GUTTMAN'S IMAGE ANALYSIS

C.P. Parel[1]

It is well known that given a set of n observed quantitative variables ($n \geqslant 3$), the multiple regression of each variable on all the remaining (n—1) variables can be directly considered. If $r_{j(n\text{-}1)}$ is the multiple correlation of x_j with the remaining (n-1) variables, then $r^2_{j(n\text{-}1)}$ is that portion of the total variation of x_j explained or accounted for by the (n-1) other variables and $r^2_{j(n\text{-}1)}$ is called the "coefficient of determination". In other words, $r^2_{j(n\text{-}1)}$ gives how much x_j has in "common" with the other (n-1) variables. If $r^2_{j(n\text{-}1)} = 0$ then it is said that x_j has nothing in common with them; if $r^2_{j(n\text{-}1)} = 1$, then x_j is linearly dependent on them.

To further examine this interrelationship, consider the regression equation

$$x_j = \sum_{i=1}^{n-1} \beta_i x_i + e_j \qquad (1)$$

where $p_j \equiv \sum_{i=1}^{n-1} \beta_i x_i$, the predicted value of x_j; and e_j is the error of prediction. The β_i's are the "weights" for predicting variable x_j from the remaining (n-1) other variables. The prediction and the error of prediction are uncorrelated, and e_j is also uncorrelated with each predictor, x_i, $i = 1, 2, \ldots, n\text{-}1$. In other words,

$$r_{ejxi} = 0; j \neq i$$

In Eq. (1), the variable x_j is split into two parts: (i) p_j, the "common part", or what Guttman calls the *image of the variable* x_j which

[1]Prof. Emeritus, UP Statistical Center

is predictable by multiple linear regression from the other (n-1) variables; and (ii) e_j, the "unique" part, called *anti-image* of x_j, or that part of x_j which cannot be predicted by the other (n-1) variables.

Guttman's images and anti-images of a set of variables are defined with respect to a universe of variables. However, in applications, it is not possible to deal with the whole universe of variables corresponding to a particular domain of characteristics under consideration. Neither can we get observations of such variables from a whole universe of subjects with those characteristics. Thus, it is not possible to determine accurately by multiple correlation techniques the images and anti-images of the variables in a finite set of variables to be analyzed. The images and anti-images determined are just "estimates" of the true images and anti-images of the true images and anti-images of the variables. Guttman calls these *partial images* and *partial anti-images.* In this paper, though, images and anti-images will understood to mean Guttman's partial images and partial anti-images, respectively.

Consider the (n x 1) vector of standardized random variable Z where z_j, the jth element of the vector is $z_j = (x_j - \overline{x})/s_j$ where $\bar{x}$ and s_j are the mean and standard deviation of x_j, respectively and where $x_1, x_2, \ldots x_n$ have a joint multivariate normal distribution. Thus,

$$E(Z) = 0 \text{ ; cov } Z = I_n.$$

Let z_{ip} be the predicted value of z_{ji} from the other (n-1) variables for the ith subject; $i = 1, 2 \ldots, N$, the number of subjects. If b_{jj} is defined to be zero; that is, the "weight" of z_j in predicting itself in the multiple regression is zero, then

$$z_{jip} = \sum_{k=1}^{n} b_{jk} \, z_{ki} \; ;$$

or replacing b_{jk} by w_{jk}, as it is a weight,

$$z_{jip} = \sum_{k=1}^{n} w_{jk} \, z_{ki} \, . \qquad (2)$$

If R is the correlation matrix of the z_j's, $j = 1, 2, \ldots, n$, it can be shown that the partial regression coefficients $\beta_{jk(n-2)}$ are:

$$\beta_{jk(n-2)} = w_{jk} = -\Delta_{jk}/\Delta_{jj} \quad ; (j \neq k) \tag{3}$$

where Δ_{jk} and Δ_{jj} are the cofactors of r_{jk} and r_{jj} in R. It is assumed here that $\Delta_{jj} \neq 0$. In matrix notation,

$$Z_p = W'Z, \tag{4}$$

where $W = [w_{jk}]$ is the weight matrix; Z is the (n x 1) vector of standardized random variables; and Z_p is the (n x 1) vector of predicted values of the corresponding variables in Z based on the (n-1) other variables in Z as predictors. In other words, Z_p is Guttman's vector of (partial) images of the variables in Z.

