Prediction interval for an individual y
A prediction interval is an interval estimate of a predicted value of y.

When an x is used to predict $\hat{y}$ from the regression line an interval can be calculated to a confidence interval for y

$$s_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} \tag{29}$$

$$ME = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0-\bar{x})^2}{n(\sum x^2) - (\sum x)^2}} \tag{30}$$

where x_0 denotes the given x value, $t_{\alpha/2}$ has $n-2$ degrees of freedom, s_e.

$$(\hat{y} - ME, \hat{y} + ME) \tag{31}$$

Example 5: What is the best predicted number of people in a household that discards 50 lb of garbage? Use $\alpha = 0.05$ therefore $\alpha/2 = 0.025$ and $s_e = 0.6283$. $df = n - 2 = 62 - 2 = 60$. This implies $t_{\alpha/2} = 2.000$, $n(\sum x^2) - (\sum x)^2 = 4.52$, $\bar{x} = 27.4$.

1. NOTE x value
2. CALCULATE (Point Estimate (PE) $\hat{y}$
3. DETERMINE $t_{\alpha/2}$
4. CALCULATE/NOTE s_e
5. CALCULATE ME
6. $LB = PE - ME$
7. $UB = PE + ME$
8. **Interpretation:** We are $1 - \alpha$ % confident that the true number $\hat{y}$ for x is between LB and UB.

What is the difference between a prediction interval and a confidence interval?

11.2.1 Multiple Linear Regression

Since I was younger I've been making the best out of nothing. - Cameron Jibril Thomaz

The regression line using the population parameters can be seen as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_1 x_{2i} + \beta_1 x_{3i} + \cdots + \beta_j x_{ji}$$

Estimates of regression line:

- β_0: population y-intercept parameter
- β_1: population 1^{st} slope parameter
- β_2: population 2^{nd} slope parameter
- β_3: population 3^{rd} slope parameter
- β_j: population j^{th} slope parameter

The regression line using the sample estimates can be seen as:

$$y_i = b_0 + b_1 x_{1i} + b_1 x_{2i} + b_1 x_{3i} + \cdots + b_j x_{ji}$$

Estimates of regression line:

- b_0: sample y-intercept estimate
- b_1: sample 1^{st} slope estimate
- b_2: sample 2^{nd} slope estimate
- b_3: sample 3^{rd} slope estimate
- b_j: sample j^{th} slope estimate

The Adjusted R^2 - proportion of variance accounted by the model, however, it is modified to account for the number of variables and the sample size.

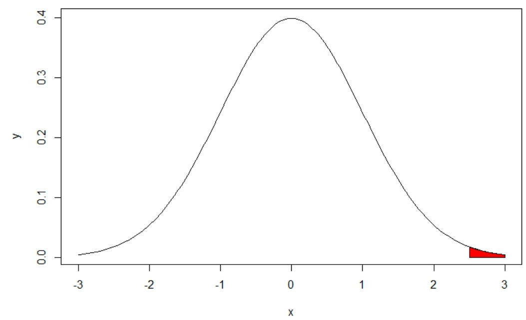
$$\text{adjusted } R^2 = 1 - \frac{(n-1)}{n-k-1}(1-R^2) \tag{32}$$

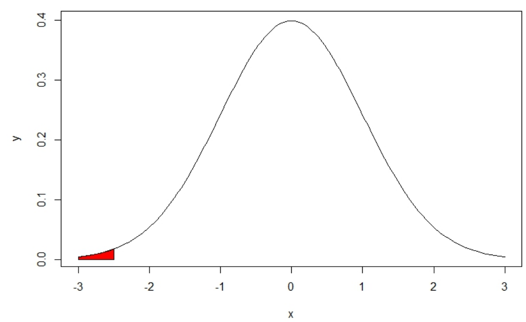
where n is the sample size and k is the number of predictors

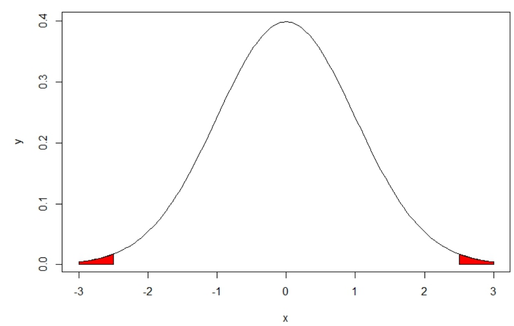
11.2.2 Hypothesis Testing for $\beta's$

Process for Hypothesis Testing for this class:

1. Identify and State the Statistical Question
   - Determine the variable(s) of interest
   - Determine the type variable(s) (i.e., quantitative or qualitative): **slope(s) are always quantitative (in this class)**
   - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand
     - $\boldsymbol{H_0 : \beta_j = 0}$ and $\boldsymbol{H_1 : \beta_j > 0}$
     - $\boldsymbol{H_0 : \beta_j = 0}$ and $\boldsymbol{H_1 : \beta_j < 0}$
     - $\boldsymbol{H_0 : \beta_j = 0}$ and $\boldsymbol{H_1 : \beta_j \neq 0}$
2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true): **will be given to you, if not assume** $\alpha = 0.05$


0.4
0.3
0.2
0.1
0.0
y
-3
-2
-1
0
1
2
3
x


0.4
0.3
0.2
0.1
0.0
y
-3
-2
-1
0
1
2
3
x



Really IMPORTANT:

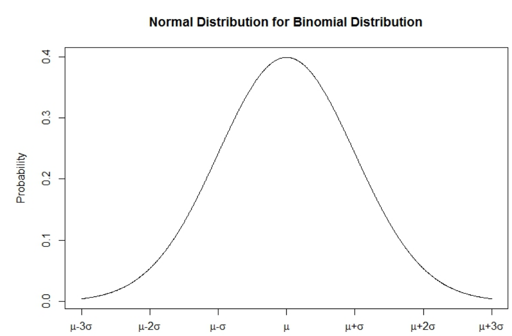
- α:
- $df = n - 2$
- Critical Value:

3. Perform Statistical Test and Interpret Results

$$TS = t = \frac{b_j}{\frac{s_e}{\sqrt{n(\sum x^2) - (\sum x)}}} \tag{33}$$

where

$$s_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}$$



- Test Statistic:
- p-value:

4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms

11.2.3 Confidence Interval for $\beta' s$

1. Define α
2. Find $\alpha/2$
3. Find the critical value ($CV = t_{\alpha/2}$) and $df = n - 2$ that corresponds to $\alpha/2$
4. Standard Error (SE)

$$SE = s_e \sqrt{\frac{1}{n(\sum x^2) - (\sum x)}} \tag{34}$$

5. Find Margin of Error ($ME = SE \text{x} CV$)
6. Lower Bound $LB = PE - ME$
7. Upper Bound $UB = PE + ME$
8. **Interpretation:** We are $1 - \alpha\%$ that the true slope is within this interval. **OR.** We are $1 - \alpha\%$ that the true slope is within the LB and UB.

11.2.4 Assumptions about Regressions

Important Assumptions:

1. There is a linear relationship between x and y, the errors all are near 0 (linear trend in Scatter Plot)
2. The residuals all have the same/constant variance (no pattern in Residual Plot)
3. The residuals are independent from each other
4. The residuals are normally distributed