MicroFACT® 2.1

A Microcomputer Factor Analysis Program for Ordered Polytomous Data and Mainframe Size Problems

Niels G. Waller

January 22, 2003

The MicroFACT® computer program and the MicroFACT User's Guide are distributed exclusively by

Assessment Systems Corporation
2233 University Ave, Suite 200
St. Paul, MN 55114 USA
support@assess.com
www.assess.com
Voice: 651/647-9220
Fax: 651/647-0412

Contents

Preface

MicroFACT contains routines licensed from the Numerical Algorithms Group (NAG), Downers Grove, Illinois, 60515, phone (708) 971-2337, email: infodesk@nag.com.

System Requirements

IBM PC running Windows 3.1/95/98/NT/2000/XP (32 bit)
Requires approximately 5 MB of free space for installation.
MicroFACT uses virtual memory for temporary array storage. Thus, to run large problems your computer's hard disk must contain a sufficient amount of unused disk space.

Acknowledgments

The most recent version of MicroFACT has benefited from comments from the following reviewers: Irene D. Feurer (Vanderbilt University), Michael Finger (University of Minnesota), Matt Henson (UCLA), Robert Pruzek (State University of New York at Albany), Steven P. Reise (UCLA).

MicroFACT 2.*i* Revision History

01/16/03 version 2.1. Fixed minor bug in Copy/Cut/Paste procedure;
Fixed minor problem in triplet testing procedure
that could cause the program to end abruptly.

Part I

MicroFACT 2.1

Chapter 1

Introduction

The linear factor analysis model produces biased results when applied to polytomous data (i.e., ordered categorical data; Bernstein & Teng, 1989; McDonald & Ahlawat, 1974; Mislevy, 1986; Muthén, 1978, 1989; Wherry & Gaylord, 1944). Nevertheless, social scientists continue to use this model on phi coefficients (i.e., Pearson correlations from binary data) or Pearson correlations of Likert items because more appropriate methods or computer programs for categorical item factor analysis are either:

1. Not widely available (Fraser, 1986);
2. Restricted to confirmatory rather than exploratory analyses,
3. Restricted to binary items (Fraser, 1986); or
4. Restricted in the number of items or factors that can be efficiently handled (Bock & Aitkin, 1981; Christofferson, 1975; Muthén, 1987; Wilson, Wood & Gibbons, 1984).

These limitations pose methodological challenges to social scientists working in disciplines where ordered categorical items and datasets with large numbers of items are commonplace. In psychology, for example, personality, social and clinical test items are commonly scored *True/False, Agree/Disagree, Yes/No*; or they are scored on multicategory rating scales (e.g., *Disagree strongly, Disagree a little, Neither agree nor disagree, Agree a little, Agree strongly*). In education, achievement and ability items are typically scored right or wrong. In political science, people frequently vote for or against political referendums, whereas in marketing research, product quality is often assessed by consumer satisfaction/dissatisfaction.

For lack of readily available alternatives, investigators in these fields–more often then not–rely on software packages such as SPSS, SAS, BMDP or SYSTAT to correlate and factor analyze polytomously scored items. These software packages provide appropriate factor analysis routines for continuously measured

data. However, they can render seriously biased results with polytomous data (Bernstein & Teng, 1989; McDonald & Ahlawat, 1974; Mooijaart, 1983).

MicroFACT provides a methodologically defensible and computationally efficient method for polytomously scored item factor analysis in either large or small data sets. The program also handles continuously scored and ipsatized item response data. MicroFACT uses virtual memory for temporary array storage and handles data matrices of virtually any size (problem size is limited by a computer's unused hard disk space). In the next sections, alternative models for binary and polytomous item factor analysis are reviewed.

1.1 Alternative Factor Models for Binary Response Data

The linear factor analysis model potentially distorts the underlying structure of dichotomous data for at least three reasons. First, Pearson correlations computed on binary data (i.e., phi coefficients) are influenced by the shape of the item distributions as well as by the item content. Even when two perfectly reliable items measure identical traits they will not be perfectly correlated if their endorsement frequencies differ (Carroll, 1961; Lord & Novick, 1968, p. 347). Second, the common factor analysis model assumes that a linear function describes the regression of an item response on the underlying factor scores. Stated otherwise, it assumes that the probability of a keyed response is a linear function of the underlying factors. This assumption is violated with binary data. Probabilities are bounded by 0.0 and 1.0, yet people with very low or very high factor scores can have endorsement probabilities that are less than 0.0 or greater than 1.0 in the linear model. Third, the dimensionality of a matrix of phi coefficients may differ from the dimensionality of the underlying continuous variables (Bernstein & Teng, 1989; Hambleton & Rovinelli, 1986). This occurs because the shape of the regression of an observed dichotomous response variable on an underlying factor is often nonlinear. McDonald & Ahlawat (1974) demonstrated that the linear factor analysis model can produce spurious factors when the shape of the nonlinear item-factor regressions differ markedly over items. In the past, these "arti-factors" were attributed to non-uniform item difficulties (i.e., differential item endorsement frequencies), and thus were labeled difficulty factors (Ferguson, 1941). To avoid so-called "difficulty factors," psychometricians have developed alternative procedures for elucidating structure in binary item data matrices. Three of these alternatives are:

1. Linear factor analysis of tetrachoric correlations;
2. Generalized least squares factor analysis of tetrachoric correlations; and
3. Full-information item factor analysis.

A brief discussion of these methods follows.

1.1.1 Linear Factor Analysis of Tetrachoric Correlations

An early solution to the binary response data problem was to perform a linear factor analysis on a matrix of tetrachoric correlations (LFATC) rather than on a matrix of phi coefficients. This approach assumes that the observed binary-scored variables are realizations of latent continuous variables that have been categorized during measurement. When the underlying distribution for the latent scores is bivariate normal, the tetrachoric correlation is the maximum likelihood estimate of the Pearson correlation among the latent responses (Divgi, 1979; Pearson & Heron, 1913).

This approach has several limitations (some of which can be surmounted by procedures outlined later). Notably, tetrachoric correlations are not sufficient statistics for the item response patterns (Bock & Lieberman, 1970). In typical applications, tetrachoric correlations are computed in a pairwise fashion, with only joint occurrence frequencies considered during estimation of the correlations. In other words, a tetrachoric correlation matrix is computed two variables at a time. Consequently, tetrachoric correlations do not utilize all of the information available in the $2n$ possible item response patterns that can be formed from n dichotomously scored items. For this reason, a factor analysis of tetrachoric correlations is called a limited-information model for binary item data analysis (Bartholomew, 1980).

1.1.2 Generalized Least Squares Factor Analysis of Tetrachoric Correlations

Generalized least squares factor analysis of tetrachoric correlations (GLSTC) uses more information than LFATC, although it is also a limited-information model. Unlike the simpler approach described above, GLSTC considers large sample variances and covariances of the estimated tetrachoric correlations during model fitting. Correlations with small standard errors are weighted more heavily than correlations with large standard errors when assessing the discrepancy between observed and model-implied correlations. Because the weight matrix of GLSTC includes estimated covariances among the correlations, the method considers four variables at a time. When more than four variables are modeled, GLSTC is a limited information procedure. Basic theory for GLSTC with tetrachorics was provided by Muthén (1978; see also Christofferson, 1975) and implemented in the LISCOMP computer program (Muthén, 1987).

