Continuous by Continuous Interactions

Joel S Steele, PhD

Adapted from: (http://stackoverflow.com/questions/18147595/plot-3d-plane-true-regression-surface)

Data

These data are a subset of the **High school and beyond** dataset used by numerous authors. For our purposes we will be grabbing a copy of these data from *UCLA*. These data are in **Stata** format and are available from http://www.ats.ucla.edu/stat/data/hsbdemo.dta.

Model with Interactions

Here we start with a model in which **Reading** scores (*read*) are predicted by **Math** (*math*) and **Social Studies** (*socst*) scores, and the **interaction** between them. For this model we leave each predictor **uncentered**. This means that the estimates of the *Intercept* as well as the influence of the predictors (i.e., *math*, *socst*, *math* × *socst*) on the outcome (*read*) are conditional on where each predictor is **zero**.

The model we fit is

$$read_i \sim math_i + socst_i + math_i \times socst_i + \epsilon_i$$

Have a look. Keep in mind that with the interaction included all of the main effects are conditional, thus they are simple effects and no longer "main" effects.

Table 1: Moderated regression estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.843	14.545	2.602	0.010
math	-0.111	0.292	-0.379	0.705
socst	-0.220	0.272	-0.810	0.419
math:socst	0.011	0.005	2.157	0.032

This model fits the data fairly well with a reported $R^2 = 0.546$. The F-table is below.

Table 2: Regression model F-table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
math	1	9175.571	9175.571	189.413	0.000
socst	1	2023.769	2023.769	41.777	0.000
math:socst	1	225.422	225.422	4.653	0.032
Residuals	196	9494.658	48.442	NA	NA

Notice something here, according the the F-table all of the predictors are significant in how much variance they explain. However, from Table 1, of the predictors, only the interaction was significant. This *should* strike you as **strange**. This results from the fact that we have a model with a *continous* by *continous* interaction in it, which makes the other estimates conditional. We investigate this more below.

Plotting the interaction

Since our interaction is between two continuous variables we need tp decide on which one to select different levels of. Which one will serve as the **moderator**? For this example we use *socst* as the **moderator** in order to plot the conditional relationship between the other predictor and the outcome.

We start by selecting *socst* values that are 1 standard deviation above and below the mean. These will be our conditional values for when we plot the relation between *math* and *read*. We can then use the model fit to the data to predict using the one standard deviation above and below values. These will be added to the plot of the points that we create.

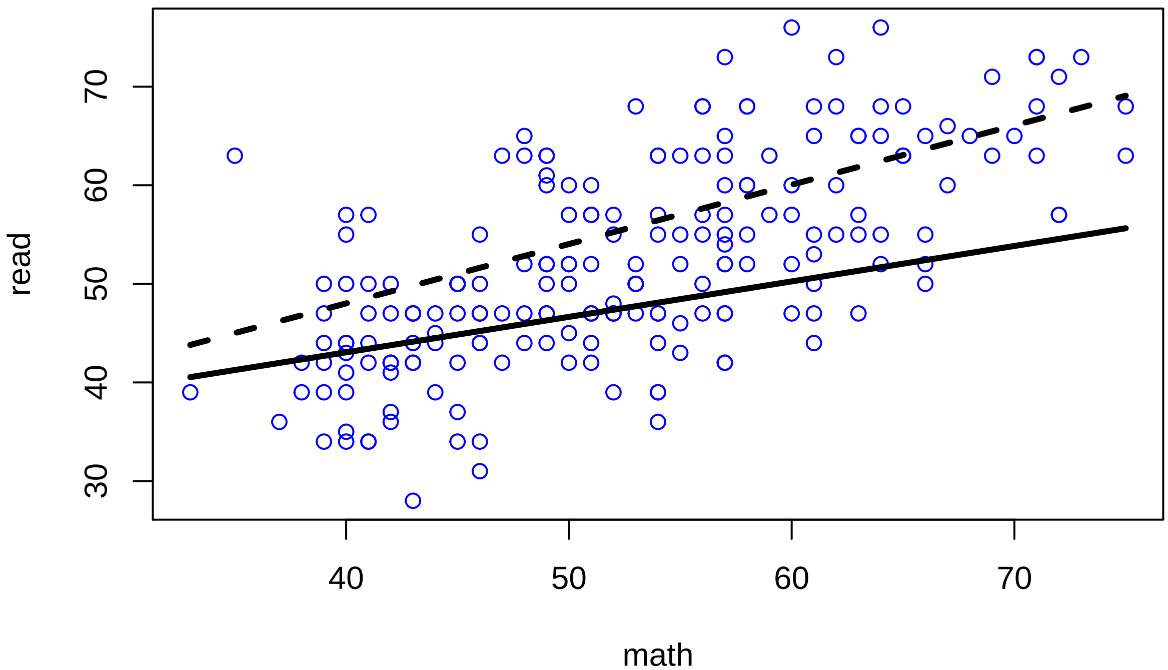

Figure 1: Scatterplot with conditional regression lines +/- 1 standard deviation from the mean on Social Studies.

