A Generalised Formulation of Multivariate Spatial Indices for Formal Regional Analysis

M. N. Pal*

Introduction

In regional analysis, often we have to identify and evaluate varying regional configurations of activities or other characteristics and factors of growth and also their inter-relationships. In identifying a simple regional configuration related to a single characteristic, a simple measure like the location factor or location quotient is in use. However, for a full-fledged regional analysis, the identification of a simple regional configuration or even a bunch of such simple configurations may be necessary but is never sufficient. The identification and evaluation of a composite regional configuration, based on many interrelated variables, become essential. Even a simple regional configuration may be the result of interactions of several other causal variables, in which case an analysis and evaluation with many interrelated variables should follow for making rational regional planning decisions. A complex or composite regional configuration is usually identified by mapping inter-related variables by the geographic technique of superposing. This technique has been found to be quite useful in evaluating the regional patterns and local peculiarities for a composite characteristic involving only two inter-related variables. But the superposition technique is not of much help when the composite characteristic is to be evaluated on the basis of many spatial variables, since, in that case, the combinations of different classes of all such variables become numerous. Again it is possible to identify the areas of different ranks in terms of a single variable, implicit in a simple regional configuration. But the composite ranking of areas with respect to a vector of several variables that are superimposed together to identify a composite regional configuration is very difficult. In view of these problems, the identification and evaluation of multivariate composite regional configuration has been dealt with the formulation of composite index (or indices) which is used in the depiction of general spatial trend, pattern or association and also the local peculiarities. For the formulation of a particular composite index, various methods of combining the constituent variables are in use. At the moment, we have no unified procedure of assessing the relative merits of the composite indices formulated from the same set of variables by various statistical and non-statistical methods. Question is : how a composite index, whatever be its method of construction, represents its different constituent variables *specifically* and *aggregatively*.

*The author is Professor in Indian Statistical Institute, Calcutta.

The statistically constructed composite spatial indices usually take into consideration, explicitly or implicitly, this aspect of representativeness. Thus Kendall's formulation [1939] is based on the *aggregate representation maximising principle* and called an *optimal formulation*, while Pal's formulation [1963, 1971] is based on the *Specific representations equalising principle* and called an *equity formulation*. Here, by specific representation of any constituent variable, we mean its linear correlation coefficient with the corresponding composite index. The squared aggregate representation is the arithmetic mean of squared specific representations of all constituent variables. However, as the specific representations are measured by the linear correlation coefficients, it is necessary to examine for the similarity of statistical distributions of the variables and make suitable mathematical transformation so that the intrinsic relation between a variable and the index could be depicted by the linear correlation coefficient.

The statistical formulations can however be compared on the basis of specific and aggregate representativeness [ref. to Pal and Chattopadhyay, 1972-73]. But, at the moment, different non-statistical formulations with the same set of variables do not permit any comparison on their relative merits in respect of representativeness. For example, Kendall's regional illustration [1939] is an optimal index formulation, which can be called an agricultural land productivity index with yield rates of different crops as its constituent spatial variables. The same land productivity index could also be constructed from the same set of spatial variables, with, say, some price weights under economic consideration or with some physical weights like relative acreages under other non-statistical consideration. A purpose of this paper is to bring under one generalised fold the different statistical and non-statistical formulations so that they could be compared and contrasted for representativeness.

The ratio of specific representation to aggregate representation would be called here as the *representativeness multiplier*, or simply the *multiplier*, for any constituent variable. The vector of all multipliers for a composite index formulation would be designated by m. By a direct computation of the specific and the aggregate representations, the vector m can be estimated for any existing formulation, whether statistical or non-statistical. The vector m becomes the sum vector S (with any of its elements equal to unity) for the equity formulation and that for the optimal formulation is designated by M. There can be various modes of generating the vector m, say, for example, by a convex combination of M and S, or by a power transformation on M, when $M \neq S$. The vector M depicts a kind of ordering of constituent variables, showing the relative importance of variables as derived from the empirical optimality consideration. The vector S treats all constituent variables as equally important Any convex combination of M and S depicts the same order of importance of constituent variables as depicted by M, but the divergence of importance falls in between those depicted by M and S.

A power transformed formulation can also be so designed as to achieve a reduced representativeness divergence compared with those of the optimal formulation. Again, by some other considerations of real world situation, the relative order and magnitudes of importance can be ascertained *a priori*, which may be different from that depicted by either M or S or their convex combinations. Thus there are possibilities of generalised formulation with different multiplier vector m, matching particular situations of multivariate regional analysis.

With any known, derived or pre-assigned multiplier vector m, the mathematical details of any generalised formulation, determining the aggregate representation and the vector of combining weights for the constituent variables, have been established in this paper. The different existing formulations are also shown as special cases of the generalised formulation. The mathematical formula for estimating the relationship between any two modes of generalised formulation has also been shown here so that one can readily examine how close a particular formulation is, say, to the optimal or any other formulation. Finally, it is emphasized here that by bringing all existing formulations under the common fold of generalised formulation through the vector m, it has been possible to get an unified procedure of assessing their relative merits statistically with the parameters estimated within the generalised framework. Before attempting for the generalised formulation, a brief review on the computation procedure is given below for the two fundamental statistical formulations.

