Lecture 18 and 19 Linear Model Evaluation

* yi is the “true” value from our data set (i.e. xi 𝜷 + ϵ i )
* ŷi is the estimate of yi from our model (i.e. xi 𝜷LS)
* ȳ is the sample mean all yi
* yi - ŷi are the estimates of ϵ i and are referred to as residuals
* A black and white math symbol

  Description automatically generated
* A black and white math equation

  Description automatically generated
* A black text with letters

  Description automatically generated with medium confidence
  + R^2 measures the fraction of variance that is explained by our model (yhat)
* Hypothesis Testing
  + P value is the probability of observing estimates of β at least as extreme as the one observed
  + A p-value smaller than a given threshold would mean the data was unlikely to be observed under H0 so we can reject the hypothesis H0. If not, then we lack the evidence to reject H0.
  + Each parameter of an independent variable x has an associated confidence interval and t-value + p-value
  + If the parameter / coefficient is not significantly distinguishable from 0 then we cannot assume that there is a significant linear relationship between that independent variable and the observations y (i.e. if the interval includes 0 or if the p-value is too large)
* Confidence Intervals
  + Goal: for a given confidence level (let’s say 90%), construct an interval around an estimate such that, if the estimation process were repeated indefinitely, the interval would contain the true value (that the estimate is estimating) 90% of the time.
  + Assume Yi ~ N(5, 25) , for 1 ≤ i ≤ 100
    - yi = μ + ϵ where ϵ ~ N(0, 25).
    - Least Squares estimator of μ (μLS) is the sample mean ȳ
    - What is the 95% confidence interval for μLS?
      * CI.95 = [ȳ - 1.96 x SE(μLS), ȳ + 1.96 x SE(μLS)] = [ȳ - 1.96 x .5, ȳ + 1
      * Where SE(μLS) = σϵ / √n = 5 / √100 = 0.5
* Z value
  + e the number of standard deviations from the mean of a N(0,1) distribution required in order to contain a specific % of values were you to sample a large number of times
  + To find the .95 z-value (the value z such that 95% of the observations lie within z standard deviations of the mean ( μ ± z \* σ )) you need to solve the pdf
    - * A mathematical equation with numbers and symbols

        Description automatically generated