Tutorial: Dummy Variables and Interaction Terms in Regression

Feedback Form: tiny.cc/hammadfeedback

Hammad Shaikh

Dummy (Indicator) Variables in Regression

$$\underbrace{I}_{\text{indicator}}(\text{statement}) = \begin{cases} 1, & \text{true} \\ 0, & \text{false} \end{cases} = \begin{cases} 1, & \text{i college grad} \\ 0, & \text{otherwise} \end{cases}$$

- Consider dummy variable of $\overbrace{I(\text{i college grad})}$

$C \longrightarrow Y$

- $Y_i = \beta_0 + \beta_1 \overbrace{I(\text{i college grad})}^{C} + \epsilon_i$
  - How to interpret β_1?

$$E(Y|C) = \beta_0 + \beta_1 C + \overbrace{E(\varepsilon|C)}^{0}$$

$$\hookrightarrow \left.\begin{cases} E(Y|C=0) = \beta_0 \\ E(Y|C=1) = \beta_0 + \beta_1 \end{cases}\right\} \beta_1 = E(Y|C=1) - E(Y|C=0)$$

Interaction Term in Regression

$C \to Y$ depend

$M = I(\text{male})$

- Recall: $Y_i = \beta_0 + \beta_1 \underbrace{I(\text{i college grad})}_{C} + \epsilon_i$
  - $\widehat{b}_1 = \overline{Y}_{college} - \overline{Y}_{HS}$

- Question: Do returns to college education differ by gender?

$Y_{males} = \beta_0^M + \beta_1^M C + \varepsilon^M \Rightarrow \beta_1^M =$ male educ. returns

$Y_{females} = \beta_0^F + \beta_1^F C + \varepsilon^F \Rightarrow \beta_1^F =$ female educ returns

$Y = \beta_0 + \beta_1 C + \beta_2 M + \beta_3 M \times C + \varepsilon$

$\beta_3 = \beta_1^M - \beta_1^F$

Dummy Variables With More Than Two Levels

$$= \begin{cases} 1, & \text{bachelor is highest educ} \\ 0, & \text{otherwise} \end{cases}$$

- $Y_i = \beta_0 + \beta_1 \overbrace{I(\text{Bachelor})} + \beta_2 I(\text{Master}) + \beta_3 I(Doctorate) + \epsilon_i$

$$I(\text{HS grad or less}) + I(\text{Bachelor}) + I(\text{Master}) + I(\text{Doctorate}) = 1$$

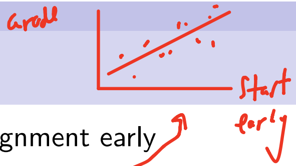
- What is the omitted educational attainment category?

↳ $I(\text{HS grad or less})$

- How to interpret β_3?

$$\left.\begin{aligned} E(Y \mid \text{Doctorate}) &= \beta_0 + \beta_3 \\ \beta_0 &= E(Y \mid \text{HS grad or less}) \end{aligned}\right\} \beta_3 = E(Y \mid \text{Doctorate}) - E(Y \mid \text{HS grad or less})$$

Factorial Experiments Application



- Goal: encourage students to start assignment early

- Interventions: 1) Typical Message, 2) Informational Message

$$\text{Class} \xrightarrow{\frac{1}{2}} \text{Typical} \qquad \text{Class} \xrightarrow{\frac{1}{2}} \text{Informational}$$

- Factor: 1) TA sends email, 2) Instructor sends email

$$\text{Typical} \xrightarrow{\frac{1}{2}} \text{TA} \qquad \text{Typical} \xrightarrow{\frac{1}{2}} \text{Instructor} \qquad \text{Info} \xrightarrow{\frac{1}{2}} \text{TA} \qquad \text{Info} \xrightarrow{\frac{1}{2}} \text{Instructor}$$

- How to measure differential impact on assignment grade?

$$Y = \beta_0 + \beta_1 I(\text{typical}) + \beta_2 I(\text{TA}) + \beta_3 I(\text{TA}) \times I(\text{typical}) + \varepsilon$$

Factorial Experiments Application

- $Y = \beta_0 + \beta_1 I(info) + \beta_2 I(Instructor) + \beta_3 I(info) I(instructor) + \epsilon$

- How to interpret β_0?

$\beta_0 = E(Y \mid info = 0, instructor = 0)$

- How to interpret β_3?

$\beta_3 = [E(Y \mid info = 1, instructor = 1) - E(Y \mid info = 0, instructor = 1)]$

$- [E(Y \mid info = 1, inst = 0) - E(Y \mid info = 0, instr = 0)]$

Dummy Variable Trap

$I(TA) + I(instructor) = 1$
multicolinearity

- $Y = \beta_0 + \beta_1 I(TA) + \beta_2 I(Instructor) + \epsilon$

↳ No omitted category

- How to interpret β_1?

$E(Y|TA=1) = \beta_0 + \beta_1$
$E(Y|Instructor=1) = \beta_0 + \beta_2$
can't isolate for β_1

- $Y = \alpha_1 I(TA) + \alpha_2 I(Instructor) + \epsilon$

- How to interpret α_1?

$\alpha_1 = E(Y|TA=1)$, $\alpha_2 = E(Y|Instructor=1)$

Alternative Representation of Factorial Experiment

- $Y = \alpha_1 I(\mathit{info}) \times I(\mathit{Instructor}) + \alpha_2 I(\mathit{info}) \times I(\mathit{TA}) + \alpha_3 I(\mathit{instructor}) + \alpha_4 I(\mathit{TA})$

- How to interpret α_3?

$$E(Y \mid instructor = 1, info = 0) = \alpha_3$$

- How to interpret α_1?

$$\alpha_1 = \underbrace{E(Y \mid info = 1, instructor = 1)}_{\alpha_1 + \alpha_3} - \underbrace{E(Y \mid instructor = 1, info = 0)}_{\alpha_3}$$

$$Y = \beta_0 + \beta_1 I(\text{collgrad}) + \beta_2 I(\text{male})$$

$$+ \underbrace{\beta_3}_{= \beta_1^M - \beta_1^F} I(\text{collgrad}) \times I(\text{male}) + \varepsilon$$

$$E(Y \mid C, M) = \beta_0 + \beta_1 C + \beta_2 M + \beta_3 C \times M$$

$$\left.\begin{array}{ll}
\text{Males } (M=1) & \left\{\begin{array}{l} \left.\begin{array}{l} E(Y \mid C=1, M=1) = \beta_0 + \beta_1 + \beta_2 + \beta_3 \\ E(Y \mid C=0, M=1) = \beta_0 + \beta_2 \end{array}\right\} \beta_1 + \beta_3 \end{array}\right. \\
\text{Females } (M=0) & \left\{\begin{array}{l} \left.\begin{array}{l} E(Y \mid C=1, M=0) = \beta_0 + \beta_1 \\ E(Y \mid C=0, M=0) = \beta_0 \end{array}\right\} \beta_1 \end{array}\right.
\end{array}\right\} \beta_3$$