VERSION 5.1
Mplus LANGUAGE ADDENDUM

In this chapter, changes to existing options and new options introduced in Version 5.1 are discussed.

THE VARIABLE COMMAND

AUXILIARY

Auxiliary variables are variables that are not part of the analysis model. The AUXILIARY option has two new settings and a new way to specify existing settings. The first new setting r is used with TYPE=MIXTURE to explore which covariates are important predictors of latent classes. This is done using pseudo-class draws, that is, posterior-probability based multinomial logistic regression of a categorical latent variable on a set of covariates. The letter r in parentheses is placed behind the variables in the AUXILIARY statement that will be used as covariates in the multinomial logistic regression. Following is an example of how to specify the r setting:

AUXILIARY = gender race (r) educ ses (r) x1-x5 (r);

where race, ses, x1, x2, x3, x4, and x5 will be used as covariates in the multinomial logistic regression. The r setting and the e setting (see the Version 5 Mplus User's Guide) cannot be used in the same analysis.

The second new setting m is used with TYPE=GENERAL with continuous dependent variables to specify which variables will be used as missing data correlates in addition to the analysis variables (Collins, Schafer, & Kam, 2001; Graham, 2003). The m setting is not available with MODINDICES, BOOTSTRAP, and models with a set of exploratory factor analysis (EFA) factors in the MODEL command.

The AUXILIARY option has an alternative specification for the e, r, and m settings that is convenient when there are several variables that cannot be specified using the list function. These are AUXILIARY = (e), AUXILIARY = (r), and AUXILIARY = (m). When e, r, or m in parentheses follows the equal sign, it means that e, r, or m applies to all of the variables that follow. For example, the following AUXILIARY statement specifies that the variables x1, x3, x5, x7, and x9 will be used as missing data correlates in addition to the analysis variables:

AUXILIARY = (m) x1 x3 x5 x7 x9;

COUNT

The COUNT option is used to specify which dependent variables are treated as count variables in the model and its estimation and the type of model to be estimated. Four new models have been added. The following models can be estimated for count variables: Poisson, zero-inflated Poisson, negative binomial, zero-inflated negative binomial, zero-truncated negative binomial, and negative binomial hurdle (Long, 1997; Hilbe, 2007).

The COUNT option can be specified in two ways for a Poisson model:

COUNT = u1 u2 u3 u4;

or

COUNT = u1 (p) u2 (p) u3 (p) u4 (p);

or using the list function:

COUNT = u1-u4 (p);

The COUNT option can be specified in two ways for a zero-inflated Poisson model:

COUNT = u1-u4 (i);

or

COUNT = u1-u4 (pi);

where u1, u2, u3, and u4 are count dependent variables in the analysis. The letter i or pi in parentheses following the variable name indicates that a zero-inflated Poisson model will be estimated.

With a zero-inflated Poisson model, two variables are considered, a count variable and an inflation variable. The count variable takes on values for individuals who are able to assume values of zero and above following the Poisson model. The inflation variable is a binary latent variable with one denoting that an individual is unable to assume any value except zero. The inflation variable is referred to by adding to the name of the count variable the number sign (#) followed by the number 1.

Following is the specification of the COUNT option for a negative binomial model:

COUNT = u1 (nb) u2 (nb) u3 (nb) u4 (nb);

or using the list function:

COUNT = u1-u4 (nb);

Following is the specification of the COUNT option for a zero-inflated negative binomial model:

COUNT = u1- u4 (nbi);

With a zero-inflated negative binomial model, two variables are considered, a count variable and an inflation variable. The count variable takes on values for individuals who are able to assume values of zero and above following the negative binomial model. The inflation variable is a binary latent variable with one denoting that an individual is unable to assume any value except zero. The inflation variable is referred to by adding to the name of the count variable the number sign (#) followed by the number 1.

Following is the specification of the COUNT option for a zero-truncated negative binomial model:

COUNT = u1-u4 (nbt);

Count variables for the zero-truncated negative binomial model must have values greater than zero.

