DOCUMENT RESUME

ED 268 134 TM 850 735

AUTHOR Mislevy, Robert J.
TITLE Recent Developments in the Factor Analysis of Categorical Variables. Research Report.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-85-24
PUB DATE Jul 85
NOTE 70p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Correlation; *Estimation (Mathematics); *Factor Analysis; *Factor Structure; Latent Trait Theory; *Least Squares Statistics; *Mathematical Models; *Maximum Likelihood Statistics; Statistical Studies
IDENTIFIERS *Categorical Data

ABSTRACT

This paper reviews recent work in factor analysis of categorical variables. Emphasis is on the generalized least squares solution. A section on maximum likelihood solution focuses on extensions of the classical model, especially the normal case. Many of the recent developments have taken place within this context, and it provides a unified framework of exposition against which other models may be introduced in contrast. Section 2 provides a brief review of factor analysis of measured variables, setting up notation and formulas in this more familiar context. Section 3 introduces the common factor model for dichotomous items. Sections 4 and 5 discuss estimation of factor loadings from matrices of tetrachoric correlations, unweighted and weighted respectively. Section 6 discusses a full information solution based on the method of maximum likelihood. Finally, section 7 outlines a number of extensions to the basic model under investigation. These include Bayesian prior distributions on unique variances, confirmatory factor analysis, comparisons of factor structures between groups, and relaxation of assumptions about response functions and population distributions. Eight pages of references are included. (PN)

RR-85-24

RESEARCH REPORT

RECENT DEVELOPMENTS IN THE FACTOR ANALYSIS OF CATEGORICAL VARIABLES

Robert J. Mislevy

Educational Testing Service
Princeton, New Jersey
July 1985

Recent Developments in the Factor Analysis of Categorical Variables

Robert J. Mislevy

Educational Testing Service

Running Head: Developments in Factor Analysis

Abstract

Despite known shortcomings of the procedure, exploratory factor analysis of dichotomous test items has been limited, until recently, to unweighted analyses of matrices of tetrachoric correlations. Superior methods have begun to appear in the literature, in professional symposia, and in computer programs. This paper places these developments in a unified framework, from a review of the classical common factor model for measured variables through generalized least squares and marginal maximum likelihod solutions for dichotomous data. Further extensions of the model are also reported as work in progress.

Key words: binary variables, categorical data, contingency tables, covariance structures, factor analysis, item response theory, latent structure, tetrachoric correlations

Recent Developments in the Factor Analysis of Categorical Variables

1. Introduction

Under classical Thurstonian factor analysis (Thurstone, 1947), values of p measured variables are modeled as linear functions of some smaller number of m continuous latent variables, the "factors" that account for the correlations among the observed variables. The usual objectives in factor analysis are to determine the number of factors that provide a satisfactory fit to the observed correlation matrix and to estimate the regression coefficients of the observed variables on the factors—all this, it is hoped, leading to a more parsimonious and meaningful explication of the patterns of interrelationship among the observed variables.

Recent interest in item response theoretical (IRT) methods of constructing and scoring tests (see, for example, Hambleton & Cook, 1977; Lord, 1980; Wright & Stone, 1979) has led to a renewed interest in the extension of classical factor analysis to dichotomous test items. In the extension, the measured variables of the classical formulation now play the role of latent response processes to each of the items; a correct response is observed only when the response process variable arising in the confrontation of a given examinee with a given item exceeds a latent threshold characterizing the item. (Modifications will

also be introduced to account for the possibility of random correct responses, as can occur when the test directions encourage examinees to guess on multiple-choice items.) While it is certain that not only the unidimensional models posited in most applications of IRT but the multidimensional models of factor analysis are strictly incorrect in any given application, a number of benefits may accrue nonetheless. It is not unreasonable to summarize into a single score, responses to a set of items fairly well explained by a single dominant factor, for example; but the appearance of clusters of items separating clearly into multiple factors suggests a need to consider reporting separate subtests scores.[1]

Early work along these lines proceeded by first obtaining the matrix of tetrachoric correlations among the test responses, an approximation of the correlation matrix among the latent response proceses among the various items under the assumption that they follow a multivariate normal distribution. Those attempts ran into difficulties, due to the occasional values of +1 and -1 that result, the fact that matrices of sample tetrachorics are not necessarily positive definite, the lack of statistical tests for the number of factors, and the failure to account for the chance successes that occur with multiple-choice items.

This paper reviews some recent work in factor analysis of categorical variables. Emphasis is on the generalized least

squares (GLS) solution developed by Christoffersson (1975) and Muthén (1978) and the maximum likelihood approach introduced by Bock and Aitkin (1981). The section on maximum likelihood solution and its extensions draws upon recent work reported in a symposium at the 1984 meeting of the Psychometric Society, including papers by Bock (1984), Gibbons (1984a), Muraki (1984), and Muthén (1984a). We focus for the most part on extensions of the classical model, especially the normal case, for convenience of presentation. Many of the recent developments have taken place within this context, and it provides a unified framework of exposition against which other models may be introduced in contrast.

Section 2 provides a brief review of factor analysis of measured variables, setting up notation and formulas in this more familiar context. Section 3 introduces the common factor model for dichotomous items. Sections 4 and 5 discuss estimation of factor loadings from matrices of tetrachoric correlations, unweighted (ULS) and weighted (GLS) respectively. Section 6 discusses a full information solution based on the method of maximum likelihood. Finally, section 7 outlines a number of extensions to the basic model currently under investigation. These include Bayesian priors on unique variances, confirmatory factor analysis, comparisons of factor structures between groups,

and relaxation of assumptions about response functions and popualtion distributions.

2. Factor Analysis of Measured Variables

Factor analysis, at its heart, is a method of data explanation through model-fitting. The matrix of covariances or correlations among a large number of variables $\underset{\sim}{y} = (y_1,\ldots,y_p)$ is the object of analysis; it is hypothesized that the interrelationships among the variables can be accounted for by a linear multiple regression model, with the y's as dependent variables. The distinguishing feature of factor analysis is that the predictors, $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)$ are not observed but must be inferred from the data. In this section, we review the basic models and procedures associated with factor analysis of measured variables. (For readable introductions to the concepts of factor analysis, see Harman (1976), Jöreskog (1979), and Lawley & Maxwell (1971).)

2.1 The Common Factor Model

The classical factor analysis model for measured variables assumes an m-dimensional latent variable $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)$ in a population of examinees. Without loss of generality, $\underset{\sim}{\theta}$ is assumed to have mean $\underset{\sim}{0}$. Observations on a random sample of N examinees, however, consist not of values of $\underset{\sim}{\theta}$ but of values of p manifest variables $\underset{\sim}{y} = (y_1,\ldots,y_p)$, where $p > m$. It is assumed that $\underset{\sim}{y}$ depends stochastically upon $\underset{\sim}{\theta}$ through the following system of linear equations:

$$y_1 = \lambda_{11}\theta_1 + \ldots + \lambda_{1m}\theta_m + e_1$$

$$y_2 = \lambda_{21}\theta_1 + \ldots + \lambda_{2m}\theta_m + e_2$$

$$\vdots$$

$$y_p = \lambda_{p1}\theta_1 + \ldots + \lambda_{pm}\theta_m + e_p$$

or, in matrix form,

$$\underset{\sim}{y} = \underset{\sim\sim}{\Lambda}\underset{\sim}{\theta} + \underset{\sim}{e} \quad . \tag{2.1}$$

$\underset{\sim}{\Lambda}$ is typically referred to as the matrix of factor loadings. Let $\underset{\sim}{\Phi}$ represent the covariance matrix of $\underset{\sim}{\theta}$ and let $\underset{\sim}{\Psi}$ represent the covariance matrix of the residuals $\underset{\sim}{e}$. The covariance matrix $\underset{\sim}{\Sigma}$ of $\underset{\sim}{y}$ is then given by

$$\underset{\sim}{\Sigma} = \underset{\sim\sim\sim}{\Lambda\Phi\Lambda'} + \underset{\sim}{\Psi} \quad .$$

Under the Thurstonian model, the residuals are assumed to be uncorrelated, and the factor loadings and the factor covariance matrix account entirely for the linear relationships among the manifest variables. The elements of the diagonal matrix $\underset{\sim}{\Psi}$ are typically referred to as the unique variances of the y's.

After incorporating constraints necessary to make the model identified (see Section 2.2), it is possible to fit a given $\underset{\sim}{\Sigma}$ with respect to $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ without additional assumptions about the distributions of $\underset{\sim}{y}$, $\underset{\sim}{\theta}$, or $\underset{\sim}{e}$ (see for example, Harman (1976) and Thurstone (1947)). In order to facilitate the transition to the discrete case, however, we shall introduce some distributional assumptions and restrict consideration to statistical estimation procedures. Suppose the residuals in Equation 2.1 are also assumed to follow a multivariate normal distribution; $\underset{\sim}{e} \sim \text{MVN}(\underset{\sim}{0},\underset{\sim}{\Psi})$. The The distribution of $\underset{\sim}{y}$, conditional on $\underset{\sim}{\theta}$, or for a specified examinee with $\underset{\sim}{\theta} = \underset{\sim}{\theta}_i$, may be inferred as

$$(\underset{\sim}{y}|\underset{\sim}{\theta}_i,\underset{\sim}{\Lambda},\underset{\sim}{\Psi}) \sim \text{MVN}(\underset{\sim}{\Lambda}\underset{\sim}{\theta}_i,\underset{\sim}{\Psi}) \quad . \tag{2.2}$$

This is the conditional distribution of y.

