LECTURE 5:   THE LISREL 8 AND PRELIS 2 PROGRAMS

I.   THE PRELIS 2 PROGRAM.
II.   THE LISREL 8 PROGRAM.
III.   SIMPLIS Command Language.

In this lecture, we will go over the PRELIS and LISREL programs, highlighting features, noting changes in the new versions, and noting limitations.   I'll also briefly overview the SIMPLIS command language, which allows one to specify models without knowing matrix algebra.

I.   THE PRELIS PROGRAM.

PRELIS is a preliminary program for preparing data to be analyzed by LISREL.   It helps screen data for departures from normality (using measures of skewness and kurtosis), and provides summary descriptive statistics, such as univariate statistics and concise bivariate contingency tables.   Moreover, it will produce sample moment matrices suitable for analysis with LISREL 8, and in particular, provide polychoric and polyserial correlations, as well as the appropriate asymptotic covariance matrix of observed sample moments for use in weighted least squares.   There are 10 lines of code needed to specify a PRELIS computer run.

1.   DA Line (Data Parameters).
2.   LA Line (Labels).
3.   RA Line (Raw Data).
4.   Definition of Variables (Scale Types).
5.   Transformation of Variables.
6.   Selecting Cases and Variables.
7.   MI Line (Missing Values).
8.   WE Line (Weight Variable).
9.   RG Line (Regression).
10.   OU Line (Output and Save Matrices).

1.   DA Line (Data Parameters).

The DATAPARAMETERS (DA) line defines the input data (number of variables and observations, treatment of missing values).   The format is

   DA   NI = 10 MI = -9 MC = 5 TR = LI

where NINPUT (NI) refers to the number of input variables; MISSING (MI) defines a global value for missing observations in the raw data; MAXIMUMCATEGORIES (MC) defines a global maximum number of categories for all ordinal variables; TREATMENT (TR) defines missing values as either pairwise or listwise.

2.   LA Line (Labels).

This line specifies the file containing variable labels and the Fortran format of the labels.

LA   FI = C:\LISREL8W\LABELS\LABELS.TXT   FO
(5A8, 2A5)

This example tells PRELIS to read the labels from a file called "labels.txt" from my subdirectory LISREL8\LABELS\ on my hard disk's C: drive.   It further specifies that a Fortran format follows on the next line, specifying the labels are written in character form with eight labels per line, the first five each occupying eight spaces, the last two occupying two spaces.   For labels an A-format for alphanumeric characters must be used.   If FO is omitted, the labels are read in free format (a blank separating labels); if FI is omitted, PRELIS will look to read the labels in the input stream starting with the next line.

PRELIS 2 also allows you to specify labels for categories of categorical (ordinal) variables.

CL *varlist* n1 = clab1 n2 = clab2 n3 = clab3...

where n1, n2, ... are numerical codes in the data referring to categories, and clab1, clab2... are labels to be assigned to the categories.   For example:

CL CLASS 1 = UPPER 2 = MIDDLE 3 = LOWER

3.   RA Line (Raw Data).

This line specifies where to read the raw data.   It has the following format:

RA   FI = C:\LISREL8\DATA\LOUSY.DAT   FO
(12F10.7, 10X, 3F2.0)

This example tells PRELIS to read the raw data from a file "lousy.dat" from my LISREL8\DATA directory on my C: drive.   It specifies a Fortran format, which says the first twelve variables which occupy 10 spaces, and the last two are decimals (this is overridden if a decimal is in the data itself), then skip ten space, and then read three variables each of which occupy two spaces.   (For numerical data, an F-format for more on Fortran formats, see the PRELIS manual.)   Again, if FO is omitted, free-format is assumed; if FO is not in the run stream, PRELIS will look for it in the first line of the data file.

4.   Definition of Variables (Scale Types).

This line defines the scales of specific variables, overriding the global value of MC on the DA line.   There are five possibilities:

CO   Continuous (this is the default value for each variable)
OR   Ordinal variables.
CA   Censored from above.
CB   Censored from below.
CE   Censored from above and below.
FI   Fixed X-variables.

For example, assume that the DA line read "DA MC = CO":

OR   1 3-6
CB   INCOME, AGE
CE   'EDUCAT' 'SES'

In this example, the first and third through sixth variables are ordinal, INCOME and AGE are censored from below, and EDUCAT and SES are censored both from above and below (the quotes are unnecessary), and the remaining variables are to be treated as continuous.   Thus, the definition of variables overrides, for specific variables, the global definition given on the DA line.

PRELIS 2 allows one to specify fixed-Xs and will compute univariate probit regressions of y* on the x-variables for each ordinal variable, where y* is the normally-distributed continuous variable underlying y.   It will also compute conditional polychoric correlations for a given X, the joint covariance matrix of the observed variables, and its corresponding asymptotic covariance matrix.   Thresholds are estimated jointly with the regression coefficients, but can also be specified to be estimated from the marginal distributions of the ordinal variables (and then treated as fixed)
MT   *variable list*

or fixed to specific values by the user specified in *filename*:

FT *filename variable list*

5.   Transformation of Variables.

This set of lines recodes and transforms variables.   There are three possibilities: PO (POWER) which transforms a variable to a power (exponential), LO (LOGARITHM) which takes the log of the variable, and RE (RECODE) which recodes values of the variable.   The power function uses the following formula:

$$y = (\alpha + \beta X)^{\gamma}$$

So that $\gamma = 2$ gives a square transformation; $\gamma = 3$ a cubed transformation; $\gamma = \frac{1}{2}$ (.5) a square root transformation; and $\gamma = -1$ an inverse.   The format is:

PO   *varlist*   AL = *α-value*   BE = *β-value*   GA = *γ-value*

The log transformation uses the following equation:

$$y = \log (\alpha + \beta X)$$

The format is:

LO   *varlist*   AL = *α-value*   BE = *β-value*

Default values of these parameters are $\alpha = 0$; $\beta = 1$ and $\gamma = 1$.

