

Simple Intercepts, Simple Slopes, and Regions of Significance in MLR 2-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 and simple slopes, the region of significance, and computes specific values to facilitate the plotting of significant two-way interactions in ordinary least squares (OLS) regression. The interaction can be between two dichotomous variables, two continuous variables, or a dichotomous and a continuous variable. We assume that the user is sufficiently knowledgeable in the testing, probing, and interpretation of interactions in multiple regression (e.g., Aiken & West, 1991; Bauer & Curran, 2004; Cohen, Cohen, West & Aiken, 2003). A more extensive treatment of interaction effects can be found [here](#).

For the purposes of this page, we define y to be the dependent variable, x to be the predictor variable, and z to be the moderator. The regression equation of interest is thus

$$\hat{y} = \hat{b}_0 + \hat{b}_1x + \hat{b}_2z + \hat{b}_3xz \quad (1)$$

where \hat{y} is the model implied value of y , x is the first predictor, z is the second predictor, and xz is the product between the two predictors. Further, \hat{b}_0 is the intercept of the equation, and \hat{b}_1 , \hat{b}_2 , and \hat{b}_3 are the respective regression parameters. This equation may be rearranged to highlight that the regression of y on x (denoted the *focal* predictor) can be understood as a function of z (denoted the *moderator*):

$$\hat{y} = (\hat{b}_0 + \hat{b}_2z) + (\hat{b}_1 + \hat{b}_3z)x \quad (2)$$

The parenthetical terms in Equation 2 are called the *simple intercept* and *simple slope*, respectively. Equation 2 can be rewritten as:

$$\hat{y} = \hat{\omega}_0 + \hat{\omega}_1x \quad (3)$$

where:

$$\hat{\omega}_0 = (\hat{b}_0 + \hat{b}_2 z)$$

$$\hat{\omega}_1 = (\hat{b}_1 + \hat{b}_3 z)$$
(4)

The values of $\hat{\omega}_0$ and $\hat{\omega}_1$ are compound coefficients, the calculation of which is our purpose here. In practice, if the interaction term \hat{b}_3 is found to be significant at a given alpha-level, the regression of y on x is typically "probed" across values of z to better understand the nature of the conditional relation. If other predictors are included in the model (e.g., demographic covariates, etc.), the simple intercepts will be calculated and tested conditioned on values of zero for these covariates. For interpretational purposes, it is thus 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. Note, however, that the computation and testing of simple slopes (often of most interest) do not depend on the scaling of other covariates in the model.

The table below provides three sets of output that allow for further probing of the xz interaction.

The Region of Significance. The first available output is the *region of significance* of the relation between y and x as a function of z . The region of significance defines the specific values of z at which the regression of y on x moves 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. Calculation of these simple intercepts and slopes at any value of z is described below.

Simple Intercepts and Simple Slopes. The second available output is the calculation of point estimates and standard errors for up to three simple intercepts ($\hat{\omega}_0$) and simple slopes ($\hat{\omega}_1$) of y on x at specific conditional values of z . A *simple slope* is defined as the regression of the outcome y on the predictor x at a specific value of the moderator z . In the table we refer to these specific values of z as *conditional values* (cv_{z1} , cv_{z2} , and cv_{z3}). We can choose from a variety of potential conditional values of z for the computation of the simple intercepts and slopes. If z is dichotomous (e.g., z is coded either 0 or 1 to denote gender), we could select $cv_{z1}=0$ and $cv_{z2}=1$ to compute the regression of y on x for males and for females (and cv_{z3} is left blank). If z is continuous, we might select values of z that are one standard deviation above the mean of z , equal to the mean of z , and one standard deviation below the mean of z . Whatever the conditional values chosen, these specific values are entered in the section labeled "Conditional Values of Z," and this will provide the corresponding simple slopes of y on x at those values of z . The calculation of simple intercepts and simple slopes at specific values of z is optional; the user may leave any or all of cv_{z1} , cv_{z2} , and cv_{z3} 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. If desired, the user enters any two values of x in order to plot the regression line between y and x at specific values of z . Although any pair of values can be used, we recommend using either the lower and upper *observed* values of x , the lower and upper *possible* values of x , or one *sd* below and above the mean of x . However, many other specific values can be chosen that may be more appropriate for a particular research application.

Using the Calculator. Simple intercepts and slopes, the region of significance, and points to plot can be obtained by following these five steps:

1. Enter the sample values of \hat{b}_0 , \hat{b}_1 , \hat{b}_2 , and \hat{b}_3 as defined above. It is *extremely* important that this numbering system consistently correspond to x and z throughout.
2. Enter the asymptotic variances of \hat{b}_0 , \hat{b}_1 , \hat{b}_2 , and \hat{b}_3 (i.e., the squared standard errors), and the asymptotic covariances of \hat{b}_2 with \hat{b}_0 and of \hat{b}_3 with \hat{b}_1 . 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](#).
3. The *region of significance*, and simple intercepts and simple slopes calculated at the boundaries of this region, are provided by default. The user must provide the degrees of freedom (df), regression parameters, and asymptotic variances and covariances. Degrees of freedom are determined by the formula $df = N - k - 1$, where N is the sample size and k is the number of predictors (including product terms and any covariates). 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).
4. If the calculation of additional *simple intercepts* and *simple slopes* is desired for specific conditional values of z , enter the degrees of freedom (df) and the conditional values of z at which to estimate the regression of y on x . If z is dichotomous and was originally coded 0 and 1 to denote group membership, simply enter 0 and 1 for cv_{z1} and cv_{z2} and leave cv_{z3} blank. If z is continuous, conditional values might be plus or minus one standard deviation around the mean of z (e.g., $cv_1 = \bar{z} + 1sd_z$, $cv_2 = \bar{z}$, and $cv_3 = \bar{z} - 1sd_z$). Note that for the calculations in this table (and unlike the computer strategy described by Aiken & West, 1991, pp. 18-19), adding one *sd* to the mean of z provides the simple slope one *sd* above the mean, and vice versa. If plus and minus one *sd* is not desired, any conditional values of interest may be used to best correspond to the research question of interest. If these fields are all left blank, no simple intercepts or simple slopes will be provided.
5. If the *points to plot* are desired, simply enter a lower and upper value of the predictor x in the appropriate box. Any values can be used, although we recommend using the lower and upper *observed* values of x , the lower and upper *possible* values of x , or one *sd* below and above the mean of x . 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 x is continuous and z is dichotomous, we suggest treating z as the moderator for the simple slopes plot (so that each line will represent the regression of y on x at conditional values of z) and treating x as the moderator for the confidence bands / regions of significance plot (so that the x-axis will represent values of the focal predictor x and the y-axis will represent the group difference in y at conditional values of x). This will require switching the roles of x and z 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 z (-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

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks: Sage.
- Bauer, D. J., & Curran, P. J. (in press). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*, 3rd ed. Hillsdale: Erlbaum.

All material on these pages not otherwise credited is ©2003 by Kristopher J. Preacher.