University of Nottingham
UK | CHINA | MALAYSIA

Introductory Econometrics BUSI2053

Dummy (Indicator) Variables

Lecture Outline

- Indicator (Dummy) Variables
- Applying Indicator Variables for Two groups
- Indicator Variables for more than Two groups
- Slope Dummy Variables
- Testing joint significance of qualitative factors

- **Suggested Reading:**
- Chapter 7: Hill, R.C., Griffiths W.E. and Lim, G.C. Principles of Econometrics, fourth edition, Wiley, 2012 (pp. 258-271)
- Chapter 9, 13: Gujarati, D.N. and Porter D.C. Basic econometrics, 5th ed., McGraw-Hill, 2009 (pp. 277-290; 467-474)
- Chapter 5: Dougherty, Christopher. Introduction to econometrics. 4th ed. Oxford University Press, 2011 (pp.224-244)
- Chapter 13: Westhoff, F. An introduction to econometrics: a self-contained approach, MIT Press, 2013

Indicator (Dummy) Variables

- Indicator variables are used to account for qualitative (categorical) factors in econometric models
- Indicator (dummy) variables allow us to construct models in which some or all regression model parameters, including the intercept, change for some observations in the sample
- They are often called **binary or dichotomous** variables, because they take just two values, usually one or zero,
- They are also called **dummy variables**, to indicate a qualitative, non-numeric characteristic
- Generally, we define an indicator (Dummy) variable D as:

$$D = \begin{cases} 1 & \text{if characteristic is present} \\ 0 & \text{if characteristic is not present} \end{cases}$$

- The value $D = 0$ defines the reference group, or base group

Indicator Variables for Two groups

- Consider a model to predict the value of a house (property) as a function of its characteristics:
  - size
  - location (desirable or not)
  - number of bedrooms
  - age
- How do we account for location, which is a qualitative variable?
- To account for location, a qualitative variable, we would have a dummy variable:

$$D = \begin{cases} 1 & \text{if property is in the desirable neighborhood} \\ 0 & \text{if property is not in the desirable neighborhood} \end{cases}$$

Indicator Variables for Two groups

- Consider the square footage at first: $PRICE = \beta_1 + \beta_2 SQFT + e$

  - β_2 is the value of an additional square foot of living area and β_1 is the value of the land alone

- Adding our indicator variable for location to our model:

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + e$$

- If our model is correctly specified, then:

$$E(PRICE) = \begin{cases} (\beta_1 + \delta) + \beta_2 SQFT & \text{when } D = 1 \text{ (desirable location)} \\ \beta_1 + \beta_2 SQFT & \text{when } D = 0 \text{ (not desirable)} \end{cases}$$

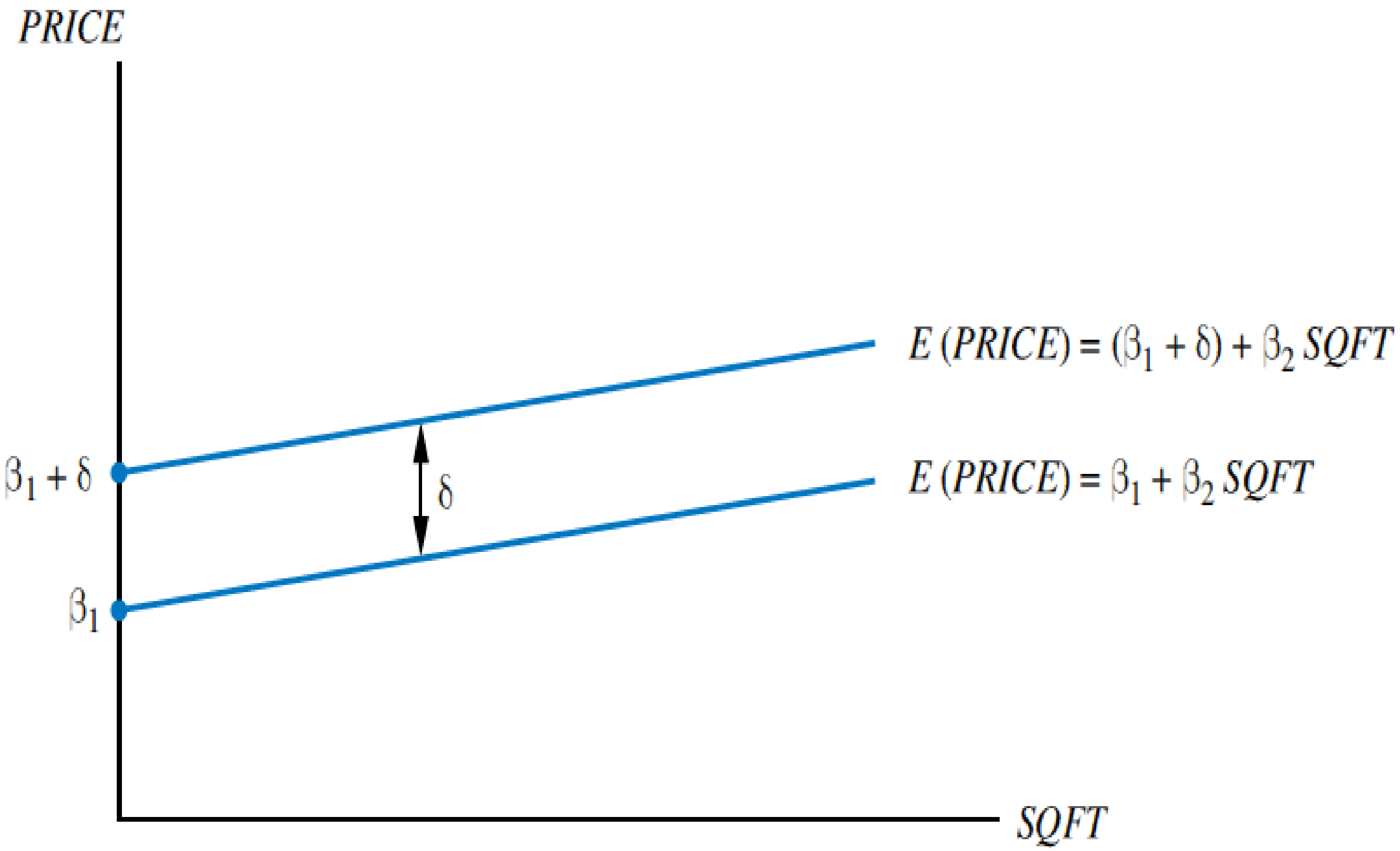
Note: δ is the population parameter for the dummy variable D.

