

Does Mplus correlate the independent variables in a regression model by default?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *Does Mplus correlate the independent variables in a regression model by default?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165056>

Mplus, a statistical software used for structural equation modeling and regression analysis, does not automatically correlate the independent variables in a regression model by default. This means that the software does not assume or enforce any relationships between the independent variables, and the user must explicitly specify any correlations between these variables. This allows for greater flexibility in the analysis and allows the user to control for potential confounding factors in the model.

Does Mplus correlate the independent variables in a regression model by default? | Mplus FAQ

In this simple example of a path model (i.e. all variables are observed) we have a single dependent variable (y), predicted by three independent or predictor variables (x1, x2 and x3). Assume that all of the variables are continuous. To fit this model we use the Mplus input file below. The Model section of the input file contains the command `y on x1 x2 x3`, which specifies that y should be regressed on the three x variables. Note that we have not specifically included correlations between the x variables in our model. We have included `tech1` under Output, this will allow us to see a listing of all parameters estimated in the model. The dataset can be downloaded [here](#) .

Data:

File is `D:datamydata.dat` ;

Variable:

Names are y x1 x2 x3;

Analysis:

Type = general ;

Model:

y on x1 x2 x3;

Output:

tech1;

Below are selected portions of the output generated by the input file. You can view the entire output by clicking [here](#). Under "MODEL RESULTS" we see estimated coefficients for the regression of y on the three x variables, as well as estimates of the intercept of y and the residual variance of y. Note that the covariances of the x variables are not listed, indicating that they were not estimated as part of the model. Although Mplus does not explicitly include the covariances between the independent variables by default, the coefficient estimates are identical to those you would receive if you explicitly modeled the covariances between the independent variables (see below for a demonstration of this). The results from this model are also identical to those you would find if you ran a regression in any statistical package.

MODEL RESULTS

Two-Tailed

Estimate S.E. Est./S.E. P-Value

Y ON

X1 0.392 0.074 5.302 0.000

X2 0.036 0.062 0.589 0.556

X3 0.348 0.072 4.810 0.000

Intercepts

Y 11.155 3.128 3.566 0.000

Residual Variances

Y 50.805 5.080 10.000 0.000

We already have a fairly good idea that the covariances of the independent variables (i.e., x1, x2 and x3) are not being modeled, but if we want to check further, we can. Any time we are unsure about which parameters are being estimated, we can use the option `tech1` to get a listing of all parameters in a model.

We can use the listing of model parameters to confirm that the covariances were not estimated. The `tech1` option produces two sets of matrices, one of which

shows all the estimated parameters (the other shows their starting values, but we won't deal with that set in this FAQ). Each parameter estimated by the model is listed in one of the matrices printed by tech1. Within the model, estimated parameters are identified with numbers, starting with one and numbering sequentially. Parameters that are not estimated are represented by 0s in the matrices. The matrices (and sometimes vectors) are identified by Greek letters. The content of each of these matrices is described in the chapter of the Mplus manual titled "Output, savedata, and plot commands," under the heading "Mplus Parameter Arrays". Note that not all matrices are required for every model, and only the relevant matrices are printed. The PSI matrix contains the covariances of the continuous variables, which is where the covariances of the independent variables in this model would be listed, if they were estimated (all other matrices were omitted to save space). The PSI matrix can be read like a correlation matrix, except that the values listed (in this case) are the parameter numbers, rather than the estimates of the values of those parameters. Looking at the PSI matrix below, we see that the variance of Y is estimated (indicated by the number 5 at the top left), but that no

other variances (on the diagonal) or covariances (the off diagonals) were estimated in the model. This is consistent with the output in the "MODEL RESULTS" section above, where the residual variance of y is listed, but no other variances or residual variances are listed.

TECHNICAL 1 OUTPUT

PARAMETER SPECIFICATION

<output omitted>

PSI

Y X1 X2 X3

Y 5

X1 0 0

X2 0 0 0

X3 0 0 0 0

What happens if covariances between the independent variables are included?

Unlike the input in the first example, the input file shown here explicitly includes the covariances between the independent variables (x_1 , x_2 and x_3), otherwise the

two models are identical.

Data:

File is D:datamydata.dat ;

Variable:

Names are y x1 x2 x3;

Analysis:

Type = general ;

Model:

y on x1 x2 x3;

x1 with x2 x3;

x2 with x3;

Output:

tech1;

Below is the output from the model that explicitly requests the covariances between the x variables. The coefficients and standard errors for the regression portion (i.e., Y ON) of this model are identical to those in the example above. The y intercept (under Intercepts) for this model is 11.153, while the y intercept in the above model is 11.155, a difference that is substantively unimportant given the scale of the y variable (the scale of a variable can be seen in the

descriptive statistics, e.g., means and standard deviations). The standard errors of the intercepts in the two models are identical. The estimate of the residual variance of y is also nearly identical. So we see that including the covariances does not seem to effect the coefficients we are likely to be most interested in, the regression coefficients. That said, there are some differences between the two models. Below the regression coefficients are the covariances between the x variables (denoted WITH) that were requested in the model input statement, these were not estimated in the first model. In addition to the covariances we requested, the model includes (and hence the output includes), estimates of the mean and variance of each of the x variables. When it estimates the covariances of the x variables, Mplus also includes their means and variances in the model, so this model includes everything in the first model, plus the extra parameters required to explicitly model the covariances. Note that the results from both models should match regression output from statistical packages such as SAS, SPSS or Stata.

MODEL RESULTS

Two-Tailed

Estimate S.E. Est./S.E. P-Value

Y ON

X1 0.392 0.074 5.302 0.000

X2 0.036 0.062 0.590 0.555

X3 0.348 0.072 4.810 0.000

X1 WITH

X2 54.489 8.057 6.763 0.000

X3 63.297 8.106 7.809 0.000

X2 WITH

X3 68.067 9.118 7.465 0.000

Means

X1 52.645 0.661 79.670 0.000

X2 52.405 0.757 69.206 0.000

X3 52.230 0.723 72.223 0.000

Intercepts

Y 11.153 3.128 3.565 0.000

Variances

X1 87.329 8.733 10.000 0.000

X2 114.681 11.468 10.000 0.000

X3 104.597 10.460 10.000 0.000

Residual Variances

Y 50.806 5.081 10.000 0.000

ARABPSYCHOLOGY.COM