Assuming further that $\underset{\sim}{\theta} \sim \text{MVN}(\underset{\sim}{0},\underset{\sim}{\Phi})$, we may derive the marginal distribution of $\underset{\sim}{y}$, or the distribution of $\underset{\sim}{y}$ from an examinee selected at random, by integrating Equation 2.2 over the examinee population:

$$p(\underset{\sim}{y}|\underset{\sim}{\Lambda},\underset{\sim}{\Psi},\underset{\sim}{\Phi}) = \int_{\underset{\sim}{\theta}} p(\underset{\sim}{y}|\underset{\sim}{\theta},\underset{\sim}{\Lambda},\underset{\sim}{\Psi})\, p(\underset{\sim}{\theta}|\underset{\sim}{\Phi})\, d\underset{\sim}{\theta} \quad . \tag{2.3}$$

Since both densities under the integral are normal, the

integration can be carried out explicitly. We find that

$$\underset{\sim}{y} \sim \text{MVN}(\underset{\sim}{0}, \underset{\sim}{\Sigma}) \tag{2.4}$$

where again

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi} \quad . \tag{2.5}$$

2.2 Parameter Estimation

The likelihood function for the responses of a random sample of N examinees under Equation 2.4 is given by

$$L(\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N | \underset{\sim}{\Lambda}, \underset{\sim}{\Phi}, \underset{\sim}{\Psi}) = \prod_{i=1}^{N} \frac{\exp(-\underset{\sim}{y}_i' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{y}_i / 2)}{(2\pi)^{p/2} |\underset{\sim}{\Sigma}|^{1/2}} \quad . \tag{2.6}$$

Maximizing Equation 2.6 with respect to the parameter matrices $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$, and $\underset{\sim}{\Psi}$ proceeds by taking the log of Equation 2.6, differentiating with respect to each parameter, equating the results to zero, then finding parameter values that satisfy these so-called likelihood equations. Unique estimates of the parameters do not exist, however, unless additional side restrictions are imposed along with Equation 2.6 in order to set the scales and

orientations of the latent θ's. It is typical to require that $\underset{\sim}{\Phi} = \underset{\sim}{I}_m$, the identify matrix of order m, and, in maximum likelihood estimation, that $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal.

The maximum likelihood (ML) estimation procedure described in the preceeding paragraph takes the form of minimizing a fitting function that is proportional to the log likelihood, namely

$$F = -\frac{N}{2}\left[\mathrm{tr}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}) - \log|\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}|\right] \quad , \tag{2.7}$$

where $\underset{\sim}{S}$ is the observed correlation matrix among the y's. It is important to note that the product over N examinees that appears in the likelihood simplifies down to expressions that involve only a summary of the response vectors, in terms of the observed covariance matrix. In other words, fully efficient estimates of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ can be obtained by utilizing only the $p(p + 1)/2$ elements of $\underset{\sim}{S}$, and that no information is lost by collapsing over the response patterns of N examinees, no matter how large N may be compared to $p(p + 1)/2$.

For later reference, we also mention two additional methods of estimating $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$. Both proceed by making the fitted $\hat{\underset{\sim}{\Sigma}}$, or the function of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$, and $\underset{\sim}{\Psi}$ given in Equation 2.5, close to $\underset{\sim}{S}$ in some sense. Let 'o' denote the "matrix stacking" operator, which

rewrites a matrix $\underset{\sim}{X} = (\underset{\sim}{x}_1\ \underset{\sim}{x}_2\ \cdots\ \underset{\sim}{x}_m)$ as the column vector $(\underset{\sim}{x}_1', \underset{\sim}{x}_2', \ldots, \underset{\sim}{x}_n')'$. The fitting methods are unweighted least squares (ULS), which minimizes

$$F = (\underset{\sim}{S}^o - \hat{\underset{\sim}{\Sigma}}^o)'(\underset{\sim}{S}^o - \hat{\underset{\sim}{\Sigma}}^o) \quad , \tag{2.8}$$

or the component-by-component sums of squared differences between the elements of $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$; and generalized least squares (GLS), which minimizes

$$F = (\underset{\sim}{S}^o - \hat{\underset{\sim}{\Sigma}}^o)'\underset{\sim}{W}^{-1}(\underset{\sim}{S}^o - \hat{\underset{\sim}{\Sigma}}^o) \quad ,$$

or the sums of squared differences between elements of $\underset{\sim}{S}$ and $\underset{\sim}{\Sigma}$ but weighted in a matter than takes into account the precision and the possibility of correlated errors in the estimation of $\underset{\sim}{S}$. In principle, the correct weight matrix required for a rigorous GLS solution is $\underset{\sim}{W} = \underset{\sim}{\Sigma} \times \underset{\sim}{\Sigma}$ where x represents the Kronecker or direct product of matrices. In practice, the consistent estimator $\underset{\sim}{S} \times \underset{\sim}{S}$ is used.

For formal justification and computational details on each of three fitting methods, the reader is referred to Anderson (1959), Browne (1974/1977), Jöreskog (1967, 1977), Jöreskog and

Goldberger (1972), and Lawley and Maxwell (1971). We merely mention a number of properties that are relevant to our presentation:

1. All three methods provide consistent estimates of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ under the assumptions noted at the beginning of this section.
2. Both ML and GLS require positive definite matrices $\underset{\sim}{S}$; ULS does not.
3. ML and GLS provide large-sample chi-square tests of model fit. Moreover, the difference between the chi-squares of nested models (e.g., a three-factor model versus a two-factor model) itself follows a chi-square distribution, with degrees of freedom equal to the number of additional parameters estimated in the less restrictive model, when the more restrictive model is correct. Thus, rigorous tests for the number of factors are available.

2.3 Rotation of a Solution

The solution provided by any of these procedures is unique but determined in part by the arbitrary imposition of rotational constraints; i.e., $\underset{\sim}{\Lambda}\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}' = \underset{\sim}{I}$ in GLS and ML, $\underset{\sim}{\Lambda}\underset{\sim}{\Lambda}' = \underset{\sim}{I}$ in ULS, and $\underset{\sim}{\Phi} = \underset{\sim}{I}$ in all three. It is easily seen that infinitely many other solutions for $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Phi}$ would combine through Equation 2.5 to produce the same $\underset{\sim}{\Sigma}$. Let $\underset{\sim}{A}$ be a square matrix of

full rank m, with normalized columns. If $\hat{\Sigma} = \Lambda\Phi\Lambda' + \Psi$, then it is also true that $\hat{\Sigma} = \Lambda^*\Phi^*\Lambda^{*\prime} + \Psi$, where $\Lambda^* = \Lambda A$ and $\Phi^* = A^{-1}\Phi A'^{-1}$. Various choices of A, though leaving the factor solution essentially unchanged in terms of model fit, can produce patterns of factor loadings that are easier to scan visually or to interpret substantively. The process of obtaining values of Λ^* and Φ^* is called factor rotation. Attention may be restricted to those A's that keep off-diagonal elements of Φ at zero (orthogonal rotations) or those that do not (oblique rotations). (See Harman (1976) and Thurstone (1947) for lucid explanations of rotation.)

2.4 Heywood Cases

It is possible to construct correlation matrices that conform to the common factor model, but for which one or more unique variances take the value of zero (Heywood, 1931). Zero uniquenesses correspond to measured variables falling completely within the factor space, or being explained perfectly by the latent variables, without measurement error at all. Negative uniquenesses are not defined within the usual context. Solutions with nonpositive uniqueness not generally palatable in practice.

Two approaches to dealing with these so-called Heywood solutions have been proposed in the literature. One is to allow such solutions, with the nonpositive uniqueness taken as a possible warning of model misfit (Jöreskog & Sörbom, 1980). In exploratory factor analysis, a Heywood solution may indicate that one is

attempting to fit a model with too many (or, occasionally, too few) factors, that one or more factors are poorly identified by the current set of observed variables, or, if sample size is small, that unfavorable sampling fluctuations have occurred. Appropriate remedies would be to fit a simpler model or to obtain data on additional variables and/or subjects. A second approach is to constrain estimation to solutions with only positive (or possibly only non-negative) unique variances. This may be done by imposing upon unique variances either arbitary constraints (e.g., Christoffersson, 1975, p. 9) or formal Bayesian prior distributions (e.g., Lee, 1981; Martin & McDonald, 1975).

3. A Common Factor Model for Dichotomous Data

In this section we outline the extension of the multiple factor model to dichotomous data. Attention is focused upon dichotomies which are reasonably considered to have arisen from a continuous latent process, but through observational constraints produce only dichotomous responses. Examples of this type would include right/wrong responses to test items, for/against votes on a referendum, and satisfied/dissatisfied judgments about a product. The model is also relaxed to allow for a fixed rate of "false positive" responses, as might occur when examinees can respond correctly to test items through lucky guesses as well as through the aptitude of interest.

There is no impediment to computing Pearson produce moment correlations among dichotomous variables ("phi coefficients," as they are called in this special case), and it might seem natural to apply the methods of the previous section to fit factor analytic models to correlation matrices so obtained. Several writers, however, have demonstrated dangers inherent in such an undertaking.

One problem is that the values of phi coefficients depend not only upon the strength of relationship among variables, but upon the means of the individual variables as well (Carroll, 1945, 1983). In the limiting case of two dichotomous variables with a perfect Guttman ordering, the value of the correlation obtained by Pearson's formula depends solely upon the means of the two variables and attains the value of 1 only when both variables have equal means.

A second problem is that the value of a dichotomous variable is bounded, implying that its regression on any continuous latent variable with infinite range cannot be linear (McDonald & Ahlawat, 1974). If applied directly to correlations from dichotomous variables, the linear factor analysis model is given by Equation 2.1 is misspecified from the start and pc ntially misleading because the best linear approximation to a true curvilinear relationship will depend on the regio: in which the data are most informative. In other words, the estimated linear relationship will depend upon the mean of the binary variable.

