Constructing, Interpreting and Presenting Interactions

Hsueh-Sheng Wu

CFDR Workshop Series

October 23, 2017

BGSU

Center for Family and Demographic Research

Outline

- What is interaction?
- Differences among control variables, moderators, and mediators
- Construct and interpret Interaction
- Examples
- Conclusions

What is Interaction?

- A situation where the simultaneous influences of two variables on a third is not additive
- A example data set for two-way interactions

Table 1. A sample data set

Situation	I-Fen's class (X1)	HS's class (X2)	ScoreA	ScoreB	ScoreC
1	0	0	0	0	0
2	1	0	6	6	6
3	0	1	2	2	2
4	1	1	8	10	4

[1] A respondent is exposed to four experiments in terms of taking I-Fen's and/or HS's classes.

[2] A value of 1 indicates taking either I-Fen's or HS's class, 0 otherwise.

[3] The value of ScoreA through ScoreC indicates how much respondents understand interaction.

What is Interaction? (Cont.)

- A two-way interaction model for X1 and X2 on Y

Interaction		
	X1	
X2	No	Yes
No	A	B
Yes	C	D

1. A,B,C,and D represent the value of a dependent variable

- If the relation between A and B is different from that of C and D, it indicates the presence of an interaction between X1 and X2 because the relation between X1 and Y differs by the level of X2. Similarly, If the relation between A and C is different from that of B and D, it indicates that the relation between X2 and Y differs by the level of X1.

What is Interaction? (Cont.)

- Two-way Interaction Model for Score A

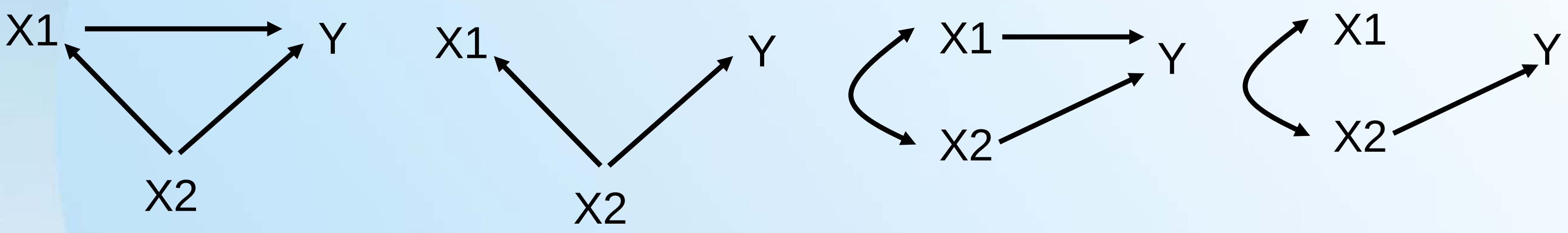
Interaction Model for Score A				Interaction Model for Score B				Interaction C		
	I-Fen (X1)				I-Fen (X1)				I-Fen (X1)	
HS (X2)	No	Yes		HS (X2)	No	Yes		HS (X2)	No	Yes
No	0	6		No	0	6		No	0	6
Yes	2	8		Yes	2	10		Yes	2	4

- Interaction may enhance, reduce, or have no impact on the relations between two variables.
- Two-way interactions looks at whether the relations between two variables differ, depending on the level of a third variable. You can also test whether the relations differ, depending on the presence of more than one variable, which means testing higher-order interactions.
- Looking at the numbers across cells in a table helps understanding what interaction conceptually means, but does not test it statistically.

