Lab 5: Interaction Terms and Effect Modification

February 2021

Objective

Today, we will be looking at 2 different multiple linear regression models, with 2 possible parameterizations for each. We will look at how to conduct tests on specific hypotheses for each model, and study the different interactions and their meanings.

Quick Discussion: Modeling Vs. Estimation

$E[Y|X] = X\beta$ = Mean Model

β_1 vs. $\hat{\beta}_1$

OLS = Estimation

Primary Analysis

We are going to be looking at the following two formulas:

Coding 1

Y: A1c at 6 months
X_1: Indicator of ANY REACH
X_2: Indicator of REACH + FAMS
X_3: Baseline A1c

Coding 2

Y: A1c at 6 months
X_1: Indicator of ONLY REACH
X_2: Indicator of REACH + FAMS
X_3: Baseline A1c

Note: $X_1 + X_2$ under Coding 2 = X_1 from Coding 1

We are going to be looking at the following two models:

Model A (No effect modification by baseline A1c)

$$E[Y|X_1 = x_1, X_2 = x_2, X_3 = x_3] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Model B (Effect modification allowed)

$$E[Y|X_1 = x_1, X_2 = x_2, X_3 = x_3] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3$$

Questions

For each of these models, we can ask the following questions (estimation and/or hypothesis testing):

(1) Is receiving ONLY REACH effective relative to control?
(2) Is (REACH + FAMS) effective relative to control?
(3) Is receiving (REACH + FAMS) effective relative to ONLY REACH?
(4) Is receiving ANY REACH effective relative to control?

For each model and for each coding, how would you answer these questions? What parameters correspond to each? Write out each hypothesis test (Hint: Writing out expressions for each of these questions, i.e. $\mathrm{E}[Y|X_1 =?, X_2 = ?, X_3 =?]$ is incredibly helpful).

Solutions

Question 1

Coding 1, Model A)

Is receiving ONLY REACH effective relative to control?

We care about comparing the situation where $X_1 = 1$ and $X_2 = 0$ to the situation where both are 0, when X_3 can equal some x_3 in both scenarios.

E[$Y|X_1 = 1, X_2 = 0, X_3 = x_3$] - E[$Y|X_1 = 0, X_2 = 0, X_3 = x_3$] $= \beta_1$

So here, test $H_0 : \beta_1 = 0$

Coding 1, Model B)

We care about the same things here as in Model A:

E[$Y|X_1 = 1, X_2 = 0, X_3 = x_3$] - E[$Y|X_1 = 0, X_2 = 0, X_3 = x_3$] $= \beta_1 + \beta_4 * x_3$

So here, we will test $H_0 : \beta_1 = \beta_4 = 0$.

Coding 2, Model A)

We care about comparing the case where $X_1 = 1$ and $X_2 = 0$ to the case where both equal 0. This ends up looking the same as in Coding 1.

E[$Y|X_1 = 1, X_2 = 0, X_3 = x_3$] - E[$Y|X_1 = 0, X_2 = 0, X_3 = x_3$] $= \beta_1$

So here, test $H_0 : \beta_1 = 0$

Coding 2, Model B)

We care about the same case here:

E[$Y|X_1 = 1, X_2 = 0, X_3 = x_3$] - E[$Y|X_1 = 0, X_2 = 0, X_3 = x_3$] $= \beta_1 + \beta_4 * x_3$

So here, we will test $H_0 : \beta_1 = \beta_4 = 0$..

These end up being the same in both codings. Models A/B make different assumptions. But coding 1 and 2 doesn't really make different assumptions. We can add the X's from Coding 2 to get X_1 from Coding 1.

Question 2

Is (REACH + FAMS) effective relative to control?

Coding 1, Model A)

We care about the case when both X_1 and X_2 are 1, compared to when both are 0:

E$[Y|X_1 = 1, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 0, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 * x_3 - (\beta_0 + \beta_3 * x_3) = \beta_1 + \beta_2$$

So here, our test of interest is one of a linear combination of parameters:

$H_0 : \beta_1 + \beta_2 = 0$.

Coding 1, Model B)

We care about same case here:

E$[Y|X_1 = 1, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 0, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 * x_3 + \beta_4 * x_3 + \beta_5 * x_3 - (\beta_0 + \beta_3 * x_3) = \beta_1 + \beta_2 + x_3 * (\beta_4 + \beta_5)$$

$H_0 : \beta_1 = \beta_2 = \beta_4 = \beta_5$.

Coding 2, Model A)

We care about the case when $X_1 = 0$ and $X_2 = 1$ are 1, compared to when both are 0:

E$[Y|X_1 = 0, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 0, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_2 + \beta_3 * x_3 - (\beta_0 + \beta_3 * x_3) = \beta_2$$

Here, our hypothesis would be $\beta_2 = 0$.

Coding 2, Model B)

We care about the same case here:

E$[Y|X_1 = 0, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 0, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_2 + \beta_3 * x_3 + \beta_5 * x_3 - (\beta_0 + \beta_3 * x_3) = \beta_2 + x_3 * \beta_5$$

$H_0 : \beta_2 = \beta_5 = 0$

Question 3

Is receiving (REACH + FAMS) effective relative to ONLY REACH?

