Cowles Commission Discussion Papers: Statistics, No. 305

November 28, 1947

A Biometric Multidimensional Model

by

George Rasch

I. <u>Physiological Background.</u> In a Danish thesis, H. Vogelius: Basal Metabolism of Girls, Copenhagen, 1945, a tentative treatment is given of a problem which may be taken as in some respects typical for a class of physiological situations.

On each individual in a certain population (e.g. girls 6-15 years old from a small town in Denmark) a set of physiological measurements are made (e.g. basal metabolism, surface area, weight and height) and we know the age of each individual.

Considering these data only, the following simple model would seem adequate

$$(1) \qquad y' = \Pi z' + v'$$

where y is the vector of the measurements while z is the age and v the vector of residuals which may be accepted as normally distributed, although not independent.

In other cases other variables such as social status may play an important role. Therefore, generally speaking, z should be regarded as a vector.

Now two quite different situations may occur: 1.) z is observed or 2.) z is not observed, possibly even not observable. Situation 1.) presents itself in the metabolism data mentioned where the age was recorded. Situation 2.) would more often occur in large population investigations where, in any fixed age group, the social state has an important bearing upon the weight and height and possibly on the metabolism as well. But the practical difficulties in securing observations on social status may be prohibitive. Now some data available point to the possibility that even though both height and weight are strongly influenced by the social state then the balance between height and weight is independent of this variable. Then, from physiological considerations it would seem reasonable to expect a similar situation with regard to metabolism.

In the data mentioned above the social state is not observed, but due to the relative uniformity in a small-town community the influence presumably is so small that the social state may be regarded as a constant. Therefore the treatment of the data as a usual multidimensional regression on the age may be expected to be adequate.

In the literature, however, observations are reported from large communities or from social strata definitely different from that of the Danish small town, and furthermore the age is not even registered for each person. Only the age range for the population is stated.

Still worse situations arise, however, easily: Neither age nor social state -- or whichever the relevant "exogenous" variables may be -- are observed in any of the populations examined, but we just know that such variables would be important.

In the most obscure cases not even the number of such relevant variables is known -- just as in Thurstone's factor theory. Thus in the worst case the model (1) may be interpreted in this way: We accept as a working hypothesis that y may be represented by the right hand term of (1) where z is an unobserved vector of order K, $K(\leq G)$ in some situations being known and denoted K, but in other situations not known, and denoted $K = \rho$; Π is a coefficient matrix of rank $K = \rho$ and v is a vector assumed to be normally distributed, but the elements are not supposed to be independent.

The factor specification therefore would seem adequate for our purposes were it not for Thurstone's assumption of the independence of the residuals. This condition seems hard to accept in physiology as in cases of observed "exogenous" variables the residuals were found to be not independent even though they were fairly small. By dropping the independence, however, we get a modification of the factor specification which seems workable.

2. A Modified Factor Specification. The formulation of this specification runs as follows: We accept as a working hypothesis (primarily not subject to test) that the observed vector

$$(2) \qquad y = (y_1, \ldots y_G)$$

may be represented linearly by some unknown factors

$$(3) \qquad \zeta = (\zeta_1, \ldots, \zeta_\rho)$$

and residuals

$$(4) \qquad v = (v_1, \ldots, v_G)$$

i.e.

$$(5) \qquad y' = \Pi \zeta' + v'$$

In the first development of the theory v shall be assumed to be normally distributed but the elements are not supposed to be independent. Π is a coefficient matrix and of course a parameter in the specification. As to ζ, it stands for unobservable quantities which may be estimated -- and in practice will have to be estimated -- from the data. Therefore, we shall also take the elements of ζ to be parameters in the problem. Hence the change of notation from z to ζ.

The problem of identifiability may be split into two problems:
1) The identifiability of Π, ζ and the variance matrix Φ of v for given ρ, and
2) the identifiability of ρ.

Re (1): Consider first a given representation of "the systematic part"

(6) $$\eta' = \Pi\zeta'.$$

Obviously, the same η may be represented by an infinity of Π's and ζ's produced by postmultiplication of Π by any non-singular square matrix Ψ which then is absorbed in ζ:

(7) $$\Pi^* = \Pi\Psi \quad , \quad \zeta^{*\prime} = \Psi^{-1}\zeta' .$$

As a pure formality we may therefore normalize, say, Π such that it is uniquely determined as a representative of the class of equivalent Π's. Then ζ is also fixed. And for the given class of Π's and ζ's the residuals are expressible in terms of y, and their variances and covariances are therefore identifiable.

