Tuesday – April 16:

# Dummy Variables for Multiple Categories:

* A categorical variable can have more than two categories
* Use multiple dummy variables to capture all categories, one for each category
* Reference category – 0
  + Ex. Interpret -1.3 For a single person, they will be expected to have a membership on average 1.3 less months
* Given the intercept term, we can exclude one of the dummy variables form the regression
  + Including all dummy variables, create a dummy variables trap (perfect multicollinearity; more later)
  + ALWAYS make (n-1) less dummy variables than the variables you have
  + Excluded variable represents reference category
* Example: mode of transportation with three categories
  + Public transportation, driving alone, or carpooling
    - Use any two dummy variables: for example
      * D1 = 1 for public transportation, and 0 otherwise
      * D2 = 1 for driving alone and 0 otherwise
      * D1=d2 = 0 indicates car pooling
* ALWAYS be specific about what your variables are

# Model Selection:

Types of “goodness-of-fit” models:

* Standard error of the estimate, se
* Coefficient of determination, R^2
* Adjusted coefficient of determination, adjusted R^2

## Standard Error of the Estimate:

* It is the standard deviation of the residuals
* The sample variance is the average of the squared residuals
* You se n-k-1 since we are using k number of predictor variables because we are estimating all the betas
* The smaller the SSE the better the fit
  + Its hard to interpret the SSE

## R^2:

* Denotes the sample variation in the dependent variable that is explained by the estimated regression equation
  + Closer to 1 🡪 better the fit
* Pros: Very easy to interpret
* Cons: the r^2 number goes down when you increase the number of predictor variables
* When comparing models, you want the model with the highest adjusted R^2

When comparing the three models (from slides) you cant use R^2 since they have different predictor variable numbers so we use Standard Error which tells us that model 3 is more accurate.