

# 高校在学時を起点とする パネル調査における初期標本 脱落とバイアスの補正

「東大社研・高卒パネル調査」を事例として

Attrition in Panel Surveys of High School Graduates and Bias Correction: Case Analysis of the Japanese Life Course Panel Survey

山口泰史 YAMAGUCHI, Yasufumi

東京大学社会科学研究所 特任研究員

高校在学時を起点とするパネル調査における、初期標本脱落に伴うバイアス補正の可能性を「東大社研・高卒パネル調査」を事例として検討した。標本脱落は、男性や成績下位、進路多様校出身など特定の層に偏っているが、Inverse Probability Weighting法によって、そのバイアスが改善された。

キーワード 標本脱落, IPW法, 高校在学時を起点とするパネル調査

## 1 はじめに

学校から職業への移行や学校教育が個人に与える効果を検討する上で、高校在学時を起点としてそれ以後を追跡するようなパネル調査は極めて有用である。しかし、高校卒業時は離家や生活習慣の変化の時期であるため、サンプルの初期脱落により追跡調査の結果に大きなバイアスが生じうるという問題が知られている。

高校在学時を起点とするパネル調査の実施方法として、①高校をまずサンプリングし、高校を通じて生徒にアプローチする、②対象者を直接サンプリングしてアプローチする、という2つが考えられる<sup>1)</sup>。在籍する学校についての情報はじめ、より充実した情報を得られるという点で①には大きな利点がある一方、一般に調査開始時点で直接対象者に調査応諾を得る訳ではないため、①においてとくに初期脱落が大規模になりかねない。

パネル調査における標本脱落については、調査への回答継続を前年度調査の各項目への回答で予測し、その逆数で分析にウェイトづけを行うInverse Probability Weighting法（以下、IPW法）によってバイアスのある程度補正しうることが指摘されている（Wooldridge, 2002など）。日本では、たとえば坂本（2006）が「消費生活に関するパネル調査」（財団法人家計経済研究所）について検討し、IPW法の使用により、標本脱落に伴うバイアスが補正される可能性を示唆した。また、三輪（2014）は「全国家族調査パネルスタディ」（日本家族社会学会）についてIPW法によるバイアス補正の可能性を検討した。彼は、「2つの度数分布の違いをみるための統計指標で、何パーセント分だけ相対度数をシフトさせれば両者が完全に一致するかを示す」（三輪, 2014: 171）ものである非類似指数を用いて、①第1波となるNFRJ08の全回答者（パネル調査への非応諾者も含む）とパネル調査（NFRJ08-Panel）応諾者、②パネル調査応諾者と第1波～第5波完



全回答者、それぞれの間の第1波回答のズレ（偏り）がIPW法による補正である程度減少することを示した。

高校在学時を起点とするパネル調査からの初期脱落については、井上（2009）が「現代高校生の生活と意識に関する調査」とその追跡調査（東京大学教育学研究科高校教育研究会）を分析し、脱落を促す属性は成人調査とかならずしも同じではなく、具体的には大学に進学した場合の偏差値や出身高校が公立かどうか、社会観などが脱落の有無と関連していることを明らかにした。しかし、このような特定の層に偏った初期脱落によって生じるバイアスをどのように補正して分析することが可能なのかについては、いまだ明らかになっていない。

そこで本稿では、東京大学社会科学研究所パネル調査プロジェクトが実施している「東大社研・高卒パネル調査（以下、JLPS-H）」を取り上げ、調査からの初期脱落を生む特性を確認した上で、それに伴って生じるバイアスに対しIPW法による補正を試みる。JLPS-Hは対象へのアプローチに際し、前述の①の方法を取っているため、より大規模な初期脱落が生じやすいが、このような調査においてもIPW法によるバイアスの補正が可能なかを議論できる。大規模な初期脱落が生じている場合でも、それに伴うバイアスの補正にIPW法が有用ならば、高校在学時を起点とするパネル調査についてIPW法を用いた分析がもっと盛んに試みられるべきである。

## 2 データと方法

### 2.1 JLPS-Hの概要

JLPS-Hでは、高卒後の進学率・無業率に基づき47都道府県から選定された4つの県に所在する全日制高校101校の生徒に対して、第1波調査が行われた<sup>2)</sup>。高校3年生を対象に2004年1月から3月までの期間で実施し、7563ケースの回答が得られている。

第1波の生徒調査票に、継続調査への協力意向を対象者本人に確認する欄を設け、承諾する場合に調査票発送先の住所を記載してもらった。応諾者は2036ケース、第2波（2004年10月～2005年1月実施）の実際の回答数は482ケースである。それ以降は概ね400～500前後で推移し、2018年8月時点で第14波まで実施済みである。

