Chapter 6
Identification of Types on Corresponding Relation

Yosuke Noda

When corresponding relations between individual codes with common connections between them exist between classification A and classification B (assuming that classification A and classification B are based on different classification systems), a collection of cases in which there are closed corresponding relations between the classification A and classification B codes is called a group. In order to make use of such groupings of corresponding relations, it is very important to know what sorts of corresponding relations exist between the two classifications. The basis for dividing correlations into types is determining how many links exist between each classification code and the classification codes of the other system.

According to Noda and Yamamoto, the combinations of links between grouped corresponding relations can be divided into four types: type 1, type 2, type 3, and type 4. Considering for the moment corresponding relations running from classification A to classification B, in type 1, there is a one-to-one correlation between the individual codes of classification A and classification B. In type 2 there is a one-to-many correlation. Multiple individual codes in classification B each have a single link with a single individual code in classification A, and the single individual code in classification A has links to multiple individual codes in classification B, constituting a collection of corresponding relations. In type 3 there is a many-to-one correlation, and the relationship is the opposite of type 2. Type 4 indicates a many-to-many correlation, and is a collection of correlations that are not type 1 through type 4.

This chapter explains how to further divide type 4 into two new types: type 4A and type 4B. It also introduces methods of calculating and estimating the distribution of weight, which is the basis for distinguishing between the two types. Calculation of the distribution of weight is how the weight of the statistical value of the classification A individual classification codes contained in a group of correlations is determined. This, in turn, makes it possible to assign the appropriate amount of weight to the various individual classification codes in classification B.

The number of classification A individual classification codes in a group is indicated as m, and the number of classification B individual classification codes in the group is n. The weight distribution structure is as follows.

	$B_1(b_1)$	$B_2(b_2)$	$\cdots$	$B_n(b_n)$
$A_1(a_1)$	w_{11}	w_{12}	$\cdots$	w_{1n}
$A_2(a_2)$	w_{21}	w_{22}	$\cdots$	w_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$A_m(a_m)$	w_{m1}	w_{m2}	$\cdots$	w_{mn}

Here A_1 is a classification A individual classification code and the a_1 in parentheses shows the statistical value related to individual classification code A_1. The other variables are analogous. The weight to be distributed from A_i to B_j is therefore the following.

$$\{w_{ij}\}\ i,j=1\cdots n$$

The distribution method is as follows. The statistical value a_1 corresponding to A_1 is divided into n parts. Using weight w_{11}, a portion is secured as the statistical value of B_1, a_1w_{11}, and a portion is secured as the statistical value of B_2, a_1w_{12}. Finally, a_1w_{1n} is secured as the statistical value of B_n, in the same manner. Therefore, the statistical value b1 of individual classification code B_1 is equivalent to the sum of: $a_1 w_{11}$ from A_1 and $a_2 w_{21}$ from A_2, ..., $a_m w_{m1}$ from A_m. Consequently, the relational expression for the corresponding relation is as follows.

$$a_1 w_{1j} + a_2 w_{2j} + \cdots + a_m w_{mj} = b_j \quad \cdots\cdots (1)$$

$$(j = 1 \cdots n)$$

If we also take into account the weighting conditions, we get the following.

$$w_{i1} + w_{i2} + \cdots + w_{in} = 1 \quad \cdots\cdots (2)$$

$$(i = 1 \cdots m)$$

We can then combine the relational expression for the correlation (1) and the weighting conditions (2) and rewrite it using a determinant as follows.

$$Xw = b \quad \cdots\cdots (3)$$

Here, the distribution weight w in relational expression (3) can be expressed as successive column vector mn, b as successive column vector $m+n$ which is the total of m parts 1 and m parts b_1 to b_n, and coefficient matrix X as matrix $(m+n)\times mn$ which is comprised of the conditional expression in which m rows from the top corresponds to (2) and the correlative expression corresponding to n rows in (1).

$$w = (w_{11} \quad \cdots \quad w_{1n} \quad w_{m1} \quad \cdots \quad w_{mn})'$$

$$X = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & & & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1 \\ a_1 & 0 & \cdots & 0 & \cdots & a_m & 0 & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & \cdots & a_1 & \cdots & 0 & 0 & \cdots & a_m \end{bmatrix}$$

$$b = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ b_1 \\ \vdots \\ b_2 \end{bmatrix} \quad \cdots\cdots (4)$$

The rank of this coefficient matrix X is $m+n-1$, so equation (3) does not have a unique solution of w. However, if the number of weights is reduced to create a corresponding relation in which the number of weight elements is equal to the rank of X or $m+n-1$, this equation has a unique solution w which can be expressed as follows.

$$w = X^{-1} b \quad \cdots\cdots (5)$$

In the actual corresponding relation, a single classification A individual classification code does not necessarily have n links which is the number of classification B individual classification codes. This signifies that some w_{ij} are 0. Consequently, a unique solution only exists when, among mn weights, $(m-1)(n-1)$ weights are 0 and $m+n-1$ weights are not 0.

