

数量化Ⅰ類

データ数を n 、要因数を R 、要因 j ($j=1\sim R$) のカテゴリ数を C_j とする。

各データ ($i=1\sim n$) について $\hat{y}_i = \sum_{j=1}^R \sum_{k=1}^{C_j} a_{j,k} \delta_i(j,k) + \varepsilon_i$ とする。

$C_{total} = \sum_{j=1}^R C_j$ とする。

任意の i, j, k について、 $\delta_i(j,k) = 0 \text{ or } 1$ であり、

任意の i, j について、 $\sum_{k=1}^{C_j} \delta_i(j,k) = 1$ であるべき。

$$\text{ベクトル } \{\hat{y}\} = \begin{Bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{Bmatrix}, \quad \{a\} = \begin{Bmatrix} a_{1,1} \\ \vdots \\ a_{1,C_1} \\ \vdots \\ a_{R,1} \\ \vdots \\ a_{R,C_R} \end{Bmatrix}, \quad \{\varepsilon\} = \begin{Bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{Bmatrix},$$

$$\text{また、} [\Delta] = \begin{bmatrix} \delta_1(1,1) & \cdots & \delta_1(1,C_1) & \cdots & \delta_1(R,1) & \cdots & \delta_1(R,C_R) \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \delta_n(1,1) & \cdots & \delta_n(1,C_1) & \cdots & \delta_n(R,1) & \cdots & \delta_n(R,C_R) \end{bmatrix}$$

とすると、上式は $\{\hat{y}\} = [\Delta] \{a\} + \{\varepsilon\}$ となる。

ここで、 $Q = \{\varepsilon\}^t \{\varepsilon\} = (\{\hat{y}\} - [\Delta] \{a\})^t (\{\hat{y}\} - [\Delta] \{a\})$

が 最少となる $\{a\}$ を求める。

$$\begin{aligned} Q &= (\{\hat{y}\} - [\Delta] \{a\})^t (\{\hat{y}\} - [\Delta] \{a\}) \\ &= (\{\hat{y}\}^t - \{a\}^t [\Delta]^t) (\{\hat{y}\} - [\Delta] \{a\}) \\ &= \{\hat{y}\}^t \{\hat{y}\} - 2\{a\}^t [\Delta]^t \{\hat{y}\} + \{a\}^t [\Delta]^t [\Delta] \{a\} \end{aligned}$$

となり $\frac{\partial Q}{\partial \{a\}} = -2[\Delta]^t \{\hat{y}\} + 2[\Delta]^t [\Delta] \{a\} = \{0\}$

結局、 $[\Delta]^t [\Delta] \{a\} = [\Delta]^t \{\hat{y}\}$ となる。

ここで 上記の $[\Delta]$ のランクについて検討事項がある。

要因 j 、カテゴリ k についての列ベクトルを

$$\{\delta_{j,k}\} = \begin{Bmatrix} \delta_1(j,k) \\ \vdots \\ \delta_n(j,k) \end{Bmatrix} \text{ と表現する。}$$

この列ベクトル $\{\delta_{j,k}\}$ について調べる。

(1) 同一の j について C_j 個の列ベクトル $\{\delta_{j,k}\}$ は 互いに一次独立。

まず、 $\sum_{k=1}^{C_j} \{\delta_{j,k}\} = \begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix}$ であり、また任意の k について

$$\{\delta_{j,k}\} \neq \begin{Bmatrix} 0 \\ \vdots \\ 0 \end{Bmatrix} \text{ であり、} \{\delta_{j,k}\} \neq \begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix} \text{ である。つまり、} \{\delta_{j,k}\}$$

の中のどれかの i 成分は 1 を取る。

もし、 $\{\delta_{j,k}\}$ が一次従属なら、例えば、 $k = 1 \sim C_j$ の中の k_1 について

$$\{\delta_{j,k_1}\} \text{ が、他のベクトル}$$

$$\{\delta_{j,k}\} \text{ (} k = 1, \dots, k_1 - 1, k_1 + 1, \dots, C_j \text{) の一次結合で}$$

表現できることになる。

しかし、 $\{\delta_{j,k_1}\}$ の中の零でない 第 i 成分について考えると、

$$\delta_i(j, k_1) \text{ は } 1 \text{ で、}$$

$$\text{それ以外の } \delta_i(j, k) \text{ (} k = 1, \dots, k_1 - 1, k_1 + 1, \dots, C_j \text{)}$$

は 0 であるから、

$$\{\delta_{j,k}\} \text{ (} k = 1, \dots, k_1 - 1, k_1 + 1, \dots, C_j \text{) による}$$

一次結合は成り立たないことになる。

(2) 異なる2つの要因 j_1, j_2 についての列ベクトル

$\{\delta_{j_1,k}\} \quad k=1 \sim c_{j_1}$ と $\{\delta_{j_2,k}\} \quad k=1 \sim c_{j_2}$ との間では1次従属である。

$$\sum_{k=1}^{c_{j_1}} \{\delta_{j_1,k}\} = \begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix}, \quad \sum_{k=1}^{c_{j_2}} \{\delta_{j_2,k}\} = \begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix} \quad \text{であるから、要因 } j_2 \text{ の}$$

中の1つのベクトル \tilde{k} 番目のベクトル $\{\delta_{j_2,\tilde{k}}\}$ は

$$\{\delta_{j_2,\tilde{k}}\} = \sum_{k=1}^{c_{j_1}} \{\delta_{j_1,k}\} - \sum_{k=1}^{c_{j_2}} \{\delta_{j_2,k}\} \quad (\text{ただし } k \neq \tilde{k})$$

