

消費者データを深掘りする方法：3相主成分分析



経営戦略研究科客員教授(経営戦略専攻) 中西正雄

消費者調査で頻繁に得られるデータには

- ・ 2次元(2相)データ：複数の個人(例 消費者)から複数の測度(変数)の測定値が得られたもの。
- ・ 3次元(3相)データ：複数の個人から複数の対象(例 ブランド)に関する複数の測度(変数)の測定値が得られたもの。

といったタイプがある。このような2相データや3相データを通常の方法で因子(主成分)分析した際にもつきまとうのは、個人によって因子(成分)構造が異なるのではないかという疑念である。しかし3相データ

が得られているのなら、以下に述べるモデルを使えば、容易に因子(成分)構造の個人差を検出することができる。3相データは、マーケティング・リサーチの様々な局面で得られるので、それを活用して消費者行動に関する新しい知見を得ることが期待される。

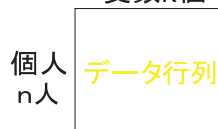
個人*i*から得られたデータ行列(Z_i とよぶ。変数ごとに中央化してあるものとする)を主成分分析すると、成分得点行列 S^*_i (対象布置という)および成分行列 F^*_i とが得られる。

$$Z_i = S^*_i F^*_i \quad (S^{*i'} S^*_i = (m-1)I; \quad F^{*i'} F^*_i = \Lambda_i^2)$$

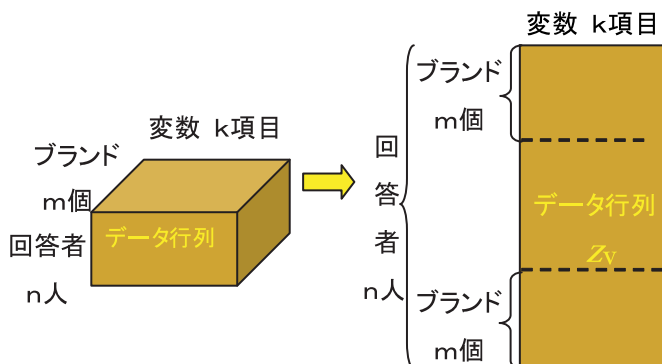
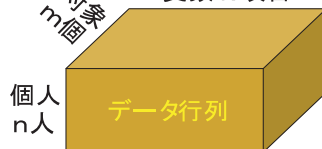
上式で記号'は倒置行列、 I は単位行列、 Λ_i^2 は固有値行列を示す。しかしこうして得られた成分空間はその個人に特有のものであるため、複数の個人間で S^*_i や F^*_i の直接比較はできない。複数個人の平均値をとることも、リンゴとオレンジの平均をとるようなもので、まったく無意味である。とは言え、多数の回答者からの2相データを一々分析するのは非効率であるだけでなく、そこから一般的なパターンを読み取ることは難しい。

3相データ(対象*m*個×変数*k*項目×個人*n*人)があれば、それを構成する2相データ(Z_i)を縦に並べて($m \times n$)×*k*の2次元行列を作る。これを Z_y 行列とよぶ(右図参照)。この行列を主成分分析すると、固有値行列 Λ (階数 $r_k \leq k$)、成分(負荷量)行列 F_0 、および成

2相データ
変数*k*個



3相データ
変数*k*項目

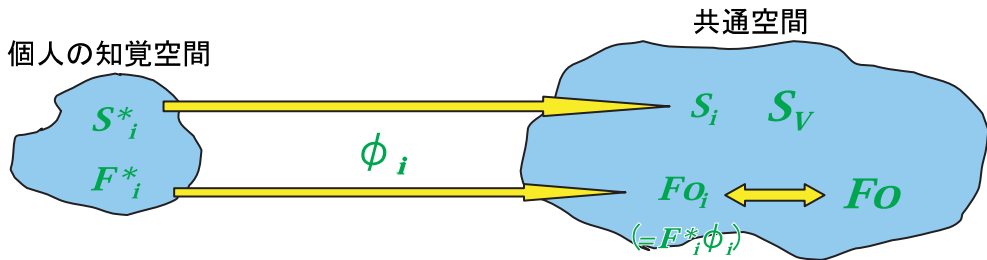


得点行列 S_V が得られる。添え字 o は「全体」(overall) の、 v は vertical の意味である。

$$Z_V = S_V F_o' \quad (\text{ここで } S_V' S_V = (nm-1)I; \quad F_o' F_o = \Lambda^2 I / (m-1))$$

この等式は Z_V が列中央化されているから成立するのであって、主成分分析で常にこれが成り立つ訳ではない。また Z_V 行列を事前に変数ごとに標準化しておけば、この主成分分析結果は因子分析の一つの解にもなっているので、 F_o を因子負荷量と解釈できる。

上の結果からは対象布置 (S_V) に関する個人・グループ間の差異を分析できても、成分 (負荷量) の比較はできない。成分行列は全体として一つ (F_o) しか得られないからである。成分行列の個人間比較を可能にするには、各個人の成分行列 F^*_i を共通空間 (F_o で規定される) に射影する必要がある。この場合、共通空間はいわば複数の個人の姿を映す「共通の鏡」の役割を果たす。各個人は自分とはすこし違う姿が映っていると感じるかもしれないが、同じ鏡に映っていることにより、複数個人間で比較ができる訳である。



個人 i の 2 相データ Z_i からの対象布置を S^*_i 、成分行列を F^*_i とし、全体データから得られた S_V の個人 i に対応する部分を S_i ($i=1, 2, \dots, n$) とする。 S^*_i から S_i への変換 ϕ_i と、 F^*_i の共通空間への写像 (F_{oi}) は次式であたえられる。

$$S_i \equiv S^*_i \phi_i \Rightarrow \phi_i = (S^*_i' S^*_i)^{-1} S^*_i' S_i = S^*_i' S_i / (m-1)$$

$$F_{oi} = F^*_i \phi_i = F^*_i S^*_i' S_i / (m-1) = Z_i' S_i / (m-1)$$

最後の等式は Z_i と S_i から F_{oi} を容易に計算できることを示している。 F_{oi} ($i=1, 2, \dots, n$) は個人間の直接比較が可能なので、標本をいくつかのグループ (セグメント) に分割した場合でも、グループごとに F_{oi} と S_i の平均をとったり、グループ間での F_{oi} や S_i の違いを分散分析したりすることが可能である。 F^*_i と S^*_i の場合直接比較ができないためにグループ間の差を統計的に検定することができなかったことを思えば大きな改善と言えよう。

参考文献

* Kroonenberg, P. M. (1985), "Three-Mode Principal Components Analysis of Semantic Differential Data: The Case of a Triple Personality", *Applied Psychological Measurement*, Vol. 9, No. 1, pp. 83-94.
 * Tucker, L. R. (1966), "Some Mathematical Notes on Three-Mode Factor Analysis", *Psychometrika*, Vol. 31, No. 3, pp. 279-311.