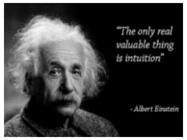
Interaction Models

(Binary x Continuous)

LI(✘) – Relationship between Y and X's are not linear



Interaction Models: Intuition

- Interaction models are very popular in predictive analytics
- An interaction between any two variables x_1 and x_2 is modeled as a multiplicative effect: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$
- Interpreting interaction effects is **straightforward** if one of the variables in the interaction term is a **dummy variable**, e.g., x_2
- But it is **not so easy** or intuitive if both variables are **continuous** (discuss later)
- **Without** the **interaction** term, the equation above is like having two **parallel** regression lines, one when $x_2 = 0$ and the other when $x_2 = 1$ → the effects of x_1 and x_2 are **"additive"**
- But **with** the **interaction** term what we also get two regression lines, but the lines are **no** longer **parallel** → the effects of x_1 and x_2 are **"multiplicative"**
- If **the interaction** effect has the **same sign** as the **main effect**, the interaction **enhances** the main effect; **otherwise** it **offsets** it

A Simple Interaction Model

- Let's illustrate the concept of interaction effect with the earlier example for gas mileage of cars:

$$mpg = \beta_0 + \beta_{Origin}(Origin) + \beta_{Size}(Size) + \beta_{Int}(Origin\ x\ Size) + \varepsilon$$

- β_{Size} and β_{Origin} are called "main effects" which are the same additive effects we saw earlier.
- β_{Int} is the "joint" interaction effect of size and origin together
- As before, $Origin = 0$ for domestic cars and $\mathbf{Origin = 1}$ for foreign cars
- You can think of the model above as two separate regression models:

For domestic cars, $\beta_{Origin} = 0 \rightarrow mpg = \beta_0 + \beta_{Size}(Size) + \varepsilon$

For foreign cars, $\beta_{Origin} = 1 \rightarrow mpg = (\beta_0 + \beta_{Origin}) + (\beta_{Size} + \beta_{Int})(Size) + \varepsilon$

β_0 is the intercept for domestic cars;

$\beta_0 + \beta_{Origin}$ is the intercept for foreign cars

β_{Size} is the main effect of engine size for domestic cars

$\beta_{Size} + \beta_{Int}$ is the main effect of engine size for foreign cars

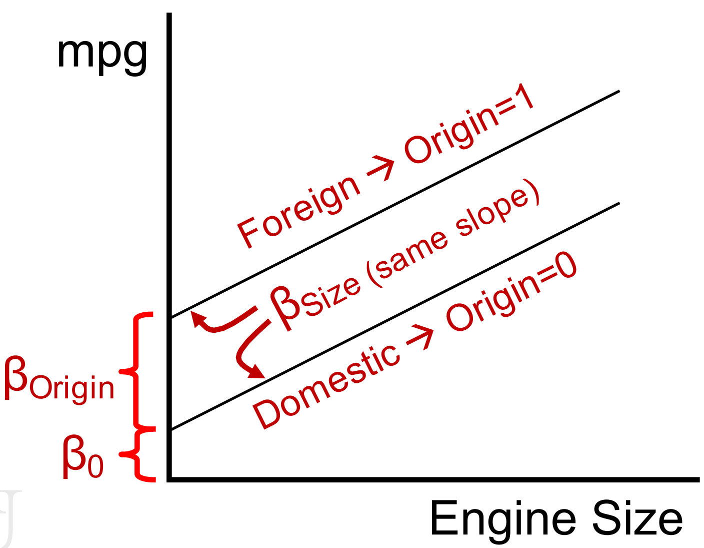
β_{Int} is the change in effect size (change in slope) from domestic to foreign cars – i.e., the interaction effect

Interpretation

- **Without** an **interaction** term, the effects of Size and Origin are **additive**
- So the effect of Origin = 1 is to shift the regression line with the **same slope**
- **With** an **interaction** term, the **slope** of the regression also **changes**
- β_{Size} → The effect of engine size on mileage for domestic cars
- $\beta_{Size} + \beta_{Int}$ → The effect of engine size on mileage for foreign cars
- β_{Int} → The effect of size changes from domestic to foreign

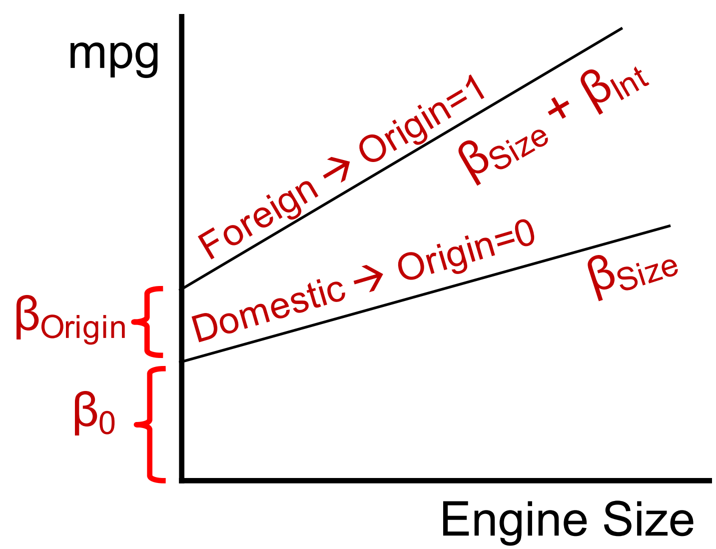
No Interaction Terms

$$mpg = \beta_0 + \beta_{Origin}(Origin) + \beta_{Size}(Size) + \varepsilon$$



With Interaction Term

$$mpg = \beta_0 + \beta_{Origin}(Origin) + \beta_{Size}(Size) + \beta_{Int}(Origin \times Size) + \varepsilon$$



Tips

`lm(y~x1+x2+x1*x2,data=dataName)` → Fits a linear model with 2 main effects and 1 interaction effect → Important!! This model works well for continuous x binary interactions; see next R Tips slide for continuous x continuous interactions

`lm(y~x1*x2,data=dataName)` → Does the same thing; by default this notation automatically models main effects, even thought they are not specified

`lm(y~x1:x2,data=dataName)` → This notation can be used if you wish to omit one or both main effects → Not recommended!! Unless you have a very good reason.

Kogod School
of
Business