A third problem is illustrated in Mooijaart's (1983) approximation of the covariance among two discretized variables (e.g., a phi coefficient) in terms of a factor model for underlying continuous variables and functions of the observed discrete variables. In the special cases of either (a) all low factor loadings in the underlying model or (b) all discrete variables having means near .5, a factor model with the same number of factors but rescaled loadings will provide a good fit to the phi coefficients. In general, however, the expression for phi coefficients is augmented by terms that depend on the skewness of the discrete variables, which, with binary variables, is a direct function of their means values. Additional factors may be required to fit the phi matrix when these additional terms are large and their patterns are unfavorable.

When binary variables are produced by dichotomizing continuous variables, then the choice of cutting points materially affects the values of the expected phi coefficients. Factor analyses of phi coefficients of binary variables produced by the same underlying correlational structure but dichotomized at different points can conform to factor models with different structures and possibly different numbers of factors. For these reasons, we shall not discuss the analysis of phi coefficients, but rather confine our attention to models and methods under which strength of relationship and mean level are not confounded.

3.1 The Model

As in the classical model, we posit m latent variables $\underset{\sim}{\theta}$. In the case of $p > m$ observed responses (e.g., to a p-item test), we also posit the corresponding structure on p "response process" variables $\underset{\sim}{y} = (y_1,\ldots,y_p)$:

$$y_j = \lambda_{j1}\theta_1 + \ldots \lambda_{jm}\theta_m + \nu_j \quad , \tag{3.1}$$

where ν_j is a residual, the density of which will be specified presently. In contrast to factor analysis of measurement variables, however, we do not observe $\underset{\sim}{y}$ directly. Instead, we observe a vector of dichotomous variables $\underset{\sim}{x} = (x_1,\ldots,x_p)$ with values determined in the following manner:

$$x_j = \begin{cases} 1 \text{ if } y_j > \gamma_j \\ 0 \text{ if } y_j < \gamma_j \end{cases}$$

where γ_j is a value associated with item j—its "threshold" parameter. (The model will be relaxed in a following section to allow for the possibility of random positive responses.) Let $\underset{\sim}{\Gamma}$ denote $(\gamma_1,\ldots,\gamma_p)$.

Suppose that the residuals v_j are distributed as $N(0,\sigma_j^2)$ and are independent over items and examinees. We shall denote the diagonal matrix $(\sigma_1^2,\ldots,\sigma_p^2)$, or the vector of unique variances, as $\underset{\sim}{\Psi}$. The conditional probability of a correct response from examinee i to item j is then given as

$$P(x_{ij} = 1|\underset{\sim}{\theta}_i) = \frac{1}{\sqrt{2\pi}\,\sigma_j}\int_{\gamma_j}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{v - \sum_s \lambda_{js}\theta_{si}}{\sigma_j}\right)^2\right] dv$$

$$= F\left(\frac{\gamma_j - \sum_s \lambda_{js}\theta_{si}}{\sigma_j}\right)$$

$$= F_j(\underset{\sim}{\theta}_i) \quad . \tag{3.2}$$

Equation 3.2 will be recognized as a multivariate generalization of the two-parameter normal item response model (Lawley, 1943, 1944; Lord, 1952). Connections between the two models are explored in Lord and Novick (1968, Chapter 24).

Suppose it is further assumed that $\underset{\sim}{\theta}$ distributes $MVN(\underset{\sim}{0},\underset{\sim}{\Phi})$ in in a population of interest. As in the classical model, it follows that the marginal distribution of $\underset{\sim}{y}$ is $MVN(\underset{\sim}{0},\underset{\sim}{\Sigma})$, where again $\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}$. The fact that neither $\underset{\sim}{\theta}$ nor $\underset{\sim}{y}$ are observed introduces indeterminacies of scale and orientation into the

model; we shall begin to resolve them by specifying $\underset{\sim}{\Phi} = \underset{\sim}{I}_m$ and $\Sigma_{jj} = 1$ for each j. This implies that

$$\underset{\sim}{\Sigma} = \underset{\sim\sim}{\Lambda\Lambda} + \underset{\sim}{\Psi} \quad , \tag{3.3}$$

and

$$\sigma_j^2 = 1 - \sum_s \lambda_{js}^2$$

or, in matrix notation,

$$\underset{\sim}{\Psi} = \underset{\sim}{I} - \text{diag}(\underset{\sim\sim}{\Lambda\Lambda}') \quad .$$

Let $\underset{\sim}{x}_i = (x_{i1},\ldots,x_{ip})$ be the vector of 0/1 responses from examinee i in a randomly selected sample of size N. The marginal likelihood of the data is given by

$$L((\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_N)|\underset{\sim}{\Lambda},\underset{\sim}{\Gamma}) = \prod_{i=1}^{N} \int_{\underset{\sim}{\theta}} p(\underset{\sim}{x}_i|\underset{\sim}{\theta},\underset{\sim}{\Lambda},\underset{\sim}{\Gamma}) f(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}$$

$$= \prod_{i=1}^{N} \int_{\underset{\sim}{\theta}} \prod_j F_j(\underset{\sim}{\theta})^{x_{ij}} [1 - F_j(\underset{\sim}{\theta})]^{1-x_{ij}} f(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} \tag{3.4}$$

where $f(\underset{\sim}{\theta})$ represents the standard MVN density function. Equation 3.4 can also be written as a product over s distinct response patterns $\underset{\sim}{x}_\ell$, observed with frequencies r_ℓ, as

$$L = \prod_{\ell=1}^{s} \{ \int_{\underset{\sim}{\theta}} \prod_j F_j(\underset{\sim}{\theta})^{x_{\ell j}} [1 - F_j(\underset{\sim}{\theta})]^{1-x_{\ell j}} f(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} \}^{r_\ell} \qquad (3.5)$$

where $s \leq \min(N, 2^p)$. In contrast to the solution in which y's are observed directly, (3.5) cannot be collapsed further.

This fact has important implications for parameter estimation. It can be known a priori, for example, that the information about $\underset{\sim}{\Sigma}$ contained in observed values of $\underset{\sim}{y}$ from one million examinees to 100 items can be summarized without loss as a covariance matrix with just 5500 elements. If responses are to 100 dichotomous items, however, a total of 2^{100} distinct response patterns are possible; even allowing for the fact that many of these patterns will not occur in any given sample, hundreds of thousands of distinct pieces of data must be maintained to produce fully efficient estimates of Λ and $\underset{\sim}{\Gamma}$. To put it another way, the information in all cells of the 2^p contingency table of responses to all items is required for fully efficient estimation of parameters in the factor model.

3.2 Accounting for Random Correct Responses

For the purposes of test analysis and construction, a useful extension to the model described above is to account for the correct responses that result from correct guesses to multiple-choice items. Under these circumstances, the probabilities of correct response from even examinees of very low ability do not approach the value of zero implied by Equation 3.2. Failure to take these effects into account can produce analyses that are misleading as to not only the elements of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Gamma}$ but as to the number of factors needed to account for the data (Carroll, 1983).

It is possible to allow for chance success on item j at the rate of g_j by taking

$$F_j(\underset{\sim}{\theta}) = g_j + (1 - g_j)F_j^*(\underset{\sim}{\theta}) \quad ,$$

where $F_j^*(\underset{\sim}{\theta})$ is the function of $\underset{\sim}{\lambda}_j$ and γ_j given in Equation 3.2, which accounts for the rate of success produced by the latent factors of interest. No further revisions are required in Equations 3.3-3.5, although the following sections will consider implications of this extension for estimation procedures.

4. An Unweighted Least Squares Solution

Under the model of Section 3 for binary responses that arise from the dichotomization of underlying MVN response process variables, without the possibility of false positive responses due to guessing effects, it is possible to write the expectation of proportions of correct response to a given item j as

$$P_j = \int_{\gamma_j}^{\infty} f(z)\, dz \qquad (4.1)$$

and the proportion of persons responding correctly to both items j and k as

$$P_{jk} = \int_{\gamma_j}^{\infty} \int_{\gamma_k}^{\infty} f(z_1, z_2 | \sigma_{jk})\, dz_1\, dz_2 \quad , \qquad (4.2)$$

where f denotes a standard normal density function, univariate or bivariate as appropriate, and σ_{jk} denotes the correlation among response process variables y_j and y_k. Denoting the expected proportion of examinees answering item j correctly but item k incorrectly as $P_{j\bar{k}}$, and defining $P_{\bar{j}k}$ and $P_{\bar{j}\bar{k}}$ analogously, we could write expressions similar to Equation 4.2 for each. (P_{jk}, $P_{\bar{j}k}$, $P_{j\bar{k}}$, and $P_{\bar{j}\bar{k}}$ are the expected proportions of response in a two-by-two contingency table.)

From the observed proportion p_j, γ_j may be estimated via Equation 4.1 by

$$\hat{\gamma}_j = F^{-1}(p_j) \quad ,$$

where F^{-1} is the inverse of the cumulative standard normal distribution. Given estimates of γ_j and γ_k and the four entries in the two-by-two table of joint response frequencies, it is possible to estimate σ_{jk} via Equation 4.2. The resulting value is called the sample tetrachoric correlation coefficient (Pearson, 1900); efficient computing approximations are given by Divgi (1979). Let $\underset{\sim}{S}^*$ be the matrix of (sample) tetrachoric correlations among a set of p test items, with responses generated in accordance with the no-guessing model of Section 3.