Recode allows one to recode values of a variable to recode missing values, collapse variables, combine categories, or define new categories.   The format is:

RE   *varlist*   OLD = *valuerange*   NEW = *value*

where valuerange refers to a range of values of the old variable, and value refers to the new value.   For example:

RE   INCOME   OLD = 0-10000   NEW = 0
RE   INCOME   OLD = 10001-50000   NEW = 1
RE   INCOME   OLD = 50001-100000   NEW = 2

This recodes income categories into a trichotomous ordinal variable.

PRELIS 2 adds some shortcuts to recoding categories.   For example, if you wanted to recode a four-category variable to reverse the scale so that 1=4, 2=3, 3=2, 4=1, simply recode as follows:

RE   LIKERT   OLD = 1,2,3,4 NEW = 4,3,2,1

Or if you wanted to dichotomize a four-category variable into a two category variable, so that 1=0, 2=0, 3=1, 4=1, recode as follows:

RE   DUMMY   OLD = 1-2, 3-4   NEW = 1, 0

PRELIS 2 allows one to define new variables as functions of old variables using the NE command:

NE   *varlabel = some function of old variables*

where the function can use addition, subtraction, multiplication indicated by *, and exponentiation indicated by **. For example, to define a new variable, $y_1 = (1 + x^2)^{1/2}$

NE Y1 = 1 + X**2
PO Y1 GA=.5

6.   Selecting Cases and Variables.

To select cases to be analyzed, use the SC (SELECT CASES) line.   It has the following format:

SC   *varlist*   *conditionlist*

where conditionlist specifies conditions of selection -- either greater than some value > c, less than some value < c or equal to some value = c.   Only cases passing the conditions will be selected; all others will be deleted from the analysis.

SCRACE   > 0
SCSEX   = 1
SCINCOME   < 10000

In the above example, race is coded 0 for minorities 1 for whites; sex is coded 0 for females, 1 for males; and income is in dollars.   Thus, the above lines will restrict the analysis to poor white males.

To delete variables from the analysis use SD (SELECT CASES AND DELETE VARIABLES).   It has the following format:

SD   *varlist*   *conditionlist*

By omitting the conditionlist parameter, this simplifies to a list of variables to be *excluded* from the analysis.   By including the conditionlist, one can both exclude variables *and* select cases at the same time.   PRELIS will automatically delete the variables included in the varlist of SD:

SD   SEX = 1

This will select only females and simultaneously drop the variable SEX from the analysis.

PRELIS 2 has added a parameter called *casecondition*, which selects cases for all variables:

SC   *casecondition*

Casecondition can be greater than a constant, less than a constant, or equal to a constant, or equal to odd or even cases.   For example

SCcase = odd
SCcase = < 100

will select all odd numbered case numbers less than 100.

7.   MI Line (Missing Values).

This optional line defines missing value codes for specific variables (as opposed to the *global* MI command in the DA line).   The format follows:

MI   -9   1 3 5

where -9 is the value of 1st, third, and fifth variables to be defined as missing.   Variable labels can also be used to specify missing values.

PRELIS 2 has increased the flexibility of specifying missing values.   It allows more than one missing value code per variable and also allows a range of values to be defined as missing:

MI   a, b, c... *varlist*

where a, b, c are values in the data to be treated as missing -- either singular values, like 8, or ranges of values such as 8-20 -- for each variable listed in the *varlist*.   Place a comma after each value except the last.

PRELIS 2 also allows one to impute values for missing data.   What it does is replace the missing value on Y with a missing value (or the mean of several missing values) from another case that is similar on some other matching variables, X.   You have to define which other variables, X, should be used as matching variables.   For example, suppose Y is income with a missing value on the fourth case, and you define X_1 (Father's income) as the matching variable, and it has a value of $x_1 = 20,000$ for the fourth case.   PRELIS 2 will examine those cases which have data on both Y and X_1, and find those that have the closet values to $x_1 = 20,000$.   If there is only one such value (say case 201 has $x_1 = 20,002$), it'll replace Y with the corresponding value of Y from case 8.   If there are more than one such case (say 8, 20, and 139), it'll replace Y with the corresponding mean of Y's from cases 8, 20, and 139).   It'll do this assuming that the x's are similar enough, measured by the variance ratio, $VR = s_m^2/s_y^2$, where s_m^2 is the variance of the y-values of the matching cases, and s_y^2 is the variance of y on all cases.   The default is $VR > .5$ for matching to take place.   Smaller values imply more precision in mean imputation; larger values are not recommended by Jöreskog and Sörbom.   The command to impute missing values is:

IM (*Ivarlist*) (*Mvarlist*) VR = n   XN   XL

Where Ivarlist is the list of variables to be imputed; Mvarlist is the list of matching variables.   By default, PRELIS 2 lists all cases with missing values, the imputed value, the number of matching cases, the variance ratio, and the reason for failure to impute.   XN makes PRELIS 2 list only successful imputations; and XL makes it skip the listing of cases altogether.

8.   WE Line (Weight Variable).

This specifies a variable used to weight the observations -- such as sampling weights from a disproportionate stratified sample.   The format is

WE   *varname*

For example,

WE   SAMPLE.WTS

where SAMPLE.WTS is the weight variable found in the dataset.

