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