M PLUS

GENERAL NOTES

1. Mplus correlates all exogenous variables by default, unless they are latent variables

2. Mplus estimates all residual variances by default

3. No line can be more than 90 characters long.

4. Many of the subcommands can be abbreviated using the first three or so characters of them

5. To see a diagram of your model, after running it from Mplus, click on the menu item in the interface called 'Diagram' and then click on 'View Diagram.'

TITLE LINE (Title:)

This line is optional. Can use as many lines as want. Just hit return and keep typing. A line can not be more than 60 characters.

DATA LINE (Data:)

This line defines where to get the input data.

FILE IS c:\temp\mplus\ex3.1.dat; tells M Plus where to get the data file. The data usually are in free format in an ascii file.

The number of variables it reads in the above commands is linked to the number of names on the names command below. Five names means it reads in five variables per case.

Here are some other options with the DATA line:

Type=Individual; is the default and reads in raw data, respondents by variables
Type=COVARIANCE; reads in lower triangular covariance matrix (with diagonals)
Type=CORRELATION; reads in lower triangular correlation matrix (with diagonals)
Type=FULLCOV; reads in full covariance matrix
Type=FULLCORR; reads in full correlation matrix
Type=MEANS; reads in means
Type=STDEVIATIONS; reads in standard deviations
Type=IMPUTATION; reads in imputed data sets to be analyzed

Here is an example of reads in a covariance matrix, means and standard deviations that are stored in a file:

```
DATA: FILE IS c:/mplus/covariance2/test3.txt ;
TYPE IS MEANS STDEVIATIONS CORRELATION ;
NOBSERVATIONS = 101;
VARIABLE:
  NAMES ARE y x1 x2 x3 t ;
```

Here is what the data look like in the input file (the five means first, then the SDs and then the correlation matrix:

```
0 0 0 0 0
1 1 1 1 .5
```

```
1.0
.20  1.00
.20  .00   1.00
.20  .00  .00   1.00
.30  .00  .00  .00  1.00
that reads
```

NOBS=2000; is the number of observations to set N at when summary data are analyzed. Not necessary to specify this when the input is for raw data

NGROUPS=3; specifies 3 groups in a multi-group solution. Default is 1 group and do not need to specify if you are analyzing only one group. See **GROUPING IS** command in the variable line section below for how to specify the groups themselves.

VARIABLE LINE (Variable:)

This line defines variable names and variable types.

NAMES ARE y x1 x2; Gives the names of the variables to be read in, in the order they occur in the input data set. A name can not be longer than 8 characters. Can not use commas in a name or blanks.

NAMES ARE y1-y5; will generate the names y1 y2 y3 y4 y5.
NAMES ARE jima-jimd; will generate the names jima jimb jimc jimd.

USEVARIABLES ARE y x1; Specifies the subset of variables to use if you only want to work with a subset of the variables input in the previous command. Default is to use all variables that were input.

USEOBSERVATIONS = ethnic EQ 1 AND gender EQ 2; uses only observations whose scores are 1 on ethnic and 2 on gender. Can invoke AND, OR, NOT (logical not) EQ (equal to) NE (not equal to) GE (greater than or equal to) LE (less than or equal to) GT (greater than) LT (less than). For example,
USEOBSERVATIONS = ethnic GT 1

MISSING ARE .; indicates a period in the entry as indicating a missing value for all variables

MISSING ARE ALL(-9999); indicates a value of -9999 is a missing value for all variables

MISSING ARE ethnic(7 99) y1(8); indicates that the values 7 and 99 represent missing values for the variable ethnic and the value 8 represents a missing value for variable y1.

MISSING ARE ALL(7 99) indicates that the values 7 and 99 represent missing values for all variables.

CENSORED ARE y(a) y1(b); Specifies which endogenous variables are censored, with the letter a or b in the parentheses indicating censoring above or below. The censoring points are determined empirically.

CENSORED ARE y(ai) y1(bi); Specifies which endogenous variables are censored inflated, with the letter ai or bi in the parentheses indicating censoring above or below with inflation. The censoring points are determined empirically.

CATEGORICAL ARE y1 y2; Specifies y1 and y2 are categorical outcomes. For binary or ordered variables, zero must be the lowest category. If they are not, then they will automatically be recoded this way by M Plus. If the values of an ordered variable are 2, 5, 8, and 9, M Plus will recode these to 0, 1, 2, 3.

NOMINAL ARE y1 y2; Specifies which endogenous variables are treated as unordered, categorical variables. Limit is 10 categories per variable. The values are automatically renumbered by M Plus to begin with the number 0. In the Model command, you sometimes must refer to a specific category on a nominal variable. This is done with the # sign. So x1#1 refers to category 1 on the variable x1.

COUNT are y1 y2; Specifies which outcome variables are to be treated as count variables for Poisson like regression.

COUNT are y1(1) y2(i); Specifies which outcome variables are to be treated as count variables for Poisson like regression with zero inflated models.

GROUPING IS gender(1=male 2=female); For multiple group analysis specifies the grouping variable and the values and labels to use for that variable.

WEIGHT is x1; Specifies the weight variable in a complex sampling design to use in the analysis. Weights are not allowed with bootstrapping.

CLUSTER is x2; Specifies, in a complex sample design, the cluster information (e.g., school, household).

CENTER x1 x2 x3 x4 (GRANDMEAN); Specifies grand mean centering for variables x1, x2, x3 and x4. Centering is performed across all cases. Can use in any program (e.g., linear regression) to mean center the variables.

CENTER x1 x2 x3 x4 (GROUPMEAN); Specifies group mean centering for variables x1, x2, x3 and x4. Usually only used in multilevel analyses.

WITHIN = y1 y2 x1; Used in a multilevel design and specifies level 1 variables. They cannot be used as level 2 variables.

BETWEEN = y1 y2 x1; Used in a multilevel design and specifies level 2 variables. They cannot be used as level 1 variables.

DEFINE LINE (Define:)

This line allows for transformations of data that are input into M Plus

x1 = log(x1); takes the log of x1 and outs it in x1.

Can use the logical operators in if statements AND, OR, NOT (logical not) EQ (equal to) NE (not equal to) GE (greater than or equal to) LE (less than or equal to) GT (greater than) LT (less than). For example,
If (gender EQ 1 and ses EQ 2) then group = 1.

Can use addition (=), subtraction (-), multiplication (*), division(/), exponenitation(**) and remainder (%), i.e., the remainder of y/x. Also have log, log10, exp, sqrt, abs, sin, cos, tan, asin, acos, atan.

CUT y1(30,40); takes the continuous variable y1 and breaks it into 3 groups, (1) less than or equal to 30, (2) greater than 30 but less than 40 and (3) greater than or equal to 40. Can do multiple variables at a time with the same cutpoints
CUT y1-y7(30,40);
CUT y1 y2 y3 (30,40);

ANALYSIS Line (Analysis:)

Esti=ML; specifies a maximum likelihood analysis

Esti=MLM; specifies the Satorra-Bentler scaled chi square analysis

Esti=MLM; specifies the Huber-White robust estimator

Esti=WLSMV; specifies a weighted least squares with mean and variance adjustments type estimator

Bootstrap=2000; specifies 2000 bootstrap samples. You can not get modification indices with bootstrap. Once you specify bootstrap, all significance tests and CIs are bootstrap based.

