VERSION 8.5
MPLUS LANGUAGE ADDENDUM

In this addendum, changes introduced in Version 8.5 are described. They include corrections to minor problems that have been found since the release of Version 8.4 in November 2019 as well as the following new features:

- Automatic reordering of latent classes for mixture models
- Bayes estimation for count, nominal, and binary variables using a logit link for both real data analysis and Monte Carlo simulation studies for single-level, two-level, and mixture models with a single categorical latent variable
- Bayes estimation of mixture models with categorical latent variable regression on observed variables
- Data imputation for count, nominal, and binary variables with a logit link
- Multilevel data saving

AUTOMATIC REORDERING OF LATENT CLASSES FOR MIXTURE MODELS

The SVALUES option of the OUTPUT command has been extended to allow the automatic reordering of latent classes for mixture models in addition to creating input statements for the MODEL command with estimated values as starting values. This is available for models with one or more categorical latent variables as long as each class has the same set of parameters, for example, the measurement parameters of LCA and LTA. The reordering does not apply to the KNOWNCLASS variable.

In this extension, the SVALUES option is used in conjunction with the OPTSEED option of the ANALYSIS command. The OPTSEED option is used to specify the optseed that gives the best loglikelihood of the original analysis for which the classes are to be reordered.

For an estimated model with one categorical latent variable and three classes, the SVALUES option is specified as follows for a three class model where the analysis reorders the classes from 1 2 3 to 3 2 1:

SVALUES (3 2 1);

For a model with three categorical variables with four classes each, the SVALUES option is specified as follows for the three categorical latent variables:

SVALUES (4 1 2 3 | 3 4 1 2 | 2 1 4 3);

The sets of numbers in parentheses are given in the order of the categorical latent variables as specified in the CLASSES option of the VARIABLE command.

BAYES ESTIMATION FOR COUNT, NOMINAL, AND BINARY VARIABLES USING A LOGIT LINK

Bayes estimation is available for count, nominal, and binary variables using a logit link for both real data analysis and Monte Carlo simulation studies for single-level, two-level, and mixture models with a single categorical latent variable.

REAL DATA ANALYSIS 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 and the Bayes estimator: Poisson and negative binomial. The negative binomial model can use either the NB-2 variance representation (Hilbe, 2011, p. 63) or the Polya-Gamma representation (Polson et al., 2013). Only the Polya-Gamma representation can be used for models with equality constraints. Count variables may not have negative or non-integer values. See the COUNT option in the Mplus User's Guide for the analysis of count variables using maximum likelihood estimation.

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);

Following is the specification of the COUNT option for a negative binomial model using the NB-2 variance representation:

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 negative binomial model using the Polya-Gamma representation:

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

or using the list function:

COUNT = u1-u4 (nbpg);

With a negative binomial model, a dispersion parameter is estimated. The dispersion parameter is referred to by using the name of the count variable. If the dispersion parameter is estimated at zero, the model is a Poisson model.

REAL DATA ANALYSIS NOMINAL

The NOMINAL option is used to specify which dependent variables are treated as unordered categorical (nominal) variables in the model and its estimation. Unordered categorical dependent variables cannot have more than 10 categories. The number of categories is determined from the data. The NOMINAL option is specified as follows:

NOMINAL ARE u1 u2 u3 u4;

where u1, u2, u3, u4 are unordered categorical dependent variables in the analysis.

For nominal dependent variables, all categories but the last category can be referred to. The last category is the reference category. The categories are referred to in the MODEL command by adding to the variable name the number sign (#) followed by a number. The three categories of a four-category nominal variable are referred to as u1#1, u1#2, and u1#3.

The estimation of the model for unordered categorical dependent variables uses zero to denote the lowest category, one to denote the second lowest category, two to denote the third lowest category,

etc. If the variables are not coded this way in the data, they are automatically recoded. When data are saved for subsequent analyses, the recoded categories are saved. For details about how the recoding is done, see the NOMINAL option of the VARIABLE command in the Mplus User's Guide.

REAL DATA ANALYSIS BINARY CATEGORICAL LOGIT

The CATEGORICAL option is used to specify which dependent variables are treated as binary or ordered categorical (ordinal) in the model and its estimation and the type of model to be estimated. For Bayes and a binary dependent variable, the default is a probit model. The LINK option of the ANALYSIS command is used to request a logistic model by specifying:

LINK = LOGIT;

For Bayes with a combination of binary and ordinal dependent variables, the default is a probit model. The logistic model is not available for ordinal dependent variables. To obtain logistic regression coefficients for the binary variables and probit regression coefficients for the ordinal variables, specify the following in the ANALYSIS command:

LINK = LOGIT PROBIT;

For categorical dependent variables, there are as many thresholds as there are categories minus one. The thresholds are referred to in the MODEL command by adding to the variable name the dollar sign ($) followed by a number. The threshold for a binary variable u1 is referred to as u1$1. The two thresholds for a three-category variable u2 are referred to as u2$1 and u2$2. Ordered categorical dependent variables cannot have more than 10 categories. The number of categories is determined from the data.

