DOCUMENT RESUME

ED 055 110		TM 000 837
AUTHOR	Joreskog, Karl G.	
TITLE	Simultaneous Factor Analysis in Several Populations.	
INSTITUTION	Educational Testing Service, Princeton, N.J.	
PUB DATE	Nov 70	
NOTE	33p.	
EDRS PRICE	MF-$0.65 HC-$3.29	
DESCRIPTORS	*Factor Analysis; *Factor Structure; Goodness of Fit; Hypothesis Testing; Mathematical Applications; *Mathematical Models; *Research Methodology; *Sampling; Statistical Analysis	

ABSTRACT

This paper is concerned with the study of similarities and differences in factor structures between different groups. A common situation is when a battery of tests has been administered to samples of examinees from several populations. A very general model is presented, in which any parameter in the factor analysis models (factor loadings, factor variances, factor covariances, and unique variances) for the different groups may be assigned an arbitrary value or constrained to be equal to some other parameter. Given such a specification, the model is estimated by the maximum likelihood method yielding a large sample chi-square of goodness of fit. By computing several solutions under different specifications one can test various hypotheses. The method is capable of dealing with any degree of invariance, from the one extreme, where nothing is invariant, to the other extreme, where everything is invariant. Neither the number of tests nor the number of common factors need to be the same for all groups, but to be at all interesting, it is assumed that there is a common core of tests in each battery that is the same or at least content-wise comparable. (Author)

RESEARCH

BULLETIN

RB-70-61

SIMULTANEOUS FACTOR ANALYSIS IN SEVERAL POPULATIONS

K. G. Jöreskog

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
November 1970

Simultaneous Factor Analysis in Several Populations

Abstract

This paper is concerned with the study of similarities and differences in factor structures between different groups. A common situation is when a battery of tests has been administered to samples of examinees from several populations.

A very general model is presented, in which any parameter in the factor analysis models (factor loadings, factor variances, factor covariances, and unique variances) for the different groups may be assigned an arbitrary value or constrained to be equal to some other parameter. Given such a specification, the model is estimated by the maximum likelihood method yielding a large sample χ^2 of goodness of fit. By computing several solutions under different specifications one can test various hypotheses.

The method is capable of dealing with any degree of invariance, from the one extreme, where nothing is invariant, to the other extreme, where everything is invariant. Neither the number of tests nor the number of common factors need to be the same for all groups, but to be at all interesting, it is assumed that there is a common core of tests in each battery that is the same or at least content-wise comparable.

Simultaneous Factor Analysis in Several Populations*

1. Introduction and Summary

This paper is concerned with the study of similarities and differences in factor structures between different groups. A common situation is when a battery of tests has been administered to samples of examinees from several populations. Traditionally this type of problem has been solved by obtaining orthogonal unrotated solutions for each group separately, rotating these to similarity and examining various similarity indices.

Perhaps the best approach to the problem is that of Meredith [1964a,b], who has shown that, under certain conditions, when the various populations are derivable as subpopulations from a parent population under selection on some external variable, there is a factor pattern that is invariant over populations. Meredith [1964b] gives two methods for estimating the common factor pattern by least squares rotation of independent orthogonal solutions for each group into a common factor pattern. If this can be achieved, the common factor pattern may be rotated further, orthogonally or obliquely, to a more readily interpretable solution.

The method to be presented is both more general and statistically more optimal. It is more general in several respects. Firstly, the method may be used regardless of whether the populations are derived by selection or not. The only requirement is that the populations be clearly defined and the samples independent. Secondly, the method is capable of dealing with

---

*This research was supported by grant NSF-GB-12959 from National Science Foundation. My thanks are due to Marielle van Thillo who checked the mathematical derivations and wrote and debugged the computer program SIFASP.

any degree of invariance, from the one extreme, where nothing is invariant, to the other extreme, where everything is invariant. Thirdly, neither the number of tests nor the number of common factors need to be the same for all groups, but to be at all interesting it is assumed that there is a common core of tests in each battery that is the same or at least content-wise comparable.

A very general model is presented, in which any parameter in the factor analysis models (factor loading, factor variance, factor covariance, and unique variance) for the different populations may be assigned an arbitrary value or constrained to be equal to some other parameter. Given such a specification, the model is estimated by the maximum likelihood method assuming the observed variables to have a multinormal distribution in each population. This yields a large sample χ^2 test of the goodness of fit of the overall model. By computing several solutions under different specifications one can test various hypotheses. For example, one can test the hypothesis of an invariant factor pattern or the hypothesis of an invariant specified simple structure factor pattern.

2. A General Model

2.1 The Model

Consider a set of m populations $\Pi_1, \Pi_2, \ldots, \Pi_m$. These may be different nations, or culturally different groups, groups of individuals selected on the basis of some known or unknown selection variable, groups receiving different treatments, etc. In fact, they may be any set of exclusive groups of individuals that are clearly defined. It is assumed that a battery of tests has been administered to a sample of individuals from each population. The battery of tests need not be the same for each group, nor need the number of tests be the same.

However, since we shall be concerned with characteristics of the tests that are invariant over populations, it is necessary that some of the tests in each battery are the same or at least content-wise equivalent.

Let p_g be the number of tests administered to group g and let x_g be a vector of order p_g, representing the measurements obtained in group g. We regard x_g as a random vector with mean vector μ_g and variance-covariance Σ_g. It is assumed that a factor analysis model holds in each population so that x_g can be accounted for by k_g common factors f_g and p_g unique factors z_g, as

(1) $$x_g = \mu_g + \Lambda_g f_g + z_g \quad ,$$

with $\mathcal{E}(f_g) = 0$ and $\mathcal{E}(z_g) = 0$ and Λ_g a factor pattern of order $p_g \times k_g$. The usual factor analytic assumptions then imply that

(2) $$\Sigma_g = \Lambda_g \Phi_g \Lambda_g' + \psi_g^2 \quad ,$$

where Φ_g is the variance-covariance matrix of f_g and ψ_g^2 is the diagonal variance-covariance matrix of z_g.

In addition to assuming that a factor analytic model holds in each population the model may specify that certain parameters in Λ_g, Φ_g, ψ_g, $g = 1,2,\ldots,m$ have assigned values and that some set of unknown elements in Λ_g, Φ_g and ψ_g are the same for all g. The most common situation is when the same battery has been administered to each group and when the whole factor pattern Λ_g is assumed to be invariant over groups. This case will be considered separately in section 3.

