3.1 Alternate Variable Types and Interactions

Dr. Bean - Stat 5100

1 Why Interactions?

Example (HO 3.1.1): Y = `cycles`, X_1 = `charge_rate`, X_2 = `temperature`

All models we have discussed in this class assume that the effects of the explanatory variables are **additive**.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

In other words, the effect of each explanatory variable can be considered **separate** from all other explanatory variables.

What if the **real** effect of X_1 on Y actually depends on X_2 as well?

What would it mean for the effect of `charge_rate` on `cycles` to depend on `temperature`?

- We "know": higher `charge_rate` → lower `cycles`, and
  higher `temperature` → higher `cycles`
- But maybe: higher `charge_rate` **and** higher `temperature` → **much** higher `cycles`
- "**much**" higher here: significantly more than could be attributed to the sum of the effects of `charge_rate` and `temperature` only (often called **synergy**)

Whenever the effect of an explanatory variable (X_k) on the response (Y) *depends on* the values of other explantory variables, you have an **interaction effect**.

Metaphor: The bachelorette - the relationship of each potential suitor (X_k) with the bachelorette (Y) is partially depends upon the other potential suitors.

(Groups) How is an interaction effect different from multicollinearity?

Define an interaction term as a new predictor variable:

$$\begin{aligned} X_3 &= X_1 \cdot X_2 \\ Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \end{aligned}$$

Note: sometimes β_{12} instead of β_3

1.1 How to interpret interaction terms?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

- if X_1 increases by 1 unit, then we expect an average change of $\beta_1 + \beta_3 X_2$ in Y
  - the effect of X_1 on Y depends on X_2
  - if the interaction term is non-zero, we *cannot* separate the effect of X_1 from the effect of X_2. We must consider them jointly (unless X_1 or $X_2 = 0$).

1.2 Best Practices

- Don't check all possible interactions. Only include an interaction term in a linear model if its output is interpretable.
- Include all lower-ordered terms that compose an interaction term, regardless of the significance of the lower interaction term.
  - Prevents forcing lower ordered coefficients to zero.
  - Maintains a flexible response surface and facilitates interpretation.

1.3 Things to remember about interactions:

- Unless the X_k are standardized, the interaction term $X_3 = X_1 * X_2$ is likely to be collinear with either X_1 or X_2.
  - This will ruin inference for the "lower order" terms, but not the interaction term.
- Two-way interactions are often interpretable, but higher order interactions (ex: $X_4 = X_1 * X_2 * X_3$) become difficult to interpret.
  - A plot of residuals from a non-interaction model against the potential interaction term may help to determine inclusion (if a trend is apparent).
- If your problem is best solved by including multiple, high-ordered, interaction terms, then regression trees/random forests is likely a better approach (more in Module 4).

1.4 Polynomial Predictors

- Up to this point, we have limited ourselves to modeling variables that share a linear relationship.
- If a variable X_k shares a quadratic, or higher-order (often called "curvilinear") relationship with Y, then that means that the effect of X_k on Y *depends upon itself* (i.e. interacts with itself).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \varepsilon$$

- Handle higher-ordered terms the same way we handle other interaction terms:
  - include lower-order terms
  - standardize to reduce multicollinearity

- coefficient interpretations important: – if X_1 increases by 1 unit (and X_2 held constant), then we expect an average change in Y of $\beta_1 + \beta_3 X_2 + \beta_4 \cdot (2X_1 + 1)$

2 Alternate Variable Types

Up to this point we have only focused on **quantitative variables**:

- Values are represented as numbers where number *order* and *magnitude* matters.
- Quantiative variables can be either:
  - Continuous: can take on any value (theoretically infinite number of decimal places) within a range.
  - Discrete: can only take on a discrete (countable) set of values.

We now wish to also consider **qualitative variables**

- Cannot be measured/ordered on a numerical scale.
- SAS can't recognize words/letters in a regression model, and it will treat a set of numbered factored levels as quantitative (and thus order the levels).
- Because of this, we use **dummy/indicator variables** to include qualitative predictors in a model.

2.1 Dummy Variables

Consider the following student demographic variables (qualitative in bold): (age, height, **Utah residency status**, weight, **major college**)

Use an indicator variable to include residency status in model

$$X = I_{\text{resident}} = \begin{cases} 1 & \text{if student is resident of Utah} \\ 0 & \text{otherwise} \end{cases}$$

Things get a little more complicated for major college as we have to create multiple dummy variables to represent a single categorical variable:

$$\begin{aligned} X_1 &= I_{\text{College of Science}} = \begin{cases} 1 & \text{if student's major is within the college of science} \\ 0 & \text{otherwise} \end{cases} \\ X_2 &= I_{\text{College of Engineering}} \\ &\vdots \\ X_7 &= I_{\text{School of Business}} \end{aligned}$$

(Groups) If there are eight colleges in the University, why would I only have seven dummy variables?

3 Example (See HO 3.1.1)

Y = months, X_1 = size, X_2 = type of firm

Note that $X_2 = I_{[\text{firm} = \text{stock}]} = \begin{cases} 1 & \text{if firm} = \text{stock} \\ 0 & \text{otherwise} \end{cases}$

Model with only qualitative predictor:

$$Y = \beta_0 + \beta_2 X_2 + \varepsilon$$

- equivalent to a two-sample t-test
- special case of one-way ANOVA model (`proc glm`, STAT 5200)

$$\begin{aligned} Y_{i,j} &= \mu_i + \epsilon_{i,j}, \qquad i = 1, 2; j = 1, \ldots, n_i \\ &= \mu + \alpha_i + \epsilon_{i,j}, \qquad \sum_{i=1}^{2} \alpha_i = 0 \\ \epsilon_{i,j} &\ iid \ \ N(0, \sigma^2) \end{aligned}$$

Model with both qualitative and quantitative predictor:

- Additive

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- Interaction

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

Note how the additive and interaction models differ:
(in the size (X_1) vs. months (Y) relationship for each firm type)

- Additive:
  - stock ($X_2 = 1$): $Y = (\beta_0 + \beta_2) + \beta_1 X_1 + \varepsilon$
  - mutual ($X_2 = 0$): $Y = \beta_0 + \beta_1 X_1 + \varepsilon$
- Interaction

- stock ($X_2 = 1$): $Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_1 + \varepsilon$
- mutual ($X_2 = 0$): $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

Note that the additive model results in *two parallel lines*, where the difference between stock and mutual firms are separated by a constant distance β_2. Whereas in the interaction model, both the slope *and* the intercept are different.

3.1 Note on interactions between qualitative predictors.

- possibly very interesting
- numerically much easier in [two-way] ANOVA setting (`proc glm`, STAT 5200), as ANOVA doesn't require the use of dummy variables.