Coding 1, Model A)

We care about the case where $X_1 = 1$ and $X_2 = 1$ compared to the case where $X_1 = 1$ and $X_2 = 0$:

E$[Y|X_1 = 1, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 1, X_2 = 0, X_3 = x_3] = \beta_2$

Our hypothesis might be $H_0 : \beta_2 = 0$

Coding 1, Model B)

We care about the same case here:

E$[Y|X_1 = 1, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 1, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 * x_3 + \beta_4 * x_3 + \beta_5 * x_3 - (\beta_0 + \beta_1 + \beta_3 * x_3 + \beta_4 * x_3) = \beta_2 + \beta_5 * x_3$$

$H_0 : \beta_2 = \beta_5 = 0$
Coding 2, Model A)

We care about the case where $X_1 = 0$ and $X_2 = 1$ compared to the case where $X_1 = 1$ and $X_2 = 0$:

E$[Y|X_1 = 0, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 1, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_2 + \beta_3 * x_3 - (\beta_0 + \beta_1 + \beta_3 * x_3) = \beta_2 - \beta_1$$

Here, our hypothesis test is one of a linear combination of parameters: $H_0 = \beta_2 - \beta_1 = 0$.

Coding 2, Model B)

We care about the same case here:

E$[Y|X_1 = 0, X_2 = 1, X_3 = x_3]$ - E$[Y|X_1 = 1, X_2 = 0, X_3 = x_3] =$

$$\beta_0 + \beta_2 + \beta_3 * x_3 + \beta_5 * x_3 - (\beta_0 + \beta_1 + \beta_3 * x_3 + \beta_4 * x_3) = \beta_2 - \beta_1 + \beta_5 * x_3 - \beta_4 * x_3 = \beta_2 - \beta_1 + (\beta_5 - \beta_4)x_3$$

We have a more complex hypothesis test here.

Question 4

Is receiving ANY REACH effective relative to control?

Preface: this question is not intuitive. We will work through each part step-by-step.

Coding 1, Model A)

For this instance, we have three effective groups we can create with different values of coefficients. We have a control group, where both X_1 and X_2 are 0. We have a reach only group, where X_1 is 1 and X_2 is 0. And we have a reach + fams group, where both are 1. We want to test if receiving *any* reach is effective relative to control. The easiest thing to do here is to write out all three cases.

$E[Y|X]$ in the control group:

$\beta_0 + \beta_3 x_3$

$E[Y|X]$ in the reach only group:

$\beta_0 + \beta_1 + \beta_3 x_3$

$E[Y|X]$ in the reach+fams group:

$\beta_0 + \beta_1 + \beta_2 + \beta_3 x_3$

So our hypothesis: EITHER $\beta_1 \neq 0$ OR $\beta_1 + \beta_2 \neq 0$ in order for there to be an effect of any REACH. However, if $\beta_1 \neq 0$ then in order for $\beta_1 + \beta_2 = 0$, $\beta_2 = 0$. So our null hypothesis can be written as:

$$H_0 : \beta_1 = \beta_2 = 0$$

This is a joint test.

Coding 1, Model B)
Let's repeat this process
$E[Y|X]$ in the control group:

$\beta_0 + \beta_3 x_3$

$E[Y|X]$ in the reach only group:

$\beta_0 + \beta_1 + \beta_3 x_3 + \beta_4 x_3$

$E[Y|X]$ in the reach+fams group:

$\beta_0 + \beta_1 + \beta_2 + \beta_3 x_3 + \beta_4 x_3 + \beta_5 x_3$

So our hypothesis: $\beta_1 = \beta_4 = 0$ AND $\beta_1 = \beta_2 = \beta_4 = \beta_5 = 0$ in order for there to not be any effect of any REACH. We can combine this into:

$$H_0 : \beta_1 = \beta_2 = \beta_4 = \beta_5 = 0$$

Note the different uses of OR vs. AND in the hypotheses that I wrote out in the two models. When writing the null hypothesis out, I used ANDs-all of the βs had to be 0 in order for the null hypothesis to be that there is no effect of the treatment. However, when writing out the alternative hypotheses, I used OR; only one β in the entire hypothesis had to be different from 0 in order for the null to not be true.

Coding 2, Model A)

$E[Y|X]$ in the control group:

$\beta_0 + \beta_3 x_3$

$E[Y|X]$ in the reach only group:

$\beta_0 + \beta_1 + \beta_3 x_3$

$E[Y|X]$ in the reach+fams group:

$\beta_0 + \beta_2 + \beta_3 x_3$

So our null hypothesis: $\beta_1 = \beta_2 = 0$ in order for there to not be any effect of any REACH.

This ends up being equivalent to the results from Coding 1 due to the extra step we took to combine hypotheses. Coding 2, Model B)

$E[Y|X]$ in the control group:

$\beta_0 + \beta_3 x_3$

$E[Y|X]$ in the reach only group:

$\beta_0 + \beta_1 + \beta_3 x_3 + \beta_4 x_3$

$E[Y|X]$ in the reach+fams group:

$\beta_0 + \beta_2 + \beta_3 x_3 + \beta_5 x_3$

So our null hypothesis: $\beta_1 = \beta_2 = \beta_4 = \beta_5 = 0$ in order for there to not be any effect of any REACH.

This is true because if any of those βs do not equal 0, then there will be a difference in mean A1c between a group that received some REACH and one that didn't.