Type=MeanStructure; Specifies you want means and intercepts in the analysis. This is the default starting in version 5.0

Type=Random; Specifies random intercept and random coefficient models.

Type=Complex; Specifies you want to compute standard errors and chi squares for cluster and complex sampling designs.

Type=Mixture; Specifies you want to compute models with a combination of continuous and categorical latent variables.

Type=TwoLevel; Specifies you want to analyze a multilevel model.

LISTWISE=ON FIML is the default missing data algorithm for analyses. To obtain listwise deletion, use **LISTWISE=ON** as a DATA subcommand.

MODEL LINE (Model:)

This line defines the model to be analyzed.

BY is short for "measured by" and defines the latent variable. For example, f1 BY y1 y2 y3;. First variable listed is set as marker variable. If pout a * after the observed variable, then this says to estimate the loading. A @ says to fix the loading at a value. For example, f1 BY y1@1 y2* y3*; The defaults of M Plus make this statement equal to f1 BY y1 y2 y3;.

ON describes a regression equation. For example, y on x1 x2; regresses y onto x1 and x2. Default is to allow exogenous variables to be correlated and to estimate an error variance for an endogenous variable.

WITH describes a correlation or covariance. f1 WITH f2; means correlate f1 with f2. Default is that all exogenous variables are correlated and the correlations are estimated.

PWITH pairs the variables on the left side with those on the right side for correlations. For example

y1 y2 y3 PWITH y4 y5 y6;

means correlate y1 WITH y4; y2 WITH y5; y3 WITH y6;

To **fix the variance** of an exogenous variable, use the @ convention. For example f1@1 f2@1 fixes the variances of the two factors f1 and f2 to be 1.0.

A list of observed or latent variables refers to the **variances or residual variances** of those variables. For example

```
y1* y2* y3*;
```

refers to the variance of y1, y2 and y3 if they are exogenous variables and the residual variances if they are endogenous variables.

y1 WITH y2 allows **correlated error** between y1 and y2 if they are both exogenous variables

Means and intercepts are referred to by variables in brackets. For example

```
[y1 y2 y3];
```

refers to the means of y1 y2 and y3 if they are exogenous and the intercepts if they are endogenous. Default is to estimate all means and intercepts of observed variables and to fix latent variables means and intercepts at 0. In multiple group analysis, default is to fix means and intercepts of first group at 0 and estimate the other groups.

Parameters can be constrained to be equal by placing the same number in parentheses at the end of a line. For example

```
y1 ON x1 x2 (1)
y1 ON x3;
```

regresses y1 onto x1, x2 and x3 and forces the x1 and x2 coefficients to be equal. Or,

```
y1 ON x1 (1)
y1 ON x2;
y2 ON x3 (1)
y2 ON x4;
```

regresses y1 onto x1 and x2 and y2 onto x3 and x4 and forces the coefficients for x1 and x3 to be equal.

As another example,

```
f1 BY y1 y2 y3 y4 y5 y6 (1)
f1 BY y7 y8 y9  (1);
```

forces factor loadings for y2 to y9 to be equal (the loading for y1 is set to 1.0 by default). But

```
f1 BY y1 y2 y3 y4 y5 y6
f1 BY y7 y8 y9  (1);
```

forces the loadings for y7-y9 to be equal but not the others (again with the y1 loading being fixed at 1.0).

```
f1 BY y1-y4 (1)
f1 BY y5-y6 (2)
f1 BY y7-y9 (3);
```

sets y1 as the marker variable and then forces y2-y4 to be equal, y5-y6 to be equal, y7-y9 to be equal.

[y1 y2 y3](1); **forces the means** (or intercepts) of y1 y2 and y3 **to be equal.**

To **label a parameter**, put the label after the parameter in parentheses

```
y on X1(p1)
y on X2(p2)
y on X3(p3);
```

This example labels the path for x1 to y as p1, x2 to y as p2 and x3 to y as p3. Only one label can appear per line.

MODEL CONSTRAINT LINE (Model Constraint:)

For **non-linear constraints on parameters**, first label the parameters, as above. Then add the statement to impose the constraint:

```
MODEL CONSTRAINT:
    p1 = p2*p3;
```

MODEL INDIRECT LINE (Model Indirect:)

```
MODEL INDIRECT:
    y3 IND x1;
```

does an indirect effect and total effect analysis using y3 as the outcome variable and x1 as the distal variable. Must specify for each distal-outcome variable you want to examine. For example

```
MODEL INDIRECT:
  y3 IND x1;
  y3 IND x2;
  y3 IND x3;
```

will do mediation analysis using y3 as outcome and x1 as distal variable; y3 as outcome and x2 as distal variable; and y3 as outcome and x2 as the distal variable.

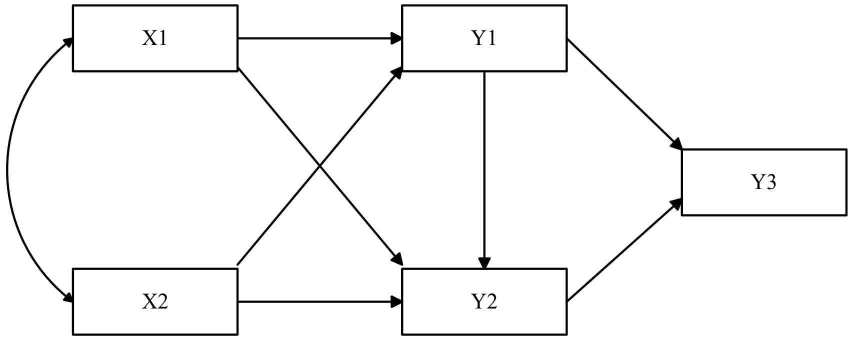
If use with bootstrap, then get bootstrapped CIs

If you want to focus on specific mediated pathways among multiple mediators, specify them using VIA:

```
MODEL INDIRECT:

  y3 VIA y1 x1;
```

This command in the following model requests all indirect effects from x1 to y3 that are mediated by y1. These include x1 to y1 to y3, and x1 to y1 to y2 to y3:



SAVEDATA Line (Savedata:)

Saves data. Missing values are saved as a *.

FILE IS c:\temp\jim.txt; saves the results in jim.txt. Saves in ascii file

FORMAT IS FREE; saves data in free format

SAVE = CPROB; saves in a mixture model the probability that an individual is in each class as well as the class number that has the highest probability for that individual.

SAVE = FSCORES; saves factor scores for latent variables.

