The Analysis, Interpretation, and Presentation of Interaction Effects

Hsueh-Sheng Wu
CFDR Workshop Series
October 23, 2023
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 an outcome is more than the sum of main effects of these two variables

• An example data set for two-way interaction:

<table>
<thead>
<tr>
<th>Situation</th>
<th>I-Fen's class (X1)</th>
<th>HS's class (X2)</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

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 values of the three scenarios indicates different types of relations between these two independent variables.
What Is Interaction? (Cont.)

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

<table>
<thead>
<tr>
<th>Interaction</th>
<th>X1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X2</td>
</tr>
<tr>
<td>No</td>
<td>A</td>
</tr>
<tr>
<td>Yes</td>
<td>C</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Interaction Model for Scenario 1</th>
<th>Interaction Model for Scenario 2</th>
<th>Interaction Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-Fen (X1)</td>
<td>I-Fen (X1)</td>
<td>I-Fen (X1)</td>
</tr>
<tr>
<td>HS (X2)</td>
<td>HS (X2)</td>
<td>HS (X2)</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

- Interaction may enhance (e.g., 1+1>2), reduce (e.g., 1+1<2), or have no impact (e.g., 1+1 =2) 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 $X_1$ on $Y$, net of the effect of $X_2$ on $Y$?

- Test: Include $X_2$ in regression and see if $X_1$ remains a significant predictor of $Y$
Differences among Control Variables, Mediating Variables, and Moderating Variables (Cont.)

**Mediating variables**

- **Research question:** Does $X_1$ exert its influence on $Y$ via its effect on $X_2$?

  - **Logic:** If $X_2$ mediates the $X_1$-$Y$ relation, then the following conditions hold
    - $X_1$ predicts $Y$, $X_1$ predicts $X_2$, and $X_2$ predicts $Y$
  - When $Y$ are predicted by both $X_1$ and $X_2$:
    - The regression coefficient of $X_2$ (i.e., $b$) should be significant
    - $X_1$ should predict $X_2$ (i.e., $a$)
    - The regression coefficient of $X_1$ differs before and after $X_2$ 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 $X_1$ on $Y$ differ for different levels of $X_2$, i.e., $\beta_1$ is different from $\beta_2$?

  - 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.
Steps of Testing the Mediator

1. Test if $X_1$ predicts $Y$

   \[ Y = B_1 + cX_1 + \varepsilon_1 \]

2. Test if $X_1$ predicts $X_2$

   \[ X_2 = B_2 + aX_1 + \varepsilon_2 \]

3. Test if $X_1$ still predicts $Y$ when $X_2$ is in the model

   \[ Y = B_3 + c'X_1 + bX_2 + \varepsilon_3 \]
Construct and Interpret Interaction

Regressions without an interaction term

\[ Y = B_0 + B_1 X_1 + B_2 X_2 + \varepsilon \]

- \( B_0 \) means the average level of \( Y \) when \( X_1 \) and \( X_2 \) are both 0
- \( B_1 \) means the predicted change in \( Y \) when \( X_2 \) is 0 and \( X_1 \) increases by one unit
- \( B_2 \) means the predicted change in \( Y \) when \( X_1 \) is 0 and \( X_2 \) increases by one unit
Regressions with a two-way interaction term

\[ Y = B_3 + B_4 X_1 + B_5 X_2 + B_6 X_1 \times X_2 + \varepsilon \]

- \( B_3 \) means the change in \( Y \) when \( X_1 \) and \( X_2 \) are both 0
- \( B_4 \) means the change in \( Y \) when \( X_2 \) is 0 and \( X_1 \) increases by one unit
- \( B_5 \) means the change in \( Y \) when \( X_1 \) is 0 and \( X_2 \) increases by one unit
- \( B_6 \) means the additional change in \( Y \) when the product of \( X_1 \) and \( X_2 \) increases by one unit

These four coefficients tell 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

\[ Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + \\
B_4 X_1 \times X_2 + B_5 X_1 \times X_3 + B_6 X_2 \times X_3 + \\
B_7 X_1 \times X_2 \times X_3 + \epsilon \]

\(B_0\) means the value of \(Y\) when \(X_1, X_2, X_3\) are all 0

\(B_1\) means the predicted change in \(Y\) when \(X_2\) and \(X_3\) are 0 and \(X_1\)
increases by one unit

\(B_2\) means the predicted change in \(Y\) when \(X_1\) and \(X_3\) are 0 and \(X_2\)
increases by one unit

\(B_3\) means the predicted change in \(Y\) when \(X_1\) and \(X_2\) are 0 and \(X_3\)
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>
<thead>
<tr>
<th>Command</th>
<th>The predictor of interest</th>
<th>predicted margins or the marginal effect</th>
<th>Adjustment of other variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>margins</td>
<td>[marginlist]</td>
<td>, [blank]</td>
<td>[blank]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>, dydx(varlist)</td>
<td>atmeans</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>at(at(atspec))</td>
</tr>
</tbody>
</table>

**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.

• The first three sets show how the `-margins-` command is used for a continuous dependent variable, a binary dependent variable, a nominal dependent variable.

• The fourth 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 wuhn@bgsu.edu