In theory, GLSTC provides an elegant means of modeling binary data. In practice, the method has limitations. The asymptotic covariance matrix for the tetrachoric correlations becomes prohibitively large in moderately sized problems, and current implementations of the method (Muthén, 1987) are limited to approximately 20-25 variables on a personal computer. Moreover, the asymptotic weight matrix is often nonpositive definite (and therefore not invertable) in realistically sized samples. Other times it is poorly estimated, yielding biased parameter estimates. Muthén cautions that use of the GLS estimator for dichotomous variables "...should probably not be attempted when the number

of variables exceeds much more than 30 variables, and... to estimate the weight matrix properly with many variables, large samples are required–at the very least, 1,000 observations when more than 10 variables are present" (1989, p. 25)

1.1.3 Full-Information Item Factor Analysis

The most statistically efficient method for item-level factor analysis is the full-information factor analysis model (FIFA) developed by Bock and Aitkin (1981), implemented by Wilson, Wood and Gibbons (1984) in the program TESTFACT (see also Bock, Gibbons & Muraki, 1988, for a recent discussion) and advanced by Schilling (1993). This approach differs from the former models in that it does not analyze observed or latent item correlations. Rather, full-information factor analysis is conceptually similar to the normal ogive (or logistic) item response theory model (Lord, 1980). Parameters are estimated by marginal maximum likelihood and the EM algorithm (Dempster, Laird & Rubin, 1977), resulting in parameter standard errors and a chi-square test of model fit. This approach uses information from all unique item response patterns in the data matrix. Consequently, the method is computationally intensive. Bock notes that "the practical limit of the number of factors is five . . . while 60 to 100 items is not excessive for a fast computer" (Bock, Gibbons & Muraki, 1988, p. 262). Unfortunately, these limitations prohibit the use of full-information factor analysis on many datasets. For example, several widely used personality assessment questionnaires, such as the MMPI (Hathaway & McKinley, 1983), MPQ (Tellegen, 1982), CPI (Gough, 1987) or ACL (Gough & Heilbrun, 1983) include 300 to 600 items each, and contain many more than five latent factors. A further limitation of full-information factor analysis is that the chi-square test of model fit is highly inaccurate when models include eight or more variables in datasets of 1,000 people or less (Reiser & VandenBerg, 1994).

1.2 Binary Item Factor Analysis with MicroFACT

In a recent large-scale Monte Carlo study, Knol and Berger (1988, 1991) compared the aforementioned procedures for binary item factor analysis. The authors found that "a common factor analysis on the matrix of tetrachoric correlations yields the best estimates" (Knol & Berger, 1988, p. 1). "Best" was defined in this study as the method that yields the smallest mean-squared error of parameter recovery. Surprisingly, the less complex iterated principal factor analysis (IPFATC) model on tetrachoric correlations outperformed GLSTC, FIFA, and a multidimensional latent trait model (McKinley & Reckase, 1983). Moreover, the performance of IPFATC improved as the number of dimensions and items in the data increased. Other investigators have found that a simple factor analysis of tetrachoric correlations yielded parameter estimates that are very close to the full-information estimates. This is particularly true when the methods are compared in large item pools (Schilling, 1993). For these

reasons, MicroFACT performs iterated (and noniterated) principal factor (and components) analysis on tetrachoric, polychoric or Pearson correlation matrices. Because MicroFACT uses virtual memory for temporary array storage it easily handles data sets with 500 to 1,000 items and an unlimited number of people.

1.2.1 Computation of Tetrachoric Correlations

MicroFACT computes tetrachoric correlations using procedures outlined in Divgi (1979). Correlations computed from frequency tables with one or more cell counts below five are corrected for bias using a procedure discussed in Brown and Benedetti (1977). Tetrachoric correlations in the presence of guessing (*i.e.,* nonzero, lower asymptotes on the item response functions) are computed with a modification due to Carroll (1945). On request, MicroFACT also computes the smoothed tetrachoric correlation matrix. A smoothed tetrachoric correlation matrix is a positive semidefinite least-squares approximation of the original tetrachoric correlation matrix. Smoothing the tetrachoric correlation matrix prior to performing a factor analysis has been found to reduce the number of Heywood cases (i.e., communalities that are greater than 1.00) in the factor solution (Knol & Berger, 1988). The smoothing procedure is as follows:

(1.1) Compute an eigendecomposition of the initial matrix

$$\Sigma^* = KDK' \tag{1.1}$$

where **K** is an ordered matrix of eigenvectors and **D** is the corresponding diagonal matrix of ordered eigenvalues of Σ.

(1.2) The least squares approximation of Σ is given by

$$\Sigma^* = \left[Diag\left(KD^+K\right)\right]^{-1/2} KD^+K' \left[Diag\left(KD^+K'\right)\right]^{-1/2} \tag{1.2}$$

where elements of D^+, δ^+, are the ordered and rescaled eigenvalues that are greater than a small constant (0.005 in MicroFACT).

1.2.2 Testing the Assumptions Underlying Tetrachoric Correlations

An important assumption of tetrachoric correlations is that the underlying (i.e., latent) distribution of the binary item data matrix is multivariate normal. It is often difficult or impossible to test this assumption in data sets with more than a trivial number of variables. Nevertheless, it is possible to test for latent, *trivariate* normality using a triplet testing procedure due to Muthén and Hofacker (1988). MicroFACT can perform this test by specifying the TRP option on the MODEL command line. Example 9 illustrates the triplet testing procedure in MicroFACT. Reise and Waller (2001) provide an empirical example.

1.2.3 Computation of Polychoric Correlations

Generalization of the tetrachoric correlation to multicategory data was first discussed by Ritchie-Scott (1918), and later by Pearson and Pearson (1922). More recently, statistical theory and computational formulae for the polychoric correlation have been provided by Lancaster and Hamdan (1964) and Olsson (1979a). MicroFACT computes two-stage maximum likelihood estimates of the polychoric correlation using equations reported in Olsson (1979a). The two-stage estimator differs from the one-stage estimator in that in the former method the item thresholds are estimated prior to the estimation of the polychoric correlations, whereas in the latter method the item thresholds and correlations are estimated simultaneously. Olsson (1979a) reported that the two methods yielded results that are virtually identical for practical purposes. MicroFACT calculates two-stage estimates because they are computationally more efficient. Polychoric correlations can be computed from data with a maximum of 15 ordered categories for each variable. The number of ordered categories can vary among variables.

Chapter 2

Running MicroFACT 2.1 in DOS

To run MicroFACT in DOS mode, simply double click on the MicroFACT icon (in Windows) or type MFACT2 at the DOS command prompt.

During execution, MicroFACT prompts you for an input file name. You must include the extension (e.g.., type EX1.IN to analyze the first example input file). The input file must be prepared in an ASCII editor or saved as a pure ASCII or text file.