2.2 Identification of Parameters

Before an attempt is made to estimate a model of this kind, the identification problem must be examined. The identification problem depends on the specification of fixed, free and constrained parameters. Under a given specification, each Λ_g, Φ_g and Ψ_g generates one and only one Σ_g but it is well known that different Λ_g and Φ_g can generate the same Σ_g. It should be noted that if Λ_g is replaced by $\Lambda_g T_g^{-1}$ and Φ_g by $T_g \Phi_g T_g'$, where T_g is an arbitrary nonsingular matrix of order $k_g \times k_g$, then Σ_g is unchanged. Since T_g has k_g^2 independent elements, this suggests that k_g^2 independent conditions should be imposed on Λ_g and/or Φ_g to make these uniquely defined and hence that $\sum_{g=1}^{m} k_g^2$ independent conditions altogether should be imposed. However, when equality constraints over groups are taken into account, all the elements of all the transformation matrices are not independent of each other and therefore a lesser number of conditions need to be imposed. It is hard to give further specific rules in the general case. For the special case when the whole factor pattern is invariant over groups, however, a more precise consideration of the identification problem is given in section 3.2. In other cases one should verify that the only transformations $T_1, T_2, \ldots, T_m$ that preserve the specification about fixed, free and constrained parameters are identity matrices.

2.3 Estimation and Testing of the Model

Let N_g be the number of individuals in the sample from the g^{th} population and let $\bar{x}_g$ be the usual sample mean vector and S_g the usual sample variance-covariance matrix with $n_g = N_g - 1$ degrees of freedom. The only requirement for the sampling procedure is that it produces independent measurements for the different groups.

If we assume that x_g has a multinormal distribution it follows that S_g has a Wishart distribution based on Σ_g and n_g degrees of freedom. The logarithm of the likelihood for the g^{th} sample is

$$(3) \qquad \log L_g = -\frac{1}{2} n_g[\log|\Sigma_g| + tr(S_g\Sigma_g^{-1})] \quad .$$

Since the samples are independent, the log-likelihood for all the samples is

$$(4) \qquad \log L = \sum_{g=1}^{m} \log L_g \quad .$$

Maximum likelihood estimates of the unknown elements in Λ_g , Φ_g , Ψ_g , $g = 1,2,\ldots,m$, may be obtained by maximizing $\log L$. However, it is slightly more convenient to minimize

$$(5) \qquad F = \frac{1}{2} \sum_{g=1}^{m} n_g[\log|\Sigma_g| + tr(S_g\Sigma_g^{-1}) - \log|S_g| - p_g]$$

instead. At the minimum, F equals minus the logarithm of the likelihood ratio for testing the hypothesis implied by the model against the general alternative that each Σ_g is unconstrained. Therefore, twice the minimum value of F is approximately distributed, in large samples, as χ^2 with degrees of freedom equal to

(6) $$d = \sum_{g=1}^{m} \frac{1}{2} p_g(p_g + 1) - t$$

where t is the total number of independent parameters estimated in the model.

2.4 Minimization Procedure

The function F will be minimized numerically with respect to the independent parameters using a modification of the method of Fletcher and Powell [1963]. The application of this method makes use of exact expressions for first-order derivatives and approximate expressions for second-order derivatives of F.

Let

(7) $$\Omega_g = \Sigma_g^{-1}(\Sigma_g - S_g)\Sigma_g^{-1} \quad , \qquad g = 1,2,\ldots,m \ .$$

Then it follows from the corresponding results for a single population (see e.g., Lawley & Maxwell, 1963, Chapter 6 or Jöreskog, 1969) that for $g = 1,2,\ldots,m$,

(8a) $$\partial F/\partial \Lambda_g = n_g \Omega_g \Lambda_g \Phi_g \quad ,$$

(8b) $$\partial F/\partial \Phi_g = n_g \Lambda_g' \Omega_g \Lambda_g - \frac{1}{2} n_g \, \mathrm{diag}(\Lambda_g' \Omega_g \Lambda_g) \quad ,$$

(8c) $$\partial F/\partial \Psi_g = n_g \, \mathrm{diag}(\Omega_g \Psi_g) \quad .$$

We shall also need expressions for $\mathcal{E}(\partial^2 F/\partial\theta_i \partial\theta_j)$, where θ_i and θ_j are any two parameters. If θ_i is an element of Λ_g , Φ_g or Ψ_g and θ_j is an element of Λ_h , Φ_h or Ψ_h , $g \neq h$, $\mathcal{E}(\partial^2 F/\partial\theta_i \partial\theta_j)$ is zero.

Otherwise, if both θ_i and θ_j are elements of Λ_g, Φ_g or ψ_g, the required second-order derivatives can be expressed in terms of the elements of

(9a) $\xi = \Sigma^{-1}\Lambda$

(9b) $\eta = \Sigma^{-1}\Lambda\Phi = \xi\Phi$

(9c) $\alpha = \Lambda'\Sigma^{-1}\Lambda = \Lambda'\xi$

(9d) $\beta = \Phi\Lambda'\Sigma^{-1} = \Phi\alpha$

(9e) $\gamma = \Phi\Lambda'\Sigma^{-1}\Lambda\Phi = \beta\Phi$,

as

(10a) $\mathcal{E}(\partial^2 F/\partial\lambda_{ir}\partial\lambda_{js}) = n(\sigma^{ij}\gamma_{rs} + \eta_{is}\eta_{jr})$

(10b) $\mathcal{E}(\partial^2 F/\partial\lambda_{ir}\partial\phi_{st}) = (n/2)(2 - \delta_{st})(\xi_{is}\beta_{rt} + \xi_{it}\beta_{rs})$

(10c) $\mathcal{E}(\partial^2 F/\partial\lambda_{ir}\partial\psi_{jj}) = 2n\sigma^{ij}\eta_{jr}\psi_{jj}$

(10d) $\mathcal{E}(\partial^2 F/\partial\phi_{rs}\partial\phi_{tu}) = (n/4)(2 - \delta_{rs})(2 - \delta_{tu})(\alpha_{rt}\alpha_{su} + \alpha_{ru}\alpha_{st})$

(10e) $\mathcal{E}(\partial^2 F/\partial\phi_{rs}\partial\psi_{jj}) = n(2 - \delta_{rs})\xi_{jr}\xi_{js}\psi_{jj}$

(10f) $\mathcal{E}(\partial^2 F/\partial\psi_{ii}\partial\psi_{jj}) = 2n(\sigma^{ij})^2\psi_{ii}\psi_{jj}$.

Here we have omitted the subscript g for simplicity of notation.

The function F is regarded as a function of the elements of Λ_g, Φ_g, ψ_g, $g = 1,2,\ldots,m$, and is to be minimized with respect to these taking into account that some elements may be fixed and some may be constrained to be equal to others. Such a minimization problem may be solved as follows.