Differences among Control Variables, Moderators, and Mediators

Control Variables

- Research question: What is the effect of X1 on Y, net of the effect of X2 on Y



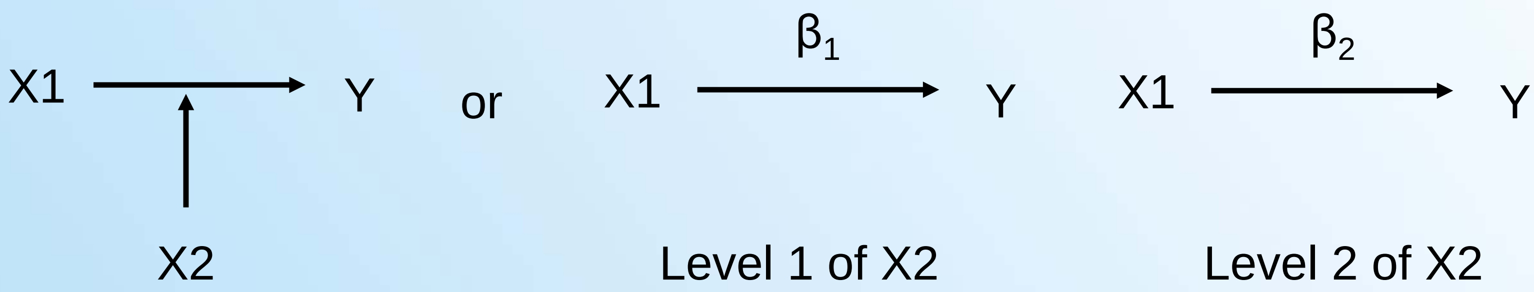
- Test: You include X2 in regression and see if X1 remains a significant predictor of Y

BGSU
Center for Family and Demographic Research

Differences among Control Variables, Moderators, and Mediators (Cont.)

Moderators (Interaction)

- Research question: If the effect of X1 on Y differs for different levels of X2, i.e., β_1 is different from β_2.

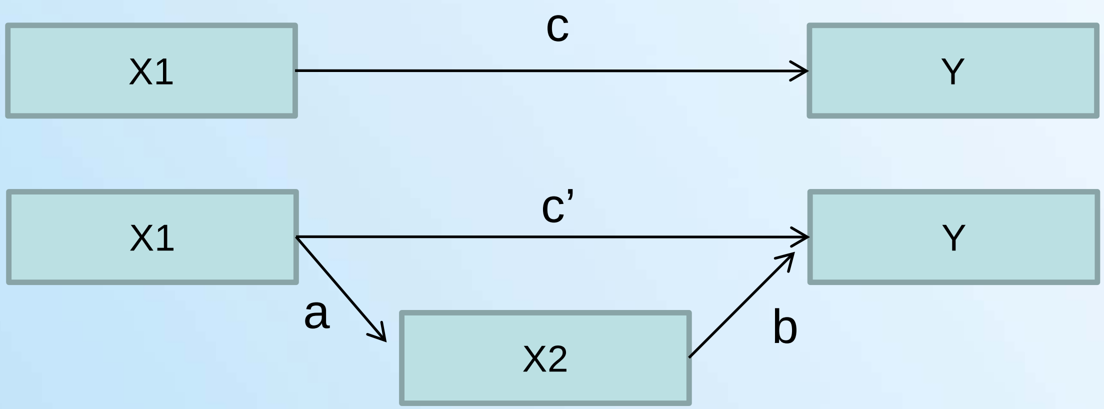


- Test: You include an interaction term (e.g., a product of two variable for a two-way interaction or a product of three variables for a three-way interaction) in regression.

Differences among Control Variables, Moderators, and Mediators (Cont.)

Mediators

- Research question: If X1 exerts its effect on Y via its effect on X2.



  - Logic: If X2 mediates the X1-Y relation, then the following conditions hold
    - X1 predict Y, X1 predict X2, and X2 predicts Y
  - When Y are predicted by both X1 and X2:
    - The regression coefficient of X2 (i.e., b) should be significant
    - X1 should predict X2
    - The regression coefficient of X1 differs before and after X2 is in the regression (i.e., c’ is different from c).

Differences among Control Variables, Moderators, and Mediators (Cont.)

Steps of Testing Mediation

1. Test if X1 predicts Y

$$Y = B_1 + cX1 + \varepsilon_1$$

2. Test if X1 predicts X2

$$X2 = B_2 + aX1 + \varepsilon_2$$

3. Test if X1 still predicts Y when X2 is in the model

$$Y = B_3 + c'X1 + bX2 + \varepsilon_3$$

Differences among Control Variables, Moderators, and Mediators (Cont.)

Decision Rules

- X2 completely mediates the X1-Y relation if all three conditions are met:
  (1) X1 predicts Y
  (2) X1 predicts X2
  (3) X1 no longer predicts Y, but X2 does when both X1 and X2 are used to predict Y

- X2 partially mediates the X1-Y relation if all three conditions are met:
  (1) X1 predicts Y
  (2) X1 predicts X2
  (3) Both X1 and X2 predict Y, but X1 have a smaller regression coefficient after X2 was included in the model.