Example:

Listing of sample input file: EX6.IN

```
* This example demonstrates a MicroFACT job listing for
* 26 measurements from Thurstone's box problem.
* The correlation matrix is taken from: Cureton, E. E.,
* & Mulaik, S. A. (1975). The weighted
* varimax rotation and the promax rotation.
* Psychometrika, 40, 183-195.
* Note that the correlation matrix is not Gramian.
* Thus, squared multiple correlations are computed by
* first including a ridge constant on the correlation
* matrix as recommended by Finkbeiner & Tucker (1982).
>TITLE: Thurstone 26 box
>DATA: box.dat 26 30
>OUTFILE: box.out Short
>METRIC: CO COR NS
>MODEL: PFA 3 Labels NoSort
>ROTATE: WVARIMAX
```

```
>SCORE: NOSCORE
>PLOT: LOADINGS
>TECHNICAL: .0001 200 .00001 200 F .10 SMC F
>FORMAT:
>LABELS:
' x'
' y'
(more labels go here)
```

Keywords

*Comment Lines: (optional)
>TITLE: (required)
>DATA: (required)
>OUTFILE: (required)
>PHI: (optional)
>MODEL: (required)
>ROTATE: (required)
>SCORE: (required)
>PLOT: (required)
>TECHNICAL: (optional)
>OUTPUT: (optional)
>FORMAT: (required)
>LABELS: (optional)

Keywords and command statements can be entered in upper or lower case.

2.1 *Comment Lines: (optional)

MicroFACT input files can contain an unlimited number of introductory comment lines. Comment lines must begin with an asterisk.

2.2 >TITLE: (required)

All MicroFACT input files begin with the >TITLE: keyword. Titles can be up to 80 characters long.

2.3 >DATA: (required)

Format: `>DATA: InFILENAME {character} NVAR {integer} NSUB {integer}`

The >DATA: line must contain three entries: (1) a FILENAME that specifies the name and location (path) of the input data file, (2) NVAR, an integer equal to the number of variables to be included in the analysis (note that the required ID variable is not counted in this number), and (3) NSUB, the number of subjects included in the data file (or if reading correlations or factor loadings, NSUB equals the number of subjects used in calculating the correlations or factor loadings). Path statements can be included with the filename. Variables are separated by spaces

InFILENAME	Name of input data file.
NVAR	Number of variables in analysis (ID not counted).
NSUB	Number of subjects in analysis.

Example 1:

```
>DATA: MMPI2.DAT 567 5313
```

Read data from file MMPI2.DAT. The file contains observations on 567 variables (and a person ID code) from 5313 people.

Example 2:

```
>DATA: C:\TESTDATA\LSAT6.DAT 5 1000
```

Read file LSAT6.DAT from directory TESTDATA. The file contains 5 variables (and a person ID code) from 1000 people.

Example 3:

```
>DATA SMCORMAT 30 500
```

SMCORMAT denotes the name for a previously smoothed and saved correlation matrix. This example demonstrates that MicroFACT can read Pearson product-moment, tetrachoric or polychoric correlation matrices from previous analyses.

2.4 >OUTFILE: (required)

Format: `>OUTFILE: OutFILENAME {character} Print_Verbosity {SHORT, LONG}`

MicroFACT results will be written to the file specified in OutFILENAME. DOS path statements are allowed. The second entry on the >OUTFILE: line controls the amount of output that will be printed in the file named in OutFILENAME.

OutFILENAME Name of output data file.

Print_Verbosity Amount of technical detail in output file.

- SHORT Print standard amount of technical detail.
- LONG Print everything (e.g., input or calculated correlation matrix and matrix of residuals).

Example 1:

```
>OUTFILE: RESULT.OUT SHORT
```

Example 2:

```
>OUTFILE: C:\TESTDATA\LSAT6.OUT LONG
```

2.5 >METRIC: (required)

Format: `Variable_Metric {CO, DI, OR, CD, IP} Data_Type {RAW, COR, FLD1,FLD2} Smooth_Option {SM, NS, PS}`

The >METRIC: command specifies how the data are to be analyzed. It expects three entries.

The first entry, Variable_Metric, tells MicroFACT whether the data are continuous, dichotomous, or ordered polytomous (i.e., ordered categorical). MicroFACT expects one of the following keywords:

CO if data are continuous. Correlations will be Pearson product-moment correlations.

DI if data are dichotomous. Correlations will be tetrachoric correlations.

OR if data are ordered polytomous. Correlations will be polychoric correlations.

CD if continuous data are to be dichotomized (at medians). Correlations will be tetrachoric correlations.

IP if continuous data are to be ipsatized (i.e., person centered) before calculating the Pearson-product moment correlations.

The second entry (Data_Type) on the >METRIC: command line informs MicroFACT whether the data are raw responses, correlations, or factor loadings. MicroFACT expects one of the following keywords:

RAW if the data file contains item response data (i.e., binary if the data are dichotomous, otherwise the actual responses).

COR if the data file contains an $N \times N$ correlation matrix (note: MicroFACT reads correlation matrices only in free-field format).

FLD1 if the data file contains an $N \times k$ matrix of factor loadings (where k equals the number of common factors) from an orthogonal factor model.

FLD2 if the data file contains an $N \times k$ matrix of factor loadings (where k equals the number of common factors) from an oblique factor model. Note that if the data file contains factor pattern coefficients (i.e., loadings from an oblique solution), the >METRIC: command line must be followed by the >PHI: command line.

The **FLD1** and **FLD2** options have been included so that published factor solutions or factor solutions from previous analyses can be rotated to alternative positions.

The third entry (Smooth_Option) on the >METRIC: command line informs MicroFACT whether to smooth the correlation matrix. MicroFACT expects one of the following keywords:

SM to smooth the correlation matrix.

NS do not smooth the correlation matrix.

PS the correlation matrix has been previously smoothed.

This keyword is ignored when the input data matrix contains factor loadings. Nevertheless, one of the above keywords must be included in all MicroFACT job listings.

Example 1:

```
>METRIC: CO RAW NS
```

Compute Pearson product-moment correlations from raw data.

Example 2:

```
>METRIC: DI RAW SM
```

Compute a smoothed tetrachoric correlation matrix from raw data.

Example 3:

```
>METRIC: OR COR PS
```

Polychoric correlations are analyzed. The correlation matrix was computed and smoothed in a previous analysis.

Example 4:

```
>METRIC: OR COR SM
```

Polychoric correlations are analyzed. The polychoric correlation matrix was computed but not smoothed in a previous analysis. MicroFACT will smooth (and save) the matrix in this run.

Example 5:

```
>METRIC: CO FLD1 NS
```

The input data matrix contains factor loadings from an orthogonal factor analysis.

Example 6:

```
>METRIC: CO FLD2 NS
```

The input data matrix contains factor loadings from an oblique factor analysis (i.e., factor pattern coefficients).

2.6 >PHI: (optional)

The >PHI: line is used to specify the filename for the factor correlation matrix. This line is used when Data_Type (on the >METRIC line) equals FLD2.

Example 1:

```
>PHI: c:\MF\phi.dat
```

2.7 >MODEL: (required)

Format: `>MODEL: Model_Option {CMT, PCA, PFA, ROT} NFAC {integer} LAB_Option {LABELS, NOLABELS} SORT_Option {SORT, NOSORT}`

The >MODEL: command informs MicroFACT how the data are to be analyzed. It expects four entries.

The first entry, Model_Option, describes the model. MicroFACT expects one of following keywords:

CMT to compute a correlation matrix (only) from raw data

PCA to specify the principal components model

PFA to specify the (possibly iterated) principal axes factor analysis model

ROT to specify that an input (possibly oblique) factor loading matrix should be rotated to an alternative position.

TRP to specify Muthén and Hofacker's triplet testing of trivariate normality.