OUTPUT Line (Output:)

Samp prints basic descriptive statistics of input variables

StdYX prints standardized solution

Residual prints the predicted covariances and the (unstandardized) residuals for the predicted and observed covariances

Mod(All 4) prints modification indices greater than 4.0. Default cut-off is 10

Cinterval prints 95% and 99% confidence intervals

Cinterval(bcbootstrap) prints 95% and 99% bootstrapped confidence intervals. To use this, you must specify bootstrap in the analysis command. You can not get modification indices with bootstrap. Can't get both regular CIS and bootstrap CIs in the same run.

Residual shows residual analysis of predicted and observed parameters. Need this to examine disparities between predicted and observed covariances

Tech4 shows estimated means and covariances and correlations among the latent variables

Tech1 shows numbering scheme used for parameter estimates by MPlus. Good for use with error messages.

Sample Programs to Illustrate Programming Code

Multiple Regression With All Single Indicators

```
TITLE: Example of a multiple regression
DATA: FILE IS c:\temp\mplus\ex3.1.dat;
VARIABLE: NAMES ARE y1 x1 x2; MISSING ARE ALL(-9999);
MODEL: y1 ON x1 x2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Multiple Regression with Bootstrapping

```
TITLE: Example of a multiple regression with bootstrap
DATA: FILE IS c:\temp\mplus\ex3.1.dat;
ANALYSIS: Estimator: ML ; Bootstrap=2000;
VARIABLE: NAMES ARE y1 x1 x2; MISSING ARE ALL(-9999);
MODEL: y1 ON x1 x2;
OUTPUT: Samp StdYX Residual Tech4 Cinterval(BCBootstrap)
```

Notes: Bootstrap=2000 indicates the number of bootstrap replicates and CInterval(BCBootstrap) indicates you want bootstrap confidence intervals. Modification indices are not available with bootstrapping

Censored Regression

```
TITLE: Example of regression with censored dependent variable
DATA: FILE IS c:\temp\mplus\ex3.2.dat;
VARIABLE: NAMES ARE y1 x1 x3;
          CENSORED ARE y1 (b);
          MISSING ARE ALL(-9999);
MODEL: y1 ON x1 x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: By default, MPlus uses a robust weighted least squares procedure. Can also used robust ML if want. This model works with a censored dependent variable called y1, with censoring below. The point of censoring is determined empirically.

Censored Inflated Regression

```
TITLE: Example of censored-inflated regression
DATA: FILE IS c:\temp\mplus\ex3.3.dat;
VARIABLE: NAMES ARE y1 x1 x3;
          CENSORED ARE y1 (bi);
          MISSING ARE ALL(-9999);
MODEL: y1 ON x1 x3;
       y1#1 ON x1 x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Note on CENSORED command we use bi instead of b in the parentheses. Two regressions are estimated, one binary (the second equation on the MODEL line) and one continuous. The continuous is for people who are above the censoring point. The binary is for those above versus below the censoring point and is a logistic regression, i.e. it predicts those above versus below the censoring point. The inflation variable is referred to by adding the # sign to the censored variable and then the number 1.

Poisson Regression

```
TITLE: Example of Poisson regression
DATA: FILE IS c:\temp\mplus\ex3.7.dat;
VARIABLE: NAMES ARE y1 x1 x3;
          COUNT IS y1;
          MISSING ARE ALL(-9999);
MODEL: y1 ON x1 x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Zero Inflated Poisson Regression

```
TITLE: Example of zero inflated Poisson regression
DATA: FILE IS c:\temp\mplus\ex3.8.dat;
VARIABLE: NAMES ARE y1 x1 x3;
          COUNT IS y1 (i);
          MISSING ARE ALL(-9999);
MODEL: y1 ON x1 x3;
       y1#1 ON x1 x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Note on COUNT command we use i in the parentheses. Two regressions are estimated, one binary (the second equation on the MODEL line) and one continuous. The inflation variable is referred to by adding the # sign to the censored variable and then the number 1.

Multinomial Logistic Regression

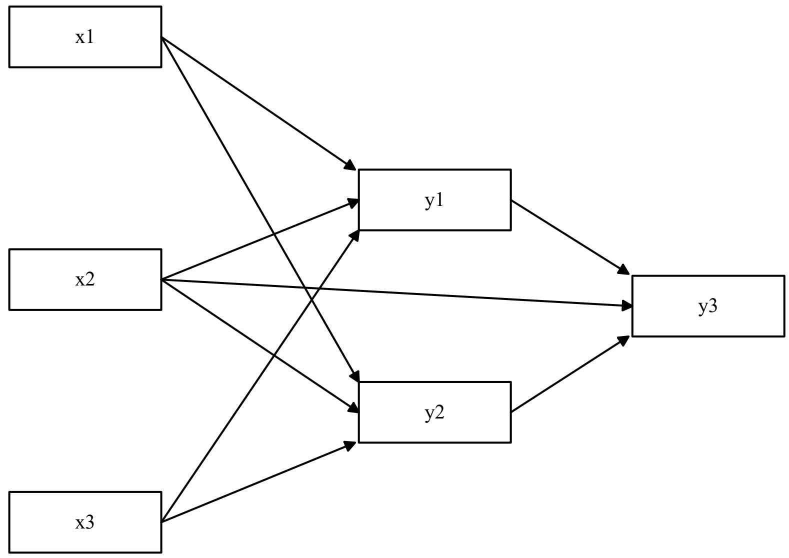
```
TITLE: Example of multinomial logistic regression
DATA: FILE IS c:\temp\mplus\ex3.6.dat;
VARIABLE: NAMES ARE u1 x1 x3;
          NOMINAL IS u1;
MODEL: u1#1 u1#2 ON x1 x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: The outcome variable has three categories. u1#1 refers to the first category and u1#2 refers to the second category. In this example the third group (u1#3) is the reference group. The model uses x1 and x3 as predictors of group membership.

All Single Indicator Path Model

The path model tested is on the next page:

```
TITLE:  Example of single indicator variables
DATA: FILE IS c:\temp\mplus\ex3.11.dat;
VARIABLE: NAMES ARE y1-y3 x1-x3;
          MISSING ARE ALL(-9999);
MODEL:  y1 y2 ON x1 x2 x3;
        y3 ON y1 y2 x2;
MODEL INDIRECT:  y3 IND x1; y3 IND x2; y3 IND x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```



All Single Indicator Path Model with Categorical/Ordinal Outcomes

The path model is the same as the one in the previous example, but the outcome variables are binary and/or ordinal instead of continuous.

```
TITLE:  Example of single indicator variables with Categorical Outcomes
DATA: FILE IS c:\temp\mplus\ex3.12.dat;
VARIABLE: NAMES ARE y1-y3 x1-x3;
          CATEGORICAL ARE y1-y3;
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL:  y1 y2 ON x1 x2 x3;
        y3 ON y1 y2 x2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Use of ML on Analysis line produces a logistic regression that involves numerical integration. Can’t use indirect effects command with logistic model, nor can one get modification indices. Could do if switched to probit analysis using the default estimator for this program instead of ML.