The error of prediction which Guttman calls the "unpredictable" portion of z_j is $e_j = z_j - z_{jp}$ or using z_{ju} for e_j

$$z_{ju} = z_j - z_{jp}. \tag{5}$$

Or, in matrix notation, $Z_u = Z - Z_p$;

or, $$Z = Z_p + Z_u. \tag{6}$$

Eq. (6) is the fundamental equation in Guttman's image analysis.

Consider the weight matrix W. Since from above, a variable predicted from (n-1) other variables in Z receives a weight zero in the prediction equation involving a full set of n variables, Eq. (3) can be written in matrix form as:

$$W = \begin{bmatrix} 0 & \frac{-\Delta_{12}}{\Delta_{22}} & \frac{-\Delta_{13}}{\Delta_{33}} & \cdots\cdots & \frac{-\Delta_{1n}}{\Delta_{nn}} \\ \frac{-\Delta_{21}}{\Delta_{11}} & 0 & \frac{-\Delta_{23}}{\Delta_{33}} & \cdots\cdots & \frac{-\Delta_{2n}}{\Delta_{nn}} \\ \vdots & \vdots & 0 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{-\Delta_{n1}}{\Delta_{11}} & \frac{-\Delta_{n2}}{\Delta_{22}} & \frac{-\Delta_{n3}}{\Delta_{33}} & \cdots\cdots & 0 \end{bmatrix} \qquad (7)$$

A more convenient form was provided by Kaiser (1963) as:

$$W = I - R^{-1} T^2 \qquad (8)$$

where I is the identity matrix; R^{-1} is the inverse of the correlation matrix R and $T^2 = [\text{diag } R^{-1}]^{-1}$. This can easily be verified. Eq. (8) can also be written as:

$$W = R^{-1}(R - T^2) . \qquad (9)$$

Earlier in the paper, it was pointed out that the error of prediction, or the anti-image of a variable z_j is uncorrelated with each of the (n-1) predictors separately; that is,

$$E\,(z_{jiu} \cdot z_{ki}) = 0; \quad j \neq k \qquad (10)$$

This can easily be shown. Let the variance of the anti-images be

$$\sigma^2_{ju} = \underset{i}{E}\,[z_{jiu}]^2 = \underset{i}{E}\,[z_{ji} - z_{jip}]^2$$
$$= E\,[z_{ji} - \sum_{k=1}^{n} w_{jk}\, z_{ki}]^2$$

Differentiating this with respect to w_{jk} and setting this derivative equal to zero,

$$\frac{\partial \underset{i}{E} [z_{ji} - \sum_{k=1}^{n} w_{jk} z_{ki}]^2}{\partial w_{ki}} = \frac{\underset{i}{E} [\partial(z_{ji} - \sum_{k=1}^{n} w_{ki} z_{ki})]^2}{\partial w_{ki}}$$

$$= \underset{i}{E} [2 (z_{ji} - \sum_{k=1}^{n} w_{ki} z_{ki}) (-z_{ki})] = 0$$

or, $$\underset{i}{E} (z_{ji} - \sum_{k=1}^{n} w_{ki} z_{ki}) z_{ki} = E (z_{jiu} z_{ki}) = 0; \quad j \neq k.$$

However, the covariance of the anti-image of a variable and the variance itself is not zero. This can be easily shown using matrices.

Consider $E\ (Z_u Z') = E\ [Z - Z_p) Z'] = E[\ (Z - W Z) Z'] = (I - W')E(Z\ Z')$. But $E\ (Z\ Z') = R$, the correlation matrix. Thus, $E\ (Z_u Z') = (I - W')R = R - W'R = R - (I - T^2 R^{-1}) R =$

$$T^2 = [\ \text{diag}\, R^{-1}\ ]. \tag{11}$$

This means, therefore, that the covariances of the anti-images of variables with the variables themselves are correspondingly equal to the inverses of the diagonal elements of the inverse of the correlation matrix, R.

