

筒井 淳也 (立命館大学産業社会学部准教授)

1 頻繁に活用されるマルチレベル分析

社会学や心理学の計量分析の分野では、「マルチレベル分析 (multilevel analysis)」とよばれるデータの分析手法が盛んに活用されるようになってきている。社会学のトップジャーナルの1つである *American Sociological Review* に目を通せば、いかにこの手法が頻繁に使われているかわかるだろう。

他方で、社会学の分野ではマルチレベル分析の性質についての理解はあまり進んでいないように思える。とくに現在はソフトウェアが対応していればその手法を使えるという事情もあり、理解よりも利用が先行しているのが現状である。

この論考ではこのような事情を鑑み、マルチレベル分析とはどのような分析手法なのか、そしてどのように使っていけばよいのかについて、できるだけ平易な言葉で解説することにしよう。

2 固定効果モデルで「面通し」

計量分析にはさまざまな手法があり、またさまざまな用途があるが、分析者が回帰分析に一番期待するのは、バイアスの小さな推定である。

数学のスコアに対する試験前の学習時間の効果について推定する例を考えよう。通常のモデルでは、

$$\text{数学スコア} = \beta_0 + \beta_1 \text{学習時間} + e$$

という式で推定がなされる。 β_1 学習時間の部分を固定効果、 e を変量効果 (攪乱項) とよぶ。この e の部分には、学習時間以外に数学スコアに影響した無数の要因が集積されている。もしこの隠れた要因の一部が学習時間と相関しているのならば、 β_1 の推定値はバイアスを被るので、そうい

った要因は観察しておくことが望ましい。たとえば就学前教育が豊富だと学習時間が長く、また数学スコアも高いことが理論的に予想できる。その場合、就学前教育を観察して変数として投入しておかないと、 β_1 には本来学習時間の効果ではない効果が入り込み、結果がバイアスをもってしまう。

一般的に、よい育ちをした子は勉強する癖もついているだろうから、試験前に勉強せずともそもそも高いスコアを期待できる。育ちに関わる変数を観察しないでたんに試験前学習時間で数学スコアを回帰してしまうと、ウソの結果が推定されてしまう。調査においてさまざまな変数に影響を与える可能性のある変数、すなわち性別、年齢、学歴、出身家庭の情報などを同時に観察し、回帰モデルに投入するのは、これらの変数自体の効果に関心がある場合もあるだろうが、これらから影響を受ける変数の推定バイアス (擬似相関) を取り除くためという意味が大きい。

とはいえ、基本属性を観察して統制することによってバイアスをすべて取り除けるわけではない。質問紙調査では観察しにくい微妙な育ちの要因がバイアスをもたらしている可能性もある。もしそうなら、何が数学スコア上昇をもたらしたかという「捜査」における試験前学習時間という容疑者は、じつは真の犯人ではない (効果をもっていない) のかもしれない。

ここで、試験前学習時間が本当に犯人なのかどうかを検証するために「面通し」をするさまざまな方法がある。もっとも信頼性が高いのは実験である。まず複数人の数学スコアを試験で測定し、次に無作為に分けた2グループの片方のみ学習をさせ、そののちふたたび試験をするのである。この2グループのスコア向上の差が、学習のバイアスのない効果であるということになる。

このような実験が難しいときにしばしばとられる代替手段が、パネル調査である。たとえば2回の試験結果の差を、それぞれの試験前の学習時間の差で回帰することで、観察期間中一定であった要因の影響を推定結果から除去することができる。「育ち」の効果は持続的なものであるから、パネルデータ分析によってこれに起因するバイアスをキャンセルできるのである。これがパネルデータを使った固定効果モデル (fixed-effects model) である (北村 2005)。

3 変量効果モデルで「犯人探し」

上記のような方法は、最初からターゲット (容疑者) を絞り込んで、その効果を厳密に推定しようという方針に基づいている。計量経済学者はこの方針をとることが多い。しかし社会科学データにおいてはとくに、被説明変数は複数の (それこそ無数の) 要因によって説明される。共犯者が無数にいるようなものである。そこで、たとえばまだ容疑者の目星がつかない段階で、だいたいのあたりをつけるための探索的な捜査をする、という方針をとることが考えられる。