Two Factor Measurement Model Each With 3 Indicators

```
TITLE:  Example of CFA
DATA: FILE IS c:\temp\mplus\ex5.1.dat;
VARIABLE: NAMES ARE y1-y6;
          MISSING ARE ALL(-9999);
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: First indicator is by default fixed at 1. Correlation between factors is assumed and estimated.

Two Factor Measurement Model Each With 3 Indicators and Correlated Error

```
TITLE:  Example of CFA with correlated error
DATA: FILE IS c:\temp\mplus\ex5.1.dat;
VARIABLE: NAMES ARE y1-y6;
          MISSING ARE ALL(-9999);
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       y2 WITH y5;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Correlated error for the y2 and y5 indicators is introduced by y2 WITH y5 statement.

Two Factor Measurement Model With Binary/Ordinal Indicators

```
TITLE:  Example of CFA with binary or ordinal indicators
DATA: FILE IS c:\temp\mplus\ex5.2.dat;
VARIABLE: NAMES ARE y1-y6;
          CATEGORICAL ARE y1-y6;
          MISSING ARE ALL(-9999);
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Uses a robust weighted least squares estimator and polychoric/tetrachoric correlations.

Higher Order Factor Model

```
TITLE:  Example of higher order factor model
DATA: FILE IS c:\temp\mplus\ex5.6.dat;
VARIABLE: NAMES ARE y1-y12;
          MISSING ARE ALL(-9999);
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f3 BY y7-y9;
       f4 BY y10-y12;
       f5 BY f1-f4*;
       f5@1
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

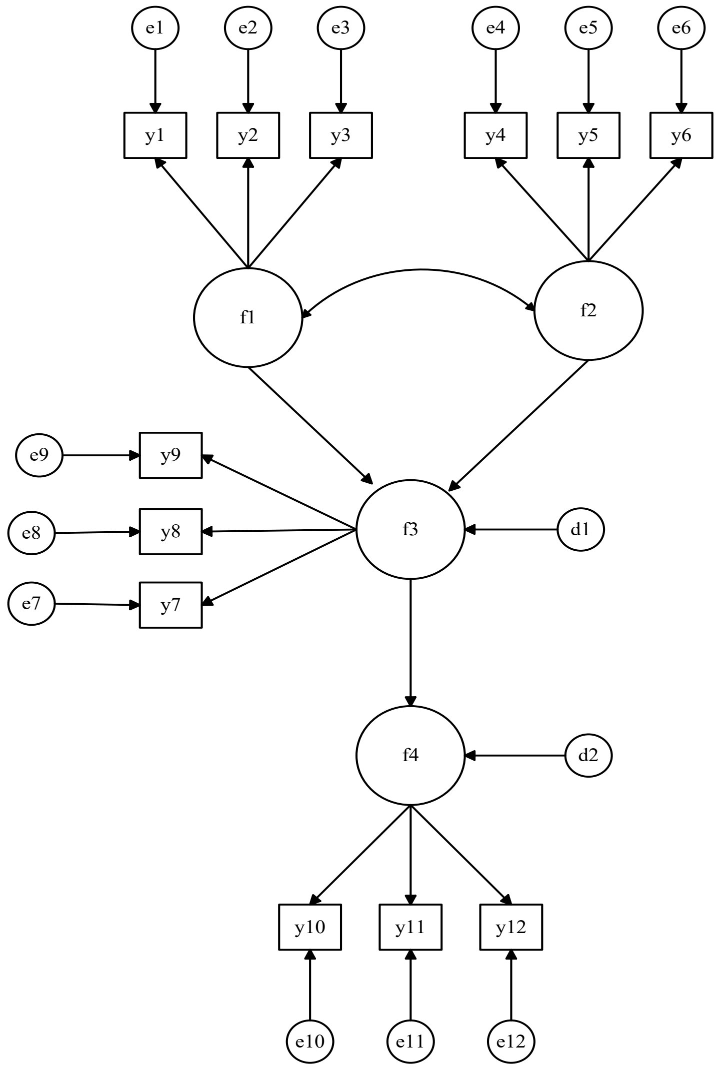
Notes: For higher order factor models, must define the lower order factors before specifying the higher order factor on the model line. This program sets variance of f5 (the higher order factor) to 1 and estimates all the loadings for it on the first order factors.

Latent Variable SEM Model

The model being estimated is on the next page.

```
TITLE:  Example of latent variable SEM model
DATA: FILE IS c:\temp\mplus\ex5.11.dat;
VARIABLE: NAMES ARE y1-y12;
          MISSING ARE ALL(-9999);
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f3 BY y7-y9;
       f4 BY y10-y12;
```

```
       f4 ON f3;
       f3 ON f1 f2;
MODEL INDIRECT:  f4 IND f1; f4 IND f2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```



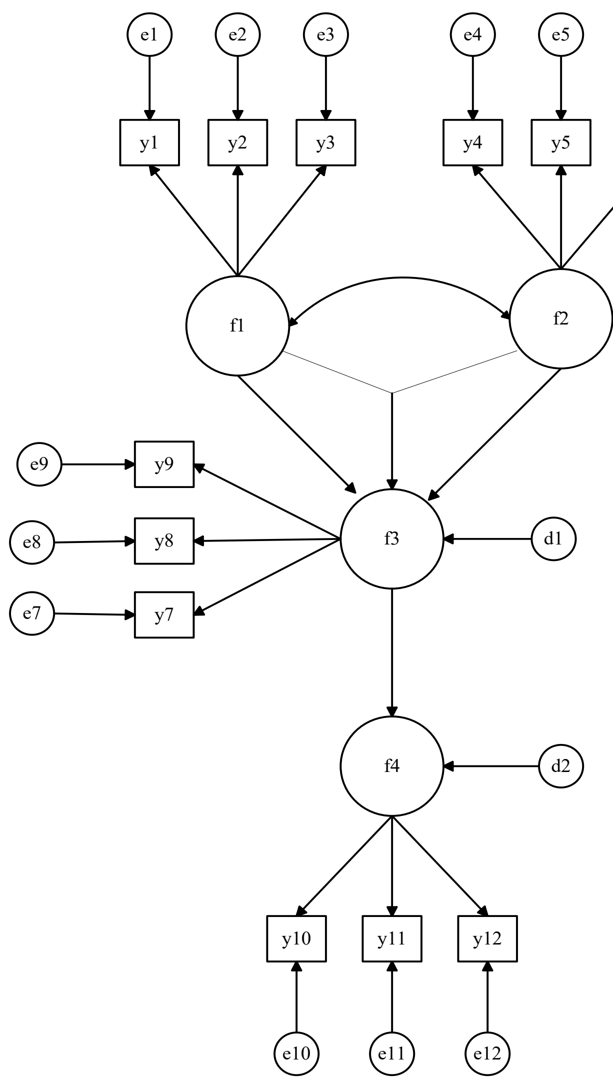
Latent Interaction Analysis Between Two Continuous Factors

The model being estimated is on the next page.

