

## 正準判別分析（重判別分析）

変量数を  $P$ 、群の数を  $g$ 、各群  $k = 1 \sim g$  のデータ数を  $n_k$  とする。

第  $k$  群（ $k = 1 \sim g$ ）のデータを、

$$\begin{bmatrix} x_{1,1}^{(k)} & \cdots & x_{1,P}^{(k)} \\ \vdots & \ddots & \vdots \\ x_{n_k,1}^{(k)} & \cdots & x_{n_k,P}^{(k)} \end{bmatrix}$$

とする。

全データ数  $n = \sum_{k=1}^g n_k$  とする。

各群の（母）分散共分散行列は等しいとする。すなわち、

$$\Sigma_{(1)} = \Sigma_{(2)} = \cdots = \Sigma_{(g)} = \begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,P}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{P,1}^2 & \cdots & \sigma_{P,P}^2 \end{bmatrix} = \Sigma$$

とする。

第  $k$  群の平均ベクトル

$$\{m\}^{(k)} = \begin{pmatrix} m_1^{(k)} \\ \vdots \\ m_P^{(k)} \end{pmatrix}, \quad m_j^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{i,j}^{(k)}, \quad j = 1 \sim P$$

全体の平均ベクトル

$$\{m\} = \begin{pmatrix} m_1 \\ \vdots \\ m_P \end{pmatrix}, \quad m_j = \frac{1}{n} \sum_{k=1}^g \left( \sum_{i=1}^{n_k} x_{i,j}^{(k)} \right), \quad j = 1 \sim P$$

とする。

$$M^{(k)} = \begin{bmatrix} m_1^{(k)} & \cdots & m_P^{(k)} \\ \vdots & \ddots & \vdots \\ m_1^{(k)} & \cdots & m_P^{(k)} \end{bmatrix} \quad (n_k, p) \text{ 行列}$$

$$C^{(k)} = X^{(k)} - M^{(k)} = \begin{bmatrix} x_{1,1}^{(k)} - m_1^{(k)} & \cdots & x_{1,p}^{(k)} - m_p^{(k)} \\ \vdots & \ddots & \vdots \\ x_{n_k,1}^{(k)} - m_1^{(k)} & \cdots & x_{n_k,p}^{(k)} - m_p^{(k)} \end{bmatrix} \quad (n_k, p) \text{ 行列}$$

$$S^{(k)} = (C^{(k)})^t C^{(k)} \quad (p, p) \text{ 行列}$$

$$V^{(k)} = \frac{S^{(k)}}{n_k - 1} \quad (p, p) \text{ 行列} \quad \text{不偏分散共分散行列}$$

$$W = \sum_{k=1}^g S^{(k)} \quad (p, p) \text{ 行列} \quad \text{群内変動}$$

$$V = \frac{W}{n-g} \quad (p, p) \text{ 行列} \quad \text{群内分散}$$

群間変動について

$$D^{(k)} = \begin{bmatrix} m_1^{(k)} - m_1 & \cdots & m_p^{(k)} - m_p \\ \vdots & \ddots & \vdots \\ m_1^{(k)} - m_1 & \cdots & m_p^{(k)} - m_p \end{bmatrix} \quad k = 1 \sim g, (n_k, p) \text{ 行列}$$

$$E^{(k)} = (D^{(k)})^t D^{(k)} \quad (p, p) \text{ 行列}$$

$$B = \sum_{k=1}^g E^{(k)} \quad (p, p) \text{ 行列} \quad \text{群間変動}$$

P 次元のベクトル  $\{a\} = \begin{Bmatrix} a_1 \\ \vdots \\ a_p \end{Bmatrix}$  を考え

群 k のデータ  $(x_{i,1}^{(k)}, x_{i,2}^{(k)}, \dots, x_{i,p}^{(k)})$  ( $i=1 \sim n_k$ ) を  $\{a\}$  に射影した変量 (スカラー) を

$$z_i^{(k)} = (x_{i,1}^{(k)}, x_{i,2}^{(k)}, \dots, x_{i,p}^{(k)}) \begin{Bmatrix} a_1 \\ \vdots \\ a_p \end{Bmatrix}$$

群内変動の拘束条件を

$$\{a\}^t W \{a\} = 1$$

とする。

群間変動を  $\{a\}^t B \{a\}$ 、群内変動を  $\{a\}^t W \{a\}$  とし、

$$h(\{a\}, \lambda) = \frac{\{a\}^t B \{a\}}{\{a\}^t W \{a\}} - \lambda (\{a\}^t W \{a\} - 1)$$

