

# LISREL 8 and PRELIS2: Getting Started

Section 1: Introduction.....	2
Section 2: Reading Raw Data Using PRELIS2 .....	2
2.1 A Detailed PRELIS2 Example .....	4
2.2 Another Detailed Example .....	6
Section 3: Specifying a LISREL Model .....	8
Section 4: Building LISREL Command Files Using LISREL Matrix Language.....	9
Section 5: Building LISREL Command Files Using SIMPLIS.....	19
Section 6: Building LISREL Command Files Using SIMPLIS.....	21
Section 7: Documentation.....	21

## Section 1: Introduction

This tutorial is for those who plan to use the LISREL software to estimate structural equation models (SEMs). You can also use this software to carry out exploratory and confirmatory factor analysis, as well as path analysis. For more information on LISREL, you might visit [Lisrel's home page](#) or our own [FAQs](#). For help with SEMs, LISREL, PRELIS, or SIMPLIS, call 475-9400 to make an appointment to meet with an SSC statistical consultant, or send e-mail to: [stats@ssc.utexas.edu](mailto:stats@ssc.utexas.edu).

In this tutorial, we will familiarize you with the basic steps to analyzing your data with LISREL:

- Enter your raw data or covariance matrix into a data file
- Create a PRELIS2 command file to process (read and transform) the raw data
- Specify your models of interest
- Create a LISREL or SIMPLIS program file to test your models of interest
- Evaluate your LISREL output

## Section 2: Reading Raw Data Using PRELIS2

Structural equation models are statistical models of linear relationships among latent (unobserved) and manifest (observed) variables. LISREL requires the input of a correlation or covariance matrix. You can use PRELIS2 to prepare either of these matrices from your raw data file.

You can read raw data using PRELIS2. The advantage of using PRELIS2 is that it is specifically formulated to handle both ordinal and continuously distributed variables. If you have a choice to use raw data or matrix data, raw data is usually the better choice because PRELIS2 can both analyze and process this data in ways that other programs cannot. Consider the following sample data file, taken from pages 1-19 of the PRELIS2 manual:

1	3	-.7	-.4
2	4	2.3	1.6
3	3	1.2	1.7
1	-9	-.4	-.3
3	2	-1.2	-.7
2	1	-.9	1.2
2	1	.8	.3
3	3	1.6	1.5
1	2	-.9	-.9
1	4	-.8	-.8
1	1	.7	.8
2	2	1.1	1.3

You can use the following PRELIS2 syntax to read this data.

```

First PRELIS2 example program
! The first line of the program can be a title statement.
! Comment statements start with exclamation marks,
! one per line. This example program is from the
! PRELIS2 manual, p. 4-3
DA NI=4 NO=12 MISSING=-9 TREATMENT=listwise
RAW-DATA-FROM FILE=sample.data
LABELS
var1 var2 var3 var4
CONTINUOUS var1 var2 var3 var4
! Output a polyserial correlation matrix
! OUTPUT MATRIX = PMATRIX SM=data.100pm
! Or you can output a Pearson correlation matrix
! if you prefer
!OUTPUT MATRIX = KMATRIX SM=junk.km
! Or a covariance matrix
OUTPUT MATRIX = CMATRIX SM=sample.cov

```

The first line of this program is a title line. It will appear at the top of each page of the output file written by PRELIS2. A number of comment statements follow. Comment statements begin with the exclamation mark (!). Comment statements are ignored by LISREL as it processes the program; they are included to clarify the program's contents for human readers. These are recommended for the same reason title lines are useful: they help clarify complex portions of your programs that may not be clear to you or others who later read your programs. Unlike title lines, however, comments are not reproduced on the program's printed output.

The DA line tells PRELIS2 that there are four items ( $NI = 4$ ) and 12 observations ( $NO = 12$ ) in the data file. Also, missing values are represented by -9 ( $MISSING = -9$ ).  $TREATMENT = listwise$  tells PRELIS2 to use listwise data deletion when cases have missing values. Listwise deletion means that an entire case's values are deleted if the case has one or more missing values on any of the variables. Another popular option for  $TREATMENT$  is pairwise. Pairwise deletion means that missing data for one or both of a pair of variables are deleted, but the rest of the case's data are retained for analyses. If possible, listwise deletion of missing data is a safe approach if you are unsure which approach to take for handling missing data. However, using listwise deletion can significantly reduce the working sample size if many cases have missing data values. There are other missing values options beyond listwise and pairwise data deletion-see the PRELIS2 manual for details.

The RAW-DATA-FROM FILE statement identifies the origin of the data. The LABELS statement labels the variables in the data file. The CONTINUOUS statement tells PRELIS2 which variables are to be treated as continuous data. ORDINAL is the default measurement level for all variables with 15 or fewer levels; CONTINUOUS is the default for variables with 16 or more levels. Depending upon the measurement level of each variable, you may choose to code the variable as CONTINUOUS or ORDINAL. According to the authors of the PRELIS2 manual, using ORDINAL coding for noncontinuous data will yield more appropriate results. See the detailed PRELIS2 example below for an example of this type of coding.

**IMPORTANT: If you use nonCONTINUOUS coding for ANY variable in your data file, PRELIS2 may produce results that do not agree with those generated by other software packages such as SAS and SPSS.**

PRELIS2 allows you to output a number of different matrix formats. PM (polyserial) is the recommended correlation matrix format. However, choosing this format may result in output that differs from the output produced by other statistical software packages because PRELIS2 has been optimized to treat ordinal data as categorical rather than continuous. PRELIS2 can also output standard Pearson correlations--use the KM keyword to obtain these. Finally, the CM keyword will output a covariance matrix.

