Business Insights through Data

Chapter 14: Introduction to Multiple Regression part 2

Adapted by
Dr. Nabanita Talukdar

LEARNING OBJECTIVE

- How to use categorical independent variables in a regression model.
- How to use interaction in a regression model.

Adapted by Dr. Nabanita Talukdar

Dummy-Variable Example (with 2 Levels)

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

Let:

Y = pie sales

X_1 = price

X_2 = holiday (X_2 = 1 if a holiday occurred during the week).

(X_2 = 0 if there was no holiday that week).

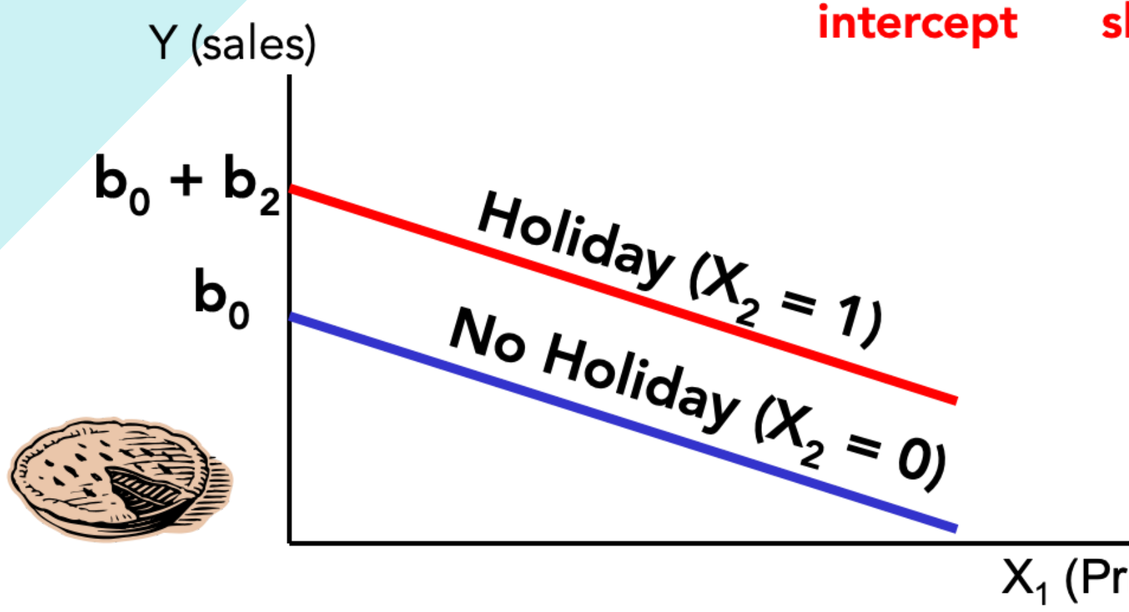
Adapted by Dr. Nabanita Talukdar

Dummy-Variable Example (with 2 Levels)

$$\hat{Y} = b_0 + b_1X_1 + b_2(1) = (b_0 + b_2) + b_1X_1$$

$$\hat{Y} = b_0 + b_1X_1 + b_2(0) = b_0 + b_1X_1$$

Different intercept **Same slope**



If $H_0: \beta_2 = 0$ is rejected, then "Holiday" has a significant effect on pie sales.

Interpreting the Dummy Variable Coefficient (with 2 Levels)

DCOVA

Example: $\widehat{Sales} = 300 - 30(Price) + 15(Holiday)$

Sales: number of pies sold per week

Price: pie price in $

Holiday: 1 If a holiday occurred during the week
0 If no holiday occurred

$b_2 = 15$: on average, sales were 15 pies greater in weeks with a holiday than in weeks without a holiday, given the same price.

Dummy-Variable Models (more than 2 Levels)

DCOVA

- The number of dummy variables is **one less than the number of levels.**
- Example:

  Y = house price ; X_1 = square feet.

- If style of the house is also thought to matter:

  Style = ranch, split level, colonial.

  Three levels, so two dummy variables are needed.

Dummy-Variable Models (more than 2 Levels) (con't)

DCOVA

Example: Let "colonial" be the default category, and let X_2 and X_3 be used for the other two categories:

Y = house price
X_1 = square feet
X_2 = 1 if ranch, 0 otherwise
X_3 = 1 if split level, 0 otherwise

The multiple regression equation is:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

Interpreting the Dummy Variable Coefficients (with 3 Levels)

DCOVA

Consider the regression equation:

$$\hat{Y} = 20.43 + 0.045X_1 + 23.53X_2 + 18.84X_3$$

For a colonial: $X_2 = X_3 = 0$

$$\hat{Y} = 20.43 + 0.045X_1$$

For a ranch: $X_2 = 1; X_3 = 0$

$$\hat{Y} = 20.43 + 0.045X_1 + 23.53$$

With the same square feet, a ranch will have an estimated average price of 23.53 thousand dollars more than a colonial.

