Using and Interpreting Interaction Terms in Regression Analysis

May 01, 2025

Introduction

This is a brief explainer on how to use and interpret interaction terms in regression analysis. We will focus on the core concepts using Huntington-Klein (2022) as a guide and illustrate them with a practical example using American National Election Studies (ANES) data from the `stevedata` package.

What are Interaction Terms?

Often in social science, the relationship between two variables is *conditional*. That is, the effect of a predictor X on an outcome Y might change depending on the level of a third variable, Z. Interaction terms allow us to explicitly model this conditionality.

Huntington-Klein (2022, 218, n. 58) defines an interaction term simply as:

> "…the product of two variables multiplied together and included in a regression, usually alongside the un-multiplied versions."

The standard regression model incorporating an interaction looks like this:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 (X \cdot Z) + \epsilon$$

Here, β_3 is the coefficient for the interaction term $(X \cdot Z)$. You have to include the *constituent terms* ($\beta_1 X$ and $\beta_2 Z$) alongside the interaction term. Omitting them would mean β_3 incorrectly absorbs not only the interaction effect but also parts of the direct effects of X and Z.

Interpreting Interaction Models

When your model includes an interaction term, the interpretation of the coefficients changes compared to a simple additive model. We need to understand two key aspects (Huntington-Klein 2022, 219):

1. How to describe the effect of X on Y.
2. How to interpret the interaction coefficient, β_3.

The Effect of X (Conditional Effect)

In the interaction model, β_1 **no longer represents the overall effect of** X. Instead, the effect of X on Y *depends on the value of* Z. We find this conditional effect using the partial derivative of Y with respect to X:

$$\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 Z$$

- **The Math:** This equation shows that the slope of the relationship between X and Y is not constant, but changes linearly with the value of Z.
- **Example:** Imagine our ANES example where Y is feelings towards Democrats (`therm_dem`), X is Republican Party ID (`pid7`), and Z is education (`educat`). The equation $\frac{\partial \text{therm\_dem}}{\partial \text{pid7}} = \beta_1 + \beta_3 \times \text{educat}$ means the effect of becoming more Republican on feelings towards Democrats isn't a single number. It's a value that potentially differs for people with low education versus high education. β_1 would represent the effect of Party ID *only* for those with an education level of $Z = 0$ (which might be outside the actual data range, making β_1 alone not directly interpretable in many cases).

Interpreting the Interaction Coefficient (β_3)

The interaction coefficient, β_3, quantifies *how the effect of* X changes as Z changes. Huntington-Klein (2022, 220) provides an explanation:

> "[β_3] tells us how our prediction of $\partial Y/\partial X$ changes when Z increases by one unit. In other words, β_3 is how much stronger the effect of X on Y gets when Z increases by one unit."

- **The Math:** β_3 is the coefficient on Z in the equation for the conditional effect of X. It tells us how much that conditional effect ($\beta_1 + \beta_3 Z$) changes for each one-unit increase in Z.
- **Example:** In our ANES example, β_3 would tell us how much the effect of moving one step more Republican (`pid7`) on Democratic ratings (`therm_dem`) changes for each one-unit increase in the education scale (`educat`). If β_3 were negative, it would mean that the negative impact of Republican ID on Democratic ratings becomes *even stronger* (more negative) as education increases. If β_3 were positive, it would mean the effect becomes weaker (less negative or potentially positive) as education increases.

Example: Partisanship, Education, and Democratic Party Ratings

The Data: `stevedata::anes_partytherms`

We use data from the American National Election Studies (ANES) time series, specifically the `anes_partytherms` dataset from the `stevedata` package. This dataset includes survey responses from 1978-2012. For this example, we focus on three key variables:

- `therm_dem`: The respondent's "feeling thermometer" rating for the Democratic party (0 = Cold/Unfavorable, 100 = Warm/Favorable). This is our outcome variable (Y).
- `pid7`: A 7-point party identification scale (1 = Strong Democrat, 4 = Independent, 7 = Strong Republican). This is our main predictor (X).
- `educat`: The respondent's education level, typically coded on a scale (e.g., 1 = Less than HS, ..., 7 = Advanced Degree). This is our moderator variable (Z).

