Simple intercepts, simple slopes, and regions of significance in HLM 2-way interactions

Kristopher J. Preacher (*Vanderbilt University*)
Patrick J. Curran (*University of North Carolina at Chapel Hill*)
Daniel J. Bauer (*University of North Carolina at Chapel Hill*)

If the Rweb server is not working

The code generated by this utility can be pasted directly into an R console window. R (a free, open-source statistical computing environment) may be obtained here: http://cran.r-project.org/.

This web page calculates simple intercepts, simple slopes, and the region of significance to facilitate the testing and probing of two-way interactions estimated in hierarchical linear regression models (HLMs). The interaction can be between two dichotomous variables, two continuous variables, or a dichotomous and a continuous variable. Further, the interaction can occur solely within level 1 (i.e., *Case 1*), solely within level 2 (i.e., *Case 2*), or result from a cross level prediction of a level 1 random effect by a level 2 covariate (i.e., *Case 3*). Because the analytic methods are identical for probing interactions in all three cases, we use the general notation $\hat{\omega}_0$ to define the simple intercept and $\hat{\omega}_1$ to define the simple slope regardless of which case we are considering. We use the standard notation of Raudenbush and Bryk (2002) to define each of these cases, 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.

Case 1

The first case we consider involves an interaction between two predictors within the level 1 equation but with no predictors of these effects at level 2. For the two predictor case, the level 1 equation is

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{1ij}x_{2ij} + r_{ij} \tag{1}$$

where y_{ij} is the value of y for observation i in group j, x_{1ij}, and x_{2ij} are the two level 1 covariates for observation i in group j, and $x_{1ij}x_{2ij}$ is the interaction between the two level 1 covariates. Further, β_{0j} is the intercept of the regression equation for group j, β_{1j} and β_{2j} are the main effects of x_{1ij} and x_{2ij}, respectively, β_{3j} is the within-level interaction between x_{1ij} and x_{2ij}, and r_{ij} is the observation- and group-specific residual. Because the regression parameters are viewed as random variables, these can be expressed in the level 2 equations as

$$
\begin{aligned}
\beta_{0j} &= \gamma_{00} + u_{0j} \\
\beta_{1j} &= \gamma_{10} + u_{1j} \\
\beta_{2j} &= \gamma_{20} + u_{2j} \\
\beta_{3j} &= \gamma_{30} + u_{3j}
\end{aligned}
\tag{2}
$$

where the γs represent the fixed regression coefficients and the u's represent the group-specific deviations from the fixed effects. This formulation is sometimes referred to as a *random effects regression model* given that the level 1 regression coefficients vary over the level 2 units, but are not conditioned on level 2 covariates. We can substitute the level 2 equations into the level 1 equation to result in the reduced form equation such that

$$
\begin{aligned}
y_{ij} = &\left(\gamma_{00} + \gamma_{10}x_{1ij} + \gamma_{20}x_{2ij} + \gamma_{30}x_{1ij}x_{2ij}\right) \\
&+ \left(u_{0j} + u_{1j}x_{1ij} + u_{2j}x_{2ij} + u_{3j}x_{1ij}x_{2ij} + r_{ij}\right)
\end{aligned}
\tag{3}
$$

The first parenthetical term represents the fixed effects and the second parenthetical term represents the random effects. If the interaction term (e.g., γ_{30}) is found to be significant, it is necessary to further probe this effect to identify the precise nature of this conditional relation.

Following the methods described in Bauer and Curran (2004), we can define the conditional regression of y on x_1 (denoted the *focal* predictor) as a function of x_2 (denoted the *moderator*). Note that this distinction between focal predictor and moderator is arbitrary given the symmetry of the interaction. Rearrangement of the expected value of the reduced-form equation highlights the conditional relation between the dependent variable y and focal predictor x_1 as a function of the moderator x_2:

$$
\mu_{y|x_2} = \left(\gamma_{00} + \gamma_{20}x_2\right) + \left(\gamma_{10} + \gamma_{30}x_2\right)x_1 \tag{4}
$$

where $\mu_{y|x2}$ denotes the model implied mean value of y as a function of x_1 at a specific value of x_2. Note that Equation (4) has the form of a simple regression of y on x_1 where the first parenthetical term is the intercept of the regression and the second parenthetical term is the slope of the regression. We will refer to the first parenthetical term as the *simple intercept* and the second term as the *simple slope*. It can be seen that the simple intercept and simple slope are *compound coefficients* that result from the linear combination of other regression parameters. To further explicate this, we can re-express Equation (4) in terms of sample estimates of population values such that

$$
\hat{\mu}_{y|x_2} = \hat{\omega}_0 + \hat{\omega}_1 x_1 \tag{5}
$$

where

$$\hat{\omega}_0 = \hat{\gamma}_{00} + \hat{\gamma}_{20} x_2$$
$$\hat{\omega}_1 = \hat{\gamma}_{10} + \hat{\gamma}_{30} x_2 \tag{6}$$

The sample estimates of the simple intercept ($\hat{\omega}_0$) and simple slope ($\hat{\omega}_1$) define the conditional regression of y on x_1 as a function of x_2. 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 calculators.