The Optimal and the Equity Formulations : Brief Review

Suppose n spatial variables, each varying over N spatial units of observations, are related to a composite characteristic. These variables, after appropriate mathematical transformations, ale designated by $x_1, x_2, \ldots x_i, \ldots x_n$. The standardized variable for x_i is designated by Z_i where $Z_i = (x_i - \bar{x}_i)/\sigma_i$, with $\bar{x}_i$ denoting the mean value and σ_i the standard deviation. *The optimal formulation* : Kendall's composite index [1939], denoted by I_h can be written algebraically as follows for the jth spatial unit of observation ; $j = 1, 2, \ldots N$.

$$I_{hj} = (a_1 x_{1j} + \ldots + a_n x_{nj})/(a_1 + \ldots + a_n), \qquad \ldots (1)$$

In this formulation, we have $a_i = r_i/\sigma_i$, where the specific representations r_i's and the corresponding aggregate representation ρ_h are solvable from the matrix equation :

$$R.r = n\,\rho^2_h.r\,, \qquad \ldots (2)$$

where $R = ((r_{ij}))$ = the known correlation matrix of the variables, r = the column vector of specific representations, $n\,\rho^2_h = r'.r$, with r' denoting the transposed vector of r. For solving the matrix equation (2), an iterative procedure of computation after

Hotelling [1933] is used. In this iterative procedure at the ith stage of iteration with the estimates of ith weight vector W_i, we go for the calculation of the next weight vector as follows :

$$R.W_i = \Sigma_{i+1} ,$$

$$\text{and} \quad W_{i+1} = \Sigma_{i+1}/H_{i+1} , \qquad ...(3)$$

where H_{i+1} is the highest element in the calculated vector Σ_{i+1}. Combining the above computational steps, we can write

$$R.W_i = H_{i+1}.W_{i+1} \qquad .. (4)$$

The initial vector W_o can be any arbitrary vector. However one can preferably take

$$W_o = \Sigma_o/H_o ,$$

where Σ_o is the vector formed by the row sums of the elements of R and H_o the highest element in Σ_o. The iteration stops when we get the stable weight vector W_s with the condition that

$$W_{s-1} = W_s \qquad .. (5)$$

Any subsequent iteration will not change the weight vector further. From the above relation we can deduce

$$R.W_s = H_s.W_s , \qquad .. (6)$$

which is comparable to relation (2) in structural form. As such vector r can be taken as proportional to the vector W_s, and we get :

$$H_s = n\, \rho^2_h = r'r \qquad .. (7)$$

$$\text{and} \quad r = W_s \sqrt{H_s/W_s'\, W_s} \qquad .. (8)$$

The equity formulation : Pal's equity index [1963, 1971], designated by I_e, can be written algebraically as follows for the jth spatial unit of observation :

$$I_{ej} = (b_1 x_{1j} + ... + b_n x_{nj})/(b_1 + ... + b_n) \qquad .. (9)$$

In this formulation, we have $b_i = w_i/\sigma_i$, where the aggregating weight vector w with elements w_i's and also the common specific representation ρ_e are solvable from the matrix equations :

$$S'.w = 1$$

$$\text{and} \quad R.w = \rho^2_e.S , \qquad . (10)$$

where S is the sum vector (column vector with each of its elements equal to unity) and S′ is the transpose of vector S.

The central critical value of a location factor (or its power transforms) is always unity. The two formulations (1) and (9) are presented here in a form that suits the uses of location factors (or their power transforms) as the constituent variables $x_1, \ldots, x_n$. In any of these two formulations, the central critical value is again unity, likewise the constituent location factors. The generalised formulation to be deduced later will however be given in the standardized form, from which the above-mentioned form can easily be arrived at by a simple linear transformation.

The two formulations are however coincident in the two variable space, i.e., when $n = 2$. Thus the question of preference between the optimal and the equity formulations does not arise when $n = 2$. For this reason, at times final composite index can be formulated by a sequential application of the common formulation through a number of intermediate stages with the consideration of only a pair of variables or sub-indices at each stage. Such a final composite index is then neither the optimal index nor the equity index constructed at a single stage with all the variables together, but stands some where in between the two. In the higher variable spaces ($n > 2$), the two formulations are different. In such cases the specific represetation r_i's are generally different in the optimal formulation. As stated already, r_i's are same always in the equity fomulation. It is clear from the very mode of optimal formulation that $\rho_n \geqq \rho_e$. The virtue of the equity formulation lies in the similarity of representativeness of all variables in the composite index, which cannot be expected for an optimal formulation with $n > 2$. For some of the constituent variables, the specific representations of the optimal formulation may have undesirably low values, much below ρ_e.

Each of the composite indices formulated above can serve as a depicter of general spatial trend or pattern and the local peculiarities of any constituent variable, not conforming the general pattern, can be identified by marking the significant departures from the regression of the spatial variable on the composite index. This provides a very useful aid to the cartographic methods for a formal regional analysis. In this approach of multivariate analysis, the unexplained variation of the variable-space is really considered for the local peculiarity analysis. It should be noted that if the divergence of specific representations is too great in an optimal index, the local peculiarity analysis by the above procedure becomes difficult for variables with low specific representations. In the equity formulation, the local peculiarity analysis can be tackled without any bias towards any constituent variable, for reasons of similarity of representations in it. Thus, even though the optimality criterion appears to be more alluring generally, the equity criterion is held more useful for spatial trend and peculiarity analysis, particularly when the loss in the aggregate representativeness of the equity index is not significantly different from that of the optimal index [Pal and Chattopadhyay 1972-73].

As stated already the multiplier vector for the equity fomulation is the sum vector S and that for the optimal formulation is designated by M. The multiplier vector m for a convex combination of the two formulations can be written as

$$m = \delta M + (1 - \delta) S ,$$

with varying values of δ, $0 \leq \delta \leq 1$. This shows particular generalisation possibilities in between the two formulations. For any arbitrary multiplier vector m, we have $m'.m = n$, since $r'r = n \rho^2$. This means that the average of all m^2_i of vector m (and hence of M^2_i) is unity. So different m^2_i's are in general above and below unity. The values of m_i's can be both positive and negative. However, through a proper transformation of the initial variables, all m_i's can be made non-negative in most of the situations.

Development of the Generalised Formulation

The Mathematical Framework :

For developing the mathematical details of the proposed generalised formulation, we start with the representativeness multiplier vector m (a column vector with elements m_i's), subject to the condition

$$m'.m = n \tag{11}$$

(i.e., the arithmetic mean of all m^2_i adjusted to unity) for the n constituent spatial variables $Z_1, Z_2, \ldots , Z_n$ in standardized form (corresponding to unstandardized variable $x_1, x_2, \quad , x_n$). As stated already, the vector m is known or derived or preassigned from some external condition. From the empirical data on N spatial units of observation for the variables, the correlation matrix R can be determined, denoted by

$$R = (r_{ij}), \text{ with } r_{ji} = r_{ji}, r_{ii} = 1, \text{ and}$$
r_{ij} denoting the correlation coefficient between Z_i and Z_j.

