THE MINIMUM AVERAGE CORRELATION BETWEEN EQUIVALENT SETS OF UNCORRELATED FACTORS*

PETER H. SCHÖNEMANN

PURDUE UNIVERSITY

A simplified proof of a lemma by Ledermann [1938], which lies at the core of the factor indeterminacy issue, is presented. It leads to a representation of an orthogonal matrix T, relating equivalent factor solutions, which is different from Ledermann's [1938] and Guttman's [1955]. T is used to evaluate bounds on the average correlation between equivalent sets of uncorrelated factors. It is found that the minimum average correlation is independent of the data.

1. Introduction

No other facet of the factor analysis model has generated more confusion than the issue of factor indeterminacy. Briefly, the issue is that the latent variables in the factor model, *i.e.*, the common and the unique factors, are indeterminate as long as none of the uniquenesses vanishes and the number of observed variables is finite. This issue was first raised by Wilson in [1928] and subsequently discussed at length by him and several other authors (e.g., Camp, 1932; Piaggio, 1933; Thomson, 1934, to name a few) in the early thirties. The most comprehensive discussion was given by Guttman (especially 1955, 1956), who also contributed several new results. A more recent and very readable treatment can be found in Heermann [1964, 1966]. In view of the long history of this issue it is surprising that we can find no current text on factor analysis [Holzinger and Harman, 1941; Thurstone, 1947; Cattell, 1952; Fruchter, 1954; Harman, 1960; Lawley and Maxwell, 1963; Harman, 1967; Pawlik, 1968; Überla, 1968; Schönemann, 1969] which even mentions it, let alone discusses it adequately.

A partial explanation for this neglect may, perhaps, be sought in the fact that the early discussions all dealt with the Spearman single general factor case. Ironically, there it is most easily understood because it cannot be confused with the conventional rotation problem. Another reason might be the relative difficulty of some of the early papers. We shall, therefore, give a considerably simplified proof of a basic lemma which was first derived by Ledermann [1938], and later used by Heermann [1966] to rederive certain results of Guttman [1955]. Our proof of this lemma makes use of the Eckart-

^{*} This paper owes much to an unknown reviewer.

Young [1936] decomposition of the total factor pattern (A, U), which allows us to shorten the proof to a few lines, while Ledermann's stretches over several pages. We then go on to sharpen quantitatively certain results by Guttman [1955] and Heermann [1966] to arrive at the conclusion that the minimum average correlation between equivalent sets of uncorrelated factors is independent of the data and decreases for a fixed number of observed variables as the number of common factors is increased, *i.e.*, the factor indeterminacy problem gets worse, rather than better, as one passes from the Spearman case to the Thurstone case.

2. Definitions

Given a set of p observed random variables y_i in $\eta' = (y_1, \dots, y_p)$ which have

(2.1) expected value $\epsilon(\eta) = \phi_p$ and covariance matrix var $(\eta) = \Sigma$, we shall say " η satisfies the model of factor analysis for (A, U) and (ξ', ζ') " if, for some $m , <math>\eta$ can be written

(2.2)
$$\eta = (A, U) \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = A\xi + U\zeta,$$

where A is a $p \times m$ matrix of constants called the "common factor pattern," U is a p.d. diagonal matrix called the "unique factor pattern," and where the x_i in $\xi' = (x_1, \dots, x_m)$ are m unobserved random variables called "(uncorrelated) common factors" and the z_i in $\zeta' = (z_1, \dots, z_p)$ are p unobserved random variables called "(standardized) unique factors" which jointly satisfy

(2.3)
$$\epsilon \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = \phi_{p+m} \quad \text{and} \quad \text{var} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = I_{p+m} .$$

The matrix (A, U) will be called the "total pattern."

In passing, we note that our argument could have been formulated, with but minor changes in notation, for the sample case. Instead of (2.2) we could have written $Y = \hat{A}X + \hat{U}Z$, where $Y(p \times N)$ are the observations of an N-fold random sample on η with the sample means removed, i.e., a $p \times N$ matrix of "observed (deviation) scores." $X(m \times N)$ is a matrix of (standardized and uncorrelated) "common factor scores" and $Z(p \times N)$ a matrix of (standardized and uncorrelated) "unique factor scores." \hat{A} and \hat{U} would then be sample estimates of A and U.