4.1 Unweighted Analysis of S^*

Now $\underset{\sim}{S}^*$ is an estimate of $\underset{\sim}{S}$, the correlation matrix among the latent $\underset{\sim}{y}$'s, which has the common factor model given in Equation 3.3. Standard procedures for factor analysis of measured variables (Section 2) may be employed, then, to estimate $\underset{\sim}{\Lambda}$. Before proceeding, however, two points require attention. First, the sample tetrachoric takes a value of -1 or +1 when either $p_{\bar{j}k}$ or $p_{j\bar{k}}$ is zero. This problem is remedied in practice by adding a small number to each cell in the two-by-two contingency table for

each pair of items--in effect, placing a mild Dirichlet prior distribution on the joint proportions of response as in Fienberg and Holland (1970). Second, unlike a true correlation matrix or even a sample correlation matrix, $\underset{\sim}{S}^*$ is not necessarily positive definite. This fact typically rules out analysis by ML or GLS, leaving ULS. That is, $\underset{\sim}{\Lambda}$ is estimated by minimizing the quantity

$$\sum_j \sum_{j<k} (S^*_{jk} - \hat{\Sigma}_{jk})^2 \quad .$$

4.2 Advantages and Disadvantages of the ULS Solution

The advantages of ULS solutions for factor models for dichotomous variables are first, its superiority over factor analysis of phi coefficients, and second, its relative economy; solutions in the measured variables case generally require far less computation than the methods specifically designed for the categorical data, as outlined in subsequent sections of this presentation.

The disadvantages of this solution can be classified into two categories. The first category arises in the attempt to compute $\underset{\sim}{S}^*$. Extreme values will be poorly determined, and those that would have been +1 or -1 take values that depend on the choice of an ad hoc remedy. And because estimation error is introduced in the production of $\underset{\sim}{S}^*$, the statistical theory for obtaining ULS standard errors (Browne, 1974/1977) does not hold. The second

category arises from the fact that unlike the case of normally distributed measured variables, summarization of dichotomous variables in terms of a covariance matrix does not retain all the information about their joint relationships. Only the information in the one-way marginals (percents-correct) and two-way marginals is used. Computational efficiency is thus achieved at the sacrifice of information.

4.3 Adjustments When Guessing Is Present

The preceding discussion considered the case in which responses were determined solely through $\underset{\sim}{\theta}$, not accounting for the possibility of chance successes. The same solution can be carried out when chance successes do occur, at prespecified rates g_j to each of the items, if the observed proportions and joint proportions are adjusted appropriately. Carroll (1945) and Samejima (in Green et al., 1982, p. 28) give formulas for this purpose. Jensema's (1976) expression for adjusted percents correct and Samejima's expressions for joint proportions are shown below. Observed values are indicated by asterisks; the adjusted values are subsequently used in Equations 4.1 and 4.2.

$$P_j = (P^*_j - g_j)/\bar{g}_j$$

$$P_{jk} = P^*_{jk} - (g_k/\bar{g}_k)P^*_{j\bar{k}} - (g_j/\bar{g}_j)P^*_{\bar{j}k} + (g_j g_k/\bar{g}_j\bar{g}_k)P^*_{\bar{j}\bar{k}}$$

$$P_{j\bar{k}} = (\bar{g}_k)^{-1} P^{*}_{j\bar{k}} - (g_j/\bar{g}_j\bar{g}_k) P^{*}_{j\bar{k}}$$

$$P_{\bar{j}k} = (\bar{g}_j)^{-1} P^{*}_{\bar{j}k} - (g_k/\bar{g}_j\bar{g}_k) P^{*}_{\bar{j}k}$$

$$P_{\bar{j}\bar{k}} = (\bar{g}_j\bar{g}_k)^{-1} P^{*}_{\bar{j}\bar{k}} \quad ,$$

where $\bar{g}_j = 1 - g_j$ and $\bar{g}_k = 1 - g_k$. These adjustments can produce proportions above 1 or below 0. Ad hoc remedies, such as the imposition of arbitrary floors and ceilings on either proportions or values of g_j are then required before the estimation of the factor model can begin.

5. Generalized Least Squares Solutions

Section 4 presented formulas for the expected values of p_j, or item proportions correct, and p_{jk}, or joint item proportions, in terms of the parameters of the extended common factor model (possibly after adjustment for prespecified rates of chance success, as in Section 4.3). ULS estimations proceeds from these formulas alone, minimizing a quantity that measures the similarity between the data (sample percents correct and sample tetrachoric correlations, the latter computed from sample joint proportions) and a fitted facsimile of the data in terms of the parameters. The similarity is judged by sum of the squared differences,

element by element, with each element weighted equally. More efficient use of data can be made by taking into account the varying magnitudes and interrelationships of sampling error among the elements. One approach by which this objective can be achieved is generalized least squares (GLS).

5.1 Christoffersson's Solution

Let $\underset{\sim}{P} = (P_1, P_2, \ldots, P_p, P_{11}, \ldots, P_{jk}, \ldots)$, with $1 \leq k < j \leq p$, be the vector of the expected values of P_j and P_{jk}, modeled as functions of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Gamma}$, and let $\underset{\sim}{p}$ be the corresponding vector of observed values. When the model is correct, the quantity $\underset{\sim}{e} = \underset{\sim}{p} - \underset{\sim}{P}$ will follow a multivariate normal distribution in large samples with expection $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Sigma}_e$. Christoffersson (1975, Appendix 2) derives an expression for a consistent estimator $\underset{\sim}{S}_e$ of $\underset{\sim}{\Sigma}_e$, and implements a GLS solution for the parameters of the factor model by minimizing

$$F = (\underset{\sim}{p} - \underset{\sim}{P})' \underset{\sim}{S}_e^{-1} (\underset{\sim}{p} - \underset{\sim}{P}) \quad .$$

The solution thus obtained provides consistent parameter estimates.

A number of additional features of Christoffersson's solution also merit comment at this point.

First, his expressions for the elements of $\underset{\sim}{S}_e$ include not only p_j and p_{jk}, terms from one-way and two-way margins of the 2^p

raw data table, but also terms from the three- and four-way margins; that is, joint proportions correct for items taken three and four at a time. This means that the GLS solution is using more information than the ULS solution, but by ignoring yet higher level interactions, still not all of the information available. (As discussed in Section 6.2, the loss may be negligible.)

Second, statistical tests of model fit are available. Asymptotically F follows a chi-square distribution, with degrees of freedom equal to $p(p + 1)/2$ minus the number of parameters in $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Gamma}$ estimated in the model (as in previous section, certain restrictions in $\underset{\sim}{\Lambda}$ are required to eliminate linear and rotational indeterminacies). This test is not usually of interest so much for itself--the model is not expected to fit--but for comparisons between models with different numbers of factors. The difference between the chi-squares for an m factor and an m + 1 factor solution for the same data also follows a chi-square distribution in large samples when the m factor model is correct, with degrees of freedom equal to the number of additional parameters estimated in the less restrictive solution. Indeed, the test of most interest in educational and psychological applications is typically the comparison of the one- and two-factor solutions.

Third, standard errors of estimation are also available. In large samples, the covariance matrix of estimation errors of the free elements of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Gamma}$ is approximated by the inverse of the

matrix of second derivatives of F with respect to these parameters. Standard errors for individual parameters are square roots of the corresponding diagonal elements. In exploratory work, these standard errors are not of major interest. They apply to the parameters only as estimated, not to rotated solutions. They prove more interesting by way of contrast to those obtained in the full information maximum likelihood solution described in the next section.

Fourth and finally, computation requirements are considerably heavier than those of the ULS solution. Solution is iterative, requiring the numerical solution of integrals of the form of Equations 4.' and 4.2 in each cycle. Further comment on this point follows a discussion of Muthén's GLS solution, asymptotically equivalent to Christoffersson's but somewhat less burdensome.

5.2 Muthén's Solution

Muthén's (1978) GLS solution bears more resemblance to the ULS solution of the preceding section, as well as the solutions for measured variables; the fitting function again produces estimates that in the appropriate sense make a fitted correlation matrix similar to an observed one. Whereas Christoffersson minimizes residuals in terms of the P's in Equation 5.1, Muthén minimizes

$$F = (\underset{\sim}{s} - \underset{\sim}{\xi})' \underset{\sim}{S}_{\delta}^{-1} (\underset{\sim}{s} - \underset{\sim}{\xi})' \qquad (5.2)$$

where $\underset{\sim}{\xi} = (\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2)$ with $\underset{\sim}{\xi}_1 = \underset{\sim}{\Gamma}$, $\underset{\sim}{\xi}_2 = (\sigma_{12}, \ldots, \sigma_{jk}, \ldots)$, and $\underset{\sim}{s}$ being the sample estimates of the quantities, i.e., the sample thresholds and sample tetrachorics—where $\underset{\sim}{S}_{\delta}$ is a consistent estimator of the covariance matrix of $\underset{\sim}{\delta} = \underset{\sim}{\xi} - \underset{\sim}{s}$. Muthén obtains an expression for $\underset{\sim}{S}_{\delta}$ from Christoffersson's expression for $\underset{\sim}{S}_{e}$ by "linearizing" the model; that is, by approximating the complex relationship between $\underset{\sim}{\xi}$ and $\underset{\sim}{P}$ by the initial terms of a Taylor series expansion. Integrals of the form of Equations 4.1 and 4.2 need then be evaluated only once. These procedures have been incorporated into the computer program LISCOMP (Muthén, 1985).

Muthén's solution shares many of the other characteristics of Christoffersson's, notably use of three- and four-way marginal information, consistent estimates, standard errors, and tests of fit. And although Muthén's solution is faster, practical limitations arise from the same source, namely, the magnitude of the matrix $\underset{\sim}{S}_{e}$. These effects are illustrated in Table 1. Computing requirements under the GLS solution increase proportionally to m and with the fourth power of p. About 25 items seems to be an upper limit with current machinery.

Insert Table 1 about here

Muthén notes that in many cases, ULS estimates are reasonable approximations to GLS estimates. The superiority of GLS, through its use of three- and four-way joint proportions, becomes more evident as one attempts to extract more from the data, so to speak; that is, with other features held constant, in solutions with fewer examinees, fewer items, or more factors.