Selecting a center point

We will now investigate how centering our continuous variables will effect the estimates from our model.

Here we center each variable by subtracting the mean without dividing by the standard deviation. So, the results are still in the original metric and are **NOT** *z*-scores, they are just mean deviation scores now. These variables are defined as:

$$math.c_i = math_i - \bar{math}$$
$$socst.c_i = socst_i - \bar{socst}$$

Using our new mean centered variables we can fit the model again. Keep and eye out for differences.

Table 3: Moderated regression with centered variables

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	51.615	0.569	90.763	0.000
math.c	0.481	0.064	7.545	0.000

	Estimate	Std. Error	t value	Pr(>|t|)
socst.c	0.374	0.056	6.730	0.000
math.c:socst.c	0.011	0.005	2.157	0.032

Ostensibly, that's a big difference! And, what's more, all of the estimates are now significant! It's worth mentioning that the fit didn't change, we still have a reported $R^2 = 0.546$.

The big differences are reflected in the standard errors between the two models. When data are mean centered the standard errors are much smaller. This is also reflected in the tolerance values, which remember are equal to 1 over the variance inflation factor. When we mean center the tolerance goes up.

Table 4: Raw and Centered model variance comparisons

	Raw S.E.	Centered S.E.	Raw tolerance	Centered tolerance
(Intercept)	14.545	0.569	NA	NA
math	0.292	0.064	0.033	0.684
socst	0.272	0.056	0.029	0.685
math:socst	0.005	0.005	0.010	0.964

Result of centering on the parameter estimates

Below we will center each variable in turn to see the differences in the model estimates. Not only does the value of the (Intercept) change, but so do the estimates of *math* and *socst* by themselves. This represents the conditional influence of these variables when the interaction term is included in the model.

Of course you don't have to mean center your data. If there is another value that is supported by theory, or just makes sense to use, you can center on that value. This is the same idea as with centering time. We can focus on different regions of the data to help interpretation or to reflect our hypotheses.

Table 5: Model estimate comparisons

	Raw no intrxn	Raw intrxn	Centered math	Centered socst	All Centered
(Intercept)	7.147	37.843	32.025	26.311	51.615
math	0.504	-0.111	-0.111	0.481	0.481
socst	0.354	-0.220	0.374	-0.220	0.374
intrxn	0.000	0.011	0.011	0.011	0.011

Notice that the interaction term did not change. This represents the curvature of the response surface that results from fitting this model. Notice also that the parameter estimates bounce around depending on if the variables are centered or not. This equates to different slopes for the variables based on where on the response surface we are focusing.

Visualizing the response surface.

Here, We go on to plot the regression surface (using *rgl*). We first plot the points and then graph a response surface for a model with no interactions.