A peculiarity should be mentioned. Apparently the partition of y into a systematic part η and a residual v is somewhat arbitrary as we might nib a bit from any η and let it join the v or the reverse, or, to be more specific, we might nib from ζ and add the corresponding term to v:

(8) $$\zeta^{*\prime} = \zeta' - \varepsilon' \quad , \quad v^{*\prime} = v' + \Pi\varepsilon'$$

thus changing both ζ and the variance matrix of the residuals.

This brings forth another difference between Thurstone's factor situation and the specification suggested here. We have taken ζ to be a parameter while v is a random variable. The ε removed from ζ therefore cannot be taken as a random variable without spoiling the parametric character of ζ^*, but if, on the other hand, ε is chosen in a systematic way, preserving ζ^* as a parameter, then the stochastic character of v^* is violated. In Thurstone's specification where the factors are considered as random variables the argument is a different one: ε now being a random variable the chances are that it is uncorrelated with both v and ζ. Thus we get

(9a) $$E\{v^{*\prime}v^{*}\} = E\{v'v\} + \Pi\, E\{\varepsilon'\varepsilon\}\,\Pi'$$

and

(9b) $$E\{v^{*\prime}\zeta^{*\prime}\} = -\Pi\, E\{\varepsilon'\varepsilon\}.$$

As the residuals are assumed to be independent of the factors only ε's for which

(10) $$\Pi\, E\{\varepsilon'\varepsilon\} = 0$$

could be used and then

(11) $$E\{v^{*\prime}v^{*}\} = E\{v'v\}.$$

Thus in both cases the identifiability of the moment matrix of the residuals seems safe against small changes in ζ.

Re 2): The identifiability of ρ presents a real difficulty. For Thurstone's factor model the problem is discussed by Reiersøl. For the present model it might formally be maintained that ρ should be taken just so large that the residuals become random variables. This affords, however, hardly any criterion for ρ unless the possible departure from "random variation" is specified. This may be done in particular cases where the order of the observations is relevant or where, say, a normal distribution of the residuals is expected, while the series of ζ's would be far from normally distributed (that they are in principle just parameters and not random variables does not prevent a large number of them from looking like a normal distribution, of course). In such cases we may proceed from $\rho = 1$ to higher values and stop when the v's become normally distributed.

In many practical cases this may not bring about a sufficiently useful result and therefore another approach has been considered. In this approach we accept for convenience the normal distribution of v but we do not insist upon it. And neither is the question whether the ζ's turn out to form a recognizable distribution or not important. These sacrifices imply, however, that theoretically ρ loses its identifiability completely.

If an element of ζ is normally distributed it may be removed from ζ and included in v and v still has the properties it should have.

In practice, however, ρ may be determined just as unambiguously as the degree of the polynomial in the problem of fitting a curvilinear regression to observed data. In this situation the analysis is carried on until no significant -- or, more crudely, no material -- reduction of the variance of the residuals is found. Similarly it seems possible to start from $\rho = 1$ and proceed picking factors up until no significant -- or material -- reduction of the variances is found. As we are here concerned with variance matrices the problem is more intricate than in the curvilinear regression problem referred to, but a closer study of the case in question presumably leads to some definite criterion.

3. Transformation to confluence. Considering now ρ, Π and Φ as given, transformations of the equations may disclose important features of the actual model. Attempts may go in two quite different directions. We may in accordance with Thurstone try to transform Π and ζ such that a particularly informative structure of Π is obtained, the "simple structure" which is discussed by Reiersøl. The other one links up with the analysis carried out in Vogelius' paper. In this case the factors are eliminated, the equation (5) thus being replaced by linear relations between the observed variables and some new residuals, confluent relations in Frisch's sense.

The main reason for this reformulation of the relations is a practical one. What is wanted in the case in question is not a construction of ingeniously defined statistical factors which in some future possibly will prove to be of great scientific value, the physiological interpretation of which is however at present obscure. The medical men needed a ready-made rule of thumb such that they on observation of a possibly normal, possibly pathological individual may easily decide whether the basal metabolism falls inside the "normal limits", i.e. they want to test the hypothesis that a new observation belongs to the same population as the observed one. They want to do this as easily and as clear cut as possible -- and any physiologically unspecified factor certainly would make them very suspicious. Relations between the observed quantities themselves are more appealing, to be sure.

Let Θ be a matrix of order (L, G) and of rank L $(= G - \rho)$ satisfying the equation

(12) $$\Theta \Pi = 0 .$$

Then we have

(13) $$\Theta y' = w'$$

where

(14) $$w' = \Theta v'$$

has the moment matrix

(15) $$\Omega = \Theta \Phi \Theta' \quad , \quad \Phi = E\{ v'v \} .$$

Since a Θ may be premultiplied by any nonsingular square matrix Θ may have been chosen such that

(16) $$\Omega = I .$$

For simplicity we shall assume this has been done.