Using LISTWISE missing-data deletion and treating all variables as continuously distributed yields the following covariance matrix.

```
FIRST PRELIS2 EXAMPLE PROGRAM
                                COVARIANCE MATRIX
                                VAR1      VAR2      VAR3      VAR4
VAR1      .750
VAR2      .000      1.278
VAR3      .300      .103      1.428
VAR4      .362      .044      1.146      1.036

MEANS
      VAR1      VAR2      VAR3      VAR4
      2.000      2.556      .556      .589

STANDARD DEVIATIONS
      VAR1      VAR2      VAR3      VAR4
      .866      1.130      1.195      1.018 (Some output omitted)
```

The covariance matrix, mean, and standard deviation values should match those produced by other statistical software. The output appears in the PRELIS2 output file. You can browse this file with a text editor.

## 2.1 A Detailed PRELIS2 Example

The three data sets used throughout the tutorial are available [here](#) for download. You can copy the files to your own computer to practice using PRELIS2.

In this example, the PRELIS2 syntax reads six variables from the raw data file "data.100".

```
Second PRELIS2 Demonstration Program
DA NI = 6 NO = 100 MISSING = -9 TREATMENT = PAIRWISE
RAW-DATA-FROM FILE = DATA.100
LABELS
CONTIN1 ORDINAL1 ORDINAL2 ORDINAL3 CONTIN2 ORDINAL4
CONTINUOUS CONTIN1 CONTIN2
ORDINAL ORDINAL1 - ORDINAL3 ORDINAL4
OUTPUT RA=data.100raw MATRIX = PMATRIX SM=data.100pm
```

The title line begins the program. The next line is the DA, or DATA line. NI refers to the number of indicator or manifest variables (e.g., survey items). In this case, there are six such variables. NO is the number of observations (cases) in the dataset: 100. MISSING = -9 tells PRELIS2 that the number -9 in the dataset represents missing values. TREATMENT = PAIRWISE tells PRELIS2 to delete missing data based on a pairwise deletion approach instead of a listwise method.

The next line contains the RAW-DATA-FROM FILE statement. It tells PRELIS2 which data file to read. Notice that PRELIS2 assumes that the data are entered in free-field format, which means that each data value in the raw data file is separated by a space, and that spaces are not also used to represent missing data. You can abbreviate this command to RA if you wish.

If you have data in a format different from free-field format, you must use a Fortran format statement to tell PRELIS2 how to read and correctly format the data file. Use of the Fortran format statement is shown in the third example, below.

The LABELS statement labels each of the variables read from the data file. Notice that the LABELS keyword appears on one line of the PRELIS2 program; the actual labels for the six variables appear on the next line.

The next statement is the CONTINUOUS statement. This statement tells PRELIS2 which variables it should treat as continuous variables. In this case, the labeled variables CONTIN1 and CONTIN2 are continuously distributed, so they are so marked with the CONTINUOUS statement. The ORDINAL statement is similar to the CONTINUOUS statement, except that it tells PRELIS2 to treat the variables ORDINAL1 through ORDINAL3, and ORDINAL4 as ordinal rather than continuous. Notice the hyphen (-) used to join contiguous variables in the CONTINUOUS and ORDINAL statements. In this instance, ORDINAL1 through ORDINAL3 are contiguous in the raw data file. However, CONTIN1 separates ORDINAL3 from ORDINAL4, so ORDINAL4 must be listed separately in the ORDINAL statement.

The PRELIS2 manual recommends using ORDINAL and CONTINUOUS statements where they are warranted because the results obtained with these options are more appropriate than those obtained for ordinal variables when the options are not used. However, the use of these options makes it highly likely that correlation and covariance matrices generated with PRELIS2 will not match those created by other statistical software packages such as SAS and SPSS. If this is a concern to you, you may want to carefully evaluate the use of the ORDINAL and other options before you create a correlation or covariance matrix using PRELIS2.

The PMATRIX option for the MATRIX keyword instructs PRELIS2 to output a matrix of Pearson product moment, polychoric, and polyserial correlations that LISREL can then use for its analyses. The type of correlation generated depends on the measurement level of the two variables being correlated. In this example, two continuous variables are correlated with a Pearson product moment correlation. On the other hand, two ordinal variables are correlated with polychoric correlations, and a continuous variable is correlated with an ordinal variable using a polyserial correlation. This is why it is very important to correctly set the measurement level of each variable in the variable list by identifying the variables as ORDINAL or CONTINUOUS.

Other MATRIX options are MMATRIX (matrix of moments about zero), AMATRIX (augmented moment matrix), CMATRIX (covariance matrix), KMATRIX (matrix of Pearson product moment correlations), and MATRIX=OM (matrix of correlations based on optimal scores).

Including an SM=*filename* option on the OUTPUT line will save the PMATRIX to an external file. Any file named with the SM option on the OUTPUT statement line of PRELIS2 will result in a file saved to your working directory named in all capital letters.

A related command, SA = *filename*, allows you to save the asymptotic covariance matrix to an external file. SR = *filename* saves the transformed raw data to an external file. SV = *filename* saves asymptotic variances to an external file. The RA= *filename* command (shown above) allows you to save a copy of the raw data file read by PRELIS2 to a new external data file. You can browse or print this file to make sure that PRELIS2 correctly read your raw data file (especially handy if you are uncertain of the accuracy of your data input statements).

PRELIS2 features many useful data manipulation and transformation functions, too many to demonstrate here. For example, PRELIS2 can recode survey data (such as reverse score questionnaire items), and it can be used to select specific groups of cases for analyses, specify multiple missing values per variable, impute missing data, and save and merge subsets of raw data.

## 2.2 Another Detailed Example

The following example illustrates how to use PRELIS2 to read a raw data file, as well as label and reverse-score several of the Likert survey variables contained in this data file. The original data file, **ex7.raw**, is available [here](#).

```
Third PRELIS2 Demonstration Program
!Example 7B from page 5 of the PRELIS2 manual
!(slightly modified)
DA NI=7
LA
YEAR NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
RA=EX7.RAW FO
(F4.0,138X,6F2.0)
!Assigning Category Labels as follows
!AS = Agree Strongly
!A = Agree
!D = Disagree
!DS = Disagree Strongly
CL NOSAY - INTEREST 1=AS 2=A 3=D 4=DS 5=DK 9=NA
! Reverse coding example
RE NOSAY - INTEREST OLD=1,2,3,4 NEW=4,3,2,1
! Select only the cases with YEAR less than 1954
SC YEAR <1954
OUTPUT MATRIX = PMATRIX SM=EX7.COR
```

As before, the DA and LA lines tell PRELIS2 the number of variables and their labels, respectively. The raw data file is EX7.RAW; it is referenced by the RA statement. The RA statement is followed by the FO statement, which tells PRELIS2 to expect a Fortran format on the next line. You must use a Fortran format when the data are not arranged in a free-field format, as was the case in the previous examples.