Here R is a *positive definite* matrix. Let us designate by Z_0 the generalised composite index to be constructed by a weighted aggregation of variables $Z_1, Z_2, \quad , Z_n$, with the relative representativeness as depicted by the vector m. We have to solve for the *aggregating weight* vector, designated by w (a column vector with elements w_i's) and also the aggregate representation ρ. Once ρ is solved, we know the specific representations vector $r = \rho m$ (a column vector with elements r_i's). By definition of specific representation, we have $r_i = r_{0i}$ = the correlation coefficient between Z_0 and Z_i for $i = 1, 2, \quad , n$.

We shall denote the extended correlation matrix of variables $Z_0, Z_1, \ldots , Z_n$ by R^*, so that we can write

$$R^* = \begin{pmatrix} 1 & \rho.m' \\ \rho.m & R \end{pmatrix} = \begin{pmatrix} 1 & r' \\ r & R \end{pmatrix}$$

Further, the *weight vector* w is extended to w* with the incorporation of $-\rho$ as its first element along with the elements of w.

$$\text{i.e. } w^* = \begin{pmatrix} -\rho \\ w \end{pmatrix}$$

Clearly we have to solve for the elements of w*.

Next consider the matrix equation :

$$R^*.w^* = 0 \qquad \text{...(12)}$$

This equation implies that

$$m'.w = 1, \text{ i.e., } \sum_j m_j w_j = 1 \qquad \text{...(13)}$$

and

$$R.w = \rho^2.m, \text{ i.e., } \sum_j r_{ij}\, w_j = \rho^2 m_i \, ; \; i = 1, 2, \dots n. \qquad \text{..(14)}$$

Eliminating m from the system of equations (13) and (14), we get

$$w' R w = \rho^2 \qquad \text{...(15)}$$

Then, if we take

$$Z_0 = \sum_j (w_j / \rho) Z_j , \qquad \text{...(16)}$$

we have the expectation $E(Z_0) = 0$,

and the variance

$$\begin{aligned} V(Z_0) &= E(Z_0^2) \\ &= \sum_i \sum_j w_i\, w_j\, r_{ij} / \rho^2 \\ &= w' R w / \rho^2 \\ &= 1, \text{ using relation (15)} \end{aligned}$$

Again

$$\begin{aligned} r_i = r_{0i} &= E(Z_0 Z_i) \\ &= \sum_j r_{ij}\, w_j / \rho \\ &= \rho\, m_i , \text{ using relation (14)} \end{aligned}$$

Thus Z_0 can be taken as the standardized weighted aggregation of the given variables $Z_1, \dots, Z_n$, so that R* can truely be taken to stand for the correlation matrix of $Z_0, Z_1, \dots, Z_n$. Clearly, if there exists a solution to the matrix equation (12) for w*, we are in a position to get the generalised formulation (16) (in standardized form) and also estimate the related statistical parameters. Note that the following linear transformation of Z_0 gives the *generalised index formulation* I_0 in the form comparable to those given in formulation (1) and (9) for any jth spatial unit of observation ; j = 1, 2, ... N.

$$I_{0j} = \left(\rho Z_{0j} + \sum_i c_i \bar{x}_i \right) / \left(\sum_i c_i \right),$$

where $c_i = w_i / \sigma_i$

Then we have

$$I_{0j} = (c_1 x_{1j} + \dots c_n x_{nj}) / (c_1 + \dots + c_n) \quad \dots(17)$$

which is in the form comparable with formulations (1) and (9). The index I_0 retains the magnitudes of all specific and aggregate representations same as those established for the standardized index Z_0.

The existence of a non-trivial solution vector w* for equation (12) will now be established in the form of theorems given below.

Theorems on the Development of Generalised Formulation

Theorem 1 : The solution vector w* always exists for the equation

$$R^* w^* = 0$$

Proof : For existence of non-trivial solution, we have simply to prove that R* is a singular matrix. Since R* is a correlations matrix of the variables of which one is a linearly dependent variable of all others, it can be easily proved to be a singular matrix. Analytically, if $R_{0\cdot12\dots n}$ is the multiple correlation coefficient with Z_0 as the regressor and $Z_1, \dots, Z_n$ as regressants, then we must have

$$R^2_{0\cdot12\dots n} = 1,$$

by the construction of Z_0. Further, we have [ref to Johnston, 1972]

$$1 - R^2_{0\cdot12\dots n} = \det R^* / \det R.$$

Since R is non-singular, it follows that

$$\det R^* = 0$$

That is, R* is a singular matrix. Hence the solution vector w* exists.

Theorem 2 : Given m'.m = n and R* singular, we must have

$$m' R^{-1} m \geqq 1$$

and

$$\rho^2 = 1 / (m' R^{-1} m).$$

Proof : Since R* is singular

$$0 = \det R^* = \rho^2. \det \begin{pmatrix} 1/\rho^2 & m' \\ m & R \end{pmatrix}$$

and by Cauchy's expansion [Aitken, 1942], we have

$$0 = \det R - \rho^2 \sum_{i,j} m_i m_j R_{ij}, \quad \dots(18)$$

where R_{ij} is the cofactor of r_{ij} in R.

From relation (18), we get

$$1/\rho^2 = \sum_{i,j} m_i m_j R_{ij} / \det R$$

$$= m' R^{-1} m,$$

where R^{-1} is the inverse matrix, which exist because R is positive definite,

Hence $\rho^2 = 1/(m' R^{-1} m)$, ...(19)

This proves the second part of the theorem.

