

数量化 2 類

グループ(群)別に項目の選択データが与えられているとき、それらのグループを区別(判別)するのに最適な項目選択枝の数量化を行うのが数量化 2 類である。データは、各項目(アイテム) j 、 $j=1, \dots, n$ 、に対する選択枝(カテゴリ) $1 \sim N_j$ から 1 つ k 、 $k=1, \dots, N_j$ 、を選ぶ形式のアイテム・カテゴリ型である。岩坪(1987)では、次の表 1 の例があげられている。1 はその選択枝が選ばれたことを、0 は選ばれなかったことを表す。

表 1 グループ別に与えられたアイテム(項目)・カテゴリ(選択枝)型データ

群	個体	項目 1		項目 2		
		選択枝 1	選択枝 2	選択枝 1	選択枝 2	選択枝 3
A	1	1	0	1	0	0
	2	1	0	1	0	0
	3	1	0	0	1	0
B	1	1	0	0	1	0
	2	0	1	0	1	0
C	1	0	1	0	1	0
	2	0	1	0	0	1
	3	0	1	0	0	1

表 1 の形式のデータを表す変数 $n_i(v, jk)$ を次のように定義する。

個体 i がグループ v に属しており、かつアイテム j のカテゴリ k が選ばれているとき

$$n_i(v, jk) = 1$$

それ以外るとき

$$n_i(v, jk) = 0$$

このとき、アイテム j のカテゴリ k に数値 $x(jk)$ を与えると、グループ v の i 番目の個体の得点 $y_i(v)$ を次式で与えることができる

$$y_i(v) = \sum_j \sum_k x(jk) \cdot n_i(v, jk) \quad (1)$$

$y_i(v)$ の値のグループ内での変動は小さく、グループ間の変動は大きくなるように $x(jk)$ の値を定める。すなわち、相関比の2乗

$$\eta^2 = \frac{V_B}{V_T}$$

が最大になるように $x(jk)$ の値を定める。ここで、 V_B は $y_i(v)$ の群間分散を、 V_T は全体の分散を表す。

いま、 $n_i(v, jk)$ の項目選択肢間の分散共分散行列を A 、群間分散共分散行列を B とおくとき、 η^2 を極大にする $x(jk)$ は次式を満たすことが導かれる(岩坪、1987)。

$$Bx = \eta^2 Ax \quad (2)$$

ここで、 x は $x(jk)$ を要素とするベクトルである。

行列 A および B の要素は以下のように与えられる。

$$a(jk, j'k') = \frac{1}{m} \sum_{v=1}^g \sum_{i=1}^{m_v} \{n_i(v, jk) - \overline{n(jk)}\} \{n_i(v, j'k') - \overline{n(j'k')}\}$$

$$b(jk, j'k') = \frac{1}{m} \sum_{v=1}^g m_v \{\overline{n(v, jk)} - \overline{n(jk)}\} \{\overline{n(v, j'k')} - \overline{n(j'k')}\}$$

ここで、 g はグループ数、 m_v はグループ v 内の個体数を表し

$$m = \sum_{v=1}^g m_v$$

$$\overline{n(v, jk)} = \frac{1}{m_v} \sum_{i=1}^{m_v} n_i(v, jk)$$

$$\overline{n(jk)} = \frac{1}{m} \sum_{v=1}^g m_v \overline{n(v, jk)} = \frac{1}{m} \sum_{v=1}^g \sum_{i=1}^{m_v} n_i(v, jk)$$

である。

η^2 は(2)式の固有値となっているので、相関比の2乗を最大にするためには(2)式

の最大固有値に対応する固有ベクトルを $x(jk)$ の値とする。複数個の解を求めるときは、(2) 式の固有値のうち大きさの順に必要な個数を取り、対応する固有ベクトルを用いる。項目選択肢 (アイテム・カテゴリ) の値 $x(jk)$ が決まると、(1) 式より個体の得点 $y_i(v)$ も定まる。

解法

(2) 式を満たす \mathbf{x} を岩坪 (1987) の式を用いて以下のように求める。

\mathbf{A} の固有値分解を次式で表す。

$$\mathbf{A} = [\mathbf{T} \ \mathbf{T}_0] \begin{bmatrix} \Lambda_A & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T}' \\ \mathbf{T}_0' \end{bmatrix}$$

このとき、(2) 式は次のように変形できる (岩坪、(3.31) 式)、

$$\mathbf{T}'\mathbf{B}\mathbf{T}\mathbf{z} = \lambda\Lambda_A\mathbf{z}, \quad \lambda = \eta^2 \tag{ 3 }$$

