

多変量解析を理解するために(5)

因子分析の基礎

高木 廣文 | Takagi Hirofumi

東邦大学医学部看護学科国際保健看護学研究室教授

■1950年生まれ。74年東京大学医学部保健学科卒。79年同大学院博士課程修了。同年米国国立衛生院奨励研究員として米国国立環境保健学研究所勤務。81年聖路加看護大学講師、82年同大助教授。89年文部省統計数理研究所助教授。99年新潟大学医療技術短期大学部看護学科教授、同年同大医学部保健学科教授。2006年4月より現職。著書に『ナースのための統計学 — データのとり方生かし方』(医学書院)、『多変量解析ハンドブック』(現代数学社)、『健康科学とコンピュータ』(共立出版)など。

1. 因子分析とは何か

世の中には、英語やフランス語など語学が得意な人もいれば、将棋や囲碁などが得意な人もいる。同じ学部学生でも、統計学はできるが生化学はできない学生、哲学はできるがドイツ語はできない学生など様々である。しかし、多くの教科間の成績には相関があり、互いに無関係ではない。すなわち、人間の持つ多様な能力には、その根底に何か共通する要因があるのではないかと推測できる。

人間の知能のように、多様な変数の背後に存在すると考えられる潜在的な因子構造を探るための方法として、今回は因子分析について説明する。

独自の要因(独自因子 unique factor)によって、その観測値が定まるものと仮定する。

通常、仮定される共通因子 $f_1, f_2, \dots, f_k$ の個数 k は、変数の個数 p より小さく、 $k < p$ とするのが普通である。このように仮定すると、各変数 z_i ($i=1, 2, \dots, p$) について、

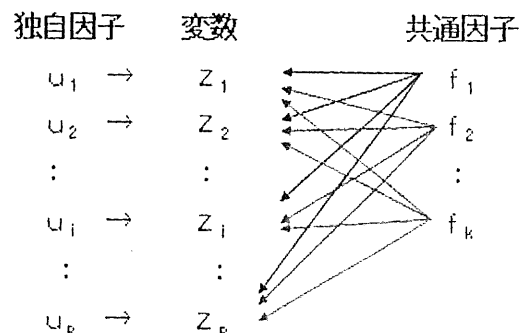
$$z_1 = a_{11} \times f_1 + a_{12} \times f_2 + \dots + a_{1p} \times f_p + e_1 \times u_1$$

$$z_2 = a_{21} \times f_1 + a_{22} \times f_2 + \dots + a_{2p} \times f_p + e_2 \times u_2$$

⋮

$$z_p = a_{p1} \times f_1 + a_{p2} \times f_2 + \dots + a_{pk} \times f_k + e_p \times u_p$$

図1 因子分析の概念図



2. 因子分析の基礎理論

今、変数が p 個あるものとする。なお、すべての変数は平均 0、分散 1 に標準化されているものとする。図 1 に示したように、各変数 z_i ($i=1, 2, \dots, p$) は、全変数に共通する要因(共通因子 common factor)と、各変数に

のように書くことができる。ここで、 $u_1, u_2, \dots, u_p$ は各変数の独自因子であり、 $e_1, e_2, \dots, e_p$ は各変数の独自因子に掛ける重みである。

因子分析では、各共通因子に掛ける重み a_{ij} ($i=1, 2, \dots, p; j=1, 2, \dots, k$) を求めることが主要な問題となる。共通因子に掛ける重みは、「(共通) 因子負荷量 (common) factor loading」と呼ばれ、各変数とその因子との相関係数である。因子負荷量の値 (正負にかかわらず) が大きいほど、その共通因子の影響が大きいことを示している。

各変数の因子負荷量と独自因子負荷量には、

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2 + e_i^2 = 1, (i=1, 2, \dots, p)$$

の関係があり、共通因子負荷量の 2 乗和は、求めた共通因子によって、どのくらいその変数を説明できるかの目安になり、「共通性 communality」と呼ばれる。

共通性

$$h_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2, (i=1, 2, \dots, p)$$

である。

共通性が大きい変数は、求めた共通因子によって十分説明できるが、共通性が小さい変数は、求めた共通因子だけでは説明が十分でないことを示している。すなわち、共通性が小さい変数は、他の変数に比べて独自性が強いことを示している。

データから共通因子負荷量を求めるために用いられる手法には、主因子法と最尤法がある。具体的な計算方法の解説は成書に譲るが、主因子法は主成分分析と同様に変数間の相関係数行列の固有値と固有ベクトルを求める方法である。すなわち、

$$R_F a = \lambda a$$

を解けばよい。ここで、 R_F は p 個の変数間の相関係数行列で、その対角成分である 1 を共通性に置き換えたものであり、 a はその固有ベクトル、 λ は固有値である。

因子分析では、因子負荷量を計算するために、あらかじめ共通性の推定値と共通因子数を決めておく必要がある。

よく用いられる共通性の初期推定値には、

① SMC (Squared Multiple Correlation) :

他の変数を説明変数として重回帰分析を行った場合の重相関係数の 2 乗値

② その変数と他の変数との相関係数のうち、絶対値が最大のもの

③ 1 (主成分分析と一致)

などがある。

実際の計算では、求めた因子負荷量から共通性を計算し、その共通性を用いて再度計算を行う。最終的に、すべての変数の共通性の値が一定になるまで反復計算を行うので、それほど共通性の初期値に神経質になる必要はない。しかし、必ずしも反復計算により全変数の共通性が収束するとは限らないので注意が必要である。

