Interpreting Interaction Effects; Interaction Effects and Centering

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised February 20, 2015

Models with interaction effects can be a little confusing to understand. The handout provides further discussion of how interaction terms should be interpreted and how centering *continuous* IVs (i.e. subtracting the mean from each case so the new mean is zero) doesn't actually change what a model means but can make results more interpretable.

Interaction Effects Without Centering. This problem is modified from Hamilton's Statistics with Stata 5 and uses data from a survey of undergraduate students collected by Ward and Ault (1990). DRINK is measured on a 33 point scale, where higher values indicate higher levels of drinking. In the sample the mean of Drink is about 19 and the observed scores range between 4 and 33. GPA is the student's Grade Point Average (higher values indicate better grades). The average gpa is about 2.81. The range of gpa theoretically goes from 0 to 4 but in actuality the lowest gpa in the sample is 1.45. MALE is coded 1 if the student is male, 0 if Female. MALEGPA = MALE * GPA. Here are the descriptive statistics:

```
. use https://www3.nd.edu/~rwilliam/statafiles/drinking.dta, clear
(Student survey (Ward 1990))
. sum male drink gpa malegpa

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        male |        243    .4485597    .4983734          0          1
       drink |        243      19.107    6.722117          4         33
         gpa |        218    2.808394    .4591705       1.45          4
     malegpa |        218    1.234679    1.390995          0       3.75
```

First, we regress drink on gpa and male.

MODEL I: DRINK REGRESSED ON GPA & MALE, WITHOUT CENTERING

```
. regress drink gpa i.male

      Source |       SS       df       MS              Number of obs =     218
-------------+------------------------------           F(  2,   215) =   18.36
       Model |  1437.71088     2  718.855442           Prob > F      =  0.0000
    Residual |  8416.31205   215  39.1456374           R-squared     =  0.1459
-------------+------------------------------           Adj R-squared =  0.1380
       Total |  9854.02294   217  45.4102439           Root MSE      =  6.2566

------------------------------------------------------------------------------
       drink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |    -3.4529   .9400734    -3.67   0.000     -5.30584    -1.59996
      1.male |   3.535818   .8649733     4.09   0.000     1.830904    5.240732
       _cons |   26.91249     2.7702     9.71   0.000     21.45226    32.37272
------------------------------------------------------------------------------
```

The model does not allow for the effects of GPA to differ by gender, but it does allow for a difference in the intercepts. Interpreting each of the regression coefficients,

* The constant term of 26.9 is the predicted drinking score for a female with a 0 gpa. No woman in the sample actually has a gpa this low. So, you can interpret this as the depths to which a woman would plunge if she was doing that badly.

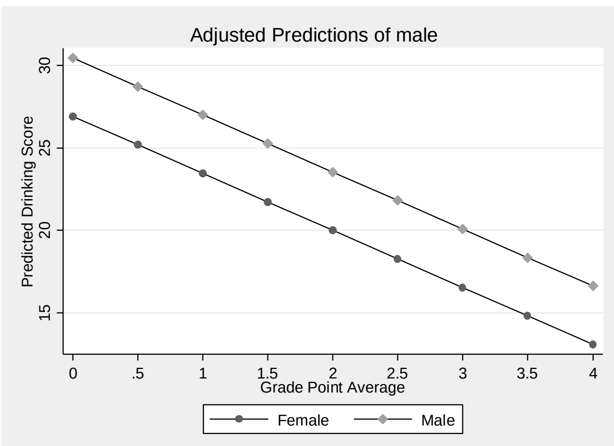
* For both men and women, each one unit increase in gpa results, on average, in a 3.4529 decrease in the drinking scale. That is, those with higher gpas tend to drink less.

* On average, men score 3.54 points higher on the drinking scale than do women with the same GPAs. As the following graph shows, the lines for men and women are parallel but the intercepts are different. Hence, with Model I, regardless of GPA, the predicted difference between a man and a woman with the same gpa is 3.54.

