FACTOR ANALYSIS - SOME CLASS NOTES

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, New York

BU-748-M* November, 1981

Abstract

A description of factor analysis and its weaknesses.

Notation: $\underset{\sim}{X}^*$ represents a vector of random variables. (The asterisk emphasizes that $\underset{\sim}{X}^*$ is a vector, in contrast to the usual convention that $\underset{\sim}{X}$ is a matrix.)

0. Introduction

Factor analysis is a method of attempting to explain the correlations that are found to exist among a set of random variables represented by the vector $\underset{\sim}{X}^*$. The attempted explanation is in terms of a number of hypothetical and unobservable random variables called factors. The underlying idea is that maybe the correlations existing among the variables we can observe can be explained by their being linear combinations of some fewer number of variables (that we cannot observe). Thus if $\underset{\sim k\times 1}{F^*}$ represents these unobservable random variables called factors, the population model for $\underset{\sim}{X}^*$ is taken to be

$$\underset{\sim p\times 1}{X^*} = \underset{\sim p\times 1}{\mu} + \underset{\sim p\times k}{L}\underset{\sim k\times 1}{F^*} + \underset{\sim}{E^*} \quad .$$

---

* In the Biometrics Unit Series, Cornell University, Ithaca, New York.

Before explaining this equation, the point must be made that principal components is not factor analysis – even though many people think it is. The concept of principal components never conjectures $\underset{\sim}{X}^*$ as being explained in terms of any other variables – as does factor analysis. Principal components is concerned only with finding linear combinations of elements of $\underset{\sim}{X}^*$ that explain maximum amounts of variation in the $\underset{\sim}{X}^*$'s, and never refers to a model involving unobservable variances – see Kendall and Lawley (1956).

The fiction that principal components and factor analysis are part of the same analysis, or worse, that principal components is just one way of doing factor analysis, comes about from the fact that both procedures do have in common the desire to explain the behaviour of observed variables. Furthermore, there is an approximation to the factor analysis model that permits of a solution using the principal components calculations. However, the lack of a model is no cause for not carrying out a principal components analysis, whereas the absolute need of a model for carrying out factor analysis is the feature that distinguishes the two analyses. Despite this, confusion still abounds as, for example, in Harman (1967), a widely used reference for factor analysis.

Also, many computer programs include both principal component and factor analysis in the same package – and this also engenders the false notion that they are the same thing. Indeed, prior to Joreskog's (1967) finding a workable algorithm for solving the maximum likelihood equations for factor analysis, this method of estimation was seldom used. What has been used, and probably still is being used, is what is sometimes called PFA, principal factor analysis. This is described in Section 4 of these notes. It is an approximation to factor analysis that makes use of the calculations of principal components analysis.

1. The Model

a. Description

The factor analysis model is

$$\underset{\sim}{X}^*_{p\times 1} = \underset{\sim}{\mu} + \underset{\sim}{L}_{p\times k}\underset{\sim}{F}^*_{k\times 1} + \underset{\sim}{E}^*_{p\times 1} \tag{1}$$

where, using $\mathcal{E}$ for expected value,

$$\underset{\sim}{\mu} = \mathcal{E}(\underset{\sim}{X}^*) \quad , \tag{2}$$

and the other terms are defined as follows.

$\underset{\sim}{F}^*_{k\times 1}$ is a vector of k random variables, $k < p$, and these random variables are called factor scores or factors. They are unobservable.

$\underset{\sim}{L}_{p\times k} = \{\ell_{ir}\}$ is a $p \times k$ matrix of coefficents to be determined that are called factor loadings.

ℓ_{ir} = loading of factor r in variable i
= loading of variable i on factor r .

$\underset{\sim}{E}^*$ is a vector of p random (error) terms representing the difference between $\underset{\sim}{X}^*$ and $\underset{\sim}{\mu} + \underset{\sim}{L}\underset{\sim}{F}^*$.

Normality assumptions usually made are

$$\begin{bmatrix} \underset{\sim}{F}^* \\ \underset{\sim}{E}^* \end{bmatrix} \sim N\left[\begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix} \quad \begin{pmatrix} \underset{\sim}{U} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{W} \end{pmatrix}\right] \tag{3}$$

so that

$$\underset{\sim}{X}^* \sim N(\underset{\sim}{\mu}, \underset{\sim}{V}), \text{ with } \underset{\sim}{V} = \underset{\sim}{L}\underset{\sim}{U}\underset{\sim}{L}' + \underset{\sim}{W} \quad . \tag{4}$$

Simplified forms of $\underset{\sim}{U}$ and $\underset{\sim}{W}$ are often assumed:

$$\underset{\sim}{U} = \underset{\sim}{I} \tag{5}$$

and/or

$$\underset{\sim}{W} = \underset{\sim}{D}, \text{ diagonal} \tag{6}$$

or

$$\underset{\sim}{W} = \sigma^2\underset{\sim}{I} \quad . \tag{7}$$

b. Rotations

Define

$$\underset{\sim}{P}_{k\times k} \text{ as an orthogonal matrix} \quad . \tag{8}$$