```
TITLE:  Example of latent variable interaction model
DATA: FILE IS c:\temp\mplus\ex5.13.dat;
VARIABLE: NAMES ARE y1-y12;
          MISSING ARE ALL(-9999);
ANALYSIS: TYPE = RANDOM;
          ALGORITHM = INTEGRATION;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
```

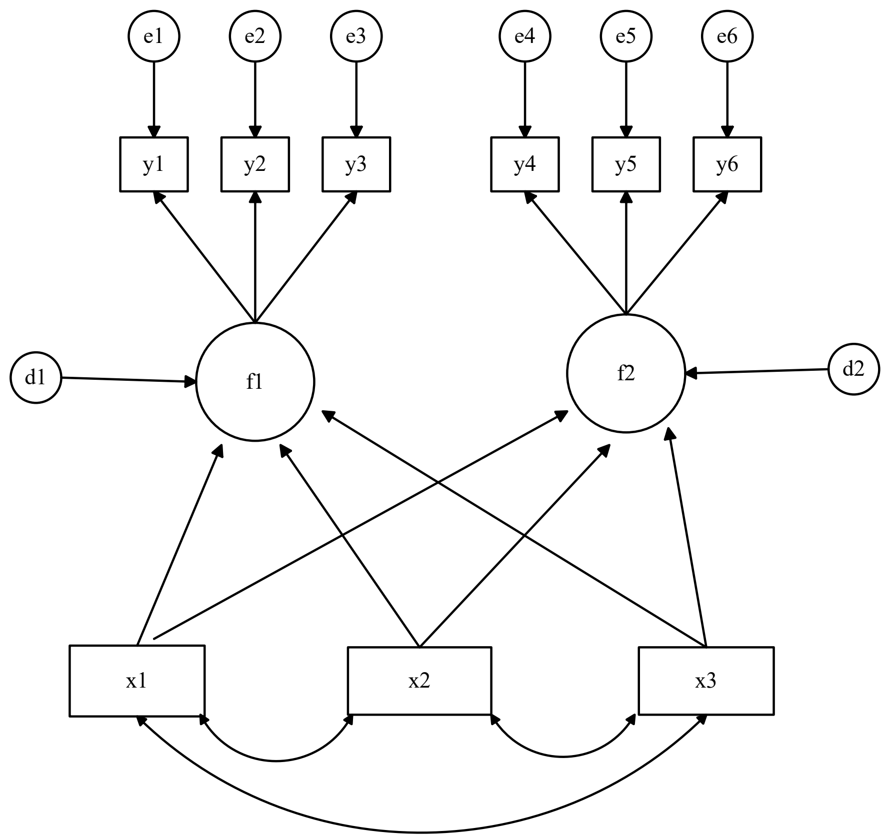
```
        f3 BY y7-y9;
        f4 BY y10-y12;
        f4 ON f3;
        f3 ON f1 f2;
        f1xf2 | f1 XWITH f2;
        f3 ON f1xf2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Strategy is based on Klein & Moosburger (2000). This routine requires the use of numerical integration, which produces robust standard errors. Uses the | line in which the left hand side specifies the name of the latent product term variable (no product terms are computed in the data set; everything is handled internally by M Plus – in this case the product term is called f1xf2) and the right hand side provides the instructions to create the product term. XWITH is short for "multiplied with." The | symbol defines random effect variables in a model. It is used mainly in growth curve models and in interaction models such as this one. Standardized and residual and tech4 output statements are not allowed in random effects models.



Multiple Groups SEM Model: Equal Form

Model we will work with is the following:




```
TITLE: Example of multiple group analysis with equal form
DATA: FILE IS c:\temp\mplus\ex5.14.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          GROUPING IS g (1=male 2=female);
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f1 f2 ON x1-x3;
MODEL female: f1 BY y2-y3;
              f2 BY y5-y6;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: M Plus specifies the general model that applies across all groups in the MODEL line. All parameters are free to vary except the factor loadings, which are set equal across all groups. A specific model is then specified for each group (e.g., MODEL female), telling M Plus what parameters to free up in that group relative to the main model that is constrained to be equal across groups. In the above example, the factor loadings for females are freed to vary from the male factor loadings. You only need to specify special labeled model lines for k-1 of the k groups. The above model has all free parameters across the groups.

Multiple Groups SEM Model: Completely Constrained Solution

```
TITLE: Example of multiple group analysis with everything forced to be equal
DATA: FILE IS c:\temp\mplus\ex5.14.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          GROUPING IS g (1=male 2=female);
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f1 ON x1  (1);
       f1 ON x2  (2);
       f1 ON x3  (3);
       f2 ON x1  (4);
       f2 ON x2  (5);
       f2 ON x3  (6);
       f1  (7);
       f2  (8);
       y1   (9);
       y2   (10);
       y3   (11);
       y4   (12);
       y5   (13);
       y6   (14);
       x1   (15);
       x2   (16);
       x3   (17);
       x1 WITH x2  (18);
       x1 WITH x3  (19);
       x2 WITH x3  (20);
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Only need to specify the one general model because it is going to be applied to all groups. Need to number each parameter in the model to force it to be equal across groups (with the exception of the factor loadings, which are held equal across groups by default). See the earlier discussion of equality constraints.

Multiple Groups SEM Model: Loading Invariance Only

```
TITLE: Example of multiple group analysis with loading invariance only
DATA: FILE IS c:\temp\mplus\ex5.14.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          GROUPING IS g (1=male 2=female);
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f1 f2 ON x1-x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Uses the defaults of M Plus to accomplish analysis with just the specification of the MODEL line. Loadings are set as invariant by default and all other parameters are free to vary across groups by default.

Multiple Groups SEM Model: Loading and Measurement Error Invariance

```
TITLE: Example of multiple group analysis with loading invariance only
DATA: FILE IS c:\temp\mplus\ex5.14.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          GROUPING IS g (1=male 2=female);
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f1 f2 ON x1-x3;
       y1   (1);
       y2   (2);
       y3   (3);
       y4   (4);
       y5   (5);
       y6   (6);
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Same as the previous program but now with the labeled lines added just before the OUTPUT line to force the error variances to be the same across all groups. Loadings are set as invariant by default. All other parameters are free to vary by default.

Multiple Groups SEM Model: Loading Invariance With One Constrained Path in a Two Group Solution

```
TITLE: Example of multiple group analysis with loading invariance and constrained
equal path with two group solution
DATA: FILE IS c:\temp\mplus\ex5.14.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          GROUPING IS g (1=male 2=female);
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f1 ON x1;
       f1 ON x2  (1);
       f1 ON x3;
       f2 ON x1 x2 x3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Uses the defaults of M Plus to accomplish analysis with just the specification of the MODEL line. Loadings are set as invariant by default and all other parameters are free to vary across groups by default. By giving a label to a path (the label "(1)" in the f1 on x2 line), the path is held equal across all groups.

Multiple Groups SEM Model: Loading Invariance With One Constrained Path in Two Groups in a Four Group Solution

```
TITLE: Example of multiple group analysis with loading invariance and
constrained equal path with two groups in four group solution
DATA: FILE IS c:\temp\mplus\ex5.14b.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          GROUPING IS g (1=cuban 2=mex 3=domin 4= puerto);
          MISSING ARE ALL(-9999);
ANALYSIS: Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       f1 ON x1;
       f1 ON x2  (1);
       f1 ON x3;
       f2 ON x1 x2 x3;
model domin:
f1 on x2;
model puerto:
f1 on x2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Uses the defaults of M Plus to accomplish analysis comparing the paths of two out of four groups by constraining them to be equal. Loadings are set as invariant by default and all other parameters are free to vary across groups by default. By giving a label to a path (the label "(1)" in the f1 on x2 line), the path is held equal across all groups. Then the two specialized model lines free up the path for the groups you are **not** interested in contrasting. The above compares f1 on x2 for Cubans and Mexicans by forcing the path for these two groups (and only these two groups) to be equal.

