INTERACTION

1 Two-Variable Interaction Model

$$Y_i = B_0 + B_1X_{i1} + B_2X_{i2} + B_3X_{i1}X_{i2} + \epsilon_i \tag{1}$$

X_1 and X_2 are said to be additive in their relationship to Y if the slope of Y with respect to X_1 does not depend on X_2 and vice versa; otherwise, they are said to interact.

The hypothesis test of $H_o : B_3 = 0$ is equivalent to a test of interaction between X_1 and X_2.

The meaning of the regression coefficients B_1 and B_2 here is not the same as that given earlier because of the interaction term $B_3X_{i1}X_{i2}$. The coefficients B_1 and B_2 no longer indicate the change in Y with a unit increase in the predictor variable, with the other predictor held constant at any given level. *The change in Y with a unit increase in X_1 given a value of X_2 is*:

$$B_1 + B_3X_2 \tag{2}$$

Or that the equation for the '*Simple Slope*' for X_1 is:

$$\hat{Y} = [B_1 + B_3X_2]X_1 + [B_2X_2 + B_0] \tag{3}$$

Similarly, the change in mean response (Y) with a unit increase in X_2 given a value of X_1 is:

$$B_2 + B_3X_1 \tag{4}$$

The equation for the '*Simple Slope*' of X_2 is:

$$\hat{Y} = [B_2 + B_3X_1]X_2 + [B_1X_1 + B_0] \tag{5}$$

What will the value of $X_1 = \frac{-B_2}{B_3}$ represent?

What will the Simple Slope of X_1 be when the value of X_2 is equal to zero? Similarly for the Simple Slope of X_2?

Why would it be valuable for the researcher to center X_1 and X_2?

Cohen and Cohen recommend that continuous variables always be centered. Reasons for centering:

- Interpretations are more straight forward for the main effects.
- Removes all nonessential collinearity.

Lets take a look at a model using SAS. First lets look at our X_1 and X_2 in order to get some values for ${X_2}_{LOW}$, ${X_2}_{MEAN}$, and ${X_2}_{HIGH}$ which we can use for our plots. Then lets run the model using **PROC REG**. Using PROC REG enables us to get estimates of the covariance matrix which will be useful later on.

2 Testing Simple Slopes, and obtaining CI's

Lets denote the covariance matrix of $B_1,\ B_2,\ and\ B_3$ by:

$$\begin{pmatrix} \sigma^2_{B_1} & \sigma^2_{B_1B_2} & \sigma^2_{B_1B_3} \\ \sigma^2_{B_2B_1} & \sigma^2_{B_2} & \sigma^2_{B_2B_3} \\ \sigma^2_{B_3B_1} & \sigma^2_{B_3B_2} & \sigma^2_{B_3} \end{pmatrix}$$

Since there is no readily available way to test simple slopes in SAS (there are available marcos on the web). Calculations will have to be done by hand. The first thing we would like to obtain is the standard error of the simple slope for X_1. We we can get from the formula:

$$\sigma_{SSX_1} = \sqrt{\sigma^2_{B_1} + 2X_2\sigma^2_{B_1B_3} + X_2^2\sigma^2_{B_3}} \tag{6}$$

Using σ_{SSX_1} and the t-distribution we can obtain the test statistic:

$$t_{X_1} = \frac{B_1 + B_3X_2}{\sigma_{SSX_1}}\ with\ n - k - 1\ df \tag{7}$$

Where k is the number of parameter (3 in our case). Furthermore the equation for t (1 - α)% confidence interval for the simple slope of X_1 at the point X_2 is:

$$(B_1 + B_3X_2) \pm t_{1-\frac{\alpha}{2},n-k-1} * \sigma_{SSX_1} \tag{8}$$

Graphing interactions

What we now want to do is obtain a graph of Y by X_1 for the three different values of X_2 $X_{2_{LOW}}$, $X_{2_{MEAN}}$, and $X_{2_{HIGH}}$ chosen earlier. This can be done using PROC GPLOT.

```
data generate;
x2low=1;
x2high=4;
x2mean=2.5;
x1min= 20;
x1max = 40;
range= x1max-x1min;
do x1 = x1min to x1max by range/5;
ylow = (b2 + b3*x2low)x1 + (b2*x2low + b0);
ymean = (b2 + b3*x2mean)x1 + (b2*x2mean + b0);
yhigh = (b2 + b3*x2high)x1 + (b2*x2high + b0);
output;
end;
run;
proc print;
run;

proc gplot data=generate;
* title1 Plot of Interaction;
symbol1 color=black interpol=join value=none;
```

```
symbol2 color=red interpol=join value=none;
symbol2 color=blue interpol=join value=none;
plot Ylow*X1 Ymean*X1 Yhigh*X1/ overlay;
run;
quit;
```

Does X_2 have a *Synergistic* effect on X_1, or does it act as a *Buffer*??

3 Three-way Interactions

The model for all three way interactions is:

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + B_3 X_{i3} + B_4 X_{i1} X_{i2} + B_5 X_{i1} X_{i3} + B_6 X_{i2} X_{i3} + B_7 X_{i1} X_{i2} X_{i3} + \epsilon_i \quad (9)$$

Note that to include higher order terms all lower order terms must be left in the model. The making of plots will be done in a similar fashion as above, but we will now have two, or three, separte graphs for multiple levels of our X_3 variable. The three-way interaction term test's if the effect of X_2 on Y as a function of X_1 will differ depending on the level of X_3.