



回顧調査とパネル調査の 特性を考える

——「教育と職業」の調査に関連して

中澤 渉

(大阪大学教授)

I 回顧調査とパネル調査をめぐる日本 社会学の現状

個人の教育歴、学校経験が将来の進路選択（就職先・職業）や所得にどう影響するのか、という問いは、社会科学の重要な関心の対象である。特に学校卒業（修了、場合によっては中退）後の初職入職経路は注目を集めており、多くの研究蓄積が存在する（石田 2014；荻谷・本田 2010；石田 2005；Rosenbaum and Kariya 1989）。こうした関心の集中は、「メンバーシップ型雇用」（濱口 2009）のもと、OJT を経て子飼いの社員を養成する日本企業のモデルを反映したものであった。筆者の専門は社会学なので、本稿では議論の対象を社会学に限定して話を進める。

社会学で、教育と職業の関係を検討する際、よく使用されてきた大規模調査データは何か。よく知られているのは、「社会階層と社会移動（social stratification and mobility）全国調査（通称：SSM 調査）」であろう。SSM 調査には、調査対象者の教育歴や職歴の情報が詳細に含まれる。また日本版総合社会調査（Japanese General Social Survey：JGSS）も、初職に関する質問があり、最終学歴と初職の関係を検討できる。いずれも東京大学社会科学研究所の SSJ データアーカイブに寄託されており、手続きを踏めば研究者は分析可能だ¹⁾。

これらは、特定の調査時点において、回答者が自らの在学期間や就職の経緯、そしてその後の職歴を思い出して回答する回顧調査である。しかし、職歴データの揃う SSM 調査であれば、その情報を駆使して、対象者をまるで毎年追跡調査したかのようなパネル・データの形式（パーソン・イヤー・データとか、ロング形式ともよばれる）に変換できる。

一方、パネル調査は、対象となる被調査者を、実際に追跡し続けるものである。長年言われてきたことだが、アメリカはパネル・データが豊富で、社会学でもパネル・データを使って分析することは、半ば当たり前になっている。因果効果分析が流行しており、そのためには経時的観察を伴うパネル・データの使用が前提となるからだろう。

ただし、日本の社会学で本格的にパネル・データが分析されるようになったのは、比較的最近である。筆者の参加してきた、東京大学社会科学研究所の若年・壮年パネル調査（Japanese Life Course Panel Survey：JLPS）はその一例である。日本家族社会学会の全国家族調査（National Family Research of Japan：NFRJ）でも、パネル調査が実施され、その重要性が認識されつつある。

ただ、パネル・データ分析の意義をどこに置くかは学問分野に依存し、用いる手法も異なる傾向

がある（山口 2004）。筆者は以前、パネル・データ分析の意義に関する論考を発表したことがあるが（中澤 2012）、これは計量経済学の議論をベースにしたもので、もちろんそれ自体誤りではないが、社会学的関心に照らして有意義な主張であったかといえ、今となっては疑問を感じる。パネル調査が重要だとしても、ただ個人を追跡すればよいデータになるわけではない。

実査にあたって無視できないパネル調査の欠点は、コストが大きいことだろう。そもそも調査協力が得にくい中で、面倒な調査を何年も協力してもらうことのハードルは大きい。仮に最初は協力してくれても、回数を重ねると協力者が減る脱落（attrition）の問題も避けられない。脱落がランダムに起これば問題ないが、脱落するサンプルには（移動が多いなど、追跡が困難な人々など）一定の傾向があると予想できる。もしこの傾向が、分析の関心（従属変数）と関連する場合、独立変数と従属変数の関係が歪められて推定される可能性がある（Elwert and Winship 2014）。データクリーニングも、前回調査との関係を検証し、回答に矛盾がないかなどを点検する必要性が生じ、管理コストも高まる。

回顧調査の場合、記憶の正確さや、回答時点の状況が過去の評価を歪めることへの懸念が生じる。特に収入や意識の項目に、その懸念は当てはまる。逆にいえば、重要な事実関係の項目は、回答のブレや誤りが生じにくいと考えられるので、回顧調査でもそれなりに信頼できるデータは集められそうだ。だとすると（過去の収入や意識を関心の対象から外せば）、なぜパネル調査を行うのかという意義は、より一層問われることになる。