6. A Maximum Likelihood Solution

The preceding sections have considered ULS and GLS estimation of the parameters of a common factor model for dichotomous responses. These are "limited information" solutions, in that they utlilize only information in lower order margins of the full 2^p contingency table that summarizes all responses, and therefore all available information, for estimation. In this section, we review a full information solution, namely the marginal maximum liklihood (ML) estimation introduced by Bock and Aitkin (1981). (The Bock-Aitkin procedure extends on an earlier solution given by Bock and Lieberman (1970) for the one-dimensional case.) The following discussion is based on this approach, which has been implemented in the TESTFACT computer program (Wilson, Wood, & Gibbons, 1983).

6.1 The Marginal Probability of a Response Pattern

Assume again the common factor model for dichotomous items given in Section 3, initially without the possibility of chance

success; that is, we posit m latent variables $\underset{\sim}{\theta}$ and $p > m$ observed binary variables x_j that take the values 1 or 0 in the following manner:

$$x_{ij} = \begin{cases} 1 \text{ if } y_{ij} > \gamma_j \\ 0 \text{ if } y_{ij} < \gamma_j \end{cases} \tag{6.1}$$

where

$$y_{ij} = \lambda_{j1}\theta_{i1} + \ldots + \lambda_{jm}\theta_{im} + \nu_j \quad . \tag{6.2}$$

The residual terms ν_j are independent over items and examinees, and follow $N(0,\sigma_j^2)$ distributions, where

$$\sigma_j^2 = 1 - \sum_k \lambda_{jk}^2 \quad .$$

Recalling Equation 3.2, this implies that

$$P(x_{ij} = 1|\underset{\sim}{\theta}) = F\left(\frac{\gamma_j - \sum_k \lambda_{jk}\theta_{ik}}{\sigma_j}\right)$$

$$= F_j(\underset{\sim}{\theta}_i) \quad , \tag{6.3}$$

where F is the cumulative standard normal distribution. It is further assumed that $\underset{\sim}{\theta} \sim \text{MVN}(\underset{\sim}{0},\underset{\sim}{I}_m)$, from which it follows that $\underset{\sim}{y} \sim \text{MVN}(\underset{\sim}{0},\underset{\sim}{\Sigma})$ where

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi} \quad . \tag{6.4}$$

It was shown that under these assumptions, the probability of a typical response pattern $\underset{\sim}{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell p})$ is given by

$$P_\ell = P(\underset{\sim}{x} = \underset{\sim}{x}_\ell) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_j F_j(\underset{\sim}{\theta})^{x_{\ell j}} [1 - F_j(\underset{\sim}{\theta})]^{1-x_{\ell j}} f(\underset{\sim}{\theta})\, d\theta_1 \cdots d\theta_m$$

$$= \int_{\underset{\sim}{\theta}} L_\ell(\underset{\sim}{\theta}) f(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} \quad .$$

(We recall that the possibility of chance successes at fixed rates g_j may be incorporated at this point by replacing $F_j(\underset{\sim}{\theta})$ above with $F_j^*(\underset{\sim}{\theta}) = g_j + (1 - g_j)F_j(\underset{\sim}{\theta})$.) This integral can be approximated to any desired degree of accuracy by m-dimensional Gauss-Hermite quadrature (Stroud & Sechrest, 1966):

$$P_\ell \approx \sum_{k_m}^{q} \cdots \sum_{k_2}^{q} \sum_{k_1}^{q} L_\ell(\underset{\sim}{X}_{\underset{\sim}{k}}) A(X_{k_1}) A(X_{k_2}) \cdots A(X_{k_m}) \quad ,$$

where integration over real m-space has been replaced by summation over a finite grid of q^m quadrature points $\underset{\sim}{X}_k = (X_{k_1},\ldots,X_{k_m})$. Because it has been assumed that the dimensions of $\underset{\sim}{\theta}$ are orthogonal in the population of interest, the weight assigned to each point is the product of the weights associated with each coordinate X_{k_t}.

6.2 Estimation Procedures

Consider the responses of a random sample of N examinees. Under the assumptions given above, it follows that the counts r_ℓ of distinct response terns follow a multinomial distribution given by

$$P(\underset{\sim}{r}|\underset{\sim}{\Lambda},\underset{\sim}{\Gamma}) = \frac{N!}{r_1!\, r_2! \cdots r_s!} P_1^{r_1} P_2^{r_2} \cdots P_s^{r_s} \quad . \qquad (6.5)$$

The full information maximum likelihood solution given by Bock and Aitkin (1981) maximizes Equation 6.5 with respect to the elements of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Gamma}$.

It proves convenient computationally to rewrite the argument of the normal probability function in Equation 6.3 in terms of slopes a_{jk} and intercepts c_j as follows:

$$-(\gamma_j - \Sigma_k \lambda_{jk}\theta_{ik})/\sigma_j = c_j + \Sigma_k a_{jk}\theta_{ik} \quad .$$

From maximum likelihood estimates of $\underset{\sim}{a}$'s and $\underset{\sim}{c}$'s, maximum likelihood estimates of γ's and $\underset{\sim}{\lambda}$'s are obtained as

$$\hat{\gamma}_j = -\hat{c}_j/\hat{d}_j \quad \text{and} \quad \hat{\lambda}_{jk} = \hat{a}_{jk}/\hat{d}_j$$

where

$$\hat{d}_j = (1 + \sum_s \hat{a}^2_{js})^{1/2} \quad .$$

Estimation proceeds by finding those values of $\underset{\sim}{a}$ and $\underset{\sim}{c}$ which maximize Equation 6.5. This is done by taking the first derivatives of the logarithm of the likelihood function Equation 6.5 with respect to each parameter in turn, setting them to zero, and solving with respect to $\underset{\sim}{a}$ and $\underset{\sim}{c}$. The interested reader is referred to Bock and Aitkin for details of the solution. The essence of the approach, however, can be seen in the form of the likelihood equations. For a typical parameter u_j from item j (either a slope or an intercept), we have

$$0 = \sum_{k_m}^{q} \cdots \sum_{k_1}^{q} \frac{\bar{r}_{jk} - \bar{N}_k F_j(X_k)}{F_j(X_k)[1 - F_j(X_k)]} \cdot \frac{\partial F_j(X_k)}{\partial u_j} \qquad (6.6)$$

$$u_j = c_1, \ldots, c_p, a_{11}, \ldots, a_{pm}$$

where

$$\bar{N}_{\underset{\sim}{k}} = \sum_{\ell=1}^{s} \frac{r_\ell}{P_\ell} L_\ell(X_k) A(X_{k_1}) \ldots A(X_{k_m}) = \sum_{\ell=1}^{s} r_\ell \, P(\underset{\sim}{X}_k | \underset{\sim}{x}_\ell, \underset{\sim}{\Lambda}, \underset{\sim}{\Gamma}) \tag{6.7}$$

is approximately proportional to the population density in the region of quadrature point $\underset{\sim}{X}_k$ and

$$\bar{r}_{\underset{\sim}{jk}} = \sum_{\ell=1}^{s} x_{j\ell} \frac{r_\ell}{P_\ell} L_\ell(\underset{\sim}{X}_k) \, A(X_{k_1}) \ldots A(X_{k_m})$$

$$= \sum_{\ell=1}^{s} x_{j\ell} r_\ell \, P(\underset{\sim}{X}_k | \underset{\sim}{x}_\ell, \underset{\sim}{\Lambda} \, \underset{\sim}{\Gamma}) \tag{6.8}$$

is approximately proportional to the probability of a correct response to item j from examinees with $\underset{\sim}{\theta}$'s in this region. (An application of Bayes theorem will be recognized in Equations 6.7 and 6.8, yielding the posterior probability of ability $\underset{\sim}{X}_k$ given $\underset{\sim}{x}_\ell$, conditional on $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Gamma}$.)

Solution of these equations is iterative, since the terms $\bar{r}_{\underset{\sim}{jk}}$ and $\bar{N}_{\underset{\sim}{k}}$ depend on the parameters $\underset{\sim}{a}$ and $\underset{\sim}{c}$ themselves through $L_\ell(\underset{\sim}{X}_k)$. In a variation of an EM algorithm (Dempster, Laird, & Rubin, 1977), Bock and Aitkin proceed in cycles with two steps each:

E-step: Using provisional estimates $\underset{\sim}{a}^t$ and $\underset{\sim}{c}^t$, evaluate Equations 6.7 and 6.8. These are the expected values of the population densities and item proportions correct in the regions of the quadrature points, conditional on the data and $\underset{\sim}{a}^t$ and $\underset{\sim}{c}^t$.

M-step: Taking the $\bar{r}_{jk}$'s and $\bar{N}_k$'s as known, solve Equations 6.6 with respect to the parameters to obtain $\underset{\sim}{a}^{t+1}$ and $\underset{\sim}{c}^{t+1}$.

Solving the so-called likelihood equations in this manner yields saddle points or relative extrema of Equation 6.5. Whether they are relative maxima can be determined by examining values of the likelihood function in the region around the solution. Whether a relative maximum is unique can be studied by iterating from a number of different starting values.

As with the GLS solution, the ML solution provides for standard errors of estimation and statistical tests of fit. The covariance matrix of estimation errors of the parameters is given by the negative inverse of the matrix of expected second derivatives of the log likelihood function; this may be approximated by the matrix of second derivatives at the ML solution. Standard errors are obtained as the square roots of the appropriate diagonal elements. For a model with m factors, the likelihood ratio chi-square approximation for a test against a general multinomial distribution is given by

$$G^2 = -2 \sum_{\ell} r_{\ell} (\log NP_{\ell}/r_{\ell})$$

with degrees of freedom equal to $2^p - p(m + 1) + m(m - 1)/2$. This value reflects the number of cells in the full contingency table layout for the data, less the number of parameters estimated plus the number of constraints imposed to effect identification. Because the expected number of examinees per cell will usually be small for more than, say, 10 items, the approximation to the chi-square distribution may be unreliable. Comparison of G^2 for nested models such an an m factor model versus m + 1 factor model, however, is more robust under these circumstances.

