Lecture 5-3: Dummy Variables

Goal: Explore dummy variables: two categories, multiple categories, and two dummy variables with their interaction, which, with an intercept, give each of four groups its own mean.

- Dummy variables
  - Why use dummy variables
  - Two categories
  - Multiple categories
  - Two dummy variables and their interaction

Worksheet

1. Suppose X_i is the region of counting area i: London, 8 other English regions, Wales, or Scotland. Estimate

$$Y_i = \beta_0 + \beta_1 D_{SE,i} + \beta_2 D_{SW,i} + \cdots + \beta_9 D_{Wales,i} + \beta_{10} D_{Scot,i} + u_i,$$

   with London omitted.

   (a) Interpret β_0

   (b) Interpret each β_j, $j \in \{1, \ldots, 10\}$.

   (c) Derive the estimand and an OLS t–test for $H_0 : E[Y|\text{Wales}] = E[Y|\text{Scotland}]$. Express the test as a linear restriction $R\boldsymbol{\beta} = \mathbf{0}$ and give the variance formula used for its standard error.

2. Let X_{White} indicate a white (non-Hispanic) voter and X_{Col} a voter with a 4-year college degree. Estimate

$$Y = \beta_0 + \beta_1 X_{White} + \beta_2 X_{Col} + \beta_3 (X_{White} \cdot X_{Col}) + u.$$

(a) Compute $E[Y|X_{White} = w, X_{Col} = c]$ for $(w, c) \in \{0, 1\} \times \{0, 1\}$.

(b) Show that β_3 equals the difference–in–differences:

$$\beta_3 = \big(E[Y|1, 1] - E[Y|1, 0]\big) - \big(E[Y|0, 1] - E[Y|0, 0]\big).$$

(c) Give a test for "no interaction" and interpret it.