Simple Intercepts, Simple Slopes, and Regions of Significance in HLM 3-Way Interactions

Kristopher J. Preacher, **Patrick J. Curran**, and **Daniel J. Bauer**
University of North Carolina at Chapel Hill

This web page calculates simple intercepts, simple slopes, and the region of significance to facilitate the testing and probing of three-way interactions estimated in hierarchical linear regression models (HLMs). The interaction can involve any combination of dichotomous and continuous variables. This web page explores only those situations involving the interaction of two level 2 predictors of level 1 slopes. We use the general notation $\hat{\omega}_0$ to define the simple intercept and $\hat{\omega}_1$ to define the simple slope. We use the standard notation of Raudenbush and Bryk (2002) to define terms in each equation, and we assume that the user is knowledgeable both in the general HLM and in the testing, probing, and interpretation of interactions in multiple linear regression (e.g., Aiken & West, 1991). The following material is intended to facilitate the calculation of the methods presented in Bauer and Curran (2004) and Curran, Bauer, and Willoughby (in press), and we recommend consulting these papers for further details, as well as our **companion web page** on 2-way interactions in HLM.

The case we consider here involves one predictor at level 1 and a two-way interaction estimated at level 2. This can be seen as an extension of what is sometimes referred to as a *slopes as outcomes* model. Using the two-level notation system of Raudenbush and Bryk (2002), the level 1 equation is expressed as

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + r_{ij} \tag{1}$$

where y_{ij} is the observed value of outcome y for observation i nested within group j, β_{0j} is the intercept for group j, β_{1j} is the regression slope of y on x within group j, and r_{ij} is the person and group specific residual. These regression coefficients can then be modeled as a function of two level 2 covariates (w_{1j} and w_{2j}) and their interaction ($w_{1j}w_{2j}$) such that

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}w_{1j} + \gamma_{02}w_{2j} + \gamma_{03}w_{1j}w_{2j} + u_{0j}\\ \beta_{1j} &= \gamma_{10} + \gamma_{11}w_{1j} + \gamma_{12}w_{2j} + \gamma_{13}w_{1j}w_{2j} + u_{1j}\end{aligned} \tag{2}$$

where, for the equation predicting individual differences in intercepts, γ_{00} is the fixed intercept, γ_{01}, γ_{02}, and γ_{03} are the fixed regression coefficients for the two main effects and the level 2 interaction, respectively, and u_{0j} is the level 2 residual. Similarly, for the equation predicting individual differences in level 1 slopes, γ_{10} is the fixed intercept, γ_{11}, γ_{12}, and γ_{13} are the fixed regression coefficients for the two main effects and the interaction, respectively, and u_{1j} is the level 2 residual. Finally, the level 2 equations can be substituted into the level 1 equation to form the reduced form equation such that

$$\begin{aligned}y_{ij} = {} & \gamma_{00} + \gamma_{01}w_{1j} + \gamma_{02}w_{2j} + \gamma_{03}w_{1j}w_{2j}\\ & + \left(\gamma_{10} + \gamma_{11}w_{1j} + \gamma_{12}w_{2j} + \gamma_{13}w_{1j}w_{2j}\right)x_{1ij} + \left(u_{0j} + u_{1j}x_{1ij} + r_{ij}\right)\end{aligned} \tag{3}$$

It can be seen that the regression of the level 1 slope on the level 2 covariates (and their product) results in a cross-level interaction among x_{1ij}, w_{1j}, and w_{2j} with regression coefficient γ_{13}.

Following the methods described in Bauer and Curran (2004), we can define the conditional regression of y on x (denoted the *focal* predictor) as a function of w_1 and w_2 (denoted the *moderators*). Note that this distinction between focal predictor and moderators is arbitrary given the symmetry of the interaction. Rearrangement of the expected value of the reduced form equation, conditioned on w_1 and w_2, results in

$$\begin{aligned}\mu_{y|w_1,w_2} = \quad &(\gamma_{00} + \gamma_{01}w_1 + \gamma_{02}w_2 + \gamma_{03}w_1w_2) \\ &+ (\gamma_{10} + \gamma_{11}w_1 + \gamma_{12}w_2 + \gamma_{13}w_1w_2)x_1\end{aligned} \tag{4}$$

where $\mu_{y|w1w2}$ denotes the model implied mean value of y as a function of x at specific values of w_1 and w_2, and where the *simple intercept* and *simple slope* for the conditional regression of y on x as a function of w_1, w_2, and their interaction are given by the first and second parenthetical expression, respectively. The sample estimates of these compound effects can be explicitly defined as

