Econometrics

Chapter 8: Multiple Regression with Binary (Dummy) Variables

In Choi

Sogang University

Dummy Variables

- A dummy variable is a variable that takes on the value 1 or 0.
- Dummy variables are also called binary variables, for obvious reasons.

A Dummy Independent Variable

- Consider a simple model with one continuous variable (x) and one dummy (d):

$$y_i = \beta_0 + \delta_0 d_i + \beta_1 x_i + u_i.$$

This can be interpreted as an intercept shift.
If $d_i = 0$, then $y_i = \beta_0 + \beta_1 x_i + u_i$. If $d_i = 1$, then $y_i = (\beta_0 + \delta_0) + \beta_1 x_i + u_i$.

A Dummy Independent Variable

Example

Earnings equation for married women

$$\ln earnings = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 education + \beta_4 kids + \varepsilon$$

$$kids = \begin{cases} 0, & \text{no kids} \\ 1, & \text{kids} \end{cases}$$

Variable	Coefficient	s.e.	t
Age	0.20056	0.08386	2.392
Age^2	–0.0023147	0.00098688	–2.345
Education	0.067472	0.025248	2.672
Kids	–0.35119	0.14753	–2.380

The earnings of women with children are 35% less than those without.

Dummies for Multiple Categories

- We can use dummy variables to control for something with multiple categories.

Example

(Seasonal effects)
Suppose y_t is a quarterly data

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \delta_1 D_{1t} + \delta_2 D_{2t} + \delta_3 D_{3t} + \varepsilon_t.$$

Let

$$D_{1t} = \begin{cases} 1, & \text{if spring} \\ 0, & \text{otherwise} \end{cases} \qquad D_{2t} = \begin{cases} 1, & \text{if summer} \\ 0, & \text{otherwise} \end{cases}$$

$$D_{3t} = \begin{cases} 1, & \text{if fall} \\ 0, & \text{otherwise} \end{cases}$$

δ_i denote seasonal effects.

Dummies for Multiple Categories

Example

(continued) Alternatively, we may use

$$y_t = \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \delta_1 D_{1t} + \delta_2 D_{2t} + \delta_3 D_{3t} + \delta_4 D_{4t} + \varepsilon_t.$$

But we should not use

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \delta_1 D_{1t} + \delta_2 D_{2t} + \delta_3 D_{3t} + \delta_4 D_{4t} + \varepsilon_t.$$

Since

$$D_{1t} + D_{2t} + D_{3t} + D_{4t} = 1,$$

we run into a multicollinearity problem.

Dummies for Multiple Categories

Example

(Threshold effects and categorical variables)

$$income_i = \beta_0 + \beta_1 age_i + \delta_B B_i + \delta_M M_i + \delta_P P_i + \varepsilon_i$$

$$B_i = \begin{cases} 1, & \text{Bachelor's degree} \\ 0, & \text{otherwise} \end{cases}$$

$$M_i = \begin{cases} 1, & \text{Master's and Bachelor's degrees} \\ 0, & \text{otherwise} \end{cases}$$

$$P_i = \begin{cases} 1, & \text{Ph.D., Master's and Bachelor's degrees} \\ 0, & \text{otherwise} \end{cases}$$

Dummies for Multiple Categories

Example

(continued) High school

$$E\left[income_i|age_i, HS\right] = \beta_1 + \beta_2 age_i$$

Bachelor's

$$E\left[income_i|age_i, B\right] = \beta_1 + \beta_2 age_i + \delta_B$$

Master's

$$E\left[income_i|age_i, M\right] = \beta_1 + \beta_2 age_i + \delta_B + \delta_M$$

Ph.D.

$$E\left[income_i|age_i, P\right] = \beta_1 + \beta_2 age_i + \delta_B + \delta_M + \delta_P$$

Dummies for Multiple Categories

Example

(continued) High school

$$\begin{aligned}
\delta_B &: \text{ marginal effect of Bachelor's degree} \\
\delta_M &: \text{ marginal effect of Master's degree} \\
\delta_P &: \text{ marginal effect of Ph.D.'s degree}
\end{aligned}$$