Our research question is: Does the effect of `pid7` on `therm_dem` depend on the level of `educat`?

Model Specification

We estimate the following OLS model:

$$\text{therm\_dem}_i = \beta_0 + \beta_1 \text{pid7}_i + \beta_2 \text{educat}_i + \beta_3 (\text{pid7}_i \times \text{educat}_i) + \epsilon_i$$

Estimation using `lm()`

```
# Load data
data(anes_partytherms, package = "stevedata")

# Estimate model using base R's lm()
model_lm <- lm(therm_dem ~ pid7 * educat, data = anes_partytherms)

# The '*' automatically includes pid7, educat, and pid7:educat
```

Results

```
# Regression table using modelsummary with kableExtra
# Ensure variable names are clear in the table
modelsummary(model_lm,
             output = "kableExtra",
             title = "Interaction of Partisanship and Education on
                      Democratic Thermometer Rating",
             coef_rename = c("(Intercept)" = "Intercept",
                             "pid7" = "Party ID (1=SDem -> 7=SRep)",
                             "educat" = "Education (1=Low -> 7=High)",
                             "pid7:educat" = "Party ID x Education"),
             gof_map = c("nobs", "r.squared"),
             stars = TRUE,
             notes = list('Data: ANES (1978-2012). OLS standard errors
                          in parentheses.',
                          '*** p<0.001, ** p<0.01, * p<0.05'))
```

Answering the Interpretation Questions with Results

Let's use the results from Table 1.

1. What is the effect of Party ID (`pid7`) on Democratic ratings?

The conditional effect is $\beta_1 + \beta_3 \times \text{educat}$. Using the estimated coefficients from Table 1:

Effect of `pid7` = -7.5 + (-0.065 × `educat`).

We need to evaluate this at specific education levels:

```
## The estimated effect of a one-unit increase in ‘pid7‘ (towards Republican) on ‘therm_dem‘ is:

## * At Education Level 1: -7.568
## * At Education Level 4: -7.763
## * At Education Level 7: -7.958
```

Table 1: Interaction of Partisanship and Education on Democratic Thermometer Rating

	(1)
Intercept	91.794***
	(0.521)
Party ID (1=SDem -> 7=SRep)	−7.503***
	(0.134)
Education (1=Low -> 7=High)	−1.049***
	(0.118)
Party ID x Education	−0.065*
	(0.029)
Num.Obs.	30 525
R2	0.459

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001
Data: ANES (1978-2012). OLS standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05

As shown, the negative effect of identifying as more Republican on feelings towards the Democratic party becomes slightly stronger (more negative) as education level increases.

2. How do we interpret the interaction coefficient (β_3)?

The coefficient for `Party ID x Education` (β_3) is estimated to be -0.065 (and is statistically significant in this case, $p < 0.05$).

This means that for each one-unit increase on the education scale (`educat`), the negative relationship between `pid7` and `therm_dem` becomes stronger (more negative) by approximately 0.065 points on the thermometer scale.

Visualization using ggplot2

Visualizing the interaction helps clarify the conditional relationship. Figure 1 plots the predicted `therm_dem` across the range of `pid7` for low (1), middle (4), and high (7) education levels, based on our model.

```r
# Get predicted values using ggeffects from the lm model
# Need to specify the model object correctly in the terms argument
pred_data <- ggpredict(model_lm, terms = c("pid7", "educat [1, 4, 7]"))

# Create plot using ggplot2
ggplot(pred_data, aes(x = x, y = predicted, color = group)) +
  geom_line(linewidth = 1) +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high, fill = group),
              alpha = 0.15, linetype = "dashed") +
  scale_color_manual(values = c("navyblue", "forestgreen", "darkred"),
                     name = "Education Level:",
                     labels = c("1 (Low)", "4 (Mid)", "7 (High)")) +
  scale_fill_manual(values = c("navyblue", "forestgreen", "darkred"),
                    name = "Education Level:",
                    labels = c("1 (Low)", "4 (Mid)", "7 (High)")) +
  labs(title = "",
       x = "Party Identification (1=Strong Dem to 7=Strong Rep)",
       y = "Predicted Thermometer Rating (Democrats)") +
```

```
theme_bw() +
theme(legend.position = "bottom",
      plot.title = element_text(hjust = 0.5)) # Center title
```