The second entry, NFAC, describes the number of dimensions (i.e., components or factors)

The third entry, LAB_Option, informs MicroFACT whether user-supplied variable labels are provided in the input file. MicroFACT expects one of the following keywords:

LABELS if user-supplied labels are provided
NOLABELS if labels are not supplied

The fourth entry, SORT_Option, informs the program whether to sort the factor loadings. MicroFACT expects one of the following keywords:

SORT if variables are to be sorted by factor. (Note: only rotated factor loadings are sorted in this version of MicroFACT.)

NOSORT if original variable order is maintained

Example 1:

```
>MODEL: PFA 5 LABELS SORT
```

Extract five factors from the data matrix, sort the factor loadings, and print variable labels in the output file.

Example 2:

```
>MODEL: CMT 0 LABELS NOSORT
```

Compute the appropriate correlation matrix as specified in the >Metric: command.

Example 3:

```
>MODEL: ROT 3 LABELS NOSORT
```

Three (possibly oblique) common factors from a previous solution will be rotated to an alternative position.

2.8 >ROTATE: (required)

Format: `>ROTATE: ROTATION {NONE, VARIMAX, WVARIMAX, PROMAX i, Oblimin` Δ`, HK r}`

The >ROTATE: command is used to select the factor rotation option. Available choices include:

NONE no rotation is performed

VARIMAX a row standardized Varimax (Kaiser, 1958) rotation is performed.

WVARIMAX a weighted Varimax (Cureton & Mulaik, 1975) rotation is performed.

PROMAX i both Promax and Varimax rotations are performed. When Promax is chosen the user must specify an integer, i, that denotes the power to be used in computing the target pattern. Integer-valued powers between 2 and 6 are allowed. If you are unfamiliar with this option, you should choose 3.

OBLIMIN Δ Direct Oblimin (Jennrich & Sampson, 1966). The Oblimin keyword must be followed by a delta parameter (Δ) that controls the degree of factor obliqueness. A requirement is that $\Delta \leq 0$. Settings of 0, -$\frac{1}{2}$, and -1 correspond to Direct Quartimin, Direct Biquartimin, and Direct Covarimin.

HK r Harris-Kaiser Ortho-Oblique. When the Harris-Kaiser option is chosen, the user must specify a positive real-valued number, r, that controls solution obliqueness. Values of r can range from 0 to 1.00. A value of 1.00 results in an orthogonal solution. Harris and others have reported that a value of .5 seems to work well in many situations.

Example 1:

```
>ROTATE: PROMAX 3
```

Perform Varimax and Promax rotations with target pattern loadings raised to the 3rd power.

Example 2:

```
>ROTATE: HK .5
```

Perform a Harris-Kaiser ortho-oblique rotation

2.9 >SCORE: (required)

Format: `>SCORE: {NOSCORE, REG, BART}`

The >SCORE: command is used to request factor scores. In MicroFACT 2.1, factor scores are computed only for continuous or ipsatized data. The following choices are available:

NOSCORE no factor scores are computed

REG Thurstone's regression method factor scores are computed

BART Bartlett method factor scores are computed

If the program runs successfully, factor scores will be written to a file named: FSCORE.OUT. The first column of this file contains the person ID.

Example 1:

```
>SCORE: BART
```

Compute Bartlett method factor scores.

2.10 >PLOT: (required)

Format: `>PLOT: {NONE, SCORES, LOADINGS, ALL}`

The >PLOT: command is used to request a variety of plots. The following choices are available:

NONE No plots will be created.

SCORES Bivariate plots of factor scores. These plots should *always* be examined whenever binary or otherwise polytomous data are analyzed as if they were continuous. Factor score plots can help determine whether so-called "nonlinear factors" (Etezadi-Amoli. & McDonald, 1983; McDonald, 1967, 1969) are present.

LOADINGS Bifactor plots of factor loadings. Note that the factor axes will be plotted at 90°angles even when the factor pattern is from an oblique solution. Factor axes are always plotted at right angles to enhance readability of the factor hyperplanes.

ALL Create all available plots.

Example 1:

```
>PLOT: SCORES
```

Generate bivariate plots of factor scores.

Example 2:

```
>PLOT: LOADINGS
```

Generate bivariate plots of factor loadings.

2.11 >TECHNICAL: (optional)

Format: `>TECHNICAL: COM_EPS {real, default=.0001}, MAXIT {integer, default=200}, ROT_EPS {real, default=.00001}, ROT_MAXIT {integer, default=200}, RNDSTRTS {logical, T {default} or F}, HYP_WIDTH {real, default=.20}, INIT_COM {character, SMC, HIR, ONE}, PAGEWIDTH {integer, default=120}, Output {logical, T or F {default}}`

The >TECHNICAL: optional command line contains 9 entries that provide finer control of program execution and program output. If the >TECHNICAL command line is omitted then MicroFACT 2.1 uses program defaults (specified below). **It is important to note that if the user wishes to changes *any* of the default settings, then all of the technical settings must be specified on the >TECHNICAL command line.**

COM_EPS Convergence criterion for the iterated communality estimation {default=.0001}. Communality estimation is terminated whenever the sum of the communalities ($\sum_{j=1}^{nvar} h_j^2$) fails to change by more than COM_EPS across two consecutive iterations. In other words, iterations stop when: $\sum_{j=1}^{nvar} h^2_{j(iter\ i+1)} - \sum_{j=1}^{nvar} h^2_{j(iter\ i)} \leq$ COM_EPS.

MAXIT This integer-valued parameter controls the maximum number of iterations during communality estimation{default=200}. Its primary purpose is to allow users to run models with noniterated communalities estimates (by setting MAXIT=1), such as squared multiple correlations.

ROT_EPS This real-valued parameter sets the convergence criterion for the rotation step {default=.00001}.

ROT_MAXIT This integer-valued parameter determines the maximum number of iterations for the rotation step {default=200}.

RNDSTRTS This logical-valued parameter determines whether the rotation step will begin from the principal axes position or from an orientation that is determined by a random spin {**T** =random spin, or **F** = principal axes}. This option is included because of the tendency of rotation algorithms to find local minima (Rozeboom, 1992; Ten Berge, 1995) {default = T}.

HYP_WIDTH This real-valued parameter controls the hyperplane width {default=.20}.

INIT_COM This character-valued parameter sets the initial method for communality estimation. Initial communalities can be calculated as squared multiple correlations (SMC) , highest off-diagonal correlations (HIR), or unities (ONE). When the input or computed correlation matrix is singular, MicroFACT computes squared multiple correlations by the method described by Finkbeiner and Tucker (1982).

PAGEWIDTH This entry controls the number of columns on the printed page. Maximum PAGEWIDTH is currently 120 columns.

OUTPUT This logical (T or F) parameter is used to notify MicroFACT 2.1 that one or more output matrices are to be written to an external file. The various output options are specified on the >OUTPUT: command line. If the OUPUT parameter is set to T the following input line must begin with the >OUTPUT: keyword.

Example 1:

```
>TECHNICAL: .0001 1 .00001 200 F .20 SMC 80 F
```

Use noniterated communality estimates with squared multiple correlations. Rotate axes from principal axes orientation.

Example 2:

```
>TECHNICAL: .0001 200 .00001 200 T .20 SMC 130 T
```

Use iterated communality estimates. Perform one (orthogonal) random spin before factor rotation. Output (to be specified on the >OUTPUT: line) will be written to external files.