ただし、漫然と回顧調査とパネル調査のどちらが優れているか、を問うのはあまり意味がない。現実には、調査の問いや関心、資金や人員などのリソースをもとに、調査設計を行うことになるからだ。本稿では、「教育と職業」というテーマに即して、回顧調査とパネル調査の違いについて、考慮すべき問題を整理しておきたい。

II データの代表性をどこに据えるのか

上述のように、一時点の調査で得られた回顧情報をもとに、パネル形式のデータを作成することは確かに可能だ。では果たして、それはパネル調査の代わりになりうるのか。しかしながら、話は単純ではない。それは主として、標本抽出の問題にかかわる。

大規模な社会調査のサンプリングは、実査の直前に行われ、層化多段抽出法が採用されることが多い。2015 年 SSM 調査は、2014 年 12 月 31 日時点で日本在住の、20～79 歳の日本国籍男女が母集団である（白波瀬 2018）。2015 年 SSM 調査により、学歴や、学歴と職業との関係、就職経路を分析するのであれば、それはそのイベントが起きたときではなく、2014 年末における 20～79 歳の日本人男女を母集団と想定して、分析する必要がある²⁾。

直ちに問題になるのは、2014 年末時点で亡くなっている人は対象から除かれるということだ。高齢になるほど、亡くなる人も増える。死亡というイベントがランダムに発生していれば大きな問題はないが、階層と健康や死亡率の関係を指摘する社会階層研究は多い（Elo 2009；Erikson and Torssander 2008）。死亡者も学校や就職の経験をしているはずだから、彼らを見捨てて生存したサンプルだけで分析すれば、何らかのバイアスが生じるだろう。また高齢者にとって、若い頃に関する記憶の間違いは発生しやすいかもしれない。転職が多いなど、変化の多い人生を歩んだ人も、記憶のミスというリスクは大きくなるだろう³⁾。

しかし具体的な標本抽出に踏み込むと、もっと厄介な問題がある。サンプリングの基準が 2014 年末ならば、それは 2014 年末の居住地に基づいて行われていることになる。しかし、回答者の現在の居住地と、学校時代や就職の場所は異なることも珍しくない。ということは、学校時代、もしくは進学・就職といったトランジションが起きている時点で抽出した標本と、SSM 調査の標本は、原理的に同じものになりえない。

SSM 調査には、調査時点の居住地（都道府県）と、回答者の 15 歳時点の居住地（都道府県）の

情報が含まれている。大雑把にまとめると、首都圏や愛知、京阪神といった大都市圏の都府県サンプルは、15歳時点の居住地より割合が大きく、進学・就職で地方から都市へという人口移動が生じたことを推測させる。一方で、その間、出身地にとどまった人も存在する。移動した人とそうでない人に、質的な違いがないと言い切れるだろうか。

雑駁にまとめると、地方→都市という移動はそれなりにあるが、逆のケースは稀である。そう考えると、回顧調査における地方のサンプルは、元々そこに住んでずっと留まっているか、Uターンしてきた者の可能性が高い。一方大都市圏のサンプルは、元々そこに住んでいたか、地方から移ってきた人が混じっている。だとすると、就職というイベントが起きる前に無作為抽出をした場合と比較して、SSM調査では、地方においては地方にとどまった（もしくは帰った）人が抽出される可能性が高まり、地方から都市に移動した人の選ばれる可能性は低くなるだろう。話はそんなに単純ではないかもしれない。しかし少なくとも、調査直前に抽出したサンプルと、回顧情報にあるイベント発生時に抽出したサンプルでは、ランダム抽出でも、その内容が異なる可能性があることは認識しておくべきだ。