We prefer to treat the population case to emphasize the fact that the factor indeterminacy is first and foremost a problem in the population. The factor score issue is simply a secondary, practical problem which reflects the population indeterminacy into the sample. We, therefore, cannot agree with McDonald and Burr [1967] who seem to think the factor indeterminacy is restricted to only one of the four models they discuss, Model IV, which

essentially corresponds to our sample case. Rather, it also applies to their Model I, which corresponds to our population case. Moreover, in the population there is no need to worry about the existence of $(\xi', \zeta')'$: if η satisfies the model of factor analysis, then at least one set of factors exists by definition. In the sample, on the other hand, one usually starts with Y, \hat{A} , and \hat{U} , and the problem arises to show that X, Z satisfying $Y = \hat{A}X + \hat{U}Z$ exist if $YY'/(N-1) = \hat{A}\hat{A}' + \hat{U}^2$. This was done by Kestelman [1952] for the uncorrelated case and by Guttman [1955] for the more general correlated case.

3. A Simple Proof of a Lemma by Ledermann

By "orthogonal right unit of a matrix B" we shall mean a matrix T which satisfies BT = B and TT' = T'T = I. We now wish to prove:

Ledermann's Lemma: Any matrix B of order $p \times (p + m)$ has an orthogonal right unit which can be expressed as a function of an arbitrary orthogonal matrix S of order $m \times m$.

Proof: Let

$${}_{p}B_{p+m} = {}_{p}V_{p}D_{p}W'_{1 \text{ (p+m)}}$$

be the Eckart-Young (1936) decomposition of B, so that

(3.2)
$$V'V = VV' = W'_1W_1 = I_p$$
, $D = diagonal$.

Consider the matrix T defined as

$$(3.3) T = W_1 W_1' + W_2 S W_2',$$

where W_2 is of order $(p + m) \times m$ and satisfies

$$(3.4) W_1'W_2 = \phi, W_2'W_2 = I_m$$

(i.e., W_2 is an orthonormalized set of basis vectors of the nullspace of W_1) and S is an arbitrary orthogonal matrix of order $m \times m$ which can be chosen at will.

T in (3.3) is clearly orthogonal and it is also a right unit for B, since $BT = (VDW_1')(W_1W_1' + W_2SW_2') = VDW_1' = B$. Since T is a function of an arbitrary orthogonal matrix S of order $m \times m$, the lemma is proved.

So as not to complicate matters unnecessarily at this stage we defer the problem whether all orthogonal right units of B must be of the form (3.3) to sec. 6 (Theorem 3). To appreciate the significance of factor indeterminacy and to understand what follows the lemma in its present form will suffice.

4. The Minimum Average Correlation Between Equivalent Sets of Uncorrelated Factors

The point of the lemma is this: if a given set of random variables η satisfies the model of factor analysis, then we have at least one set of factors

 $(\xi', \, \zeta')'$ which satisfies (2.1)–(2.3). From this set we can get another set of "equivalent factors" $\begin{pmatrix} \xi^* \\ \zeta^* \end{pmatrix} = T' \begin{pmatrix} \xi \\ \zeta \end{pmatrix}$ as in

$$(4.1) \qquad (A, U) \binom{\xi}{\zeta} = (A, U)TT' \binom{\xi}{\zeta} = (A, U)T' \binom{\xi}{\zeta} = (A, U) \binom{\xi^*}{\zeta^*}$$

which also satisfies (2.1)–(2.3) provided T is chosen in accordance with (3.3). Since T is a right unit of (A, U), the equivalent factors satisfy (2.2), as (4.1) shows. And since T is also orthogonal, the equivalent factors $(\xi^*, \zeta^*)'$ also obey the covariance strictures (2.3). For the correlation matrix between both sets of equivalent factors one finds

(4.2)
$$\operatorname{cov}\left\{\begin{pmatrix} \xi \\ \zeta \end{pmatrix}, \begin{pmatrix} \xi^* \\ \zeta^* \end{pmatrix}\right\} = \left[\operatorname{var}\left(\frac{\xi}{\zeta}\right)\right]T = T.$$