The L residuals w are now independent and of unit variance, and so far a rule might be given: Form for an observed individual the residual for each equation, compute their sum of squares, apply the χ^2-test with $\mathcal{L}$ degrees of freedom (this test would even be invariant against the orthagonal transformation of Θ which would preserve (16)).

For one thing this suggestion would be met with a claim for a simpler procedure and for another thing a point of great practical importance may be made: The system of equations formed expresses the normal balance between the said variables (in the example discussed basal metabolism, surface area, weight and height (age is eliminated)), but lack of balance of the total system may be due to lack of balance between weight and height only, say. And if that is a possibility the doctor's action probably would be quite different from what it would be if he was sure that the deficiency was not due to, say, malnutrition, but that it was really a question of glandular disturbance.

Therefore, a closer analysis of the equation system is of great practical importance, the object being to isolate, if possible, the different groups of variables from each other. Thus we have to split the confluence (13) into subconfluences each of which contains fewer variables than the total system. This of course may be carried out in numerous ways, but in so doing we should notice that if we got two equations like these, for instance:

$$(17) \qquad \begin{aligned} \theta_{11} y_1 + \theta_{12} y_2 \qquad\qquad &= w_1 \\ \theta_{22} y_2 + \theta_{23} y_3 &= w_2 \end{aligned}$$

then the practical handling of them will be severely hampered if w_1 and w_2 are not independent. In a new observed case with w_2 within the normal limits the judgment about w_1 -- i.e. the balance between y_1 (which may be the really important variable, in the case quoted the standard metabolism) and y_2 --

depends on just where within the normal limits w_2 fell. Therefore, in a partition of (13) in subconfluences we shall require that the residuals of the separate groups of equations are mutually independent. The independence of the residuals within each group is unimportant; that may be secured by "local" transformations.

4. Decomposition of confluences in subconfluences.

Let

$$\Theta y' = w' \quad , \quad E\{w' w\} = \Omega , \tag{18}$$

where Ω may be a diagonal matrix; we shall take it as the unit matrix. The first problem to be discussed here is whether y may be split into two parts

$$y = (y_I , y_{II}) \tag{19}$$

and the confluence relation accordingly be broken into two sets of equations, one of which contains y_{II} only:

$$\Theta^*_{1I} y'_I + \Theta^*_{1II} y'_{II} = w_1^{*\prime} \tag{20a}$$

$$\Theta^*_{2II} y'_{II} = w_2^{*\prime} \tag{20b}$$

and in such a way that w_1 and w_2 are independent. A particularly interesting point is of course the possibility of

$$\Theta^*_{1II} = 0 . \tag{21}$$

The way in which (20) is produced from (18), to which it is supposed to be equivalent, is by premultiplication by an invertible matrix

$$\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_{2} \end{pmatrix} . \tag{22}$$

For the purpose of discussing the possibilities Θ is broken up in

$$\Theta = (\Theta_I , \Theta_{II}) . \tag{23}$$

multiplying (18) by Ψ we thus obtain

$$\begin{pmatrix} \Psi_1 \Theta_I , & \Psi_1 \Theta_{II} \\ \Psi_2 \Theta_I , & \Psi_2 \Theta_{II} \end{pmatrix} \begin{pmatrix} y'_I \\ y''_{II} \end{pmatrix} = \begin{pmatrix} w_1^{*\prime} \\ w_2^{*\prime} \end{pmatrix} . \tag{24}$$

The problem is now whether it is possible to find such matrices Ψ_1 and Ψ_2 that at the same time

(25) $$\Psi_2 \, \Theta_I = 0$$

and the moment matrix of $w_1^* w_2^*$ is unit which would mean that Ψ had to be orthogonal. As to the last point it is, however, not necessary to require more than independence of w_1^* and w_2^* in the first instance. If that operation has proved possible we may afterwards normalize each of the equations (20) separately. Therefore we shall only require

(26) $$\Psi_1 \, \Psi_2' = 0 .$$

Consider first the solubility of (25). Θ_I is a matrix of order (L, G_I) and of a certain rank R_I, which is $\leqq$ both L and G_I. An equation of the form

(27) $$x \, \Theta_I = 0.$$

thus means G_I equations with L unknowns, but as the rank is R_I it is equivalent to only R_I equations with L unknowns. Thus the equation (27) has $L - R_I$ fundamental vector solutions. Necessary for the existence of Ψ_2 is therefore that the number of row vectors in Ψ_2 does not exceed $L - R_I$, i.e.