What can be said about the covariances of the images of the variables and the variables themselves? To examine this, consider

$$E\,(Z_p\, Z') = \text{E}\,[\,(W'Z)Z'] = W'\,E(Z\,Z') = W'R$$
$$=(I - T^2 R^{-1})\,R.$$

Hence, $E\,(Z_p Z') = R - T^2$. (12)

This means, therefore, that the covariances of the images of the variables with the variables themselves are equal to the diagonal elements of R^{-1}, and the covariances of the images with the remaining (n-1) variables are equal to the correlation coefficients of the variables with the remaining (n-1) variables. That is,

$$\underset{i}{E}\,[z_{jip} z_{ji}] = 1 - \Delta/\Delta_{jj}. \; ; \; j = 1, 2, \ldots, n, \text{ where } \Delta = \text{determinant}$$

of R, and, $\underset{i}{E}\,[z_{jip}\; z_{ki}] = r_{jk} \;, \quad j \neq k.$

The concern in image analysis is to obtain the covariance matrices of the different images and the different anti-images.

Let Λ = the covariance matrix of the images of the variables in Z and let Ω = the covariance matrix on the anti-images of the variables in Z. By definition,

$$\begin{aligned} \Lambda &= E\,(Z_p\, Z'_p) \\ &= E\,[W'ZZ'W] = W'E\,(ZZ')\,W \\ &= W'R\,W = (I - T^2R^{-1})'\,R(I - T^2R^{-1}) \\ &= R - 2T^2 + T^2R^{-1}T^2 \; ; \end{aligned} \tag{13}$$

whereas,

$$\begin{aligned} \Omega &= E\,(Z_u Z'_u) = E[\,(Z - Z_p)\,(Z - Z_p')\,] \\ &= E\,[\,(I - W')\,ZZ'(\,I - W)\,] \\ &= (I - W')\,R\,(I - W) \\ &= [\,I - (I - T^2R^{-1})'\,]\;\; R\,[\,I - (\,I - T^2R^{-1})\,] \end{aligned}$$

or, $$\Omega = T^2R^{-1}T^2 = [\text{diag. } R^{-1}\,]^{-1}R^{-1}\,[\text{diag. } R^{-1}\,]^{-1} \,. \tag{14}$$

It should be noted that the diagonal elements of Ω are the same as the diagonal elements of T^2, since the diagonal elements of $T^2R^{-1}T^2$ are equal to the diagonal of T^2. This means, therefore that the variances of the anti-images are equal to the diagonal elements of T^2. In other words,

$$\underset{i}{E}\,(z_{jiu})^2 = \frac{\Delta}{\Delta_{jj}} \tag{15}$$

From Eqs. (13) and (14), $\Lambda = R - 2T^2 + \Omega$ or, $R = \Lambda - \Omega + 2T^2$ (16)

Let us examine the elements of these matrices. Let $\lambda_{jk} \equiv$ the jk^{th} element of Λ and $\gamma_{jk} \equiv$ the jk^{th} element of Ω. By definition, $\lambda_{jk} = E(z_{jip}\, z_{kip})$ and $\gamma_{jk} = E(z_{jiu}\, z_{kiu})$. According to Eq. (5), $z_{jiu} = z_{ji} - z_{jip}$. Multiplying this equation by z_{ki} and taking expectations:

$$\underset{i}{E}(z_{ki}\, z_{jiu}) = \underset{i}{E}(z_{ki}\, z_{ji}) - \underset{i}{E}(z_{ki}\, z_{jip}).$$

But $\underset{i}{E}(z_{ki}\, z_{jiu}) = 0 \;, j \neq k.$

Therefore, $E(z_{ki}\, z_{jip}) = E(z_{ki}\, z_{ji}) = r_{kj} \;, j \neq k.$

$$\text{Hence, } r_{kj} = r_{jk} = \underset{i}{E}(Z_{ki}\, Z_{jip}) = \underset{i}{E}(Z_{jip}\, Z_{ki}) \qquad (17)$$