- X2 does not mediate the X1-Y relation if any of the condition holds:
  (1) X1 does not predict Y
  (2) X1 does not predict X2, or
  (3) The regression coefficient of X1 remain the same before and after X2 is used to predict Y

BGSU

Center for Family and Demographic Research

Construct and Interpret Interaction

Regressions without an interaction term

$$Y = B_0 + B_1 X1 + B_2 X2 + \varepsilon$$

B_0 means the average level of Y when X1 and X2 are both 0

B_1 means the predicted change in Y when X2 is 0 and X1 increases by one unit

B_2 means the predicted change in Y when X1 is 0 and X2 increases by one unit

BGSU

Center for Family and Demographic Research

Construct and Interpret Interaction (Cont.)

Regressions with an two-way interaction term

$$Y = B_3 + B_4 X1 + B_5 X2 + B_6 X1 * X2 + \varepsilon$$

B_3 means the change in Y when X1 and X2 are both 0

B_4 means the change in Y when X2 is 0 and X1 increases by one unit

B_5 means the change in Y when X1 is 0 and X2 increases by one unit

B_6 means the additional change in Y when the product of X1 and X2 increases by one unit

BGSU Center for Family and Demographic Research

Construct and Interpret Interaction (Cont.)

Regressions without an Interaction term

$$Y = B_0 + B_1 X1 + B_2 X2 + B_3 X3 + \varepsilon$$

B_0 means the value of Y when X1, X2, X3 are all 0

B_1 means the predicted change in Y when X2 and X3 are 0 and X1 increases by one unit

B_2 means the predicted change in Y when X1 and X3 are 0 and X2 increases by one unit

B_3 means the predicted change in Y when X1 and X2 are 0 and X3 increases by one unit

Construct and Interpret Interaction (Cont.)

Regressions without an interaction term

$$Y = B_0 + B_1 X1 + B_2 X2 + B_3 X3 +$$
$$B_4 X1 * X2 + B_5 X1 * X3 + B_6 X2 * X3 +$$
$$B_7 X1 * X2 * X3 + \varepsilon$$

B_0 means the value of Y when X1, X2, X3 are all 0

B_1 means the predicted change in Y when X2 and X3 are 0 and X1 increases by one unit

B_2 means the predicted change in Y when X1 and X3 are 0 and X2 increases by one unit

B_3 means the predicted change in Y when X1 and X2 are 0 and X3 increases by one unit

BGSU

Center for Family and Demographic Research

Construct and Interpret Interaction (Cont.)

B_4 means the additional change in Y when X3 = 0 and the product of X1 and X2 increases by one unit

B_5 means the additional change in Y when X2 = 0 and the product of X1 and X3 increases by one unit

B_6 means the additional change in Y when X1 = 0 and the product of X2 and X3 increases by one unit

B_7 means the additional change in Y when the product of X1, X2, and X3 increases by one unit

BGSU Center for Family and Demographic Research

Examples

Description of the data

Description of a Sample Data						
Variable Name	Variable Label	Obs	Mean	Std. Dev.	Min	Max
age	age	1,500	39.91333	11.5127	20	60
female	female	1,500	0.506	0.500131	0	1
education	education	1,500	1.816667	0.76819	1	3
stress	level of stress	1,500	57.1908	178.9915	0.25	748.8927
female_age	Interaction betv	1,500	22.04	23.16351	0	60
creativity	level of creativi	1,500	382.1054	895.4859	47.51422	3899.28
aggression	Being Aggressiv	1,500	0.25	0.433157	0	1