Then the model (1) is unchanged if written as

$$\underset{\sim}{X}^* = \underset{\sim}{\mu} + \underset{\sim\sim}{L}\underset{\sim\sim}{P}'\underset{\sim\sim}{P}\underset{\sim}{F}^* + \underset{\sim}{E}^* = \underset{\sim}{\mu} + \underset{\sim\sim}{M}\underset{\sim}{G}^* + \underset{\sim}{E}^* \tag{9}$$

for

$$\underset{\sim}{M} = \underset{\sim\sim}{L}\underset{\sim\sim}{P}' \quad \text{and} \quad \underset{\sim}{G}^* = \underset{\sim\sim}{P}\underset{\sim}{F}^* \tag{10}$$

with

$$\begin{bmatrix} \underset{\sim}{G}^* \\ \underset{\sim}{E}^* \end{bmatrix} \sim N\left[\begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix} \begin{pmatrix} \underset{\sim\sim\sim}{P}\underset{\sim\sim\sim}{U}\underset{\sim\sim\sim}{P}' & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{W} \end{pmatrix}\right]$$

and

$$\underset{\sim}{X}^* \sim N(\underset{\sim}{0}, \underset{\sim}{V})$$

with

$$\underset{\sim}{V} = \underset{\sim}{M}(\underset{\sim\sim\sim}{P}\underset{\sim\sim\sim}{U}\underset{\sim\sim\sim}{P}')\underset{\sim}{M}' + \underset{\sim}{W} = \underset{\sim\sim}{L}\underset{\sim\sim\sim}{P}'\underset{\sim\sim\sim}{P}\underset{\sim\sim\sim}{U}\underset{\sim\sim\sim}{P}'\underset{\sim\sim}{P}\underset{\sim\sim}{L}' + \underset{\sim}{W} = \underset{\sim\sim\sim}{L}\underset{\sim\sim\sim}{U}\underset{\sim\sim\sim}{L}' + \underset{\sim}{W} \text{ of (4)} \quad .$$

Hence, in concept, the rotated model (9) is not different from the original model (1). Each is in terms of a set of factors, $\underset{\sim}{F}^*$ and $\underset{\sim}{G}^*$, respectively, each uncorrelated with $\underset{\sim}{E}^*$ and in both cases $\underset{\sim}{X}^*$ has the same covariance structure, namely $\underset{\sim}{V} = \underset{\sim\sim\sim}{L}\underset{\sim\sim\sim}{U}\underset{\sim\sim\sim}{L}' + \underset{\sim}{W}$. The only difference between the two models is that $\underset{\sim}{L}$ and $\underset{\sim}{M}$ do not have the same value. Since neither $\underset{\sim}{F}^*$, nor $\underset{\sim}{G}^* = \underset{\sim\sim}{P}\underset{\sim}{F}^*$, can be observed. This means that the factors represented by either $\underset{\sim}{F}^*$ or $\underset{\sim}{G}^*$ can be conceived of only to within the limits of an orthogonal transformation (or "rotation") – but this transformation is reflected in the coefficient matrices $\underset{\sim}{L}$ and $\underset{\sim}{M} = \underset{\sim\sim}{L}\underset{\sim\sim}{P}'$, which we seek to estimate.

Non-orthogonal transformations are also considered sometimes. These would involve considering the model (1) in the form

$$\underset{\sim}{X}^* = \underset{\sim}{\mu} + \underset{\sim\sim}{L}\underset{\sim\sim}{Q}^{-1}(\underset{\sim\sim}{Q}\underset{\sim}{F}^*) + \underset{\sim}{E}^*$$

where $\underset{\sim}{Q}$ is not orthogonal. Such transformations are sometimes called oblique rotations.

c. Numbers of factors

The whole objective of factor analysis is to explain the p X^*-variables in terms of fewer F^*-variables, namely k F^*-variables, with $k < p$. The variance structure given in (4) imposes limitations on k .

$$\underset{\sim}{V}_{pxp} = \underset{\sim}{L}_{pxk}\underset{\sim}{U}_{kxk}\underset{\sim}{L}'_{kxp} + \underset{\sim}{W}_{pxp} \quad . \tag{11}$$