But $r_{kj} = \sigma_{kj}/\sqrt{\sigma_{kk}\, \sigma_{jj}} = \sigma_{kj}$ since $\sigma_{kk} = \sigma_{jj} = 1$. Thus, the correlation of coefficient of z_i and z_k, $j \neq k$, is equal to the covariance of z_k and the image of z_j, and vice versa. If Eq. (5) is multiplied by z_{kip}, and expectations are taken,

$$\underset{i}{E}(z_{kip}\, z_{jiu}) = \underset{i}{E}(z_{kip}\, z_{ji}) - \underset{i}{E}(z_{kip}\, z_{jip})$$

$$\text{Hence, } \underset{i}{E}(z_{kip}\, z_{jiu}) = r_{jk} - \lambda_{jk} \qquad (18)$$

Again, multiplying Eq. (5) by z_{ku} and taking expectations:

$$\underset{i}{E}(z_{kiu}\, z_{jiu}) = \underset{i}{E}(z_{kiu} z_{ji}) - \underset{i}{E}(z_{kiu}\, z_{jip})$$

or,

$$\gamma_{jk} = -\underset{i}{E}(z_{kiu}\, z_{jip}), \qquad (19)$$

since $E(z_{kiu} z_{ji}) = 0, k \neq j$. Using Eq. (18), $\gamma_{jk} = -(r_{jk} - \lambda_{jk})$

$$\text{or, } \quad r_{jk} = \lambda_{jk} - \gamma_{jk} \;, j \neq k. \qquad (20)$$

The equation states that the correlation coefficient between any observed variables z_j and z_k with unit variance is equal to the covariance of their respective images diminished by the covariance of their respective anti-images. Thus, in Eq. (16), the off-diagonal elements r_{jk} of R, the correlation matrix, are equal to that given in Eq. (20). It should be noted, therefore, that both the covariances of the images and of the anti-images of any two variables z_j and z_k affect the correlation between the variables. Also from Eq. (16),

$$\lambda_{jj} - \gamma_{jj} + 2\ \Delta/\Delta_{jj} = 1 \qquad (21)$$

In Eq. (19), the covariance between the anti-image of z_k and the image of z_j is the negative of the covariance between the anti-images of z_k and z_j, $j \neq k$. Let us see the form of this covariance matrix of the images and anti-images of the variables in Z. Consider

$$\begin{aligned} E(Z_u Z_p') &= E\,[(I - W')\,Z Z' W] \\ &= (I - W')\,E(Z Z')\,W \\ &= (I - W')\,RW \\ &= RW - W'RW \\ &= RW - \Lambda \\ &= R\,(I - R^{-1}T^2) - (R + \Omega - 2T^2). \end{aligned}$$

Hence,

$$E(Z_u Z_p') = -\Omega + T^2 \qquad (22)$$

But the diagonal elements of Ω are the same as the diagonal elements of T^2. Hence, the diagonal elements of the covariance matrix of the anti-images and images of the variables in Z are all zero. This means that the covariance of the anti-image and the image of the same variable is zero. This is as expected. The off-diagonal elements of the covariance matrix of the anti-images and the images are the negatives of the off-diagonal elements of Ω, the covariance matrix of the anti-images. Let us examine this further.

Let R_u = the correlation matrix of the anti-images of the variables in Z.

$R_u = [\text{diag}\,\Omega]^{-½}\ \Omega\ [\text{diag}\ \Omega]^{-½} = (T^2)^{-½}\ \Omega\ (T^2)^{-½} = T^{-1}\ \Omega\ T^{-1} = T^{-1}\ [\ T^2\ R^{-1}\ T^2\ ]\ T^{-1}$, from Eq.(14). Therefore,

$$R_u = T R^{-1} T. \qquad (23)$$

This means that the jk^{th} element $r_{jk(u)}$ of R_u is $r_{jk(u)} = t_{jj} r^{jk} t_{kk}$, where t_{jj} and t_{kk} are the j^{th} and k^{th} diagonal elements of T and r^{jk} is the jk^{th} element of R^{-1}. Here, the t's are the standard deviations of the anti-images, since $T^2 \equiv [\text{diag}\ \Omega]$. But $t_j = (\Delta/\Delta_{jj})^{1/2}$; $t_k = (\Delta/\Delta_{kk})^{1/2}$ and $r^{jk} = \Delta_{jk}/\Delta$, where as defined earlier, Δ_{jj}, Δ_{kk} and Δ_{jk} are the cofactors of the elements r_{jj}, r_{kk} and r_{jk} of the correlation matrix R and Δ is the determinant of R. Therefore,