Here is a visual presentation of the results. [NOTE: The `scheme(sj)` option creates graphs that are formatted for publication in *The Stata Journal* and that are good for black and white printing.]

```
. quietly margins male, at(gpa=(0(.5)4))
. marginsplot, scheme(sj) noci ytitle(Predicted Drinking Score) name(intonly)

  Variables that uniquely identify margins: gpa male
```



Now see what happens once we add the interaction term.

MODEL II: DRINK REGRESSED ON GPA, MALE, MALEGPA, WITHOUT CENTERING

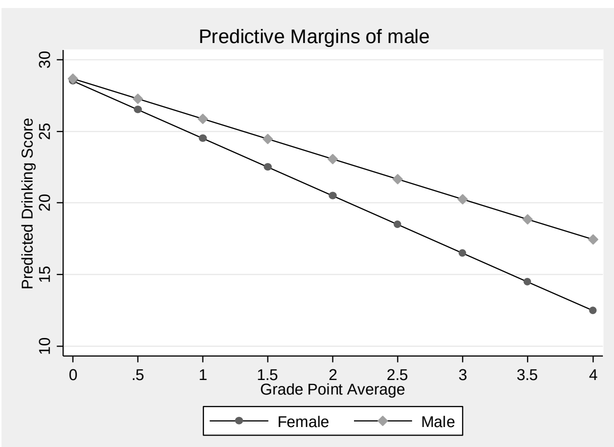
```
. regress drink gpa male i.male#c.gpa

      Source |       SS       df       MS              Number of obs =     218
-------------+------------------------------           F(  3,   214) =   12.35
       Model |  1453.87872     3  484.626241           Prob > F      =  0.0000
    Residual |  8400.14421   214  39.2530103           R-squared     =  0.1475
-------------+------------------------------           Adj R-squared =  0.1356
       Total |  9854.02294   217  45.4102439           Root MSE      =  6.2652

------------------------------------------------------------------------------
       drink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |  -4.011209   1.281774    -3.13   0.002    -6.537728   -1.484691
        male |    .148815    5.34808     0.03   0.978    -10.39285    10.69048
             |
  male#c.gpa |
          1  |   1.212068   1.888589     0.64   0.522    -2.510551    4.934686
             |
       _cons |   28.52206   3.739645     7.63   0.000     21.15081    35.89332
------------------------------------------------------------------------------

. quietly margins male, at(gpa=(0(.5)4))
. marginsplot, scheme(sj) noci ytitle(Predicted Drinking Score) name(intgpa)

  Variables that uniquely identify margins: gpa male
```



For convenience, we'll ignore the fact that the effect of malegpa is insignificant (otherwise I'd have to scrounge around for another example.) Note that

* The effects of gpa and malegpa show you that the effect of gpa is greater in magnitude for women than for men, i.e. higher gpas reduce the drinking of women more than they reduce the drinking of men. Hence, the male/female lines are no longer parallel. As a result, the difference between a man and a woman with the same gpa depends on what the gpa is. The higher the gpa, the greater the expected difference between a man and a woman is.

* The intercept is still the predicted drinking score for the non-existent lazy or idiotic woman with a gpa of 0. This number is actually slightly higher than it was in Model I, which reflects the fact that

the estimated effect of gpa on women is now greater since it is no longer being diluted by the weaker effect that gpa has on men.

* The coefficient for male in Model II, .148815, is much smaller than it was in Model I (3.535818). But, this is because it now has a different meaning. Before we added interaction effects, the male/female lines were parallel, and the predicted difference between a man and a woman with the same gpa was always 3.54 regardless of what the gpa actually was. Now, however, the coefficient for male is the predicted difference between a man and a woman who both have a 0 gpa. Since no such people exist, this isn't particularly interesting. I guess you could say that a man and a woman who were doing so poorly would both hit the bottle about as much. For a man and a woman who both have average gpas of about 2.81, the predicted difference is still about 3.5. (You can compute this from the Model II coefficients.) For a man and a woman with perfect gpas, the guy is predicted to score about 5 points higher on the drinking scale.

