

スパース因子分析の研究

西田 翔

因子分析とは、観察される変数間の相互関係を、因子と呼ばれる少数の隠れている変数によって説明することを目指すモデルである。一般的な因子分析には解釈する側の主観的な決定が必要とされる。これに対し、新たな因子分析の手法、スパース因子分析が提案された。スパース(sparse)とは、「疎な」という意味で、分析対象のパラメータを部分的に疎にする、すなわち零にする分析手法である。このスパース因子分析によって最適な因子解が推定され、余分なパラメータを無視するよう適切に零要素をもつ。その零要素の位置も最適な方法で推定される。そして、この手法は最適解が特定数の零要素、すなわちスパース負荷量と同時に得られる。この分析法によって因子回転が不用になり、因子負荷量が一意に定まる。そのため、どの回転法を採用するかという問題が回避できるようになった。結果として客観性も向上する。この提案された分析法をスパース直交因子分析(Sparse Orthogonal Factor Analysis)と呼び、各英単語の頭文字を取って SOFA と表し、最適な直交因子解が推定される(Adachi & Trendafilov, 2015)。

本論文では、R 言語(R Core Team, 2015)による SOFA の実装と、人間科学における SOFA の有用性の検証が目的である。その方法として、SOFA のプログラムを R 言語によって作成し、人間科学分野において実際に用いられるデータを使用して実データ解析を行う。Adachi & Trendafilov(2015)における SOFA のプログラムは Fortran によって作成されている。これは、科学技術計算以外の分野ではあまり一般的ではなく、ソフトウェア購入にも一定の費用が必要となる。一方で R 言語は、フリーかつオープンソースソフトウェアの統計解析向けのプログラミング言語である。R 言語は Fortran に比べ計算速度では劣るものの、高い汎用性と信頼性から幅広く用いられている。そのため、本研究では Fortran ではなく R 言語を用いることで SOFA の使いやすさが向上すると考えられる。

実際のデータをプログラムに適用して解析を行った。一つ目のデータは学生 168 人に対して行われた、24 種類の知能テストデータ(<http://astro.ocis.temple.edu/~alan/MMST/datasets.html>)である(Holzinger & Swineford, 1939)。二つ目のデータは大学生 200 名を対象とした性格テストのデータである(柳井ら, 1987)。因子数はスクリーテストを用いて決定し、スパース数は変数と因子数をかけて割った値とした。またスパース因子分析と比較するために、因子分析に VARIMAX 回転を加えたものを比較した。これらの結果を比較すると、両方のデータに対して、どちらも注目する因子負荷量は同様であり、変数と因子の関係も類似している。一方で、通常の因子分析は 0.01 という小さい値が存在しているのに対し、SOFA ではそのような値は 0 にあるため解釈が容易である。

実データ解析を行ったところ、通常の因子分析と同様の分析結果が得られ、かつ零要素が一定数あることによる解釈の容易さが生じる、という有効性を示した。今回の用いたデータは人間科学に用いられるようなデータであるため、人間科学にも SOFA の有効であると言えよう。しかしながら、SOFA においては直交解のみが求められ、斜交解は求められない。つまり、分析によって生じた共通因子に相関がないと仮定してデータの解釈を行うことになる。そのため今後は、同様の方法で斜交解を求められるようなスパース因子分析にも発展させるという課題が残る。また、因子数に関して分析者が主観的に決めるという点にも課題が残る。(行動統計科学)