Consider the following data:

```
38.60 13.63 16.96
```

To save time and effort, you can enter these data in the raw data file without the decimal points.

```
3860 1363 1696
```

You can include a decimal point in your format specification. The format specification for this set of data would be:

```
3F5.2
```

The 3 refers to the three values present in each row of data. If you had ten variables instead of three, you would use a format of 10F5.2 (assuming each variable had the same format). The 5.2 portion of the format specification means that each data value will have a field spanning 5 total columns in the data file, with two of those columns appearing to right of the decimal point. The format statement appears in your PRELIS2 program on the line directly below the RAW-DATA-FROM FILE line. The parentheses are required with this statement.

The Fortran format in this case tells PRELIS2 to read a variable called YEAR in the first four columns of the raw data file. This is the F4.0 portion of the format. The 138X then tells PRELIS2 to skip over the next 138 columns of the data file. Finally, the 6F2.0 tells PRELIS2 to read six variables in sequence, each with a format of 2.0: two columns of data, with no values after a decimal point.

The CL (Category Labels) statement tells PRELIS2 to label the categories or values of each of the variables in the list from NOSAY to INTEREST.

The RE statement invokes reverse coding, which is particularly useful for reverse scoring of Likert survey items. Following the RE keyword is a list of the variables that are reversed scored, followed by the keyword OLD and the list of values to be reverse scored. Then another keyword, NEW, appears, followed by the new, reverse coded values.

The SC statement allows you to select cases for inclusion in the PRELIS2 analyses and any output matrices you choose to save. In this example, all cases with YEAR values less than 1954 are used in the PRELIS2 analyses and saved to the output correlation matrix, EX7.COR. This correlation matrix can now be directly read into LISREL for further processing.

### Section 3: Specifying a LISREL Model

Once you produce your preprocessed data from PRELIS2 or another statistical software package, you are almost ready to analyze the data using LISREL. But before you can write a LISREL program, you first must specify your models of interest. One of the more straightforward ways of doing this is to draw your model visually, either by hand or with a computer graphics program. For instance, suppose you were interested in testing a model of the stability of alienation over time, as measured by anomia and powerlessness feelings at two measurement occasions, 1967 and 1971, as well as education level and a socioeconomic index. (The original data for this example come from Wheaton, B., Muthén, B., Alwin, D., and Summers, G., 1977, "Assessing reliability and stability in panel models." In D.R. Heise (Eds.): *Sociological Methodology 1977*. San Francisco: Jossey-Bass.)

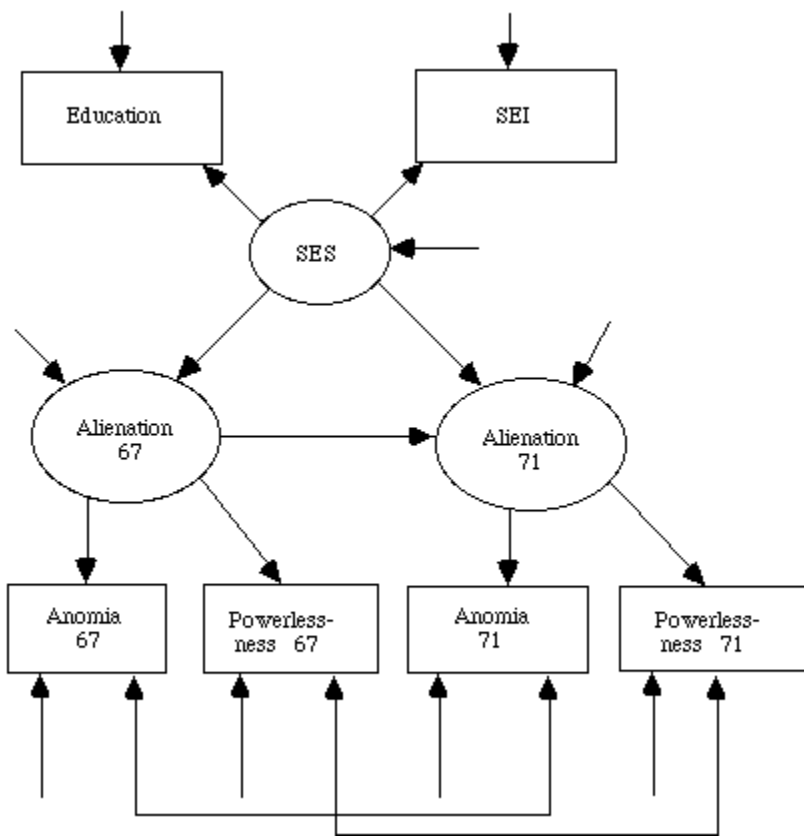


Figure 1: Specifying a model of stability of alienation over time

Figure 1 demonstrates several common characteristics of structural equation models. First, measured (manifest) variables such as Anomia67 are represented by square or rectangular boxes, while nonmeasured (latent) constructs such as Alienation67 are symbolized by circles or ovals.

Single arrows represent a causal impact of one variable upon another, with the head of the arrow pointing towards the variable being influenced by the second variable. For instance, Alienation71 influences or "causes" the responses to the survey items comprised by the measured variable

Anomia71. Of course, no manifest measurement of alienation, powerlessness, or any other latent construct is one hundred percent perfect: there is always measurement error to consider. Measurement error for each variable is represented by the single arrows pointed towards the variable, but with no corresponding variable at the other end of the arrow exerting a causal influence.