In truth we know none of the elements of this equation. But $\underset{\sim}{X}^*$ can be observed and $\underset{\sim}{V}$ estimated, so let us suppose we know $\underset{\sim}{V}$. Then, ignoring rank, (11) represents $\frac{1}{2}p(p+1)$ equations in $pk + \frac{1}{2}k(k+1) + \frac{1}{2}p(p+1) = \frac{1}{2}p(p+1) + \frac{1}{2}k(k+1+2p)$ unknowns, i.e., always more unknowns than equations. Clearly, even if $\underset{\sim}{V}$ were known, there would be no way of solving (11). If $\underset{\sim}{W}$ is confined to being a diagonal matrix than we have $\frac{1}{2}p(p+1)$ equations in $pk + \frac{1}{2}k(k+1) + p = \frac{1}{2}(k+1)(2p+k)$ unknowns, which permits of solution if

$$(k+1)(2p+k) \leq p(p+1)$$

i.e.,

$$k^2 + 2pk + k - p(p-1) \leq 0 \quad . \tag{12}$$

Since

$$k^2 + 2pk + k - p(p-1) = 0$$

when

$$k = \tfrac{1}{2}[-(2p+1) \pm \sqrt{(2p+1)^2 + 4p(p-1)}] = \tfrac{1}{2}[-(2p+1) \pm \sqrt{[-(8p^2+1]} \quad ,$$

(12) is satisfied when

$$k < \tfrac{1}{2}[\sqrt{8p^2+1} - (2p+1)] \quad ,$$

i.e.,

$$k < \sqrt{2p^2 + \tfrac{1}{4}} - (p + \tfrac{1}{2}) \quad .$$

Examples

# of observable variables,	p =	1	2	3	4	5	6	7	8	9	10	21	100
# of factors,	k <	0	0	0	2	2	2	3	3	4	4	7	41

On taking $\underset{\sim}{U}$ as an identity matrix and $\underset{\sim}{W}$ as a diagonal matrix, as in (5) and (6), then (11) implies $\frac{1}{2}p(p+1)$ equations in $pk+p$ unknowns and we need

$$\tfrac{1}{2}(p+1) > k+1 \quad ,$$

i.e.,

$$k < \tfrac{1}{2}(p-1) \quad . \tag{13}$$

On further excluding orthogonal rotations (which for a $k \times k$ matrix implies $\frac{1}{2}k(k-1)$ relationships – orthogonality of k rows) we would then have $\frac{1}{2}p(p+1)$ equations in $pk - \frac{1}{2}k(k-1) + p$ unknowns requiring

$$\tfrac{1}{2}p(p+1) > pk - \tfrac{1}{2}k(k-1) + p \quad ,$$

i.e.,

$$k^2 - 2pk - k + p^2 - p \geq 0 \quad .$$

Since

$$k^2 - (2p+1)k + p(p-1) = 0$$

when

$$k = \tfrac{1}{2}[(2p+1) \pm \sqrt{(2p+1)^2 - 4p(p-1)}] = [p + \tfrac{1}{2} + \sqrt{2p+\tfrac{1}{4}}] \quad ,$$

the inequality is satisfied when

$$k < p + \tfrac{1}{2} + \sqrt{2p+\tfrac{1}{4}} \quad .$$

Examples

# of observed variables,	p =	1	2	3	4	5	6	7	8	9	10	20	100
# of factors,	k <	0	1	2	2	3	4	4	5	6	7	14	86

Other than these results there seems to be inherently no part of factor analysis methodology that determines (or even gives guidance on) how many factors, k, there will be in a factor analysis representation of p observable random variables.

2. Maximum Likelihood Estimation

The factor analysis model is (1):

$$\underset{\sim}{X}^{*}_{p\times 1} = \underset{\sim}{\mu} + \underset{\sim\sim}{L}\underset{\sim}{F}^{*} + \underset{\sim}{E}^{*} \quad . \tag{1}$$

$\underset{\sim}{X}^{*}$ represents p random variables that can be observed: $\underset{\sim}{X}_{p\times N}$ represents a matrix of N realized values of $\underset{\sim}{X}^{*}$. Even though $\underset{\sim}{F}^{*}$ cannot be observed, and despite there being no procedure for deciding how many elements are in $\underset{\sim}{F}^{*}$ (other than some upper limit), the estimation problem is to estimate the factor loadings defined as $\underset{\sim}{L}$.

a. Equations to be solved

We confine attention to estimation by maximum likelihood and take assumptions (5) and (6) so that

$$\underset{\sim}{V} = \underset{\sim\sim}{L}\underset{\sim\sim}{L}' + \underset{\sim}{D} \quad . \tag{15}$$

