

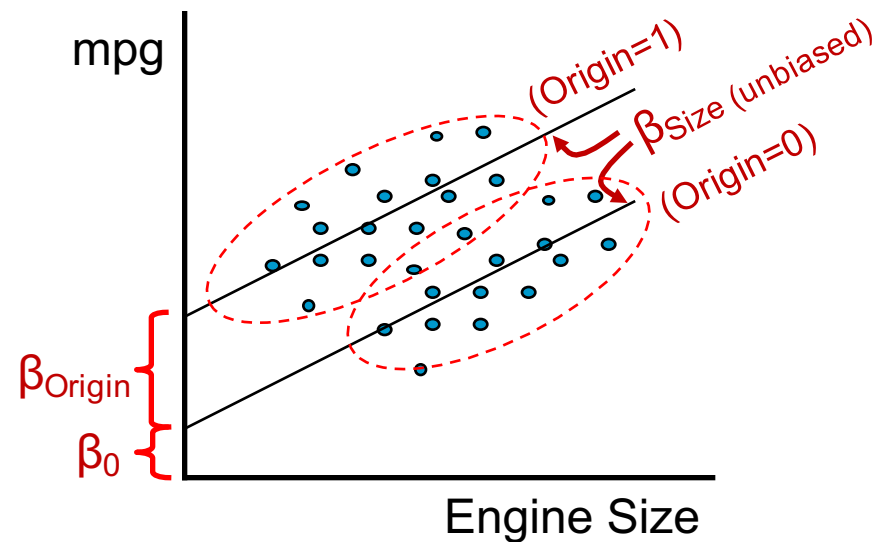
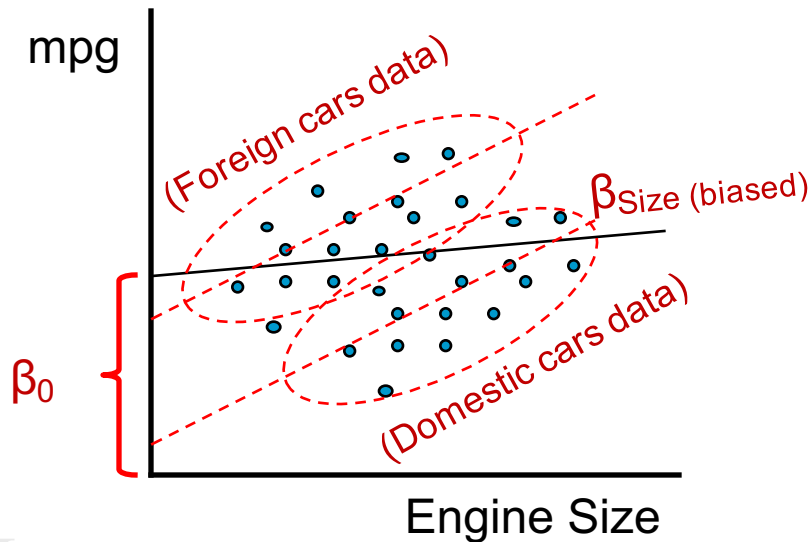
Dummy Variables

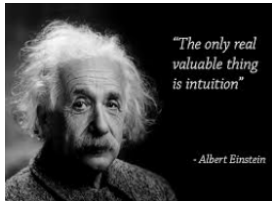
Adding a Binary Variable

- What if a car's engine size is a marginal predictor of mileage (mpg) but you suspect that this effect may be different between domestic and foreign cars?
- In this case, you can add a binary variable "Origin" to the model, which takes the value of 1 if the car is foreign and 0 if it is domestic.
- The model will then look like (X_1 is continuous; X_2 is 0 or 1):

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

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





Dummy Variables: Intuition

- A regression with one or more dummy variables is very common in predictive modeling.
- As the diagram in the previous slide illustrates a dummy variable shows a parallel shift in the regression line. For example, a model that predicts miles per gallon in a vehicle based on the vehicle origin and engine size is:

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

- β_{Size} is the effect of engine size on mileage “controlling for” or “taking into account” vehicle origin”.
- β_{Origin} is the effect of origin on mileage “controlling for” engine size.
- $\text{Origin} = 0$ for domestic cars and $\text{Origin} = 1$ for foreign cars
- Thus, you can think of the model above as two separate regression models:

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

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

So, β_0 is the intercept for domestic cars;

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

β_{Origin} is the effect difference between domestic and foreign cars



KOGOD SCHOOL
of
BUSINESS