Following is the specification of the COUNT option for a negative binomial hurdle model:

COUNT = u1-u4 (nbh);

With a negative binomial hurdle model, two variables are considered, a count variable and a hurdle variable. The count variable takes on values for individuals who are able to assume values of one and above following the truncated negative binomial model. The hurdle variable is a binary latent variable with one denoting that an individual is unable to assume any value except zero. The hurdle variable is referred to by adding to the name of the count variable the number sign (#) followed by the number 1.

THE DEFINE COMMAND

Three new functions have been added to the DEFINE command. The first function creates a variable that is the sum of a set of variables. The second function creates a variable that is the average of a set of variables. The third function creates a variable that is the average for each cluster of an individual-level variable.

SUM

The SUM function is used to create a variable that is the sum of a set of variables. It is specified as follows:

sum = SUM (y1 y3 y5);

where the variable sum is the sum of variables y1, y3, and y5. Any observation that has a missing value on one or more of the variables being summed is assigned a missing value on the sum variable.

The list function can be used with the SUM function as follows:

ysum = SUM (y1-y10);

where the variable ysum is the sum of variables y1 through y10.

MEAN

The MEAN function is used to create a variable that is the average of a set of variables. It is specified as follows:

mean = MEAN (y1 y3 y5);

where the variable mean is the average of variables y1, y3, and y5. Averages are based on the set of variables with non-missing values. Any observation that has a missing value on all of the variables being averaged is assigned a missing value on the mean variable.

The list function can be used with the MEAN function as follows:

ymean = MEAN (y1-y10);

where the variable ymean is the average of variables y1 through y10.

CLUSTER_MEAN

The CLUSTER_MEAN function is used with the CLUSTER option to create a variable that is the average of the values of an individual-level variable for each cluster. It is specified as follows:

clusmean = CLUSTER_MEAN (x);

where the variable clusmean is the average of the values of x for each cluster. Averages are based on the set of non-missing values for the observations in each cluster. Any cluster for which all observations have missing values is assigned a missing value on the cluster mean variable. A variable created using the CLUSTER_MEAN function cannot be used in subsequent DEFINE statements in the same analysis.

TYPE=TWOLEVEL EFA

The UW and UB settings for TYPE=TWOLEVEL EFA have been expanded to include the settings of UW* and UB*. When UW and UB are specified, the unrestricted models are not estimated but instead the model parameters are fixed at the sample statistic values. When UW* and UB* are specified, the unrestricted models are estimated.

ROWSTANDARDIZATION

The ROWSTANDARDIZATION option is used with exploratory factor analysis (EFA) and when a set of EFA factors is part of the MODEL command to request row standardization of the factor loading matrix before rotation. The ROWSTANDARDIZATION option has three settings: CORRELATION, KAISER, and COVARIANCE. The CORRELATION setting rotates a factor loading matrix derived from a correlation matrix with no row standardization. The KAISER setting rotates a factor loading matrix derived from a correlation matrix with standardization of the factor loadings in each row using the square root of the sum of the squares of the factor loadings in each row (Browne, 2001). The COVARIANCE setting rotates a factor loading matrix derived from a covariance matrix with no row standardization. The default is CORRELATION. The COVARIANCE setting is not allowed for TYPE=EFA. If factor loading equalities are specified in a model for EFA factors, the CORRELATION and KAISER settings are not allowed.

ROTATION

The default rotation for TYPE=EFA has changed from the direct QUARTIMIN rotation to the GEOMIN oblique rotation. The GEOMIN epsilon (Browne, 2001) default setting varies as a function of the number of factors. With two factors, it is .0001. With three factors, it is .001. With four or more factors, it is .01. The default can be overridden using the GEOMIN option.

The TARGET setting of the ROTATION option has been added (Browne, 2001). It is used with models that have a set of EFA factors in the MODEL command. This setting allows the specification of target factor loading values to guide the rotation of the factor loading matrix. Typically these values are zero. The default for the TARGET rotation is oblique. Following is an example of how to specify an orthogonal TARGET rotation:

ROTATION = TARGET (ORTHOGONAL);

For TARGET rotation, a minimum number of target values must be given for purposes of model identification. For the oblique TARGET rotation, the minimum is m(m-1) where the m is the number of factors. For the orthogonal TARGET rotation, the minimum is m(m-1)/2. The target values are given in the MODEL command using the tilde (~) symbol.