さらに、日本の就職慣行は、中卒、高卒、（大学や短大などの）高等教育卒と、学校段階で異なる。その上、時代によって制度的枠組みも変化し、採用活動は景気の影響も受ける。就職活動の実態は、男女差も大きく、性別を考慮しない分析は実態に見合わない。SSM2015年調査の有効回収サンプルは7817で一見多そうだが、以上のような性、学歴、就職年をすべて考慮して分析するのは想像以上に難しい。男女、学歴別で、各年度の就職者、というようにグループ分けすると、各グループのサンプルは僅かとなり、推定の誤差が大きくなってしまう。これは必ずしも回顧調査に起因する問題ではないが、対象者の年齢が幅広い大規模調査は、単年度で変化する制度的、経済的要因を考慮した分析を行いにくい。

個人のライフコースを、学校から就職へ、という形で追うパネル調査を想定してみよう。海外に

目を向けると、そうした教育や労働市場への移行に着目したパネル調査は、SSM調査やJGSSのように幅広い年齢層を対象とするのではなく、特定のコーホート（世代）を母集団として標本を抽出し、そのサンプルを継続して追跡するものが多い。広い年齢層の対象者をピックアップすると、調査票の質問構成や分析が中途半端になる可能性が高い⁴⁾。それより、特定のコーホートに焦点を当てれば、進学や就職といった特定のライフステージに差し掛かったとき、詳細の質問項目を入れることができる。毎年でなくとも、何年かごとに同様の調査を繰り返せば、世代間の比較もできるだろう。

Ⅲ 変数の時間的な変化をどう見るか

回顧データで限界に直面する情報は、①（家庭の）収入や資産⁵⁾、②アスピレーションに代表される意識項目、③成績である。いずれも遠い昔の記憶があやふやになりがちで、信頼性に疑問符が付く。①②については、パネル調査によってより正確な回答が期待できる。

教育現場の進路決定で、成績が最重要ファクターの一つであることに、議論の余地はない。しかしその把握や分析について、社会調査では真剣に検討されてきたとは言い難い。これまでよく用いられてきたのは、簡便な自己評価である。具体的には、学校間の格差が相対的に少ないと想定できる中学3年時点进行を思い出し、クラス内での相対的位置（上位から下位の5段階）で評価してもらう。主観的で信頼性に怪しさが残る上、回顧調査の場合、回答時の地位が過去の成績評価に影響する可能性も考えられる。

では仮に、学校から正確なテストの点数の提供を受ければ問題解決するのか。しかしテストの難易度は毎回異なるため、点数はそのまま比較できない。そこで日本人の間に膾炙する「偏差値が想起されるが、偏差値は統計的には「標準得点（z-score）」とよばれるもので、異なる平均や分散の得点を、平均0、分散1の分布に変換した得点を意味するに過ぎない。偏差値の変化は、特定の母集団内における相対的位置の変化であり、真の

実力や能力の変化を反映するわけではない⁶⁾。当然、異なる母集団間の偏差値を比較する意味はない。

教育心理学分野の、特にテスト理論では、この問題が古くから意識されてきた。近年注目を集めているのが、項目反応理論 (Item Response Theory: IRT) である (加藤・山田・川端 2014; 豊田 2012)。ただし、IRT を教育調査に取り込む際には、社会調査の片手間として、あるいは単なる統計分析の技法として理解すべきではない。というのも、信頼性の高いテストを実行するには、テスト理論や各教科教育法の専門家とのコラボレーションが欠かせず、周到な準備を要するからだ (川口・松尾・樋口 2016)。

とはいえ、2015 年より、埼玉県の学力・学習状況調査では、IRT が導入されており、今後のこの分野の研究に影響をもたらすことが考えられる。これにより、パネル調査で、異時点間の成績の変化を追跡することが可能となるからだ。

なお、学校から職業への移行の情報を含むパネル調査が実施されたとき、何が分析できるだろうか。計量経済学で用いられる固定効果モデル (fixed effect model) は、回帰分析における観察されない異質性によるバイアスを除去するため、複数時点のデータがあることを利用し、観察されない異質性と一緒に変数の独立変数の影響を除去し、可変の独立変数の変動と、従属変数の変動の関連を検討するものである (中澤 2012)。したがって、質問紙は、毎回同じ質問を繰り返す設計になっている必要がある。