Since in the model (1) for $\underset{\sim}{X}^{*}$ the random variables in $\underset{\sim}{F}^{*}$ cannot be measured, the estimation problem is to estimate $\underset{\sim}{L}$. And because $\underset{\sim}{L}$ occurs in $\underset{\sim}{V}$, as does $\underset{\sim}{D}$ also, the overall estimation problem is to estimate $\underset{\sim}{L}$ and $\underset{\sim}{D}$ of $\underset{\sim}{V} = \underset{\sim\sim}{L}\underset{\sim\sim}{L}' + \underset{\sim}{D}$. The logical statistic for this purpose is the sum of squares and products matrix $\underset{\sim}{S}$,

$$N\underset{\sim}{S} = \underset{\sim\sim}{X}\underset{\sim\sim}{X}' - N\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}' \quad ,$$

where $\bar{\underset{\sim}{x}}$ is the vector of means $\underset{\sim}{x} = \underset{\sim\sim}{X}\underset{\sim}{1}_{N}/N$. Then $\underset{\sim}{S}$ has the Wishart distribution $W(\underset{\sim}{V},p,N)$ with density

$$\frac{|\underset{\sim}{S}|^{\frac{1}{2}(N-p-1)}\exp -\frac{1}{2}\mathrm{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{S})}{|\underset{\sim}{V}|^{\frac{1}{2}N}\, 2^{\frac{1}{2}NP}\pi^{\frac{1}{4}p(p-1)}\prod_{i=1}^{p}\Gamma\left(\frac{n+1-i}{2}\right)} \quad .$$

Therefore, for estimating elements of $\underset{\sim}{L}$ and $\underset{\sim}{D}$, the log likelihood based on this density is used, namely

$$\theta = \text{constant} - \tfrac{1}{2}N\log|\underset{\sim}{V}| - \tfrac{1}{2}N\mathrm{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{S}) \quad .$$

This has to be minimized with respect to each d_r and each ℓ_{ij}. The result (see Appendix) is

$$\sum_j \hat{\ell}_{ij}^2 + \hat{d}_i = s_{ii} \qquad \text{for } i = 1, \cdots, p \tag{16}$$

and

$$\hat{\underset{\sim}{L}}'\hat{\underset{\sim}{D}}^{-1}\underset{\sim}{S} = (\underset{\sim}{I} + \hat{\underset{\sim}{L}}'\hat{\underset{\sim}{D}}^{-1}\hat{\underset{\sim}{L}})\hat{\underset{\sim}{L}}' \tag{17}$$

with s_{ii} being the i'th diagonal element of $\underset{\sim}{S}$. Clearly, there is no analytic solution to (16) and (17).

b. No unique solution

Had we used $\underset{\sim}{M} = \underset{\sim}{L}\underset{\sim}{P}'$ for P orthogonal these equations would be unchanged: (16), which for $\underset{\sim}{\ell}_i'$ being the i'th row of $\underset{\sim}{L}$ and for $\underset{\sim}{m}_i'$ being the i'th row of $\underset{\sim}{M}$, would become

$$s_{ii} = \hat{\underset{\sim}{m}}_i'\hat{\underset{\sim}{m}}_i + \hat{d}_i = \hat{\underset{\sim}{\ell}}_i'\underset{\sim}{P}'\underset{\sim}{P}\hat{\underset{\sim}{\ell}}_i + \hat{d}_i = \hat{\underset{\sim}{\ell}}_i'\hat{\underset{\sim}{\ell}}_i + \hat{d}_i \quad ,$$

which is still (16); and (17) would be

$$\hat{\underset{\sim}{M}}'\hat{\underset{\sim}{D}}^{-1}\underset{\sim}{S} = (\underset{\sim}{I} + \hat{\underset{\sim}{M}}'\hat{\underset{\sim}{D}}^{-1}\hat{\underset{\sim}{M}})\hat{\underset{\sim}{M}}' \quad ,$$

which is

$$\underset{\sim}{P}\hat{\underset{\sim}{L}}'\hat{\underset{\sim}{D}}^{-1}\underset{\sim}{S} = \underset{\sim}{P}\hat{\underset{\sim}{L}}' + \underset{\sim}{P}\hat{\underset{\sim}{L}}'\hat{\underset{\sim}{D}}^{-1}\hat{\underset{\sim}{L}}\underset{\sim}{P}'\underset{\sim}{P}\hat{\underset{\sim}{L}}'$$

and this reduces back to (17).

Thus we see that these maximum likelihood equations have no unique solution. They can be solved only to within an orthogonal transformation or rotation. Of this Lawley and Maxwell (p. 11, 1st ed.) write "In this situation all the statistician can do is to select a particular solution, one which is convenient to find, and leave the experimenter to apply whatever rotation he thinks desirable." Part of this advice ("a solution which is convenient to find") is familiar in the context of the general linear model not of full rank. But in that case the statistician can go on and tell the

experimenter how to usefully use his non-unique solution to yield unique results (estimates of estimable functions and so on). But this kind of advice is not possible in the factor analysis situation. Equations (16) and (17) have no unique solution, not because of any rank deficiency as in the linear model case, but simply by virtue of the nature of the model and analysis proposed. The model and analysis therefore seem open to question.

