

Why is Mplus excluding cases with missing values when the model does not specify listwise deletion?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *Why is Mplus excluding cases with missing values when the model does not specify listwise deletion?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165082>

Mplus, a statistical modeling software, excludes cases with missing values even when the model does not explicitly specify listwise deletion. This is because Mplus follows the default approach of pairwise deletion, where missing values in individual variables are ignored and the remaining values are used for analysis. This method allows for a larger sample size and can provide more accurate results. However, it is important for researchers to carefully consider the implications of using pairwise deletion and whether it is appropriate for their specific research question and data.

Why is Mplus excluding cases with missing values when the model does not specify listwise deletion? | Mplus FAQ

Mplus can be used to estimate a model in which some of the variables have missing values using full information maximum likelihood (FIML). Starting in version 5 this is done by default, in earlier versions this type of estimation could be requested using `type = missing;`. However, for some models, Mplus drops cases with missing values on any of the predictors. Below is an example of this using count models, but we have encountered this behavior with other types of models (e.g., models with categorical outcomes). We hope that this page will help you recognize this when it happens, and help you understand what Mplus is doing. A method of specifying the model so that cases with missing values on the predictor are included is presented at the bottom of the page.

Below are the descriptive statistics for a small dataset (the output is from another package). The variable d1 is a binary variable we will use as a predictor, x1 and x2 are continuous predictors, and count is the outcome (true to its name it is a count variable). We know from working with the data that there are 150 cases with complete data on all of the variables. Note that x1, x2, and count all have missing values. The Mplus version of the dataset can be downloaded [here](#).

Variable | Obs Mean Std. Dev. Min Max

```
-----+-----  
d1 | 200 .545 .4992205 0 1  
x1 | 180 52.71111 9.388861 33 75  
x2 | 188 52.35106 10.83193 26 71  
count | 177 1.615819 1.796265 0 7
```

Below is the Mplus input file to run a model with x1, x2, and d1 predicting count.

Data:

File is d:datafiml_count.dat;

Variable:

Names are d1 x1 x2 count;

Missing are all (-9999);

count is count;

Model:

count on x1 x2 d1;

When we run the model we receive the following output. A warning message tell us that Mplus has excluded 31 cases because they have missing values on the x-variables (predictors). Mplus has also excluded 19 cases with missing values on "all variables except the x-variables," that is, cases missing on the outcome. Below the warning messages we see that the number of the observations used to estimate the model was 150.

***** WARNING**

Data set contains cases with missing on x-variables.

These cases were not included in the analysis.

Number of cases with missing on x-variables: 31

***** WARNING**

Data set contains cases with missing on all variables except

x-variables. These cases were not included in the analysis.

Number of cases with missing on all variables except x-variables: 19

2 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS

SUMMARY OF ANALYSIS

Number of groups 1

Number of observations 150

Why has Mplus excluded cases with missing values on the predictor variables, when it typically includes such cases? For some models (including count models) the predictor variables (called observed covariates by Mplus) are not included in the model in the same manner as other variables, and hence their missing values cannot be handled using maximum likelihood based techniques. Note that in some models, for example a model similar to this one, but with a continuous outcome, predictor variables are considered part of the model, and cases with missing values on these variables are included in the analysis. It is possible to add predictors that would not otherwise be considered part of a model to the model (see below), which will allow for missing values. Note that when the

predictors are included in the model the same distributional assumptions that are made about other variables in the model (e.g., normality) are now also made about the predictor variables.

There are two differences between the model shown below and the one shown above. Looking at the bottom of the input file, we have added to the model command, including the name of a predictor variable in square brackets includes the mean of the variable in the model. We have also added the analysis command with the integration = montecarlo option because this model requires the use of monte carlo integration. If we leave out this option, Mplus will prompt us to include it.

Data:

file is d:datafiml_count.dat ;

Variable:

names are d1 x1 x2 count;

missing are all (-9999) ;

count is count;

Analysis:

integration = montecarlo;

Model:

```
count on x1 x2 d1;  
;
```

When we run the model, we find that Mplus does not print any error messages, and the number of observations is 200 (i.e., all cases were included in the analysis).

SUMMARY OF ANALYSIS

Number of groups 1

Number of observations 200

ARABPSYCHOLOGY.COM