一方、共通因子数の決め方には、

① 相関係数行列を主成分分析した場合の 1 以上の固有値の個数

② 相関係数行列の対角要素を SMC で置き換えた行列の負でない固有値の個数

などがある。

実際の解析では、共通性の推定とともに適当に因子数を変えて、試行錯誤的に分析しないと良い結果が得られないことも多い。

3. 因子の回転について

因子分析では、最初に求めた共通因子をそのまま使用するのではなく、解釈しやすいように「因子の回転」を行うことが多い。すなわち、回転後の因子負荷量行列は実際に解釈しやすいことが必要である。そのような因子構造を満たすものとして、「単純構造 simple structure」が知られている。これは、①各変数はできるだけ少数の因子に高く負荷すること、②少数の因子とのみ相関が高くなること、という基準を満たす構造である。

因子の回転方法には、回転後の共通因子の相関がない「直交回転」の方法と、共通因子間に相関のある「斜交回転」の方法がある。

直交回転では、「オーソマックス法」と呼ばれる方法のひとつである「バリマックス法 varimax method」がよく用いられる。バリマックス法によって抽出された因子は、他の手法によるものより、因子の意味付けがしやすいという長所を持つと言われている。なお、尺度構成（次回解説予定）を行う場合は、下位尺度間に相関があるのが普通であり、直交回転ではなく斜交回転の方が因子の解釈がしやすい場合もある。

4. 因子の寄与率について

共通因子に関して寄与率、累積寄与率を、主成分分析と同じように求めることが多い。すなわち、変数の個数を p 、第 k 因子の固有値を λ_k とした場合、その寄与率 CR_k ($k=1, 2, \dots, p$) は、

$$CR_k = \frac{\lambda_k}{p} \times 100 \quad (\%)$$

により求められる。

上記のように、寄与率は全変数の変動に対す

る寄与を求めるのが普通だが、因子分析の場合、「全変数の共通性の合計」に対する寄与率や累積寄与率を求める場合もある。すなわち、共通因子 f_k の寄与率 CR_k ($k=1, 2, \dots, p$) を、

$$CR_k = \frac{\lambda_k}{\sum_{i=1}^p h_i^2} \times 100 \quad (\%)$$

によって求める。この場合、共通性の合計は変数の個数 p より常に小さいので、普通の定義よりも見かけ上、寄与率を大きくすることができる。ただし、累積寄与率はすべての共通因子について求めれば、必ず100%になる。

寄与率や累積寄与率の計算方法は、上記のように2とおりあり、どちらを用いたかを明確に示す必要がある。

5. 因子得点について

因子分析では、因子負荷量を基に、変数間の構造を検討するのが主な目的であるが、主成分分析のように各個人の共通因子の標準化得点を求めることがある。これは「因子得点 factor score」と呼ばれる。因子得点の計算方法については、主成分分析の場合の主成分と同様に、

$$f = w_1 \times z_1 + w_2 \times z_2 + \dots + w_p \times z_p$$

のように、各変数に重みを掛けて合計する。具体的な重みの計算については、幾つかの方法が提唱されているが、ここではその詳細は省略する。

共通因子が何を表しているのかを解釈するためには、全変数の因子負荷量を表にまとめたり、共通性や寄与率などを付加して表中に記したりする必要がある。また相関図を描く場合と同様に、2つの因子負荷量を使って平

面上に各変数をプロットすることもある。また、各ケースを2つの因子得点を基に平面上にプロットする場合もある。作成した図を見比べることで、因子の意味付けが行えることもあるが、必ずしも因子の意味付けが可能なわけではない。無理な因子の解釈は、誤った結論を導くことにもなりかねないので、注意が必要である。

6. 主成分分析と因子分析の違いは何か

前回説明した主成分分析と、今回の因子分析は非常によく似た方法である。実際の計算においては、主成分分析は因子分析の主因子法で共通性を1にした場合、すなわち独自因子がないと仮定した場合と数学的に完全に一致する。しかし、主成分分析と因子分析では基本的な考え方に大きな差がある。

因子分析は、各変数を潜在的に仮定される幾つかの共通因子と独自因子に分解して考えることが基本的モデルになっている。これに対して、主成分分析は各変数を用いて、新たな合成得点、すなわち主成分を求めようという考えである。このように、因子分析は変数を分解し、主成分分析は合成しようという、全く逆の方向からデータを分析するものである。

このように、主成分分析と因子分析には、基本モデルに根本的な違いがあり、研究目的に応じて使い分ける必要がある。しかし、現実のデータ解析では、主成分分析では共通性の大きさや因子数（主成分数）を決める必要がないなど便利な点が多いので、主成分分析を因子分析と同様の目的で用いることも多い。従って、このような目的で主成分分析を用いる場合、基本的な考え方が因子分析とは異なるという点をよく理解して用いるべきであろう。

*参考文献

- [1] 高木廣文・林 邦彦(2006): エビデンスのための看護研究の読み方・進め方: 中山書店, 東京.
- [2] 高木廣文(2007): 多変量解析を理解するために (4)主成分分析の基礎: エストレーラ, 4月号, pp.48-51.
- [3] 柳井晴夫・高木廣文編著(1986): 多変量解析ハンドブック: 現代数学社, 京都.
- [4] 高木廣文・柳井晴夫編著(1995): HALBAUによる多変量解析の実践: 現代数学社, 京都.
- [5] 高木廣文(2006): HALBAU7によるデータ解析: シミック(株).