Indicator Variables for Two groups

- Adding an indicator variable causes a parallel shift in the relationship by the amount δ; but slope remains the same

- The least squares estimator's properties are not affected

- We can test the significance of its least squares estimate



Indicator Variables for Two groups

- Suppose houses located near a university command higher price
- Regression equation for house prices: $PRICE = \beta_1 + \delta D + \beta_2 SQFT + e$

$D = 1$ if the house is located near a university
$D = 0$ if if the house is not located near a university

- Use *utown.gdt* from POE 4th ed. [*PRICE* (in \$1000); *SQFT* (in 100)]

$$\widehat{PRICE} = 5.681 + 60.369D + 8.356SQFT$$
s.e. (0.983) (0.186)

$$E(PRICE) = 66.05 + 8.356SQFT \text{ if } D = 1$$
$$E(PRICE) = 5.681 + 8.356SQFT \text{ if } D = 0$$

- Interpretation

$\hat{\delta} = 60.369$: The price of a house is located near a university is \$60369 higher than a house with a similar size but is not located near a university.

$\hat{\beta}_2 = b_2 = 8.365$: The price per **100** square feet is **\$8356** and it is not affected by the location.

Extension to more than Two groups

- If the data fall naturally into *s* subgroups, then *s*-1 dummy variables can be created:
- For example, we want to estimate a model with 3 different locations, rural, urban or city centre
- Include 2 dummy variables (3-1) as D_1 and D_2

$$PRICE = \beta_1 + \gamma D_1 + \theta D_2 + \beta_2 SQFT + e$$

$$D_1 = 1 \text{ and } D_2 = 0 \text{ if location} = \text{urban}$$
$$D_1 = 0 \text{ and } D_2 = 1 \text{ if location} = \text{city centre}$$
$$D_1 = 0 \text{ and } D_2 = 0 \text{ if location} = \text{rural}$$

Location	D1	D2	
Urban	1	0	
City Centre	0	1	
Rural	0	0	

- Note: The value of the intercept β_1 represents the reference group (in this example "rural").

Controlling Time-specific effect with a dummy

- Quiz: To estimate the demand for ice cream for which the demand fluctuates according to the season, need to control for the impact of 4 seasons; Summer, Autumn, Winter and Spring. How many dummy variables would you include in the model?

- Annual indicator variables are used to capture year effects not otherwise measured in a model.

- An indicator variable can be used to control for the potential impact of an unusual event (war, pandemic, change in law, etc.) in a time-series model

Slope Dummy Variable

- Assume that house location affects both the intercept (land value) and the slope (price per square feet), then both effects can be incorporated into a single model:

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + \gamma(SQFT \times D) + e$$