Bi-directional arrows refer to correlated or bi-directional relationships (as opposed to causal relationships). In these cases, such as the relationship between the errors of Anomia67 and Anomia71, no statements about causality are being made; only a relationship is discussed. Variables which are at the top or left portion of the model and which exert a causal impact on other variables are said to be *exogenous* or "upstream variables." By contrast, variables which are influenced by other variables are said to be *endogenous* or "downstream variables."

## Section 4: Building LISREL Command Files Using LISREL Matrix Language

Once you have prepared your data and specified your models of interest, you can analyze the data using LISREL. LISREL features two primary ways you can build programs: LISREL matrix syntax and SIMPLIS. The former method allows you a great deal of control in specifying your structural equation models. Furthermore, you can adapt pre-existing LISREL7 programs to run under LISREL8 with few changes.

LISREL8 will run most LISREL7 syntax and generate output similar to that of LISREL7. One notable exception to this general rule is that LISREL8 prints standard errors (in parentheses) and t-values by default underneath each parameter estimate appearing in the output file:

```
LAMBDA-X
          SES
          -----
EDUC          1.00
SEI           .52
              (.04)
              12.36
```

Thus, the TV option on the OUTPUT line (as was the case under LISREL7) is no longer necessary. This change and other new features specific to LISREL8 are documented in the LISREL8 manual.

The program below uses LISREL matrix syntax to construct and test a model of the stability of alienation over time, as measured by anomia and powerlessness feelings at two measurement occasions, 1967 and 1971, and education level and a socioeconomic index.

```
Stability of Alienation
! See LISREL8 manual, p. 207
! Chapter 6.4: Two-wave models
```

```

DA NI=6 NO=932
LA
ANOMIA67 POWERL67 ANOMIA71 POWERL71 EDUC SEI
CM FI=ex64.cov
MO NY=4 NX=2 NE=2 NK=1 BE=SD TE=SY,FI
LE
ALIEN67 ALIEN71
LK
SES
FR LY(2,1) LY(4,2) LX(2,1) TE(3,1) TE(4,2)
VA 1 LY(1,1) LY(3,2) LX(1,1)
OU

```

The LISREL program starts with a title statement, followed by several comment lines. The title and comment lines in LISREL follow the same conventions as those used in PRELIS2, described above.

The DA line contains a NI=6 statement, which tells LISREL that there are 6 indicator (observed) variables. NO refers to the number of observations in the dataset, which is 932 in this example. The LA line labels each of the indicator variables in the order in which they are read from the covariance matrix. In this example, the covariance matrix is read from an external file called **ex64.cov**. The CM portion of this line tells LISREL that you are using a covariance matrix. (You would start the line with the keyword KM if you were reading a correlation matrix from the external file.) Raw data can also be read here by using the RA and FO commands described previously, but it is generally recommended that you preprocess your data using PRELIS2, if possible, to output the data as a covariance matrix for analysis by LISREL.

Be sure to double-check the "Covariance matrix to be analyzed" portion of the LISREL output to verify that LISREL correctly read the covariance matrix. LISREL can also read Pearson and polyserial correlation matrices. The syntax to read these matrices appears below, preceded by LISREL comment statements (indicated by !).

```

! Covariance matrices are usually the best choice to use as
! LISREL input, but if you must use a correlation matrix, the
! polyserial correlation matrix is usually your best choice.
! It codes ordinals as ordinals, continuous as continuous,
! etc.
! PM FI=ex64.pcorr
! OU SE SS!
Or you can use Pearson correlations
! KM FI=ex64.corr
! OU SE SS
! Or you can use covariances
CM FI=ex64.cov
OU SE SS

```

**Important: If you generated your correlation or covariance matrix using SAS, SPSS, or another statistical software package, be sure to add the FU keyword to the KM or CM line of your LISREL program.**

PRELIS2 produces a *lower-triangular* correlation or covariance matrix by default. This is the type of matrix LISREL expects to read. By contrast, other packages produce a *full* correlation or covariance matrix. If you plan to have LISREL read full matrices you create with other software products, you must use the FU keyword.

```
CM FI=ex64.cov FU
```

If you do not add the FU keyword to LISREL, it will try to read a full matrix as a lower-triangular matrix. This will lead to incorrect results.

The MO line defines the model. NY and NX define the number of measurement or observed variables present. Notice that you can have *two* sets of measured variables, one set on each end of a *structural model*. A structural model is the portion of your model composed exclusively of latent variables (as opposed to measurement variables). NX identifies the "starting" or "upstream" side of your model; NY refers to the "finishing" or "downstream" side of your model.

NE and NK define the number of latent variables associated with the observed variables of NY and NX, respectively. In the example program, there are four observed variables (NY) influenced by two latent NE variables. There are two observed variables (NX) influenced by one NK latent variable.

The LE and LK commands allow you to label the latent variables numbered in the NE and NK commands, respectively. In this example, the two NE latent variables are called ALIEN67 and ALIEN71. The single NK variable is called SES.

At this point, we have defined six total measurement or observed variables, which correspond to the indicator variables mentioned in the earlier LA command. We have also defined the three latent variables that we believe influence or determine (at least to some extent) the values of the observed variables.

However, we have not yet defined the paths (relationships) between the latent variables and the observed variables, and the paths between latent variables with other latent variables. This is done with the BE, VA, and FR commands shown above. To use these commands properly, you need to know that LISREL specifies the relationships between variables using eight matrices. By default, most of these matrices have zero values for each element, which means no relationship between variables is present. You must change this default element by element, or use LISREL shortcuts to correctly specify the paths of your model. See the LISREL8 manual for a full description of the eight matrices and the various shortcut options available to you.

You may find it helpful to superimpose the eight LISREL matrices onto your own structural equations model. Notice that some models may not require the use of all eight LISREL matrices. For instance, confirmatory factor analysis models only require the use of the TD, LX, and PH matrices.