The number of weights in the group which are not 0 equals the total number of correlative links in the group. Therefore, using this information allows type 4 to be divided into groups for which a unique solution can be found using an algebraic equation and groups for which a unique solution cannot be found. Groups for which the rank of coefficient matrix X in equation (3) is greater than $m+n-1$ and which do not have a unique solution w are classified as type 4B. Groups with a unique solution w are classified as type 4A. Applying this distribution weight identification method also allows types 1 to 3 to be identified as special patterns of type 4A.

If grouped correlations are type 4A, the number of weights which are not 0 is $m+n-1$. The successive column vector $m+n-1$ obtained by removing the 0 elements from this w is labeled ane w as w.

$$w = (w_{11} \cdots w_{1n} \; w_{22} \cdots w_{mm})'$$

Next, the $m+n$ element which is the lowest element is removed from b resulting in the successive column vector $m+n-1$.

$$b = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ b_1 \\ \vdots \\ b_{n-1} \end{bmatrix}$$

For coefficient matrix X, first the $m+n$ row which has the same number as b is removed to yield $m+n-1$ rows. Also, w_{ij} is moved from the upper left to the right. When the end of the row is reached, w_{ij} is moved from the left to right of the next row and so on to attach a series number. The series number corresponding to weight w_{ij} is $(i-1)\ n+j$. Removing the rows which correspond to w_{ij} 0 elements from the rows in X produces the successive matrix $(m+n-1)\times(m+n-1)$.

$$X=\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ a_1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & a_1 & 0 & \cdots & 0 & 0 & a_2 & \cdots & 0 \\ \vdots & & & & & & \vdots & & \vdots \\ 0 & 0 & & & & & 0 & \cdots & a_m \\ \vdots & \vdots & & & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Since $m<n$ in this example, values below am are 0. Whether the rank of X equals $m+n-1$ can now be checked and the w of weights which are not 0 can be derived using equation (5).

Type 3 corresponds to equation (4) with $m=1$. In other words,

$$w=(w_{11}\cdots w_{1n}\quad 0\cdots 0)'$$
$$b=(1\ 0\cdots 0\ b_1\cdots b_{n-1}\ 0)'$$

Here, the w elements which are not 0 are the n elements from w_{11} to w_{1n}, and the b_1 is the total of 1 part and b_1 to b_{n-1}. The series number rows corresponding to $b=0$ and series number columns corresponding to $w=0$ are removed from X. Next, the 0 elements are removed from w and b the results are labeled anew as w and b. Thus, X, w and b are as follows.

$$w=(w_{11}\quad \cdots \quad w_{1n})'$$

$$X=\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ a_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & & & a_1 & 0 \end{bmatrix}$$

$$b=\begin{bmatrix} 1 \\ b_1 \\ b_2 \\ \vdots \\ b_{n-1} \end{bmatrix}$$

Here, the rank of coefficient matrix X is n. Solving this produces the following values.

$$w_{1j}=b_j/(b_1+\cdots+b_n)\qquad j=1\cdots n$$

Type 2 corresponds to equation (5) with $n=1$. Here, the rank of coefficient matrix X is m. Solving this produces the following values.

$$w_{j1}=1\quad j=1\cdots m$$

For type 1, w and b are as follows in equation (4).

$$w=(w_{11}\ 0\cdots 0)'$$
$$b=(\ 1\quad 0\cdots 0)'$$

Therefore, solving for X, w and a yields $w=w_{11}$, $b=1$ and $X=1$, with w_{11} equaling 1. Thus, the rank of coefficient matrix X for type 1 is 1.

For type 4B corresponding relations where weight w does not have a unique solution, methods of obtaining the weight through repeated calculations using the weight obtained as a priori information as an initial value can be considered. One of these methods is the RAS approximation method conceived by R. Stone.

In conclusion, assuming the number of classification A and classification B individual classification codes to be m and n, respectively, correlation types can be identified by the fact that the rank of coefficient matrix X is 1 for type 1, m for type 2, n for type 3, $m+n-1$ for type 4A, and greater than $m+n-1$ for type 4B corresponding relations.