JGSS データによる父学歴の欠損メカニズムの分析
―〈わからない〉と〈無回答〉の違い―

保田 時男

大阪商業大学総合経営学部

An Analysis on the Missing Mechanism of Fathers' Education in JGSS:

The difference between DK and NA

Tokio YASUDA

For some studies of intergenerational social mobility, you sometimes need to know education experienced by fathers of survey respondents. But, fathers' education is one of the questions that often occur much missing data. Missing data may cause some biases for statistical analyses. You need to specify the missing mechanism of fathers' education in order to resolve the problem. Some studies examined the mechanism, but the analyses by those studies only made unclear conclusion. In this paper, the missing mechanism of fathers' education was made clear by the analysis of JGSS data. The reason for which the analysis was possible was that JGSS had made different codes indicating each of 'Do not know' and 'No answer.' The result of the analysis indicated that respondents' cohort and education had significant effects on the missing of fathers' education. The result implied that the missing of fathers' education would be 'ignorable' when you control those variables.

Key words: JGSS, missing data, intergenerational social mobility

世代間社会移動の研究においては、しばしば調査対象者の父親の学歴を知る必要がある。しかし、父学歴は比較的欠損することが多い質問項目であり、統計的分析の結果に偏りを与える原因になる。その問題を解消するためには、父学歴の欠損メカニズムを特定する必要がある。そのため、これまで SSM 調査のデータを用いて、父学歴の欠損メカニズムを検討する研究がなされてきたが、その結果は曖昧なものであった。本稿では、JGSS データを用いて、これまで曖昧であった父学歴の欠損メカニズムを明らかにしている。それが可能であったのは、「わからない」と「無回答」が区別してコーディングされているためである。分析の結果、調査対象者本人のコウホートと学歴が欠損の原因として効果を持つことが明らかになった。この結果は、これらの変数を統制すれば父学歴の欠損が「無視できる」ことを意味している。

キーワード：JGSS、欠損データ、世代間社会移動

１．目的

1.1　欠損データによる偏り

本稿は、社会調査における欠損データ（missing data）の問題を扱っている。欠損データとは、調査計画上は得られるべきであったけれども、実際には得ることができなかったデータのことである。具体的には、調査対象者の不在や協力拒否によって 1 ケース分の回答がまるごと得られなくなってしまうこと（unit nonresponse）や、一部の質問への回答が拒否や不注意などにより無回答となってしまうこと（item nonresponse）により発生するものが、社会調査における欠損データである。近年は特に調査環境の悪化により、欠損データの増加が問題となっている。

欠損データが多く発生すると、分析結果に偏りが発生することがありうる。つまり、欠損のないデータだけを扱って分析をすることによって、データに欠損がなければ可能であったはずの分析結果からは、かけ離れた結果が得られてしまうことがありうる。社会調査データの統計的な分析は、対象者がランダム・サンプリングにより選ばれていることを前提にしているが、欠損データの発生はこの前提を破壊し、データに想定外の偏りを生んでしまうことがあるのである。

社会調査の統計的分析では、当初この問題に対して非常に単純な対処法を取ってきた。それは、分析に関わる変数について一部でも欠損しているケースは、分析の対象外とするという方法（listwise deletion）である。また、やや工夫をこらして、欠損データを平均値で置き換える方法が取られることも多かった。しかし、これらの方法は便宜的なものなので、欠損データによる偏りの問題を何ら解決するものではない。

これに対して、ここ 2、30 年の間に、社会調査の方法論者や応用統計学者の間で欠損データの問題を正面から解決しようとする取り組みが盛んになってきた。その動きは大きく 2 つに分けて考えることができる。1 つは、そもそもの欠損データの発生を可能な限り防ごうとする動きである。つまり、社会調査における無回答を減らすための研究である。もう 1 つの動きは、発生した欠損データに対して統計的処理を施すことにより、その偏りを事後的に補正しようとする研究である。この種の研究は特に社会調査における欠損データだけを念頭に置いたものではなく、様々な実験・調査で発生した欠損データへの対処法として開発されてきた。最近の展開については、Groves et al.（2001）などが参考になる。

本稿では、これらのうち後者の研究展開による成果を利用して、社会調査における 1 つの具体的な欠損データの偏りを補正することを試みる。この種の補正は、他に得られている情報の活用が不可欠なので、unit nonresponse の分析には基本的にそぐわない。分析結果の偏りに与える重要性はどちらも同じではあるが、本稿では item nonresponse のみを扱うことをあらかじめ断っておく。