A comparison of the standard errors for estimated parameters obtained from GLS and ML provides a measure of a loss of information in GLS when joint information for more than four items at a time is neglected. Comparisons reported by Gibbons (1984a) indicate the differences are slight; not only standard errors comparable within .01 were found for a data set amenable to solution by both ML and GLS, but similar parameter estimates and chi-square values were obtained.

6.3 ML Versus GLS

Given that both ML and GLS provide standard errors, tests of fit, and comparable and consistent parameter estimates, it might be asked whether one method is to be preferred over the other.

The answer is yes, at least with present computing machinery; the computational algorithms of ML and GLS present clear and distinct advantages of one solution over the other under appropriate circumstances. As noted in the previous section, the demands of GLS increase linearly with the number of factors but with the fourth power of the number of items. The numerical integration over the factor space required in ML, on the other hand, implies geometric increases in computation with the number of factors, although the item by item computations required in the M-steps increase only linearly with the number of items. The practical implications are these: ML is preferable for long tests with few factors; GLS is preferable for short tests with many factors; both are acceptable for short tests and few factors; and at present, neither is very good for long tests and many factors. (Bock (1984) quantifies the current meaning of the phrase "many factors" saying that with 60 items, 1-3 factor models are quite reasonable with ML, 4 factors are possible, and 5 is about as much as currently feasible.)

7. Further Extensions of the Models

The preceding sections of this review have considered the extension of classical factor analysis to dichotomous variables, concentrating on the basic models and on estimation procedures. In this final section, we briefly survey a number of additional directions in which these models may be further extended, and direct the reader to work in progress in these areas.

7.1 Polytomous Responses

Discussion thus far has concentrated on analyses of dichotomous data. Data received in the form of ratings on n_j-point ordinal scales can also be addressed in much the same manner, if it is reasonable to suppose that the data arise from cut points on underlying continuous normal variables. Let the probability of a response in a category less than or equal to category k be given by

$$F_{jk}(\underset{\sim}{\theta}) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \int_{\gamma_{jk}}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{y - \Sigma_s \lambda_{js}\theta_{is}}{\sigma_j} \right)^2 \right] dy \qquad k=1,\ldots,n-1$$

and $F_{jo}(\underset{\sim}{\theta})$ is defined as 0 and $F_{j,n_j}(\underset{\sim}{\theta})$ is defined as 1. Then the probability of a response in category k is given by

$$P(x_{ij} = k|\underset{\sim}{\theta}) = F_{jk}(\underset{\sim}{\theta}) - F_{j,k-1}(\underset{\sim}{\theta}) \quad .$$

Under this model, either of two approaches toward parameter estimation can be taken. Under ULS or GLS, one first estimates the correlations among supposed underlying MVN variables $\underset{\sim}{y}$; these are called the sample polychoric correlations (Olsson, Drasgow, & Dorans, 1982). From this point estimation proceeds as in the dichotomous case. Such solutions are provided in Jöreskog and

Sörbom's (1984) LISREL program and Muthén's (1985) LISCOMP. Under ML, solutions are available for both the unidimensional case (Muraki, 1983, and Thissen, 1984) and the multidimensional case (Muraki, 1985). In principle, all of the extensions mentioned in the following sections are applicable to polytomous response data.

7.2 Simultaneous Estimation of Asymptotes

The marginal probability of a sample of response patterns $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_N$ was given in Section 3 as

$$P = \prod_i \int \prod_j F_j(\underset{\sim}{\theta})^{x_{ij}}[1 - F_j(\underset{\sim}{\theta})]^{1-x_{ij}} f(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} \quad , \tag{7.1}$$

where the item response functions $F_j(\underset{\sim}{\theta})$ were given by either

$$F_j(\underset{\sim}{\theta}) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \int_{\gamma_j}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{y - \sum_s \lambda_{js}\theta_{is}}{\sigma_j} \right)^2 \right] dy \quad , \tag{7.2}$$

the cumulative normal distribution, or by

$$F_j^*(\underset{\sim}{\theta}) = g_j + (1 - g_j)F_j(\underset{\sim}{\theta}) \quad , \tag{7.3}$$

with g_j a fixed constant indicating a possibly nonzero lower asymptote for the probability of a correct response from even

examinees with low values of $\underset{\sim}{\theta}$ in every component. Under ULS and GLS estimation, use of Equation 7.3 rather than Equation 7.2 led to adjustments of the observed proportions and pairwise proportions of correct responses to items. Under ML estimation, the adjustment for chance correct responses need not be limited to fixed values $\underset{\sim}{g}$; in principle there is no reason that Equation 7.1 cannot be maximized with respect to the g's as well as the a's and c's. One simply includes additional likelihood equations, one for each g_j (or only one if it is desired to estimate a common g for all items) of the form given as Equation 6.6. This possibility is currently under investigation by Bock and Muraki (1984).

Preliminary results reported by Muraki (1984) with fixed asymptotes indicate caution may be required in interpreting the results of such an endeavor. Muraki examined simulated responses to 25 items from a randomly generated sample of 1000 subjects from a standard normal population, with the true item response model having one dimension and including an asymptote of .20 for all items. In a preliminary analysis, a one-factor item response model with estimated lower asymptotes was fit to the data using the BILOG computer program (Mislevy and Bock, 1982) in order to obtain values of g which could be input to common factor runs. Four factor models were fit to these data by means of the ML solution in the TESTFACT program:

1. One factor, g's fixed at zero.

2. One factor, g's fixed at nonzero preliminary estimates.

3. Two factors, g's fixed at zero.

4. Two factors, g's fixed at nonzero preliminary estimates.

It was found that models 2 and 3 both provided a good fit to the data, as well as model 4.

This finding suggests that the likelihood surface of a more general model that includes both 2 and 3, namely a two-factor model in which asymptotes are also estimated, is nearly equally high in regions around at least two possible parameter vectors, and may even exhibit relative maxima at these points. This finding is not disturbing from a data-analytic point of view; it is not surprising to have obtained a good fit from model 3, even though it was not the model under which the data were generated, in view of the fact that more parameters were estimated. Practical considerations give one pause, however; two solutions (2 and 3) from the plausible general model (4) both explain the data nearly equally well, but have quite different implications for action. Without careful examination, the decision of whether or not to split the items into two different tests might depend on the starting values that one might happen to supply to the iterative solution. One must conclude, not surprisingly, that model fitting alone, without consideration of the nature of the data and the properties of the models being used, should not be the sole guide to test construction decisions.

7.3 Bayesian Prior Distributions

As noted in Section 2, the occasional appearance of Heywood solutions, or occurrences of zero or negative unique variances, has led various researchers to incorporate Bayesian prior distributions on these parameter (Lee, 1981; Martin & McDonald, 1975). Under the ML solution presented in Section 6, unique variances do not appear as parameters to be estimated; their values are implied through values of the a's through

$$\sigma_j^2 = 1 - \sum_s \frac{a_{js}^2}{1 + a_{js}^2} \quad . \tag{7.4}$$

Under these circumstances, a Heywood solution takes the appearance of one or more a's becoming infinite. To avoid this problem, it might seem appropriate at first blush to impose prior distributions on the a's. The difficulty arises, however, when comparing the fit of competing models, that the strengths of the priors imposed on different solutions may vary as a function of the number of parameters being estimated.

A more satisfactory solution, developed by Mislevy and Bock and reported in Bock (1984), is to impose prior distributions on a's implicitly, by imposing them on unique variances and inferring the implied distributions on the joint distribution of a's through

Equations 7.4. Independent beta distributions on unique variances are proposed, with parameters (r,1) where $1 < r < 2$. This distribution takes the form

$$p(\pi) = B^{-1}(r,1)\pi^{r-1} \tag{7.5}$$

where B(r,1) represents the beta function. A choice of r near 1 results in a prior distribution that runs nearly flat across the unit inverval, but drops suddenly and steeply to zero as π approaches zero. This is tantamount to saying that one knows little about the value of the unique variance, except that it is not zero or negative. Substituting the expression Equation 7.4 into Equation 7.5, we have a joint prior distribution for the slope parameters of item j:

$$p(\underset{\sim}{a}_j|r) = B^{-1}(r,1)[1 - \sum_s \frac{a_{js}^2}{1 + a_{js}^2}]^{r-1} \quad . \tag{7.6}$$

Multiplying the marginal likelihood function Equation 6.5 by the prior distribution Equation 7.6 on a's yields an expression proportional to the posterior distribution of the a's and c's, with a diffuse prior distribution on c's implicit. The result

is then maximized as in the straight maximum likelihood solution, except the maxima are now modal points of the posterior.

By similar methods, prior distributions could also be introduced for $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$, and, when included in the model, the guessing parameter $\underset{\sim}{g}$. A fully Bayesian approach would allow for the incorporation of prior knowledge about items and hypotheses about their interrelationships. While such a treatment has yet to appear, the stage has been well set; the marginal maximum likelihood solution described in Section 5 provides a satisfactory starting point for dealing with the likelihood term, and experience with forms and procedures for prior distributions gained in the measured variables case (e.g., Lee, 1982; and Martin & McDonald, 1975) appears readily transferrable.

7.4 Relaxation of Distributional Assumptions

The usual factor analytic formulation for discrete variables assumes normal distributions for both the response functions (or conditional distributions of $\underset{\sim}{y}$ given $\underset{\sim}{\theta}$), and for the distributions of $\underset{\sim}{\theta}$. These assumptions are motivated by convenience; the marginal distribution resulting from the mixture (Equation 2.3) is itself normal, simplifying to expressions of the type shown as Equation 2.4 and 2.5. There is no reason, however, not to consider other distributional forms. Use of the logistic function for the conditional distribution, for example, leads in the one-dimensional case to certain item response models considered in Lord and Novick

(1968). Bartholomew (1980) suggests a model in which both $f(\underset{\sim}{y}|\underset{\sim}{\theta})$ and $\underset{\sim}{g}(\theta)$ are logistic; i.e.,

$$P(x_{ij} = 1|\underset{\sim}{\theta}) = [1 + \exp(c_j + \sum_k a_{ij}\theta_{ik})]$$

and

$$P(\theta_1 < x_1,\ldots,\theta_m < x_m) = \Pi\ [1 + \exp(x_k - \theta_k)]^{-1} \quad .$$

Due to the similarities in the shapes of the logistic and normal distributions, results from this logit factor model can be expected to agree well with results from the normal model discussed in the preceeding sections. Computations appear simpler under the logit model in the one-dimensional case, but simpler under the normal in the multivariate case.