Again by Cauchy's expansion

$$\det \begin{pmatrix} 0 & m' \\ m & R \end{pmatrix} = - \sum_{i,j} m_i m_j R_{ij}$$

Hence relation (18) can be rewritten as

$$0 = \det R + \rho^2 \det \begin{pmatrix} 0 & m' \\ m & R \end{pmatrix},$$

and using relation (19) we get

$$m' R^{-1} m = - \det \begin{pmatrix} 0 & m' \\ m & R \end{pmatrix} \Big/ \det R$$

By applying Schweinsian expansion [Aitken, 1942] on the ratio of symmetric determinants on the right, we get

$$m' R^{-1} m = m_1^2 + \sum_{k=2}^{n} [\{\det R\,(k,m)\}^2/\{\det R\,(k-1) \det R\,(k)\}], \qquad ...(20)$$

where R (k) is the leading principal minor matrix of kth order of R ; k=2, , n, and R (k, m) is the matrix obtained by replacing the last column of R (k) by the first k elements of the column vector m. Clearly the right hand side of (20) gives the sum of all positive terms and hence we can write

$$m' R^{-1} m \geqq m_1^2$$

Now, since $m'm = n$, the average of all m_i^2 is unity and values of m_i^2 are above and below unity. So, without any loss of generality, we can rearrange the suffixes of the variables in such a way that m_i^2's are in decreasing order. Then we have $m_1^2 \geqq 1$, and hence $m' R^{-1} m \geqq 1$. Hence the proof is complete.

Note that the first part of the theorem proves the essential property of the aggregate representation that : $\rho^2 \leqq 1$.

Cor. 1 : ρ^2 can as well be expressed as follows :

$$\rho^2 = \det R / \sum_{k} m_k R_{k(m)} , \qquad ...(21)$$

where $R_{k(m)}$ is the determinant of matrix R with its kth column replaced by the column vector m.

This readily follows, since we have

$$m' R^{-1} m = \sum_k \sum_i m_k m_i R_{ik}/\det R$$

$$= \sum_k m_k \sum_i m_i R_{ik}/ \det R$$

$$= \sum_k m_k R_{k(m)} / \det R ,$$

Theorem 3 : Any element w_j of the weight vector w is given by

$$w_j = R_{j(m)} / \sum_k m_k R_{k(m)} .$$

Proof : From the matrix equation $R^* w^* = 0$,

$$m' w = 1 \text{ and } R w = \rho^2.m$$

$$w = \rho^2 R^{-1} m$$

Using relation (21), we get

$$w = (\det R). R^{-1} m/ \sum_k m_k R_{k(m)}$$

$$\text{or} \quad w = \frac{(R_{ij}).m}{\sum_k m_k R_{k(m)}},$$

where (R_{ij}) is the matrix of cofactors R_{ij}.

$$\therefore \quad w_j = \sum_i m_i R_{ij}/ \sum_k m_k R_{k(m)}$$

$$\text{or} \quad w_j = R_{j(m)} / \sum_k m_k R_{k(m)} , \qquad \text{.. (22)}$$

Cor. 2 : The solution vector w^* for the matrix equation $R^* w^* = 0$ is given dy the relations (21) and (22).

This completes the determination of the generalised formulation in any of the forms (16) and (17). Formula to measure the relationship between any two arbitrary generalised formulations is established in the next sub-section.

Correlation Between Any Two Generalised Formulations

Theorem 4 : The correlation coefficient between any two generalised formulations, designated with symbols α and β, is given by any of the following relations :

$$r^2_{\alpha\beta} = \{ m'(\alpha).w(\beta) \} \{ m'(\beta).w(\alpha) \},$$

or $$r_{\alpha\beta} = r'(\beta).w(\alpha)/\rho_\alpha ,$$

or $$r_{\beta\alpha} = r'(\alpha).w(\beta)/\rho_\beta .$$

Proof : Following the generalised formulation as shown in (16), the standardized composite indices can be written as

$$Z_0(\alpha) = \sum_j w_j(\alpha) Z_j/\rho_\alpha$$

and $$Z_0(\beta) = \sum_i w_i(\beta) Z_i/\rho_\beta$$

Then the correlation between $Z_0(\alpha)$ and $Z_0(\beta)$:

$$r_{\alpha\beta} = E\{ Z_0(\alpha).Z_0(\beta) \}$$

$$= \sum_i \sum_j w_j(\alpha).w_i(\beta) r_{ij} / \left\{\rho_\alpha \rho_\beta\right\}$$

$$= \left\{\sum_i w_i(\beta) \sum_j w_j(\alpha) r_{ij}\right\} \Big/ \left\{ \rho_\alpha \rho_\beta \right\}$$

$$= \left\{\sum_i w(\beta).m_i(\alpha) \rho_\alpha\right\} \Big/ \rho_\beta$$

$$\text{i.e. } r_{\alpha\beta} = \rho_\alpha m'(\alpha).w(\beta)/\rho_\beta$$

We can alternatively deduce :

$$r_{\alpha\beta} = \rho_\beta m'(\beta).w(\alpha)/\rho_\alpha$$

From these two relations, we get

$$\left.\begin{array}{lll} & r^2_{\alpha\beta} & = \{ m'(\alpha).w(\beta) \} \{ m'(\beta).w(\alpha) \}, \\ \text{or} & r_{\alpha\beta} & = r'(\alpha).w(\beta)/\rho_\beta , \\ \text{or} & r_\alpha & = r'(\beta).w(\alpha)/\rho_\alpha . \end{array}\right\} \quad \ldots (23)$$

Cor. 3 : If $\rho_\alpha < \rho_\beta$, then we have

$$m'(\beta).w(\alpha) < r_{\alpha\beta} < m'(\alpha).w(\beta).$$

Some Special Cases of the Generalised Formulation

Equity Formulation : In the equity index referred to in section 2, we have

$$r_1 = r_2 = \ldots = r_n = \rho_e$$

i.e., $m_1 = m_2 \ldots = m_n = 1.$

Hence we can conclude :

Cor. 4 : The generalised formulation reduces to the equity formulation when the multiplier vector coincides with the sum vector S. The equity formulation will be designated hereafter words with the symbol e. Then it readily follows from relations (19), (21) and (22) that