In particular, equivalent pairs of factors have correlation $t_{kk}(k=1, p+m)$. We collect these observations in a

Corollary: Suppose a vector of random variables η satisfies the model of factor analysis for (A, U) and a set of uncorrelated factors $(\xi', \zeta')'$. A sufficient condition that η also satisfies the model of factor for another set of uncorrelated factors $(\xi^{*}, \zeta^{*})'$ is that both sets are related by T in (3.3) which is a function of an arbitrary orthogonal matrix S of order $m \times m$, where m is the number of common factors. The correlation matrix between both sets of equivalent, uncorrelated factors is T.

Starting with this observation, but a different representation for T, Heermann [1966] showed that by an appropriate choice of S the sum of the correlations between equivalent factor pairs related by T in (3.3) can be either minimized or maximized. Here we go a bit further and evaluate the minimum and maximum so obtained:

(4.3)
$$\operatorname{tr}(T) = \operatorname{tr} W_1 W_1' + \operatorname{tr} W_2 S W_2' = p + \operatorname{tr} S.$$

Since tr $S = \sum_{i=1}^{m} s_{i,i}$ and $|s_{i,i}| \leq 1.0$ imply

$$(4.4) -m = \operatorname{tr} (-I) \le \operatorname{tr} S \le \operatorname{tr} I = m,$$

one finds

$$(4.5) p-m \le \operatorname{tr} T \le p+m,$$

i.e., one minimizes the sum of correlations if one chooses S = -I so that

$$(4.6) T_{\min} = T(S = -I) = W_1 W_1' - W_2 W_2' = 2W_1 W_1' - I_{p+m} ,$$

and one maximizes this sum if one chooses, perhaps not surprisingly,

$$(4.7) T_{\max} = T(S = I) = I_{p+m}.$$

We are interested in the worst possible case and define as a measure of the average amount of factor indeterminacy the minimum average correlation between equivalent solutions related by T in (3.3), *i.e.*,

(4.8)
$$\tau = \frac{1}{p+m} \min_{T} \sum_{k=1}^{p+m} t_{kk} = \frac{1}{p+m} \operatorname{tr} T_{\min}.$$

In view of (4.5) one finds

(4.9)
$$\tau = (p-m)/(p+m) = (1-m/p)/(1+m/p).$$

We have proved, therefore,

Theorem 1: The minimum average correlation τ between equivalent sets of uncorrelated factors related by T in (3.3) is given by (4.9) and does not depend on (A, U) (or, in the sample, (\hat{A}, \hat{U})).

To illustrate the meaning of this result, assume the ratio m/p is in the vicinity of a third, as it often is in practical work. One finds $\tau = (2/3)/(4/3) = .5$. This means that in this case we can be sure, without looking at the data, that we can find for any given set of factors $\begin{pmatrix} \xi \\ \zeta \end{pmatrix}$ which satisfy (2.1)-(2.3) another set with at least one factor predicting no more than 25 percent of the variance of its equivalent twin.

Equation (4.9) also yields immediately a theoretical result which has been proved by several writers (e.g., Piaggio, 1933; Guttman, 1956): as $\lim_{p\to\infty} (m/p)$ approaches zero, the average minimum correlation approaches unity. Loosely, if the number of variables increases indefinitely, and if it increases faster than m, the number of common factors, then the factor indeterminacy disappears. Since our intention was to raise, rather than diminish, concern about the factor indeterminacy issue, we should add that Camp [1932, p. 425] gives a numerical illustration for the Spearman case where it "would require the measurement of 297 aptitudes," none having more than one common factor, namely g, to push the indeterminacy back into tolerable bounds, i.e., to render g "almost unique."