Case 2

The second case arises when there are no predictors at level 1 and there is a two-way interaction estimated within level 2. This is sometimes referred to as a *means as outcomes* model. Using the two-level notation system of Raudenbush and Bryk (2002), the level 1 equation is expressed as

$$y_{ij} = \beta_{0j} + r_{ij} \tag{7}$$

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, and r_{ij} is the person and group specific residual. Because there are no predictors, the intercept represents the model implied mean of y within group j. These group means 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

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + \gamma_{02} w_{2j} + \gamma_{03} w_{1j} w_{2j} + u_{0j} \tag{8}$$

where 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 interaction, respectively, and u_{0j} is the level 2 residual. Finally, the level 2 equation can be substituted into the level 1 equation to form the reduced form equation such that

$$y_{ij} = \left(\gamma_{00} + \gamma_{01} w_{1j} + \gamma_{02} w_{2j} + \gamma_{03} w_{1j} w_{2j}\right) + \left(r_{ij} + u_{0j}\right) \tag{9}$$

As was described for Case 1, if the interaction term (i.e., γ_{03}) is found to be significant, it is necessary to further probe this effect. We can again (arbitrarily) define the conditional regression of y on w_1 (the *focal* predictor) as a function of w_2 (the *moderator*). Rearrangement the expected value of the reduced form equation results in

$$\mu_{y|w_2} = \left(\gamma_{00} + \gamma_{02} w_2\right) + \left(\gamma_{01} + \gamma_{03} w_2\right) w_1 \tag{10}$$

where $\mu_{y|w2}$ represents the model implied value of y as a function of w_1 at a specific value of w_2. As before, the first term represents the simple intercept and the second the simple slope. The sample estimates of these compound effects can be explicitly defined as

$$\hat{\mu}_{y|w_2} = \hat{\omega}_0 + \hat{\omega}_1 w_1 \tag{11}$$

where

$$\begin{aligned} \hat{\omega}_0 &= \hat{\gamma}_{00} + \hat{\gamma}_{02} w_2 \\ \hat{\omega}_1 &= \hat{\gamma}_{01} + \hat{\gamma}_{03} w_2 \end{aligned} \tag{12}$$

The sample estimates of the simple intercept ($\hat{\omega}_0$) and simple slope ($\hat{\omega}_1$) define the conditional regression of y on w_1 as a function of w_2.

Case 3

The third and final case arises when there is a single main effect predictor at level 1 and a single main effect predictor at level 2 which is manifested in the reduced form equation as a cross-level interaction. This type of conditional relation may be the most commonly encountered in many HLM applications and is sometimes referred to as a *slopes as outcomes* model. The level 1 equation is

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + r_{ij} \tag{13}$$

where x_{1ij} is the observed predictor for observation i nested within group j, β_{1j} is the regression slope of y on x_1 within group j, and all else is defined as above. The level 2 equations are

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} w_{1j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} w_{1j} + u_{1j} \end{aligned} \tag{14}$$

where w_{1j} is the observed predictor for group j, γ_{00} and γ_{10} are the fixed intercepts, γ_{01} and γ_{11} are the fixed regression coefficients for w_{1j}, and u_{0j} and u_{1j} are the residual terms. Finally, substituting the level 2 equation into the level 1 equation results in the reduced form equation such that

$$y_{ij} = \left(\gamma_{00} + \gamma_{10} x_{1ij} + \gamma_{01} w_{1j} + \gamma_{11} x_{1ij} w_{1j}\right) + \left(u_{0j} + u_{1j} x_{1ij} + r_{ij}\right) \tag{15}$$

It can be seen that the regression of the level 1 slope on the level 2 covariate results in a cross-level interaction between x_{1ij} and w_{1j} with regression coefficient γ_{11}. Rearrangement of the expected value of the reduced form equation results in

$$\mu_{y|w_1} = \left(\gamma_{00} + \gamma_{01} w_1\right) + \left(\gamma_{10} + \gamma_{11} w_1\right) x_1 \qquad (16)$$

where the simple intercept and simple slope for the conditional regression of y on x_1 as a function of w_1 are given by the first and second parenthetical expression, respectively. The sample estimates of these compound effects can be explicitly defined as

$$\hat{\mu}_{y|w_1} = \hat{\omega}_0 + \hat{\omega}_1 x_1 \qquad (17)$$

where

$$\begin{aligned} \hat{\omega}_0 &= \hat{\gamma}_{00} + \hat{\gamma}_{01} w_1 \\ \hat{\omega}_1 &= \hat{\gamma}_{10} + \hat{\gamma}_{11} w_1 \end{aligned} \qquad (18)$$

The sample estimates of the simple intercept ($\hat{\omega}_0$) and simple slope ($\hat{\omega}_1$) define the conditional regression of y on x_1 as a function of w_1. It is sometimes of interest to estimate a cross-level interaction in which the question of interest revolves around the simple slope of y on w_1 as a function of x_1, but we do not address such situations here. The Case 3 table may be used for such situations, switching x_1 and w_1 (and parameters associated with them) where appropriate.