しかし社会学で、固定効果モデルにより回答しうる問いは多くない。社会学では、性、出身階層、人種民族、一旦獲得した学歴など、個人内で通常変化させることのできない (変化させるのが難しい) 属性や性質により個人をグループ分けし、そのグループ間の処遇の違いを検討する、という問いを立てることが多い (岸ほか 2018)。固定効果モデルからは、不変の変数がモデルから除かれてしまうので、扱いにくいのである。

ランダム効果モデル (random effect model) は不変の変数をモデルに含められるが、これは独立変数と観察されない異質性に相関がないという、

やや非現実的な強い仮定を置くことになる。それを避けるために、Hausman-Taylor 法や、ハイブリッド法も提唱されている (Allison 2009; Halaby 2004)。しかし、これらの回帰モデルは変化の対称性を前提としており (有田 2013)、成長のような右上がりの、一方向的な変化に着目しやすい教育をテーマにした分析には馴染みにくい。

とはいえ、教育現場のみを関心の対象とするパネル調査であれば、学習環境の変化とパフォーマンスの変化の関係を検討できるので、同じ質問を繰り返して個人内変動を分析できる構造にしておくともよいかもしれない。しかし異なる学校段階への進学や、教育から職業へという移行に焦点を当てたパネル調査の場合、それぞれのライフステージに沿った情報を詳細に集める、という設計を意識した方が、学校時代と就職後の関係を様々な観点から分析する可能性が広がるだろう。その場合、繰り返しになるが、対象年齢層を幅広くとることはせず、特定のコーホートに焦点を絞って追跡するほうが、結果として有益なデータとなるように思われる。

- 1) ただし本稿執筆時点において、SSM 調査のうち利用できるのは 1955 年から 2005 年までの 6 回分の調査 (SSM 調査は 10 年おきに実施) で、最新の 2015 年調査についてはまだ公開されていない。
- 2) SSM 調査のようなデータを、世代間移動表の分析に用いる際には注意が必要である。すべてのサンプルは「親」をもつが、「子」がいるわけではない。したがって回答者からみた親の情報は集められるが、子の場合は、子がいなければ不可能である。SSM 調査のような回顧データをもとに世代間移動分析をする際、回答者とその親の関係に着目することになる。この場合、子を持たなかった親世代は、対象となりうる回答者が存在しないので、分析に含まれ得ない。こうした回顧データは、回答者にとっての出身階層と到達階層とか、教育機会の不平等を問う場合には問題ないが、世代間の階層移動とか、階層再生産というマクロな構造を問う場合には注意が必要である。なぜなら、結婚や子の有無と階層の間に関連があるのに、この点を考慮せずに「親子間の階層再生産」を論じるとすれば、それは子を持たなかった特定の集団を外した偏った分析結果となりかねないからである (余田 2018)。
- 3) もっとも、変化の多いライフコースを送る人は、パネル調査であっても、移動があったりして捕捉が難しいことが多い。
- 4) 質問紙の内容は無限に増やせないで、対象者の多くが回答できる項目が優先されることが多い。進学や就職という重大なライフイベントがあれば、それに特化した細かい質問を含めたほうがよい。しかし、対象年齢を広くとってしまうと、進学や就職というイベントを経験しないサンプルも多く含まれる。すると、特定のライフステージに沿う質問を入れにく