As one extracts more factors from a given set of observed variables the average indeterminacy gets worse. It cannot hurt, therefore, to reread some of the earlier discussions of this indeterminacy, even if one no longer believes in Spearman's model, which, as it turns out, represents the conservative instance of this issue. Suppose Spearman succeeded in factoring q variables for a single general factor and achieved a minimum average correlation $\tau = \tau^*$. To match this τ , Thurstone, if he wanted to extract m common factors, would have had to factor p variables, so that (p - m)/(p + m) = (q - 1)/(q + 1),

or p=mq. Of course, he also would have had to increase the sample size accordingly in order to estimate his Σ with the same precision as Spearman. We record this result as a

Corollary: If q variables are needed to achieve a given minimum average correlation τ^* for a single general factor solution, then exactly p = mq variables are needed to achieve the same minimum average correlation τ^* for an m common factor solution.

5. Connection with Earlier Results

The T_{\min} obtained by Ledermann [1938] is of the form

$$T_{\min} = I_{n+m} - 2Q_m(Q'_m Q_m)^{-1} Q'_m,$$

where Q_m is a matrix of order $(p+m) \times m$ of full column rank. As Heermann showed in [1966], it is possible to write (5.1) in the form

(5.2)
$$T_{\min} = 2 \binom{A'}{U} \Sigma^{-1}(A, U) - I_{p+m},$$

which is also the representation used by Guttman [1955]. This representation has the advantage that it gives the diagonal elements t_{kk} as explicit functions of the squared multiple correlation coefficients

$$\rho_{x_j,y_1,\dots,y_p}^2 = a_i' \Sigma^{-1} a_i \quad \text{and} \quad \rho_{z_i,y_1,\dots,y_p}^2 = u_i' \Sigma^{-1} u_i$$

(where a_i is the j'th column of the common factor pattern A and u_i the i'th column of the unique factor pattern U). These multiple correlations, estimates of which can be, and probably should be, computed in the sample case, have an immediate geometrical interpretation as the cosines of an angle θ which a given factor x_i (or z_i) spans with the best linear combination of the y_i . All equivalent factors x_i^* lie on a cone generated by rotating x_i (through T) about a fixed axis representing this linear combination. Thus, the worst case for a given factor x_i is obtained upon locating x_i^* on the surface of this cone diametrically opposed to x_i , so that x_i and x_i^* are separated by an angle 2θ . Since $\cos 2\theta = 2\cos^2\theta - 1$, the diagonal elements t_{kk} in T_{\min} are simply the cosines of twice the generating angle of the cone. This basic geometrical interpretation is already implicit in Wilson's first paper [1928] and was made explicit by Thomson in 1935 [p. 251]. For a more recent lucid statement of the geometry of factor indeterminacy see Heermann [1964].

Using the Eckart-Young decomposition (3.1) of (A, U), one can rewrite our T_{\min} in (4.6) as

(5.3)
$$T_{\min} = 2W_1W'_1 - I_{p+m}$$
$$= 2(W_1DV')(VD^{-2}V')(VDW'_1) - I_{p+m}$$
$$= 2\binom{A'}{U}\Sigma^{-1}(A, U) - I_{p+m},$$

where we assume, as Heermann did, that Σ^{-1} exists, which is not necessary, but convenient. The result in (5.3) is the Guttman-Heermann T_{\min} in equation (5.2).

The matrix $\binom{A'}{U}\Sigma^{-1}(A, U) = W_1W_1' = P$ is idempotent and symmetric. It is the orthogonal projector for the row space of (A, U), and it can be obtained from the Moore-inverse of (A, U), which is $(A, U)^+ = \binom{A'}{U}\Sigma^{-1}$.

Any symmetric idempotent matrix which is a deficient rank right unit P of some matrix C can be transformed into an orthogonal (and hence full rank) right unit T = 2P - I, as is easily verified.

Since the trace of an idempotent matrix equals its rank, which is p for the idempotent in (5.2), we could have evaluated tr $T_{\min} = 2p - (p+m) = p - m$, using the Guttman-Heermann representation. Similarly, using Ledermann's representation in (5.1), tr $T_{\min} = p + m - 2m = p - m$. The main advantage of our present representations (3.3), (4.6), we think, is that the associated proofs are shorter.