Let θ_g be a vector of all the elements in Λ_g, Φ_g and ψ_g arranged in a prescribed order. Since Φ_g is symmetric, only the elements in the lower half and the diagonal are counted. Then θ_g is of order $r_g = p_g k_g + \frac{1}{2} k_g(k_g + 1) + p_g$. Let $\theta' = (\theta_1', \theta_2', \ldots, \theta_m')$. Then θ consists of all the elements of all the parameter matrices and is of order $r = r_1 + r_2 + \ldots + r_m$. The function F may now be regarded as a function $f(\theta)$ of $\theta_1, \theta_2, \ldots, \theta_r$, which is continuous and has continuous derivatives $\partial F/\partial\theta_i$ and $\partial^2 F/\partial\theta_i \partial\theta_j$ of first and second order, except where any Σ_g is singular. The totality of these derivatives is represented by a gradient vector $\partial F/\partial\theta$ and a symmetric second order derivative matrix $\partial^2 F/\partial\theta\partial\theta'$.

The vector $\partial F/\partial\theta$ of order r is formed by arranging the elements of the derivative matrices (8a)-(8c) in the same order as the elements of Λ_g, Φ_g and ψ_g, $g = 1,2,\ldots,m$, in θ. As an approximation to the $r \times r$ matrix $\partial^2 F/\partial\theta\partial\theta'$ we use $\mathcal{E}(\partial^2 F/\partial\theta\partial\theta')$ which is of the form

$$(11) \qquad \mathcal{E}(\partial^2 F/\partial\theta\partial\theta') = \begin{bmatrix} \mathcal{E}(\partial^2 F/\partial\theta_1\partial\theta_1') & 0 & \cdots & 0 \\ 0 & \mathcal{E}(\partial^2 F/\partial\theta_2\partial\theta_2') & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \mathcal{E}(\partial^2 F/\partial\theta_m\partial\theta_m') \end{bmatrix}$$

where $\mathcal{E}(\partial^2 F/\partial\theta_g\partial\theta_g')$ is a symmetric matrix of order $r_g \times r_g$ formed by computing (10a)-(10f) and arranging these so that the order of rows and columns corresponds to the order of the parameters in θ_g.

Now let some $r - s$ of the θ's be fixed and denote the remaining θ's by $\pi_1, \pi_2, \ldots, \pi_s$, $s \leq r$. The function F is now regarded as a function $G(\pi)$ of $\pi_1, \pi_2, \ldots, \pi_s$. Derivatives $\partial G/\partial\pi$ and $\mathcal{E}(\partial^2 G/\partial\pi\partial\pi')$ are obtained from $\partial F/\partial\theta$ and $\mathcal{E}(\partial^2 F/\partial\theta\partial\theta')$ by omitting rows and columns corresponding to the fixed θ's. Among $\pi_1, \pi_2, \ldots, \pi_s$ let there be some t distinct and independent parameters denoted $\kappa_1, \kappa_2, \ldots, \kappa_t$, $t \leq s$, so that each π_i is equal to one and only one κ_j but possibly several π's equal the same κ. Let $K = (k_{ij})$ be a matrix of order $s \times t$ with elements $k_{ij} = 1$ if $\pi_i = \kappa_j$ and $k_{ij} = 0$ otherwise. The function G (or F) is now a function $H(\kappa)$ of the independent arguments $\kappa_1, \kappa_2, \ldots, \kappa_t$ and we have

$$\partial H/\partial\kappa = K'(\partial G/\partial\pi) \tag{12}$$

$$\mathcal{E}(\partial^2 H/\partial\kappa\partial\kappa') = K'\mathcal{E}(\partial^2 G/\partial\pi\partial\pi')K \quad . \tag{13}$$

Thus, the first-order and expected second-order derivatives of H are simple sums of the corresponding derivatives of G.

For the minimization of $H(\kappa)$ we use a modification of the method of Fletcher and Powell [1963] for which a computer program has been written by Gruvaeus and Jöreskog [1970]. This method makes use of a symmetric matrix E of order $t \times t$, which is evaluated in each iteration. Initially E is any positive definite matrix approximating the inverse of $\partial^2 H/\partial\kappa\partial\kappa'$. In subsequent iterations E is improved, using information built up about the function so that ultimately E converges to an approximation of the inverse of $\partial^2 H/\partial\kappa\partial\kappa'$ at the minimum. If t is large, the number of iterations may be excessive but can be considerably decreased by the provision of a good starting point for κ and a good initial estimate of E .

In principle, a good initial estimate of E may be obtained by computing $\partial^2 H/\partial\kappa\partial\kappa'$ at the starting point and then inverting this matrix. However, in our problem, the second-order derivatives are rather complicated and time-consuming to compute. Instead, we therefore use estimates of the second-order derivatives provided by the information matrix

$$(14) \qquad \mathcal{E}(\partial^2 H/\partial\kappa\partial\kappa') = \mathcal{E}(\partial H/\partial\kappa\ \partial H/\partial\kappa') \quad .$$

In addition to being more easily evaluated, this matrix also yields other valuable information. The inverse of $\mathcal{E}(\partial^2 H/\partial\kappa\partial\kappa')$ evaluated at the minimum minimum of H is an estimate of the variance-covariance of the estimated parameters $\hat{\kappa}_1, \hat{\kappa}_2, \ldots, \hat{\kappa}_t$. This may be used to obtain standard errors of the estimated parameters.

The starting point κ may be chosen arbitrarily but the closer it is to the final solution the fewer iterations will be required to find the solution. The minimization method converges quadratically from an arbitrary starting point to a local minimum of the function. If several local minima exist there is no guarantee that the method will converge to the absolute minimum.

2.5 Computer Program

A computer program, SIFASP, that performs all the computations described in the previous sections has been written in FORTRAN IV and a write-up for this is available [van Thillo & Jöreskog, 1970]. This program reads an observed covariance matrix or a correlation matrix and a vector of standard deviations for each group, a set of pattern matrices specifying the fixed, free and constrained parameters and a set of matrices of start values for the minimization. It then minimizes the function F as described in the previous section to obtain the maximum likelihood solution for each group. These are then printed together with residuals, i.e., differences between observed and reproduced variances and covariances, and a χ^2 measure of overall fit.

The computer program assumes that the number of variables and the number of common factors are the same for each group. This is no loss of generality, since it can always be achieved by the introduction of pseudovariables and pseudofactors in some groups as follows. Each pseudovariable has unit observed variance, zero observed covariances with every other variable, zero factor loadings on each factor including the pseudofactors and unit unique variance. Each pseudofactor has unit variance and zero covariance with every other factor and pseudofactor. It is readily verified that such pseudovariables and pseudofactors have no effect on the likelihood function whatsoever.