くんだり、仮に入れられたとしても回答する資格のない非該当サンプルが多くなり、結局、計量分析可能なサンプル数が集まらない、ということが起きる。

- 5) 進学のエconomic格差を検討するには、進学を決定する直前の家計情報を把握する必要がある。ところがそうした全国調査は日本で稀なため、(多くの人が関心を抱いているはずだが)意外なほど実証研究は少ない。回顧調査であるSSM調査には、在学時点の家計情報は入っていない。仮にわかっても、回答者の年齢幅が広いので、回答者の進学時点での家計収入を当時の物価水準で評価せねばならず、分析は困難を伴うだろう。教育機会の不平等に関する実証研究で、親の学歴や職業が考慮されながら、家計収入を考慮していない分析が多いのは、そうしたデータの制約による部分が大きいの。
- 6) 折角、学校でのパネル調査が可能になり、テスト得点の変化を追跡することが可能になったとしても、その得点の変化が何を意味するのか解釈できなければ意味がない。テストが前回50点で、今回70点だから実力が伸びた、という解釈は、常に成立するわけではない。点数が伸びたのは、テストの難易度が下がったからかもしれないからだ。偏差値も同じ問題を抱えており、極端な場合、真の実力は伸びていなくても、周囲の子が皆怠けていて点数を落としたため、相対的な順位が上がれば、偏差値は上昇する(逆に、真の実力は伸びているのに、偏差値が下がることも起こりうる)。つまり複数の時点間のテスト得点の変化を、あくまで個人の実力や能力の変化として測定する技術が求められている。IRTはそうした要望に応える手法である。

参考文献

- Allison, Paul D. (2009) *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.
- Elo, Irma T. (2009) "Social Class Differentials in Health and Mortality: Patterns and Explanations in Comparative Perspective." *Annual Review of Sociology* 35: 553-572.
- Elwert, Felix and Christopher Winship (2014) "Endogenous Selection Bias: The Problem of Censoring on a Collider Variable." *Annual Review of Sociology* 40: 31-53.
- Erikson, Robert and Jenny Torssander (2008) "Social Class and Cause of Death." *European Journal of Public Health* 18 (5): 473-478.
- Halaby, Charles N. (2004) "Panel Models in Sociological Research: Theory Into Practice." *Annual Review of*

Sociology 30: 507-544.

- Rosenbaum, James E and Takehiko Kariya (1989) "From High School to Work: Market and Institutional Mechanisms in Japan." *American Journal of Sociology* 94 (6): 1334-1365.
- 有田伸 (2013) 「変化の向き・経路と非変化時の状態を区別したパネルデータ分析——従業上の地位変化がもたらす所得変化を事例として」『理論と方法』28: 69-85.
- 石田賢示 (2014) 「学校から職業への移行における『制度的連結』効果の再検討——初職離職リスクに関する趨勢分析」『教育社会学研究』94: 325-344.
- 石田浩 (2005) 「後期青年期と階層・労働市場」『教育社会学研究』76: 41-57.
- 加藤健太郎・山田剛史・川端一光 (2014) 『Rによる項目反応理論』オーム社.
- 荻谷剛彦・本田由紀編 (2010) 『大卒就職の社会学——データからみる変化』東京大学出版会.
- 川口俊明・松尾剛・樋口裕介 (2016) 「テストをめぐる学際的検討—教育方法学・教育心理学・教育社会学の対話」『福岡教育大学紀要 第4分冊 教職科編』65: 1-13.
- 岸政彦・北田暁大・筒井淳也・稲葉振一郎 (2018) 『社会学はどこから来てどこへ行くのか』有斐閣.
- 白波瀬佐和子 (2018) 「2015年『社会調査と社会移動に関する全国調査(SSM調査)』実施の概要」保田時男編『2015年SSM調査報告書1 調査方法・概要』2015年SSM調査研究会, 1-12.
- 豊田秀樹 (2012) 『項目反応理論(入門編)第2版』朝倉書店.
- 中澤渉 (2012) 「なぜパネル・データを分析するのが重要なのか」『理論と方法』27 (1): 23-40.
- 濱口桂一郎 (2009) 『新しい労働社会——雇用システムの再構築へ』岩波新書.
- 山口一男 (2004) 「パネルデータの長所とその分析方法——常識の誤りについて」『家計経済研究』62: 50-58.
- 余田翔平 (2018) 「出生力と学歴再生産——前向きアプローチによる検討」荒牧草平編『2015年SSM調査報告書2 人口・家族』2015年SSM調査研究会, 13-25.

なかざわ・わたる 大阪大学大学院人間科学研究科教授。最近の主な著作に『日本の公教育——学力・コスト・民主主義』中公新書(2018年)。教育社会学専攻。