を最大にする  $\{a\}$  を求める。

$$\frac{\partial h(\{a\}, \lambda)}{\partial \{a\}} = 2 B \{a\} - 2 \lambda W \{a\} = \{0\} \quad \text{となり、}$$

$\{a\} \neq \{0\}$ 、 $B \{a\} = \lambda W \{a\}$  であるから、 $W^{-1} B \{a\} = \lambda \{a\}$  という固有値問題となる。

$W^{-1} B$  の rank は  $q = \min(p, g-1)$  なので、

$$W^{-1} B \text{ の固有値を } \{\lambda\} = \begin{Bmatrix} \lambda_1 \\ \vdots \\ \lambda_q \end{Bmatrix},$$

固有ベクトルを  $(\{a_1\}, \dots, \{a_q\})$  とする。

$$\ell = 1 \sim q, \quad \{\widetilde{a}_\ell\} = \frac{\{a_\ell\}}{\sqrt{\frac{\{a_\ell\}^t W \{a_\ell\}}{n-g}}} = \left\{ \begin{array}{c} \widetilde{a}_1^\ell \\ \vdots \\ \widetilde{a}_p^\ell \end{array} \right\} \quad \text{とする。}$$

$k=1 \sim g, \ell=1 \sim q, i=1 \sim n_k$  の正準スコア  $z_{i,\ell}^{(k)}$  は

$$z_{i,\ell}^{(k)} = \left\{ \begin{array}{c} x_{i,1}^{(k)} - m_1 \\ \vdots \\ x_{i,p}^{(k)} - m_p \end{array} \right\}^t \{\widetilde{a}_\ell\} = \left\{ \begin{array}{c} x_{i,1}^{(k)} \\ \vdots \\ x_{i,p}^{(k)} \end{array} \right\}^t \{\widetilde{a}_\ell\} - \{m\}^t \{\widetilde{a}_\ell\}$$

$$a_{0,\ell} = -\{m\}^t \{\widetilde{a}_\ell\} \quad \text{とおけば、} \quad z_{i,\ell}^{(k)} = a_{0,\ell} + \left\{ \begin{array}{c} x_{i,1}^{(k)} \\ \vdots \\ x_{i,p}^{(k)} \end{array} \right\}^t \{\widetilde{a}_\ell\}$$

となる。

群ごとの正準スコアの平均値を、 $k=1 \sim g, \ell=1 \sim q$  について

$$m_{z\ell}^{(k)} = a_{0,\ell} + (\{m\}^{(k)})^t \{\widetilde{a}_\ell\}, \quad m_z^{(k)} = \left\{ \begin{array}{c} m_{z1}^{(k)} \\ \vdots \\ m_{zq}^{(k)} \end{array} \right\}$$

とし、 $\ell_1=1 \sim q, \ell_2=1 \sim q, k=1 \sim g$  について

$$\left[ t_{\ell_1, \ell_2}^{(k)} \right] = (\{\widetilde{a}_{\ell_1}\})^t S^{(k)} (\{\widetilde{a}_{\ell_2}\}) = \begin{bmatrix} t_{1,1}^{(k)} & \cdots & t_{1,q}^{(k)} \\ \vdots & \ddots & \vdots \\ t_{q,1}^{(k)} & \cdots & t_{q,q}^{(k)} \end{bmatrix}$$

$q \times q$  行列の  $(\ell_1, \ell_2)$  成分

$$\widetilde{V}_k = \frac{\begin{bmatrix} t_{1,1}^{(k)} & \cdots & t_{1,q}^{(k)} \\ \vdots & \ddots & \vdots \\ t_{q,1}^{(k)} & \cdots & t_{q,q}^{(k)} \end{bmatrix}}{n_k - 1} \quad \text{また} \quad r_{\ell_1, \ell_2}^{(k)} = \frac{t_{\ell_1, \ell_2}^{(k)}}{\sqrt{t_{\ell_1, \ell_1}^{(k)} \cdot t_{\ell_2, \ell_2}^{(k)}}}$$

である。

今、任意のデータ  $\{x\} = \begin{Bmatrix} x_1 \\ \vdots \\ x_p \end{Bmatrix}$  が与えられたときの正準スコアは

$$\{z\} = \begin{Bmatrix} z_1 \\ \vdots \\ z_q \end{Bmatrix} = \begin{Bmatrix} a_{0,1} + \{x\}^t \{\widetilde{a}_1\} \\ \vdots \\ a_{0,q} + \{x\}^t \{\widetilde{a}_q\} \end{Bmatrix}$$

となる。

また、各群の正準スコアの平均からのマハラノビス距離は  $k=1 \sim g$  について

$$D_k^2 = \{z - m_z^{(k)}\}^t \widetilde{V}_k^{-1} \{z - m_z^{(k)}\}$$

である。