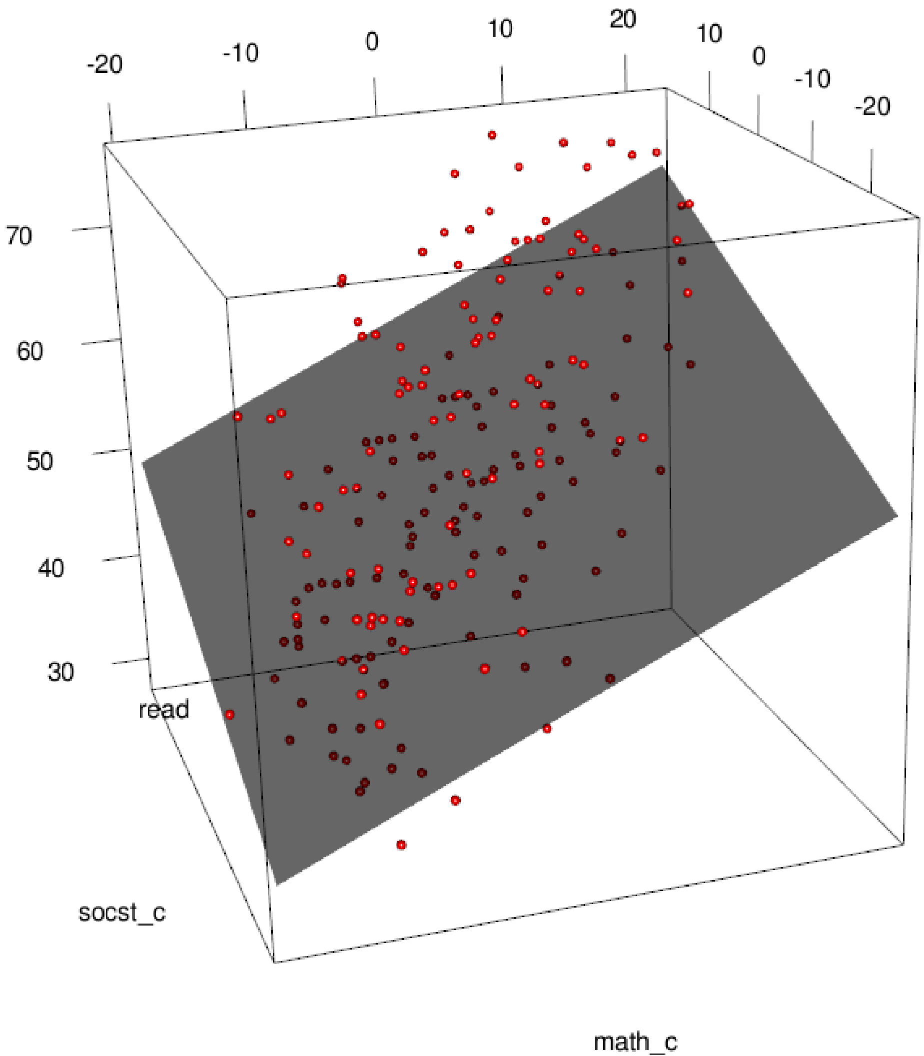

Figure 2: Regression response surface.

Graphing the interaction

Next, we move onto graphing the interaction between the two continuous variables. This is done by first plotting the points as above, but then adding the response surface. Since the interaction is significant, we should expect to see some curvature due to the interaction.

This is great, but... we are only seeing the surface around the neighborhood of the data that we actually have. This is not a bad thing, but it doesn't really illustrate what the surface looks like out of this range.

When we mean center the data we are focusing our estimation around our available data. When predictors are uncentered or raw, the zero points estimated by the model may reflect regions of the response surface that are far out of the range of the available data. We may want to see what to expect if we estimated the effects when each predictor is zero on the original scale. In our case someone who scored zero on both *math* and *socst*. So, when someone has zero for both *math* and *socst*, rather than the averages. In order to accomplish this, we will open up the range of the plot.

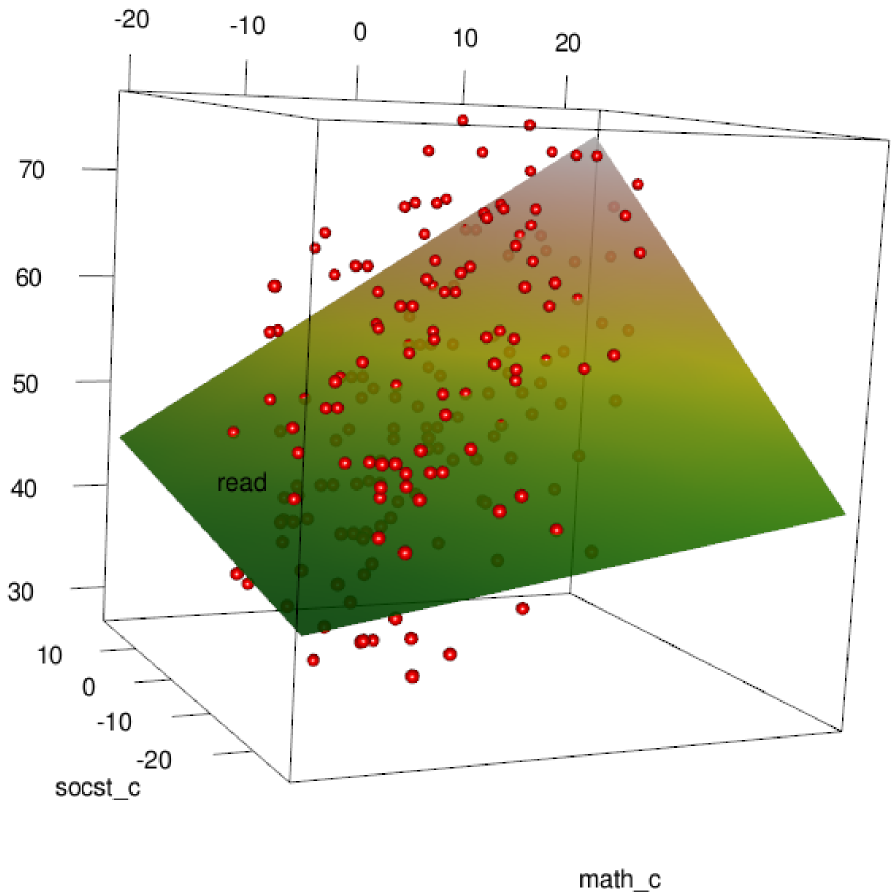

Figure 3: Interaction response surface.