Cor. 5 : The aggregate or any specific representation of the equity formulation ρ_e is given by

$$\rho_e^2 = 1/(S' R^{-1} S) = \det R / \sum_k R_{k(S)}, \quad \text{.. (24)}$$

and the elements aggregating weight vector w (e) are given by

$$w_j(e) = R_{j(S)} / \sum_k R_{k(S)}, \quad j = 1, \ldots, n. \quad \text{.. (25)}$$

Theorem 5 : The generalised matrix equation $R^* w^* = 0$ reduces to $(R^2 - \rho_e^2 J)\, w(e) = 0$ with $S'w(e) = \sum_j w_j(e) = 1$, in the equity formulation, where J is the nth order matrix containing the sum vector in each of its columns. Further, ρ_e^2 can as well be expressed as follows :

$$\rho_e^2 = \det R / \{ \det(R + J) - \det R \},$$

Proof : In the equity formulation, $R^* w^* = 0$ reduces to

$$S' w(e)\ 1 \text{ and } R.w(e) = \rho_e^2.S$$

Then

$$R.w(e) = \rho_e^2. \begin{pmatrix} S'w(e) \\ \vdots \\ S'w(e) \end{pmatrix} = \rho_e^2. \begin{pmatrix} S' \\ \cdot \\ S' \end{pmatrix} \cdot w(e)$$

$$\left.\begin{array}{l} \text{or} \quad R.w(e) = \rho^2_e.J.w(e), \\ \text{or,} \quad (R - \rho^2_e J).w(e) = 0, \\ \quad \text{with} \quad S'w(e) = 1 \end{array}\right\} \qquad ...(26)$$

Again if λ is any scalar, we can prove

$$\det(R + \lambda J) = \det R + \lambda \sum_k R_{k(S)},$$

Putting $\lambda = 1$, we have $\det(R + J) = \det R + \sum_k R_{k(S)}$,

So we can write

$$\det(R + \lambda J) = \det R + \lambda \{\det(R + J) - \det R\}, \qquad ...(27)$$

Now for the existence of non-trivial solution of equation (26), we must have

$$\det(R - \rho^2_e J) = 0$$

Using relation (27), we get

$$\det R - \rho^2_e \{\det(R + J) - \det R\} = 0.$$

Hence,

$$\rho^2_e = \det R / \{\det(R + J) - \det R\}, \qquad ...(28)$$

Theorem 6 : ρ^2_e decreases with the increase in the number of constituent variables n.

Proof : This follows readily from the relation (20) shown in Theorem 2 with the replacement of vector m by vector S.

Note : This result shows that for large n, it is better to orient the multivariate regional and cartographic analysis in two (or more) stages by dividing the set of variables related to a composite characteristic into suitable disjoint subsets of variables reflecting two (or more) sub-characteristics of the composite characteristic, particularly when ρ_e is not high enough for large n. The sub-equity indices can be formed from these subsets at the initial stage and finally the equity index can be constituted, combining the sub-equity indices. The corresponding regional and cartographic analysis is thus oriented in two (or more) stages.

Lemma 1 : To prove that, for any general formulation $\bar{m}^2 = (\Sigma m_i/n)^2$ cannot exceed unity, and the maximum value of unity is attained only in the equity formulation.

Proof : For the maximum value of $\bar{m}^2$ subject to $\Sigma m_i^2 + n$, we put

$\Phi + (\Sigma m_j)^2 / n^2 - \mu (\Sigma m_i^2 - n)$,

where μ is a Lagrange multiplier.

By the first order condition of maximisation

$$O = \frac{1}{2} \frac{\partial \Phi}{\partial m_j} = \Sigma m_i / n^2 - \mu m_j ; \quad j = 1, \ldots, n$$

or $\quad \mu m_j = \bar{m} / n$, for all $j = 1, \ldots, n$.

Summing over j equating, we get

$$\mu = 1 / n.$$

So for the maximum value, we have

$$m_j = \bar{m}, \text{ for all } j = 1, \ldots, n$$

As $\Sigma m_j^2 = n$, we have $\sum_j \bar{m}^2 = n$, i.e., $m^2 = 1$.

So $m^2_{max} = 1$, which is attained when all m_j's are equal. Thus, in general, we have

$$m^2 \leqq 1 \text{ and } \bar{m}_{max} = 1 \text{ only when } m = S.$$

Note · $m^2 \leqq 1$ implies $-1 \leqq \bar{m} \leqq 1$. But by our choice and suitable mathematical transformations of variables, all m_i's can be expected to be positive. Thus we can usually get $0 < \bar{m} \leqq 1$ for a general index formulation, while the arithmetic mean of all m_j^2 must be unity. So for a preassigned multiplier vector these conditions must be met.

Optimal Formulation

Theorem 7 : The generalised formulation reduces to the optimal formulation, if

$$w = m / n = r / (n \rho)$$

Proof : From relation (19), we have

$$\frac{1}{\rho^2} = m' R^{-1} m$$

$$= \sum_i \sum_j m_i m_j R_{ij} / \det R,$$

subject to constraint $m'm = \Sigma m^2_j = n$.

Putting

$$\Phi = \sum_i \sum_j m_i m_j R_{ij} / \det R - \mu (\sum_i m_i^2 - n),$$

where μ is a Lagrange multiplier, and using the first order condition for maximisation of ρ^2 (i.e., minimisation of $1 / \rho^2$) under the constraint, we get

$$O = \frac{1}{2} \frac{\partial \Phi}{\partial m_k} = \sum_j m_j R_{kj} / \det R - \mu m_k$$

or $$R_{k(m)} / \det R = \mu m_k \; ; \; k = 1, \dots, n, \qquad \dots(29)$$

and also $\Sigma m_k^2 = n$.

From these relations, we can deduce

$$\sum_k m_k R_{k(m)} / \det R = \mu \sum_k m_k^2$$

$$= n \mu$$

Dividing relation (29), by this relation, we get

$$R_{j(m)} / \sum_k m_k R_{k(m)} = m_j / n \; ; \quad j = 1, \dots, n.$$

Using relation (22), we get

$$w_j = m_j / n$$

Hence $$w = m / n,$$

and as $m \rho = r$, we have $w = r / (n \rho)$.