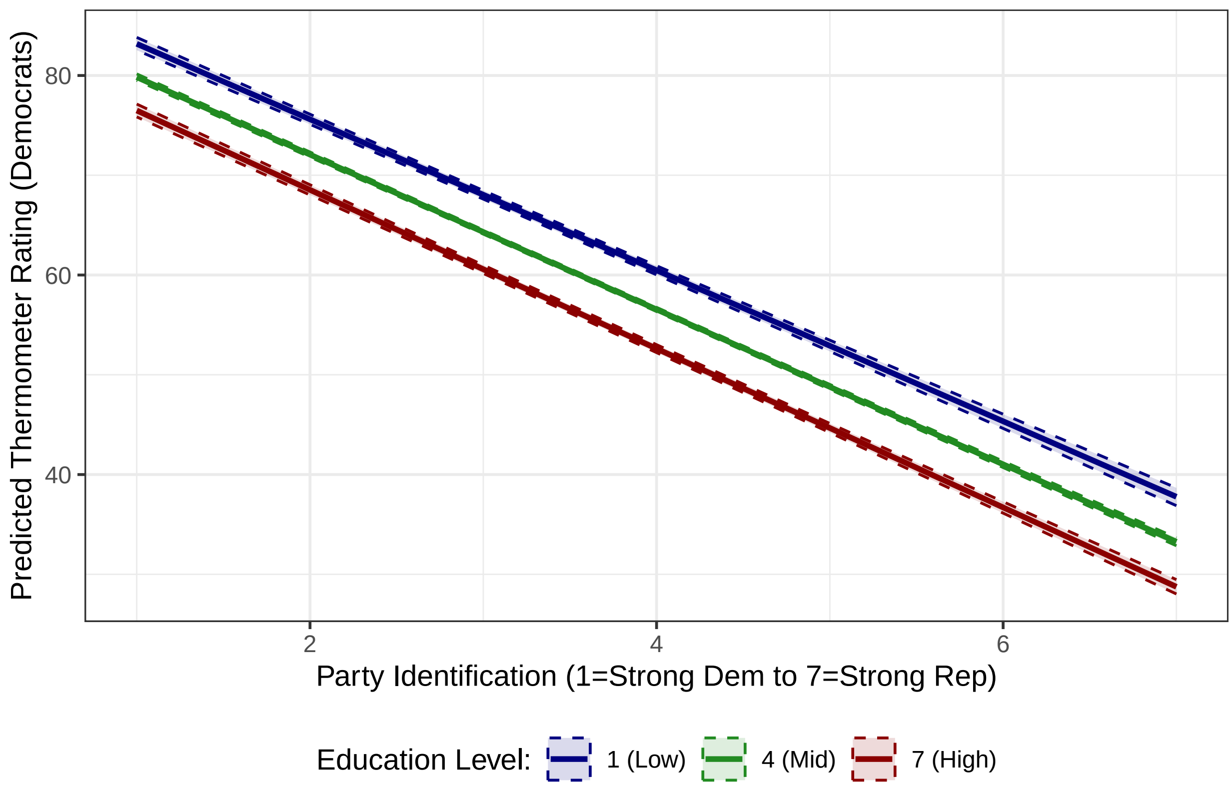

Figure 1: Predicted Democratic Thermometer Rating by Partisanship and Education Level.

Figure 1 visually confirms the interaction: the downward slope representing the effect of moving towards Republican ID is slightly steeper for the highly educated group (red line) compared to the less educated group (blue line).

Conclusion

Interaction terms are essential for modeling conditional relationships in regression analysis, but they require careful interpretation and parsimony. Remember that interpreting models with interactions can be achieved by calculating *conditional effects* (the effect of X at specific values of Z) and understanding the *interaction coefficient* (β_3) as the rate at which the effect of X changes per unit change in Z.