2.12 >OUTPUT: (optional)

Format: `>OUTPUT: NUM_OUTPUT {integer} COR SMO EIG FAC RSD`

The >OUTPUT: command line allows users to write various output matrices to external files. One or more of the following keywords are used to save MicroFACT 2.1 output:

COR The computed (Pearson, Tetrachoric, or Polychoric) correlation matrix.

SMO The smoothed correlation matrix.

EIG The eigenvalues of the (possibly smoothed) correlation matrix.

FAC The rotated or unrotated factor pattern matrix (if an oblique solution is requested, the factor correlation matrix will be saved in a file with the .phi extension).

RSD The matrix of model residuals.

!!! Note that the first entry on the >OUTPUT: command line is an integer (NUM_OUTPUT) specifying the number of output files to be created.

The output files created by the above keywords are easily identified by their three-letter extensions. For example, suppose that the input file is named MyInput.in. Further suppose that the >OUTPUT: command line contains the following keywords: COR, EIG, and FAC. MicroFACT will write the appropriate results to files called: MyInput.cor, MyInput.eig, and MyInput.fac.

Example 1:

```
>OUTPUT: 3 COR EIG FAC
```

Write the three output matrices to files with extensions .cor, .eig, and .fac.

2.13 >FORMAT: (required)

The >FORMAT: command line must contain a legitimate, fixed field FORTRAN FORMAT statement. The first nonblank parameter (e.g., I5 as contrasted to 3X) must refer to a person ID. The ID must be an integer of at most 15 digits. The entire format statement is limited to 80 columns (i.e., one line). Correlations are read in free-field format only. When the data matrix contains correlations, an asterisk (*) can be used to denote free-field format.

The parameters for the FORMAT statement must be separated by commas and enclosed in parentheses. The following parameters, or descriptors, may be used:

In Read the next n digits in the data as an integer

Fn.m Read the next digits as an n-column floating-point number, with m digits after the decimal

nX Skip the next n digits

The first two descriptors, In and F$n.m$, can be preceded by a number representing the number of times each command should be performed. See the example below.

Example 1:

The data matrix contains a 5-column person ID and item responses for 12 binary items:

```
10001 101101110111
10002 001101010001
10003 101001000111
```

Suppose only items 1-5, 7-9, and 12 are to be included in the factor analysis. A possible format statement would be:

```
>FORMAT: (I5,1X,5F1.0,1X,3F1.0,2X,1F1.0)
```

Where:

I5 means that the ID code is five columns in length

1X means skip the next column

5F1.0 means that each of the next five variables is one column long and there are no decimal places

1X means skip the next column

3F1.0 means that each of the next three variables is one column long without decimal places

2X means skip two columns

1F1.0 means that the next variable is one column long with no decimal place

2.14 >LABELS: (optional)

The >LABELS: command line is the only optional command line in MicroFACT. If the keyword LABELS is present on the >MODEL: command line, MicroFACT expects N variable labels, enclosed in single quotes, to follow the >LABELS: keyword. The length of each label is limited to 15 characters.

2.15 Using MicroFACT 2.1 in Monte Carlo simulations

MicroFACT 2.1 has been enhanced to allow users to run monte carlo simulations. When MicroFACT 2.1 is called from an external program, input files can be specified by placing an input filename directly after MFACT2.EXE in a call to the program. For instance:

```
SHELL(''c:/programs/mfact/MFACT2.EXE MYINPUT.IN'')
```

Desired output (e.g., tetrachoric correlation matrices, factor pattern matrices) can be specified on the >OUTPUT: command line.

Chapter 3

Displayed Output

3.1 Displayed Output includes:

Command Summary and Preliminary Statistics:

A reproduction of the Command File

A description of the Model Options

A description of the Technical Options

Input or computed correlation matrix

Eigenvalues of the (possibly smoothed) correlation matrix

The Scree Plot of Eigenvalues

An Analysis of Model Fit Including:

Number of iterations during communality estimation

A GFI index $= 1.0 -$ mean-squared residual/mean-squared$r$
(see McDonald, 1999, p. 83).

Residual covariance matrix

Mean squared residual

Root Mean Squared Residual

Mean residual

Standard deviation of residuals

Coefficient of skewness of residuals

Coefficient of kurtosis of residuals

5-point summary of residuals

Residual distribution plot

Unrotated Factor Loadings

Variance explained by each unrotated factor

Final communality estimates

Rotated (and possibly sorted) Factor Pattern Matrix

Rotated (and possibly sorted) Factor Structure Matrix

Factor Correlation Matrix

Factor Hyperplane Count

Factor Score Coefficient Matrix

Correlations of Factors with Factor Score Estimates

Bivariate Plot of Factor Scores

Bivariate Plot of Factor Loadings

3.2 Extended Output for Dichotomous Item Factor Analysis Models

When estimating the parameters of dichotomous item factor analysis models, MicroFACT 2.1 produces three additional output files with information that can be used to judge the quality of the model output. These files have one of the following file extensions:

DFF

FRQ

SER

DFF Files with the DFF extension report classical item difficulties for dichotomous items. Items with extreme difficulty values (e.g., $p > .95$ or $p < .05$) may be associated with poorly estimated tetrachoric correlations unless the sample size is extremely large. Example output from EX1B.DFF is reproduced below:

```
1 0.5936
2 0.5033
3 0.2357
4 0.6112
5 0.2797
6 0.2159
7 0.4967
.
.
.


29 0.4659
30 0.2720
```

FRQ Files with the FRQ extension report the bias corrected cell frequencies, the Pearson and Tetrachoric correlations, and the (maximum likelihood estimate of the) standard error of the tetrachoric correlation for each 2 x 2 table calculated from the data. Warnings are issued when cell values fall below 5.00 for any cell in any 2 x 2 table. Example output from EX1B.FRQ is reproduced below:

Bias Corrected Cell Frequencies and Correlations

Variables 1 and 2

	0	1	
1	191.00	348.00	r(tet) = 0.520 std err 0.042
0	260.00	109.00	r(phi) = 0.344

Variables 1 and 3

	0	1	
1	364.00	175.00	r(tet) = 0.456 std err 0.051
0	330.00	39.00	(phi) = 0.253

Variables 2 and 3

```
     0        1
1  304.00   153.00     r(tet) = 0.402 std err 0.051
0  390.00   61.00      r(phi) = 0.235
```

SER Files with the SER extension report the maximum likelihood standard errors of the tetrachoric correlations. These standard errors have been assembled into a matrix. Diagonal values of this matrix equal zero. Example output from EX1B.SER is reproduced below:

```
0.000000000 0.042185450 0.050677994 0.044689695
0.042185450 0.000000000 0.051039511 0.047094375
0.050677994 0.051039511 0.000000000 0.053260763
0.044689695 0.047094375 0.053260763 0.000000000
```

Part II

*Win*MFACT2

Chapter 4

Running MicroFACT 2.1 with *Win*MFACT2

*Win*MFACT2 is a graphical user interface, front-end for MicroFACT that simplifies the process of generating and running command files, and viewing and editing output files. *Win*MFACT2 is installed automatically when users install MicroFACT 2.1. To run *Win*MFACT2, simply double click the program icon:

Figure 4.1: *Win*MicroFACT Icon

Previously created MicroFACT 2.1 command files can by executed in *Win*MFACT2 by selecting `Open` and `Run` under the `File` menu in the *Win*MFACT2 document window. If the commands run successfully, the job output will be loaded automatically into the *Win*MFACT2 document window. The document window includes `Cut` and `Paste` functions, under the `Edit` menu, that can be used to edit the output prior to printing.

