモデルより積極的に採用する根拠はない。そのため脱落した層の配偶変化の遷移確率の値について断言はできないが、それぞれのモデルの $q_{01}$ と $q_{00}$ の値の比較から判断すれば、脱落して

図表-3 モデルに基づく条件付き確率と周辺確率の推計

	コンプリート・データ	MAR	HW	AN
コーホートA	-	$q_{00}=0.359$	$q_{00}=0.193$	$q_{00}=0.328$
	-	(0.024)	(0.021)	-
	-	$q_{10}=0.971$	$q_{10}=0.917$	$q_{10}=0.965$
	-	(0.006)	(0.031)	-
	配偶率 = 0.781 (0.012)	配偶率 = 0.782 (0.017)	配偶率 = 0.751 (0.017)	配偶率 = 0.765 -
コーホートB	-	$q_{00}=0.371$	$q_{00}=0.230$	$q_{00}=0.343$
	-	(0.033)	(0.057)	-
	-	$q_{10}=0.942$	$q_{10}=0.857$	$q_{10}=0.934$
	-	(0.019)	(0.057)	-
	配偶率 = 0.615 (0.025)	配偶率 = 0.601 (0.031)	配偶率 = 0.568 (0.029)	配偶率 = 0.595 -
コーホートC	-	$q_{00}=0.316$	$q_{00}=0.194$	$q_{00}=0.344$
	-	(0.026)	(0.047)	-
	-	$q_{10}=0.916$	$q_{10}=0.810$	$q_{10}=0.924$
	-	(0.017)	(0.039)	-
	配偶率 = 0.583 (0.020)	配偶率 = 0.567 (0.023)	配偶率 = 0.533 (0.023)	配偶率 = 0.571 -
コーホートD	-	$q_{00}=0.322$	$q_{00}=0.166$	$q_{00}=0.343$
	-	(0.027)	(0.060)	-
	-	$q_{10}=0.939$	$q_{10}=0.837$	$q_{10}=0.947$
	-	(0.018)	(0.041)	-
	配偶率 = 0.551 (0.023)	配偶率 = 0.533 (0.026)	配偶率 = 0.499 (0.025)	配偶率 = 0.537 -

( )内は標準誤差  
ANの配偶率は「国勢調査」より算出

いる層の方が「無配偶→有配偶」による遷移が大きいことを示す結果は得られていない。つまり結婚に伴う脱落が発生しているとしても、それは無配偶対象者の脱落理由の一つにすぎず、残存サンプルの配偶率に影響を与えるほどではないといえよう<sup>7)</sup>。「結婚」という調査の拒否理由は対象者の側にも、データを収集する側にも認識されやすいため、全体の脱落の中で過大に評価されてきたのかもしれない。実際、未婚率が高い後生コーホートの方が早生のコーホートより高い脱落傾向にあることは、「結婚」を中心にして対象者の脱落を説明できなくなってきたことを意味していよう。

## 5. まとめ

本稿では、対象者の脱落が残存サンプルの配偶構成にどのような影響を与えているか検証してき

た。JPSCでは結婚を理由とした調査の継続拒否が一定数、存在している。このことが、残存サンプルの配偶率にバイアスを生じさせていることが予想された。しかし、各追跡コーホートの調査開始から4年後の状況を検証したところ、残存サンプルと「国勢調査」の配偶率との間には、大きなズレは観測できなかった。分析からは、1) 無配偶の対象者は結婚の有無にかかわらず、もとより脱落しやすいこと、2) 脱落した無配偶対象者の多くが4年後も無配偶の状態を保っていることも示唆されている。結婚に伴う脱落はこれらの効果と相殺しあい、結果として残存サンプルの配偶率にはバイアスが生じていないものと考えられる。4年のスパンに関していえば、コンプリート・データを用いても母集団の配偶率との間に大きなズレは観測できなかったことの意義は大きい。配偶状態および、配偶状態のみに影響を受ける変数の集計に