One way out of the predicament is to introduce more equations. Lawley and Maxwell do this by requiring $\underset{\sim}{L}$ to be such that

$$\underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{L} \text{ is diagonal} \quad . \tag{18}$$

However, justification for this seems to be slight and indeed they "... ignore the possibility that any of the diagonal elements of $\underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{L}$ are equal", a fact which would negate the effectiveness of (18) ensuring a unique solution for (16) and (17).

3. Interpretation

a. The factor loadings

Having obtained $\hat{\underset{\sim}{L}}$, a solution for (16) and (17), what can be done with it? The definition of $\underset{\sim}{L}$ given at the outset is

$$\underset{\sim}{L} = \{\underset{\sim}{\ell}_{ir}\}$$

for

$$\underset{\sim}{\ell}_{ir} = \text{factor loading of variable i on factor r} \quad .$$

In the words of Lawley and Maxwell, the experimenter, after obtaining one $\hat{\underset{\sim}{L}}$ is left "to apply whatever rotation he thinks desirable". This simply produces another $\hat{\underset{\sim}{L}}$ and, presumably, his ideas on what is "desirable" imply that one settles for an $\hat{\underset{\sim}{L}}$ that one likes the look of, meaning no doubt an $\hat{\underset{\sim}{L}}$ that one can interpret, and hopefully, interpret in a useful manner. Thus it is

that in books and papers on factor analysis one sees injunctions such as "rotate until all loadings, or as many as possible, are positive". And then, it seems, factors are interpreted by inspection. For example, suppose data on high school examinations in algebra, geometry, English and French are subjected to a factor analysis and the loadings come out as follows

	Factor 1	Factor 2	Factor 3
Algebra	.95	-.03	-1.01
Geometry	.87	.04	-1.04
English	-.01	.88	.96
French	.02	.87	.87

then it is not difficult to interpret the factors as (1) an aptitude in mathematics, (2) an aptitude in language, and (3) the difference between them. But this is based purely on the outcome of the data. Had factor 3 not been as shown, but had yielded loadings -.03, .96, -.46 and .59, what kind of interpretation would that bring forth? And if one didn't like the factors so produced, or was unable to interpret them, is it then a statistically sound procedure to simply keep rotating until one does like the outcome and can interpret it? What assurance is there that future data will yield concomitant interpretation? Shouldn't modelling be done prior to data analysis and not post hoc?

Examples in the literature of conducting a factor analysis on "data" constructed from known factors, or with known relationships, are not reassuring: factor analysis does not always reveal the factors that are known to be the basis of the "data". Furthermore, the whole problem is additionally complicated by the question: how many factors? Allowing for differing numbers of factors does not always produce similar results: e.g., a four-factor analysis does not always yield factors in common with a three-factor analysis. As Press (p. 312) says, "... interpretation of results ... is an extremely important but often troublesome aspect of factor analysis." See also Armstrong (1967).

b. The factors

Interpretation of factor analysis is usually in terms of the relative magnitudes of the estimated factor loadings (elements of $\hat{\underset{\sim}{L}}$). Thus a study reported in Press (p. 314) consisted of a factor analysis of seven factors on the basis of 63 securities. Interpretation based on $\hat{\underset{\sim}{L}}$ of order 63×7 directed attention initially to the fact that the entries in the first column of $\hat{\underset{\sim}{L}}$ were all large. The conclusion from this was that factor 1 affected each of the 63 variables to a large extent and could therefore be interpreted as some kind of overall market factor that affects all securities.

But why cannot the factors themselves be estimated directly? Lawley and Maxwell (p. 88, 1st ed.) suggest

$$\underset{\sim}{F}^* = \hat{\underset{\sim}{L}}^{-1}\underset{\sim}{S}^{-1}\underset{\sim}{X}^* \quad .$$

However, since the model equation is

$$\underset{\sim}{X}^* = \underset{\sim}{\mu} + \underset{\sim\sim}{L}\underset{\sim}{F}^* + \underset{\sim}{E}^*$$

with $\text{var}(\underset{\sim}{X}^*) = \underset{\sim}{V}$, and because $\hat{\underset{\sim}{L}}$ will have full column rank (presumably $\underset{\sim}{L}$ does also, for otherwise $\underset{\sim}{F}^*$ would include variables that were utilized only as linear combinations of other variables), why not estimate $\underset{\sim}{F}^*$ as

$$\underset{\sim}{F}^* = (\hat{\underset{\sim}{L}}'\underset{\sim}{S}^{-1}\hat{\underset{\sim}{L}})^{-1}\hat{\underset{\sim}{L}}'\underset{\sim}{S}^{-1}(\underset{\sim}{X}^* - \underset{\sim}{\mu}) \quad ,$$

analogous to GLS procedures; or perhaps as

$$\underset{\sim}{F}^* = (\hat{\underset{\sim}{L}}'\hat{\underset{\sim}{D}}^{-1}\hat{\underset{\sim}{L}})^{-1}\hat{\underset{\sim}{L}}'\hat{\underset{\sim}{D}}^{-1}(\underset{\sim}{X}^* - \underset{\sim}{\mu})$$