*Win*MFACT2 can also be used to build MicroFACT 2.1 command files using a series of point-and click menus. To generate a command file, choose `Build` under the `File` menu (see Figure 4.2) .

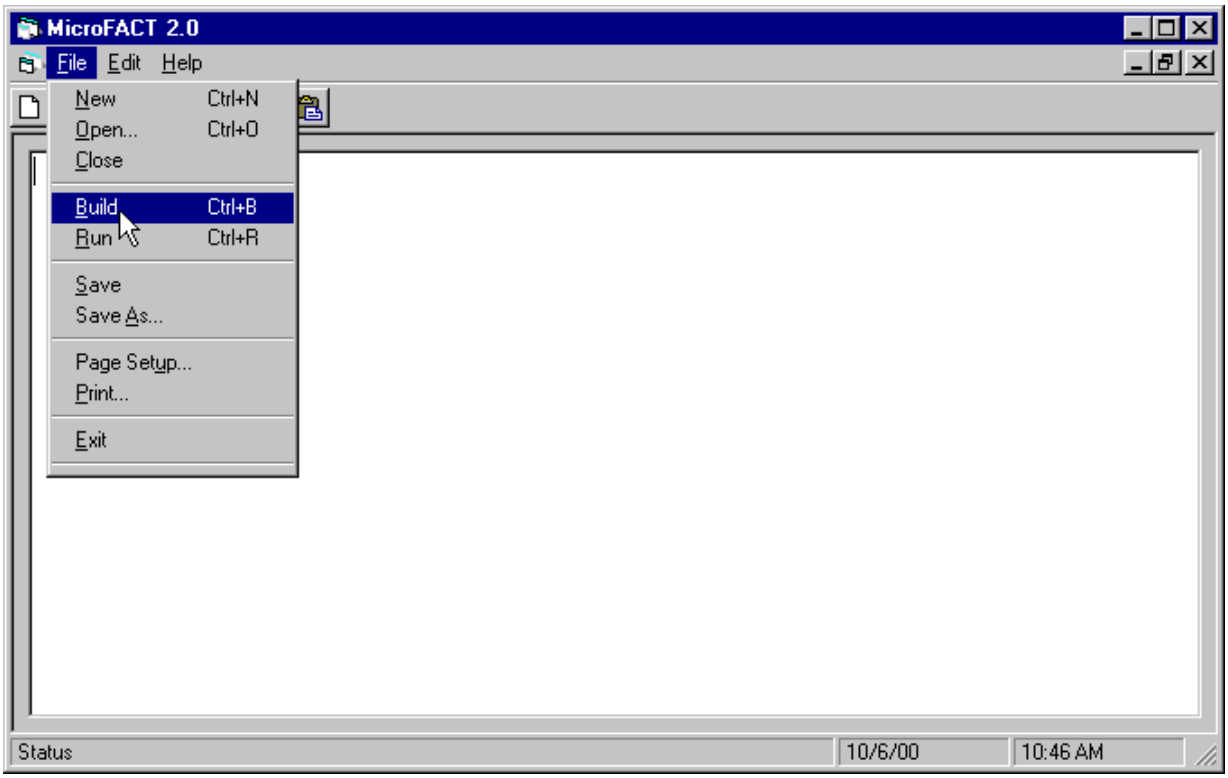

Figure 4.2: The Build Menu

After choosing `Build`, *Win*MFACT2 responds by displaying a new window (see Figure 4.3) with 13 command buttons. Ten of these buttons open additional forms that present users with various options for creating input files.

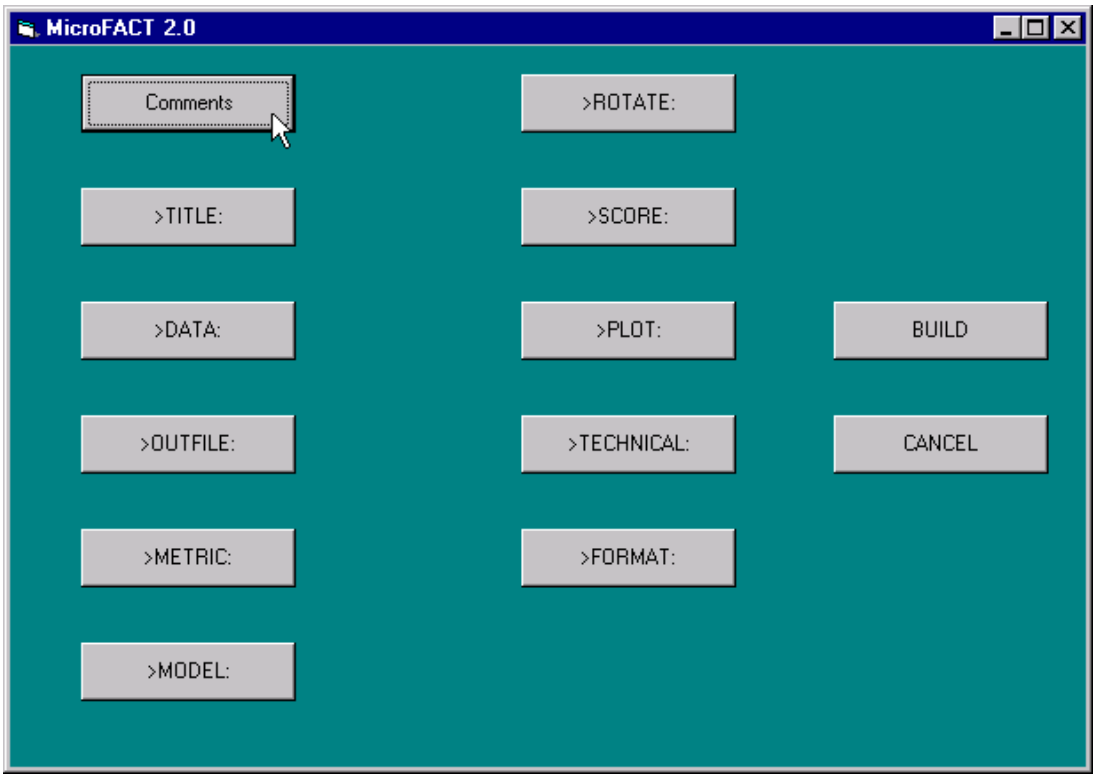

Figure 4.3: The Build Menu

Although the various command buttons can be selected in any order, it is a good practice to select them sequentially by first choosing the `Comments` button and working through the list of options. After an option is chosen, a yellow asterisk will appear next to the command button to indicate that the command has been completed (see Figure 4.4).