1.2 学歴移動表における欠損データの問題

本稿で具体的に取り上げる問題は、世代間の学歴移動表における欠損データである。世代間の学歴移動表とは、調査対象者の最終学歴がその親の学歴によってどのような影響を受けるのか、その社会移動の様子を表したクロス表のことである。世代間の学歴移動は、教育機会の平等性や開放性を判断する材料として、しばしば階層論的な研究で問題にされる。親の世代の学歴は、父親の学歴と母親の学歴のそれぞれが取り上げられうるが、親世代の階層を表す指標として父親の属性が重視されることが多いので、本稿でも、世代間の学歴移動で、父親の学歴から本人の学歴への移動を指すことにする。

父学歴と本人学歴のうち、本人学歴の欠損は一般的にほとんど発生しないのに対して、父学歴の欠損は比較的多く発生する。その割合は戦後一貫して上昇傾向にあり、例えば1955 年から 10 年ごとに行われている SSM 調査では、1995 年までの間にその欠損率は、4.7%、9.9%、8.9%、14.8%、18.3%と上昇を続けている。このような父学歴への無回答の発生メカニズムを考察し、それによる偏りを補正することは、階層論的な研究の結果を適切に解釈する上でますます重要になっている。

この問題に対して、保田（2000a; 2000b）はログリニア・モデルの応用による分析を行っている。まず、1985 年 SSM 調査における父学歴の欠損が分析された（保田,2000b）。その分析では、父学歴の欠損が発生する原因として、対象者本人の性別と学歴、および欠損している父学歴そのもの(1)の 3 つを候補とし、欠損原因メカニズムを特定することが試みられた。

その結果、以下の 2 つのモデルがデータに適合的であることが分かった。

（1）「父学歴」と「本人学歴」が欠損の発生原因
（2）「本人性別」と「本人学歴」が欠損の発生原因

特に、（1）のメカニズムが正しければ、学歴移動表の解釈に及ぼす影響が大きく、世代間の移動がより開放的な（親の学歴に子の学歴が規定されない）ものとして解釈されることがわかった。しかし、その一方で（2）が正しければ、学歴移動表の解釈に及ぼす影響はほとんどない。欠損データの分析においては、その情報が欠損しているがゆえに、欠損の発生メカニズムを断定することが難しく、いくつかの可能性が残ってしまう傾向がある。その影響力が大きく異なるこれら 2 つの欠損原因メカニズムのいずれが正しいのかを特定できなかった点が、保田（2000b）の不十分な点であった。

そこで、1995 年 SSM 調査のデータを利用して、ふたたび同様の分析が行われた（保田 2000a）(2)。ただし、こちらの分析では、以下の 2 点で工夫が凝らされた。第 1 に、性別によりそのメカニズムが異なる可能性があるため、男女別に欠損原因の特定が試みられた。第 2 に、1985 年調査の分析結果に見られた父学歴の効果が世代による学歴構造の違いを反映したものである可能性があるために、10 歳刻みの年齢層が新たに欠損原因の候補と

して加えられた。第 2 の点について説明を補足しておこう。調査対象者の父親には、戦前から戦後に渡る幅広い時期の教育を受けた者が含まれている。その間、急激な高学歴化が進んだために、父親の学歴の程度はその出生コウホートに大きく規定されている。また、戦前の旧制学校教育から戦後の新制学校教育への切り替えが、父学歴の理解のし易さに及ぼす影響も大きいと考えられるので、この点でも父親の出生コウホートが重要と考えられる。父親の出生コウホートを直接統制することが望ましいが、父親の出生年は調べられていなかったために、調査対象者本人の年齢層でおおまかな代用を行ったということである。

その結果、以下の 2 つのモデルが析出された。

(1)「父学歴」と「本人学歴」が欠損の発生原因
(2)「本人年齢層」と「本人学歴」が欠損の発生原因

正確に述べると、対象者が女性の場合には、(2) のメカニズムのみが適合的と特定することができたが、対象者が男性の場合には、(1) と (2) のどちらのメカニズムが適切なのかを特定することができなかった。この場合もやはり、(1) のメカニズムの場合には学歴移動表の解釈に及ぼす影響が甚大である。その一方で、(2) のメカニズムの場合には、対象者の年齢層を統制した分析を行っていれば、解釈に及ぼす欠損データの影響はほとんどないことになる。この分析においても、父学歴の欠損原因メカニズムが特定できないという問題への解答は曖昧なまま残された。

さらにまた、その分析により確認された本人年齢層による効果は、60 代の高齢層（＝父親の出生コウホートが古いと想定される人々）において、欠損の発生率が下がるというものであり、あまり歯切れのよい解釈ができないものであった。欠損の原因メカニズムが特定できないだけではなく、そのモデルの内容の解釈についても課題が残された。