In the stability of alienation model, the LY matrix joins the observed variables on the downstream side of the model (NY=4) to the latent variables on that side of the model (NE=2). LY starts out with no values (paths) estimated: you must tell it what relationships to estimate. In

this example, we use the FR command to *free* elements (2, 1) and (4, 2) of the LY matrix to be estimated. The first number in the matrix sequence (the "2" in the (2, 1)) refers to row of the matrix, or the observed variable number. The second number in the matrix sequence (the "1" in (2, 1)) refers to the column of the matrix, or the latent variable number. Thus, LY (2,1) refers to the path between the second observed variable present from the total of four (recall NY=4) and the first latent variable of two (NE=2). Similarly, freeing LY (4, 2) allows LISREL to estimate a relationship (path) between the fourth observed variable and second latent variable.

Observant readers will no doubt have noticed that LY (1, 1) and LY (3, 2) are not freed to be estimated and might wonder why. After all, why not join the first observed variable to the first latent variable and the third observed variable to the second latent variable? The answer is that you must account for *scale indeterminacy*. Latent variables have no units of measurement, so you must set their scale of measurement. You usually do this by fixing the path between the first observed variable and the corresponding latent variable to be 1.00 via the LISREL VA (value) command. Alternatively, you can fix the diagonal values of the PH or PS matrix to 1.00 (this is the SIMPLIS default). If you are undertaking a higher-order factor analysis, two scales are fixed, one in the LY matrix for the lower-order factors and one in the GA matrix for the higher-order factors.

Notice that the LX matrix works in much the same way as the LY matrix, except that it accounts for the relationships between upstream or starting latent variables and observed variables.

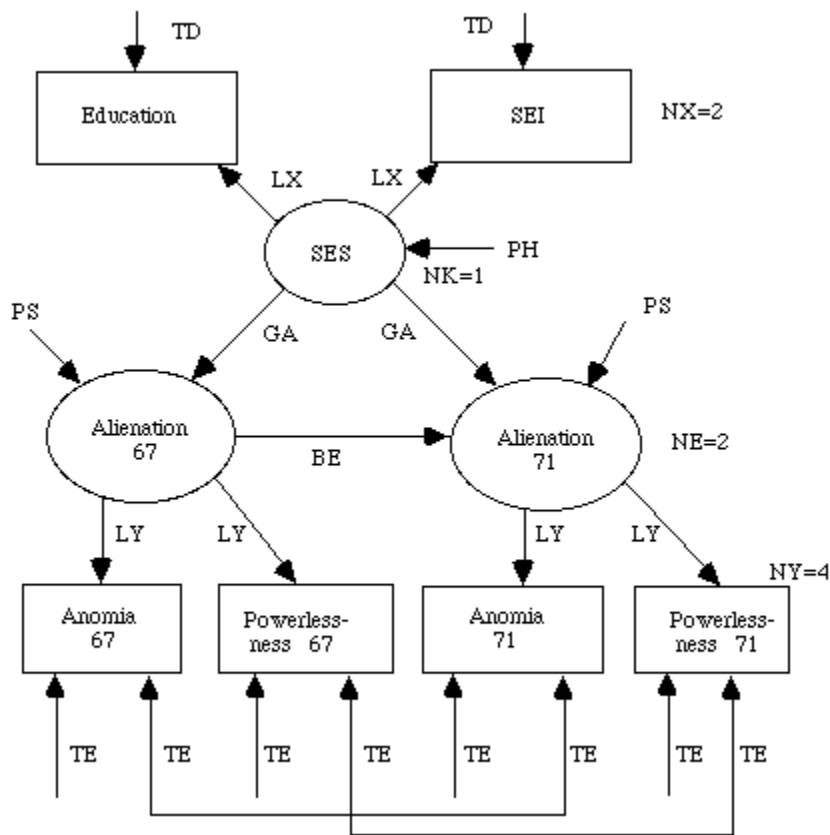


Figure 2: LISREL's eight matrices fitted to the stability of alienation model

The TE and TD matrices are the matrices of errors or residual values for the Y (downstream) and X (upstream) observed variables, respectively. Ordinarily, these matrices are set by default to be *diagonal*. That is, only their diagonal elements, which represent the error variances of each observed variable, are allowed to be estimated. However, in this example TE (3, 1) and TE (4, 2) are allowed to be estimated in what is commonly known as the *correlated errors approach*. Though you would not use the correlated errors approach ordinarily, in this case it is justified due to the repeated-measures nature of the anomia and powerlessness variables. Since the same participants complete both measurements, and both measurements are the same measurement instrument, this strategy makes sense in this instance.

Once we specify the matrices that define the relationships among the error terms and the relationships between the observed variables and latent variables at each side of the model, all that is left to specify is the structural model. The remaining LISREL matrices are used for this purpose. PS defines the intercorrelations and variances among the downstream latent variables. In this case, the default for matrix PS is PS=DI, which fixes the PS matrix to be diagonal--only the variances of the downstream latent variables will be estimated by LISREL. PH defines the intercorrelations and variances among the upstream latent variables, but because there is only one such variable here, SES, the default form of PH=SY (symmetric matrix) is sufficient to estimate the variance associated with SES.

If you were interested in modeling the intercorrelations among the downstream latent variables, as is the case in a confirmatory factor analysis, you would free the off-diagonal elements of the PS matrix for estimation. In our example, however, we hypothesize that ALIEN 67 influences ALIEN 71, and that the inverse is not true--it would make little sense to estimate the effect of alienation at 1971 on alienation at 1967. Thus, the ALIEN67-ALIEN 71 relationship is allowed to be estimated in the BE matrix, which is defined as SD, or *sub-diagonal*. Sub-diagonal means that only the lower off-diagonal elements are estimated. In this case, BE (2, 1) comprises the single lower off-diagonal element, which represents the relationship of ALIEN 67 to ALIEN 71. Put another way, BE (2, 1) represents the "effect of" ALIEN 67 on ALIEN 71.

Finally, the GA matrix defines the relationship between upstream or X latent variables with downstream or Y latent variables. Notice that GA is not mentioned in this example because its default form is FU or *full*. That is, all elements of this matrix are allowed to be estimated by default. In this case, it estimates the "effect of" SES on both ALIEN67 and ALIEN71 using two separate paths represented by GA (1, 1), the effect of SES on ALIEN67, and GA (2, 1), the effect of SES on ALIEN71.

