Probing Interactions among Continuous Variables (Example from Class)

The regression equation we obtained in class was:

$$\widehat{Grade} = 61.86 + 3.27Books + 1.53Attendance + .72(Books \times Attendance)$$

We can rearrange this equation as follows:

$$\widehat{Grade} = (61.86 + 1.53Attendance) + (3.27 + .72Attendance) \times Books$$

The term in green is called the simple intercept. It is the predicted value of grades when books equals zero, at a given value of attendance.
The term in red is called the simple slope. It is the predicted change in grades associated with a one unit increase in books, at a given value of attendance.

We will now calculate the simple slopes to see how the slopes differ at various values of attendance. So let's investigate values of the slopes at low, medium, and high values of attendance – values 1 SD below the mean, at the mean, and 1 SD above the mean of attendance. In order to calculate these values, we obtain SDs using **proc means**. Given that the SD of attendance is 4.28, and that our variables are centered (and thus the mean is zero), the values corresponding to 1 SD below the mean, at the mean, and 1 SD above the mean are: -4.28, 0, and 4.28. So the simple slopes are calculated as follows:

$$\text{Simple slope at 1 SD below the mean: } \big(3.27 + .72(-4.28)\big) = 0.19$$
$$\text{Simple slope at the mean: } \big(3.27 + .72(0)\big) = 3.27$$
$$\text{Simple slope at 1 SD above the mean: } \big(3.27 + .72(4.28)\big) = 6.35$$

```
proc means data=center;
var books attend;
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
books	40	5.551115E-17	1.7327908	-2.1500000	4.8500000
attend	40	3.108624E-16	4.2775502	-8.1000000	5.9000000

We want to get two things: (1) hypothesis tests and plots of the relationship between books and grade at the three values of attendance that we selected (which are -4.28, 0, and +4.28), and (2) the range of values of attendance for which the relationship between books and grade is significant.

We will use the online calculator by Kris Preacher: http://quantpsy.org/interact/mlr2.htm. The background information appears at the top of the website page, and it is very helpful.

We can fill it in the required information using output from **proc reg** by specifying the covb option. This requests the variance-covariance matrix of the parameters. Basically, these are the variances and covariances of the parameter estimates (i.e., the variance and covariance of the sampling distribution that these parameters are presumed to be drawn from). The variances are the squares of the standard errors, which we know are the standard deviations of the sampling distribution.

Degrees of freedom are determined by $df = n - p$, where n is the sample size and p is the number of parameters estimated.

Be careful when inputting values. Here the moderator is attend (z variable), which corresponds to the regression coefficient of B_2. But we could have chosen books to be the moderator.

On the next page, I've shown the output, as well as the populated fields for the online calculator, so you can see where the values come from. The conditional values of z are just the ones we chose. The conditional values of x are just the upper and lower limits of the X axis – I've chosen -10 and 10, since that seems to encompass the entire range of our centered attendance variable.

Everything else has a box around it. If a field has a red box around it, the information came from the standard proc reg output; if it has a blue box around it, it came from the covariance matrix.

```
proc reg data=center;
model grade=books attend bk_att_c / covb; /* prints asymptotic covariance
matrix */
run; quit;
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4165.56823	1388.52274	7.44	0.0005
Error	36	6718.33177	186.62033		
Corrected Total	39	10884			

Root MSE	13.66091	R-Square	0.3827
Dependent Mean	63.55000	Adj R-Sq	0.3313
Coeff Var	21.49631		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	61.85618	2.27841	27.15	<.0001
books	1	3.27263	1.38377	2.37	0.0235
attend	1	1.52561	0.54128	2.82	0.0078
bk_att_c	1	0.71772	0.30720	2.34	0.0252

Covariance of Estimates

Variable	Intercept	books	attend	bk_att_c
Intercept	5.1911316194	-0.262298276	0.0112966321	-0.222721801
books	-0.262298276	1.9148195666	-0.241627578	0.1111433374
attend	0.0112966321	-0.241627578	0.2929807494	-0.004786709
bk_att_c	-0.222721801	0.1111433374	-0.004786709	0.0943736443