or, using LS ideas as

$$\underset{\sim}{F}^* = (\hat{\underset{\sim}{L}}'\hat{\underset{\sim}{L}})^{-1}\hat{\underset{\sim}{L}}'(\underset{\sim}{X}^* - \underset{\sim}{\mu})$$

or, using the ideas of predicting random variables from mixed models (see Searle, 1971, p. 462) as

$$\text{predicted } \underset{\sim}{F}^* = \underset{\sim}{L}'\underset{\sim}{V}^{-1}(\underset{\sim}{X}^* - \underset{\sim}{\mu}) \quad .$$

Any of these ways gives a matrix whose row elements indicate the extent to which the variables in $\underset{\sim}{X}^*$ get utilized in forming the elements of $\underset{\sim}{F}^*$. In this way an indication is directly available as to the composition of $\underset{\sim}{F}^*$.

c. Correlation

It is often argued that factor analysis should be free of the effects of scale and dispersion. The analysis is then carried out using the correlation matrix $\underset{\sim}{R}$ in place of $\underset{\sim}{S}$. And to take account of intraclass correlations among the $\underset{\sim}{X}$'s, these are sometimes estimated as r_i . Then values $1 - r_i$ for $i = 1, \cdots, p$ are used in place of the diagonal unities in R^2 . $r_i d_i$ is then called a communality.

4. Factor Analysis and Principal Components

Suppose that the error terms $\underset{\sim}{E}^*$ in the factor analysis model (1) are ignored – or at least that their variances, the d_i of $\underset{\sim}{D}$, are deemed small enough to be ignored. Then in (15) the equation for $\underset{\sim}{V}$ is

$$\underset{\sim p\times p}{V} = \underset{\sim p\times k}{L}(\underset{\sim}{L}')_{k\times p} \quad . \tag{19}$$

In principal components analysis we dealt with the eigen roots and vectors of $\underset{\sim}{V}$ summarixed in the equation

$$\underset{\sim}{B}'\underset{\sim\sim}{V}\underset{\sim}{B} = \underset{\sim}{D}_\lambda$$

where $\underset{\sim}{B}$ is an orthogonal matrix, its columns being eigen vector of $\underset{\sim}{V}$, and $\underset{\sim}{D}_\lambda$ is a diagonal matrix of the corresponding eigen roots. Thus

$$\underset{\sim}{V} = \underset{\sim\sim}{B}\underset{\sim\sim}{D}_\lambda\underset{\sim}{B}' = \underset{\sim\sim}{B}\underset{\sim\sim}{D}_\lambda^{\frac{1}{2}}(\underset{\sim\sim}{B}\underset{\sim\sim}{D}_\lambda^{\frac{1}{2}})' \tag{20}$$

where $(\underset{\sim}{D}_\lambda^{\frac{1}{2}})^2 = \underset{\sim}{D}_\lambda$, which exists because $\underset{\sim}{V}$ is p.d. and has positive eigen roots. Comparing (19) and (20) we see that a possible value of $\underset{\sim}{L}$ is

$$\underset{\sim}{L} = \underset{\sim\sim}{B}\underset{\sim\sim}{D}_\lambda^{\frac{1}{2}} \quad .$$

And if for $m < p$, $\underset{\sim}{B}$ and $\underset{\sim}{D}_\lambda$ are partitioned as

$$\underset{\sim}{B} = \underset{p \times m}{(\underset{\sim}{B}_1 \quad \underset{\sim}{B}_2)} \qquad \text{and} \qquad \underset{\sim}{D}_\lambda^{\frac{1}{2}} = \begin{bmatrix} \overset{m \times m}{\downarrow} & \\ \underset{\sim}{D}_1^{\frac{1}{2}} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{D}_2^{\frac{1}{2}} \end{bmatrix}$$

then a comparable analysis based on just m of the eigen vectors is

$$(\underset{\sim}{L}_1)_{p \times m} = \underset{\sim}{B}_1 \underset{\sim}{D}_1^{\frac{1}{2}} \quad .$$

This is known as PFA, <u>principal components analysis</u>, or sometimes as the principal components solution to factor analysis. It then yields

$$\underset{\sim}{X}^* = \underset{\sim}{\mu} + \underset{\sim}{L}_1 \underset{\sim}{F}^* + \underset{\sim}{E}^*$$

as

$$\underset{\sim}{X}^* = \underset{\sim}{\mu} + \underset{\sim}{B}_1 \underset{\sim}{D}_1^{\frac{1}{2}} \underset{\sim}{F}^* + \underset{\sim}{E}^*$$

or, equivalently, because $\underset{\sim}{B}_1$ has orthonormal columns

$$\underset{\sim}{F}^* = \underset{\sim}{D}_1^{-\frac{1}{2}} \underset{\sim}{B}_1' (\underset{\sim}{X}^* - \underset{\sim}{\mu} - \underset{\sim}{E}^*) \quad .$$

In the form

$$\underset{\sim}{F}^* = \underset{\sim}{D}_1^{-\frac{1}{2}} \underset{\sim}{B}_1' \underset{\sim}{X}^*$$

they are simply normalized forms of the principal components. This is the sense in which factor analysis and principal components are sometimes put together.