9.   RG Line (Regression).

This estimates regression models as a preliminary step to construct a structural equation model.   The format is:

RG   *y-varlist*   ON   *x-varlist*

where y-varlist is a list of dependent variables and x-varlist is a list of independent variables:

RG   ED   INCOME ON   FATHOCC FATHED

This line specifies two regressions: one regressing ED on FATHOCC and FATHED; the other regressing INCOME on FATHOCC and FATHED

10.   OU Line (Output and Save Matrices).

This line specifies the output to be printed, the moment matrices to be estimated, and the parameters (matrices) to be printed and saved.   The format is:

OU   MA = *type of matrix   save-parameters   print-parameters*

The following are types of matrices (MA) to be estimated:

MM   Matrix of moments about zero.
AM   Matrix of moments augmented by a vector of constants (for estimating means using LISREL 6).
CM   Covariance matrix.
KM   Correlation matrix (Pearson product-moment correlations) for either continuous or ordinal variables (the latter on normal scores determined from marginal distributions.
OM   Matrix of correlations based on optimal scores.
PM   Matrix of correlations, including polychoric or polyserial correlations (used in estimating a model with ordinal or dichotomous dependent variables), along with Pearson correlations.

The default is to estimate no matrix.   Both OM and PM produce Pearson correlations between continuous variables. But for correlations between continuous and ordinal variables, OM produces a *consistent by not asymptotically efficient* estimate of the polyserial correlation, whereas PM produes an ML estimator.   For correlations between two ordinal variables, PM provides ML estimates of polychoric correlations, whereas OM provides product-moment correlations among optimal scores (cannonical correlations).   Jöreskog and Sörbom present Monte Carlo evidence that in terms of mean square error, for correlations between ordinal variables, polychoric correlations perform better than cannonical correlations, Pearson correlations on normal scores, and Pearson correlations assuming continuous scores.

The following save parameters, which specify where to save the estimated matrix, may be included:

SM = *filename* to save the matrix to be analyzed MA.*
SR = *filename* to save the transformed raw data matrix.
RA = *filename* to save a raw data submatrix being analyzed.   Use WI = *width-of-format* and ND = *number-of-decimals* (e.g., WI=1, ND=0 specifies data that can be read with format (*n*F1.0) (PRELIS 2 only).
SA = *filename* to save the asymptotic covariance matrix of MA.
SV = *filename* to save the asymptotic variances of MA.

For example, to save the asymptotic covariance matrix of MA in file C:\LISREL\SOL\ACOV.DAT,

SA = C:\LISREL\SOL\ACOV.DAT

   *To conform to LISREL, PRELIS 2 allows you to specify the specific kind of moment matrix (MA) to be saved, rather than simply using the generic SM command.   These include CM, KM, PM, RM, OM, TM.   Also, one can save the asymptotic covariance matrix using AC, rather than SA.

Note that PRELIS 1 saved the asymptotic covariance matrix (AC) in an ASCII file with the format (6D13.6). PRELIS 2 saves AC as a binary file in format (3D23.16) for greater precision and less space.   LISREL 8 only reads

asymptotic covariance matrices in binary format.   To "look at" the binary file, you must use the utility program BIN2ASC to convert the binary file to ASCII and store it:

BIN2ASC   *infile*   *outfile*

To convert the ASCII file to binary form, use the utility program ASC2BIN to convert the file and store it in binary form:

ASC2BIN   *infile*   *outfile*

The following print parameters control what is printed in the printout.

PA   Print the asymptotic covariance matrix (assuming SA was specified).
PV   Print the asymptotic variances (assuming SV was specified).
PK   Print Mardia's measure of relative multivariate kurtosis.
WP   Print output to 132 columns, rather than the default 80.
XB   Exclude the bivariate frequency tables from the printout.
XT   Exclude test statistics (e.g. test of the whether the underlying distribution is bivariate normal; test of whether the correlation is zero).

II. THE LISREL 8 PROGRAM.

There are basically thirteen lines of code needed to specify and estimate a covariance structure model in LISREL 8. These lines tell LISREL where to obtain the data, in what form the data can be found, how to read the data, the specification of the model, the method of estimating and testing the model, selected options needed, and aspects of the solution to be included in the printout.

1.   DA Line (Data Parameters).
2.   Input Data.
3.   SE Line (Selection of Variables).
4.   MO Line (Model Parameters).
5.   LK and LE lines (Labels for Latent Variables).
6.   FR and FI lines (Fixed and Free Matrix Elements).
7.   PA Line (Pattern Matrix Specifying Fixed and Free Parameters).
8.   EQ Line (Equality Constraints).
9.   VA and ST Lines (Fixed Values and Starting Values).
10.   MA Line (Matrix Specifying Values of Fixed and Free Parameters).
11.   PL Line (Parameter Plots).
12.   NF Line (Never Free These Parameters).
13.   OU line (Output Requested).

1.   DA Line (Data Parameters).

The DATA line (abbreviated DA) specifies the form of the moment matrix to be analyzed (after it is read).   The format of this line is

DA   NG = 1 NI = 5   NOBS = 1000 MA = KM

where NGROUPS (NG) refers to the number of groups to be analyzed (default = 1), NINPUT (NI) refers to the number of input variables in the sample moment matrix, NOBSERVATIONS refers to the number of observations in the sample, and MATRIX (MA) refers to the form of the matrix to be *analyzed* (not necessarily read).   There are six options for MA (default is CM):

MM   Matrix of moments about zero.
AM   Matrix of moments augmented by a vector of constants (for estimating means using LISREL 6).
CM   Covariance matrix.
KM   Correlation matrix (Pearson product-moment correlations).
OM   Matrix of optimal scores from PRELIS.
PM   Matrix of correlations, including polychoric or polyserial correlations (used in estimating a model with ordinal
        or dichotomous dependent variables).