As you can see, properly specifying your model using LISREL's matrices can be very difficult. You can make the task considerably easier by completing the following five steps:

**1. Draw your model before you start to use LISREL (see Figure 1 above).**

**2. Translate your drawn paths, observed variables, and latent variables into matrix notation. For example, the LY matrix above could be represented in the following notation.**

O	Alien67	Alien71
Anomia67	1.00	
Powerless67	*	
Anomia71		1.00
Powerless71		*

where asterisks refer to the paths (relationships) you have freed to allow LISREL to estimate them and 1.00s refer to the fixed values of 1 you include to account for scale indeterminacy. In this instance, rows refer to measured, manifest variables, and columns refer to the latent variables. For the structural model matrices PS, PH, BE, and GA, both rows and columns are latent variables.

The LX matrix contains one fixed value to account for scale indeterminacy, plus one parameter to be estimated.

	SES
Educ	1.00
SEI	*

The TD matrix estimates two error variance values and is a diagonal matrix.

	Educ	SEI
Educ	*	
SEI		*

By contrast, the TE matrix is more complex, due to the correlated error terms.

	Anomia67	Powerless67	Anomia71	Powerless71
Anomia67	*			
Powerless67		*		
Anomia71			*	
Powerless71				*

The PH matrix estimates only one parameter, the variance of the SES latent variable.

	SES
SES	*

The PS matrix estimates the variances of the Alien67 and Alien71 latent variables.

	Alien67	Alien71
Alien67	*	
Alien71		*

The BE matrix estimates the effect of Alien67 on Alien71.

	Alien67	Alien71
Alien67		
Alien71	*	

The GA matrix estimates two effects: the impact of SES on Alien67 and Alien71.

	SES	
Alien67	*	
Alien71	*	

Now compose a matrix table for each of LISREL's eight matrices. Then take your initial drawing (see Figure 1 above) and map the matrix elements onto the drawing, as is shown in Figure 3. Correlated errors are shown in italics.

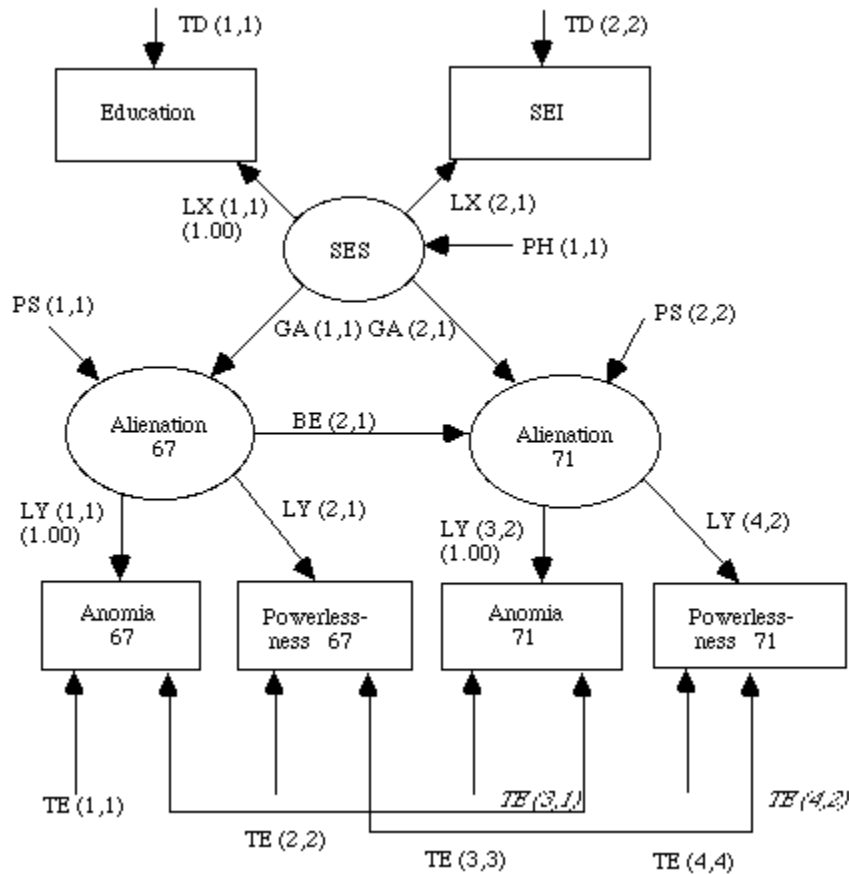


Figure 3: Superimposing LISREL matrix elements onto a model

**3. Keep a reference of the default form of each of LISREL's eight matrices handy.**

A beginner could use the VA command to set all elements of each of the eight LISREL matrices to zero, and then selectively free elements to be estimated, one element at a time. While straightforward, use of this method is not only tedious and time-consuming; it also increases the risk of typographical errors in specifying the correct matrix elements. Each LISREL matrix has a default form and a number of possible forms that you can set. A condensed table of the default matrix forms appears below; a more detailed table appears in the LISREL8 manual. Judicious use of these default and optional matrix forms speeds your programming time and reduces the risk of typographical errors, particularly with large models that have many parameters to be estimated.

Matrix Forms	Default Form	Default Mode	Possible
LY DI, FU	FU	FI	ID, IZ, ZI,
LX DI, FU	FU	FI	ID, IZ, ZI,
BE FU	ZE	FI	ZE, SD,
GA DI, FU	FU	FR	ID, IZ, ZI,
PH ST	SY	FR	ID, DI, SY,
PS SY	DI	FR	ZE, DI,
TE SY	DI	FR	ZE, DI,
TD SY	DI	FR	ZE, DI,

*Table 1: Default and Possible Forms of LISREL Matrices*

In Table 1, ZE refers to a zero-matrix. ID refers to an identity matrix (values of 1.00 appear in the main diagonal, with zero values in every other element of the matrix). IZ and ZI refer to partitioned identity and zero matrices, and zero and partitioned identity matrices, respectively. In practice, these forms are not often used. DI is a diagonal matrix--only the diagonal elements are free to be estimated. This typically occurs in the TE and TD matrices, as the diagonal values of these matrices estimate error variances for measured variables. The off-diagonal elements of these matrices would only be freed under the rare circumstance of correlated errors. SD refers to a subdiagonal matrix, which is the same as the aforementioned lower triangular matrix, a matrix containing values only in the elements that appear below the main diagonal. SY is a symmetric matrix that is not diagonal; ST refers to a symmetric matrix with 1's in the diagonal. Finally, FU refers to a rectangular or square nonsymmetric matrix.