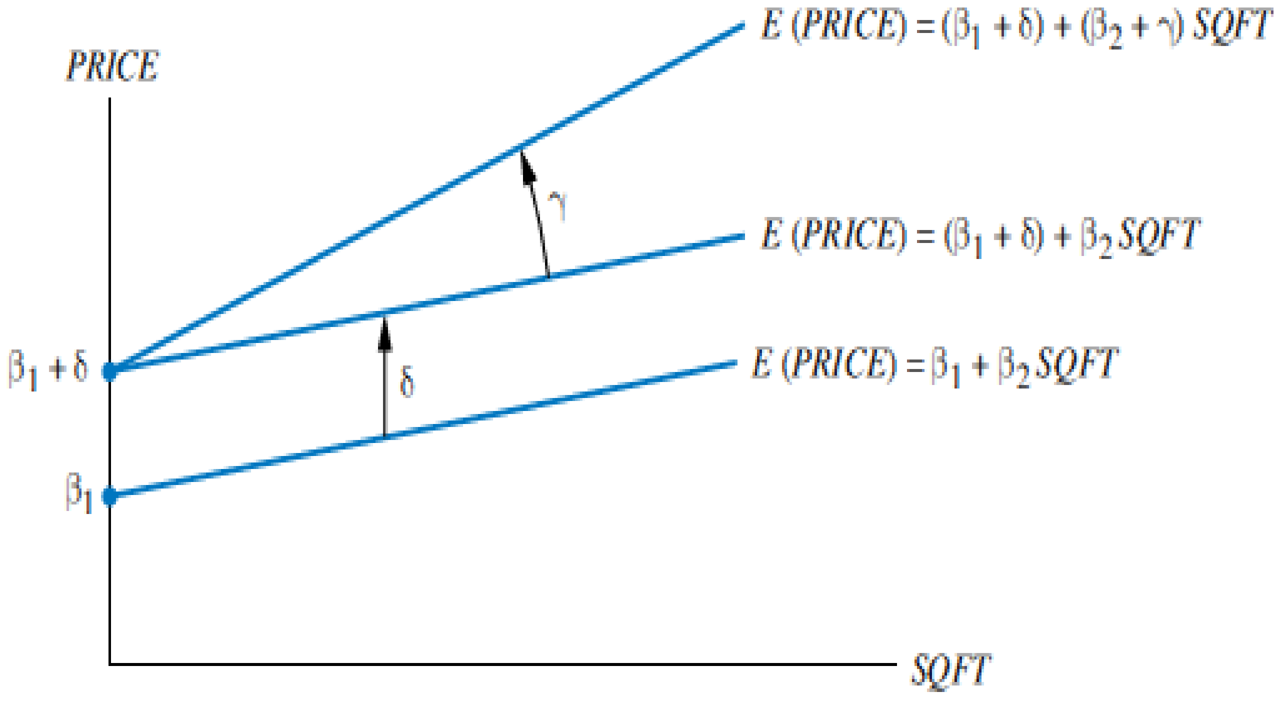
- In this model, the coefficient of dummy represents the difference of prices in two locations (D=0 and D=1) when $SQFT$ =0.
- The new variable ($SQFT$ x D), the product of house size and the indicator variable captures the interaction effect of location and size on house price. Alternatively, it is called a **slope-indicator variable** or a **slope dummy variable**, because it allows for a change in the slope of the relationship

$$E(PRICE) = \begin{cases} (\beta_1 + \delta) + (\beta_2 + \gamma)SQFT & \text{when } D = 1 \\ \beta_1 + \beta_2 SQFT & \text{when } D = 0 \end{cases}$$

Slope Dummy Variable

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + \gamma(SQFT \times D) + e$$

$$E(PRICE) = \begin{cases} (\beta_1 + \delta) + (\beta_2 + \gamma)SQFT & \text{when } D = 1 \\ \beta_1 + \beta_2 SQFT & \text{when } D = 0 \end{cases}$$



Slope Dummy Variable

- Use utown.gdt from POE 4th ed. to estimate:

$$\widehat{PRICE} = 23.062 + 28.124D + 7.664SQFT + 1.28(D \times SQFT)$$

s.e. (8.511) (0.247) (0.335)

- The estimated regression equation for a house near the university is:

$$\widehat{PRICE} = (23.062 + 28.124) + (1.28 + 7.664)SQFT$$

$$\widehat{PRICE} = 51.116 + 8.944SQFT$$

- The estimated regression equation for a house not near the university (D=0) is:

$$\widehat{PRICE} = 23.062 + 7.664SQFT$$

- Interpretation of the estimated coefficient of dummy: The estimated price difference between a plot ($SQFT$=0) near the university (D=1) and not near the university (D=0) is 28.124 (the estimated coefficient of D)

Joint significance of qualitative factors

- We can apply an F-test to check the joint significance of δ and γ (location) in our slope dummy model:
- Unrestricted model:

$$PRICE = \beta_1 + \delta D + \beta_2 SQFT + \gamma(SQFT \times D) + e$$

- Restricted model:

$$PRICE = \beta_1 + \beta_2 SQFT + e$$

$$H_0: \delta = \gamma = 0; H_1: \text{ At least one of them is not zero}$$

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(n-k)}$$

Joint significance of qualitative factors

- We can apply an F-test to check the joint significance of δ and γ in our slope dummy model:
- SSE of Unrestricted model= 236761.6
- SSE of Restricted model= 1149512
- J=2; k=4; n=1000

$$F = \frac{(1149512 - 236761.6)/2}{236761.6/(1000 - 4)} = 1919.86$$

- Using gretl omitted variable test

```
Null hypothesis: the regression parameters are zero for the variables
  D, D_sqft
Test statistic: F(2, 996) = 1919.86, p-value 0
Omitting variables improved 0 of 3 model selection statistics.
```

Summary

After this lecture you should be able to:

- include dummy variables (including slope dummies) in your regression model
- interpret dummy variable coefficients in the regression
- understand the concept of slope-dummy variable
- test the joint significance of dummy variables

University of Nottingham
UK | CHINA | MALAYSIA

Thank you.
Any question?