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 moderator. When comparing the calculation of the simple intercepts and slopes across the three cases above, it is clear that these all share a common computational form, and this is why we have used the same notation to define the simple intercept and slope for each case. However, to simplify the use of our tables in practice, we have developed calculators separately for each of the three cases, although the underlying analytics are all identical (see Bauer & Curran, 2004, for details). We now turn to a brief description of the values that can be calculated using our tables 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 as a function of the moderator. 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. 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. By default, the region is calculated at $\alpha = .05$, 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.

Simple Intercepts and Simple Slopes

The second available output is the calculation of point estimates and standard errors for up to three simple intercepts and simple slopes of the regression of *y* on the focal predictor at specific levels of the moderator. In the table we refer to these specific values of the moderator as *conditional values*. There are a variety of potential conditional values of the moderator that may be chosen for the computation of the simple intercepts and slopes. If the moderator is dichotomous (e.g., 0 or 1 to denote gender), we could select the first and second conditional values to be equal to 0 and 1 to compute the regression of *y* on the focal predictor for males and for females (leaving the third conditional value blank). If the moderator is continuous, we might select values of the moderator that are one standard deviation above the mean, equal to the mean, and one standard deviation below the mean. Whatever the conditional values chosen, these specific values are entered in the section labeled "Conditional Values," and this will provide the corresponding simple intercepts and simple slopes of the regression of *y* on the focal predictor at those specific values of the moderator. 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.

Points to Plot

Given the calculation of one or more simple slopes, it is common to plot these relations graphically to improve interpretability of effects. The final available output is the calculation of a lower and upper value associated with each of the simple slopes to aid in the graphing of these using any standard software package (e.g., Excel, SPSS, etc.). These are provided to simply aid in the graphing of effects; no inferential tests apply here. For the regression of *y* on the focal predictor at specific levels of the moderator, the user enters any two values of the focal predictor in order to plot the regression line between *y* and the predictor at specific values of the moderator. Although any pair of moderator values can be used, we recommend using either the lower and upper *observed* values of the moderator, the lower and upper *possible* values of the moderator, or one *sd* below and above the mean of the moderator. However, many other specific values can be chosen that may be more appropriate for a particular research application.

Using the Calculators

Simple intercepts, simple slopes, and the region of significance can be obtained by following these eight steps. Use as many significant digits as possible for optimal precision.

1. Select whether the interaction takes the form of Case 1, 2, or 3 described above, and select the relevant table for calculations. Again, the underlying computations are identical

across the three cases, and we present three separate tables to ease the use of these methods in practice.

2. We strongly suggest writing out by hand the equation provided at the top of each table for the given application at hand (these equations are essentially the same as Equations 4, 10, or 16, depending on the table). This will significantly aid in keeping track of the necessary values to enter into the tables. 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 moderator of interest for these tables to yield meaningful results.
3. 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 moderator throughout the calculations.
4. 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 any standard computer package. More information on obtaining the ACOV matrix can be found here.
5. 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 $\alpha = .05$, but this can be changed to any appropriate value (e.g., .10 or .025).
6. If the calculation of additional simple intercepts and simple slopes is desired for specific conditional values of the moderator other than the values defined as part of the region of significance, enter the conditional values of the moderator at which to estimate these values. If the moderator 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 the moderator is continuous, up to three conditional values may be selected as described above (results for more than three 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.
7. 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.
8. 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.

9. If the *points to plot* are desired, simply enter a lower and upper value of the focal predictor in the appropriate box. Any values can be used. If these fields are left blank, no points to plot will be provided.

Once all of the necessary information is entered into the table, simply 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. The results in the output window can be 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 and z are entered, clicking on "Calculate" will also generate R code for producing a plot of the interaction effect (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 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 is continuous and the moderator is dichotomous, we suggest treating x_2 (or w_2) as the moderator for the simple slopes plot (so that each line will represent the regression of y on x_1 (or w_1) at conditional values of the moderator) and treating x_1 (or w_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. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*, 373-400.

Curran, P. J., Bauer, D. J, & Willoughby, M. T. (2006). Testing and probing interactions in hierarchical linear growth models. In C. S. Bergeman & S. M. Boker (Eds.), *The Notre Dame Series on Quantitative Methodology, Volume 1: Methodological issues in aging research* (pp. 99-129). Mahwah, NJ: Lawrence Erlbaum Associates.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437-448.