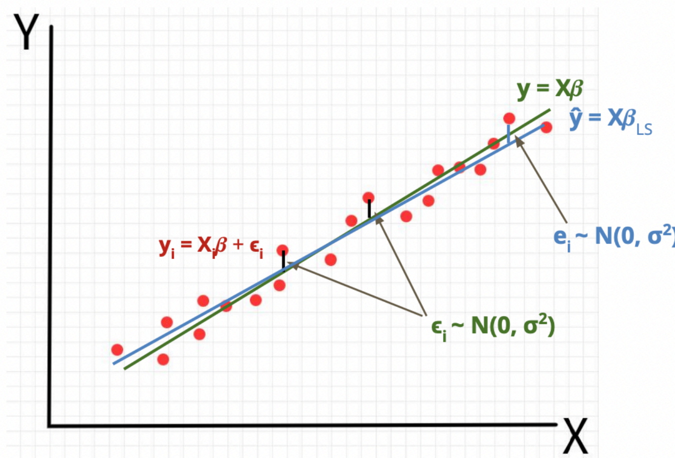
CAS CS 506
Lec 17

Linear Model Evaluation

1. Evaluating Our Regression Model
    a. Some Notation
        i. $\mathbf{y}_i$ is the “true” value from our data set (i.e. $\mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$)

            $\hat{\mathbf{y}}_i$ is the estimate of y_i from our model (i.e. $\mathbf{x}_i\boldsymbol{\beta}_{LS}$)

            $\bar{\mathbf{y}}$ is the sample mean all $\mathbf{y}_i$

            $\mathbf{y}_i - \hat{\mathbf{y}}_i$ are the estimates of $\boldsymbol{\epsilon}_i$ and are referred to as residuals
        ii. 

2. Metric for Evaluation for Fit of Our Model?
    a. Is the value of the loss function sufficient? I.e.

$$\|y - X\beta\|_2^2 = \sum_i^n (y_i - \hat{y}_i)^2$$

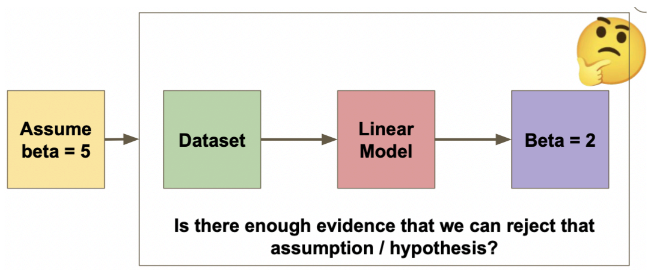
        i. Does not take into account the scales (income is higher and latitude is lower)
3. Evaluating Our Regression Model
    a. $$TSS = \sum_i^n (y_i - \bar{y})^2$$ This is a measure of the spread of y_i around the mean of y
    b. $$ESS = \sum_i^n (\hat{y}_i - \bar{y})^2$$ This is a measure of the spread of our model’s estimates of y_i around the mean of y

c. R^2 = ESS / TSS
   i. It measures the refraction of variance that is explained by our model (y-hat)

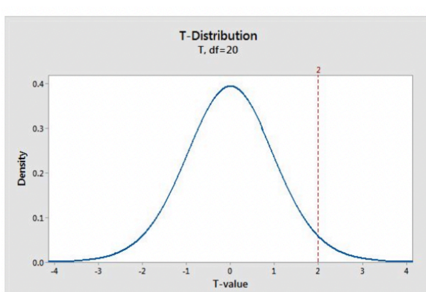
d. $$RSS = \sum_{i}^{n} (y_i - \hat{y}_i)^2$$ This is what our linear model is minimizing

e. $$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

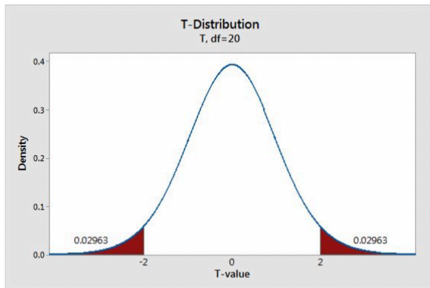
4. Hypothesis Testing

a. 

b. Each parameter of an independent variable x has an associated confidence interval and t-value + p-value
c. 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)
d. We want to know if there is evidence to reject the hypothesis H0: B = 0 (i.e. that there is no linear relation between X and Y) using the information from B hat.
e. We want to know the largest probability of obtaining the data observed, under the assumption that the null hypothesis is correct.
f. How do we obtain that probability?
   i. A:
g. Under the null hypothesis what should be the distribution of the normalized estimates? T-Distribution (parametrized by the sample size)



h. We can then compute the t-value that corresponds to the sample we observed
i. And then compute the probability of observing estimates of B at least as extreme as the one observed (i.e. trying to find evidence against H0)
j. The probability is called a p-value



k. 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.