この捜査方針は、具体的にどのように行われるのだろうか? 鍵は、

数学スコア = $\beta_0 + \beta_1$ 学習時間 + e
の e (攪乱項) の部分にある。なにしろここには無数の共犯者がうようよしているからである。まずやれることは、理論的に想定できる要因を観察して、固定効果のパートに次々と移していくことである。しかしこれは少々効率が悪いやり方だ。数学スコアに強い影響を与える要因の特徴について、まずはざっくりと知りたいものである。そんな上手い方法があるのかと思われるかもしれないが、じつはある。

まず、2回の試験を一定のインターバル (たとえば半年) を設けて実施する。そうすると、1人の生徒につき2つの観察値を得ることができる。これにより、

数学スコア = $\beta_0 + u + e$
のように、変量パート (攪乱項) を2つに分解できるようになる (β_0 は全体平均値)。 u は観察期間中変化しなかった要因の集積効果の推定量、 e は観察期間中変化した要因の集積効果の推定量であ

り、これらは変量効果モデル (random-effects model) を使えば両方の大きさを推定することができる。もし u の大きさが e に比べて大きいのなら、数学スコアに大きく貢献した要因は、半年間では変化しないような要因のグループのなかにある、ということがわかる。もし e が大きいのなら、数学スコアはそのような生徒個人に帰属できるような要因では説明できないということになる。

この場合インターバルは半年だが、その間に変化しない要因には、たとえば育ちに起因するものがあるだろう。たった2回試験をして、しかも (学習時間などの) 基本要因を観察しないようなずさんな調査であっても、「育ち」要因が効いていそうかどうかのあたりをつけることができちゃうわけである。

試験を複数回行っていなくても、受験者の所属する集団や地域の情報が得られれば、同じような絞り込みが可能である。たとえば地域、学校、クラス、その他で変量パートを分割すると、

数学スコア = $\beta_0 + u_1 + u_2 + u_3 + e$
のようになる。 u_1 が地域、 u_2 が学校、 u_3 がクラスである。もしこのうち u_2 の大きさが突出して大きいのなら、数学スコアに影響する要因は (地域でもクラスでも個人でもなく) 学校単位で変動する要因である、ということになる。学校偏差値、公立か私学か、学校の教育方針などがそれにあたるだろう。

変量効果モデルのこの機能は、個人がおかれた集団や環境による要因を探索したい社会学者にとっては大いなる福音であるといえよう。

4 混合効果モデルでさらに絞り込み

さきほどの変量効果モデルでは、性別や学習時間などの固定効果はいっさい使われていなかった。しかしこういった要因は観察しやすいので、投入するに越したことはない。2回試験をやるパネル調査の例に戻って、今度は固定効果を投入したモデルを考えてみよう。

数学スコア = $\beta_0 + \beta_1$ 性別 + β_2 学習時間 + $u + e$
これにより、まず変量効果モデルでは u の中に隠れていた性別の効果が u から取り除かれ、また e の中に隠れていた学習時間の効果が e から除去される。そのうえであらためて u と e の大き

さを見ているのである。もし、固定効果の投入前では u が大きかったのに、投入後は u が極端に小さくなったとすれば、 u のなかにはもうあまり有力な要因が残っていない、ということになる。したがって次に探索すべきは e 、つまり半年の間に変化する要因である、ということになる。

このように固定効果の影響を取り除いた上で変量効果の影響を推定することができるのが、混合効果モデル (mixed-effects model) である。なぜ「混合」なのかというと、固定効果と変量効果の両方がモデルに含まれているからである。そして一般にマルチレベル分析といわれている分析手法は、ほとんどの場合この混合効果モデルを指している。