ここで、

$$\mathbf{z} = \mathbf{T}(1k_1, \dots, nk_n)' \cdot \mathbf{U}(1k_1, \dots, nk_n)' \cdot \mathbf{x}$$

である (岩坪、1987、(3.30) 式)、 $\mathbf{T}(1k_1, \dots, nk_n)$ は、 \mathbf{T} の jk_j 行 (項目 j のカテゴリ k_j に

対応する行) を除いて縦に詰めた行列である。 $\mathbf{U}(1k_1, \dots, nk_n)$ は、 $\sum_{j=1}^n N_j$ 次単位行列から jk_j

列を取り除いて横に詰めた行列において jk_j 行の項目 j に対応する要素を -1 とおいたもの

である。項目 j における k_j は適当に選ぶ。

(3) 式を次のように変形する。

$$\Lambda_A^{-1/2} \mathbf{T}'\mathbf{B}\mathbf{T}\Lambda_A^{-1/2} \Lambda_A^{1/2} \mathbf{z} = \lambda \Lambda_A^{1/2} \mathbf{z}$$

いま

$$\mathbf{u} = \Lambda_A^{1/2} \mathbf{z} \tag{ 4 }$$

とおけば

$$\Lambda_A^{-1/2} \mathbf{T}'\mathbf{B}\mathbf{T}\Lambda_A^{-1/2} \mathbf{u} = \lambda \mathbf{u} \tag{ 5 }$$

と書けるので、 λ と \mathbf{u} は $\Lambda_A^{-1/2} \mathbf{T}'\mathbf{B}\mathbf{T}\Lambda_A^{-1/2}$ の固有値と固有ベクトルとして求めることができ

る。 \mathbf{u} が求まると、岩坪の (3.35) 式

$$\mathbf{w} = \mathbf{Tz} \quad (6)$$

および (3.39) 式

$$x(jk) = w(jk) - w(jk_j)$$

を用いて、 \mathbf{x} を求めることができる。すなわち、式 (4) と式 (5) から \mathbf{w} が

$$\mathbf{w} = \mathbf{T}\Lambda_A^{-1/2}\mathbf{u}$$

と求まる。

\mathbf{x} は、項目内のカテゴリに同じ値を加えても (2) 式の解であるという性質がある (付録参照)。この性質によって、 $x(jk)$ の項目内の平均が 0 になるように基準化する。すなわち、

$$x^*(jk) = w(jk) - \sum_{k=1}^{N_j} \bar{n}(jk)w(jk) \quad (7)$$

として基準化した最適解 $\mathbf{x}^* = (x^*(jk))$ を求める。

手順をまとめると、

(5) 式により、 \mathbf{u} を求める。

(4) 式より、 $\mathbf{z} = \Lambda_A^{-1/2}\mathbf{u}$ を求める。

(6) 式より、 \mathbf{w} を求める。

(7) 式より、 $x^*(jk)$ を求める。

となる。

(7) 式で与えられる解が平均値が 0 になるように基準化されていることは以下のように確かめられる。

$$\begin{aligned} \sum_k \bar{n}(jk)x^*(jk) &= \sum_k \bar{n}(jk)w(jk) - \sum_k \left[\bar{n}(jk) \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \right] \\ &= \sum_k \bar{n}(jk)w(jk) - \sum_k \left[\bar{n}(jk) \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \right] \\ &= \sum_k \bar{n}(jk)w(jk) - \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \sum_k \bar{n}(jk) \\ &= \sum_k \bar{n}(jk)w(jk) - \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \sum_k \left\{ \frac{1}{m} \sum_v \sum_i n_i(v, jk) \right\} \end{aligned}$$

$$\begin{aligned}
 &= \sum_k \bar{n}(jk)w(jk) - \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \left[\frac{1}{m} \sum_v \sum_i \left\{ \sum_k n_i(v, jk) \right\} \right] \\
 &= \sum_k \bar{n}(jk)w(jk) - \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \left[\frac{1}{m} \sum_v \sum_i 1 \right] \\
 &= \sum_k \bar{n}(jk)w(jk) - \left\{ \sum_{k'} \bar{n}(jk')w(jk') \right\} \times 1 \\
 &= 0
 \end{aligned}$$

付 録

$\sum_{j=1}^n N_j$ 次列ベクトルにおいて項目 j に対応する要素を c 、それ以外の要素を 0 とおいたものを次式のように \mathbf{d} とおく。

$$\mathbf{d}' = \begin{bmatrix} 0 & \cdots & 0 & \underbrace{c \cdots c}_{\text{項目}j\text{の}カテゴリ} & 0 & \cdots & 0 \end{bmatrix}$$