BGSU
Center for Family and Demographic Research

Examples

Ordinary Least Square (OLS) Regression for a continuous dependent variable

. regress creativity female age female_age i.education stress

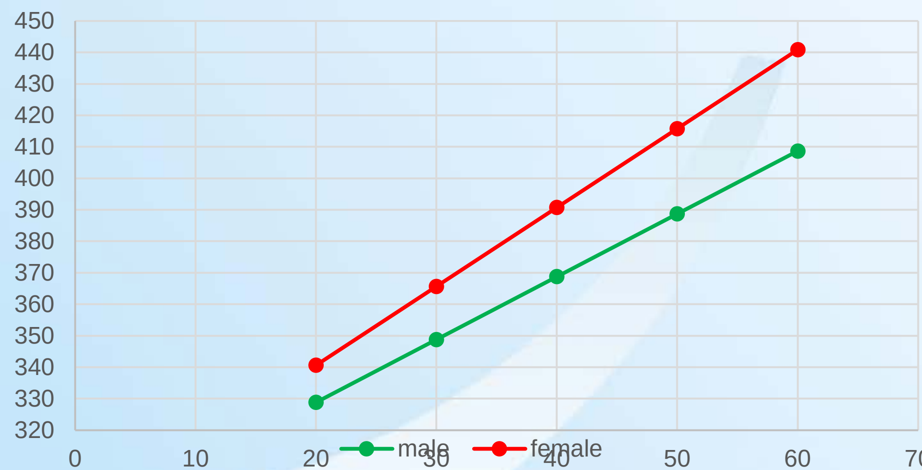
```
. regress creativity female age female_age       i.education stress

      Source |       SS           df       MS      Number of obs   =     1,500
-------------+----------------------------------   F(6, 1493)      >  99999.00
       Model |  1.2020e+09         6   200339844   Prob > F        =    0.0000
    Residual |   1454.5023     1,493  .974214537   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total |  1.2020e+09     1,499  801894.944   Root MSE        =    .98702

------------------------------------------------------------------------------
  creativity |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   1.601123    .200691     7.98   0.000     1.207456    1.994789
         age |   1.996848   .0036687   544.30   0.000     1.989652    2.004044
  female_age |    .509618   .0047726   106.78   0.000     .5002562    .5189798
             |
   education |
          2  |   1.054777   .0666471    15.83   0.000     .9240452    1.185509
          3  |   5.031624   .0898828    55.98   0.000     4.855314    5.207934
             |
      stress |   4.999865   .0001425  3.5e+04   0.000     4.999585    5.000144
       _cons |   2.911853   .1657604    17.57   0.000     2.586705    3.237001
------------------------------------------------------------------------------
```

BGSU
Center for Family and Demographic Research

female	Age	female#c.age	education	stress	female	age	female#c.age	Education_1	Education_2	Education_3	stress	_cons	Predicted Creativity
0	20	0	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	328.7950922
0	30	0	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	348.7635722
0	40	0	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	368.7320522
0	50	0	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	388.7005322
0	60	0	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	408.6690122
1	20	20	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	340.5885752
1	30	30	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	365.6532352
1	40	40	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	390.7178952
1	50	50	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	415.7825552
1	60	60	1	57.1908	1.601123	1.996848	0.509618	0	1.055	5.032	4.9999	2.911853	440.8472152

Age	male	female
20	328.7951	340.5886
30	348.7636	365.6532
40	368.7321	390.7179
50	388.7005	415.7826
60	408.669	440.8472

Predicted Plot from OLS regression



Examples

Logistic Regression for a dichotomous dependent variables

. logit aggression female age female age

```
. logit    aggression female age female_age

Iteration 0:   log likelihood = -843.50272
Iteration 1:   log likelihood = -629.73498
Iteration 2:   log likelihood = -615.49455
Iteration 3:   log likelihood = -615.34985
Iteration 4:   log likelihood =  -615.3498

Logistic regression                             Number of obs     =      1,500
                                                LR chi2(3)        =     456.31
                                                Prob > chi2       =     0.0000
Log likelihood =  -615.3498                     Pseudo R2         =     0.2705

------------------------------------------------------------------------------
  aggression |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -2.923804   .7150147    -4.09   0.000    -4.325207   -1.522401
         age |   .0508773   .0112015     4.54   0.000     .0289228    .0728318
  female_age |   .0928026   .0158041     5.87   0.000     .0618272     .123778
       _cons |  -4.084649   .4724848    -8.65   0.000    -5.010702   -3.158596
------------------------------------------------------------------------------
```