You can use combinations of the default and user-specified possible matrix forms to save time in writing your LISREL model on the MO line. You can then refine the model further by the selective use of the FR, VA, and FI statements until you have your LISREL model correctly specified.

#### **4. Use the LISREL output to double-check your LISREL model specification.**

The LISREL output contains several useful diagnostic sections. First, it displays the covariance or correlation matrix to be analyzed. In the stability of alienation example, variable names and covariances are listed, as shown below.

You can check these values against your raw data covariance matrix to make sure that your LISREL program reads the values correctly. If you generate your covariance or correlation matrix via PRELIS2, this issue is less of a concern, since you presumably checked the accuracy of the data when you preprocessed it using PRELIS2. On the other hand, had you read the raw data or a covariance or correlation matrix directly into LISREL, you should double-check that the matrix to be analyzed is accurate.

COVARIANCE MATRIX TO BE ANALYZED

	ANOMIA67	POWERL67	ANOMIA71	POWERL71	EDUC	SEI
ANOMIA67	11.83					
POWERL67	6.95	9.36				
ANOMIA71	6.82	5.09	12.53			
POWERL71	4.78	5.03	7.50	9.99		
EDUC	-3.84	-3.89	-3.84	-3.62	9.61	
SEI	-2.19	-1.88	-2.17	-1.88	3.55	4.50

LISREL also displays the values of each of its active matrices. Values not estimated are labeled with a zero value, while values to be estimated are assigned numbers ranging from the number one for the first free parameter to be estimated through k, where k is the final parameter to be estimated.

PARAMETER SPECIFICATIONS

LAMBDA-Y

	ALIEN67	ALIEN71
ANOMIA67	0	0
POWERL67	1	0
ANOMIA71	0	0
POWERL71	0	2

LAMBDA-X  
SES

	SES
EDUC	0
SEI	3

BETA

	ALIEN67	ALIEN71
ALIEN67	0	0
ALIEN71	4	0

GAMMA  
SES

	SES

```

ALIEN67          5
ALIEN71          6

  PHI
  SES
  -----
                7

  PSI
  ALIEN67        ALIEN71
  -----
                8          9

  THETA-EPS
  ANOMIA67        POWERL67    ANOMIA71    POWERL71
  -----
ANOMIA67          10
POWERL67          0          11
ANOMIA71          12          0          13
POWERL71          0          14          0          15

  THETA-DELTA
  EDUC            SEI
  -----
                16          17

```

This model estimates a total of 17 parameters. Figure 3 shows a total of 20 matrix elements, each corresponding to a particular path coefficient or variance to be estimated. However, LX (1, 1), LY (1, 1), and LY (3,2) are each fixed to a value of 1.00 to prevent scale indeterminacy. Subtracting these three parameters from the original 20 yields the 17 free parameters to be estimated, as shown in the LISREL output.

## 5. Count degrees of freedom.

Another method you can use to check your model specification is to count the degrees of freedom used in several of the goodness-of-fit statistics produced by LISREL. The first chi-square statistic shown in the output below is labeled, "Chi-square with 4 degrees of freedom". This test statistic tests the null hypothesis that the specified model is a good fit to the data. Thus, a *small* chi-square value with a correspondingly *large* p-value is desirable for this statistic.

```

GOODNESS OF FIT STATISTICS
  CHI-SQUARE WITH 4 DEGREES OF FREEDOM = 4.73 (P = 0.32)
  ESTIMATED NON-CENTRALITY PARAMETER (NCP) = 0.73
  90 PERCENT CONFIDENCE INTERVAL FOR NCP = (0.0 ; 10.53)

  MINIMUM FIT FUNCTION VALUE = 0.0051
  POPULATION DISCREPANCY FUNCTION VALUE (F0) = 0.00079
  90 PERCENT CONFIDENCE INTERVAL FOR F0 = (0.0 ; 0.011)
  ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.014
  90 PERCENT CONFIDENCE INTERVAL FOR RMSEA = (0.0 ; 0.053)
  P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 0.93

  EXPECTED CROSS-VALIDATION INDEX (ECVI) = 0.042

```

90 PERCENT CONFIDENCE INTERVAL FOR ECVI = (0.041 ; 0.052)  
ECVI FOR SATURATED MODEL = 0.045  
ECVI FOR INDEPENDENCE MODEL = 2.30

CHI-SQUARE FOR INDEPENDENCE MODEL WITH 15 DEGREES OF FREEDOM = 2131.40  
(Further output not shown)

The chi-square test for the goodness of fit of the alienation model has 4 degrees of freedom. The maximum number of degrees of freedom that any model can use is  $k(k+1)/2$ , where  $k$  is equal to the number of manifest items or variables in the analysis. In this example, there were a total of six variables used in the analysis: ANOMIA67, POWERL67, ANOMIA71, POWERL71, EDUC, and SEI. There are  $6(6+1)/2 = 6*7/2 = 42/2 = 21$  total degrees of freedom available to estimate parameters. There are 21 available degrees of freedom - 17 used for parameter estimation = 4 remaining degrees of freedom, which is precisely the number of degrees of freedom appearing in the chi-square test.

This chi-square test should not be confused with the "Chi-square for independence model," which appears farther down in the goodness-of-fit statistics printout. This test statistic, with 15 degrees of freedom, is a test of one so-called "null model," or "starting model." Some researchers employ *nested models* to test their hypotheses, wherein they compare their models of interest to a baseline or starting model. The null model provided by LISREL as its "independence" model is one such model: it consists only of estimates of the error variances associated with each of the measured variables.