Finally, since our discussion is restricted to uncorrelated factors, we can state what happens to the minimum average correlation between equivalent sets of uncorrelated common factors when $A(\hat{A})$ is rotated. Let

(5.4)
$$\tau_x = \frac{1}{m} \sum_{i}^{m} t_{ii} = \frac{1}{m} \operatorname{tr} T_{11} ,$$

where T_{11} is the upper left hand partition of order $m \times m$ of T_{\min} in (5.2), i.e., $T_{11} = 2A'\Sigma^{-1}A - I_m$. Since tr $L'T_{11}L = \text{tr } T_{11}$ for any orthogonal matrix L, we have

Theorem 2: The minimum average correlation τ_x between equivalent sets of uncorrelated common factors is invariant under "orthogonal rotation" of the common factor pattern $A(\hat{A})$.

Note, however, that τ_x , in contrast to τ , is not independent of the data.

6. Necessitu

A reviewer reminded us rather persistently that Ledermann's lemma, as used so far, gives only a sufficient condition for the relation between equivalent sets of uncorrelated factors. He also pointed out that "a proof of necessity, as well as sufficiency, for a class of transformations is given in Theorem 2 of Guttman's [1955] paper."

This point is well taken and since our present approach is somewhat different from Guttman's, it remains to be shown that it also can be used to establish the fact that all transformations relating equivalent factors must necessarily be of the form (3.3). We deferred this aspect of the problem so far because much can be learned about the basic problem of factor indeterminacy without considering this somewhat more difficult issue. If T in (3.3)

were only sufficient, but not necessary for relating two equivalent sets of uncorrelated factors, then all the results developed so far would still hold, except that possibly other sets of equivalent factors might exist which are not related by T in (3.3). In this case τ would have been an upper bound to the minimum average correlation. We feel the basic problem of factor indeterminacy has been ignored long enough to make it worthwhile to discuss it in the simplest possible terms. Having done this, we now return to Ledermann's Lemma and the associated Corollary and show that in fact they also state a necessary condition for factor equivalence.

To this end we first note that $(\xi', \zeta')'$ and $(\xi^{*'}, \zeta^{*'})'$ are linearly related because both relate linearly to η in (2.2). It is equally clear that the matrix of the linear mapping T, whether of the form (3.3) or not, must be orthogonal since to assume otherwise $(T'T \neq I)$ would imply var $\begin{pmatrix} \xi * \\ \zeta^* \end{pmatrix} \neq I_{p+m}$, violating (2.3). To see that T must also be a right unit of (A, U) we note that

(6.1)
$$(A, U) \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = (A, U) T' \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = (A, U) \begin{pmatrix} \xi * \\ \zeta * \end{pmatrix}$$

must hold for all values of the random variables x_i in ξ and z_i in ζ . If we were to assemble N = p + m linearly independent columns of such values in a matrix K, then $(A, U) \neq (A, U)T'$ would contradict (A, U)K = (A, U)T'K so that T' must be a right unit of (A, U). Orthogonality of T then implies that T must also be a right unit of (A, U). We thus have a

Lemma: If T relates two equivalent sets of uncorrelated factors $(\xi', \zeta')'$ and $(\xi^{*'}, \zeta^{*'})'$, then it must be an orthogonal right unit of (A, U). We now wish to show

Theorem 3: All orthogonal right units of B of order $p \times (p + m)$ with Eckart-Young decomposition (3.2) are of the form (3.3) if all d_i in D are nonzero.

(Guttman, 1955, is able to show necessity without this mild qualification that Σ be nonsingular.) Proof: Let us write the identity

$$(6.2) T = W(W'TW)W' = WS^*W'$$

where $W = (W_1, W_2)$ as defined in (3.1)-(3.4) and

(6.3)
$$S^* = W'TW = \begin{bmatrix} S_{11}^* & S_{12}^* \\ S_{21}^* & S_{22}^* \end{bmatrix}$$

is partitioned so that S_{11}^* is of order $p \times p$. Since T is an orthogonal right unit of B, by definition

(6.4)
$$BT = B \quad \text{and} \quad TT' = T'T = I_{p+m}$$

which implies that S^* in (6.3) must be orthogonal. The other condition in (6.4),

if written

(6.5)
$$VDW'_1(W_1, W_2)S^*\binom{W'_1}{W'_2} = VDW'_1,$$

is seen to reduce to

(6.6)
$$(I_{\mathfrak{p}}, \phi) S^* \binom{I_{\mathfrak{p}}}{\phi} = I_{\mathfrak{p}}$$

after premultiplication by $D^{-1}V'$ and postmultiplication by W_1 . Therefore,

$$S_{11}^* = I_n, \quad S_{12}^* = S_{21}^{*'} = \phi,$$

while $S_{22}^* = S(m \times m)$ can be chosen at will as long as it is orthogonal, so that S^* is. Hence T in (6.2) must be of the form (3.3), which is what we set out to prove.