1.3 JGSS データを分析に用いることの利点

本稿では、近年行われた全国規模の社会調査である第 1〜4 回 JGSS のデータを用いて保田（2000a; 2000b）と同様の分析を行う。JGSS のデータを分析することには、SSM 調査の分析からは得られない利点が 2 つある。第 1 の利点は、そのサンプル数が膨大なことである。JGSS では 2000〜2003 年にかけて毎年調査が行われているので、調査時期のずれが少ないわずかな期間に 4 回もの調査データが取られている。これらの調査データを 1 つのデータとして累積すれば、そのサンプル数は 12,299 にもなる。サンプル数が多いことは、本稿の分析にとって非常に有効である。なぜならば、欠損データの分析では、その情報が欠損しているがゆえに、通常の分析に比べて少量のサンプルでは結論が出しにくいことがあるからである。実際に、保田（2000a; 2000b）は欠損の原因に複数の解釈が成り立ちうることを示しており、曖昧な結論を出している。JGSS の十分なサンプル数を用いれば、父学歴の欠損についてより明確な結論を出すことができると期待できる。

JGSS データを用いることの第 2 の利点は、より重要な点である。保田（2000a; 2000b）が分析に用いている SSM 調査（1985 年調査および 1995 年調査）では、調査対象者が「わからない」と回答したことによる欠損と、単純に回答がないことによる欠損が区別されずにコーディングされている。これに対して、JGSS データでは「わからない」と「無回答」を区別してコーディングがなされている。欠損データの分析においては、この区別は非常に重要である。なぜならば、「わからない」と「無回答」はいずれも欠損となるものであるが、その発生メカニズムは異なるものと予想されるからである。単純に考えるならば、答えるつもりがあるのだけれども情報を持っていないという場合には、「わからない」と答え、情報を持っているのだけれども答えたくないという拒否の表明がなされた場合には「無回答」となるものと考えられる。SSM 調査データの分析においては、これらが混在した欠損を 1 つのメカニズムで無理に説明する必要がある。これに対して、JGSS データの分析では、これらを別々のメカニズムで説明することができる。そのめ、今回の分析では、より正確に父学歴が欠損するメカニズムを把握することができると期待できる。

JGSS データのこれらの特徴を活かし、本稿では、父学歴の欠損原因メカニズムについて曖昧さのない解答を示すこと、および、その際に「わからない」と「無回答」とを区別したメカニズムを明らかにすることを目的とする。この目的は、JGSS データを保田（2000a; 2000b）と同様の分析方法で分析することにより達成できる。その方法の概略は後の 2.2 節で示す。

２．分析方法

2.1 データ

分析には、JGSS-2000、2001、2002、2003 のデータを累積して用いる。JGSS（日本版 General Social Surveys）は、大阪商業大学比較地域研究所が東京大学社会科学研究所と共同で実施した全国規模の社会調査であり、2000～2003 年の間、毎年 10～11 月に満 20～89 歳の男女を対象として継続的に行われたものである（今後も、およそ 2 年に一度のペースで継続される予定である）。

各年の有効回収率は 64.9%、62.4%、62.3%、51.5%であり、有効回答数は 2,893、2,790、2,953、3,663 であるので、これらを累積すると、12,299 ケースのデータが得られる。ただし、後述する条件により、実際に分析対象とするのは、このうち 10,537 ケースである。

調査年度の異なるデータを累積して用いることから問題が発生する可能性はある。しかし、意識を尋ねる質問項目とは異なり、学歴という事実を尋ねる項目への回答は、調査時点にほとんど左右されないことが期待される。また、調査時点が非常に近い年度ということもあり、データを累積することの問題性は小さいと判断した。

JGSS では、毎回、同じ方式で対象者本人の最終学歴およびその父親の最終学歴を尋ね

ている（同時に、配偶者の最終学歴と母親の最終学歴も尋ねている）。「あなたが最後に通った（または現在通っている）学校は次のどれにあたりますか。あなたの配偶者やご両親についてもわかる範囲でお答えください。なお、中退も卒業と同じ扱いでお答えください。」という質問文で、12 種類の学校＋「わからない」の中から面接調査により回答を得ている。

本稿で分析に用いる変数は、本人学歴、父学歴および、本人の性別、出生コウホートの 4 つである。出生コウホートは、1930 年ごろ（1926～1935 年）、1940 年ごろ（1936～1945 年）、1950 年ごろ（1946～1955 年）、1960 年ごろ（1956～1965 年）、1970 年ごろ（1966～1975 年）の 5 カテゴリとし、この範囲に入らない 1,698 ケースは分析の対象外とした。このようなコウホートを分析に用いるのは、1995 年 SSM 調査のデータを用いた保田（2000a）の分析との比較を容易にするためである。保田は、1995 年調査時点で対象者の年齢層を 60 代、50 代、40 代、30 代、20 代に分類して分析を行っている。この年齢層の分類は、上記のコウホートの分類とほぼ一致する。