Thus, using the symbol h for optimal formulation, the condition for this formulation can be written as

$$\left. \begin{array}{l} n \rho_h w(h) = r(h) \\ \text{or} \quad n w(h) = M \end{array} \right\} \qquad \dots(30)$$

Note that the vector *M is not predetermined, but to be solved with* the use of this condition.

Cor. 6 : In the optimal formulation, the generalised equation,

$R^* w^* = 0$, reduces to

$$(R - n \rho_h^2 I) r(h) = 0$$

with the constraint $r'(h).r(h) = n \rho_h^2$,

where I represents the nth order indentity matrix.

Proof : In the optimal formulation $R^* w^* = 0$ reduces to

$M'.w(h) = 1$ and $R.w(h) = \rho_h^2.M$

Now by use of condition (30), it follows readily that

$R.r(h) = n\ \rho_h^2.r(h)$

i.e. $(R - n\ \rho_h^2\ I).h) = 0$

and $r'(h).r(h) = n\ \rho_h^2$ .. (31)

These equations correspond to the estimation procedure already reviewed in section (2).

Cor. 7 : Obviously, $\rho^2 \leqq \rho_h^2$,

and, in particular, we have $\rho_e^2 \leqq \rho_h^2$.

Theorem 8 : The equity formulation becomes indentical with the optimal formulation, if

$$\sum_j r_{ij} = C \geqq 0, \text{ for all } i = 1, \quad , n$$

where C is a constant quantity. (i.e., if R is a symmetric circulant matrix [Aitken, 1942] with a positive sum C of its n independent elements).

Proof : Let us suppose that the two formulations are identical. For optimal formulation we have $w = m / n$.

As the two formulations are identical, their multiplier vectors must be same and thus we have $w = S / n$.

For any general formulation, we have

$R.w = \rho^2.m$

$R.S/n = \rho^2.S$

or $R.S = n\ \rho^2.S$

i.e. $\sum_j r_{ij} = n\ \rho^2 = C$, say, for all $i = 1, \quad , n$.

Hence R is a symmetric circulant matrix. Here the common aggregate representation is given by $\rho = \sqrt{C/n}$.

Conversely, let $\sum r_{ij} = C > 0$, for all $i=1, \quad , n$.

i.e. $R\ S = C\ S$ or $S'R = CS'$ .. (32)

$\therefore \quad S = C\ R^{-1}\ S$

or $S'S = C\ S'R^{-1}S$

or $n = C/\rho^2_e$

or $C = n\ \rho_e^2$, .. (33)

Now for optimal formulation, we have

$$R.r(h) = n\ \rho_h^2.r(h) \quad \ldots(34)$$

or $$S'R.r(h) = n\ \rho_h^2.S'r(h)$$

Using relation (32), we have

$$C\ S'.r(h) = n\ \rho_h^2.S'r(h)$$

$$\therefore\quad C = n\ \rho_h^2.$$

From relation (33), we have

$$\rho_e^2 = \frac{C}{n} = \rho_h^2 \quad \ldots(35)$$

Again from relation (32), we get

$$R\ S = C\ S \quad \text{or} \quad R\ S = n\ \rho_h^2.S$$

and comparing this equation with equation (34), we infer that

r(h) is proportional to S

$$\therefore\quad M = S.$$

Hence the formulations are identical.

Cor. 8 : If $r_{ij} = a$ (constant) for all i,j ; $i \neq j$ in R, then the optimal and the equity formulations coincide and the common solution is given by

$$w = S/n = M/n$$

and $$\rho = \sqrt{\{1 + (n-1)\,a\}/n} = \rho_e = \rho_h$$

provided $a > 0$.

Note : It is possible to accommodate a negative r_{12} for $n=2$. In such a case, one of the variables, say, Z_2 can be changed to $-Z_2$, so that, again, we have $a > 0$. So, without any loss of generality, we shall assume $a > 0$ for $n = 2$ in Cor. 8.

Cor. 9 : The equity formulation is always coincident with the optimal formulation for $n = 2$.

In such a case $w_1 = w_2 = \frac{1}{2}$

and $\rho_e^2 = \rho_h^2 = (1 + r_{12})/2$, with $r_{12} > 0$.

Cor. 10 : If $R = I$ (i.e., all variables are mutually independent), then the specific representations of n variables in the common equity or optimal index are all equal to $1 / \sqrt{n}$.

Cor. 11 : Correlation coefficient r_{he} between the optimal and the equity formulations is given by any of the following relations :

$$r_{he} = \rho_e \bar{M} / \rho_h, \text{ where } \bar{M} = \Sigma M_i / n,$$

or

$$r^2_{he} = \{ M'.w(e) \} \{ S'.w(h) \} = \{ M'.w(e) \} . \bar{M}$$

Proof : The relation readily follows from theorem 4 and also from the fact that

$$\bar{M} = \sum_i M_i/n = \sum_i w_i (h) = S'.w(h)$$

Cor. 12 : We have $M'.w(e) = \rho_e^2 . \bar{M} / \rho_h^2$, $S'.w(h) = \bar{M}$ and also we have the value intervals :

$$M'.w(e) \leqq r_{he} \leqq S'w(h)$$

and $\quad r_{he} \leqq \bar{M} \leqq 1.$

Proof : The results readily follow from corollaries (3), (7) and (11).

A Generalised Formulation by a Convex Combination

If R is not a symmetric circulant matrix (implying the existence of distinct optimal and equity formulations), then for a preassigned value of δ ; $0 \leqq \delta \leqq 1$, a generalised formulation can be generated by a convex combination (designated by the symbol c) of the optimal and the equity formulations. The estimates for its aggregating weight vector and the aggregate representation are deduced in the following theorem.