となるから。

したがって、要因 $j=1$ のベクトルはそのままに、 $j=2 \sim R$ については、 $k=1$ のベクトルを除いた行列を作成する。

$$[\tilde{\Delta}] = \begin{bmatrix} \delta_1(1,1) \dots \delta_1(1,c_1) & \delta_1(2,2) \dots \delta_1(2,c_2) & \delta_1(R,2) \dots \delta_1(R,c_R) \\ \dots & \dots & \dots \\ \delta_n(1,1) \dots \delta_n(1,c_1) & \delta_n(2,2) \dots \delta_n(2,c_2) & \delta_n(R,2) \dots \delta_n(R,c_R) \end{bmatrix}$$

として、 $\{b\}^t = \{b_{1,1} \sim b_{1,c_1}, b_{2,2} \sim b_{2,c_2}, \dots, b_{R,2} \sim b_{R,c_R}\}$

$$[\tilde{\Delta}]^t [\tilde{\Delta}] \{b\} = [\tilde{\Delta}]^t \{\hat{y}\}$$

を解く。

さらに $b_{2,1} = b_{3,1} = \dots = b_{R,1} = 0$ として $\{b\}$ を編集しなおす。

つまり

$$\{b\}^t = \{b_{1,1} \sim b_{1,c_1}, b_{2,1} \sim b_{2,c_2}, \dots, b_{R,1} \sim b_{R,c_R}\}$$

とする。

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^R \sum_{k=1}^{c_j} b_{j,k} \delta_i(j,k) = \sum_{j=1}^R \sum_{k=1}^{c_j} b_{j,k} \left(\frac{1}{n} \sum_{i=1}^n \delta_i(j,k) \right)$$

ここで $\bar{\delta}(j,k) = \frac{1}{n} \sum_{i=1}^n \delta_i(j,k)$ とする。

一方 $\hat{y}_l = \sum_{j=1}^R \sum_{k=1}^{C_j} b_{j,k} \delta_l(j, k) \quad (l = 1 \sim n)$ とする。

$$\hat{y}_l - \bar{y} = \sum_{j=1}^R \sum_{k=1}^{C_j} b_{j,k} \delta_l(j, k) - \sum_{j=1}^R \sum_{k=1}^{C_j} b_{j,k} \left(\frac{1}{n} \sum_{i=1}^n \delta_i(j, k) \right) \\ (l = 1 \sim n)$$

$$\hat{y}_l = \bar{y} + \sum_{j=1}^R \sum_{k=1}^{C_j} b_{j,k} \left(\delta_l(j, k) - \frac{1}{n} \sum_{i=1}^n \delta_i(j, k) \right) \quad (l = 1 \sim n) \quad \text{となる。}$$

これから 次式が導かれる。

$$\hat{y}_l = \bar{y} + \sum_{j=1}^R \sum_{k=1}^{C_j} \left(b_{j,k} - \sum_{m=1}^{C_j} b_{j,m} \bar{\delta}(j, k) \right) \delta_l(j, k) \quad (l = 1 \sim n)$$

そこで、

$$\cdot \quad a_0 = \bar{y} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^R \sum_{k=1}^{C_j} b_{j,k} \delta_i(j, k)$$

$$\cdot \quad a_{j,k} = b_{j,k} - \sum_{m=1}^{C_j} b_{j,m} \bar{\delta}(j, k) \quad \text{但し } \bar{\delta}(j, k) = \frac{1}{n} \sum_{i=1}^n \delta_i(j, k)$$

と置くことで $a_0, a_{j,k} \quad (j = 1 \sim R, k = 1 \sim C_j)$ を定める。

次に重相関係数は以下となる。

$$R = \frac{(\sum_{i=1}^n (\hat{y}_i - \bar{y})^2)^{1/2}}{(\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2}}$$

また偏相関係数は以下となる。

$P=1 \sim R, \quad Q=1 \sim R$ において、

$$x_{i,p} = \sum_{k=1}^{C_p} b_{p,k} \delta_i(p, k) \\ \bar{x}_p = \frac{1}{n} \left(\sum_{i=1}^n x_{i,p} \right) \\ \sigma_{p,q} = \frac{1}{n} \sum_{i=1}^n (x_{i,p} - \bar{x}_p) (x_{i,q} - \bar{x}_q) \\ \sigma_{p,y} = \frac{1}{n} \sum_{i=1}^n (x_{i,p} - \bar{x}_p) (y_i - \bar{y}) \\ \text{但し } \bar{y} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^R \sum_{k=1}^{C_j} b_{j,k} \delta_i(j, k) \\ \sigma_{y,y} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

として、

$$r_{pq} = \frac{\sigma_{pq}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{qq}}}$$

$$r_{py} = \frac{\sigma_{py}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{yy}}}$$

$$[R] = \begin{bmatrix} r_{11} & \cdots & r_{1R} & r_{1y} \\ \vdots & \vdots & & \vdots \\ r_{R1} & \cdots & r_{RR} & r_{Ry} \\ r_{y1} & \cdots & r_{yR} & r_{yy} \end{bmatrix}$$

$$[R]^{-1} = \begin{bmatrix} r^{11} & \cdots & r^{1R} & r^{1y} \\ \vdots & \vdots & & \vdots \\ r^{R1} & \cdots & r^{RR} & r^{Ry} \\ r^{y1} & \cdots & r^{yR} & r^{yy} \end{bmatrix}$$

とした時の、偏相関係数 $\rho(y,p)$ は 以下となる。

$$\rho(y,p) = \frac{-r^{p,y}}{\sqrt{r^{p,p}} \sqrt{r^{y,y}}}$$