本人および父親の学歴については、単純化のため高等・中等・義務の 3 つに分類して用いた。それぞれの分類に含まれる学歴は以下のとおりである。

高等……新制短大・高専・大学・大学院、旧制高校・専門学校・高等師範学校・大学・大学院

中等……新制高校、旧制中学校・高等女学校・実業学校・師範学校

義務……新制中学校、旧制尋常小学校（国民学校を含む）・高等小学校

本人の学歴について「わからない」者や「無回答」の者は、64 ケース（0.6%）とごく少数であったため分析対象から除いた。父学歴が「わからない」者や「無回答」の者は、本分析の中核となるので、当然、それぞれ 1 つのカテゴリとして扱った。父学歴が「わからない」者は 16.1%、父学歴が「無回答」の者は 3.6%おり、欠損が 2 割近くとかなり大きな割合を占めている。

これら 4 つの変数（*A*＝性別、*B*＝コウホート、*C*＝父学歴、*D*＝本人学歴）により構成される表 1 のクロス表が分析の対象となる。

2.2 ログリニア・モデルを応用した分析技法の概要

表 1 のクロス表に対して、父学歴の欠損原因メカニズムを明らかにするための分析技法を適用する。この種の技法には様々なタイプがあるが（保田, 2000b などを参照）、ここで用いるのは、ログリニア・モデルを応用した技法について、簡単な概要を示しておこう。より詳しい説明は、Fay（1986）、Baker & Laird（1988）、保田（2000b）などを参照してほしい。

欠損データの発生の仕組みに沿って表 1 の構造を解釈すると、表 1 は、本来得られるべ

表１　JGSS-2000～2003 の累積学歴移動表

		A: 性別							
		男性				女性			
		D: 本人学歴				D: 本人学歴			
B: コウホート	C: 父学歴	高等	中等	義務	計	高等	中等	義務	計
1926～1935	高等	33	12	2	47	23	48	5	76
	中等	36	43	9	88	16	74	11	101
	義務	76	216	338	630	28	231	396	655
	わからない	14 (8.6)	40 (12.5)	85 (18.6)	139 (14.8)	8 (10.7)	58 (13.7)	126 (22.3)	192 (18.0)
	無回答	4 (2.5)	9 (2.8)	24 (5.2)	37 (3.9)	0 (0.0)	12 (2.8)	28 (4.9)	40 (3.8)
	計	163	320	458	941	75	423	566	1064
1936～1945	高等	73	32	4	109	61	71	3	135
	中等	68	81	24	173	42	119	32	193
	義務	95	276	250	621	52	297	314	663
	わからない	21 (7.9)	76 (15.8)	91 (23.2)	188 (16.5)	13 (7.6)	127 (19.6)	119 (24.2)	259 (19.8)
	無回答	10 (3.7)	17 (3.5)	24 (6.1)	51 (4.5)	3 (1.8)	33 (5.1)	24 (4.9)	60 (4.6)
	計	267	482	393	1142	171	647	492	1310
1946～1955	高等	86	16	1	103	118	51	5	174
	中等	128	108	8	244	103	164	18	285
	義務	120	332	114	566	94	386	136	616
	わからない	40 (10.5)	106 (18.3)	53 (29.1)	199 (17.4)	38 (10.5)	145 (19.0)	59 (26.6)	242 (17.9)
	無回答	6 (1.6)	17 (2.9)	6 (3.3)	29 (2.5)	10 (2.8)	18 (2.4)	4 (1.8)	32 (2.4)
	計	380	579	182	1141	363	764	222	1349
1956～1965	高等	112	20	1	133	120	37	2	159
	中等	119	104	4	227	168	166	4	338
	義務	108	185	28	321	100	245	21	366
	わからない	25 (6.7)	68 (17.7)	11 (25.0)	104 (12.9)	45 (10.1)	116 (19.9)	9 (24.3)	170 (15.9)
	無回答	11 (2.9)	8 (2.1)	0 (0.0)	19 (2.4)	13 (2.9)	19 (3.3)	1 (2.7)	33 (3.1)
	計	375	385	44	804	446	583	37	1066
1966～1975	高等	117	25	0	142	149	27	1	177
	中等	158	139	6	303	211	185	6	402
	義務	59	134	23	216	69	166	9	244
	わからない	24 (6.6)	59 (16.1)	9 (23.1)	92 (11.9)	28 (6.1)	68 (14.6)	6 (27.3)	102 (10.8)
	無回答	8 (2.2)	10 (2.7)	1 (2.6)	19 (2.5)	4 (0.9)	19 (4.1)	0 (0.0)	23 (2.4)
	計	366	367	39	772	461	465	22	948

注：父学歴の「わからない」「無回答」欄の括弧内は、それぞれの性・コウホート・本人学歴内における比率（%）を示している。

きであった（父学歴が欠損していない）データと、父親の欠損状況を表す変数を組み合わせた表のうち、一部をマージンしたものと考えることができる。つまり、仮にすべての対象について父学歴が欠損していないと考えるならば、そのデータは 2×5×3×3 のクロス表で表現できるはずである。ところが、実際には父学歴の一部が欠損している。その欠損