$$L_{II} \leqq L - R_I = L_I + L_{II} - R_I$$

or

$$R_I \leqq L_I$$

The inequality should rather be read the other way around:

(28) $$L_I \geqq R_I ,$$

meaning that when trying to separate y in y_I and y_{II} we may state the rank of Θ_I and then we have to choose L_I according to (28) if that is possible. If not, y cannot be split as anticipated. If, however, $L > R_I$ there are a series of possibilities:

(28a) $$L - R_I , \; R_I - L , \; L$$

G_I comes in rather indirectly through the obvious inequality

(29) $$R_I \leqq \begin{cases} G_I \\ L \end{cases} .$$

Now take for L_I any of the values (28a) so far possible. The general solution of (27) may be written

(30) $$x = c \, X$$

where c is an arbitrary vector of order $(1, L - R_I) = (1, L_{II}')$ and X is the matrix of order(L_2, L_2') formed of the $L_2'(\geq L_2)$ fundamental vectors. Ψ_2 then is of the form

(31) $$\Psi_2 = CX$$

where C is a matrix of order(L_2, L_2') and rank $L_2 (\leq L_2')$. The remaining question is whether it is possible to choose Ψ_1 such that (26) is fulfilled. As Ψ_1 should be of rank L_1 this means: whether the equation

(32) $$x\Psi_2' = 0$$

for a chosen C has at least L_1 independent solutions. As Ψ_2^* is of order (L, L_2) and of rank L_2 the equation (32) has just $L - L_2 = L_1$ fundamental solutions. Therefore a Ψ_1 of rank L_1 satisfying (26) does exist for any L_1 among the values (28a).

This, however, is not quite sufficient. It seems feasible that e.g. some vector in Ψ_1 were linearly dependent on the vectors in Ψ_2 or more general that Ψ were singular, i.e. a solution of the equation

$$c\,\Psi = (c_1, c_2)\begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix}$$

exists. The equation may be written

$$c_1\Psi_1 + c_2\Psi_2 = 0$$

and by multiplication by Ψ_2' and application of (26) it follows that we should have

$$c_2\,\Psi_2\Psi_2' = 0$$

but as $\Psi_2\Psi_2'$ is a square matrix of rank L_2 and order L_2 it is invertible such that the only solution is

$$c_2 = 0,$$

and consequently

$$c_1\Psi_1 = 0,$$

i.e. the vectors in Ψ_1 would not be independent in contradiction to the assumption.

The conclusion is that <u>if $L > R_I$ the confluence (18) may be split into two subconfluences with any L_1 in the interval</u>

$$R_I \leq L_1 \leq L - 1.$$

For $L_1 > R_I$ this is possible in several ways but if no other reasons are against it, it seems natural to require L_2 as large as possible, i.e. $L_1 = R_I$.

As to the possibility (21) of splitting the confluence into two non-overlapping subconfluences it is obvious that first it should be possible to split into two subconfluences (20) with some Θ^*_{1II}. Whether this is possible or not we have decided above.

Such as the equation (20) has been derived it may, however, happen that Θ_{1I} is of rank $< \ell_1$; if so y_I could be eliminated from one of the equations (20a) and we were left with $\ell_1 - 1$ equations containing y_I and one equation which could be included in (20b). Continuing this process we obviously may reduce the original system (20) to a system in which the rank of the coefficient matrix of y_I' just equals the number of rows.

Supposing now Θ^*_{1I} of rank ℓ_1 the question is whether the system (20) could be premultiplied by a square matrix of order ℓ

(33) $$\Psi^* = \begin{pmatrix} \Psi^*_{11}, & \Psi^*_{12} \\ \Psi^*_{21}, & \Psi^*_{22} \end{pmatrix}$$

reducing (20) to a system of the same type in which the new Θ^{**}_{1II} vanishes. The coefficient matrix in the new system is

(34) $$\Theta^{**} = \begin{pmatrix} \Psi^*_{11} & \Psi^*_{12} \\ \Psi^*_{21} & \Psi^*_{22} \end{pmatrix} \begin{pmatrix} \Theta^*_{1I} & \Theta^*_{1II} \\ 0 & \Theta^*_{2II} \end{pmatrix} = \begin{pmatrix} \Psi^*_{11}\Theta^*_{1I}, & \Psi^*_{11}\Theta^*_{1II} + \Psi^*_{12}\Theta^*_{2II} \\ \Psi^*_{21}\Theta^*_{1I}, & \Psi^*_{21}\Theta^*_{1II} + \Psi^*_{22}\Theta^*_{2II} \end{pmatrix}$$