Multiple Group Solution to Compare Latent Means in One Group Versus Three Others

```
TITLE: Example of multiple group analysis with loading invariance and
estimate of mean differences comparing first group versus three others
DATA: FILE IS c:\temp\mplus\ex5.14b.dat;
VARIABLE: NAMES ARE y1-y6 x1-x3 g;
          USEVARIABLES ARE y1-y6 g;
          GROUPING IS g (1=cuban 2=mex 3=domin 4= puerto);
          MISSING ARE ALL(-9999);
ANALYSIS: Type=Mean; Esti=ML;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
       [f1@0]    (1);
       [f2@0]    (2);
Model mex:
[f1];
[f2];
Model domin:
[f1];
[f2];
Model puerto:
[f1];
[f2];
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: By default, M Plus sets factor loadings equal across groups and sets all the intercepts equal across groups. The MODEL line fixes all group factor means at 0. The more specific group model lines free the means up for the groups listed.

Model With Multiple Imputed Data Sets

```
TITLE:  Example of CFA with Multiple Imputed Data Sets
DATA: FILE IS c:\temp\mplus\impute.dat;
           TYPE = IMPUTATION;
VARIABLE: NAMES ARE y1-y6;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: First indicator is by default fixed at 1. The file called “impute.dat” has the files names (and paths) of the different imputed data sets. For example, it might have the following contents

```
c:\temp\data1.dat
c:\temp\data2.dat
c:\temp\data3.dat
c:\temp\data4.dat
c:\temp\data5.dat
```

The data are read in from each file and then standard errors and significance tests are computed using the appropriate pooling formulas.

Model With Weights

```
TITLE:  Example of CFA with Weights
DATA: FILE IS c:\temp\mplus\weights.dat;
VARIABLE: NAMES ARE y1-y6 psu gswght2;
          MISSING ARE ALL(-9999);
          WEIGHT is gswght2;
          CLUSTER is psu;
ANALYSIS: TYPE = Complex;
MODEL: f1 BY y1-y3;
       f2 BY y4-y6;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: First indicator is by default fixed at 1. Weight is individual weight and cluster is the cluster weight.

Linear Growth Curve Model

```
TITLE:  Example of linear growth curve
DATA: FILE IS c:\temp\mplus\ex6.1.dat;
VARIABLE: NAMES ARE y11-y14;
MODEL: i s | y11@0 y12@1 y13@2 y14@3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: The | symbol defines latent random effect variables in the model. The names to the left of the line give the labels for the latent variables. In this case, there are two latent variables, a slope (labeled s) and an intercept (labeled i). The right hand side of the line associates input variables with time points. In the above, y11 is at time 0, y12 is at time 1, y13 is at time 2 and y14 is at time 3. The @ sign fixes the value of the path from the slope latent variable to the observed measure. In the above example, the measures might be obtained at intervals of years. To express the results in units of months, would use:

```
MODEL: i s | y11@0 y12@12 y13@24 y14@36;
```

The intercept latent variable, by default, has a path fixed at 1 to every observed variable. The latent variables are correlated by default and their means and variances are estimated. The intercepts to the observed variables are fixed at zero.

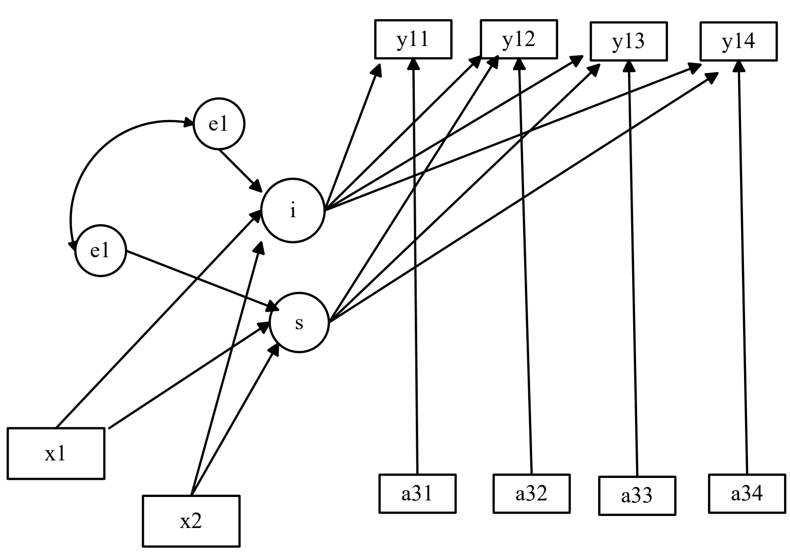
Quadratic Growth Curve Model

```
TITLE:  Example of quadratic growth curve
DATA: FILE IS c:\temp\mplus\ex6.9.dat;
VARIABLE: NAMES ARE y11-y14;
MODEL: i s q | y11@0 y12@1 y13@2 y14@3;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Everything is the same as the previous example with the linear growth curve, except an additional latent variable for the quadratic term has been added in the model line.

Linear Growth Curve Model with Time Invariant Predictors and Time Varying Covariates

```
TITLE:  Example of growth curve with covariates
DATA: FILE IS c:\temp\mplus\ex6.10.dat;
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
MODEL: i s | y11@0 y12@1 y13@2 y14@3;
       i s ON x1 x2;
       y11 ON a31;
       y12 ON a32;
       y13 ON a33;
       y14 ON a34;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```



Notes: Everything is the same as in the previous example with the linear growth curve. However, x1 and x2 are time invariant predictors and both i and s are regressed onto them. a31 to a34 are time varying covariates and y is regressed onto a at each time period e.g., at time o y11 is regressed onto a31; at time 1, y12 is regressed onto a32, and so on). Residuals of s and i are correlated by default.