関しては、残存サンプルのみのデータを用いることについてさほど過敏になる必要はないといえる。

ただし、今回の結果は残存サンプルの配偶率が数字として帳尻が合っていることを示したにすぎない。調査に残っている有配偶、無配偶それぞれの層について、年齢、学歴、所得といった他の属性レベルでもバイアスが発生していないことを保証するものではない。無配偶の方が脱落しやすい傾向にあることは、結果として早婚の対象者がサンプルに残りやすいことを意味している。また、無配偶の高い脱落傾向は、残存率が対象者の出生年によって大きく異なる可能性も示唆する。JPSCは調査の対象として20歳代後半の女性を抽出しているが、結婚に関してはこの年齢層で1歳の差が持つ意味は大きい。24歳と29歳では初回時に結婚している確率も大きく違い、無配偶の場合は今後、結婚までに要する平均期間も異なると考えられる。つまり、コーホート内では最も後生の対象者ほど脱落が起こりやすいことが予想される。

さらに、今回の分析は単一コーホートを基準にした検証であることにも留意する必要がある。JPSCを用いた分析は、A、B、C、Dすべてのコーホートのデータを結合して解析を行っているものが多い。本来的には、各コーホートは調査期間も年齢構成も異なる別個の調査の追跡集団である。異なるデータを単純に結合してよいかどうか、その判断は今回の結果からは何もわからない。コーホート間による脱落率の差異は、世代を反映した効果か時代性を反映したものか、それとも抽出時の年齢構成の影響によるものなのか。それらを明らかにした上で、結合の方策を考えていく必要がある。

このようなコーホート内、コーホート間における脱落傾向の差異の検証は今後の研究上の課題であるとともに、それらの情報の開示はデータの提供を行っていく際の責務ともいえる。

#### 注

- 1) 関数  $g$  の満たすべき性質は狭義単調増加関数、および  $g(-\infty) = 0$ ,  $g(+\infty) = 1$  である。この性質下で、制約式は  $q_{wz}$  と  $r_{wz}$  を用いて各係数  $a$  について closed-form の形に書き直すことができる。

- 2) MAR および non-ignorable の詳細については、例えば、Little and Rubin (2002) を参照。
- 3) リフレッシュメント・サンプルとは、サンプルの一部を入れ替えながら継続するパネル調査(ローターティング・パネル調査)での新規追加サンプルや、経年で磨耗した際に補充を行ったサンプルのことを指す。その抽出時点での母集団情報を反映していると考えられる。
- 4) 各コーホートとも「国勢調査」の年齢構成、配偶関係に基づき、抽出を行っている。ただし、コーホート B、C、D は無配偶の単身比率までを考慮して対象数を割り付けているが、コーホート A では考慮していない。
- 5) 4年に設定したのは、最も後生のコーホートである D の調査期間が、第20回調査時の時点で4年目だからである。
- 6) JPSC では調査に回答している対象者については、配偶状態の情報欠損(無回答)はない。
- 7) 4年というスパンでみた配偶変化についての結果であり、脱落のタイミングは考慮していない。脱落した対象者で配偶変化があった者について、脱落するタイミングが結婚(離婚)の前後に集中していることは考えられる。

#### 文献

- 坂本和靖, 2006, 「サンプル脱落に関する分析——「消費生活に関するパネル調査」を用いた脱落の規定要因と推計バイアスの検証」『日本労働研究雑誌』551: 55-70.
- 村上あかね, 2003, 「なぜ脱落したのか——『消費生活に関するパネル調査』における脱落分析」財団法人家計経済研究所編『家計・仕事・暮らしと女性の現在——消費生活に関するパネル調査 第10年度』国立印刷局, 115-122.
- Hausman, J. and D. Wise, 1979, "Attrition Bias in Experimental and Panel Data: Gary Income Maintenance Experiment," *Econometrica*, 45: 455-473.
- Hirano, K., G. Imbens, G. Ridder, and D. Rubin, 1998, "Combining Panel Data Sets with Attrition and Refreshment Samples," NBER Working Paper No. 230.
- Hirano, K., G. Imbens, G. Ridder, and D. Rubin, 2001, "Combining Panel Data Sets with Attrition and Refreshment Samples," *Econometrica*, 69: 1645-1659.
- Little, R. J. A. and D. B. Rubin, 2002, *Statistical Analysis with Missing Data*, Second Edition, New York: John Wiley & Sons.

さかぐち・なおふみ 公益財団法人 家計経済研究所 次席研究員。主な論文に「パネルデータからみた女性の仕事・結婚・出産」(『季刊家計経済研究』92, 2011)。労働経済学専攻。(sakaguchi@kakeiken.or.jp)