For a split level: $X_2 = 0; X_3 = 1$

$$\hat{Y} = 20.43 + 0.045X_1 + 18.84$$

With the same square feet, a split-level will have an estimated average price of 18.84 thousand dollars more than a colonial.

Adapted by Dr. Nabanita Talukdar

Interaction Between Independent Variables

DCOVA

- Hypothesizes interaction between pairs of X variables.
  - Response to one X variable may vary at different levels of another X variable.

- Contains two-way cross product terms.

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

$$= b_0 + b_1X_1 + b_2X_2 + b_3(X_1X_2)$$

Adapted by Dr. Nabanita Talukdar

Effect of Interaction

DCOVA

- Given:

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

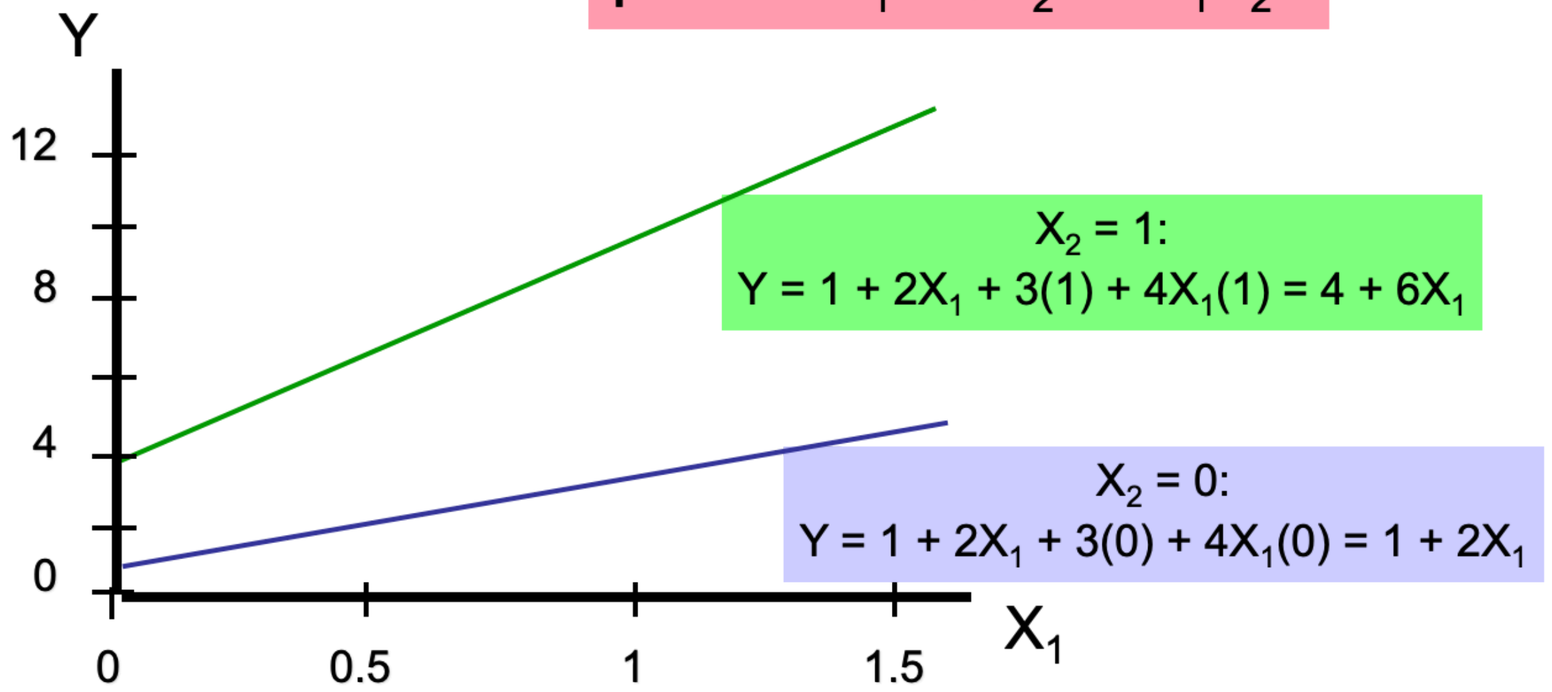
- Without interaction term, effect of X1 on Y is measured by β1.
- With interaction term, effect of X1 on Y is measured by β1 + β3 X2.
- Effect changes as X2 changes.

DCOVA

Interaction Example

Suppose X_2 is a dummy variable and the estimated regression equation is:

$$\hat{Y} = 1 + 2X_1 + 3X_2 + 4X_1X_2$$



Slopes are different if the effect of X_1 on Y depends on X_2 value

SIGNIFICANCE OF INTERACTION TERM

DCOVA

- Can perform a partial F test for the contribution of a variable to see if the addition of an interaction term improves the model.

- Multiple interaction terms can be included:
  - Use a partial F test for the simultaneous contribution of multiple variables to the model.

SUMMARY

- How to use categorical independent variables in a regression model.
- How to use interaction in a regression model.