このとき \mathbf{Ad} の $(j_1 k_1)$ 行目の要素は以下ようになる。

$$\begin{aligned}
 \sum_{k=1}^{N_j} a(j_1 k_1) c &= \frac{c}{m} \sum_k \left[\sum_{v=1}^g \sum_{i=1}^{m_v} \{n_i(v, j_1 k_1) - \bar{n}(j_1 k_1)\} \{n_i(v, jk) - \bar{n}(jk)\} \right] \\
 &= \frac{c}{m} \sum_v \sum_i \left[\{n_i(v, j_1 k_1) - \bar{n}(j_1 k_1)\} \left\{ \sum_k n_i(v, jk) - \sum_k \bar{n}(jk) \right\} \right] \quad (\text{A1})
 \end{aligned}$$

ここで、

$$\sum_k n_i(v, jk) = 1$$

および

$$\sum_k \bar{n}(jk) = \sum_k \left\{ \frac{1}{m} \sum_{v'} \sum_{i'} n_{i'}(v', jk) \right\}$$

$$\begin{aligned}
 &= \frac{1}{m} \sum_{v'} \sum_{i'} \sum_k n_{i'}(v', jk) \\
 &= \frac{1}{m} \sum_{v'} \sum_{i'} 1 \\
 &= 1
 \end{aligned}$$

であることに注意すると、(A1) 式は

$$\frac{c}{m} \sum_v \sum_i \{[n_i(v, j_1 k_1) - \bar{n}(j_1 k_1)]\{1 - 1\}\} = 0$$

となる。すなわち、次式が成り立つ。

$$\sum_{k=1}^{N_j} a(j_1 k_1) c = 0$$

上式は **Ad** の任意の $(j_1 k_1)$ 行について成り立つ。したがって、

$$\mathbf{Ad} = \mathbf{0} \tag{A2}$$

である。

同様に、**Bd** の $(j_1 k_1)$ 行目の要素は次のように計算できる。

$$\begin{aligned}
 \sum_k b(j_1 k_1, jk) c &= \frac{c}{m} \sum_k \left[\sum_v m_v \{ \bar{n}(v, j_1 k_1) - \bar{n}(j_1 k_1) \} \{ \bar{n}(v, jk) - \bar{n}(jk) \} \right] \\
 &= \frac{c}{m} \sum_v \left[m_v \{ \bar{n}(v, j_1 k_1) - \bar{n}(j_1 k_1) \} \left[\sum_k \{ \bar{n}(v, jk) - \bar{n}(jk) \} \right] \right] \tag{A3}
 \end{aligned}$$

ここで、次式が成り立っている。

$$\begin{aligned}
 \sum_k \{ \bar{n}(v, jk) - \bar{n}(jk) \} &= \sum_k \left\{ \frac{1}{m_v} \sum_{i'} n_{i'}(v, jk) \right\} - \sum_k \left\{ \frac{1}{m} \sum_{v'} \sum_{i'} n_{i'}(v', jk) \right\} \\
 &= \frac{1}{m_v} \sum_{i'} \sum_k n_{i'}(v, jk) - \frac{1}{m} \sum_{v'} \sum_{i'} \sum_k n_{i'}(v', jk) \\
 &= \frac{1}{m_v} \sum_{i'} 1 - \frac{1}{m} \sum_{v'} \sum_{i'} 1 \\
 &= 1 - 1 \\
 &= 0
 \end{aligned}$$

故に (A3) 式は次のようになる。

$$\frac{c}{m} \sum_v [m_v \{ \bar{n}(v, j_1 k_1) - \bar{n}(j_1 k_1) \} \times 0] = 0$$

すなわち、

$$\sum_k b(j_1 k_1, jk)c = 0$$

が成り立つ。

上式は \mathbf{Bd} の任意の $(j_1 k_1)$ 行について成り立つ。したがって、

$$\mathbf{Bd} = \mathbf{0} \quad (\text{A4})$$

が成り立つ。

(A2) 式と (A4) 式より、 \mathbf{x} が

$$\mathbf{Bx} = \lambda \mathbf{Ax} \quad (\text{A5})$$

を満たすとき、

$$\mathbf{B}(\mathbf{x} + \mathbf{d}) = \mathbf{Bx} + \mathbf{Bd} = \mathbf{Bx} = \lambda \mathbf{Ax} = \lambda \mathbf{Ax} + \lambda \mathbf{Ad} = \lambda \mathbf{A}(\mathbf{x} + \mathbf{d})$$

が成り立つ。すなわち、 $\mathbf{x} + \mathbf{d}$ も (A5) 式を満たす。

参考文献

岩坪秀一「数量化法の基礎」朝倉書店、1987