Theorem 9 : Given the distinct optimal and equity formulations with solution weight vectors $w^*(h)$ and $w^*(e)$, the solution weight vector $w^*(c)$ for a convex combination given by

$$m(c) = \delta M + (1 - \delta) S ; \quad 0 \leqq \delta \leqq 1,$$

is obtained from the following relations :

$$\frac{1}{\rho_c^2} = \frac{\delta^2}{\rho_h^2} + \frac{(1-\delta)^2}{\rho_e^2} + \frac{2 \delta (1 - \delta) \bar{M}}{\rho_h^2}$$

$$= \frac{\delta^2}{\rho_h^2} + \frac{(1-\delta)^2}{\rho_e^2} + \frac{2 \delta (1 - \delta) r_{he}}{\rho_h \rho_e},$$

and

$$w(c) = \rho_e^2 \left[\frac{\delta\, w(h)}{\rho_h^2} + \frac{(1 - \delta)\, w(e)}{\rho_e^2} \right].$$

Proof : We have

$$R.w(h) / \rho_h^2 = M \; ; \quad M'w(h) = 1,$$

$$R.w(e) / \rho_e^2 = S \; ; \quad S'w(e) = 1,$$

and also $R\,w(c) / \rho_c^2 = m(c) \; ; \; m'(c) = 1.$

From theorem 2, we have

$$\begin{aligned} 1 / \rho_c^2 &= m'(c).R^{-1}.m(c) \\ &= [\, \delta M' + (1 - \delta) S' \,] [\, \delta R^{-1} M + (1 - \delta) R^{-1} S \,] \\ &= [\, \delta M' + (1 - \delta) S' \,] [\, \delta.w(h) / \rho_h^2 + (1 - \delta).w(e) / \rho_e^2 \,] \\ &= \delta^2 M'.w(h)/\rho_h^2 + (1 - \delta)^2 S'.w(e) / \rho_e^2 \\ &\quad + \delta (1 - \delta) [\, S'.w(h) / \rho_h^2 + M'w(e) / \rho_e^2 \,] \end{aligned}$$

i.e., $1 / \rho_c^2 = \delta^2 / \rho_h^2 + (1 - \delta)^2 / \rho_e^2 + 2\,\delta (1 - \delta) \bar{M} / \rho_h^2$,

(using cor. 12), .. (36a)

or $1 / \rho_c^2 = \delta^2 / \rho_h^2 + (1 - \delta)^2 / \rho_e^2 + 2\,\delta\, (1 - \delta\; r_{he} / (\rho_h \rho_e)$,

(using cor. 11) .. (36b)

Again $w(c) / \rho_c^2 = R^{-1} m(c) = \delta R^{-1} M + (1 - \delta) R^{-1} S$

or $w(c) = \rho_c^2 [\, \delta.w(h)/\rho_h^2 + (1 - \delta)\, w(e) / \rho_e^2 \,].$ ...(37)

Cor. 13 : The aggregate representation ρ_c for the combination lies in between ρ_e and ρ_h.

Proof : By definition of optimal formulation,

$\rho_c \leqq \rho_h$ and also $\rho_e \leqq \rho_h$

Again since, $1 / \rho_h^2 \leqq 1 / \rho_e^2$ and $\bar{M} \leqq 1$, it follows from theorem 9 that

$$1 / \rho_c^2 \leqq \delta^2 / \rho_e^2 + (1 - \delta)^2 / \rho_e^2 + 2\delta (1 - \delta) / \rho_e^2$$

i.e., $1 / \rho_e^2 \leqq 1 / \rho_e^2$

Hence $\rho_e^2 \leqq \rho_c^2$, i.e., $\rho_e \leqq \rho_c$

Thus, we have $\rho_e \leqq \rho_c \leqq \rho_h$ .. (38)

Note : If ρ_c is preassigned in the interval (ρ_e, ρ_h), the estimate δ for the convex combination can be deduced from the relation (36a) or (36b).

Cor. 14 : The correlation coefficients between a convex formulation and any of the optimal and the equity formulations are given by

$$r_{hc} = \rho_c [\delta / \rho_h + (1 - \delta) r_{he} / \rho_e], \quad \text{and}$$

$$r_{ec} = \rho_c [\delta r_{he} / \rho_h + (1 - \delta) \rho_e .$$

Proof : Using theorems (4) and (9), we have

$$r_{hc} = \rho_h M'.w(c) / \rho_c$$

$$= \rho_h \rho_c [\delta M'.w(h) / \rho_h^2 + (1 - \delta) M'w(e) / \rho_e^2]$$

$$= \rho_h \rho_c [\delta / \rho_h^2 + (1 - \delta) \bar{M} / \rho_h^2], \text{ by Cor. 12,}$$

$$= \rho_c [\delta / \rho_h + (1 - \delta) M / \rho_h]$$

i.e. $$r_{hc} = \rho_c [\delta / \rho_h + (1 - \delta) r_{he} / \rho_e], \text{ by Cor. 11,} \quad \text{.. (39)}$$

Again we can deduce

$$r_{ec} = \rho_e S'w(c) / \rho_c$$

$$= \rho_e \rho_c [\delta S'w(h) / \rho^2_h + (1 - \delta) / S'w(e) / \rho_e^2]$$

$$= \rho_e \rho_c [\delta.\bar{M} / \rho^2_h + (1 - \delta) / \rho_e^2], \text{ by Cor. 12,}$$

i.e., $$r_{ec} = \rho_c [\delta r_{he} / \rho_h + (1 - \delta) / \rho_e], \text{ by Cor. 11} \quad \text{....(40)}$$

Note : If r_{hc} is preassigned, then δ and ρ_c can be determined by use of relations (39) and (36) and hence the particular convex formulalation can be ascertained. The value of r_{hc} should obviously be very near to unity, if the particular convex formulation is to be chosen very close to the optimal formulation. The closeress of this covex formulation to the equity formulation can be estimated from relation (40). So the different convex formulations can be selected by the degree of closeness to the optimal formulation r_{hc} (or alternatively by r_{ec}).