In the stability of alienation model, there are six total measurement variables: ANOMIA67, POWERL67, ANOMIA71, POWERL71, EDUC, and SEI. Each of these has an estimate of measurement error associated with it, which means that LISREL estimates 6 parameters for the null model. Using the same 21 available degrees of freedom as we did in the previous example,  $21 - 6 = 15$  degrees of freedom remaining, which is equivalent to the number of degrees of freedom shown in the independence (null) model chi-square test statistic.

## Section 5: Building LISREL Command Files Using SIMPLIS

As you can see, building and testing LISREL models can be a complex endeavor. To address this problem, the creators of LISREL created SIMPLIS, a new command language exclusive to LISREL8. It was added to LISREL8 as a new feature, designed to make structural equation modeling easier than using previous releases of LISREL.

The following sample program appears on page 30 of the *LISREL8: SIMPLIS Command Language* manual. The name of the file (which includes both SIMPLIS command syntax and the data in covariance matrix form) is called **ex6a.spl**. This program is functionally equivalent to the LISREL program shown above (both programs should yield a model with 4 degrees of freedom and a chi-square value of 4.73).

```
STABILITY OF ALIENATION
OBSERVED VARIABLES
  ANOMIA67   POWERL67   ANOMIA71   POWERL71   EDUC   SEI
```

```

COVARIANCE MATRIX
  11.834
   6.947   9.364
   6.819   5.091  12.532
   4.783   5.028   7.495   9.986
  -3.839  -3.889  -3.841  -3.625   9.610
  -2.190  -1.883  -2.175  -1.878   3.552   4.503
SAMPLE SIZE 932
LATENT VARIABLES ALIEN67 ALIEN71 SES
RELATIONSHIPS
  ANOMIA67 POWERL67 = ALIEN67
  ANOMIA71 POWERL71 = ALIEN71
  EDUC SEI = SES
  ALIEN67 = SES   ALIEN71 = ALIEN67 SES
LET THE ERRORS OF ANOMIA67 AND ANOMIA71 CORRELATE
LET THE ERRORS OF POWERL67 AND POWERL71 CORRELATE
LISREL OUTPUT
END OF PROBLEM

```

As you can see, SIMPLIS derives its name from the easy-to-use syntax. Like the examples shown above, the first line of the program is a title. The next line of the program, OBSERVED VARIABLES, tells LISREL what the observed variables are called and how many exist.

In this example, the COVARIANCE MATRIX is read directly in the program. You could read in the data from an external file such as a correlation matrix if your data were in that format by substituting the syntax

CORRELATION MATRIX FROM *file*

where *file* refers to the name of the data file in the current working Unix directory.

The SAMPLE SIZE tells LISREL the number of cases to use in its goodness-of-fit test statistics. The LATENT VARIABLES line is where you specify the names of the latent variables in your model. Once your OBSERVED VARIABLES and LATENT VARIABLES are defined, the next step is to join them with paths. That is, you allow specific observed and latent variables to correlate with each other. LISREL implements this procedure with its RELATIONSHIPS command. For example, in this model of alienation stability, the observed variables ANOMIA67 and POWERL67 are each allowed to have a relationship with the latent variable ALIEN67. Similarly, ANOMIA71 and POWERL71 have a relationship with the latent variable ALIEN71. Education level (EDUC) and the Socioeconomic Index (SEI), have a relationship with a latent variable called SES (Socio-Economic Status).

Notice that latent variables may have relationships not only with observed variables, but also with latent variables, as is the case in the relationship between ALIEN67 and SES. Since these data are repeated measurements on the same individuals, it makes sense to correlate the errors of anomia and powerlessness measures for 1967 and 1971. The LISREL OUTPUT statement is optional: if you use this statement in your program, LISREL will write its results in the matrix format shown in the previous section. This can be a useful diagnostic tool, particularly if you are unsure which parameters LISREL estimates. The END OF PROBLEM command tells LISREL that the program is complete.

## **Section 6: Building LISREL Command Files Using SIMPLIS**

Since you can obtain identical results by using SIMPLIS or LISREL, which approach should you take to constructing and testing your models?

There is no definitive right or wrong approach to use, but each approach has its advantages and disadvantages. Among the advantages of SIMPLIS are its ease of use and flexibility. It also represents the future of LISREL computing. On the other hand, the original LISREL language, while having the disadvantage of being somewhat arcane and difficult to decipher, has several advantages. First, it is backwards compatible. This means that if you wrote programs on the UTXVM mainframe or some other computer for LISREL7, these programs will still run (with a few minor modifications) on LISREL8. Second, the original LISREL language gives you much more control over the exact type of model you test. For example, instead of allowing the SIMPLIS program to decide for you, you can choose which relationships to estimate, and where to account for scale indeterminacy. There can be a danger in placing faith in user-friendly programs that go too far in making assumptions for the end user--a program may make an assumption for you that you are unwilling to make.

Ultimately, perhaps the best answer to the question, "Which LISREL language should I use?" is "Both." If you can successfully model your data with both SIMPLIS and the original LISREL language, obtain the same results with both analyses, and understand the assumptions and decisions you make when you use each approach, you can be more confident that your results truly reflect the model you wish to test. If you obtain discrepancies between the results from the two approaches, you know that at least one (and perhaps both) of your programs are not estimating your model of interest.

## **Section 7: Documentation**

LISREL users will find several documents worth acquiring.

*LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*

*PRELIS 2: User's Reference Guide*

*LISREL 8: User's Reference Guide*

These manuals are written and published by:

Karl Joreskog and Dag Sorbom  
Scientific Software International, Inc.  
7386 N. Lincoln Avenue, Suite 100  
Lincolnwood, IL 60712-1747

Since structural equations modeling is a very complex multivariate statistical approach, you should also consider acquiring a textbook on the topic. Several authors have written overviews,

including John Loehlin (*Latent Variable Models*, 1992) and Ken Bollen (*Structural Equations with Latent Variables*, 1989).