



# サンプルセレクションとセルフセレクション

末石 直也

(神戸大学准教授)

## I 偏りを持つ標本

統計学の目的は、標本を用いて母集団の性質を理解することであり、標本は母集団から偏りなく抽出されることが望ましい。例えば、全国各地の大企業に勤務する男性1000人を対象に、賃金に関する調査を行ったとする。この標本が母集団の特徴を適切に反映しているかどうかは、何を母集団として考えるかに依存する。日本国内の大企業に勤務する男性全体を母集団と考えるならば、調査対象を作為的に選んでいない限り、これは適切な標本である。しかし、中小企業に勤務する人々や女性も含む日本の全就業者を母集団と考えるならば、明らかに偏った標本である。

実際にデータを分析する際には、興味のある母集団から意図したような標本（無作為標本）が得られないことは多く、そのような問題をサンプルセレクションという。サンプルセレクションは様々な理由から生じる。上述の賃金の例は、調査対象の選定に起因するものであるが、これも広義にはサンプルセレクションである。本稿では、サンプルセレクションの重要な要因であるセルフセレクションを中心として、サンプルセレクションが引き起こす問題を考えることにする。

## II 処置効果とセルフセレクション

セルフセレクションとは、個人が自らの意思によってどのような行動を取るかを選択すること、あるいは、その選択の結果、ある行動を取る人たちのグループと取らない人たちのグループの間で特性の差が生じることを指す。サンプルセレクションとセルフセレクションは同義のように語られることもある<sup>1)</sup>。しかし、すべてのサンプルセレクションがセルフセレクションによるものではなく、セルフセレクションはサンプルセレクションの一因である。

本節では、プログラム評価（政策評価）の例を通じて、セルフセレクションの問題を考察する。プログラム評価とは、何らかの処置が行われたときに、それが興味のある結果変数に与える効果（処置効果）を分析するものである。処置効果は、ある処置を受けた場合の結果変数と受けなかった場合の結果変数の値の差として定義される。

やや使い古された例ではあるが、職業訓練が賃金に与える効果を例として考えることにする。ある国の政府が、その国の労働者を対象に、職業訓練のプログラムを提供するとしよう。各個人にとっての職業訓練の処置効果は、職業訓練を受けたときに得られる賃金と受けなかったときに得られる賃金の差である。ところが、各個人については、職業訓練を受けるか受けないかのどちらか一方しか選択できないため、両方の結果は同時には観測されず、処置効果もわからない。

そこで、個人レベルでの処置効果を求めるのではなく、母集団における平均的な処置効果を推定することにする。ここで母集団は、職業訓練を提供する対象として想定する全労働者としておこう。処置効果のもっともらしい推定方法としては、実際に職業訓練を受けた人たちのグループと職業訓練を受けなかった人たちのグループを母集団から抽出し、2つのグループの平均的な賃金の差を求めるということが考えられる。ところが、このような方法では、必ずしも興味のある対象を推定することができない。なぜなら、2つのグループは、ともに母集団を適切に代表していないからである。通常、職業訓練を受ける（受けない）人たちは、そのことによって賃金の上昇が見込める（見込めない）と考える人たちである。つまり、職業訓練を受けるかどうかの選択は、選択の結果を見越しつつ自分の意思で決定される。そのため、受ける人たちと受けない人たちの中で特性の違いが生じ、標本の偏りをもたらしてしまう。これがセルフセレクションが引き起こす問題である。

セルフセレクションが生じているとき、標本は全く役に立たないかという点、必ずしもそうではない。ひとつの解決策として、似たもの同士を比較するということが考えられる。仮に、職業訓練を受けるグループと受けないグループの違いが、学歴、職歴、年齢などの観測可能な変数によってある程度説明されるとしよう。そのような場合、同じような変数の値を持つ人たち、つまり似たような特性の人たちを集めてきて、その中で職業訓練を受けた人と受けなかった人の比較を行えば、処置効果を求めることができる。このような分析は、重回帰やマッチング等の手法を用いて行うことができる。また、セルフセレクションの主たる問題

は、処置を受けるかどうかの選択が、選択の結果起こりうる結果と結びついている点にある。そこで、別の解決策として、職業訓練を受けるかどうかの選択には影響を与えるが、賃金には直接的には影響を与えないような変数を探してきて、処置の効果を識別する方法もある。これは操作変数法と呼ばれる方法である。

### Ⅲ 実験と脱落

セルフセレクションの別の解決法は、実験によって自由に選択をさせないことである。実験参加者を適当に募り、処置を受けるグループと受けないグループにランダムに振り分ける。ここでのポイントは、どちらのグループに入るかを、実験参加者に選ばせないことである。このような実験は無作為化比較試験と呼ばれ、実際に行われている。処置を受けるかどうかの選択と個人の特性が無関係になるので、どちらのグループにも偏りが生じない。ただし、実験を行うことすべての問題が解決されるとは限らない。例えば、ランダムに振り分けられる実験に参加することに躊躇する人たちは多い。そうであれば、ランダムに振り分けることには成功しても、実験参加者がそもそも母集団を適切に代表していない可能性がある。また、実験に参加する人たちは、自分たちの行為や結果が分析されていると知ると、そうではない場合と比べて、異なる行動をとってしまうこともある。卑近な表現ではあるが、いつも以上に頑張ってしまうのである。

実験や追跡調査が長期間に及ぶ場合、別のサンプルセレクションの問題が生じる可能性がある。最初は実験や調査に協力していても、途中で協力をやめてしまう人たちが存在し、データの欠損が生じてしまうことがある。このような問題を脱落 (attrition) という。脱落も広義のサンプルセレクションに含まれる。

脱落によるサンプルセレクションが問題を引き起こすかどうかは、どのようにして脱落が起きているかに依存する。脱落する人たちに特定の傾向がな

く、あたかもランダムに抜け落ちていると考えてもよいならば、特に問題は生じない。必ずしも当初意図した標本が得られたわけではないが、残った人たちも脱落した人たちも同じような特性を持つならば、残った人たちだけのデータを母集団からの無作為標本と同様に扱うことが可能である。それに対し、脱落するかどうか個人に依存しているような場合には、残った標本には偏りが生じる恐れがある。

### Ⅳ おわりに

本稿では、セルフセレクションを中心にサンプルセレクションの問題点について考察した。特にプログラム評価との関連で論じたが、古典的な回帰分析においても同様の問題は生じうる。冒頭の賃金の例のように、サンプルセレクションは、調査対象の選定など、調査する側の標本の採り方に起因して起こる場合もある。一方、セルフセレクションは、調査される側の行動によって引き起こされる問題である。また、脱落の例で述べたように、すべてのサンプルセレクションが問題となるわけではなく、サンプルセレクションがある標本でも無作為標本と同様に扱える場合もある。それに対し、セルフセレクションは一般に標本に偏りをもたらす、そのような標本を用いて統計的推測を行うには、何らかの工夫が必要となる。このように対比してみると、サンプルセレクションとセルフセレクションという2つの概念は、包含関係にありつつも、見方によっては似て非なるものと言えなくもないだろう。

- 1) 特に、サンプルセレクションモデルという場合には、セルフセレクションの問題を扱った Heckman のモデルを指すことが多い印象を受ける。

すえいし・なおや 神戸大学大学院経済学研究科准教授。最近の主な著作に "Identification Problem of the Exponential Tilting Estimator under Misspecification," *Economics Letters*, 2013, 118, 509-511. 計量経済学専攻。