Interpretation of Interactions With Binary Predictors

Tutorial 11: Lecture 25

Ravleen Bajaj

What is an Interaction?

- An **interaction** means the effect of one predictor **depends on the value** of another predictor.
- In a linear regression model with one numeric predictor and one binary predictor, an interaction term allows the slope of the line relating $E[Y]$ to x to be **different across groups**.
- For a **binary predictor**:
  - No interaction: lines differ only in intercept.
  - Interaction: lines differ in slope and intercept.

Quadratic Model With Group Interaction

We model nightly sleep hours Y_i using average daily stress level x_{1i} in two groups:

- Group 0: No anxiety disorder
- Group 1: Anxiety disorder
- Let x_{2i} be a binary indicator for each group:

$$x_{2i} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ subject is in Group 1,} \\ 0, & \text{if the } i^{\text{th}} \text{ subject is in Group 0.} \end{cases}$$

We fit the model:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{2i} + \beta_4 (x_{1i} x_{2i}) + \beta_5 (x_{1i}^2 x_{2i}) + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and the ε_i's are independent.

Interpretation

- If person i has **no anxiety**, then $x_{2i} = 0$ and

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

- If person i has **anxiety**, then $x_{2i} = 1$ and

$$Y_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) x_i + (\beta_2 + \beta_5) x_i^2 + \varepsilon_i$$

- We are fitting a quadratic regression model for each group
  - The intercept and the coefficients of the linear and quadratic terms may differ by group.
  - SD of the errors are common.

Main and Interaction Effects

- Here, the effect of stress level depends on whether the person has an anxiety disorder.
    - β_4 and β_5 jointly characterize the effect of the interaction between stress and group.
- Similarly, the effect of group depends on the stress level.

Case 1: Two Numeric + Two Binary Predictors

We want to model the systolic blood pressure Y_i of 500 individuals using their Age (x_{1i}), BMI (x_{2i}), if they're active smokers or not and if they consume any prescribed medication to regulate their blood pressure.

- Let x_{3i} be a binary indicator for smokers:

$$x_{3i} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ subject is a smoker,} \\ 0, & \text{if the } i^{\text{th}} \text{ subject is not a smoker.} \end{cases}$$

- Let x_{4i} be a binary indicator for prescribed medication consumers:

$$x_{4i} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ subject is prescribed medication,} \\ 0, & \text{if the } i^{\text{th}} \text{ subject is not prescribed medication.} \end{cases}$$

Case 1: Two Numeric + Two Binary Predictors

We begin with the preliminary model with just the main effects of the predictors.

We would fit the model:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and the ε_i's are independent.

Main Effects (No Interaction)

- When no interaction terms are included, the effect of each predictor does not depend on the values of the others. An *interaction* occurs when the effect of one predictor depends on the value of another.
- Here, the effects of age and BMI do *not* depend on smoking or medication status. $\Rightarrow$ Effects are **additive**.
- β_1 and β_2 are the main effects of the numeric predictors.
- β_3 and β_4 are the main effects of the binary predictors.
- β_3: mean BP of smokers differs from that of non-smokers by a constant amount, holding age and BMI fixed.
- β_4: mean BP of medication consumers differs from that of non-consumers by a constant amount, holding age and BMI fixed.
- β_1, β_2: age and BMI change mean BP at the same rate within each smoking/medication group.
- There are **no interaction effects** because the model assumes that:
  - Smokers and non-smokers have the same age effect
  - Medication users and non-users have the same BMI effect

Case 2: Binary Interaction: Smoker × Medication

Now, let's assume the doctor prescribes different medications to smokers vs non-smokers based on some new information on side-effects:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 (x_{3i} x_{4i}) + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and the ε_i's are independent.

This allows the effect of medication to differ between smokers and non-smokers.

Interpretation

- If non-smoker, not medicated ($x_{3i} = 0, x_{4i} = 0$):

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

- If smoker only ($x_{3i} = 1, x_{4i} = 0$):

$$Y_i = (\beta_0 + \beta_3) + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

- If medicated only ($x_{3i} = 0, x_{4i} = 1$):

$$Y_i = (\beta_0 + \beta_4) + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

- If both smoker and medicated ($x_{3i} = 1, x_{4i} = 1$):

$$Y_i = (\beta_0 + \beta_3 + \beta_4 + \beta_5) + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

- Again:
  - Intercepts and effects of age and BMI may differ across groups.
  - SD of the errors are common.

Main and Interaction Effects

- Two binary predictors interact if the effect of one depends on the value of the other.
- Here, the effect of medication depends on whether the person is a smoker.
  - β_5 is the effect of the interaction between smoking and medication.
  - β_3 and β_4 are the main effects for smoking and medication.
- Similarly, the effect of smoking depends on medication use.

Case 3: Binary and Numeric Interactions

A researcher is given this data for this study and she realizes that the effect of BMI and the effect of age depend on smoking and medication status.

So, she recommends the following model:

$$
\begin{aligned}
Y_i = \beta_0 &+ \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 (x_{3i} x_{4i}) \\
&+ \beta_6 (x_{1i} x_{3i}) + \beta_7 (x_{1i} x_{4i}) + \beta_8 (x_{2i} x_{3i}) + \beta_9 (x_{2i} x_{4i}) + \varepsilon_i
\end{aligned}
$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and the ε_i's are independent.

Interpretation

- The effect of age differs for smokers (β_6).
- The effect of age differs for medicated individuals (β_7).
- The effect of BMI differs for smokers (β_8).
- The effect of BMI differs for medicated individuals (β_9).
- The effect of smoking depends on medication status (β_5).

- We are fitting separate planes for each combination of medication and smoking status:
  - SD of errors are common

Main and Interaction Effects

- Two predictors *interact* if the effect of one depends on the value of another.
- Here, the effects of age and BMI depend on smoking and medication status.
  - β_6 and β_8 are the interaction effects for age and BMI with smoking.
  - β_7 and β_9 are the interaction effects for age and BMI with medication.
  - β_5 is the interaction effect between the two binary predictors, i.e., smoking and prescribed medication consumption.
- $\beta_1, \beta_2, \beta_3, \beta_4$ are the main effects.

Takeaways

- Interactions allow regression coefficients to differ across binary groups.
- Without interactions: coefficients do not change.
- With interactions: groups can have different coefficients.

Thank You!

Questions?