Analyzing, Interpreting and Presenting Interactions

Hsueh-Sheng Wu

CFDR Workshop Series

September 14, 2020

BGSU
Center for Family and Demographic Research

Outline

- What is interaction?
- Differences among control (confounding) variables, mediating variables, and moderating variables
- Construct and interpret interaction
- Use of the -margins- command in Stata
- Examples
- Conclusions

What Is Interaction?

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

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 Scores A, B, and C

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

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

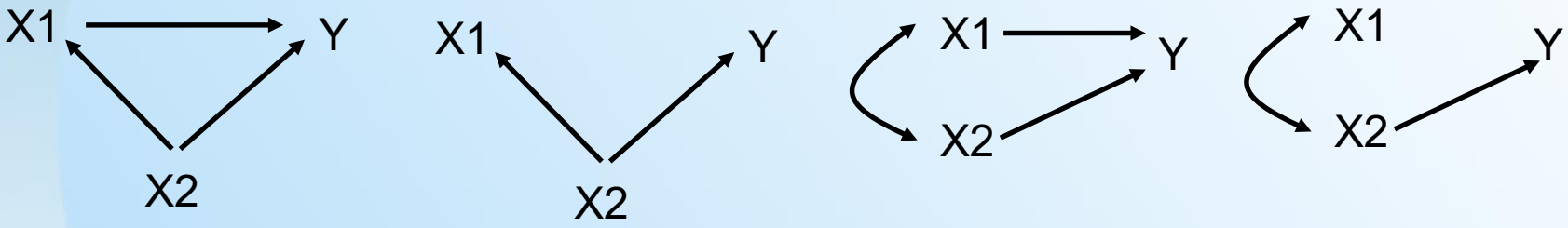
Interaction C		
	I-Fen (X1)	
HS (X2)	No	Yes
No	0	6
Yes	2	4

- Interaction may enhance, reduce, or have no impact on the relations between two variables.
- Two-way interactions looks at whether the relation between two variables differ, depending on the level of a third variable. You can also test whether the relation 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 understand what interaction conceptually means but does not test it statistically.

Differences among Control Variables, Mediating Variables, and Moderating Variables

Control (Confounding) Variables

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

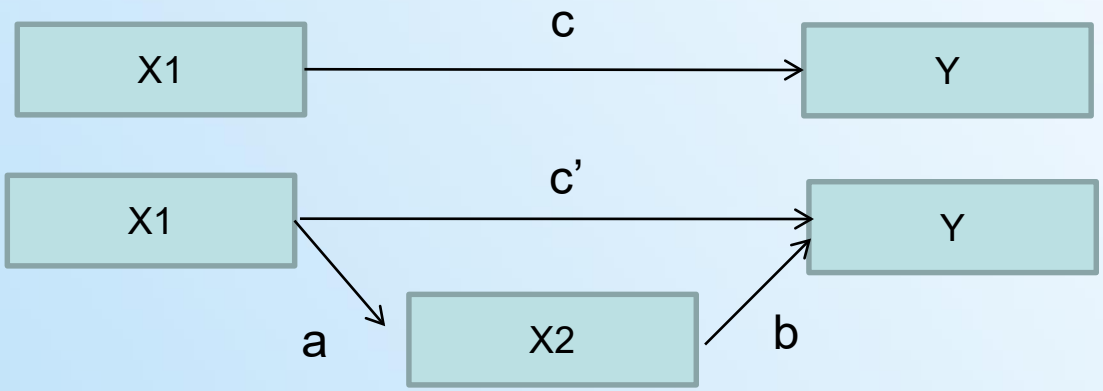


- Test: Include X2 in regression and see if X1 remains a significant predictor of Y

Differences among Control Variables, Mediating Variables, and Moderating Variables (Cont.)

Mediating variables

- Research question: Does X1 exert its influence on Y via its effect on X2?

