Quantitative Analysis and Empirical Methods

Interaction Effects

Jan Rovny

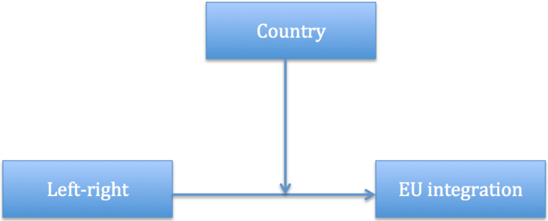
Sciences Po, Paris, CEE / LIEPP

Overview

- What are interactions
- Modeling interactions
- Demonstration in R!

Interaction effects

- A variable moderates the effect of another variable on the dependent variable.
- Capture the differential influence of X on Y because of a moderator Z
- For example:



Interaction effects

- Different levels of measurement of the predictors will require different approaches

1. Interaction between dummy variables
2. Interaction between dummy and continuous variables
3. Interaction between continuous variables

Interaction between dummy variables 1

Scenario s and b are both categorical predictors capturing state of residence (s) with values of Alpha, Beta and Chi, and race (b) with values 0=white and 1=black:

- To model this, need to 'dummy out' state
  Alpha: $s_1 = 0$, $s_2 = 0$
  Beta : $s_1 = 1$, $s_2 = 0$
  Chi: $s_1 = 0$, $s_2 = 1$
- Regular regression model:
  $y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 b_i + \epsilon_i$

Interaction between dummy variables 2

- To model an interaction between state and race:
- Substitute the original model:
  $y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 b_i + \epsilon_i$ for:

  $$y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 b_i + \beta_4 s_{1i} * b_i + \beta_5 s_{2i} * b_i + \epsilon_i,$$

  where $s * b$ is the interaction term.

Interaction between dummy variables 3

Equation:

$$y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 b_i + \beta_4 s_{1i} * b_i + \beta_5 s_{2i} * b_i + \epsilon_i$$

Interpretation:

Race	State	b	s_1	s_2	Model
White	Alpha	0	0	0	$y_i = \beta_0 + \epsilon_i$
White	Beta	0	1	0	$y_i = \beta_0 + \beta_1 + \epsilon_i$
White	Chi	0	0	1	$y_i = \beta_0 + \beta_2 + \epsilon_i$
Black	Alpha	1	0	0	$y_i = \beta_0 + \beta_3 + \epsilon_i$
Black	Beta	1	1	0	$y_i = \beta_0 + \beta_1 + \beta_3 + \beta_4 + \epsilon_i$
Black	Chi	1	0	1	$y_i = \beta_0 + \beta_2 + \beta_3 + \beta_5 + \epsilon_i$

Interaction between dummy variables 4

Equation:

$$y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 b_i + \beta_4 s_{1i} * b_i + \beta_5 s_{2i} * b_i + \epsilon_i$$

- Assessing the race gap (*black* − *white*):

  The effect of race in Alpha:

  $(\beta_0 + \beta_3 + \epsilon_i) - (\beta_0 + \epsilon_i) = \beta_3$

  The effect of race in Beta:

  $(\beta_0 + \beta_1 + \beta_3 + \beta_4 + \epsilon_i) - (\beta_0 + \beta_1 + \epsilon_i) = \beta_3 + \beta_4$

  The effect of race in Chi:

  $(\beta_0 + \beta_2 + \beta_3 + \beta_5 + \epsilon_i) - (\beta_0 + \beta_2 + \epsilon_i) = \beta_3 + \beta_5$

- Effect of race is constant across states only if:
  $\beta_4 = \beta_5 = 0$

Interaction between dummy variables 4

- We obtained the point estimates, but what about the uncertainty around them?
- We need to test the significance of the race gap
  - In Alpha, this is easy, we have the t-test on β_3
  - In Beta and Chi this is harder, we need to test the significance of $\beta_3 + \beta_4$ and of $\beta_3 + \beta_5$
  - To do this, we need the **Combined T-test**:
- $H_0 : \hat{\beta}_1 + \hat{\beta}_2 = 0$
- $Combined - T = \frac{\hat{\beta}_1 + \hat{\beta}_2}{\sqrt{\sigma^2_{\hat{\beta}_1} + \sigma^2_{\hat{\beta}_2} + 2\sigma^2_{\hat{\beta}_1, \hat{\beta}_2}}}$

See my demonstration

Interaction between dummy and continuous variables 1

Scenario We expect that the effect of education (x), measured in years of schooling, is moderated by the state of residence (s) with values of Alpha, Beta and Chi

- To model this, need to 'dummy out' state
  Alpha: $s_1 = 0$, $s_2 = 0$
  Beta : $s_1 = 1$, $s_2 = 0$
  Chi: $s_1 = 0$, $s_2 = 1$
- Regular regression model:
  $y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 x_i + \epsilon_i$
- Regression model with interaction:
  $y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 x_i + \beta_4 s_1 * x_i + \beta_5 s_2 * x_i + \epsilon_i$

Interaction between dummy and continuous variables 2

Equation:

$$y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + \beta_3 x_i + \beta_4 s_{1i} * x_i + \beta_5 s_{2i} * x_i + \epsilon_i$$

Interpretation:

State	s_1	s_2	Model	Constant	Simple Slope
Alpha	0	0	$y_i = \beta_0 + \beta_3 x_i + \epsilon_i$	β_0	β_3
Beta	1	0	$y_i = \beta_0 + \beta_1 + (\beta_3 + \beta_4) * x_i + \epsilon_i$	$\beta_0 + \beta_1$	$\beta_3 + \beta_4$
Chi	0	1	$y_i = \beta_0 + \beta_2 + (\beta_3 + \beta_5) * x_i + \epsilon_i$	$\beta_0 + \beta_2$	$\beta_3 + \beta_5$

Interaction between dummy and continuous variables 3

- Now we need to test the significance of our simple slopes
  - In Alpha, this is easy, we have the t-test on β_3
  - In Beta and Chi we need to test $\beta_3 + \beta_4$ and $\beta_3 + \beta_5$
- **Combined T-test**:
  - $H_0 : \hat{\beta}_1 + \hat{\beta}_2 = 0$
  - $Combined - T = \frac{\hat{\beta}_1 + \hat{\beta}_2}{\sqrt{\sigma^2_{\hat{\beta}_1} + \sigma^2_{\hat{\beta}_2} + 2\sigma^2_{\hat{\beta}_1, \hat{\beta}_2}}}$

See my demonstration

Interaction between two continuous variables 1

Scenario We expect the effect of x to be moderated by z, both of which are continuous predictors.

- Regular regression model:
  $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$
- Regression model with interaction:
  $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i * z_i + \epsilon_i$

Interaction between two continuous variables 2

Equation:

$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i * z_i + \epsilon_i$

Interpretation:

- The *effect* of x on y should continuously change as z changes
- Need to calculate the marginal effect of x on y as z changes:
  $\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 Z$
- And we need to calculate the standard errors of this marginal effect over the values of z:
  $se = \sqrt{\sigma^2_{\hat{\beta}_1} + Z^2\sigma^2_{\hat{\beta}_3} + 2Z\sigma^2_{\hat{\beta}_1,\hat{\beta}_3}}$

See my demonstration