More restraints than are actually needed to obtain an identified model are still being imposed, however (see Bartholomew, 1980, 1984, 1985). Indeed, if the response functions are sufficiently flexible, the distribution $\underset{\sim}{\theta}$ can be arbitrarily specified within broad limits. Suppose that the marginal distribution of response y is given as

$$p(\underset{\sim}{y}) = \int_{\underset{\sim}{\theta}} f(\underset{\sim}{y}|\underset{\sim}{\theta})\ \underset{\sim}{g}(\underset{\sim}{\theta})\ d\underset{\sim}{\theta}$$

where g is the continuous distribution of the latent variable $\underset{\sim}{\theta}$. Let g* be any other density over the same latent space that can be obtained by suitable stretching, expanding, or rotation. That is, $g^*(\underset{\sim}{\theta}) = h(g(\underset{\sim}{\theta}))$, where h is continuous and strictly increasing in all components. Define f* by $f^*(\underset{\sim}{y}|\underset{\sim}{\theta}) = f(\underset{\sim}{y}|h^{-1}(\underset{\sim}{\theta}))$. Then

$$p(\underset{\sim}{y}) = \int_{\underset{\sim}{\theta}} f(\underset{\sim}{y}|\underset{\sim}{\theta})\, g(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}$$

$$= \int_{\underset{\sim}{\theta}} f^*(\underset{\sim}{y}|\underset{\sim}{\theta})\, g^*(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} \quad .$$

This result suggests three ways by which distributional assumptions in the normal model for categorical variables might be relaxed.

First, one might wish to maintain the normal linear regression model for the response functions, but allow the $\underset{\sim}{\theta}$ distribution to to take forms other than the standard normal. The idea here would be to maintain response functions similar in form to IRT models contemplated for subsequent use, but avoid distortions in $\underset{\sim}{\Lambda}$ due to additional and unnecessary assumptions about the shape of the population distributions. Bock and Aitkin (1981) mention this possibility, and methods for estimating latent distributions that could be incorporated into the ML solution are found in Mislevy (1984).

Second, one might choose to relax even further by fixing the $\underset{\sim}{\theta}$ population distribution in some tractible manner--e.g., uniform density on the unit interval--but allowing very flexible or even nonparameteric forms for the response functions. The idea here would be to obtain more detailed diagnostic information about items, such as the presence of non-monotonic response functions. Work along these lines has been begun in the unidimensional case by Winsberg, Thissen, and Wainer (1982), who fit spline functions to item response data. Again, these extensions can be incorporated into the ML solution in a straightforward manner.

Third, one can specify the form of $f(\underset{\sim}{x}|\underset{\sim}{\theta})$ to achieve desired properties. The next subsection considers a line of work with this motivation.

7.6 Foundations of Factor Analysis

In a more general setting that includes the factor analysis of categorical variables, Bartholomew (1980, 1984, 1985) began by considering implications for the conditional distribution $h(\underset{\sim}{\theta}|\underset{\sim}{x})$ (what one knows about the latent variables after having observed the manifest variables) imposed by the choice of the form of $f(\underset{\sim}{x}|\underset{\sim}{\theta})$. He shows that if (i) conditional or local independence is satisfied, i.e.,

$$f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{j=1}^{p} f_j(x_j|\underset{\sim}{\theta})$$

so that the m latent variables account completely for relationships among the p manifest variables, and (ii) each $f_j(x_j|\underset{\sim}{\theta})$ belongs to the exponential family, i.e.,

$$f_j(x_j|\underset{\sim}{\theta}) = F_j(x_j)G_j(\underset{\sim}{\theta}) \exp\{ \sum_k [\sum_j u_{jk}(x_j)] \phi_k(\underset{\sim}{\theta})\} \quad (7.7)$$

with the special restriction that

$$u_{jk}(x_j) = \gamma_{jk} + \alpha_{jk}u_j(x_j)$$

then there exists an m-dimensional sufficient statistic $\underset{\sim}{X}$ for $\underset{\sim}{\theta}$, in the form of m functions of the p responses in $\underset{\sim}{x}$:

$$X_k = \sum_j \alpha_{jk}u_j(x_j) \quad .$$

If each f_j is normal, Poisson, or binomial, then each $u_j(x_j)$ is proportional to x_j. In the normal case introduced in Section 3, the sufficient statistics are given by

$$\underset{\sim}{X} = \underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{x} \quad .$$

Equivalently,

$$\underset{\sim}{X} = \underset{\sim\sim}{\Xi}\underset{\sim}{\theta} + \underset{\sim}{\Xi}^{1/2}\underset{\sim}{v}$$

where $\underset{\sim}{\Xi} = \underset{\sim\sim}{\Lambda}\underset{\sim\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ and $\underset{\sim}{v}$ is a random vector of independent standardized variables. The sufficient statistics may thus be thought of as a weighted average of the latent variables of interest and residuals, the latter of which contain variation specific to individual variables and random error.

Attention is focused upon estimable linear combinations of observed variables, which contain all the information in the data about the latent variables. Bartholomew points out that these statistics remain unchanged with monotonic transformations of any coordinate of the latent distribution. It may be inferred that in the absence of additional external reasons to specify the exact form of the latent marginal distribution g or the conditional distribution f, factor analysis models provide at best ordinal information within dimensions about persons' values on latent variables. The margial orderings are not invariant with respect to rotation, so even ordinal information is conditional on the arbitrary specification of orientation whenever $m > 1$.

The dependence of factor analytic solutions upon such arbitrary choices as scaling and orientation of coordinates has long been a source of dissatisfaction with analytic procedures. A degree of specification on the form of f sufficient to eliminate

these indeterminacies in case of binary variables is found in Stegelmann's (1983) multidimensional Rasch model. In its general form,

$$f_j(x_j|\underset{\sim}{\theta}) = \{1 + \exp[-\sum_s \alpha_{js}(\theta_s - \eta_j)]\}^{-1} \quad , \qquad (7.8)$$

where the α_{js} take prespecified values of 1 or 0.

A submodel of Equation 7.7, Equation 7.8 leads to sufficient statistics of the form

$$\underset{\sim}{X}' = (\sum_j \alpha_{j1}x_j, \ldots, \sum_j \alpha_{jm}x_j) \quad .$$

Note that since the α's are prespecified these are functions of data alone—not of parameters to be estimated. Rotational indeterminacy is eliminated by the fixed values of the factor loadings. Scaling indeterminacy is eliminated by Rasch's requirement of "specific objectivity," i.e., that the marginal likelihood $h(\underset{\sim}{x}|\underset{\sim}{\eta})$ be expressed in a form in which the person parameter $\underset{\sim}{\theta}$ can be separated from the item parameters $\underset{\sim}{\eta}$ as follows:

$$h(\underset{\sim}{x}|\underset{\sim}{\eta}) = \int_{\underset{\sim}{\theta}} f(\underset{\sim}{x}|\underset{\sim}{\theta},\underset{\sim}{\eta})g(\underset{\sim}{\theta})\ d\underset{\sim}{\theta}$$

$$= [\text{Prob}(\underset{\sim}{x}|\underset{\sim}{X},\underset{\sim}{\eta})] \times [\int_{\theta} p(\underset{\sim}{X}|\underset{\sim}{\theta},\underset{\sim}{\eta})g(\underset{\sim}{\theta})\ d\underset{\sim}{\theta}] \quad .$$

The only transformations to f and g that maintain this property are linear; hence, interval-scaled measurement is assured—at the cost of very strict assumptions about the form of f and the value of $\underset{\sim}{\alpha}$.

7.7 Confirmatory Factor Analysis, Multiple Group Solutions, and Structural Equations Modeling

The focus of this review has been on exploratory factor analysis; it is not known a priori how many factors are required to explain the data, much less their composition and interrelationships. Hypotheses about such matters may be entertained, however, and it proves useful to be able to fit common factor models under which certain parameter elements (factor loadings, unique variances, factor variances and covariances) are set to predetermined values or constrained to equal one another. By comparing chi-square indices of fit of competing models, one could then test hypotheses suggested by the content of the observed variables in light of psychological or sociological theories. Jöreskog (1969) describes maximum likelihood procedures by which this may be accomplished in the

setting of measured variables. Similar procedures have been developed for the setting of categorical variables by Gibbons (1984b), using ML, and by Muthén (1978), using GLS.

Gibbons (1984a) and Muthén and Christoffersson (1981), again working with ML and GLS respectively, perform confirmatory factor analysis over several examinee populations simultaneously. This work is also an extension of procedures developed by Jöreskog (1971) and Sörbom (1974) for measured variables. The interest here is in testing hypotheses about whether certain features of a common factor model can be taken as invariant across populations; e.g., whether factor loadings of items can be construed as invariant, suggesting a similar framework for approaching a questionnaire, while factor distributions and unique variances differ from one group to the next, suggesting varying population distributions and measurement precision.

Muthén (1979, 1984b) has extended Jöreskog's work in yet another area, namely that of modeling structural relationships among latent variables (Jöreskog, 1974, 1977; Jöreskog & Sörbom, 1984). Not only are latent variables $\underset{\sim}{\theta}$ posited to account for interrelationships among manifest variables, but relationships in the form of linear regression functions may be posited among latent variables. Analyses may consider several populations simultaneously, thus allowing for a wide variety of hypotheses

about the relationships of variables within and between groups to be studied.