Linear Growth Curve Model With Curve of One Variable Impacting the Curve of Another

```
TITLE: Example of one growth curve influencing another
DATA: FILE IS c:\temp\mplus\ex6.13.dat;
VARIABLE: NAMES ARE y11 y12 y13 y14 y21 y22 y23 y24;
MODEL: i1 s1 | y11@0 y12@1 y13@2 y14@3;
       i2 s2 | y21@0 y22@1 y23@2 y24@3;
       s1 ON i2 s2;
       s2 ON i1;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: See example on linear growth model for explanation of basic terms. i1 and s1 are one growth curve and i2 and s2 are the other. The slope of s1 is a linear function of the slope of s2 and the intercept i2. The slope s2 is a linear function of the intercept i1. Residuals of latent slopes and intercepts are assumed to be correlated, by default.

Piecewise Growth Model

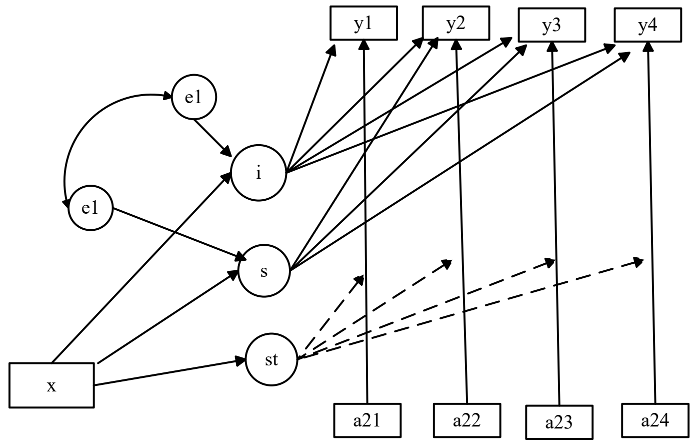
```
TITLE:  Example of piecewise growth curve
DATA: FILE IS c:\temp\mplus\ex6.11.dat;
VARIABLE: NAMES ARE y1-y5;
MODEL: i s1 | y1@0 y2@1 y3@2 y4@2 y5@2;
       i s2 | y1@0 y2@0 y3@0 y4@1 y5@2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: See example on linear growth model for explanation of basic terms. s1 is the first slope and covers the first three time periods. s2 is the second slope and covers the last three time periods.

Linear Growth Curve with Individually Varying Times of Observation

```
TITLE:  Example of growth curve with covariates and varying times
DATA: FILE IS c:\temp\mplus\ex6.12.dat;
VARIABLE: NAMES ARE y1-y4 x a21-a24 a11-a14;
          TSCORES = a11-a14;
ANALYSIS: TYPE = RANDOM;
MODEL: i s | y1-y4 AT a11-a14;
       st | y1 ON a21;
       st | y2 ON a22;
       st | y3 ON a23;
       st | y4 ON a24;
       i s st ON x;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: The model estimated is as follows:

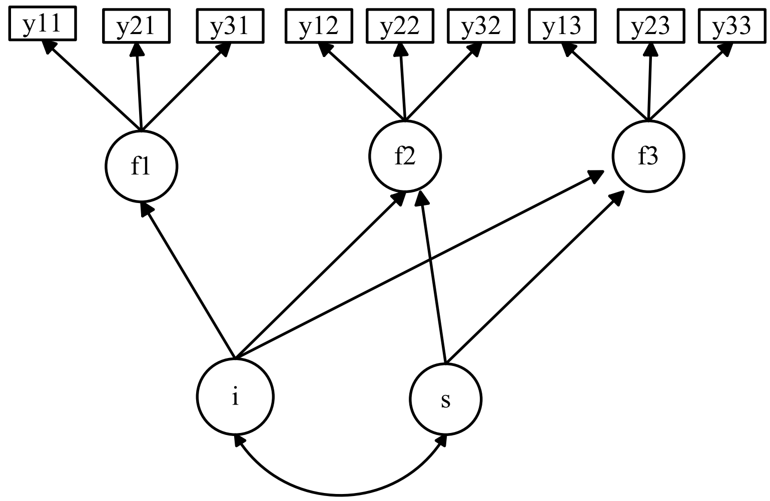


However, the times at which the measures are obtained vary a little bit around the different assessments. The TSCORES option identifies the variables in the data set that have the time varying information in them (e.g., month number at which assessment was made). This model needs the TYPE=RANDOM; option. On the model line, the AT command after y1-y4 links each outcome assessment to the time information. A latent variable called st is the basis for making the adjustment for varying times in the paths that the dashed arrows are drawn to.

Linear Growth Curve for Latent Variables

```
TITLE:  Example of piecewise growth curve
DATA: FILE IS c:\temp\mplus\ex6.14.dat;
VARIABLE: NAMES ARE y11 y21 y31 y12 y22 y32 y13
          y23 y33;
MODEL: f1 BY y11
             y21 (1)
             y31 (2);
      f2 BY  y12
             y22 (1)
             y32 (2);
      f3 BY  y13
             y23 (1)
             y33 (2);
      [y11 y12 y13] (3);
      [y21 y22 y23] (4);
      [y31 y32 y33] (5);
      i s | f1@0 f2@1 f3@2;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: The model being estimated is on the next page. The factor loadings are set invariant across time and the intercepts also are invariant across time.



Two Level HLM Random Coefficient Model with Random Slope and Random Intercept

```
TITLE:     Example of a two-level analysis
DATA:      FILE IS c:\temp\mplus\rc3.dat;
VARIABLE:  NAMES ARE attitude tcare pride resource size school;
           WITHIN =  tcare pride;
           BETWEEN = resource size;
           CLUSTER = school;
           CENTER tcare pride (GROUPMEAN) GRANDMEAN(resource size);
           CENTER resource size (GRANDMEAN) ;
ANALYSIS:  TYPE = TWOLEVEL RANDOM;
MODEL:
           %WITHIN%
           slope1 | attitude ON tcare;
           slope2 | attitude ON pride;
           %BETWEEN%
           attitude slope1 slope2 ON resource size;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Data file is organized so that scores for between level variables are repeated on each line of within level replicates. On VARIABLE line, for WITHIN line, list all within predictor variables (do not list outcome variables unless you want to fix their intercepts). For BETWEEN line, list all between level variables, but not the id variable for the clusters. CLUSTER line provides the id variable for the between level clusters. CENTERING lines applies group or grand centering, as desired. For MODEL line, %WITHIN% specifies the within level model and %BETWEEN% specifies the between level model. For %WITHIN%, the | symbol defines the latent random variable. The name to the left of the line gives the label for the random effect. The right hand side is the within level equation component the left hand term focuses upon. For %BETWEEN%, the random effects (including the name(s) of the within level outcome(s) so that its intercept is random) are regressed onto the between level predictors of interest. On OUTPUT line, StdYX, residual and Tech4 are not allowed.

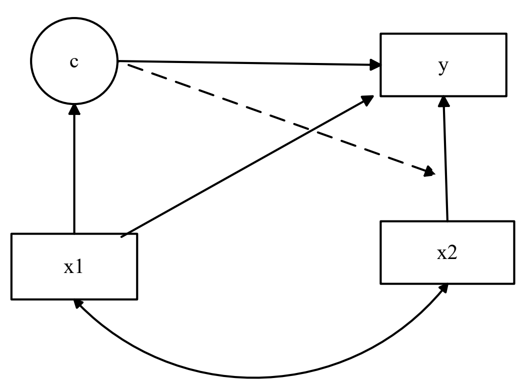
Two Level HLM Random Coefficient Model with Random Slope and Random Intercept but One Fixed Slope

```
TITLE:      Example of a two-level analysis
DATA:       FILE IS c:\temp\mplus\rc3.dat;
VARIABLE:   NAMES ARE attitude tcare pride resource size school;
            WITHIN =  tcare pride;
            BETWEEN = resource size;
            CLUSTER = school;
            CENTER tcare pride (GROUPMEAN) GRANDMEAN(resource size);
            CENTER resource size (GRANDMEAN) ;
ANALYSIS:   TYPE = TWOLEVEL RANDOM;
MODEL:
            %WITHIN%
            slope1 | attitude ON tcare;
            attitude on pride;
            %BETWEEN%
            attitude slope1 ON resource size;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: Same as previous program but pride is fixed rather than random.