状況を変数 R で表すとすると、R は 3 つの値を取りうる変数である（欠損でない、わからない、無回答）。したがって、表 1 は（2×5×3×3）×3 のクロス表のうち、R＝2 または 3 の場合についてだけ、父学歴をマージンした表とみなせる。

ここで、マージンする前の（2×5×3×3）×3 のクロス表を想定し、その各セル度数を

$$f_{abcdr}\left(a=\begin{cases}1=\text{男性}\\5=\text{女性}\end{cases}, b=\begin{cases}1=1930\text{年ごろ}\\ \vdots \\5=1970\text{年ごろ}\end{cases}, c=\begin{cases}1=\text{高等}\\2=\text{中等},\\3=\text{義務}\end{cases} d=\begin{cases}1=\text{高等}\\2=\text{中等};\\3=\text{義務}\end{cases} \quad r=\begin{cases}1=\text{父学歴が欠損でない}\\2=\text{父学歴が「わからない」}\\3=\text{父学歴が「無回答」}\end{cases}\right)$$

で表すことにする。f_{abcd2} の値や f_{abcd3} の値は父学歴（C）が不明なので、当然データを集計するだけではわからない。分析者が定めるモデルの仮定に基づいて推定する必要がある。

このとき、分析者が定めるモデルとは、通常のログリニア・モデルで用いる変数 A, B, C, D 間の関連性（独立性・非独立性）を表現するモデルと、欠損の原因メカニズムを特定するモデルの合成モデルとして組み立てられる。本稿では、父学歴の欠損原因メカニズムを特定することを目的としているので、前者についてはバリエーションを設けずに、飽和モデルを想定することにする。一方、後者のモデルには複数のバリエーションを設け、その適合度を検討する。欠損の原因メカニズムを表すモデルは、A, B, C, D の各変数と変数 R との間の関連性モデルによって表現される。例えば、コウホート（B）と父学歴（C）の単独効果によって、欠損状況が決定されるというモデルは、通常のログリニア・モデルにおけるモデル表記法に従えば、[BR][CR]モデルと表すことができる。

モデルに基づいた各セル度数の推定値 $\hat{f}_{abcdr}$ は、EM アルゴリズム（Dempster et al., 1977; MacLachlan & Krishnan, 1997）と呼ばれる最尤推定法の一種により導き出される。EM アルゴリズムは、繰り返し計算によって、モデルの条件下で最もデータに適合的な推定値を導き出す。

そうして導き出された推定値からなるクロス表が、観察データと十分に適合的かどうかは、χ^2 値を用いた通常の適合度検定を行えばよい。ただし、このとき検定統計量の算出に用いるセル度数はやや変則的で、欠損を含むセルの度数はマージンされたものを用いる。つまり、最尤推定値からなる（2×5×3×3）×3 のクロス表の一部をマージンして、表 1 と同じ形のクロス表を作り、そこで適合度を確認する。本稿では、尤度比統計量 L^2 を χ^2 分布に近似する検定統計量として用いているので、その算出式は、

$$L^2 = 2\sum f_{abcd1} \bullet \log\left(f_{abcd1} / \hat{f}_{abcd1}\right) + 2\sum f_{ab\bullet d2} \bullet \log\left(f_{ab\bullet d2} / \hat{f}_{ab\bullet d2}\right) + 2\sum f_{ab\bullet d3} \bullet \log\left(f_{ab\bullet d3} / \hat{f}_{ab\bullet d3}\right)$$

となる。

複数の欠損原因モデルを想定し、それぞれの適合度を対比することにより、父学歴の欠損メカニズムを特定することができる。また、そのモデルの中で、ログリニア・モデルの

パラメータを確かめることにより、具体的にどのような変数でどのような値を持つことが、「わからない」や「無回答」の発生率を高めているのかを知ることができる。

３．分析結果

3.1 「わからない」と「無回答」の原因モデル

いま示したログリニア・モデルの応用により、実際に「わからない」と「無回答」が発生する欠損原因モデルを明らかにした。表 2 は、さまざまな欠損原因モデルの適合度について、主な結果を示したものである。この表は、それぞれの欠損原因モデルによって実際のデータを説明しようとした場合に、どの程度データに適合的な推定値を得ることができるかを表している。適合度検定の結果、推定値と実際のデータの間に有意な乖離がない（p 値が大きい）モデルが、欠損の発生を適切に説明できているモデルということになる。

表 2　主な欠損原因モデルの適合度

欠損原因モデル	自由度	L^2	p
[DR]	54	77.37	0.020
[BR]	50	282.75	0.000
[AR][CR]	52	88.77	0.001
[AR][BR]	48	275.97	0.000
[BR][DR]	46	38.71	0.768
[AR][BR][CR]	44	46.13	0.384
[AR][BR][DR]	44	35.72	0.809
[AR][CR][DR]	48	70.16	0.020
[BR][CR][DR]	42	32.93	0.840
[AR][BR][CR][DR]	40	32.42	0.797