Cor. 15 : The convex formulation which is equally related to both equity and optimal formulations is given by :

$$\delta = \rho_h / (\rho_h + \rho_e)$$

Proof : For such convex formulation, we have

$$r_{hc} = r_{ec},$$

Now using relations (39) and (40) and equating, we get

$$\delta = \rho_h / (\rho_h + \rho_e), \quad \text{.. (41)}$$

Another Generalised Formulation by a Power Transformation

The representativeness divergence as present in the optimal index can be reduced in a generalised formulation by a power transformation on the multiplier vector M. Such a generalised formulation will be designated with symbol d. If the optimal formulation is distinct from the equity formulation, the values of M_i's are above and below unity. If we have a power function of M_i's with a fractional power λ, $0 \leqq \lambda \leqq 1$, then the relative importance depicted by multiplier vector m (d), with

$$m_j(d) \propto M_j^{\lambda} \; ; \quad j = 1, 2, \quad n,$$

shows reduced divergence of the ralative importance of constituent variables. Then from the condition

$$m'(d).m(d) = n,$$

we can write

$$m_j(d) = m_j^{\lambda} \Big/ \sqrt{\sum_i M_i^{2\lambda} / n}$$

Clearly, for different choices of λ, we can generate different multiplier vectors m(d) and thus generate different representativeness divergence reducing generalised formulations. However we propose to determine λ from a preassigned value of r_{hd}, the correlation coefficient between the optimal and the proposed formulations. The relation between λ and r_{hd} is deduced in the following theorem, so that when one of them is preassigned, the other can be estimated.

Theorem 10 : The muliplier vector m(d) of a power transformed formulation having been defined as :

$$m_j(d) = M_j^{\lambda} \Big/ \sqrt{\sum_i M_i^{2\lambda} / n}$$

$$j = 1, 2, \ldots n, \text{ and } 0 \leq \lambda \leq 1,$$

the correlation coefficient r_{hd} between this and the optimal formulations is given by :

$$r_{hd}^2 = \frac{\det R \left\{ \sum_j M_j^{\lambda} \; r_j(h) \right\}^2}{n^2 \rho_h^4 \left\{ \sum_{i,j} M_i^{\lambda} \; M_j^{\lambda} \; R_{ij} \right\}}$$

Proof : By theorem 4, we get

$$r_{hd} = r'(d)\ w(h) \ / \ \rho_h = \rho_d\ m'(d).r(h) \ / \ (n\ \rho_h^2)$$

$$\text{or, } n\ \rho_h^2\ r_{hd} = \rho_d \{ m'(d)\ r(h) \}$$

$$\text{or, } n^2\ \rho^4_h\ r^2_{hd} = \rho^2_d \{ m'(d).r(h) \}^2$$

$$= \frac{\{ m'(d).r(h) \}^2}{m'(d).R^{-1}.m(d)},$$

by use of relation (19)

$$= \frac{\det R \left\{ \sum_j M_j^{\lambda} .r_j(h) \right\}^2}{\sum_{i,j} M_i^{\lambda}\ M_j^{\lambda}\ R_{ij}},$$

by use of the defining relation of m(d)

Hence $$r^2_{hd} = \frac{\det R \left\{ \sum_j M_j^{\lambda} . r_j(h) \right\}^2}{n^2\ \rho^4_h \left\{ \sum_{i,j} M_i^{\lambda} M_j^{\lambda}\ R_{ij} \right\}}. \quad .. (42)$$

Note that when r_{hd} is given, λ is determinable from the relation (42) and vice versa. If r_{hd} is preassigned, λ can be solved by use of this relation by iterative numerical methods.

Concluding Remarks

We are now in a position not only to generate various new types of generalised formulations, but also to recast any old types of formulations, derived from the same set of constituent variables, whether differently transformed or untransformed, into the common generalisation framework as designed here. As a result, we have now the tools for making comparisons on the relative merites between any two formulations in terms of the statistical parameters involved. The illustrative examples as worked out in this connection are not presented here for the sake of brevity. This paper is concluded with the claim that the approach offered in handling any generalised formulation through the extended correlation matrix R* is quite elegant—so much so that even the estimating relation (31) for the special case of optimal formulation could be derived much more readily in this, compared with what followed by Kendall [1939] himself.

References :

1. Aitken, A. C. (1942) : "Determinants and Matrices" *Oliver and Boyd*, London.

2 Hotelling, H. (1933) : "Analysis of a Complex of Statistical Variables into Principal Components". *Journal of Educational Psychology*, 24, pp. 417-420 and 498-520.

3. Johnston, J. (1972) : "Econometric Methods" Second Edition, *Mc-Grew-Hill*, New York.

4. Kendall, M. G. (1939) : "The Geographical Distribution of the Crop Productivity in England". *Journal of Royal Statistical Society*, 102, pp. 21-48.

5. Pal, M N. (1963a) : "Zur Berechnung eines Kombinierten Konzentration-sindexes" *Raumforschung und Raumordnung, 21, pp. 87-93.*

6. Pal, M. N. (1963b) : "A Method of Regional Analysis of Economic Development with Special Reference to South India". *Journal of Regional Science*, 5, pp. 41-58.

7. Pal M. N. (1971) : "Quantitative Techniques for Regional Planning". *Indian Journal of Regional Science*, 3, pp. 1-33.

8. Pal, M. N. and R. N. Chattopadhyay (1972, 1973) : "Some Comparative Studies on the Construction of Composite Regional Indices". *Indian Journal of Regional Science*, 4, pp. 132-142 and 5, pp. 101-102.

9. Pal, M. N. (1973) : "Regional Studies and Research for Cosistent and Optimal Plan Formulations" : The Need for a Right Kind of Orientation. *Indian Journal of Regional Science*, 5, pp. 1-23.

10. Pal, M. N. (1974) : "Regional Information, Regional Statistics and Regional Planning in India", in *Regional Information and Regional Planning*, edited by A. Kuklinski, *United Nations Research Institute for Social Development*, Publication Series on Regional Planning, Vol. 6, *Mouton* The Hague, pp. 291-385.

11. Pal, M. N. (1975) : "Regional Disparities in the Level of Development in India", *Indian Journal of Regional Science*, 7, pp. 35-52 and 195.