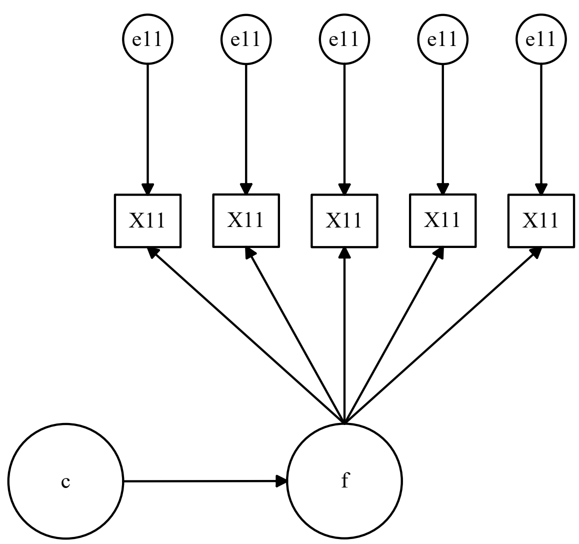
Mixture Regression Analysis

```
TITLE: Mixture regression analysis
DATA: FILE IS c:\temp\mplus\ex7.1.dat;
VARIABLE: NAMES ARE y x1 x2 c;
          USEVARIABLES ARE y x1 x2;
          CLASSES = c (2);
ANALYSIS: TYPE = MIXTURE;
MODEL:
      %OVERALL%
      y ON x1 x2;
      c#1 ON x1;
      %c#2%
      y ON x2;
      y;
SAVEDATA: FILE IS c:\temp\jim.txt;
          FORMAT IS Free;
          SAVE = CPROB;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```



Notes: This path diagram is read a little differently than traditional path diagrams. In this model, a categorical variable of two groups is thought to be influenced by x1, it is thought to influence y in the sense that the intercept for each group of c for the regression of y onto x1 and x2 is thought to differ as a function c, and it moderates the impact of x2 on y in a multiple group solution sense. M Plus classifies individuals into the two groups for this variable (in the spirit of cluster analysis) so as to maximize these conditions. The CLASSES = c(2) line provides the label for the latent categorical variable (in the case, c) as well as the number of levels of that variable or the number of clusters to create (in this case 2). The %OVERALL% statement describes the overall model that applies to all the groups or classes (i.e., all groups defined by c). It says to regress y onto x1 and x2. All parameters in this model are set to be equal across the groups of c. The second ON statement tells M Plus to conduct a multinomial logistic regression regressing c onto x1. For how this syntax works, see the example earlier on multinomial logistic regression. The %c#2% command tells M Plus about how the model should be treated for group 2 of c. It tells M Plus that the path of x2 to y will differ for group 2 relative to the other groups and that the intercept of y will differ for group 2 relative to the other groups. In other words, it tells M Plus what equality constraints to relax. It operates much like the multi-group solution syntax. The Savedata line saves the raw data that is analyzed as well as the probability that an individual is in each class as well as the class number that has the highest probability for that individual. Missing data is saved as a *. The order in which the data are saved appears at the end of the output. Note that could do the same model but where the x and y are replaced by latent variables with multiple indicators. An example is in the M Plus manual, Example 7.20.

Mixture CFA




```
TITLE: CFA mixture modeling
DATA: FILE IS c:\temp\mplus\ex7.17.dat;
VARIABLE: NAMES ARE y1-y5 g;
      USEVARIABLES ARE y1-y5;
      CLASSES = c(2);
ANALYSIS: TYPE = MIXTURE;
MODEL: %OVERALL%
       f BY y1-y5;
       %c#1%
```

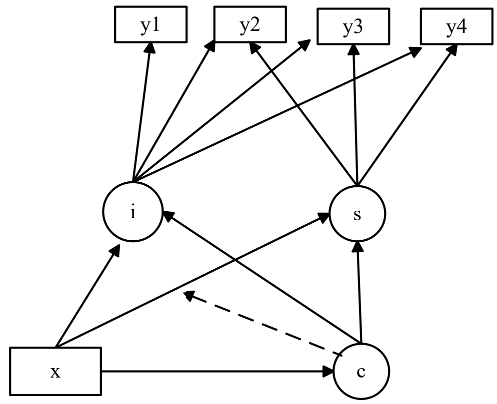
```
       [f];
SAVEDATA: FILE IS c:\temp\jim.txt;
          FORMAT IS Free;
          SAVE = CPROB;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

Notes: In this model, a categorical variable of two groups is thought to differ in the factor mean. M Plus classifies individuals into the two groups for this variable (in the spirit of cluster analysis) so as to maximize the mean differences. The CLASSES = c(2) line provides the label for the latent categorical variable (in the case, c) as well as the number of levels of that variable or the number of clusters to create (in this case 2). The %OVERALL% statement describes the overall model that applies to all the groups or classes (i.e., all groups defined by c). It specifies the CFA model in this case. The %c12% command tells M Plus about how the model should be treated for group 1 of c. It tells M Plus that the mean on f can be different for the first group relative to the other groups. In other words, it tells M Plus what equality constraints to relax. It operates much like the multi-group solution syntax. The Savedata line saves the raw data that is analyzed as well as the probability that an individual is in each class as well as the class number that has the highest probability for that individual. Missing data is saved as a *. The order in which the data are saved appears at the end of the output.

Mixture Linear Growth Curves

```
TITLE: Growth curve mixture modeling
DATA: FILE IS c:\temp\mplus\ex8.1.dat;
VARIABLE: NAMES ARE y1-y4 x c;
          USEVARIABLES = y1-y4 x;
          CLASSES = c (2);
ANALYSIS: TYPE = MIXTURE;
          STARTS = 20 2;
MODEL:
      %OVERALL%
      i s | y1@0 y2@1 y3@2 y4@3;
      i s ON x;
      c#1 ON x;
      %c#1%
      s on x;
SAVEDATA: FILE IS c:\temp\jim.txt;
          FORMAT IS Free;
          SAVE = CPROB;
OUTPUT: Samp StdYX Mod(All 4) Residual Cinterval Tech4;
```

The syntax is straightforward based on previous examples. The model tested is:


y1
y2
y3
y4
i
s
x
c