* Also, note that the coefficient for male in model II is not significant, whereas it was in Model I. But again, this reflects the fact that the coefficient has a different meaning now. In Model II, the coefficient for Male tests whether a man and woman who both have 0 gpas significantly differ in their drinking. The results show that they don't. But, at higher levels of gpa, the difference between men and women may be significant. In fact, we'll show that it is down below.

* The implication is that, once you add interaction effects, the main effects may or may not be particularly interesting, at least as they stand, and you should be careful in how you interpret them. For example, it would be wrong in this case to attach some profound meaning to the change in the effect of Male; the change just reflects the fact that the Male coefficient has different meanings in the two models. Likewise, the fact that Male becomes insignificant is not particularly interesting, because it is only testing the difference between men and women at a specific point, when gpa = 0. Once interaction terms are added, you are primarily interested in their significance, rather than the significance of the terms used to compute them.

Interaction Effects with Centering. If you want results that are a little more meaningful and easy to interpret, one approach is to *center* continuous IVs first (i.e. subtract the mean from each case), and then compute the interaction term and estimate the model. (Only center continuous variables though, i.e. you don't want to center categorical dummy variables like gender. Also, you only center IVs, not DVs.) Once we center GPA, a score of 0 on gpacentered means the person has average grades, i.e. a gpa of about 2.81. In SPSS, you would run descriptive statistics to determine the means of variables. In Stata, centering is more easily accomplished.

```
. sum gpa, meanonly
. gen gpacentered = gpa - r(mean)
(25 missing values generated)
. label variable gpacentered "Grade Point Average Centered"
```

First, we'll estimate the model without the interaction term.

MODEL III: DRINK REGRESSED ON GPA & MALE, WITH CENTERING

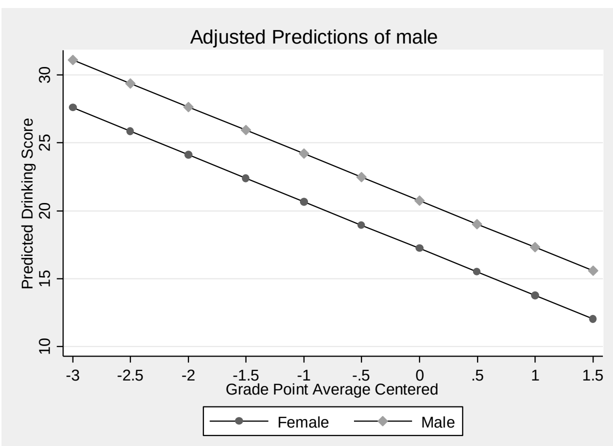
```
. regress drink gpacentered i.male

      Source |       SS       df       MS              Number of obs =     218
-------------+------------------------------           F(  2,   215) =   18.36
       Model |  1437.71088     2  718.855441           Prob > F      =  0.0000
    Residual |  8416.31205   215  39.1456375           R-squared     =  0.1459
-------------+------------------------------           Adj R-squared =  0.1380
       Total |  9854.02294   217  45.4102439           Root MSE      =  6.2566

------------------------------------------------------------------------------
       drink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 gpacentered |    -3.4529   .9400734    -3.67   0.000     -5.30584    -1.59996
      1.male |   3.535818   .8649733     4.09   0.000     1.830904    5.240732
       _cons |   17.21539   .5778114    29.79   0.000     16.07648    18.35429
------------------------------------------------------------------------------

. quietly margins male, at(gpacentered=(-3(.5)1.5))
. marginsplot, scheme(sj) noci ytitle(Predicted Drinking Score) name(intonlycntr)

  Variables that uniquely identify margins: gpacentered male
```



Note that everything is pretty much the same as before we centered (in Model I), except the intercept has changed. In Model I, the intercept of 26.9 was the predicted score of the nonexistent destitute woman who was failing everything (no wonder she drinks so much). In Model III with gpa centered, the intercept (17.215) is the predicted drinking score of a woman with average grades. A score of 0 on gpa corresponds to a score of about -2.81 on gpacentered, so it is still the case that a woman with 0 gpa would have a predicted drinking score of 26.9. Hence, centering doesn’t change what the model predicts, but it changes the interpretation of the intercept.