ここで注意すべき点が1つある。それは、上記で説明したパネル固定効果モデルと違って、混合効果モデルによる固定効果の推定では、そのままではバイアスが残っていることがある、ということである。固定効果モデルでは少なくとも半年間で変化しない要因によって引き起こされる擬似相関は除去できるのだが、混合効果モデルではそういった要因は u のなかにそのまま残っている (上記モデルでは性別の効果だけは除去できるが)。その効果は、混合効果モデルでは間違っただけで学習時間の効果 (β_2) に帰属されてしまう。

もちろん擬似相関を引き起こす要因を u や e から観察によって取り出していくにつれて固定効果のバイアスは小さくなっていくが、攪乱項のなかに隠れた要因の探索ではなく個体内で変化する要因のバイアスのない推定 (不偏推定) が目的であるのなら、混合効果モデルではなく固定効果モデルを優先させるべきである。

5 混合効果の柔軟なモデリング

さて、混合効果モデルは固定効果と変量効果の両方を同時に推定することができる柔軟さを誇っているが、さらに固定効果と変量効果の交互作用を推定することもできる。まずは交互作用を含まない

数学スコア = $\beta_0 + \beta_1$ 学習時間 + $u + e$
 という単純なモデルを推定すると、図1のようになる。

実線で示したのが、学習時間が数学のスコアに

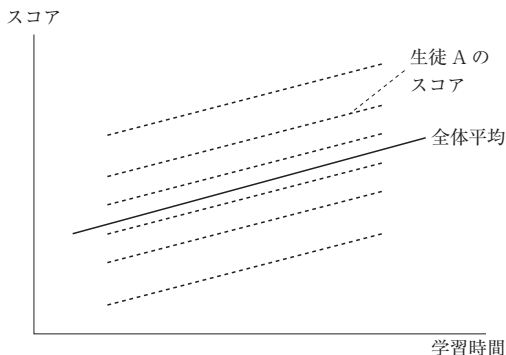


図1 数学のスコアと学習時間の関係における切片の変量効果

与える影響の平均傾向である。生徒 A は、この平均よりも高いスコアを保持していることがわかる。このようなモデルを変量切片モデルとよぶことがある。

この時点で少し補足しておこう。まず図で示されているのは変量効果の推定結果であって、もとのデータそのものではないということに気をつけよう。変量効果については、生徒個体ごとの観察数とその個体内のばらつきの大きさに応じて、全体平均からの距離が計算されている。個体内観察数 (ここでは数学テストの受験回数) が少ないほど、そのデータは信頼できないので、変量効果 (点線) は全体平均に近づく。個体内でのデータのばらつきが大きい場合も、そのデータは信頼性が低いと判定され、やはり変量効果の推定値は全体平均に近づいていく。この性質から、変量効果の推定量 (一般的に Best Linear Unbiased Predictor: BLUP とよばれている) は「縮約推定量 (shrinkage estimator)」とよばれることがある (Hox, 2002: 28)。したがって、個々の生徒個体のダミー変数から推定される値 (固定効果) と変量効果とは一致しない。

さて、図1では変量効果が全体的なレベル (切片) のみにあるという想定で推定されていた。つまり勉強時間と変量効果は相関していないという前提であった。しかし個人によって勉強の効果が異なる、ということも十分に考えられる。同じ勉強時間でもある生徒は別の生徒よりスコアの伸びが大きくなる、という想定である。数式では