<u>Estimation</u>

$$\underset{\sim}{\tilde{B}}' \underset{\sim}{S} \underset{\sim}{\tilde{B}} = \underset{\sim}{\tilde{D}} \quad .$$

The diagonal elements of $\underset{\sim}{\tilde{D}}$ are $\tilde{\lambda}$'s, eigen values of $\underset{\sim}{S}$.

Appendix

Derivation of Maximum Likelihood Equations

We need to maximize

$$\theta = \text{constant} - \tfrac{1}{2}N\log|\underset{\sim}{V}| - \tfrac{1}{2}N\text{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{S}) \tag{A1}$$

with respect to elements of $\underset{\sim}{L} = \{\ell_{ij}\}$ and of $\underset{\sim}{D} = \text{diag}\{d_r\}$ where

$$\underset{\sim}{V} = \underset{\sim}{L}\underset{\sim}{L}' + \underset{\sim}{D} \quad . \tag{A2}$$

The following results from matrix differentiation are used:

$$\frac{\partial}{\partial x}\log|\underset{\sim}{A}| = \text{tr}\left(\underset{\sim}{A}^{-1}\frac{\partial \underset{\sim}{A}}{\partial x}\right) \quad \text{and} \quad \frac{\partial}{\partial x}\underset{\sim}{A}^{-1} = -\underset{\sim}{A}^{-1}\frac{\partial \underset{\sim}{A}}{\partial x}\underset{\sim}{A}^{-1} \quad . \tag{A3}$$

Also,

$$\frac{\partial \underset{\sim}{A}}{\partial a_{ij}} = \underset{\sim}{F}_{ij} = \left\{\begin{array}{l}\text{null matrix except that}\\ \text{element } (i,j) \text{ is unity}\end{array}\right\} \tag{A4}$$

and

$$\text{tr}(\underset{\sim}{A}\underset{\sim}{F}_{ij}) = a_{ji} \quad . \tag{A5}$$

First, from (A1),

$$\frac{\partial\theta}{\partial d_r} = \frac{-N}{2}\text{tr}\left(\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial d_r}\right) + \frac{N}{2}\text{tr}\left(\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial d_r}\underset{\sim}{V}^{-1}\underset{\sim}{S}\right)$$

and equating this to 0 gives

$$\begin{aligned}\text{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{F}_{rr}) &= \text{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{F}_{rr}\underset{\sim}{V}^{-1}\underset{\sim}{S})\\ &= \text{tr}(\underset{\sim}{V}^{-1}\underset{\sim}{S}\underset{\sim}{V}^{-1}\underset{\sim}{F}_{rr}) \quad .\end{aligned}$$

Hence, using (A5)

$$(r,r)\text{'th element of } \underset{\sim}{V}^{-1} = (r,r)\text{'th element of } \underset{\sim}{V}^{-1}\underset{\sim}{S}\underset{\sim}{V}^{-1}$$

$$\therefore (r,r)\text{'th element of } \underset{\sim}{V} = (r,r)\text{'th element of } \underset{\sim}{S}$$

$$\therefore \sum_j \ell_{rj}^2 + d_r = s_{rr} \quad . \tag{A6}$$

Second, from (A1) again

$$\frac{\partial\theta}{\partial\ell_{ij}} = \frac{-N}{2}\,\mathrm{tr}\left(\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial\ell_{ij}}\right) + \frac{N}{2}\,\mathrm{tr}\left(\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial\ell_{ij}}\,\underset{\sim}{V}^{-1}\underset{\sim}{S}\right)$$

and equating this to 0 gives

$$\mathrm{tr}\left(\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim\sim}{LL'}}{\partial\ell_{ij}}\right) = \mathrm{tr}\left(\underset{\sim}{V}^{-1}\underset{\sim\sim}{S}\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim\sim}{LL'}}{\partial\ell_{ij}}\right) \quad . \tag{A7}$$