2.   Input Data.

The specification of input data typically takes two lines each: FI indicates filename; FO indicates the format of the
data.   The following are types of input data:

LA   Variable labels.
RA   Raw data.
MM   Matrix of moments about zero.
CM   Covariance matrix.
KM   Correlation matrix (Pearson product-moment correlations).
OM   Matrix of optimal scores from PRELIS.
PM   Matrix of correlations, including polychoric or polyserial correlations (used in estimating a model with
                ordinal or dichotomous dependent variables).
ME   Vector of means.
SD   Vector of standard deviations.
AC   Asymptotic covariance matrix of the covariance or correlation matrix (in binary form for LISREL 8).
AV   Asymptotic variances of the covariance or correlation matrix.
DM   User defined weight matrix.

The input data line has the following format (for e.g., a covariance matrix):

CM   FI = *filename* FO
(*format*)

For example, to read a correlation matrix, vector of standard deviations, and means from disk, and labels from the
runstream, specify:

KM   FI = C:\LISREL\DATA\CORR.DAT   FO
(5F10.7)

SD   FI = C:\LISREL\DATA\SD.DAT   FO
(5F10.7)

ME   FI = C:\LISREL\DATA\MEANS.DATFO
(5F10.7)

LA
*
'FO' 'FED' 'ED' 'OCC' 'OCC2'

3.   SE Line (Selection of Variables).

The SE line selects and reorders a subset of variables in the input matrix to be included in the analysis.   The *y-variables* are always listed *first*, followed by x-variables.   The variables can be referred to by the numerical position

in the input matrix or by their variable labels.   The following orders the above labeled variables in the same way (where FO and FED are x-variables; the others are y-variables):

SE   'ED' 'OCC' 'OCC2' 'FO' 'FED'

SE 3 4 5 1 2

4.   MO Line (Model Parameters).

This line defines the parameters of the model, including specifying the number of observed and latent variables, and specifying the structure of the parameter matrices Φ, Γ, Λ_x, Θ_δ, B, Λ_y, Θ_ε, Ψ.

The format of the MO line is

MO NY=p NX=q NE=m NK=n MN=*matrix-form, fixed-or-free*

where   NY = number of y-variables (p)
          NX = number of x-variables (q)
          NE = number of eta-variables (m)
          NK = number of ksi-variables (n)

and MN refers to a parameter matrix (PH, GA, TD, TE, LY, LX, BE, and PS) and *matrix-form* takes the following alternatives:

ZE = zero matrix (same as omitting the matrix name entirely)
ID = identity matrix (1s on diagonals, zeros elsewhere)
IZ = partitioned identity matrix and zero [I   0]
ZI = partitioned zero and identity matrix [0   I]
DI = diagonal matrix
SD = a square beta matrix with elements under the diagonal free; all others fixed to zero.
SY = a symmetric matrix.
ST = a symmetric matrix with fixed ones on the diagonal.
FU = a rectangular or square non-symmetric matrix.

and *fixed-or-free* takes one of the following two alternatives:

FI = fixed values.
FR = free values.

Jöreskog and Sörbom try to simplify general models by specifying four sub-models:

Submodel 1: Specify only NX and NK and get a confirmatory factor model:

   $x = \Lambda_x \, \xi + \delta$

Submodel 2: Specify only NY and NX and get a structural equation or path model in observables only:

   $y = B \, y + \Gamma \, x + \zeta$

this sets NE = NY, NK = NX, Λ_y = I, Λ_x = I Θ_ε = 0, Θ_δ = 0, and $\Phi = S_{xx}$, leaving only three parameter matrices B, Γ, and Ψ.

Submodel 3A: Specify only NY, NE, and NK, and eliminate x-variables:

$$\eta = B\,\eta + \gamma\,\xi + \zeta$$

$$y = \Lambda_y\,\eta + \varepsilon$$

Submodel 3b: Specify only NY and NE:

$$\eta = B\,\eta + \zeta$$

$$y = \Lambda_y\,\eta + \varepsilon$$

This eliminates all xs and ξs.

The matrices above define the specific structure of the covariance structure model. Some parameter matrices are common, including:

1. If LY and LX = ID and TD and TE = 0, the model has no measurement error.
2. If BE = SD (symmetric and below diagonal are free) model is recursive; if FU (with free elements above and below the diagonal, model is nonrecursive.
3. If PSI = SY, disturbances are correlated across equations; if PSI = SY, they are not.
4. If TE or TD = SY, measurement errors are correlated; if TE or TD = DI they are not.
5. In LISREL 8, one can define TH to refer to the covariances between TD and TE. This is defined as fixed, but then one can free specific elements within the matrix to allow covariances between δs and εs.

5.   LK and LE lines (Labels for Latent Variables).

LK refers to labels for ξ-variables; LE refers to labels for η-variables. The format follows that of labels of observables:

LK   FI = *filename* FO
(*format*)

LEFI = C:\LISREL8\LABELS\LABELS.ETA   FO
(5A8)

6.   FR and FI lines (Fixed and Free Matrix Elements).

These lines define specific parameters within a matrix to be fixed or free. The format is to follow FI or FR with the matrix name and elements:

FR PS 1 1 PS 2 2 PS 3 3
FI LY 1 1 LY 3 2 LY 5 3
FI LX 1 1 LX 3 2
FR BE 1 1 BE 2 1 BE 3 2 BE 3 1

If in the MO line, Ψ was specified to be symmetric and fixed, the above line frees the three diagonal elements.