Now, we’ll see what happens when we add the interaction:

MODEL IV: DRINK REGRESSED ON GPA, MALE, MALEGPA, WITH CENTERING

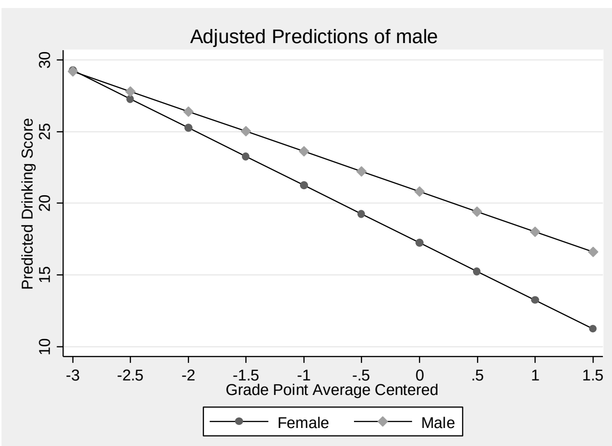
```
. regress drink gpacentered i.male i.male#c.gpacentered

      Source |       SS       df       MS              Number of obs =     218
-------------+------------------------------           F(  3,   214) =   12.35
       Model |  1453.87872     3   484.62624           Prob > F      =  0.0000
    Residual |  8400.14422   214  39.2530104           R-squared     =  0.1475
-------------+------------------------------           Adj R-squared =  0.1356
       Total |  9854.02294   217  45.4102439           Root MSE      =  6.2652

------------------------------------------------------------------------------------
             drink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
       gpacentered |  -4.011209   1.281774    -3.13   0.002    -6.537728   -1.484691
            1.male |   3.552779   .8665619     4.10   0.000     1.844689    5.260869
                   |
male#c.gpacentered |
                1  |   1.212068   1.888589     0.64   0.522    -2.510551    4.934686
                   |
             _cons |   17.25701   .5822263    29.64   0.000     16.10937    18.40464
------------------------------------------------------------------------------------

. quietly margins male, at(gpacentered=(-3(.5)1.5))
. marginsplot, scheme(sj) noci ytitle(Predicted Drinking Score) name(intgpacntr)

  Variables that uniquely identify margins: gpacentered male
```



Note that

* Except for Male and the Constant, the various model terms in Model IV are the same as before we centered in Model II. Likewise, the plot is the same, except everything has been shifted to the left because we centered gpa. (If we used the uncentered GPA, the plots would be identical.) Centering does not change the substantive meaning of the model or the predictions that are made; but it may make the results more easily interpretable.

* The intercept in Model IV, 17.26, now reflects the average drinking score for a woman with an average gpa, rather than the predicted score for the non-existent drunkard who has failed everything. Since such a person (or somebody close to her) actually does exist, the intercept is more meaningful than it was in Model II.

* The coefficient for male (3.55) is now the average difference between a male with an <u>average</u> gpa and a female with an <u>average</u> gpa. This is probably more meaningful than looking at the difference between the nonexistent man and woman who are flunking everything.

* With gpa centered, adding the interaction term produces much less change in the estimated effect of male between Models III and IV than it did when gpa was not centered (Model I versus Model II). At least in this case, this is because the predicted difference between the average man and woman is about the same regardless of whether the model includes interaction terms or not, whereas the predicted difference between a man and a woman who are failing everything changes quite a bit once you add the interaction term. Further, the Model IV difference in drinking between the average man and the average woman is statistically significant, even though the Model II difference between the 0 gpa man and woman is not.

Other Issues and Options to Be Aware of

- If you do center, be consistent throughout, i.e. different sample selections could produce different means, so comparing results produced by different centerings could be deceptive.