Theorem 3, the lemma preceding it, and the corollary in sec. 4 combine to give the stronger

Theorem 4: Suppose a vector of random variables η satisfies the model of factor analysis for (A, U) and a set of uncorrelated factors $(\xi', \zeta')'$ and $\Sigma = \text{var}(\eta)$ is nonsingular. A necessary and sufficient condition that η also satisfies the model of factor analysis for another set of uncorrelated factors $(\xi^{*'}, \zeta^{*'})'$ is that both sets are related by T in (3.3), which is a function of an arbitrary orthogonal matrix S of order $m \times m$, where m is the number of common factors. The correlation matrix between both sets of equivalent, uncorrelated factors is T.

In closing we wish to reiterate two limitations of the present work: (i) our conclusions are restricted to uncorrelated factors and (ii) the minimum average correlation τ includes both common and unique factors. We are presently investigating the effect of removing either one of these restrictions.

REFERENCES

Camp, B. H. The converse of Spearman's two-factor theorem. Biometrika, 1932, 24, 418-427.
Cattell, R. B. Factor analysis: An introduction and manual for the psychologist and social scientist. New York: Harper, 1952.

Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211-218.

Fruchter, B. Introduction to factor analysis. Princeton, New Jersey: Van Nostrand, 1954.

Guttman, L. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. British Journal of Statistical Psychology, 1955, 8, 65-81.

Guttman, I. "Best possible" systematic estimates of communalities. Psychometrika, 1956, 21, 273-285.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Harman, H. H. Modern factor analysis. Second edition (revised). Chicago: University of Chicago Press, 1967.

Harris, C. W. On factors and factor scores. Psychometrika, 1967, 32, 363-379.

Heermann, E. F. The geometry of factorial indeterminacy. *Psychometrika*, 1964, 29, 371-381.

Heermann, E. F. The algebra of factorial indeterminacy. *Psychometrika*, 1966, 31, 539–543. Holzinger, K. J. & Harman, H. H. *Factor analysis*. Chicago: University of Chicago Press,

Kestelman, H. The fundamental equation of factor analysis. British Journal of Psychology, Stat. Sec., 1952, 5, 1-6.

Lawley, D. N. & Maxwell, A. E. Factor analysis as a statistical method. London: Butterworths, 1963.

Ledermann, W. The orthogonal transformation of a factorial matrix into itself. Psychometrika, 1938, 3, 181-187.

McDonald, R. P. & Burr, E. J. A comparison of four methods of constructing factor scores. *Psychometrika*, 1967, 32, 381-401.

Pawlik, K. Dimensionen des Verhaltens. Bern: H. Huber, 1968.

Piaggio, H. T. H. Three sets of conditions necessary for the existence of a g that is real and unique except in sign. British Journal of Psychology, 1933, 24, 88-105.

Schönemann, P. H. Book review: Überla, K. Faktorenanalyse; Pawlik, K. Dimensionen des Verhaltens. *Biometrics*, 1969, 25, 604–607.

Thomson, G. H. The meaning of "i" in the estimate of "g." British Journal of Psychology, 1934, 25, 92-99.

Thomson, G. H. The definition and measurement of "g" (general intelligence). Journal of Educational Psychology, 1935, 26, 241–262.

Thomson, G. H. Some points of mathematical technique in the factorial analysis of ability. Journal of Educational Psychology, 1936, 27, 37-54.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Überla, K. Faktorenanalyse. Berlin: Springer, 1968.

Wilson, E. B. On hierarchical correlation systems. Proc. N.A.S., 1928, 14, 283-291.

Manuscript received 6/26/69 Revised manuscript received 3/13/70