クラスタリングを伴う多変量データ解析法の新展開

宇野　光平

はじめに

心理学を含む多くの分野で用いられている因子分析は、得られた変数を説明できるような因子を抽出する分析法である。 通常用いられている変数モデルで因子分析は、変数と因子の関係を表す因子負荷量と独自分散のみ推定でき、個体と因子の関係を表す因子得点は2段階のステップで推定しなくてはいけない。

因子負荷量と因子得点、両方を同時に推定するモデルは母数モデル因子分析と呼ばれるが、母数モデルは解が得られないことがわかっている。しかし、ある制約を因子得点に加えることで、母数モデルは解を得られるようになる。本研究では、個体がいくつかのクラスターにわけられ、同じクラスターに所属する個体は同一の因子得点となる、という制約を設けた。この制約によって、因子の抽出だけでなく個体の分類も行えることが判明した。

モデル

変数モデル因子分析は以下のように表せる。

$$\mathrm{X} = \mathrm{F}\Lambda' + E$$

ここで、X は n 行 p 列のデータ行列とし、F は n 行 m 列の因子得点行列、Λ は p 行 m 列の因子負荷量行列、E は n 行 p 列の誤差行列とする。また、Λ'は Λ の転置行列を表す。

それに対して、本研究で提案するモデルは

$$\mathrm{X} = \mathrm{GC}\Lambda' + E$$

ここで、G は n 行 K 列のメンバーシップ行列で、個体がどのクラスターに所属するかを表す行列である。C は K 行 m 列のクラスター得点行列で、クラスターと因子との関係を表す。

結果と考察

提案手法をいくつかのデータに適用した結果、類似の研究である RKM や FKM より分類精度が良いことがわかった。特に RKM との比較は重要である。なぜなら、RKM のモデルも

$$\mathrm{X} = \mathrm{GCA}' + E$$

と表すことができるからである。提案手法と RKM の違いは、RKM においては

$$E'E = I$$

となるのに対して、提案手法では

$$E'E = \Psi$$

となる点である。I は p 行 p 列の単位行列で、Ψは p 行 p 列の独自分散である。よって、独自分散が分類に影響していることが示唆される。今後の課題は、理論的な考察が不十分なので、提案手法の理論的な研究を進めることである。（行動統計科学）