Now we can see that the interaction results in a curved surface. The reason the interaction doesn't change is because it represents the curvature, which is constant along the surface. But, depending on where you are on the surface, the predictor effect lines will have different slopes. Thus the estimates of the other terms are conditional!

Interpretation in applied terms

What this, hopefully, shows is that, when the model contains an interaction, the magnitude and direction of the relation between a continous predictor and the outcome depends on the level of the other predictor in the model. So, for example, we might find that when someone scores a zero on both *math* and *socst*, the relation of either of these values on *read* is not and different from zero (no relation at all). However, as the scores in either measure increase, the relation with *read* changes. We see this in the difference between the estimates from the raw score model versus the estimates from the model with centered variables. In the first case the effects of *math* and *socst* were non significant. This was because we were focusing on the part of the curve where these values were zero on the raw scale. By centering the data, and making each predictor into a mean-deviation score, we move the focus of the model to be within the middle of our cloud of data. As a result the influence of each predictor on the outcome was worth noting.

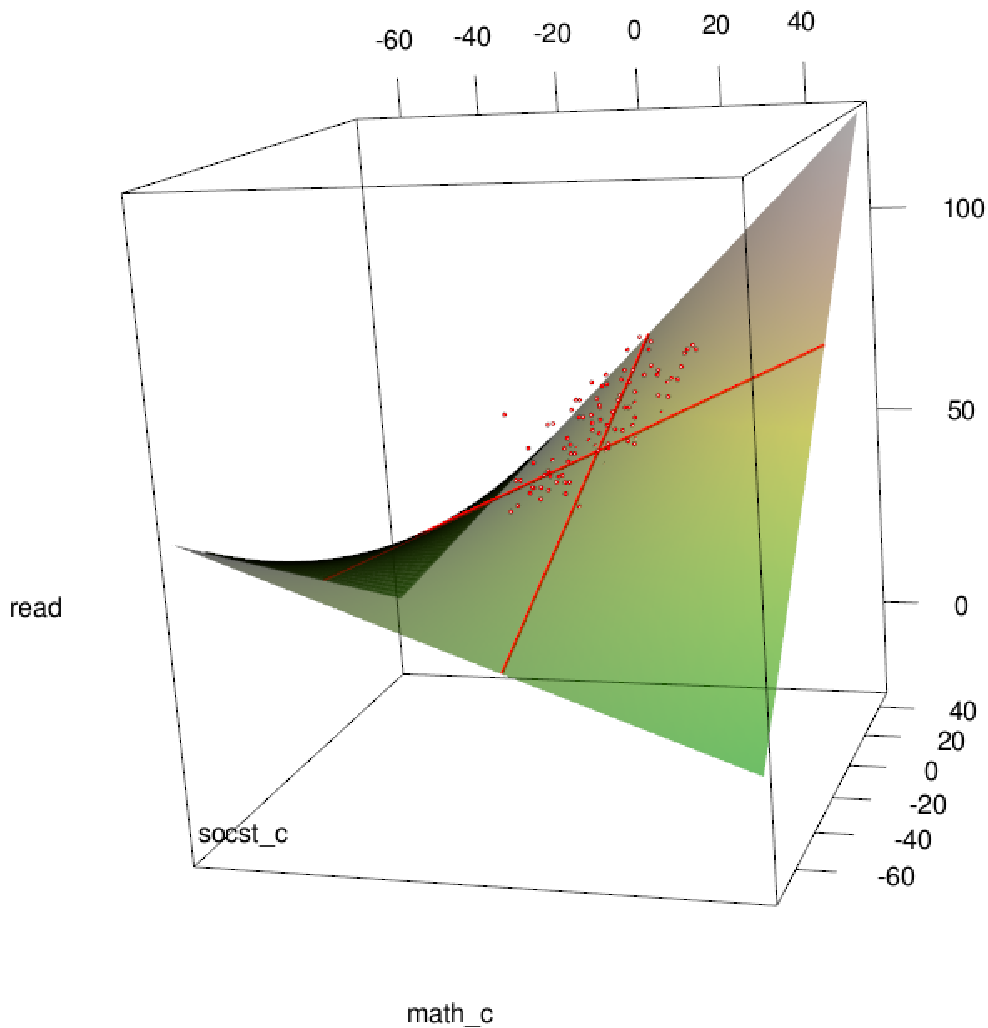

Figure 4: Expanded interaction response surface with predictor effect lines.

The other major result was that the model could be estimated much more precisely, which was reflected in the smaller standard errors between the two models.