# バイプロット分析

- 機能

ケースと変数の 2 次元同時プロットを行う。

1. タイプ 1 のバイプロットは，ケース間のユークリッド距離の近似を可能とする解である。この解は，主成分分析によって得られる解と類似点がある。
2. タイプ 2 のバイプロットは，ケース間のマハラノビス距離の近似を可能とする解である。

- 解析理論

$p$ 個の変数からなる $n$ 個のケースを表すデータ行列を $\boldsymbol{X}$，その階数を $r$ とする。$\lambda_1$，$\lambda_2, \ldots$，$\lambda_r$（$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$）を対角要素とする対角行列を $\boldsymbol{D}$，正規直交ベクトルを行ベクトルとする行列を $\boldsymbol{U}$，$\boldsymbol{V}$ とするならば，行列 $\boldsymbol{X}$ は以下のように分解できる。

$$\boldsymbol{X} = \boldsymbol{U}\,\boldsymbol{D}\,\boldsymbol{V}' \tag{1}$$

このような分解は特異値分解，$\lambda_1, \lambda_2, \ldots, \lambda_r$ は特異値と呼ばれる。もし，$\lambda_3$ 以降の値が小さいならば，上述の $\boldsymbol{X}$，$\boldsymbol{U}$，$\boldsymbol{V}$，$\boldsymbol{D}$ に対応する階数 2 の行列を $\boldsymbol{X}_2$，$\boldsymbol{U}_2$，$\boldsymbol{V}_2$，$\boldsymbol{D}_2$ としたとき，データ行列 $\boldsymbol{X}$ は以下のように近似できる。

$$\boldsymbol{X} \fallingdotseq \boldsymbol{X}_2 = \boldsymbol{U}_2\,\boldsymbol{D}_2\,\boldsymbol{V}_2' \tag{2}$$

近似の程度は，$(\lambda_1 + \lambda_2) / \sum \lambda_i$ で判定できる。

バイプロットにはいくつかの種類があるが，よく使われるのは以下の 2 種類である。

**タイプ 1・・・・・ケースのプロットにより有効**

（**??**）式において，$\boldsymbol{F} = \boldsymbol{U}_2\,\boldsymbol{D}_2$，$\boldsymbol{G} = \boldsymbol{V}_2$ とすれば，$\boldsymbol{F}$ は $n \times 2$，$\boldsymbol{G}$ は $p \times 2$ の行列であり，2 次元平面において，前者は $n$ 個のケース，後者は $p$ 個の変数を表すことになる。2 次元平面に $\boldsymbol{F}$ を点で，$\boldsymbol{G}$ をベクトルとして同時に表現するものをバイプロットと呼ぶ。

タイプ 1 のバイプロットは以下のような性質を持つ。

1. 個体 $i$ を表す $\boldsymbol{f}_i$ と変数 $j$ を表す $\boldsymbol{g}_j$ の内積 $\boldsymbol{f}_i'\,\boldsymbol{g}_j$ は $\boldsymbol{X}_{ij}$ に対する近似値を表す。
2. 個体間の相対的位置関係（ユークリッド距離），が 2 次元平面上のプロットで近似的に表現される。

**タイプ 2・・・・・変数のプロットにより有効**

（**??**）式において，$\boldsymbol{F}= \sqrt{n}\,\boldsymbol{U}_2$，$\boldsymbol{G} = \boldsymbol{V}_2\,\boldsymbol{D}_2 / \sqrt{n}$ と表すこともできる。

タイプ 2 のバイプロットは以下のような性質を持つ。

1. 個体 $i$ を表す $\boldsymbol{f}_i$ と変数 $j$ を表す $\boldsymbol{g}_j$ の内積 $\boldsymbol{f}_i'\,\boldsymbol{g}_j$ は $\boldsymbol{X}_{ij}$ に対する近似値を表す。
2. $\boldsymbol{g}_j$ の長さは変量 $\boldsymbol{X}_j$ の分散，$\boldsymbol{g}_j$ と $\boldsymbol{g}_k$ の内積は $\boldsymbol{X}_j$ と $\boldsymbol{X}_k$ の共分散，$\boldsymbol{g}_j$ と $\boldsymbol{g}_k$ のなす角の余弦は $\boldsymbol{X}_j$ と $\boldsymbol{X}_k$ の相関係数と近似的に等しい。
3. $\boldsymbol{f}_i$ と $\boldsymbol{f}_m$ の間の距離は，個体間のマハラノビスの汎距離の近似になっている。

- 補足説明

変数とケースを 2 次元平面に表示したとき，変数（またはケース）のプロットが原点付近に集中する場合には，変数（またはケース）のスケールを拡大してやってもよい。ただし，このようにした場合には，図から変数とケースの関係を読みとる際に注意が必要である。

- 参考文献


1) K. R. Gabriel: The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453–467, 1971.