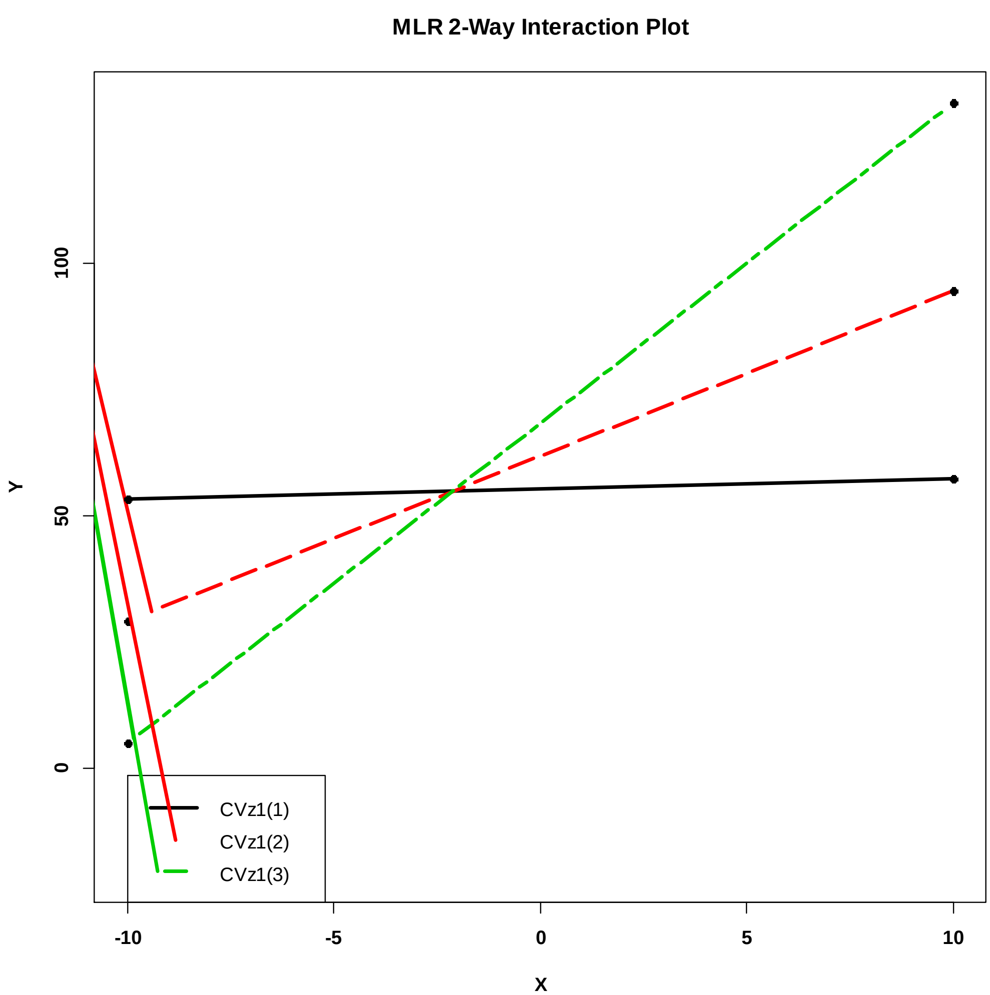
Regression Coefficients		Coefficient Variances		Conditional Values of Z	
$\hat{b}_0$	61.856	$\hat{b}_0$	5.191	cv_{Z1}	-4.28
$\hat{b}_1$	3.273	$\hat{b}_1$	1.915	cv_{Z2}	0
$\hat{b}_2$	1.526	$\hat{b}_2$	.293	cv_{Z3}	4.28
$\hat{b}_3$	.718	$\hat{b}_3$	.094	Points to Plot	
Other Information		Coefficient Covariances		x_1	-10
df	36	$\hat{b}_2, \hat{b}_0$	.011	x_2	10
α	.05	$\hat{b}_3, \hat{b}_1$	.111	Calculate	

When you click “Calculate,” three frames below will populate. Scroll down a bit in the first one, and you will get information about significance, which is pasted below. First, it tells us that:

One standard deviation below our mean (attendance = -4.28), the relationship between books and grade **is not significant**.
At our mean (attendance = 0), the relationship between books and grade **is significant**.
One standard deviation above our mean (attendance = 4.28), the relationship between books and grade **is significant**.

If we want to see what this looks like graphically, we can click **Submit above to Rweb** on the second frame down. The plot shown below is what comes up.

```
Simple Intercepts and Slopes at Conditional Values of Z
=======================================================
  At Z = cv1...
    simple intercept = 55.3247(3.2348), t=17.1028, p=0
    simple slope     = 0.2(1.6391), t=0.122, p=0.9036
  At Z = cv2...
    simple intercept = 61.856(2.2784), t=27.1492, p=0
    simple slope     = 3.273(1.3838), t=2.3652, p=0.0235
  At Z = cv3...
    simple intercept = 68.3873(3.2638), t=20.9532, p=0
    simple slope     = 6.346(2.1417), t=2.963, p=0.0054
```



If you scroll through the first frame, you'll see the boundaries for the region of significance, as well as the values of our simple slope at the boundaries of this region.

These boundaries correspond to the **values of attendance where the simple slope between books and grade becomes significant at the alpha = .05 level**.

So **when attendance is less than -28.6131 or more than -0.769, the relationship between books and grades is significant**. Note that, while -0.769 is right below the mean of attendance, -28.6131 is well below the observed minimum value in our data – it's so low that we didn't even plot it (since we plotted a range of -10 to 10). So for our purposes, what this means is that at values of attendance higher than just below the mean, the relationship between books and grades is significant.

Note: there are times – based on the region of our X variable that we choose to examine, and the points in our data where observations are more heavily concentrated – where the boundaries will refer to the area *inside of which* simple slopes are significant. In such a case, between the upper and lower bounds would be significant. So pay attention to this note, which tells you how to interpret the boundaries.

Finally, we can plot the region of significance by clicking **Submit code to RWeb** on the third frame down. Again, our plot does not even go as low as the lower bound (-28.6131), so it is not pictured.

```
Region of Significance
=======================================================
  Z at lower bound of region = -28.6131
  Z at upper bound of region = -0.769

  (simple slopes are significant *outside* this region.)
```

```
Simple Intercepts and Slopes at Region Boundaries
=======================================================
  Lower Bound...
    simple intercept = 18.1924(15.6347), t=1.1636, p=0.2522
    simple slope     = -17.2712(8.516), t=-2.0281, p=0.05
  Upper Bound...
    simple intercept = 60.6825(2.3124), t=26.2418, p=0
    simple slope     = 2.7209(1.3416), t=2.0281, p=0.05
```