この結果から明らかなように、データに適合的な欠損原因モデルの中で、もっとも単純なメカニズムでデータを説明できているのは、[BR][DR]モデルである。このモデルよりも単純なメカニズムを想定した[BR]モデルや[DR]モデルでは、適合的な推定値は得られないし、逆に[BR][CR][DR]モデルのように、より複雑なモデルを想定しても、適合度の大きな改善は見られない。よって、父学歴の欠損の発生は、調査対象者本人の出生コウホート（B）と学歴（D）によって十分に説明ができると結論付けることができる(3)。

では、本人の出生コウホートや学歴は、父学歴の欠損の発生とどのように関連しているのであろうか。表 3 は、[BR][DR]モデルにおける、ログリニア・モデルの各パラメータの推定値である。パラメータの読み方は、通常のログリニア・モデルと同様である。例え

ば、λ_{br} (1930, わからない)＝-0.14 は、1930 年前後のコウホートにおいて、父学歴に「わからない」と答える確率が、平均的な確率の exp(-0.14)＝0.87 倍と、小さいことを意味している。

表３　[BR][DR]モデルの欠損原因パラメータ

		λ	exp(λ)
λ_r	R = わからない	-0.06	0.94
λ_{br}	(B, R) = (1930, わからない)	-0.14	0.87
	(B, R) = (1940, 〃)	-0.07	0.94
	(B, R) = (1950, 〃)	0.18	1.20
	(B, R) = (1960, 〃)	0.08	1.08
	(B, R) = (1970, 〃)	-0.05	0.95
λ_{dr}	(D, R) = (高等, わからない)	-0.34	0.71
	(D, R) = (中等, 〃)	0.08	1.08
	(D, R) = (義務, 〃)	0.26	1.30
λ_r	R = 無回答	-1.57	0.21
λ_{br}	(B, R) = (1930, 無回答)	0.09	1.09
	(B, R) = (1940, 〃)	0.19	1.20
	(B, R) = (1950, 〃)	-0.19	0.83
	(B, R) = (1960, 〃)	-0.04	0.96
	(B, R) = (1970, 〃)	-0.05	0.96
λ_{dr}	(D, R) = (高等, 無回答)	-0.02	0.98
	(D, R) = (中等, 〃)	-0.03	0.97
	(D, R) = (義務, 〃)	0.06	1.06

「わからない」と「無回答」の発生メカニズムの特徴を、順に確認しよう。「わからない」の発生メカニズムは明確である。出生コウホートの影響については、1950 年ごろの世代を頂点として「わからない」が発生しやすく、それより前のコウホートや逆に後のコウホートでは、徐々に発生しにくくなる。また、本人学歴の影響については、学歴が低いほど「わからない」が発生しやすいという明らかな傾向が見られる。

これに対して、「無回答」は、戦前生まれ（1930 年ごろ～1940 年ごろ）のコウホートで多く発生する傾向にある。また、本人学歴は「無回答」の発生とはほとんど関連が見られない。本人学歴が影響するのは「わからない」の発生のみであり、「無回答」の発生には何ら影響を与えないという修正モデルについて、実際に適合度を調べて見たところ、適合

度検定の結果は、df＝48、L^2＝39.76、p＝0.795 となったので、[BR][DR]モデルと比べてほとんど適合度が変わらないことがわかった。したがって、「無回答」の発生に影響を与えているのはコウホートのみであるとする修正モデルを受け入れることができる(4)。

４．考察とまとめ

4.1 欠損メカニズムの解釈

ここで行った分析により、父学歴の欠損が発生する基本的なメカニズムが非常に明確になった。「わからない」の発生は、コウホートと本人学歴に規定され、「無回答」の発生はコウホートのみに規定されるという結果である。この節では、このモデルに対する妥当な解釈を検討する。

順序が逆になるが、まず「無回答」の発生メカニズムについて解釈する。本稿の 1.2 節において、「無回答」は父親の学歴を知っているが答えたくないという回答拒否を反映しているにちがいない、という見解を示した。しかし、分析結果はこの見解が間違いであったことを示唆している。なぜならば、1940 年ごろの出生コウホートの人々（調査時点で 60 歳前後の人々）が、特別に父親の学歴についてプライバシー意識を持ち、回答を拒否するとは考えられないからである。一般的に考えるならば、プライバシー意識はより新しいコウホートで広まっていると考えられる。