BGSU
Center for Family and Demographic Research

Examples

Logistic Regression for a dichotomous dependent variables

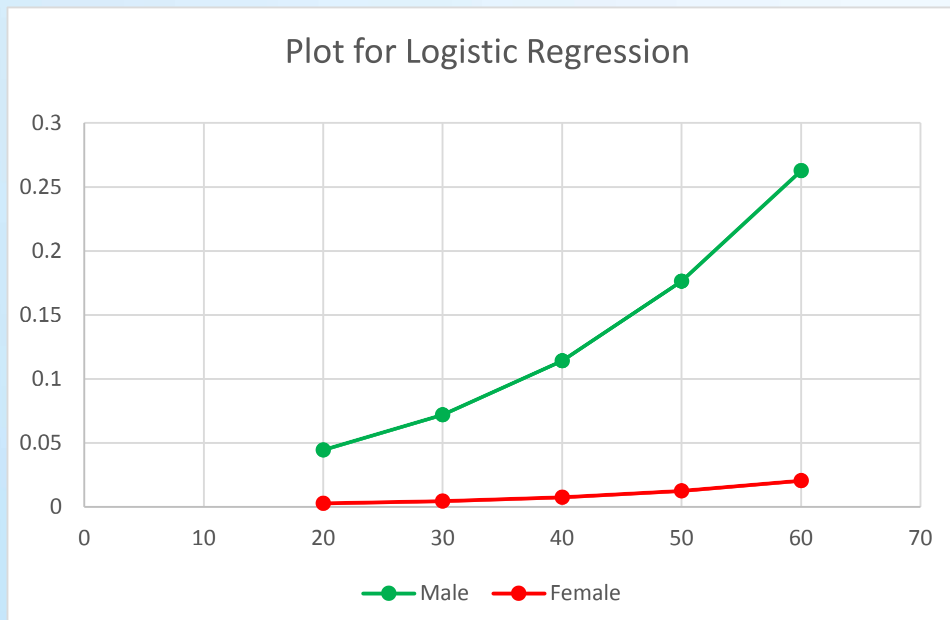
. logistic aggression female age female_age

```
. logistic aggression female age female_age

Logistic regression                             Number of obs     =      1,500
                                                LR chi2(3)        =     456.31
                                                Prob > chi2       =     0.0000
Log likelihood =  -615.3498                     Pseudo R2         =     0.2705

------------------------------------------------------------------------------
  aggression | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .0537289    .038417    -4.09   0.000     .0132308    .2181875
         age |   1.052194   .0117861     4.54   0.000     1.029345     1.07555
  female_age |   1.097245   .0173409     5.87   0.000     1.063779    1.131765
       _cons |    .016829   .0079515    -8.65   0.000     .0066662    .0424854
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

BGSU
Center for Family and Demographic Research

female	Age	female#c.age	female	age	female#c.age	_cons	Odds	Probability
0	20	0	-2.9238	0.050877	0.092803	-4.08465	0.046556	0.044484805
0	30	0	-2.9238	0.050877	0.092803	-4.08465	0.077434	0.071868858
0	40	0	-2.9238	0.050877	0.092803	-4.08465	0.128792	0.114097152
0	50	0	-2.9238	0.050877	0.092803	-4.08465	0.214213	0.176421333
0	60	0	-2.9238	0.050877	0.092803	-4.08465	0.35629	0.262694416
1	20	1	-2.9238	0.050877	0.092803	-4.08465	0.002745	0.00273713
1	30	1	-2.9238	0.050877	0.092803	-4.08465	0.004565	0.00454428
1	40	1	-2.9238	0.050877	0.092803	-4.08465	0.007593	0.007535557
1	50	1	-2.9238	0.050877	0.092803	-4.08465	0.012629	0.012471179
1	60	1	-2.9238	0.050877	0.092803	-4.08465	0.021005	0.020572512

Age	Male	Female
20	0.044485	0.002737
30	0.071869	0.004544
40	0.114097	0.007536
50	0.176421	0.012471
60	0.262694	0.020573



BGSU
Center for Family and Demographic Research

Conclusions

- Interaction is one of the most commonly used model for researchers to examine how variable may moderate the relation between other variables. Examining interaction is different from examining control or mediator variables.
- When analyzing a continuous dependent variable, it is easier to identify how interaction influences the level of the dependent variable. This is not the case for a dichotomous dependent variables, and usually you need to transform the results into predicted probability.
- Stata has two useful commands, margins and marginslpot, that help quickly generate and plot predicted values or probability for predictors in OLS or Logistic Regression. You may want to check out these two commands.
- Although regression is commonly used to test interaction between two variables, it is unclear which variable is the moderator and which one is being moderated. A better approach for testing interaction is to use Structural Equation Modeling because it directly tests if the regression coefficient between two different variables vary, depending on the value of a third variable.
- If you run into problems testing interaction, please contact me (wuh@bgsu.edu) or stop by my office during the office hours.