7.   PA Line (Pattern Matrix Specifying Fixed and Free Parameters).
An alternative way of specifying fixed and free parameters is by specifying a pattern matrix, with 1s denoting free parameters and 0s denoting fixed parameters. The format is:

PA   *parameter matrix*
(format)

For example, to specify a 3 x 3 diagonal Ψ matrix, the following are equivalent:

PA   PS
(3I1)
100
010
001

PA   PS
*
1 0 0 0 1 0 0 0 1

8.   EQ Line (Equality Constraints).

This line specifies equality restrictions on parameters.   The format is EQ, followed by the parameter matrix name, followed by the position of the parameters constrained to be equal:

EQ   LY 2 1 LY 3 2 LY 5 3

This constrains three factor slopes to be equal.

EQ   BE 2 1 BE 1 2

This constrains the reciprocal effects in a simultaneous equation model to be equal.

9.   CO Line (Linear and Non-linear Constraints).

LISREL 8 allows you to specify linear and non-linear constraints.   Thus, you can specify additive and multiplicative functions among parameters.   For example, one can specify:

$$\gamma_{11} = 10 + \gamma_{21}{}^2 + \gamma_{31}{}^2$$

To specify this, use the CO command:

CO   GA (1 1) = 10 + GA (2 1)**2 + GA (3 1)**3

where * stands for multiplication, ** stands for exponentiation, and parentheses refer to matrix elements.   While you can't divide here, you can use negative exponents, and noninteger exponents.

10.   VA and ST Lines (Fixed Values and Starting Values).

These lines specify start values for free parameters to be estimated, and constrained values of fixed parameters. The default value for all fixed or free parameters is zero.   The formats are:

VA   *value*   *matrix element*
ST   *value*   *matrix element*
VA   1.0   LY 1 1 LY 3 2 PS 1 1 PS 2 2 PS 3 3 PS 4 4

If Ψ had been defined as a 4 x 4 diagonal matrix, then the above line would be equivalent to:

VA   1.0   LY 1 1 LY 3 2 PS 1 1-PS 4 4

To give *all* parameters the same start value of 1.0, type:

ST   1.0   ALL

11.   Interval Restrictions (Constraining Parameters to Lie Within a Range of Constants).

IR allows you to specify that a parameter falls within the range specified by one or two constants.   The general form is $a \leq \theta \leq b$ or $a \geq \theta \geq b$.   This allows the user to determine inadmissable estimates.   The syntax is:

IR *parameter list condition*

where condition refers to less than or equal to a constant $< c$ or greater than or equal to a constant $> c$.   For example,

IR TD(1, 1) > 0 < 1

specifies that $0 \leq \theta_{11}{}^{\delta} \leq 1$.

12.   MA Line (Matrix Specifying Values of Fixed and Free Parameters).

This line specifies values of fixed and free parameters with a matrix of 1s and 0s, rather than by specifying each element.   It may be an easier way when there are many different nonzero start values.   The format is:

MA   *parameter matrix*
(*format*)

If the Fortran format line is omitted (or if an * is used), the program assumes free format.   For a 2 x 4 Γ matrix, the following are equivalent:

MA   GA
*
1.5 2.0 3.0 1.0 2.5 2.0 0 0

MA   GA
(4F4.2)
 15   20   30   10
 25   20   0   0

MA   GA
*
1.5 2.0 3.0 1.0
2.5 2.0/

In the last case, the slash says, "the remaining elements of the matrix will have start values of zero.

13.   PL Line (Parameter Plots).

This is an optional step that plots various values of the likelihood function by the values of either (1) a fixed parameter (Lagrangian multiplier test) or (2) a free parameter (yielding a confidence interval for the parameter).   This is very computer-intensive, but may have some limited use in helping to fit a model or determining whether a particular estimate is asymptotically normal, since (2) should be symmetric under normality.

14.   NF Line (Never Free These Parameters).

This line specifies parameters that should always be constants, given that theory suggests that they not be free. This means that (a) LISREL will not compute modification indices for them; and (b) LISREL will not free them under its mindless "automatic model modification" procedure.

15.   OU line (Output Requested).

This line specifies some options of the model (such as method of estimation) as well as what should be included in the printout. The first option is method of estimation ME:

ME = *method of estimation*

> IV   for instrumental variable estimation.
> TS   for two-stage least squares estimation (TSLS)
> UL   for unweighted least squares estimation (ULS).
> GL   for generalized least squares estimation (GLS).
> ML   for maximum likelihood estimation (ML, the default estimator).
> WL   for weighted least squares (WLS).
> DW   for diagonally weighted least squares (DWLS).

The second option RO refers to the ridge option; RC specifies a ridge constant:

RO   Use the ridge option.
RC   Ridge constant = c (default = .001)

This is directly analogous to ridge regression for models containing unstable estimates due to multicollinearity, empirical underidentification, etc. It operates by adding a constant to the diagonal elements of the sample covariance matrix -- $S + c(\text{diag } S)$ -- and then analyzing the new matrix. If this matrix is not positive definite, the program will multiply c by 10 repeatedly until it become positive definite. Note that this technique will introduce bias into the parameter estimates, but is justified on the grounds of minimizing mean squared error.