Now, for any symmetric matrix $\underset{\sim}{M}$

$$\begin{aligned}\mathrm{tr}\left(\underset{\sim}{M}\,\frac{\partial \underset{\sim\sim}{LL'}}{\partial\ell_{ij}}\right) &= \mathrm{tr}\left(\underset{\sim\sim}{M}\underset{\sim\sim ij}{F}\underset{\sim}{L'} + \underset{\sim\sim}{ML}\underset{\sim\sim ji}{F}\right)\\ &= \mathrm{tr}\left(\underset{\sim}{L'}\underset{\sim\sim}{M}\underset{\sim\sim ij}{F} + \underset{\sim\sim}{ML}\underset{\sim\sim ji}{F}\right)\\ &= (j,i)\text{'th element of } \underset{\sim}{L'}\underset{\sim}{M} + (i,j)\text{'th element of } \underset{\sim\sim}{ML}\\ &= 2[(i,j)\text{'th element of } \underset{\sim\sim}{ML}] \quad . \end{aligned} \tag{A8}$$

Therefore

$$\left\{\mathrm{tr}\left(\underset{\sim}{M}\,\frac{\partial \underset{\sim\sim}{LL'}}{\partial\ell_{ij}}\right)\right\} = 2\underset{\sim\sim}{ML} \quad .$$

Applying this to (A7) gives

$$\underset{\sim}{V}^{-1}\underset{\sim}{L} = \underset{\sim}{V}^{-1}\underset{\sim\sim}{S}\underset{\sim}{V}^{-1}\underset{\sim}{L} \quad ,$$

i.e.,

$$\underset{\sim}{L} = \underset{\sim\sim}{S}\underset{\sim}{V}^{-1}\underset{\sim}{L} \quad .$$

It is convenient to work this into an alternative form that does not involve $\underset{\sim}{V}^{-1}$.

$$\underset{\sim\sim}{LL'} = \underset{\sim\sim}{S}\underset{\sim}{V}^{-1}\underset{\sim\sim}{LL'} = \underset{\sim\sim}{S}\underset{\sim}{V}^{-1}(\underset{\sim}{V} - \underset{\sim}{D}) = \underset{\sim}{S} - \underset{\sim\sim}{S}\underset{\sim}{V}^{-1}\underset{\sim}{D}$$

$$\therefore \underset{\sim\sim}{V}\underset{\sim}{D}^{-1}\underset{\sim\sim}{LL'} = \underset{\sim\sim}{V}\underset{\sim}{D}^{-1}\underset{\sim}{S} - \underset{\sim}{S} \quad ,$$

and so

$$(\underset{\sim}{L}\underset{\sim}{L}' + \underset{\sim}{D})\underset{\sim}{D}^{-1}\underset{\sim}{L}\underset{\sim}{L}' = (\underset{\sim}{L}\underset{\sim}{L}'\underset{\sim}{D}^{-1} + \underset{\sim}{I})\underset{\sim}{S} - \underset{\sim}{S}$$

or

$$\underset{\sim}{L}\underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{L}\underset{\sim}{L}' + \underset{\sim}{L}\underset{\sim}{L}' = \underset{\sim}{L}\underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{S} \quad .$$

Then, using $\underset{\sim}{L}\underset{\sim}{L}'\underset{\sim}{P} = \underset{\sim}{L}\underset{\sim}{L}'\underset{\sim}{Q} \Rightarrow \underset{\sim}{L}'\underset{\sim}{P} = \underset{\sim}{L}'\underset{\sim}{Q}$ for any real $\underset{\sim}{P}$, $\underset{\sim}{Q}$ and $\underset{\sim}{L}$, we have

$$\underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{L}\underset{\sim}{L}' + \underset{\sim}{L}' = \underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{S}$$

or

$$(\underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{L} + \underset{\sim}{I})\underset{\sim}{L}' = \underset{\sim}{L}'\underset{\sim}{D}^{-1}\underset{\sim}{S} \quad .$$

This and

$$\Sigma \ell_{rj}^2 + d_r = s_{rr} \text{ for } r = 1, \cdots, p$$

of (A6) are the maximum likelihood equations that have to be solved for ℓ_{rj} and d_r .

References

Armstrong, J. Scott (1967). Derivation of theory by means of factor analysis, or Tom Swift and his electric factor analysis machine. The American Statistician 21, 17-21.

Harmon, H. H. (1967). Modern Factor Analysis, 2nd Ed. Chicago University Press.

Joreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika 32, 443-482.

Joreskog, K. G. and van Thillo, M. (1971). New rapid algorithms for factor analysis by unweighted least squares, generalised least squares and maximum likelihood. Research Memorandum, Educational Testing Service, Princeton, N. J. (May 1971)

Kendall, M. G. and Lawley, D. N. (1956). The principles of factor analysis. J. Roy. Stat. Soc. A, 119, 83-84.

Lawley, D. N. and Maxwell, A. E. (1971). Factor Analysis as a Statistical Method, 2nd Ed. American Elsevier Publishing Co., N. Y. (Review, JASA, March 1973, p. 224.)

Press, S. J. (1972). Applied Multivariate Analysis. Holt, Rinehart and Winston, Inc., N. Y.

Searle, S. R. (1971). Linear Models. Wiley, N. Y.