The observed variables may be rescaled initially as described in section 3.4. This is sometimes convenient when the observed variables have arbitrary units of measurements. In the special case of an invariant factor pattern, as described in the next section, the factors in the maximum likelihood solutions may be rescaled as shown in section 3.3.

The implementation of the minimization algorithm is simpler if all matrices are stored as singly subscripted arrays. This saves space, since only the lower halves of symmetric matrices need to be stored, and makes the program more efficient. The program makes use of a set of subroutines for matrix algebra on matrices stored as singly subscripted arrays. A further important advantage with this technique is the flexibility in the choice of m , p and k . Thus, in the same space as one can have $m = 4$, $p = 12$ and $k = 5$ one can also have $m = 2$, $p = 17$ and $k = 5$ or $m = 1$, $p = 24$ and $k = 12$.

The computer program works with one group ($m = 1$) as well as with more groups. When m is one the model is the same as that of Jöreskog [1969] but it is now possible to handle not only fixed parameters but also equality constraints between parameters. The new program SIFASP, therefore, makes the old program RMLFA obsolete. SIFASP can handle many types of factor analytic solutions.

3. A Model of Factorial Invariance

3.1 The Model

Perhaps the most common application of the method just described will be the case when the same tests have been administered in each population and when it is hypothesized that the factor pattern Λ is invariant over populations. Meredith [1964a] has shown that such a model will occur under certain conditions, when the populations are subpopulations derived from a parent population by selection on some external variables. Although this model is a special case of the general model described in the previous section, it deserves a separate discussion.

In this case $p_1 = p_2 = \cdots = p_m = p$ and $k_1 = k_2 = \cdots = k_m = k$ and the matrices Σ_g and Ψ_g , $g = 1,2,\ldots,m$ are all of the order $p \times p$ and the Φ_g , $g = 1,2,\ldots,m$ are all of order $k \times k$. The common factor pattern Λ is of order $p \times k$. The regression of x_g on f_g is [c.f. (1)].

(15) $$x_g = \mu_g + \Lambda f_g + z_g$$

and the variance-covariance matrix Σ_g is

(16) $$\Sigma_g = \Lambda\Phi_g\Lambda' + \psi_g^2 \quad .$$

In the special case of two populations, $m = 2$, a stricter form of invariance was considered by Lawley and Maxwell [1963, Chapter 8]. This requires not only the regression matrix Λ in (15) to be invariant but also the variances about the regression, i.e., $\psi_1^2 = \psi_2^2$. This type of restriction can easily be incorporated using the general approach of the preceding section.

3.2 Identification of Parameters

Suppose that the Λ in (16) is replaced by $\Lambda^* = \Lambda T^{-1}$ and each Φ_g is replaced by $\Phi_g^* = T\Phi_g T'$, $g = 1,2,\ldots,m$, where T is an arbitrary nonsingular matrix of order $k \times k$. Then each Σ_g remains the same so that the function F in (5) is unaltered. Since the matrix T has k^2 independent elements, this means that at least k^2 independent conditions must be imposed on the parameters in Λ, $\Phi_1, \Phi_2, \ldots, \Phi_m$ to make these uniquely defined.

Within the framework of the general procedure of the previous section, the most convenient way of doing this is to let all the Φ_g be free and to fix one nonzero element and at least $k - 1$ zeros in each column of Λ. In an exploratory study one can fix exactly $k - 1$ zeros in almost arbitrary positions. For example one may choose zero loadings where one thinks there should be "small" loadings in the factor pattern. The resulting solution may b· .otated further, if desired, to facilitate better interpretation. In a confirmatory study, on the other hand, the ·ositions of the fixed zeros, which often exceed $k - 1$ in each column, ·re given *a priori* by an hypothesis and the resulting solution cannot be rotated without destroying the fixed zeros.

3.3 Scaling of Factors

The fixed nonzero loading in each column of Λ can have any value. This is only used to fix a scale for each factor that is common to all groups. When the maximum likelihood solution has been obtained, the factors may be rescaled so that their average variance is unity. This rescaling is obtained as follows. Let

(17) $$\hat{\Phi} = (1/n) \sum_{g=1}^{m} n_g \hat{\Phi}_g \quad ,$$

with $n = \sum_{g=1}^{m} n_g$, and

(18) $$D = (\mathrm{diag}\ \hat{\Phi})^{-1/2} \quad .$$

Then the rescaled solution is

(19) $$\hat{\Lambda}^* = \hat{\Lambda} D^{-1}$$

(20) $$\hat{\Phi}^*_g = D \hat{\Phi}_g D \quad , \qquad g = 1,2,\ldots,m \ .$$

The matrix $\hat{\Lambda}^*$ has zeros wherever $\hat{\Lambda}$ has zeros but the fixed nonzeros in $\hat{\Lambda}$ have changed their values. The weighted average of the $\hat{\Phi}^*_g$ is a correlation matrix.

3.4 Scaling of Observed Variables

When the units of measurements in the different tests are arbitrary, it is usually convenient, though not necessary, to rescale the observed variables, before the factor analysis. Let

$$(21)\qquad S = (1/n) \sum_{g=1}^{m} n_g S_g \quad ,$$

with $n = \sum_{g=1}^{m} n_g$ as before and let

$$(22)\qquad D = (\mathrm{diag}\ S)^{-1/2} \quad .$$

Then the variance-covariance matrices for the rescaled variables are

$$(23)\qquad S_g^* = DS_gD \quad .$$

The weighted average of the S_g^* is a correlation matrix. The advantage of this rescaling is that, when combined with the rescaling of the factors of the previous section, the factor loadings are of the same order of magnitude as usual when correlation matrices are analyzed and when factors are standardized to unit variances. This makes it easier to choose start values for the minimization (see section 3.5) and interpret the results.

It should be pointed out that it is not permissible to standardize the variables in each group and to analyze the correlation matrices instead of the variance-covariance matrices. This violates the likelihood function (4) which is based on the distribution of the observed variances and covariances. Invariance of factor patterns is expected to hold only when the standardization of both tests and factors are relaxed.

3.5 Choice of Start Values

In a medium-sized study of say 4 groups, 20 variables and 5 factors, the number of free parameters to estimate may well exceed 200. To obtain the maximum likelihood estimates, a function of over 200 variables has to be minimized. This is not an easy task even on today's large computers. To reduce the computer time as much as possible it is necessary to choose good start values for the minimization. This can be done by doing some preliminary runs with the same computer program before the overall estimation is attempted.