THE MODEL COMMAND

The measurement part of the MODEL command has been expanded to include not only traditional confirmatory factor analysis (CFA) factors but also sets of exploratory factor analysis (EFA) factors (Asparouhov & Muthén, 2008). One of the differences between CFA and EFA factors is that CFA factors are not rotated. For a set of EFA factors, the factor loading matrix is rotated as in conventional EFA using the rotations available through the ROTATION option of the ANALYSIS command including a new targeted rotation. A set of EFA factors must have the same factor indicators. A set of EFA factors can be regressed on the same set of covariates. An observed or latent variable can be regressed on a set of EFA factors. EFA factors are allowed with TYPE=GENERAL with observed dependent variables that are continuous, censored, binary, ordered categorical (ordinal), and combinations of these variable types. EFA factors are not allowed when summary data are analyzed and when the MLM and MLMV estimators are used.

BY

The BY option has three new features that are used with sets of EFA factors in the MODEL command. One feature is used to define sets of EFA factors. The second feature is a special way of specifying factor loading matrix equality for sets of EFA factors. The third feature is used in conjunction with the TARGET setting of the ROTATION option of the ANALYSIS command to provide target factor loading values to guide the rotation of the factor loading matrix for sets of EFA factors.

DEFINING EFA FACTORS

Following is an example of how to define a set of EFA factors using the BY option:

```
f1-f2 BY y1-y5 (*1);
```

where the asterisk (*) followed by a label specifies that factors f1 and f2 are a set of EFA factors with factor indicators y1 through y5.

Following is an alternative specification:

```
f1 BY y1-y5 (*1);
f2 BY y1-y5 (*1);
```

where the label 1 specifies that factors f1 and f2 are part of the same set of EFA factors. Rotation is carried out on the five by two factor loading matrix. Labels for EFA factors must follow an asterisk (*). EFA factors must have the same factor indicators.

More than one set of EFA factors may appear in the MODEL command. For example,

```
f1-f2 BY y1-y5 (*1);
f3-f4 BY y6-y10 (*2);
```

specifies that factors f1 and f2 are one set of EFA factors with the label 1 and factors f3 and f4 are another set of EFA factors with the label 2. The two sets of EFA factors are rotated separately.

Factors in a set of EFA factors can be regressed on covariates but the set of covariates must be the same, for example,

f1-f2 ON x1-x3;

or

f1 ON x1-x3;
f2 ON x1-x3;

A set of EFA factors can also be used as covariates in a regression, for example,

y ON f1-f2;

EQUALITIES WITH EFA FACTORS

The BY option has a special convention for specifying equalities of the factor loading matrices for more than one set of EFA factors. The equality label is placed after the label that defines the set of EFA factors and applies to the entire factor loading matrix not to a single parameter. Following is an example of how to specify that the factor loading matrices for the set of EFA factors f1 and f2 and the set of EFA factors f3 and f4 are held equal:

f1-f2 BY y1-y5 (*1 1);
f3-f4 BY y6-y10 (*2 1);

The number 1 following the labels 1 and 2 that define the EFA factors specifies that the factor loadings matrices for the two sets of EFA factors are held equal.

ROTATIONS WITH EFA FACTORS

The BY option has a special feature that is used with the TARGET setting of the ROTATION option of the ANALYSIS command to specify target factor loading values for a set of EFA factors (Browne, 2001). The target factor loading values are used to guide the rotation of the factor loading matrix. Typically these values are zero. For the TARGET rotation, a minimum number of target values must be given for purposes of model identification. For the default oblique TARGET rotation, the minimum is m(m-1) where the m is the number of factors. For the orthogonal TARGET rotation, the minimum is m(m-1)/2. The target values are given in the MODEL command using the tilde (~) symbol. The target values are specified in a BY statement using the tilde (~) symbol as follows:

f1 BY y1-y5 y1~0 (*1);
f2 BY y1-y5 y5~0 (*1);

where the target factor loading values for the factor indicator y1 for factor f1 and y5 for factor f2 are zero.

THE MONTECARLO COMMAND

GENERATE

The GENERATE option has several new settings related to generating count variables. Variables can be generated for count data for the following models: Poisson, zero-inflated Poisson, negative binomial, zero-inflated negative binomial, zero-truncated negative binomial, and negative binomial hurdle (Long, 1997; Hilbe, 2007).