A number of parameters govern which portions of the solution to be included in the printout:

SE   print standard errors (prints by default in LISREL 8).
TV   print t-values (z-statistics) (prints by default in LISREL 8).
PC   print correlation matrix of parameter estimates.
PT   print starting values (default in LISREL 8 is *not* to print them).
RS   print estimated covariance matrix, residuals, standardized residuals, Q-plot.
EF   print total effects and indirect effects (and asymptotic standard errors).
MR   print covariances.
MI   print modification indices.
FS   print factor score regressions.
SS   print LISREL standardized solution (standardizing η and ξ only).
SC   print complete standardized solution (standardizing observables as well).
ALL print everything.
TO   print 80 characters/record (default on PC).
WP   print 132 characters/record (default on mainframe).
ND   number of decimals (0-8) on printout (default is ND=2).

A number of parameters govern the iteration process of estimation:

TM = seconds   maximum seconds of CPU allowed (default is 172800 seconds).
IT = n   maximum number of iterations allowed (default = number of free parameters times 3).
AD = n   check the admissibility of the solution after n iterations (default is 20).

AD = OFF   turn off the admissibility check (for example, for second-order factor models in which a factor has
                no observable indicator).
EP   convergence criterion (default is .000001).

Finally, to save portions of the solution after a run, specify on the output line:

MN = *filename*

where *filename* refers to the VAX filename in which to save the matrix and MN refers to the specific matrix to be
saved, including the LISREL parameter matrices (LY, LX, BE, GA, PH, PS, TE, TD) as well as:

MA   Matrix to be analyzed.
SI   SIgma, the matrix implied by the estimates $\Sigma(\hat{\theta})$.
RM   Regression Matrix of coefficients from the factor score regression of latent variables on observed variables.
EC   Estimated asymptotic covariance matrix of the parameters, in which standard errors are the square roots of
     diagonal elements. (unstandardized version of matrix PC).

Each matrix is written in the format (6D13.6) and preceeded by a Fortran format.   For example:

OU LY = C:\LISREL\SOL\LY.DAT BE = C:\LISREL\SOL\BE.DAT EC = C:\LISREL\SOL\ACOV.DAT

The saved matrices can then be used in a future LISREL run as start values using the MA line:

MA   LY   FI = C:\LISREL\SOL\LY.DAT FO
MA   BE   FI = C:\LISREL\SOL\BE.DAT FO

II. THE SIMPLIS COMMMAND LANGUAGE.

SIMPLIS is a simplified command language used by LISREL to specify models using scalar algebra or arrows (as in
a path model).   This originated from a SYSTAT program called "EZPATH," and has been adopted by most SEM
programs, so you can run models without knowing matrix algebra.   It makes running models easier, but has the
danger of allowing the novice user to run models that make little statistical or substance sense.   The order of
commands follows (brackets indicate optional commands):

1.   [Title]
2.   Observed Variables [from File *filename*]
3.   Covariance Matrix [from File *filemane*]
4.   Sample Size
5.   [Latent Variables]
6.   Relationships
7.   Paths
8.   [Scaling Latent Variables]
9.   [Starting Values]
10. Error Variances and Covariances
11. [Uncorrelated Factors]
12. [Equality Constraints]
13. Options [RS WP ND=3 ME=GLS IT=100 AD=OFF SI=*filename*]
14. [LISREL Output: [SS SC EF SE VA MR FS PC PT]]
15. [Path Diagram]
16. End of Problem

1.   Title

SIMPLIS runs typically begin with a title line(s), which is optional.   The first lines are all considered title lines until a keyword "Observed Variables" or "Labels" is encountered, which begins the SIMPLIS commands.

2.   Observed Variables or Labels

This line lists the observed variables in order of appearance in the data file.   The following is the format:

Observed Variables from File = *filename*

or, equivalently,

Labels from File = *filename*

The labels or observed variables determines the number and order of the variables.   The labels are entered in free format, and can be of any length (although the program prints out only the first 8 characters):

Observed Variables: FOCC IQ ED INC

Observed Variables from File=C:\SEM\DATA\LABELS.DAT

3.   Data

Data can be of different forms, including raw data, a covariance matrix, a covariance matrix and means, a correlation matrix, a correlation matrix and standard deviations, or a correlation matrix, standard deviations, and means.

Raw Data

Raw data are usually read from an external file according to the following format

Raw Data from FILE *filename*

Here is an example;

Raw Data from C:\SEM\STATUS.DAT

Correlation Matrix, Standard Deviations, and Means

For covariance or correlation matrices, the default is free format and a longer triangular matrix:

Covariance Matrix
36.0
22.2   33.8
12.2   22.5   43.9

or

Covariance Matrix: 36.0 22.2 33.8 12.2 22.5 43.9

If you wanted to read a full matrix, write full on the line before the data begin:

Covariance Matrix

Full
36.0 22.2 …

Means and standard deviations are read in similarly:

Means: 1.3 2.9 4.5

Standard Deviations
6.0 5.8 6.6

Asymptotic Covariance Matrix

To estimate models using WLS or ML with robust standard errors (Satorra-Bentler scaled statistics), you'll need to read in the asymptotic covariance matrix.   The format is:

Asymptotic Covariance Matrix from File *filename*

For example:

Asymptotic Covariance Matrix from File C:\SEM\ACOV.PM

One can also read in asymptotic variances for DWLS:

Asymptotic Variances from File *filename*

4.   Sample Size

Here one specifies the sample size.   This is necessary for the program to run.   As an example, if the sample size is 500, one can write

Sample Size = 500

Sample size 500

Sample size: 500

5.   Latent Variables (or Unobserved Variables)

This line determines the order and names of the latent or unobserved variables.   The labels cannot be the same as observed variables.

Latent Variables: SES CRIME

or

Unobserved Variables
SES CRIME

If there are no latent variables, simply omit the line.