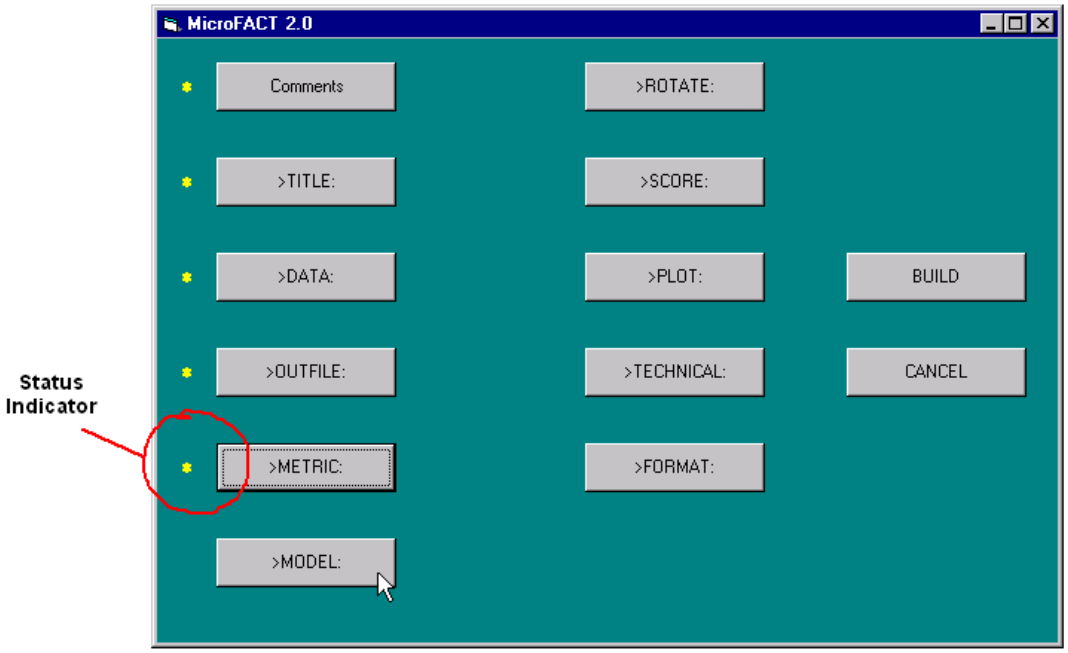

Figure 4.4: The Build Menu

When all status indicators are visible, *Win*MFACT2 has sufficient information to build a command file. To build the command file, single-click the `Build` command button.

IMPORTANT NOTE: At this point, do not try to run your command file by selecting `Run` under the `File` menu. Doing so will generate a program Warning Message. It is necessary to first save your job control statements in an external file (see Figure 4.5) To run a command file, you must `Open` the previously saved file before selecting `Run`. Whenever a file is modified, it must be saved and then re-opened before submitting the file to MicroFACT 2.1.

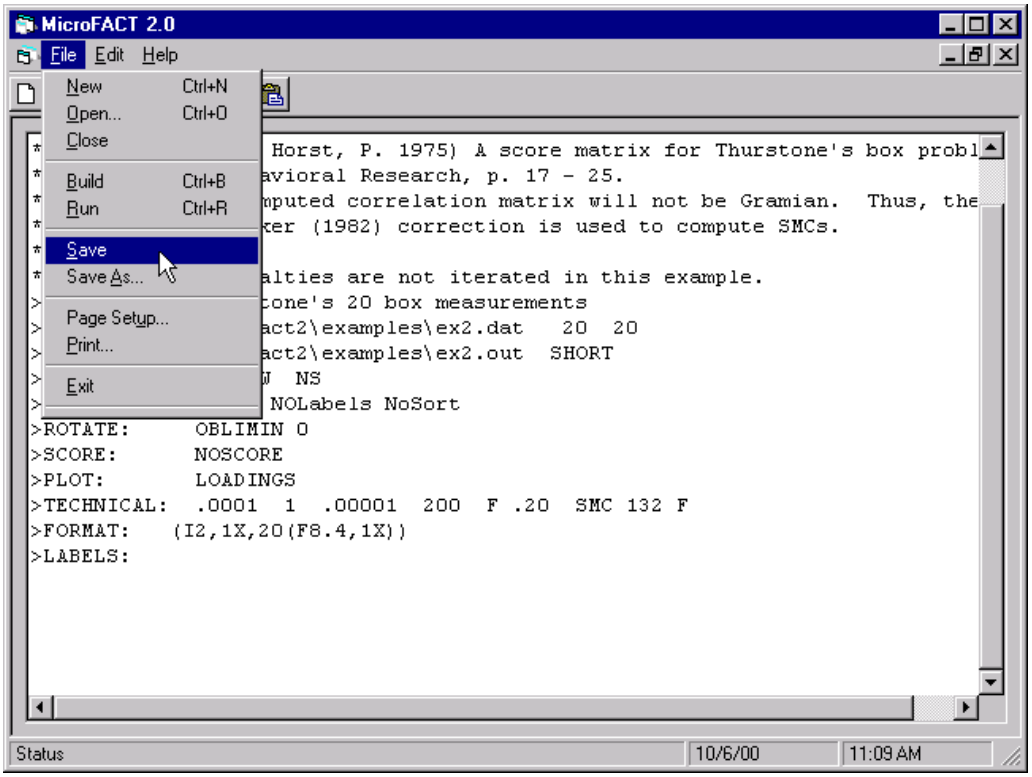

Figure 4.5: The Save Option

Part III

Bibliography

Chapter 5

References and Extended Bibliography

Alexander, R.A., Alliger, G.M., Carson, K.P., & Barrett, G. (1985). The empirical performance of measures of association in the 2 x 2 table. *Educational and Psychological Measurement, 45,* 79-87.

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

Bartholomew, D.J. (1987). *Latent variable models and factor analysis.* New York: Oxford University Press.

Bernstein, I.H. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin,* 105, 467-477.

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

Bock, R.D., Gibbons, R.D. & Muraki, E. (1988). Full-information factor analysis. *Applied Psychological Measurement, 12,* 261-280.

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

Brown, M.B. & Benedetti, J. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika, 42*, 347-355.

Burt, C. (1950). The factor analysis of qualitative data. *British Journal of Psychology (Statistical Section), 3,* 166-185.

Carroll, J.B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika, 10*, 1-19.

Carroll, J.B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika, 26*, 347-372.

Cattell, R. B., & Gorsuch, R. L. (1963). The uniqueness and significance of simple structure demonstrated by contrasting organic "natrual structure" and "random structure" data *Psychometrika, 28*, 55-67.

Cattell, R. B., & Sullivan, W. (1962). The scientific nature of factors: A demonstration by cups of coffee. *Behavioral Science, 7*, 184-193.

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

Collins, L.M., Cliff, N., McCormick, D., & Zatkin, J.L. (1986). Factor recovery in binary data sets: A simulation. *Multivariate Behavioral Research, 3*, 377-392.

Cureton, E. E., & Mulaik, S. A. (1975). The weighted Varimax rotation and the Promax rotation. *Psychometrika, 40*, 183-195.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *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.

Etezadi-Amoli, J. & McDonald, R.P. (1983). A second generation nonlinear factor analysis. *Psychometrika, 48,* 315-342.

Ferguson, G.A. (1941). The factorial interpretation of test difficulty. *Psychometrika, 6*, 323-329.

Finkbeiner, C. T., & Tucker, L. R. (1982). Approximate uniqueness estimates for singular correlation matrices. *Psychometrika, 47*, 517-521.

Fraser, C. (1986). *NOHARM: An IBM PC computer program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory.* Center for Behavioral Studies, The University of New England, Armindale, New South Wales, Australia 2351.

Fraser, C. & McDonald, R.P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research, 23*, 267-269.

Gough, H.G. (1987). *The California Psychological Inventory administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H.G. & Heilbrun, A.B., Jr. (1983*). The Adjective Checklist manual–1983 edition.* Palo Alto, CA: Consulting Psychologists Press.