The GENERATE option can be specified in two ways for generating count variables for a Poisson model:

GENERATE = u1 (c) u2 (c) u3 (c) u4 (c);

or

GENERATE = u1 (p) u2 (p) u3 (p) u4 (p);

or using the list function:

GENERATE = u1-u4 (p);

The GENERATE option can be specified in two ways for generating count variables for a zero-inflated Poisson model:

GENERATE = u1-u4 (ci);

or

GENERATE = u1-u4 (pi);

Following is the specification of the GENERATE option for generating count variables for a negative binomial model:

GENERATE = u1-u4 (nb);

Following is the specification of the GENERATE option for generating count variables for a zero-inflated negative binomial model:

GENERATE = u1-u4 (nbi);

Following is the specification of the GENERATE option for generating count variables for a zero-truncated negative binomial model:

GENERATE = u1-u4 (nbt);

Following is the specification of the GENERATE option for generating count variables for a negative binomial hurdle model:

GENERATE = u1-u4 (nbh);

COUNT

The COUNT option is used to specify which dependent variables are treated as count variables in the model and its estimation and the type of model to be estimated. The following models can be estimated for count variables: Poisson, zero-inflated Poisson, negative binomial, zero-inflated negative binomial, zero-truncated negative binomial, and negative binomial hurdle (Long, 1997; Hilbe, 2007).

The COUNT option can be specified in two ways for a Poisson model:

COUNT = u1 u2 u3 u4;

or

COUNT = u1 (p) u2 (p) u3 (p) u4 (p);

or using the list function:

COUNT = u1-u4 (p);

The COUNT option can be specified in two ways for a zero-inflated Poisson model:

COUNT = u1-u4 (i);

or

COUNT = u1-u4 (pi);

where u1, u2, u3, and u4 are count dependent variables in the analysis. The letter i or pi in parentheses following the variable name indicates that a zero-inflated Poisson model will be estimated.

With a zero-inflated Poisson model, two variables are considered, a count variable and an inflation variable. The count variable takes on values for individuals who are able to assume values of zero and above following the Poisson model. The inflation variable is a binary latent variable with one denoting that an individual is unable to assume any value except zero. The inflation variable is referred to by adding to the name of the count variable the number sign (#) followed by the number 1.

Following is the specification of the COUNT option for a negative binomial model:

COUNT = u1 (nb) u2 (nb) u3 (nb) u4 (nb);

or using the list function:

COUNT = u1-u4 (nb);

Following is the specification of the COUNT option for a zero-inflated negative binomial model:

COUNT = u1- u4 (nbi);

With a zero-inflated negative binomial model, two variables are considered, a count variable and an inflation variable. The count variable takes on values for individuals who are able to assume values of zero and above following the negative binomial model. The inflation variable is a binary latent variable with one denoting that an individual is unable to assume any value except zero. The inflation variable is referred to by adding to the name of the count variable the number sign (#) followed by the number 1.

Following is the specification of the COUNT option for a zero-truncated negative binomial model:

COUNT = u1-u4 (nbt);

Count variables for the zero-truncated negative binomial model must have values greater than zero.

Following is the specification of the COUNT option for a negative binomial hurdle model:

COUNT = u1-u4 (nbh);

With a negative binomial hurdle model, two variables are considered, a count variable and a hurdle variable. The count variable takes on values for individuals who are able to assume values of one and above following the truncated negative binomial model. The hurdle variable is a binary latent variable with one denoting that an individual is unable to assume any value except zero. The hurdle variable is referred to by adding to the name of the count variable the number sign (#) followed by the number 1.

REFERENCE

Asparouhov, T. & Muthén, B. (2008). Exploratory structural equation modeling. Technical report. www.statmodel.com.

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36, 111-150.

Collins, L.M, Schafer, J.L., & Kam, C-H (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods, 6, 330-351.

Graham, J.W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. Structural Equation Modeling: A Multidisciplinary Journal, 10, 80-100.

Hilbe, J.M. (2007). Negative binomial regression. New York: Cambridge University Press.

Long, J.S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage Publications, Inc.