Interpreting the Coefficients on Dummy Variables

Definition of a Dummy Variable

A dummy variable is an artificial variable constructed such that it takes the value unity (one) whenever the category it represents occurs, and zero otherwise.

Interpretation of the Coefficient on a Dummy Variable, Reference Category Method

Suppose D is a 0-1 variable in the regression model

$$\hat{Y} = b_0 + b_1 D + b_2 X$$

Relative to the regression line for the reference category, where D=0, the line for the category where D=1 is parallel and b_1 units higher.

Interpretation of the Coefficients on Dummy Variables, All Categories Included

Suppose $D_1, D_2, \cdots, D_k$ are a set of mutually exclusive and exhaustive dummy variables for a factor that has *k* categories in the regression model

$\hat{Y} = b_1 D_1 + b_2 D_2 + \cdots + b_k D_k + b_{k+1} X$.

Then each of the coefficients, $b_1, b_2, \cdots, b_k$, is the intercept estimate for the category that the respective dummy variable represents.

Interpretation of the Interaction Coefficient, Reference Category Method

Suppose D is a 0-1 dummy variable, and DX is the product of the dummy variable D and the continuous variable X in the regression model

$$\hat{Y} = b_0 + b_1 D + b_2 X + b_3 DX .$$

Then b_3 is the estimate of the difference in the slope between the two categories, that is, the estimated slope for the reference category where $D = 0$ is b_2, and the estimated slope for the category where $D = 1$ is $b_2 + b_3$.