この分析結果から考えられる「無回答」の妥当な解釈は、おそらく父親の学歴がわからないのではなく、あてはまる選択肢がないことによる無回答であろう。戦前の教育制度は現在の制度よりも複雑なので、調査票の中に選択肢が存在しない種類の学校が多く存在する（例えば青年学校）。父親の学歴に当てはまる選択肢がなかった回答者は、「わからない」と回答することもできず、「無回答」になってしまったものと考えられる。戦後の世代（1950 年ごろのコウホート）において、「無回答」が減少する傾向にあるのは、戦前の制度を知らないがゆえにむしろ大まかな枠組みで回答ができるためと解釈できる。

次に、「わからない」の発生メカニズムについて解釈する。父学歴が「わからない」者は、なぜ 1950 年ごろの出生コウホートで多く発生する傾向にあるのであろうか。その理由については、保田（2000b）がすでに示している解釈をそのまま当てはめることが可能である。すなわち、この世代に父親の学歴がわからない者が多くなるのは、戦後に教育制度が大きく変化したためである。1950 年ごろ（1946～1955 年）生まれのコウホートは、戦後生まれのコウホートであり、戦前の教育制度を知る機会が少ない。同じことはより新しいコウホートについても言えることであるが、新しいコウホートの場合、父親も戦後の新制度の下で教育を受けている可能性が高まるので、戦前の制度を知らないことは、父学歴を回答する上で問題にならない。

1995 年 SSM 調査の分析（保田, 2000a）においては、年齢層＝コウホートをモデルの中に含めていながら、このような傾向がはっきりとは観察されなかった。その理由は、おそ

らく SSM 調査において、すべての欠損が混在したままコーディングされており、複数のメカニズムが混じり合ってしまったためと考えられる。本稿の分析から明らかなように、父学歴が「わからない」者は、戦後すぐの生まれのコウホートに多いのに対して、「無回答」の者は、戦前生まれのコウホートに多い。これらの効果の混在が、SSM 調査の分析結果をわかりにくくしていたものと考えられる。

本稿の分析結果において、父学歴が「わからない」者の発生に影響を与えていたもう 1 つの要因は本人学歴が低いことであった。この結果は、1985 年、1995 年 SSM 調査の分析結果の双方と一致している。保田（2000b）は、この結果を 2 通りに解釈していた。1 つは、自らの学歴があまり高くない者は、全般的に学歴への関心が低いために親の学歴についても無関心であり、よく把握していなかった、という解釈である。学歴が低いということは、学歴に依拠した生活スタイルを送っていないと予想されるので、親の学歴についても意識する機会が少ないはずである。もう 1 つの解釈は、親の学歴が高かったにもかかわらず自らの学歴が低いことを恥じていることによる秘匿の可能性であった。しかしながら、この解釈のためには欠損した父学歴自体が欠損の発生に影響しているモデルが適合している必要がある。父学歴が高いことが欠損の発生する可能性を高めていなければ、このような解釈はできないからである。SSM 調査の分析（保田, 2000b）においては、そのようなモデルが適合的である可能性が残されていたが、本稿のより精緻な分析においては、その可能性は否定された。したがって、やはり 1 つ目の解釈を妥当と考えるべきであろう。

4.2 出生コウホートの効果が持つ含意

この節では、分析の結果明らかになった出生コウホートが「わからない」の発生に対して持つ効果の含意について、2 つの注意点を指摘しておきたい。第 1 の指摘は、コウホートの効果がコウホートの推移にしたがって単調に増減するものではなく、戦後すぐのコウホートにおける効果を最大点としていることへの注意である。コウホートの効果が確実になったので、父学歴を含む階層研究を行う場合、コウホートの統制は欠損データによる偏りの回避に必須である。しかし、コウホートを統制する場合にも、例えば線形回帰分析の独立変数にコウホートを加えるだけでは意味をなさない。なぜならば、1950 年ごろのコウホートを最大点とするその効果は、線形の効果ではないからである。2、30 年前のデータを分析するのであれば、単純に若いコウホートほど欠損が発生しやすいことになるので問題はないが、最近のデータを分析する際にはこの点に注意が必要である。

第 2 の指摘は、コウホートの効果によって、ここ 50 年ほどの間に見られる父学歴の欠損率の上昇傾向がある程度説明できるということである。1.2 節で述べたとおり、SSM 調査における父学歴の欠損率は、1955 年～1995 年までの間に 4.7%、9.9%、8.9%、14.8%、18.3%と上昇傾向が続いており、本稿で分析した JGSS-2000～2003 データにおいては、「わからない」と「無回答」を足し合わせた欠損率は 19.7%に達している。この欠損率

の上昇は、1950 年ごろのコウホートの成長とほぼ同期して捉えることができる。このことから考えると、今後の調査では、父学歴の欠損率は上げ止まり、あと 10〜20 年もすれば逆に減少し始める可能性がある。調査対象者のコウホート構成によって父学歴の欠損による偏り方は変わってくるので、異なる調査年度のデータを時系列的に比較する際には注意が必要である。