1. Using $m = 1$ and the pooled correlation matrix $R = DSD$, where D is given by (22), obtain an oblique maximum likelihood solution with the fixed zeros in Λ and the diagonal elements of Φ equal to unity. Let the estimate of Λ so obtained be denoted $\hat{\Lambda}^{(0)}$.
2. For each group <u>separately</u>, using $m = 1$ and S_g^*, obtain an oblique maximum likelihood solution with the <u>whole</u> Λ fixed equal to $\hat{\Lambda}^{(0)}$ and with Φ_g and ψ_g free. Let the resulting estimates be denoted $\hat{\Phi}_g^{(0)}$ and $\hat{\psi}_g^{(0)}$, $g = 1,2,\ldots,m$.
3. Then $\hat{\Lambda}^{(0)}$, $\hat{\Phi}_1^{(0)},\hat{\Phi}_2^{(0)},\ldots,\hat{\Phi}_m^{(0)}$, $\hat{\psi}_1^{(0)},\hat{\psi}_2^{(0)},\ldots,\hat{\psi}_m^{(0)}$ provide good start values for the overall minimization with the largest element in each column of $\hat{\Lambda}^{(0)}$ fixed, in addition to the fixed zeros.

If the model also specifies that the ψ_g should be invariant over groups, one uses as start values $\hat{\psi}^{(0)}$ for the common ψ, the estimate obtained in step 1 and step 2 is done with ψ fixed at $\hat{\psi}^{(0)}$.

3.6 Testing of Hypotheses and Strategy of Analysis

Suppose H_0 and H_1 represent two models under different specifications of fixed, free and constrained parameters, both models fitting the general framework of section 2.1. Then it is possible, in large samples, to test the model H_0 against the model H_1, by estimating each of them separately and comparing their χ^2 goodness of fit values. The difference in χ^2 is asymptotically a χ^2 with degrees of freedom equal to the corresponding difference in degrees of freedom.

In an exploratory study there are various hypotheses that may be tested and it seems best to proceed stepwise in a certain order.

One begins by testing the hypothesis of equality of covariance matrices, i.e.,

$$(24) \qquad H_\Sigma : \Sigma_1 = \Sigma_2 = \dots = \Sigma_m \quad .$$

This may be tested by using the test statistic

$$(25) \qquad M = n \log|S| - \sum_{g=1}^{m} n_g \log|S_g| \quad ,$$

where S is given by (21). Under the hypothesis, M is distributed approximately as χ^2 with $d_\Sigma = \frac{1}{2}(m - 1)p(p + 1)$ degrees of freedom. As shown by Box [1949], the approximation to the χ^2 distribution is improved if M is multiplied by a certain constant. When p or m is larger than 4, Box suggests a transformation to an F distribution.

It should be noted that the test statistic (25) may be obtained in SIFASP by specifying $k_g = p$, $\Lambda_g = I$, $\psi_g = 0$, $g = 1,2,\ldots,m$ and $\Phi_1 = \Phi_2 = \cdots = \Phi_m$. The maximum likelihood estimate of the common Φ will then be the pooled S as defined in (21). If the tests are scaled originally as described in section 3.4, this S is a correlation matrix R .

If the hypothesis is found to be tenable every characteristic common to all groups can be obtained from the pooled covariance matrix S or the correlation matrix R and there is no need to analyze each group separately or simultaneously.

If, on the other hand, the hypothesis of equality of covariance matrices is untenable, one may want to investigate similarities and difference in factor structures. For this purpose, a sequence of hypotheses, such that each hypothesis is a special case of the preceding, will now be considered. The first hypothesis is <u>the hypothesis of equality of number of common factors</u>, i.e.,

(26) $$H_k: \; k_1 = k_2 = \cdots = k_m = \text{a specified number } k \quad .$$

This may be tested by doing an unrestricted factor analysis [Jöreskog, 1969] on each S_g (or S_g^* or the corresponding correlation matrix) separately, using the same number of common factors for each group. The analyses may be done by Jöreskog's [1967a,b] method UMLFA but can also be done with the computer program SIFASP. In SIFASP one uses $m = 1$ and fixes k^2 elements in Λ_g and/or Φ_g ; for example, to obtain an orthogonal solution one can choose $\Phi_g = I$ and $\frac{1}{2}k(k-1)$ zeros in Λ_g and to obtain an oblique solution one can choose $\operatorname{diag}\Phi_g = I$ and $k(k-1)$ zeros in Λ_g. Each analysis gives a χ^2 with $\frac{1}{2}[(p-k)^2-(p+k)]$ degrees of freedom. Since these χ^2's are independent they may be added for each group to obtain a χ^2, χ^2_k say, with $d_k = \frac{1}{2}m[(p-k)^2-(p+k)]$ degrees of freedom, which may be used to test the overall hypothesis.

If the hypothesis of a common number of factors is found tenable, one may proceed to test <u>the hypothesis of an invariant factor pattern</u>, i.e.,

$$H_\Lambda\text{:}\quad \Lambda_1 = \Lambda_2 = \cdots = \Lambda_m \quad . \tag{27}$$

The common factor pattern Λ may either be completely unspecified or be specified to have zeros in certain positions. If Λ is unspecified, one fixes $k-1$ zeros and one nonzero value in each column almost arbitrarily. If Λ is specified to have zeros in certain positions, one fixes an arbitrary nonzero element in each column in addition. There will then be k^2 fixed elements in Λ in the unspecified case and $q \geq k^2$ in the specified case. To obtain a χ^2 for this hypothesis, one estimates Λ, $\Phi_1,\Phi_2,\ldots,\Phi_m$, $\Psi_1,\Psi_2,\ldots,\Psi_m$ from $S_1,S_2,\ldots,S_m$ simultaneously, yielding a minimum value of the function F. Twice this minimum value is a χ^2, χ^2_Λ say, with degrees of freedom

$$\frac{1}{2}mp(p+1) - pk + q - \frac{1}{2}mk(k+1) - mp \quad ,$$

where $q = k^2$ in the unspecified case. To test the hypothesis H_Λ, given that H_k holds, one uses $\chi^2_{\Lambda\cdot k} = \chi^2_\Lambda - \chi^2_k$ with $d_{\Lambda\cdot k} = d_\Lambda - d_k$ degrees of freedom.