- You don't have to use the mean when centering; you could use any value that was of substantive interest.

    - For example, if you were particularly interested in comparing male and female C students to each other, you could subtract 2.0 from each gpa. Then, a score of 0 on gpacentered would correspond to a C gpa. The intercept would be the predicted drinking score for a C woman, and the male coefficient would be the predicted difference between C men and C women.

    - Or, in our Income/Education example, you could subtract 12 from education so that a score of 0 on centered education corresponded to a high school degree. You would then modify your interpretations accordingly, i.e. the main effect of Income would be the effect of Income for people who had a high school degree, the main effect of Education would still be the effect of education for a person with average income, and the intercept would be the predicted Y score for a person with average income and a high school degree.

    - Basically, the key is to have a score of 0 on the IV correspond to something that is substantively interesting, rather than have it be a value that could not (or at least does not) actually occur in the data.

Conclusions

- You don't have to center continuous IVs in a model with interaction terms. It won't actually change what the model means or what it predicts. But, centering continuous IVs and/or presenting plots may make your coefficients more interpretable.

- If you don't center, don't get hung up looking at changes in the main effects of the variables used to compute the interactions. These are to be expected, because the meaning of these terms changes once you add the interaction terms.

- Also, don't be concerned if the main effect of the dummy is insignificant once you've added the interaction; this just means that, when the IVs = 0, the difference between groups is insignificant, but it may be significant when the IV does not = 0. Once interaction effects are added, the more critical thing is the significance of the interaction terms, not the terms that were used to compute the interactions. Whether you center or not, the interaction terms will stay the same.

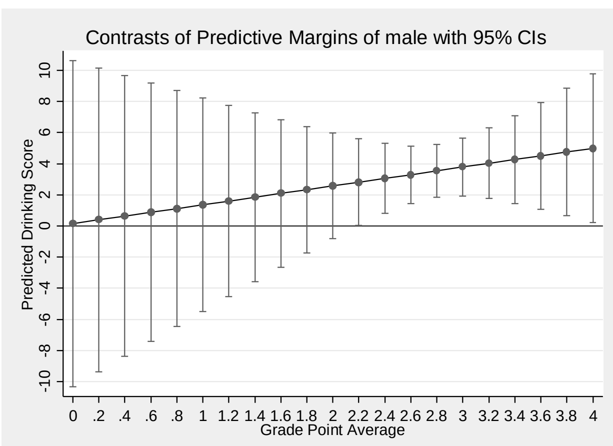
Appendix: Marginal Effects and Confidence Intervals

Here is another approach that may be useful. Rather than plot separate lines for men and women, we can plot a single line that shows the difference between the predicted values for each gender. This is known as the *marginal effect* of gender. In an OLS regression analysis the Marginal Effect for a categorical variable shows how E(Y) changes as the categorical variable changes from 0 to 1, after controlling in some way for any other variables in the model. With a dichotomous independent variable, the ME is the difference in the adjusted predictions for the two groups, in this case men and women.

Also, I have been excluding confidence intervals in my graphs; but if you include them, you can easily see whether and when the differences in adjusted predictions are statistically significant, i.e. if the confidence interval for the marginal effect includes 0 the difference is not statistically significant, otherwise it is. Here is an example:

```
. quietly regress drink gpa male i.male#c.gpa
. quietly margins r.male, at(gpa=(0(.2)4))
. marginsplot, scheme(sj) ytitle(Predicted Drinking Score)  ///
>         yline(0) ylabel(#10) xlabel(#20) name(margeffect)

  Variables that uniquely identify margins: gpa
```



As before, this shows us that the predicted difference between a man and a woman who both have a gpa of 0 is almost zero. For a man and a woman with average gpa (2.81) the predicted difference is about 3.5; and for a 4.0 gpa the predicted difference is around 5. The confidence intervals, however, reveal that the predicted differences are not statistically significant until gpa is about 2.2 or greater.