As we want to retain the zero already present we must have

(35) $$\Psi^*_{21}\Theta^*_{1I} = 0$$

Since Θ^*_{1I} is of rank ℓ_1 the square matrix $\Theta^*_{1I}\Theta^{*\prime}_{1I}$ of order ℓ_1 is nonsingular; it follows that

(36) $$\Psi^*_{21} = 0$$

Furthermore Ψ^*_{11} of order (ℓ_1, ℓ_1) must be nonsingular as otherwise Ψ^* itself is singular. We may as well put $\Psi^*_{11} = I_{\ell_1}$ similarly we may take $\Psi^*_{22} = I_{\ell_2}$, the transformed coefficient thus becoming

(37) $$\Theta^{**} = \begin{pmatrix} \Theta^*_{1I} & , & \Theta^*_{1II} + \Psi^*_{12}\Theta^*_{2II} \\ 0 & , & \Theta^*_{2II} \end{pmatrix}.$$

What we need is then a solution of the equation

(38) $$\Psi^*_{12}\Theta^*_{2II} = -\Theta^*_{1II},$$

i.e. all vectors in Θ^*_{1II} should be linear combinations of the vectors in Θ^{**}_{2II}, the condition for which is

(39) $$\rho\begin{pmatrix} \Theta^*_{1II} \\ \Theta^*_{2II} \end{pmatrix} = \rho(\Theta^*_{2II}).$$

We require, however, still more, viz, that the transformed residuals are independent. According to our derivation of (20) we have secured w_1^* and w_2^* to be independent; thus the variance matrix of (w_1^*, w_2^*) may be partitioned into

(40) $$\Omega^* = \begin{pmatrix} \Omega_{11}^* & 0 \\ 0 & \Omega_{22}^* \end{pmatrix}$$

where Ω_{11}^* and Ω_{22}^* are nonsingular square matrices of order L_1 and L_2, respectively. The transformation performed consists in multiplying (20) by

(41) $$\Psi^* = \begin{pmatrix} 1 & \Psi_{12}^* \\ 0 & 1 \end{pmatrix}.$$

Thus (21) is transformed into

(42) $$\begin{pmatrix} 1 & \Psi_{12}^* \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Omega_{11}^* & 0 \\ 0 & \Omega_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \Psi_{12}^{*\prime} & 1 \end{pmatrix} = \begin{pmatrix} \Omega_{11}^* + \Psi_{12}^*\Omega_{22}^*\Psi_{12}^{*\prime} & , \Psi_{12}^*\Omega_{22}^* \\ \Omega_{22}^*\Psi_{12}^{*\prime} & , \Omega_{22}^* \end{pmatrix}$$

which leads to the condition

(43) $$\Psi_{12}^*\Omega_{22}^* = 0$$

or, since Ω_{22}^* is invertible,

(44) $$\Psi_{12}^* = 0$$

and hence

(45) $$\Theta_{1II}^* = 0.$$

Therefore: <u>if a decomposition of the confluence (18) into two non-overlapping subconfluences</u>

(46) $$\begin{cases} \Theta_{1I}^* \, y_I' = w_1^{*\prime} \\ \Theta_{2II}^* \, y_{II}' = w_2^{*\prime} \end{cases}$$

<u>is possible with independent residuals then it will show by derivation of the subconfluences (20) with Θ_{1I}^* having its rank equal to the number of rows.</u>

We may proceed to the next problem, viz. decomposition of (18) into two partly overlapping subconfluences:

(47) $$\begin{cases} \Theta_{1I}^* y_I' + \Theta_{1II}^* y_{II}'' & = w_1^{*\prime} \\ \qquad\qquad \Theta_{2II}^* y_{II}' + \Theta_{2III}^* y_{III}' & = w_2^{*\prime} \end{cases}$$

The argument is similar to the preceding one: We start by deriving equations of the form (20) and on splitting y_{II} into two vectors which we may write (y_{II}, y_{III})

we get the form

$$(48)\quad \begin{cases} \Theta^{*}_{1\,I}\, y'_{I} + \Theta^{*}_{1\,II}\, y'_{II} + \Theta^{*}_{1\,III}\, y'_{III} = W^{*}_{1} \\ \qquad\qquad \Theta^{*}_{2\,II}\, y'_{II} + \Theta^{*}_{2\,III}\, y'_{III} = W^{*}_{2} \end{cases}$$

And by a repetition of our argument we obtain the condition

$$(49)\qquad \Theta^{*}_{1\,III} = 0 .$$

Therefore the conclusion is analogous to that above: <u>If a decomposition of (18) into two partly overlapping subconfluences is possible then it will show when the form (20) has been derived.</u>