$$r_{jk(u)} = (\Delta^{1/2}/\Delta_{jj}^{1/2})\ (\Delta_{jk}/\Delta)\ (\Delta^{1/2}/\Delta_{kk}^{1/2})$$

$$= \Delta_{jk}/(\Delta_{jj}\Delta_{kk})^{1/2}. \tag{24}$$

But the partial correlation coefficient of z_j and z_k with the other $(n\text{-}2)$ variables in Z held fixed is:

$$r_{jk(n\text{-}2)} = \frac{\Delta_{jk}}{(\Delta_{jj}\ \Delta_{kk})^{1/2}};\ j \neq k$$

Hence, $$r_{jk(u)} = -r_{jk(n\text{-}2)}\ . \tag{25}$$

This means, therefore, that the correlation coefficient of the anti-images of z_j and z_k is the negative of the partial correlation coefficient of z_j and z_k with the other $(n\text{-}2)$ variables held constant.

But $$r_{jk(u)} = \gamma_{jk}/\sqrt{\sigma_{jj}\ \sigma_{kk}}$$

or, $$\gamma_{jk} = -r_{jk(u)}\ \sqrt{\sigma_{jj}\ \sigma_{kk}}$$

or, $$\gamma_{jk} = -r_{jk(n\text{-}2)}\ \sqrt{\sigma_{jj}\ \sigma_{kk}}\ ;$$

where σ_{jj} and σ_{kk} are the variances of z_{ju} and z_{ku} respectively. Therefore, from Eq. (20),

$$r_{jk} = \lambda_{jk} + r_{jk(n\text{-}2)} \sqrt{\sigma_{jj}\ \ \sigma_{kk}}\ . \tag{26}$$

In words, the correlation coefficient r_{jk} of the variables z_j *and* z_k can be thought of as consisting of two parts: the covariance between the images of "common parts" (as it is called in common factor analysis) and a "special pairwise linkage" that may remain between the two variables after the remaining (n-2) variables are "partialled out".

How does image analysis relate to common factor analysis? Guttman's image theory, like that of common factor analysis, is based on the assumption that each variable in the set of variables in the domain of the characteristics of interest, can be split into two parts: the "image" of the variable (or, the "common part") and the "anti-image" of the variable (or, the "unique part"). However, common-factor analysis also imposes another restriction: the unique parts of two different variables are uncorrelated. It is the imposition of these two restrictions in factor analysis that give rise to the "indeterminacy" of the common factors.

According to Guttman, image analysis is the "more basic and inclusive approach" in the study of factor structures, and it includes, as a special case, common factor theory. Factor theory is concerned only with that special case wherein the partial correlations tend to zero, but in general, a partial correlation can increase or decrease as the number of variables partialled out increases. But, in general, a multiple correlation increases, and can never decrease, as the number of predictor variables increases. Image theory makes use of this non-decreasing property of the multiple correlation. In view of the fact the multiple correlations tend to approach an asymptotic value quite fast as the number of predictor variables increases. Guttman recommends image analysis as an approximation to factor analysis. According to him, if the number of variables is greater than 15, partial images and anti-images will not differ significantly from their respective images and anti-images. However, the number of observations must also be large for this statement to hold.

REFERENCES

1. Guttman, Louis: "Image Theory for the Structure of Quantitative Variables": *Psychometrika*, Vol. 18, No. 4. Dec. 1953 pp. 277-296.

2. Guttman, Louis: "Multiple Rectilinear Prediction and the Resolution into Components": *Psychometrika*, Vol. 5, 1940; pp. 75-99.

3. Mulaik, Stanley: *The Foundations of Factor Analysis.* Mc-Graw Hill Book Co., New York, 1972.

4. Cattel, Raymond B.: "*The Scientific Use of Factor Analysis in Behavioral and Life Sciences*". Plenum Press, New York, 1978.