4.3 まとめ

以上の分析結果および考察の要点をまとめると、以下のとおりである。父学歴の欠損率の高さは、階層研究に偏りを与える恐れがあるが、その欠損原因モデルはこれまで複数の可能性が示唆され、曖昧なものであった。しかし、JGSS データを用いた今回の分析によって、その曖昧さは払拭された。父学歴が「わからない」者の発生率は、本人の学歴が低いことと、その出生コウホートが戦後の 1950 年ごろに近いことによって高まり、また、父学歴が「無回答」の者の発生は、戦前のコウホートに多い、という結果である。これらの結果は、戦争を挟んだ日本の教育制度の変化に照らし合わせて、十分に妥当な解釈が可能なものであった。

コウホートの効果を制御するために注意が必要なものの、父学歴の欠損は、対象者のコウホートおよび学歴を統制すれば十分に補正できることがわかったことは、重要である。このことは、父学歴の欠損が「無視できる（ignorable）」欠損であることを意味している。欠損の原因に欠損している父学歴そのものが影響している場合には、その欠損は「無視できない（nonignorable）」と呼ばれ、欠損データの特殊な分析を踏まえなければ、その偏りは決して補正できない。一方、「無視できる」欠損の場合には、適切な変数を統制した分析を行えば、結果に偏りを生じない。「わからない」ことによる欠損と「無回答」による欠損の原因メカニズムが、それぞれに明らかになったこととともあいまって、父学歴の欠損は比較的楽観的に取り扱うことができるようになったと言うことができよう。

[Acknowledgement]

日本版 General Social Surveys（JGSS）は、大阪商業大学比較地域研究所が、文部科学省から学術フロンティア推進拠点としての指定を受けて（1999-2003 年度）、東京大学社会科学研究所と共同で実施している研究プロジェクトである（研究代表：谷岡一郎・仁田道夫、代表幹事：佐藤博樹・岩井紀子、事務局長：大澤美苗）。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SSJ データアーカイブがデータの作成と配布を行っている。

[注]

(1) 欠損している父学歴そのものが欠損の原因になるとは、次のような意味である。調査データにおける父学歴が欠損している場合でも、それは父学歴の情報が存在しないということ

ではなく、その情報を得ることができなかったのにすぎない。つまり、本来得られるべきであった父学歴の情報は、知ることができないものの存在している。その知ることのできなかった情報が欠損の原因となることは、当然ありうる。

(2) 1985 年調査の分析（保田 2000b）が 1995 年調査の分析（保田 2000a）に先行して行われたが、発行手続きの都合上、1995 年調査の分析の方が先に発行されたために、後に分析された方が「2000a」、先に分析された方が「2000b」となっている。

(3) 表には示していないが、相互作用項を含むモデルについても適合度を検討した。しかしながら、相互作用によって欠損の発生を説明することで有意に適合度が改善することはなかった。例えば、コウホートによって本人学歴の影響の仕方が異なることを意味する[BDR]モデルは、適合度検定の結果が $df=30$、$L^2=23.68$、$p=0.786$ となるので適合的であるが、[BR][DR]モデルと比べて有意な改善はない。相互作用を含んだ複雑な欠損のメカニズムを想定する必要はないということである。

(4) 念のため、コウホートについても、それが影響するのは「わからない」の発生のみで、「無回答」の発生には影響しないというモデルについて適合度を調べてみたが、適合度検定の結果は、$df=50$、$L^2=51.82$、$p=0.403$ であり、適合度が有意に悪化する。したがって、コウホートが「無回答」の発生に与える影響は無視できない。

[参考文献]

Baker, Stuart G. and Laird, Nan M. 1988, “Regression Analysis for Categorical Variables With Outcome Subject to Nonignorable Nonresponse,” *Journal of the American Statistical Association*, 83(401), 62-69.

Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977, “Maximum Likelihood From Incomplete Data Via the EM Algorithm (With Discussion),” *Journal of the Royal Statistical Society, Ser.B*, 39(1), 1-38.

Fay, Robert E. 1986. “Causal Models for Patterns of Nonresponse,” *Journal of the American Statistical Association*, 81(394), 354-365.

Groves, Robert M., Dillman, Don A., Eltinge, John L., and Little, Roderick J. A. (eds.), 2001, *Survey Nonresponse*, John Wiley & Sons, Inc.

McLachlan, Geoffrey J. and Krishnan, Thriyambakam, 1997, *The EM Algorithm and Extensions*, John Wiley & Sons, Inc.

保田時男, 2000a, 「欠損データの分析がもたらす新たな知見: 1995 年 SSM 調査の学歴移動表分析を例として」, 『大阪大学教育学年報』, 5, 139-152.

保田時男, 2000b, 「クロス集計表における欠損データの分析: 学歴移動表を例として」, 『理論と方法』, 15(1), 165-180.