Multi-Variate Regression

Multi-Variate Regression

(or Multiple Linear Regression)

- Most regression models include more than one **"predictor"** variable and possibly more than one **"control"** variable
- Predictor and control variables are ALL **"independent variables"**

$$Y = \beta_0 + \underbrace{\beta_1(X_1) + \beta_2(X_2) + \beta_3(X_3)}_{\text{Predictors}} + \ldots\ldots + \underbrace{\beta_4(X_4) + \beta_5(X_5) + \beta_5(X_5)}_{\text{Controls}} + \varepsilon$$

Independent Variables

- **Interpretation:** β_1 is the **effect size** of X_1, that is, how much Y changes, when X_1 increases by 1 unit, holding X_2 …….X_5 **constant**. Same for β_2, etc.
- Predictors and controls work in the exact same way. They are just labels:
  - A **predictor** is a variable we believe will predict Y
  - A **control** is a variable that may affect Y, so we include it for control

Tips

`lm.fit = lm(y~x1+x2+x3+etc.,data=mtcars)` → Linear regression model

`summary(lm.fit)` → Getting summary data from the linear model

`predict(lm.fit)` → Display predicted values, the training data set is used by default

`predict(lm.fit, newdata=dataName)` → To make predictions using a different test data set

`plot(lm.fit)` → To display various key regression plots

`plot(lstat,medv)` → Plot two variables

`abline(lm.fit)` → Draw a regression line in the plot (useful for simple bi-variate regression only)

KOGOD SCHOOL
of
BUSINESS