$$\hat{\mu}_{y|w_1,w_2} = \hat{\omega}_0 + \hat{\omega}_1 x_1 \tag{5}$$

where

$$\begin{aligned}\hat{\omega}_0 &= \hat{\gamma}_{00} + \hat{\gamma}_{01}w_1 + \hat{\gamma}_{02}w_2 + \hat{\gamma}_{03}w_1w_2 \\ \hat{\omega}_1 &= \hat{\gamma}_{10} + \hat{\gamma}_{11}w_1 + \hat{\gamma}_{12}w_2 + \hat{\gamma}_{13}w_1w_2\end{aligned} \tag{6}$$

The sample estimates of the simple intercept ($\hat{\omega}_0$) and simple slope ($\hat{\omega}_1$) define the conditional regression of y on x as a function of w_1, w_2, and their interaction. Because these are sample estimates, we must compute standard errors to conduct inferential tests of these effects. The computation of these standard errors is one of the key purposes of our calculator.

Summary. We are primarily interested in the estimation of the simple intercept ($\hat{\omega}_0$) and the simple slope ($\hat{\omega}_1$) of the conditional regression of the outcome on the focal predictor as a function of the moderators. We now turn to a brief description of the values that can be calculated using our table below.

The Region of Significance. The first available output is the region of significance of the simple slope describing the relation between the outcome y and the focal predictor x as a function of level 2 moderators w_1 and w_2. We do not provide the region of significance for the simple intercept given that this is rarely of interest in practice. The region of significance defines the specific values of the moderator at which the slope of the regression of y on the focal predictor transitions from non-significance to significance. Although this region can be easily obtained when testing a two-way interaction, these are much more complex to compute for a three-way interaction (see Curran, Bauer, & Willoughby, in press for further details). As is proposed in Curran et al. (in press), the table allows for the calculation of the region of significance of the regression of y on x across values of w_1 at a particular value of w_2. This is a melding of the simple slopes and region approach. There are lower and upper bounds to the region. In many cases, the regression of y on the focal predictor is significant at values of the moderator that are *less than* the lower bound and *greater than* the upper bound, and the regression is non-significant at values of the moderator falling *within* the region. However, there are some cases in which the opposite holds (e.g., the significant slopes fall within the region). Consequently, the output will explicitly denote how the region should be defined in terms of the significance and non-significance of the simple slopes. There are also instances in which the region cannot be mathematically obtained, and an error is displayed if this occurs for a given application. However, this region is calculated for a *specific conditional value of* w_2. The region can be re-calculated at several different conditional values of w_2 (e.g., ±1SD) to gain a better understanding of the structure of the three-way interaction. By default, the region is calculated at α= .05, but this may be changed by the user. Finally, the point estimates and standard errors of both the simple intercepts and the simple slopes are automatically calculated precisely at the lower and upper bounds of the region.

<u>*Simple Intercepts and Simple Slopes*</u>. The second available output is the calculation of point estimates and standard errors for up to two simple intercepts and simple slopes of the regression of *y* on the focal predictor at specific levels of the moderators. In the table we refer to these specific values of the moderators as *conditional values*. We can choose from a variety of potential conditional values of w_1 and w_2 for the computation of the simple intercepts and slopes. If w_1 or w_2 is dichotomous, we could select conditional values of 0 and 1 to compute the regression of *y* on *x* within group 0 and group 1. If w_1 or w_2 is continuous, we might select conditional values that are one standard deviation above the mean of w_1 or w_2 and one standard deviation below the mean of w_1 or w_2. Whatever the conditional values chosen, these specific values are entered in the sections labeled "Conditional Values of w_1" and "Conditional Values of w_2," and this will provide the corresponding simple slopes of *y* on *x* at those values of w_1 and w_2. The calculation of simple intercepts and slopes at specific values of the moderator is optional; the user may leave any or all of the conditional value fields blank.