この調査は現在も継続中だが、今回はこのうち第1波、第2波を用いて初期脱落の検討を行う。

### 2.2 標本脱落に基づくバイアスの補正

標本脱落に伴うバイアスについては、前述のようにIPW法を用いて補正するのが一般的である（Wooldridge, 2002；坂本, 2006；三輪, 2014など）。本稿ではIPW法をJLPS-Hに対して用い、標本脱落に伴うバイアスの補正可能性を検討する。成人の調査対象を追跡するパネル調査で検討した三輪（2014）では、IPW法がバイアスの補正効果を持つと示されているが、高校在学時を起点とするパネル調査では脱落を促す特性が異なりうるため、IPW法の効果は定かではない。

バイアスが補正されたかどうか、またどの程度補正されたかについては、前述の非類似指数を用いて、三輪（2014）の方法を部分的に踏襲して検討する。非類似指数を用いることで、第1波回答者（ $n=7563$ ）と比した第2波継続者（ $n=482$ ）の「第1波回答」の偏りを、IPW法によるウェイトなしの場合とウェイトありの場合で比較できる。「ウェイトあり」の方が非類似指数が小さければ、IPW法がバイアスの補正効果を持つと解釈できる。

## 3 分析

### 3.1 標本脱落を生む属性とウェイトの作成

どのような属性を持つ回答者が第2波に回答するのかをまず確認する。第2波への回答の有無を従属変数とした二項ロジスティック回帰分析の結果が表1である。

表1を見ると、高校所在県 (B県,C県)、学校ランク・種別 (進学校)、高卒後希望/予定進路 (大学・短大、専門学校、未定・その他・無回答)、高3時校内成績 (上位) が正に有意、高3時校内成績 (下位) が負に有意である。第2波回答の有無は、これらの要因によって異なることが分かる<sup>3)</sup>。つまり、標本脱落の確率はこれらの属性によって違い、第2波以降の分析においてバイアスを生じさせていると考えられる。

そこでIPW法に従い、バイアスの補正を目的とするウェイトを作成する。第1波時点でほぼ確定しており高卒後には変化しないと考えられる、表1に示した変数だけでどの程度バイアスを補正することが可能かを示すためにも、今回は表1に示した推定結果から、以降で用いるウェイトを作成した。

### 3.2 ウェイトの効果：非類似指数を用いた検討

このウェイトは標本脱落で生じたバイアスを

実際に改善するのだろうか。第1波の全変数359個について、第1波回答者 (n=7563) と第2波継続者 (n=482) の間の回答分布のズレを表す非類似指数を、ウェイトを使用する場合と使用しない場合の2パターンで算出し、図1を作成した。斜線はウェイト使用あり/なしの場合の非類似指数が一致した際の位置を示している。図1からは、ウェイトを使用すると多くの変数で非類似指数が減少していることが伺える。実際、288個 (80.2%) の変数で非類似指数が減少しており、359個の変数の平均減少幅は2.3である<sup>4)</sup>。

ウェイトの有無別に非類似指数の分布を示したのが表2である。ウェイトを使用しない場合、非類似指数が10を超える変数が50個あるのに対し、ウェイトを使用すると0個になり、約81%の変数の非類似指数が5以下に収まっている。ウェイトの使用により、第2波での標本脱落に伴うバイアスがある程度改善されたことが分かる。ところで、ウェイトの使用はどのような変数

表1 第2波への回答ありを予測する  
二項ロジスティック回帰分析

	B	S.E.
性別(基準:男性・無回答)		
女性	0.193 †	(0.099)
高校所在県(基準:A県)		
B県	0.336*	(0.142)
C県	0.284*	(0.132)
D県	0.219	(0.161)
第1波時年齢(基準:17歳)		
18歳	0.006	(0.136)
19歳以上・無回答	0.272	(0.353)
学校ランク/種別(基準:進路多様校)		
進学校	0.433***	(0.116)
専門学校・専門学科	0.188	(0.150)
高卒後希望/予定進路(基準:就職)		
大学・短大	0.717***	(0.170)
専門学校	0.730***	(0.178)
未定・その他・無回答	0.612*	(0.250)
高3時校内成績(基準:中位)		
上位	0.264*	(0.111)
下位	-0.319*	(0.128)
無回答	-0.915	(0.592)
切片	-3.857***	(0.226)