If this hypothesis is found tenable one may proceed to test the hypothesis

(28) $\qquad H_{\Lambda\psi}: \ \Lambda_1 = \Lambda_2 = \cdots = \Lambda_m \ ; \ \ \psi_1 = \psi_2 = \cdots = \psi_m \quad .$

To do so one has to estimate Λ, $\Phi_1, \Phi_2, \ldots, \Phi_m$, ψ under $H_{\Lambda\psi}$. This again gives a minimum value of F which when multiplied by two gives $\chi^2_{\Lambda\psi}$ with

$$d_{\Lambda\psi} = \frac{1}{2} mp(p + 1) - pk + q - \frac{1}{2} mk(k + 1) - p$$

degrees of freedom. To test $H_{\Lambda\psi}$ against H_Λ one uses $\chi^2_{\psi \cdot \Lambda} = \chi^2_{\Lambda\psi} - \chi^2_\Lambda$ with $d_{\psi \cdot \Lambda} = d_{\Lambda\psi} - d_\Lambda$ degrees of freedom.

If the hypothesis $H_{\Lambda\psi}$ is found tenable one may want to test the hypothesis

(29) $\qquad H_{\Lambda\Phi\psi}: \ \Lambda_1 = \Lambda_2 = \cdots = \Lambda_m \ ; \ \ \Phi_1 = \Phi_2 = \cdots = \Phi_m \ ; \ \ \psi_1 = \psi_2 = \cdots = \psi_m \ .$

This hypothesis is included in H_Σ but is stronger than H_Σ since H_Σ includes also the cases when the common Σ is not of the form

(30) $\qquad \Sigma = \Lambda\Phi\Lambda' + \psi^2 \quad .$

This hypothesis $H_{\Lambda\Phi\psi}$ can be tested directly on the basis of the pooled S in (21). The test of $H_{\Lambda\Phi\psi}$ against H_Σ uses a χ^2 with

$$d_{\Lambda\Phi\psi \cdot \Sigma} = \frac{1}{2} p(p + 1) - pk + q - \frac{1}{2} k(k + 1) - p$$

degrees of freedom.

Various other types of hypotheses may also be tested. For example, one may assume that some factors are orthogonal and some are oblique (see Jöreskog, 1969).

It should be emphasized that even if a χ^2 is significant, there may still be reasons to consider the model. After all, the basic model with its assumptions of linearity and normality is only regarded as an approximation to reality. The true population covariance matrix will not in general be exactly of the form specified by the hypothesis, but there will be discrepancies between the true population covariance matrix and the formal model postulated. These discrepancies will not get smaller when the sample size increases but will tend to give large χ^2 values. Therefore, a model may well be accepted even though χ^2 is large. Whether to accept or reject a model cannot be decided on a purely statistical basis. This is largely a matter of the experimenter's interpretations of the data, based on substantive theoretical and conceptual considerations. Ultimately the criteria for goodness of the model depends on the usefulness of it and the results it produces.

3.7 A Numerical Illustration

To illustrate the methods previously discussed we use the same data as Meredith [1964b] used to illustrate his rotational procedure. The data consist of nine tests selected from a battery of 26 psychological tests described by Holzinger and Swineford [1939]. The tests were administered to 7th and 8th grade children in two schools, the Pasteur and the Grant-White Schools in the Chicago area. The nine tests were selected so that each of the three factors--space, verbal and memory--would be represented by three tests. The nine tests used, with their original code numbers in parentheses, were: Visual Perception (1), Cubes (2), Paper Form Board (3), General Information (5), Sentence Completion (7), Word Classification (8), Figure

Recognition (16), Object Number (17) and Number-Figure (18). On the basis of a speeded addition test, Meredith divided each of the samples from the two schools into two approximately equal groups by splitting at the median score within each school. This yielded four groups that will be used for this illustration. The correlation matrices taken from Meredith's table 2 are shown in Table 1a with unscaled and scaled standard deviations in Table 1b. The sample sizes are: Group 1: Pasteur Low $N_1 = 77$, Group 2: Pasteur High $N_2 = 79$, Group 3: Grant-White Low $N_3 = 74$ and Group 4: Grant-White High $N_4 = 71$. Because of the way the two groups within schools were selected, it is doubtful that the assumption of multinormality is valid. This departure from multinormality will have no great effect on the estimates but may be more serious for the χ^2 values. In particular, the χ^2 test of H_Σ is known to be sensitive to departures from multinormality. For this reason and also because the sample sizes are relatively small, the χ^2 values that will be reported should be interpreted very cautiously. It should be emphasized that these data have been chosen merely to illustrate the procedures of this paper. Another application, with more substantive interest and with larger and widely varying sample sizes, are given by McGaw and Jöreskog [1970].

We begin by testing the hypothesis H_Σ that $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4$. This gives the test statistic $M = 146.95$ with 135 degrees of freedom. Transformation to Box's F -statistic gives $F_{135,\infty} = 1.03$. In view of the remark just made, this value is inconclusive. However, for the purpose of illustrating a simultaneous analysis of all four scaled dispersion matrices, we shall follow the procedure of section 3.6 and test various hypotheses of interest. The results are summarized in Table 2.

The first hypothesis, $H_{k=3}$, is that three factors adequately reproduce the correlation in each population. This gives $\chi^2 = 47.73$ with 48 degrees of freedom. This χ^2 is the sum of four χ^2's one from each population and each with 12 degrees of freedom. These are $\chi_1^2 = 15.33$, $\chi_2^2 = 10.44$, $\chi_3^2 = 14.40$ and $\chi_4^2 = 7.56$. Thus we cannot reject the hypothesis that the number of factors is three for each population. We therefore proceed by investigating whether there is an invariant factor pattern or not.

The next hypothesis, H_{Λ_u} , is that there is an invariant unspecified (unrestricted) factor pattern Λ_u . To test this hypothesis we fix one nonzero element and two zero elements in each column of Λ_u and leave Φ_1 , Φ_2 , Φ_3 , Φ_4 and ψ_1 , ψ_2 , ψ_3 , ψ_4 completely unconstrained. A convenient way to choose the fixed elements in Λ_u is to use a reference variables solution as, for example,

$$(31) \qquad \Lambda_u = \begin{bmatrix} 1 & 0 & 0 \\ x & x & x \\ x & x & x \\ 0 & 1 & 0 \\ x & x & x \\ x & x & x \\ 0 & 0 & 1 \\ x & x & x \\ x & x & x \end{bmatrix} .$$