8. Conclusion

Factor analyses of dichotomous data were first undertaken as a diagnostic tool in test construction. Deficiencies in available methods of analysis, mainly unweighted least squares factor analysis of phi coefficients or tetrachoric correlations, prevented these attempts from fulfulling their objectives satisfactorily. In particular, these problems included computational inaccuracies, failure of requisite assumptions, and lack of rigorous statistical foundation. Recent developments of generalized least squares (GLS) and maximum likelihood (ML) procedures have overcome these problems, albeit at the cost of heavier computational burden.

The developments reviewed here were intended to provide a conceptual framework and rigorous estimation procedures for the factor analysis of categorical data. They foreshadow two likely directions of future development.

The first is the extension beyond factor analysis; the models, concepts, and estimation procedures are clearly applicable to a much broader class of problems involving categorical data. Muthén's models for structural equations among latent variables for categorical observations, and Gibbon's (1981) longitudinal models for time-structured categorical data are cases in point.

The second stems from Muraki's analysis of factor analytic models in which guessing parameters are also estimated (Section 7.2). It gives one pause to realize that two models, distinct beyond rotation and holding different implications for test construction, offer nearly equally good fit to a given data set. The limitations of purely exploratory factor analyses in the classical tradition, when applied to categorical data—even after conceptual and estimation problems have been resolved—are apparent. Continued development can be expected, therefore, along lines that allow the researcher to incorporate prior information and scientific hypotheses into the process at the stage of modeling, rather than interpreting results from a minimally restrictive model. Initial efforts along this line from the sampling statistics perspective are exemplified by the confirmatory and structural equations models discussed in Section 7.7, and may be contemplated from the Bayesian perspective by the approach sketched in Section 7.

References

Anderson, T. W. (1959). An introduction to multivariate statistical analysis. New York: Wiley.

Bartholomew, D. J. (1980). Factor analysis for categorical data (with discussion). Journal of the Royal Statistical Society, Series B, 42, 293-321.

Bartholomew, D. J. (1984). The foundations of factor analysis. Biometrika, 71, 221-232.

Bartholomew, D. J. (1985). Foundations of factor analysis: Some practical implications. British Journal of Mathematical and Statistical Psychology, 38, 1-10.

Bock, R. D. (1984). Full information item factor analysis. Paper read at the 1984 meeting of the Psychometric Society. Santa Barbara, CA.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443-459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. Psychometrika, 35, 179-197.

Bock, R. D., & Muraki, E. (1984). Full information item factor analysis of the ASVAB power tests. Paper read at the 1984 meeting of the Office of Naval Contractors meeting, on model-based psychological measurement.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. South African Journal of Journal of Statistics, 8, 1-24. Reprinted in D. J. Aigner and A. S. Goldberger, (Eds.), (1977), Latent variables in socio-economic models. Amsterdam: North-Holland.

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items and between tests. Psychometrika, 26, 347-372.

Carroll, J. B. (1983). The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.), Principals of modern psychological measurement. Hillsdale, NJ: Erlbaum.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. Psychometrika, 40, 5-32.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Series B, 39, 1-38.

Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. Psychometrika, 44, 169-172.

Fienberg, S. E., & Holland, P. W. (1970). Methods for eliminating zero counts in contingency tables. In G. P. Patil (Ed.), Random counts on models and structures. University Park, PA: Pennsylvania State University Press.

Gibbons, R. D. (1981). The analysis of discrete time-structured data. Unpublished doctoral dissertation, University of Chicago.

Gibbons, R. D. (1984a). Multivariate probit analysis: A general model. Paper read at the 1984 meeting of the Psychometric Society, Santa Barbara, CA.

Gibbons, R. D. (1984b). MVPROBIT: A FORTRAN IV computer program for multivariate probit analysis [Computer program]. Chicago: University of Illinois.

Hambleton, R., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 14, 75-96.

Harman, H. H. (1976). Modern factor analysis (3rd ed.). Chicago: University of Chicago Press.

Heywood, H. B. (1931). On finite sequences of real numbers. Proceedings of the Royal Society, Series A, 134, 486-501.

Jensema, C. (1976). A simple technique for estimating latent trait mental test parameters. Educational and Psychological Measurement, 36, 705-715.

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

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-220.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2). San Francisco: Freeman.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: specification, estimation and testing. In P. R. Krishnaiah (Ed.), Applications of statistics. Amsterdam: North Holland.

Jöreskog, K. G. (1979). Basic ideas of factor and component analysis. In K. G. Jöreskog & D. Sörbom (Eds.), Advances in factor analysis and structural equation models. Cambridge: Abt.

Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. Psychometrika, 37, 243-259.

Jöreskog, K. G., & Sörbom, D. (1980). EFAP II: Exploratory factor analysis program [Computer program]. Chicago: Interational Educational Services.

Jöreskog, K. G., & Sörbom, D. (1984). LISREL: Analysis of linear structural relationships by the method of maximum likekihood [Computer program]. Chicago: Scientific Software.

Lawley, D. N. (1943). On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh, 61-A, 273-287.

Lawley, D. N. (1944). The factorial analysis of multiple item tests. Proceedings of the Royal Society of Edinburgh, 62-A. 74-82.

Lawley, D. N., & Maxwell, A. E. (1971). Factor analysis as a statistical method (2nd ed.). London: Butterworth.

Lee, S. Y. (1981). A Bayesian approach to confirmatory factor analysis. Psychometrika, 46, 153-160.

Lord, F. M. (1952). A theory of test scores. Psychometric Monograph, No. 7. Psychometric Society.

Lord, F. M. (1980). Applications of item response theory to to practical testing problems. Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Re iing, MA: Addison-Wesley.

Martin, J. K., & McDonald, R. P. (1975). Bayes estimates in restricted factor analysis: A treatment of Heywood cases. Psychometrika, 40, 505-517.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82-99.

Mislevy, R. J. (1984). Estimating latent distributions. Psychometrika, 49, 359-381.

Mislevy, R. J., & Bock, R. D. (1982). BILOG: Item analysis and test scoring with binary logistic models [Computer program]. Mooresville, IN: Scientific Software.

Mooijaart, A. (1983). Two kinds of factor analysis for ordered categorical variables. Multivariate Behavioral Research, 18, 423-441.

Muraki, E. (1983). Marginal maximum likelihood estimation for three-parameter polychtomous item response models: Applications of an EM algorithm. Unpublished doctoral dissertation, University of Chicago.

Muraki, E. (1984). Implementing full information factor analysis: The TESTFACT program. Paper read at the 1984 meeting of the Psychometric Society, Santa Barbara, CA.

Muraki, E. (1985). Full information factor analysis for polytomous item response. Paper read at the 1985 meeting of the American Educational Research Association, Chicago.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560.

Muthén, B. (1979). A structural probit model with latent variables. Journal of the American Statistical Association, 74, 807-811.

Muthén, B. (1984a). Multiple-group item factor analysis. Paper read at the 1984 meeting of the Psychometric Society, Santa Barbara, CA.

Muthén, B. (1984b). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Muthén, B. (1985). LISCOMP [Computer program]. Chicago: Scientific Software.

Muthén, B., & Christoffersson, A. Simultaneous factor analysis of dichotomous variables in several populations. Psychometrika, 46, 407-419.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. Psychometrika, 47, 337-347.

Pearson, K. (1900). On the correlation of characters not quantitatively measureable. Royal Society Philosophical Transactions, Series A, 195, 1-47.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. Psychometrika, 49, 425-435.

Samejima, F. (1982). Footnote on page 28 of Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. Evaluation plan for the Computerized Adaptive Vocational Aptitude Battery (Research Report 82-2). Baltimore, MD: Johns Hopkins.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

Stegelmann, W. (1983). Expanding the Rasch model to a general model having more than one dimension. Psychometrika, 48, 259-267.

Stroud, A. H., & Sechrest, D. (1966). Gaussian quandrature formulas. Englewood Cliffs, NJ: Prentice Hall.

Thissen, D. (1984). MULTILOG, Version 4.0 [Computer program]. Chicago: Scientific Software.

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

Wilson, D., Wood, R. L., & Gibbons, R. TESTFACT: Testing scoring and item factor analysis [Computer program]. Chicago: Scientific Software.

Winsberg, S., Thissen, D., & Wainer, H. (1982). Fitting item characteristic curves with spline functions. Paper read at the 1982 meeting of the Psychometric Society, Los Angeles.

Wright, B. D., & Stone, M. (1979). Best test design. Chicago: Mesa Press.

Footnote

[1]The exploratory nature of this use of factor analytic models, and the implicit expectation of subsequent use of item response models of similar forms, must be stressed here. That one unidimensional model of a specified parametric form will not fit a data set does not preclude the possibility that another unidimensional model of a different form will. If the question is whether the data can be explained in terms of any unidimensional monotonic latent variable model, with conditional independence, including ones quite different from the familiar and convenient IRT models in current use, then the nonparametric approach found in Rosenbaum (1984) is more appropriate.

Table 1

Numbers of Elements in GLS Factor Analysis

Number of variables	Number of elements in matrix of tetrachoric correlations	Number of elements in error covariance matrix
5	10	45
10	45	990
20	190	17,995
40	780	303,810
60	1770	1,565,565
80	3160	4,991,220
100	4950	12,248,775

Acknowledgment

The author is grateful to R. Darrell Bock, Robert Gibbons, Eiji Muraki, and Bengt Muthén for copies of materials from their 1984 Psychometric Society symposium. Many improvements to the original paper resulted from comments from Frederic Lord, Ledyard Tucker, an associate editor, and a referee.

Author: ROBERT J. MISLEVY, Research Scientist, Educational Testing Service, Princeton, New Jersey 08541. Specializations: item response theory, educational assessment