注)n=7563, \*\*\*p<0.001 \*\*p<0.01 \*p<0.05 †p<0.1

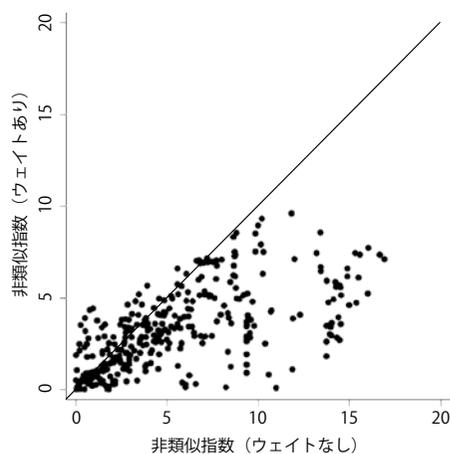


図1 ウェイトによる補正と非類似指数

表2 ウェイトの使用と非類似指数の分布

	~1	~5	~10	~15	~20
ウェイトなし	41	161	107	42	8
ウェイトあり	68	221	70	0	0



におけるバイアスを改善させるのか。第一波の359個の変数を大きく、(個人・学校・進路などの)「属性」についての変数、「行動・経験」についての変数、「考え方・意識」についての変数の3種類に分類し、ウェイト使用の有無による非類似指数の変化の様子を表3に示した。これをみると、属性変数でとくにウェイト使用の効果を確認できる。行動・経験変数や考え方・意識変数は属性変数ほど非類似指数の減少幅が大きいがないが、それでもウェイト使用の効果があることが分かる。

## 4 結論

本稿ではJLPS-Hのデータを用いて、高校在学時を起点とするパネル調査においてどのような属性が標本脱落につながるか、それに伴うバイアスはIPW法によって補正可能かどうかを検討した。

結果は以下の通りである。

(1) 主な属性変数で第2波回答の有無を予測する限り、男性、特定の県、進路多様校、就職者、成績下位層で脱落が生じやすい。(2) 非類似指数でみる限り、そのような標本脱落によって第2波データには一定のバイアスが生じている。(3) ウェイトの使用により非類似指数が多くの変数で減少しており、IPW法を用いることで、上記

のバイアスはある程度補正できる。

もちろん、IPW法を用いたバイアスの補正にも限界があり、第1波回答者と比べた第2波継続回答者の回答の歪みが完全になくなるわけではない。また、観察されない特性に依存した脱落が生じている可能性は無視できるものではない。

しかしながら、このような手法によってバイアスのある程度補正しうることも事実である。調査開始当初に想定した母集団に基づいて、何らかの示唆を提示する分析を目指すのであれば、限界は踏まえつつも、IPW法などを用いて、標本脱落によるバイアスをできるかぎり補正して、分析を進めることが望ましいと考えられる。

## 付記

本研究は、日本学術振興会科学研究費補助金基盤研究(S)(18103003,22223005)、日本学術振興会科学研究費補助金基盤研究(C)(25381122)、日本学術振興会科学研究費補助金基盤研究(B)(16H03778)、奨学寄付金・株式会社アウトソーシング、厚生労働科学研究費補助金、東京大学社会科学研究所研究資金による研究成果の一部である。パネル調査データの利用にあたっては社会科学研究所パネル調査企画委員会の許可を得た。

表3 変数の種類別に見た非類似指数とウェイトの効果

変数の種類	個数	"ウェイトなし 平均値(A)"	"ウェイトあり 平均値(B)"	A-B
属性	14	8.550	2.492	6.058
行動・経験	185	4.477	2.566	1.911
考え方・意識	160	6.263	3.875	2.388

## 注

1) 前者の例としてJLPS-Hの他に井上(2009)が検討した「現代高校生の生活と意識に関する調査」とその追跡調査、後者の例としては「高校生の進路についての追跡調査」(東京大学大学経

営・政策研究センター)や「高卒者の初期キャリア調査」(日本労働研究機構)が挙げられる。

2) 4県で全162校に調査実施を打診し、受諾を得られたのが107校(66%)、実際に生徒調査票を回

取できたのが101校(62%)であった。また、調査票の返送があった高校でもすべての生徒の回答があった訳ではない。そのため、第一波の時点で想定サンプルからの脱落は生じていることになるが、ここで脱落した生徒については情報がほとんど得られず対処が困難である。

3) この結果は、高校在学時を起点とするパネル調

査で、サンプリング時に特定の層(進路多様校、就職希望者層など)を相対的に多く抽出することで、標本脱落によるバイアスが小さくなる可能性もあることを示唆する。

4) JLPS-Hは標本脱落が大きいこともあってか、ウェイトの使用により非類似指数が大きくなる変数も一部存在することには注意が必要である。

## 文献

---

井上公人, 2009, 「パネル調査における初期脱落サンプルの傾向——高校生を対象としたパネル調査データをもとに」『立教大学大学院社会学研究科年報』16: 21-33.

三輪哲, 2014, 「NFRJ-08Panelにおけるウェイトによる脱落への対処」『家族社会学研究』26 (2): 169-178.

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.

Wooldridge, Jeffrey M., 2002, “Inverse probability weighted M-estimators for sample selection, attrition, and stratification” *Portuguese Economic Journal*, 1 (2): 117-139.