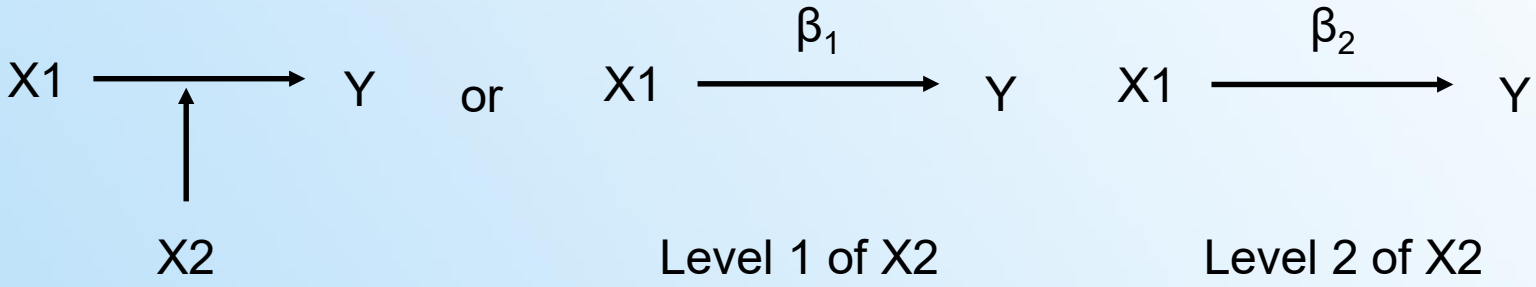


  - Logic: If X2 mediates the X1-Y relation, then the following conditions hold
    - X1 predicts Y, X1 predicts X2, and X2 predicts Y
  - When Y are predicted by both X1 and X2:
    - The regression coefficient of X2 (i.e., b) should be significant
    - X1 should predict X2
    - The regression coefficient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

Differences among Control Variables, Moderating Variables, and Mediating Variables (Cont.)

Moderating variables (Interaction)

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



- Test: 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, Moderating Variables, and Mediating Variables (Cont.)

Steps of Testing the Mediator

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

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

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

Construct and Interpret Interaction (Cont.)

Regressions with a 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

These four coefficients tells the expected values of Y in situations 1 through 4 on Slide #4.

Construct and Interpret Interaction (Cont.)

Regressions with a three-way 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

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

Use of the -margins- Command in Stata

- -margins- command is a post-estimation technique that generates predicted margins and estimates marginal effects, using estimated coefficients and estimated variance of the residual from the previously estimated model
- -margins- command is especially useful with the analysis involved categorical dependent variable, the squared term of a predictor, or interaction of predictors
- -margins- command works after most Stata estimation commands except those that use the alternative-specific estimation commands (generalized method of moments estimation) that organize data differently (e.g., alternative-specific conditional logit, nested logit regression) or do not produce full variance matrices (e.g., exact logistic regression or exact poisson regression)
- -margins- command can be applied to the data that are collected with the complex survey design or from multiple imputation

Factor Variables in the Analysis Command

- Why is it necessary to use factor variable in the analysis command?
  - It provides the information on the attribute of a predictor and allows the -margins- command treats categorical predictors and continuous predictors differently
  - It allows Stata to recognize whether a predictor is a function of other variables, for example, an interaction term is the product of two variables

- How to specify factor variables?
  - i for indicator variables, binary variables, dummies
  - c for continuous variables
  - # for the squared term of a predictor or the interaction term between two predictors
  - ## for both main effects and interaction of the predictors in the analysis model

Steps of Using the -margins- Command

- The analysis part:
  - Decide the research question which dictates the predictor(s), the outcome, and control variables
  - Specify and estimate the analysis model
- The -margins- part:
  - Consider how each control variable should be adjusted, whether the predicted margins or marginal effect will be estimated, and whether plotting the margins is needed
  - Specify the -margins- command that uses information from the analysis part to generate, test, or plot the predicted margin of the responses

Syntax of the -margins- Command

Table 1. Basic Syntax of the -Margins- Command

Command	The predictor of interest	predicted margins or the marginal effect	Adjustment of other variables
margins	[marginlist]	, [blank]	[blank]
			atmeans
			at(atspec)
		, dydx(varlist)	[blank]
			atmeans
			at(atspec)

Note:

1. These six combinations correspond to what researchers called (1) Average Adjusted Predictions, (2) Adjusted predictions at the Means; (3) Adjusted Predictions at Representative values; (4) Average Marginal Effects; (5) Marginal Effects at the Means; (6) Marginal Effects at Representative values.
2. Only factor variables and their interactions are allowed in the marginlist.
3. The pound sign"#" in the marginlist means the combinations of two predictors even when the analysis does not have the interaction term of these two variables in the data.

Stata Commands File

The Stata command file has five sets of commands.

- Three sets show how the -margins- command is used for a continuous dependent variable, a binary dependent variable, a nominal dependent variable.
- One set shows how to plot the results from the -margins- command
- The last set shows how the -margins- command helps interpret the three-way interactions.

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 (confounding) or mediating 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 binary or categorical dependent variables because the predicted probability of such variables depends on the level of the predictor. Consequently, it is better to transform the results of analyzing such variables into predicted probability.
- Stata has two useful commands, -margins- and –marginslpot-, that help quickly generate and plot predicted values or probability for predictors in various Regression analyses.
- Although regression is commonly used to test interaction between two variables, it is unclear for the interaction term, 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 variables differs, depending on the value of a third variable.
- If you run into problems testing interaction, please contact me at wuh@bgsu.edu