6.   Relationships (or Equations)

This is a scalar representation of the equations of the model, following the format:

left-hand side variables = right-hand side variables

where left-hand side variables are latent or observed variables with arrows pointing to it and right-hand side variables are latent or observed variables from where the arrows originate.   The general case is:

Relationships
varlist1 = varlist2

which specifies that all variables in *varlist1* depend on all variables in *varlist2*.   To avoid confusion, I suggest specifying each equation separately:

Relationships
ED = FOCC IQ
INC = FOCC IQ ED

7. Paths

You can also specify relationships using paths, according to the following format:

FROM *variables* -> TO *variables*

This means there are paths from each of the FROM variables to each of the TO variables.   Be sure to put spaces before and after variable names.   I suggest writing out each path, to be sure you know exactly what you are running. That will also make modifications to the model easy and clear:

FOCC -> ED
IQ -> ED
FOCC -> INC
IQ -> INC
ED -> INC

Then if you wanted to constrain the direct effect of IQ on INC to be zero, just eliminate the line IQ -> INC

8.   Scaling Latent Variables

The usual method of scaling latent variables—giving them a metric—is to choose one of its indicators to be a reference indicator, fix its unstandardized slope to be 1.0, and therefore fix the metric of the latent variable to be that of the reference indicator.   If INCOME and EDUCATION are indicators of the latent variable, SES, we can fix the metric of SES to be that of INCOME by making INCOME the reference indicator:

INCOME = 1*SES
EDUCATION = SES

This will fix the metric of SES to be that of INCOME.   We would estimate the slope for EDUCATION.

If no reference indicator is selected, the metric of the latent variable will be fixed by fixing its variance to 1.0 (and putting it in standard deviation units).

9.   Starting Values

LISREL supplies automatic starting values for free parameters, but also allows you to specify your own.   In SIMPLIS, this can be done by putting the starting value in parentheses followed by an asterisk (*) and then the variable name:

INCOME = 1*SES
EDUCATION = (1.0)*SES

The first line fixes INCOME to be the reference indicator for SES; the second line specifies the slope of EDUCATION to be free, and gives it a start value of 1.0.

10.   Error Variances and Covariances

There are three kinds of error terms: measurement errors in observed variables-x, (*errors in x*), measurement errors in observed variables-y (*errors in y*), and measurement errors in structural equations (*errors in Eta*).   By default, error variances are free parameters to be estimated.   Occasionally, one may want to fix an error variance to zero or some other value α.   This can be done in SIMPLIS:

Set the Error Variance of INCOME equal to α

By default, errors are assumed to be uncorrelated.   It is possible to free (and estimate) covariances between *errors in x*, *errors in y*, or *errors in Eta*.

Set the Error Covariance between INCOME and EDUCATION free

One can set the error covariance to be constant:

Set the Error Covariance between INCOME and EDUCATION to .05

To free and estimate correlations between Etas, free the error covariance among the Etas:

Set the Error Covariance between SES and LIFE free

11.   Uncorrelated Factors

By default the exogenous latent variables (Ksi-variables) are allowed to be freely correlated.   One can fix these covariances to be zero (orthogonal factors):

Set the covariances of Ksi1 – Ksi3 to 0

12. Equality Constraints

Path coefficients and error variances that otherwise would be estimated can be constrained to be equal (gaining one degree of freedom.   For example, the following three lines are equivalent:

Set Path from SES1 to LIFE1 = Path from SES2 to LIFE2

Set Path SES1 -> LIFE1 = Path from SES2 -> LIFE2

Set SES1 -> LIFE1 = SES2 -> LIFE2

To constrain error variances to be equal, the following are equivalent:

Set the Error Variances of INCOME and EDUCATION equal

Equal Error Variances: INCOME EDUCATION

13. Options

A number of options for the model and output can be specified on a single line with the following format:

Options: RS WP ND=3 ME=GLS AD=OFF IT=100 SI=*filename*

Print Residuals (RS): This asks for residuals $R = S - \Sigma(\hat{\theta})$, standardized residuals, and fitted covariance matrix $\Sigma(\hat{\theta})$ to be printed in matrix form.

Wide Print (WP):   Asks for the output to be printed with 132 characters per line (rather than the default of 80).

Number of Decimals (ND=*n*):   Specifies the number of digits after the decimal point in the output to be *n* (the default is *n=2*).

Method of Estimation (ME=GLS):   Specifies the method of estimation, such as GLS.   The options are:
   Instrumental Variables (IV)
  Two-Stage Least Squares (TSLS)
  Unweighted Least Squares (ULS)
  Generalized Least Squares (GLS)
  Maximum Likelihood (ML)
  Weighted Least Squares (WLS)
  Diagonally Weighted Least Squares (DWLS)

The default is ML (when no asymptotic variances or covariances are read). The default is WLS when an asymptotic covariance matrix is read, and DWLS when asymptotic variances are read.   This default can be changed by specifying an ME on the options line:

Options ME=TSLS

Admissibility Check (AD=off):   The admissibility check stops iterations if the model fails to be admissible after 20 iterations—such as underidentifed, having a zero row or column in the lambda matrix, etc.   Sometimes, you have the right model but it won't pass the admissibility check (such as a second-order factor model), so you want to turn it off or allow more iterations.   This can be done by turning it off or increasing the iterations (say, to 50):

AD=off
AD=50

Maximum number of iterations (IT=n):   This provides an upper limit to the number of iterations ML or WLS uses. The default is 3 times the number of parameters estimated.   At times, ML needs more iterations to reach convergence.   To increase iterations to 100, set:

IT=100

Save Sigma: This allows you to save the fitted covariance matrix $\Sigma(\hat{\theta})$, called Sigma for subsequent use On the options line, write SI=*filename*.   For example, to save $\Sigma(\hat{\theta})$ in C:\SEM\Sigma1, write on the options line:

SI=C:\SEM\Sigma1

Alternatively, you can use a separate command before the options line:

Save Sigma in File C:\SEM\Sigma1

14. LISREL Output

By default, SIMPLIS presents the model estimates in equation form in the output.   You can ask it to present the model in matrix form (as in LISREL output), by specifying (in the input file):

LISREL Output

You can also ask for additional output from LISREL output (see discussion of LISREL Program above):

LISREL Output: SS SC EF SE VA MR FS PC PT

15. Path Diagram

You can ask for a path diagram by using the line Path Diagram.   If you specify y and x variables, but eliminate all model parameter statements (relationships or paths), you will get an empty path diagram, and can specify the model by drawing arrows between boxes (see example below).

16. End of Problem

This line comes at the end of a run and signals that the problem ends.   One can stack multiple problems within the same input file, with each ending with an End of Problem line.

Examples

The following command files all estimate the same model.   The first one uses LISREL syntax:

```
STATUS ATTAINMENT AND LIFE SATISFACTION
 DA NI=5 NO=1000
 SD
 *
 20 15 3 60.00 3
 KM
 *
 1.00
 .500 1.00
 .750 .600 1.00
 .450 .390 .535 1.00
 .460 .440 .580 .605 1.00
 LA
 *
 FOCC IQ ED INCOME LIFE
 SE
 3 4 5 1 2 \
 MO NX=2 NY=3 GA=FU,FR BE=FU,FI PS=DI,FR
 FR BE 3 2 BE 3 1 BE 2 1
 PD
 OU ME=ML RS EF SC
```

The next one uses SIMPLIS syntax specifying relations in equation (scalar) form:

```
STATUS ATTAINMENT AND LIFE SATISFACTION
 Observed Variables: FOCC IQ ED INCOME LIFE
 Standard Deviation:
 20 15 3 60.00 3
 Correlation Matrix:
 1.00
 .500 1.00
 .750 .600 1.00
```

```
 .450 .390 .535 1.00
 .460 .440 .580 .605 1.00
 Sample Size
 1000
 Relationships
 ED = FOCC IQ
 INCOME = ED FOCC IQ
 LIFE = INCOME ED FOCC IQ
 Method: Maximum Likelihood
 End of Problem
```

The next example uses SIMPLIS and specifies effects using paths:

```
STATUS ATTAINMENT AND LIFE SATISFACTION
Observed Variables: FOCC IQ ED INCOME LIFE
Standard Deviation:
20 15 3 60.00 3
Correlation Matrix:
1.00
.500 1.00
.750 .600 1.00
.450 .390 .535 1.00
.460 .440 .580 .605 1.00
Sample Size
1000
Paths
FOCC IQ -> ED
FOCC IQ ED ->INCOME
FOCC IQ ED INCOME -> LIFE
Method: Maximum Likelihood
End of Problem
```

Here is the path diagram from these runs:
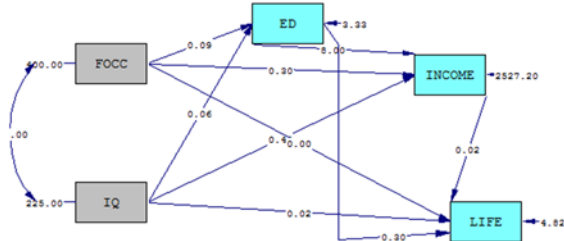
Chi-Square=-0.00, df=0, P-value=1.00000, RMSEA=0.000

Finally, one can in principle specify y and x variables (and latent variables), but omit the model specifications (either relationships or paths), and obtain an empty path diagram on the screen with boxes for observed variables and ellipses for latent variables.   Then you can then specify the model by drawing arrows on the screen:

```
STATUS ATTAINMENT AND LIFE SATISFACTION
Observed Variables: FOCC IQ ED INCOME LIFE
Standard Deviation:
20 15 3 60.00 3
Correlation Matrix:
1.00
.500 1.00
.750 .600 1.00
.450 .390 .535 1.00
.460 .440 .580 .605 1.00
Sample Size
```

```
1000
y-variables:  ED INCOME LIFE
Path Diagram
End of Problem
```

This will give you an empty path diagram on your screen, and you can use the box of arrows and boxes to draw in a path diagram to specify your model.
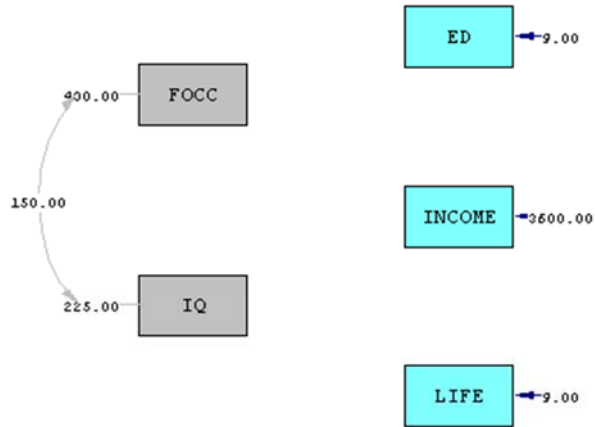
Chi-Square=2918.03, df=10, P-value=0.00000, RMSEA=0.000

Note the bad fit of this empty model!   You can now specify the model's parameters by drawing arrows connecting boxes on your screen.