Here the zeros and ones stand for fixed values and x's for parameters to be estimated. Tests 1, 4 and 7 have been chosen to be pure in their respective factors. The test of H_{Λ_u} gives $\chi^2 = 90.57$ with 102 degrees of freedom, which is not significant. Thus, we cannot reject the hypothesis that there is an invariant factor pattern with three factors. To strengthen the model we now make use of our knowledge about the tests and hypothesize that the invariant factor pattern has a specific form, namely

$$(32) \qquad \Lambda = \begin{bmatrix} 1 & 0 & 0 \\ x & 0 & 0 \\ x & 0 & 0 \\ 0 & 1 & 0 \\ 0 & x & 0 \\ 0 & x & 0 \\ 0 & 0 & 1 \\ 0 & 0 & x \\ 0 & 0 & x \end{bmatrix},$$

i.e., we assume that Λ has a nonoverlapping group structure, where the first three tests are loaded on the first factor only, the next three tests on the second factor only and the last three tests on the third factor only. As before, we put no constraints on the Φ's and the ψ's. A test of this hypothesis gives $\chi^2 = 131.24$ with 114 degrees of freedom. This has a probability level of about 0.13. Thus we cannot reject the hypothesis that the invariant factor pattern is of the specified form. An examination of the ψ's, in relation to their standard errors in the solution under H_Λ, revealed that many of these were not sufficiently different to be considered different. This suggests that one should also examine the hypothesis $H_{\Lambda\psi}$, that a stricter form of invariance holds, namely where also the ψ's are the same for all populations. A test of $H_{\Lambda\psi}$ gives $\chi^2 = 172.14$ with 141 degrees of freedom. This is just significant at the 5% level. The maximum likelihood solution under $H_{\Lambda\psi}$ is shown in Table 3. Finally, to complete the sequence of hypotheses we consider the hypothesis $H_{\Lambda\Phi\psi}$, that the whole factor structure is invariant, with the same factor pattern Λ as before. This gives $\chi^2 = 212.80$ with 159 degrees of freedom which is highly significant. This would seem to contradict the test of H_Σ. This is not so, however, since $H_{\Lambda\Phi\psi}$ is a much more restrictive hypothesis than H_Σ. The hypothesis $H_{\Lambda\Phi\psi}$ requires that the common Σ has a factor structure with three common factors with a factor pattern of the restricted type (32), but there may be many other possible representations of the common Σ. In fact, if Σ is represented by the unrestricted factor pattern Λ_u in (31) instead, one obtains $\chi^2 = 36.2$[illegible] with 147 degrees of freedom, so that $H_{\Lambda_u\Phi\psi}$ cannot be rejected although $H_{\Lambda\Phi\psi}$ was rejected.

Altogether these results suggest two alternative descriptions of the data. One is that the whole factor structure is invariant over populations with a three-factor solution of a fairly complex form. The other is to represent the tests in each population by three factors of a particularly simple form, but these factors have different variance-covariance matrices in the different populations. Additional studies with larger sample sizes are needed to discriminate statistically between the two models. Perhaps, the second alternative has the most intuitive appeal. Inspecting the factor variances in Table 4, it is seen that for the Pasteur school they tend to be higher for the Low group than for the High group, whereas for the Grant-White school generally the opposite holds. Also for the two High groups the variances are generally lower for the Pasteur school than for the Grant-White school. Note also the low covariance of 0.08 between S and M for the Low Grant-White group and the corresponding high covariance 1.03 for the Low Pasteur group. This seems to indicate that the Low Pasteur group cannot fully discriminate between the spacial and the memory tasks whereas the Low Grant-White group can do so clearly.

4. Estimation of Factor Means

A stricter form of invariance than (15) is obtained if in (1) we require that not only the factor matrix Λ_g but also the vector μ_g be invariant over populations. Then

$$x_g = \mu + \Lambda f_g + z_g \quad . \tag{33}$$

In this case it is not reasonable to assume, as before, that $\mathcal{E}(f_g) = 0$ for all g, since this implies that $\mathcal{E}(x_g) = \mu$ for all g. Instead we assume that each f_g has its own mean $\mathcal{E}(f_g) = \nu_g$ and propose to estimate these ν_g, $g = 1,2,\ldots,m$. However, since each f_g may be replaced by $f_g + b$ if μ is replaced by $\mu - \Lambda b$, without changing x_g in (33), some rule is necessary for fixing the origin of the f_g. It seems most convenient to fix the origin such that

$$\sum_{g=1}^{m} N_g \nu_g = 0 \quad . \tag{34}$$

To estimate μ and ν_g we assume that Λ, Φ_g, Ψ_g, and hence also Σ_g, are known and equal to their estimates $\hat{\Lambda}$, $\hat{\Phi}_g$, $\hat{\Psi}_g$ and $\hat{\Sigma}_g$. Let $\bar{x}_g$, as before, be the sample mean vector in group g and let $\bar{\bar{x}}$ be the overall mean, i.e.,

$$\bar{\bar{x}} = (1/N) \sum_{g=1}^{m} N_g \bar{x}_g \tag{35}$$

with $N = \sum_{g=1}^{m} N_g$. We take $\bar{\bar{x}}$ to be an estimator of μ. Then each of the following three estimators of ν_g seem reasonable:

(36) $\hat{\nu}_g = (\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\Lambda}'(\bar{x}_g - \bar{\bar{x}})$

(37) $\hat{\nu}_g = \hat{\Phi}_g\hat{\Lambda}'\hat{\Sigma}_g^{-1}(\bar{x}_g - \bar{\bar{x}})$

(38) $\hat{\nu}_g = (\hat{\Lambda}'\hat{\Sigma}_g^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Sigma}_g^{-1}(\bar{x}_g - \bar{\bar{x}})$.

Formula (36) is obtained by fitting the theoretical means $\mu + \hat{\Lambda}\nu_g$ to the observed means $\bar{x}_g$. $g = 1,2,\ldots,m$, by least squares. The advantage of this formula is that the weighting matrix in front of $\bar{x}_g - \bar{\bar{x}}$ is independent of g .

Formula (37) is obtained if one applies the mean vectors to the regression formula for correlated factor scores.

Formula (38) is the maximum likelihood estimator of ν_g for given $\Lambda = \hat{\Lambda}$, $\Sigma_g = \hat{\Sigma}_g$, $g = 1,2,\ldots,m$ and $\mu = \bar{\bar{x}}$. The latter is obtained from the minimization of

$$\sum_{g=1}^{m} \sum_{\alpha=1}^{N_g} (x_{\alpha g} - \bar{\bar{x}} - \hat{\Lambda}\nu_g)'\hat{\Sigma}_g^{-1}(x_{\alpha g} - \bar{\bar{x}} - \hat{\Lambda}\nu_g) \quad ,$$

where $x_{\alpha g}$ is the vector of observed test scores for person α in group g .

Formula (36) satisfies (34) for the estimates, but (37) and (38) have to be scaled afterwards so that (34) holds.