$$\text{数学スコア} = \beta_0 + \gamma_1 \text{ 学習時間} + u_0 + u_1 \text{ 学習時間} + e$$

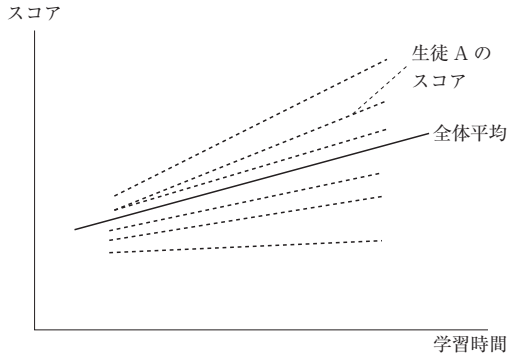


図2 数学のスコアと学習時間の関係における切片と傾きの変量効果

となる。 u_0 は切片の変量効果、 u_1 は学習時間の効果（傾き）の変量効果である。この場合、図2のような推定結果になる。

図2では、切片のみならず傾きに変量効果が想定されている（この場合でも全体平均への収縮が生じていることに留意しよう）。生徒Aにおける学習時間の効果は、全体平均よりも高いという推定結果になっている。全体的に学習時間の効果には個体ごとのバラつきが大きく、したがって半年で変化しない要因のなかにこのバラつきを説明する要因が含まれている、というアタリをつけることができる。このようなモデルを変量係数モデルとよぶことがある。

変量係数モデルでは、交互作用をもつ固定効果の変数を中心化して推定することがしばしばなされる。図2をみるとわかるが、何もしない状態では u_0 （切片の変量効果）は学習時間がゼロのときの値になり、情報としてあまり有用ではないからだ。中心化すれば平均値がゼロになり、したがって u_0 の効果も解釈しやすくなる（Luke, 2004: 49）。

ここでふたたび注意を喚起しておこう。上記のような単純なモデルだと、固定効果にはバイアスが含まれている可能性が高い。そうすると当然、交互作用効果もバイアスをもつことになる。解釈には十分に慎重になる必要がある。

6 サンプルの水増し防止

混合効果モデルには、以上のような探索的な用途以外にも重要な役割がある。それは、サンプルの水増し防止対策である。

極端な例で説明しよう。近年盛んに行われるようになってきているパネル調査であるが、たいていの場合は1年間程度のインターバルをおいて観察がなされる。ここでインターバルを仮に1時間にしたらどうだろうか。協力者の方を訪問して「あなたの昨年1年間の税込の年収は？」などと尋ねておいて、面接調査終了後一度その場を去るが、1時間の間をおいてふたたび訪問し、同じ質問をするのである。これで一気にデータの数が増える。サンプルサイズが大きければ誤差も小さくなるので、有意差も出やすくなる。調査コストもかからないし、いいことばかりである。

もちろんこれはインチキな調査法だ。なぜ1年間の間を空ければインチキではないのに、1時間だとインチキになるのだろうかという問いに答えるのは意外に難しいのでここでは説明しないが、いずれにしろ現代の計量分析では、個体ごとにあまりに変動のない観察値については「別個の（2つあれば2つの）観察」であるとはみなされず、誤差を調整することになっている。これを攪乱項のクラスター相関、あるいは個体内相関を考慮した誤差の推定とよぶ。やり方はおもに3つある。クラスターロバスト推定、FGLS（Feasible Generalized Least Squares）推定、そして混合効果最尤推定である。分散不均一のとときと同じく、後二者は基本的にロバスト推定よりも有効な（つまり精度の高い）推定量を生む。

この意味では、上でみてきたような攪乱項の分割は探索的な用途があるときにのみ行えばよい、というわけではない。データが何らかの形でクラスター化されており、そのクラスターの情報（個人、クラス、学校、地域など）が得られるのなら、とくに探索的な目的がない場合でもクラスターロバスト推定や混合効果推定を用いるほうが望ましい、ということである。

7 まとめ

以上で固定効果モデル、変量効果モデル、そして混合効果モデルの3つの分析手法の説明がひととおりできたと思う。肝心なのは、目的に応じてモデルを使い分けることである。そのためには個々のモデルがどういった性質をもっているのかについて、最低限の理解をしておく必要がある。

この論考がめざしたのは、こういった分析方針策定のためのインストラクションである。以上をよく理解されたうえで、充実した計量分析のひとつを過ごしてほしい。

文献

- Hox, J., 2002, *Multilevel Analysis: Techniques and Applications*, Mahwah, N. J. : Lawrence Erlbaum Associates.
北村行伸, 2005, 『パネルデータ分析』 岩波書店。
Luke, D. A., 2004, *Multilevel Modeling*, Thousand Oaks: Sage.