The estimation of the model for binary or ordered categorical dependent variables uses zero to denote the lowest category, one to denote the second lowest category, two to denote the third lowest category, etc. If the variables are not coded this way in the data, they are automatically recoded. When data are saved for subsequent analyses, the recoded categories are saved. For details about how the recoding is done, see the CATEGORICAL option of the VARIABLE command in the Mplus User's Guide.

MONTE CARLO SIMULATION STUDIES

For Bayes, count variables can be generated for the following models: Poisson, negative binomial with an NB-2 variance representation, and negative binomial with a Polya-Gamma representation. The letter c or p in parentheses following the variable name indicates that the variable is generated using a Poisson model. The letters nb in parentheses following the variable name indicate that the variable is generated using a negative binomial model with an NB-2 representation. The letters nbpg in parentheses following the variable name indicate that the variable is generated using a negative binomial model with a Polya-Gamma representation.

For unordered categorical (nominal) variables, the letter n followed by the number of intercepts is put in parentheses following the variable name. The number of intercepts is equal to the number of categories minus one because the intercepts are fixed to zero in the last category which is the reference category.

For Bayes, binary and ordered categorical (ordinal) variables can be generated for a probit model. Binary variables can also be generated for a logistic model. For binary and ordered categorical (ordinal) variables, the number of thresholds followed by the letter p in parentheses following the variable name specifies that a probit model will be used. If no letter is specified, a probit model is used. The number of thresholds is equal to the number of categories minus one. For binary variables, the number of thresholds followed by the letter l in parentheses following the variable name specifies that a logistic model will be used.

The GENERATE option is specified as follows:

GENERATE = u1-u2 (p) u3 (nb) u4 (nbpg) u5 (n 2) u6 (1 l);

where the information in parentheses following the variable name or list of variable names defines the scale of the dependent variables for data generation. In this example, the variables u1 and u2 are count variables generated using a Poisson model. The variable u3 is a count variable generated using a negative binomial model with an NB-2 representation. The variable u4 is a count variable generated using a negative binomial model with a Polya-Gamma representation. Count variables for the Poisson model and the negative binomial model with an NB-2 representation can be generated separately or together. Count variables for the negative binomial model with a Polya-Gamma representation cannot be

generated together with other count models. The variable u5 is a three-category unordered categorical (nominal) variable with two intercepts. The variable u6 is a binary variable with one threshold generated using a logistic model. For further information about data generation, see the GENERATE option of the MONTECARLO command in the Mplus User's Guide.

DATA IMPUTATION FOR COUNT, NOMINAL, AND BINARY VARIABLES WITH A LOGIT LINK

The DATA IMPUTATION command has been extended to include the imputation of count, nominal, and binary variables with a logit link. This is allowed only for H0 imputation (MODEL command) using ESTIMATOR=BAYES. To specify that the variable being imputed is a count variable, the letters ct in parentheses are included after the variable name on the IMPUTE list. To specify that the variable being imputed is a nominal variable, the letter n in parentheses is included after the variable name on the IMPUTE list. To specify that the variable being imputed is binary with a logit link, the letter c in parentheses is included after the variable name on the IMPUTE list and LINK=LOGIT must be included in the ANALYSIS command.

IMPUTE = u1 (ct) u2 (ct) u3 (n) u4 (n) u5 (c);

specifies that u1 and u2 will be imputed as count variables, u3 and u4 will be imputed as nominal variables, and u5 will be imputed as a categorical variable with a logit link when LINK=LOGIT is used.

MULTILEVEL DATA SAVING AND PLOTS

The SAVEDATA command can be used in conjunction with TYPE=TWOLEVEL, TYPE=THREELEVEL, and TYPE=CROSSCLASSIFIED to save between-level factors and random latent variables, between-level observed variables specified on the BETWEEN list, and between-level auxiliary variables specified on the AUXILIARY list in separate data sets that can be analyzed using TYPE=IMPUTATION. For TYPE=TWOLEVEL and TYPE=TRHEELEVEL, variables from the highest level are saved. For TYPE=CROSSCLASSIFIED, variables from level 2b are saved. To obtain these data sets, specify the following:

SAVEDATA:

```
FILE = save*.dat;
SAVE = FSCORES (200 10);
```

where the asterisk (*) after the file name indicates that individual data sets will be saved. The names of the saved data sets will be the file name followed by the number of the data set. The SAVE option specifies that factors scores for these variables will be saved. The 200 in parentheses specifies that 200 imputed data sets will be saved. The 10 in parentheses specifies thinning of 10.

When the data sets are used in TYPE=IMPUTATION, the data file specified in the FILE option of the DATA command is named savelist.dat. See Example 13.13 in the Mplus User's Guide for an example of how to use TYPE=IMPUTATION.

REFERENCES

Hilbe, J.M. (2011). Negative binomial regression. Second edition. New York: Cambridge University Press.

Polson, N.G., Scott, J.G., & Windle, J (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. Journal of the American Statistical Association, 108(504), 1339-1349.