References

Box, G. E. P. A general distribution theory for a class of likelihood criteria. *Biometrika*, 1949, *36*, 317-346.

Fletcher, R. & Powell, M. J. D. A rapidly convergent descent method for minimization. *The Computer Journal*, 1963, *6*, 163-168.

Gruvaeus, G. T. & Jöreskog, K. G. A computer program for minimizing a function of several variables. Research Bulletin 70-14. Princeton, N. J.: Educational Testing Service, 1970.

Holzinger, K. J. & Swineford, F. A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monograph No. 48*. Chicago: University of Chicago, 1939.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, *34*, 183-202.

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

McGaw, B. & Jöreskog, K. G. Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. Research Bulletin 70-63. Princeton, N. J.: Educational Testing Service, 1970.

Meredith, W. Rotation to achieve factorial invariance. *Psychometrika*, 1964, *19*, 187-206. (a)

Meredith, W. Notes on factorial invariance. *Psychometrika*, 1964, *29*, 177-185. (b)

van Thillo, M. & Jöreskog, K. G. A general computer program for simultaneous factor analysis in several populations. Research Bulletin 70-62. Princeton, N. J.: Educational Testing Service, 1970.

TABLE 1a-b

Intercorrelation Matrices (a)

Group 1 Above Main Diagonal
Group 2 Below Main Diagonal

	1	2	3	4	5	6	7	8	9
Visual Perception	--	.32	.48	.28	.26	.40	.42	.12	.23
Cubes	.24	--	.33	.01	.01	.26	.32	.05	-.04
Paper Form Board	.23	.22	--	.06	.01	.10	.22	.03	.01
General Information	.32	.05	.23	--	.75	.60	.15	-.08	-.05
Sentence Completion	.35	.23	.18	.68	--	.63	.07	.06	.10
Word Classification	.36	.10	.11	.59	.66	--	.36	.19	.24
Figure Recognition	.22	.01	-.07	.09	.11	.12	--	.29	.19
Object-Number	-.02	-.01	-.13	.05	.08	.03	.19	--	.38
Number-Figure	.09	-.14	-.06	.16	.02	.12	.15	.29	--

Group 3 Above Main Diagonal
Group 4 Below Main Diagonal

	1	2	3	4	5	6	7	8	9
Visual Perception	--	.34	.41	.38	.40	.42	.35	.16	.35
Cubes	.32	--	.21	.32	.16	.13	.27	.01	.27
Paper Form Board	.34	.18	--	.31	.24	.35	.30	.09	.09
General Information	.31	.24	.31	--	.69	.55	.17	.31	.34
Sentence Completion	.22	.16	.29	.62	--	.65	.20	.30	.27
Word Classification	.27	.20	.32	.57	.61	--	.31	.34	.27
Figure Recognition	.48	.31	.32	.18	.20	.29	--	.31	.38
Object-Number	.20	.01	.15	.06	.19	.15	.36	--	.38
Number-Figure	.42	.28	.40	.11	.07	.18	.35	.44	--

Standard Deviations (b)

	Unscaled				Scaled			
	1	2	3	4	1	2	3	4
Visual Perception	7.4	6.7	6.6	7.2	1.06	0.96	0.95	1.03
Cubes	5.6	4.0	4.8	4.0	1.20	0.86	1.03	0.86
Paper Form Board	2.9	2.8	2.6	3.0	1.02	0.99	0.92	1.06
General Information	11.8	11.0	11.3	11.5	1.03	0.96	0.99	1.01
Sentence Completion	5.2	5.2	4.7	4.5	1.08	1.06	0.96	0.91
Word Classification	5.2	5.3	5.0	5.5	0.99	1.01	0.95	1.05
Figure Recognition	8.8	7.6	6.1	7.4	1.17	1.01	0.81	0.98
Object-Number	4.7	5.2	3.9	4.9	1.00	1.10	0.83	1.04
Number-Figure	4.6	4.4	3.9	4.7	1.04	1.00	0.88	1.07

TABLE 2

Summary of Analyses

Hypothesis	χ^2	No. par.	d.f.	P
H_Σ	146.95	45	135	0.23
$H_{k=3}$	47.73	132	48	0.47
H_Λ	90.57	78	102	0.78
$H_\Lambda{}^u$	131.24	66	114	0.13
$H_{\Lambda\psi}$	172.14	39	141	0.04
$H_{\Lambda\Phi\psi}$	212.80	21	159	0.00
$H_{\Lambda_u\Phi\psi}$	36.20	33	147	1.00

TABLE 3

Maximum Likelihood Solution under $H_{\Lambda\psi}$
(Asterisks Denote Parameter Values Specified by Hypothesis)

	$\hat{\Lambda}$ S	V	M	$\hat{\psi}$
Visual Perception	.72	0*	0*	.69
Cubes	.43	0*	0*	.90
Paper Form Board	.51	0*	0*	.86
General Information	0*	.80	0*	.60
Sentence Completion	0*	.85	0*	.33
Word Classification	0*	.75	0*	.67
Figure Recognition	0*	0*	.58	.81
Object-Number	0*	0*	.48	.88
Number-Figure	0*	0*	.57	.83

$$\hat{\Phi}_1 = \begin{array}{c c} & \begin{array}{ccc} S & V & M \end{array} \\ \begin{array}{c} S \\ V \\ M \end{array} & \begin{bmatrix} 1.02 & & \\ 0.53 & 0.91 & \\ 1.03 & 0.36 & 1.30 \end{bmatrix} \end{array} \qquad \hat{\Phi}_2 = \begin{array}{c c} & \begin{array}{ccc} S & V & M \end{array} \\ \begin{array}{c} S \\ V \\ M \end{array} & \begin{bmatrix} 0.89 & & \\ 0.62 & 0.93 & \\ 0.59 & 0.50 & 0.58 \end{bmatrix} \end{array}$$

$$\hat{\Phi}_3 = \begin{array}{c c} & \begin{array}{ccc} S & V & M \end{array} \\ \begin{array}{c} S \\ V \\ M \end{array} & \begin{bmatrix} 0.72 & & \\ 0.52 & 1.06 & \\ 0.08 & 0.20 & 0.90 \end{bmatrix} \end{array} \qquad \hat{\Phi}_4 = \begin{array}{c c} & \begin{array}{ccc} S & V & M \end{array} \\ \begin{array}{c} S \\ V \\ M \end{array} & \begin{bmatrix} 1.38 & & \\ 0.42 & 1.12 & \\ 0.71 & 0.27 & 1.25 \end{bmatrix} \end{array}$$