Guilford, J. P., & Zimmerman, W. S. (1963). Some variable-sampling problems in the rotation of axes in factor analysis. *Psychological Bulletin, 60*, 289-301.

Hambleton, R.K. & Rovinelli, R.J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10*, 287-302.

Harris, C. W. & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometric, 29*, 347-362.

Hathaway, S.R. & McKinley, J.C. (1983*). The Minnesota Multiphasic Personality Inventory manual*. New York: Psychological Corporation.

Hendrickson, A.E., & White, P.O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Mathematical and Statistical Psychology, 17*, 65-70.

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

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika, 31*, 313-323.

Jöreskog, K.G. & Sörbom, D. (1989). *LISREL VII Users guide*. Chicago IL. Scientific Software.

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Kaiser, H.F., & Horst, P. (1975). A score matrix for Thurstone's box problem. Multivariate Behavioral Research, 10, 17-25.

Knol, D.L. & Berger, M.P.F. (1988). *Empirical comparison between factor analysis and item response models*. Research Report 88-11, Department of Education: Division of Educational Measurement and Data Analysis: University of Twente.

Knol, D.L. & Berger, M.P.F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457-477.

Lancaster, H. O. & Hamdan, M. A. (1964). Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. *Psychometrika, 29,* 383-391.

Lee, H. B. & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research, 14*, 301-321.

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

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

McDonald, R.P. (1965). Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology, 18*, 11-23.

McDonald, R.P. (1967). Nonlinear factor analysis. *Psychometric Monograph, No. 15(a).*

McDonald, R.P. (1969). The common factor analysis of multicategory data. *British Journal of Mathematical and Statistical Psychology, 22,* 165-175.

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

McDonald, R.P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 4,* 379-396.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

McKinley, R.L., & Reckase, M.D. (1983). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. *Behavior Research Methods & Instrumentation, 15*, 389-390.

Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11,* 3-31.

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

Muthén, B. (1989). Dichotomous factor analysis of symptom data. *Sociological Methods & Research, 18*, 19-65.

Muthén, B. (1987). *LISCOMP. Analysis of linear structural equations with a comprehensive measurement model.* Chicago, IL. Scientific Software.

Muthén, B. & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46,* 407-419.

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

Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.

Muthén, B. & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika, 53*, 563-578.

Muraki, E., & Englehard, G. (1985). Full-information item factor analysis: Applications of EAP scores. *Applied Psychological Measurement, 9*, 417-430.

NAG (1991). *Library* (Mark 15). Oxford, UK: Numerical Algorithms Group.

Olsson, U. (1979a). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443-460.

Olsson, U. (1979b). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research.* 14, 485-500.

Pearson, K. & Heron, D. (1913). On theories of association. *Biometrika, 9*, 159-315.

Pearson, K., & Pearson, E.S. (1922). On polychoric coefficients of correlation. *Biometrika, 14*, 127-156.

Perry, C.D.H. & McArdle, J.J. (1991). An applied comparison of methods for least-squares factor analysis of dichotomous variables. *Applied Psychological Measurement, 15,* 35-46.

Reise, S. P., & Waller, N.G. (2001). Item response theory for dichotomous assessment data. In F. Drasgow & N. W. Schmitt (Eds). *Measuring and Analyzing Behavior in Organizations: Advances in Measurement and Data Analysis* (pp. 88-122). San Francisco: Jossey-Bass.

Reiser, M., & VandenBerg, M. (1994). Validity of the chi-square test in dichotomous variable factor analysis when expected frequencies are small. *British Journal of Mathematical and Statistical Psychology, 47*, 850197.

Ritchie-Scott, A. (1918). The correlation coefficient of a polychoric table. *Biometrika, 12*, 93-133.

Rozeboom, W. W. (1992). The glory of suboptimal factor rotation: Why local minima in analytic optimization of simple structure are more blessing than curse. *Multivariate Behavioral Research, 27*, 585-599.

Schilling, S. (1993). *Advances in full information item factor analysis using the Gibbs sampler*. Paper delivered at the 1993 meeting of the Psychometric Society, Berkeley, California.

Takane, Y., & DeLeeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*, 393-408.

Tellegen, A. (1982*). Manual for the Multidimensional Personality Questionnaire*. Unpublished manuscript. University of Minnesota.

Ten Berge, J. M. F. (1995). Suppressing permutations or rigid planar rotations: A remedy against nonoptimal varimax rotations. Psychometrika, 60, 437-446.

Waller, N.G. (1993). Seven confirmatory factor analysis programs: EQS, EzPATH, LINCS, LISCOMP, LISREL 7, SIMPLIS, and CALIS. *Applied Psychological Measurement, 17,* 73-100.

Waller, N. G., Tellegen, A., McDonald, R. P., & Lykken, D. T. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a Negative Emotionality scale. *Journal of Personality, 64*, 545-576.

Wherry, R.I. & Gaylord, R.H. (1944). Factor pattern of test items and tests as a function of the correlation coefficient: Content, difficulty and constant error factors. *Psychometrika, 9*, 237-244.

Wilson, D., Wood, R., & Gibbons, R.D. (1984). *TESTFACT*: Test scoring, item statistics, and item factor analysis. Mooresville IN: Scientific Software.

Part IV

Examples

Chapter 6

Example Data Sets

EX1A Title: Thirty Negative Emotionality Items from The Multidimensional Personality Questionnaire.
Ref: Waller, N. G., Tellegen, A., McDonald, R. P., & Lykken, D. T. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a Negative Emotionality scale. *Journal of Personality, 64,* 545-576.

EX1B Title: Thirty Negative Emotionality Items from The Multidimensional Personality Questionnaire. Analysis of Tetrachoric Correlations.
Ref: Waller, N. G., Tellegen, A., McDonald, R. P., & Lykken, D. T. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a Negative Emotionality scale. *Journal of Personality, 64,* 545-576.

EX2 Title: Thurstone 20 box measurements plasmode.
Ref: Thurstone, L. L. (1947). *Multiple Factor Analysis*.
Chicago, University of Chicago Press.
The raw data for this example are taken from:
Kaiser, H. F., & Horst, P. (1975) A score matrix for Thurstone's box problem. *Multivariate Behavioral Research, 10,* p. 17 - 25.

EX3 Title: Cattell and Dickman Ball problem
Ref: Cattell, R. B., & Dickman, K. A. (1962). A dynamic model of Physical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin, 59,* 389-400.

EX4 Title: Cattell's coffee cup data.
Ref: Cattell, R. B., & Sullivan, W. (1962). The scientific nature of

factors: A demonstration by cups of coffee. *Behavioral Science, 7,* 184-193.

EX5 Title: Holzinger and Swineford's 24 Psychological variables.
Ref: Holzinger, K. J. & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs,* no. 48: Chicago: University of Chicago, Dept. of Education.

EX6 Title: Thurstone 26 box measurements plasmode.
Ref: Thurstone, L. L. (1947). *Multiple Factor Analysis.* Chicago, University of Chicago Press.
The correlation matrix for this example is taken from:
Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika, 40,* 183-195.

EX7 Title: Factor analysis of Big Five personality items using polychoric correlations.
Ref: Unpublished data collected by Niels Waller.

EX8 Title: LSAT6 data
Ref: Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. Psychometrika, 35, 179-197.

EX9 Title: LSAT7 data
Ref: Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179-197.