<u>*Using the Calculator*</u>. Simple intercepts, simple slopes, and the region of significance can be obtained by following these six steps:

1. We strongly suggest writing out by hand the equation provided at the top of the table (this equation is essentially the same as Equation 4). This will significantly aid in keeping track of the necessary values to enter into the table. Note that any other covariates that are included at any level in the equations can be ignored in the calculation of the simple slopes; that is, one or more covariates can be included in the estimation of the HLM analysis, but these do not play a role in the probing of the specific interaction. For interpretational purposes, however, it is essential that values of zero be within the bounds of the data. We recommend that continuous covariates be mean centered prior to analysis and that a useful reference group be chosen for categorical covariates. However, there can be only one interaction effect in the equation that involves the focal predictor and the moderators of interest for this table to yield meaningful results.
2. Enter the sample values for the regression coefficients that correspond to the simple intercept and simple slope of interest. It is *extremely* important that this numbering system consistently correspond to the focal predictor and moderators throughout the calculations.
3. Enter the asymptotic variances of the required regression parameters under "Coefficient Variances"; note that these are the squared standard errors. Also enter the necessary asymptotic covariances under "Coefficient Covariances." All of these values can be obtained from the asymptotic covariance matrix of the regression parameters available in most standard multilevel modeling packages. More information on obtaining the ACOV matrix can be found **here**.

4. The region of significance and the simple intercept and simple slope calculated at the boundaries of this region are provided by default. At a minimum, the user must provide the sample regression parameters, and asymptotic variances and covariances. One available option is the selection of the probability value upon which to calculate the region. The default value is α= .05, but this can be changed to any appropriate value (e.g., .10 or .025).
5. If the calculation of additional *simple intercepts* and *simple slopes* is desired for specific conditional values of w_1 and w_2, enter the conditional values of w_1 and w_2 at which to estimate these values. If w_1 or w_2 is dichotomous and was originally coded 0 and 1 to denote group membership, enter 0 and 1 for the first and second conditional values, and leave the third cell blank. If w_1 or w_2 is continuous, any two conditional values may be selected as described above (results for more than two conditional values may be obtained by re-entering additional conditional values and recalculating). If these conditional value fields are all left blank, no simple intercepts or simple slopes will be provided.
6. You have the option of entering custom degrees of freedom for tests of simple intercepts, tests of simple slopes, or both simple intercepts and simple slopes. If either of these boxes are left blank, asymptotic z-tests will be conducted instead of t-tests.

Once all of the necessary information is entered into the table, click "Calculate." The status box will identify any errors that might have been encountered. If no errors are found, the results will be presented in the output window. Although the results in the output window cannot be saved, the contents can be copied and pasted into any word processor for printing.

R Code for Creating Simple Slopes Plot. Below the output window are two additional windows. If conditional values of x_1 (Points to Plot) and w_1, as well as at least one conditional value of w_2, are entered, clicking on "Calculate" will also generate R code for producing a plot of the interaction between x_1 and w_1 at the lowest value of w_2 (R is a statistical computing language). This R code can be submitted to a remote Rweb server by clicking on "Submit above to Rweb." A new window will open containing a plot of the interaction effect. The user may make any desired changes to the generated code before submitting, but changes are not necessary to obtain a basic plot. Indeed, this window can be used as an all-purpose interface for R.

R Code for Creating Confidence Bands / Regions of Significance Plot. Assuming enough information is entered into the interactive table, the second output window below the table will include R syntax for generating confidence bands, continuously plotted confidence intervals for simple slopes corresponding to all conditional values of the moderator. The x-axis of the resulting plot will represent conditional values of the moderator, and the y-axis represents values of the simple slope of *y* regressed on the focal predictor.

If the moderator w_1 is dichotomous, only two values along the x-axis (corresponding to the codes used for grouping) would be interpretable. Therefore, in cases where the focal predictor x_1 is continuous and the moderator w_1 is dichotomous, we suggest treating w_1 as the moderator for the simple slopes plot (so that each line will represent the regression of *y* on x_1 at conditional values of the moderator) and treating x_1 as the moderator for the confidence bands / regions of significance plot (so that the x-axis will represent values of the focal predictor and the y-axis will represent the group difference in *y* at conditional values of the focal predictor). This will require switching the roles of the focal predictor and the moderator in the interactive table, requiring the entry of some new values from the **ACOV matrix** and re-entering old values in new places.

Regardless of what variable is treated as the moderator, the user is expected to supply lower and upper values for the moderator (-10 and +10 by default). As above, this R code can be submitted to a remote Rweb server by clicking on "Submit above to Rweb." A new window will open containing a plot of confidence bands.

References

Bauer, D. J., & Curran, P. J. (in press). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*.

Curran, P. J., Bauer, D. J, & Willoughby, M. T. (in press). Testing and probing interactions in hierarchical linear growth models. To appear in C.S. Bergeman & S.M. Boker (Eds.), *The Notre Dame Series on Quantitative Methodology, Volume 1: Methodological Issues in Aging Research*. Mahwah, NJ: Lawrence Erlbaum Associates.