Simple Linear Regression
Model Assessment

In this lecture

- Simple linear regression
  - Model assessment
  - Identifying significant coefficients in the linear model

Model assessment

- How good is the linear model?
- Which coefficients of the linear model are significant (Identify important variables)
- Can we improve quality of linear model?
  - Are there bad measurements in the data (outliers)

Model summary

```
Call:
lm(formula = BidPrice ~ CouponRate, data = bonds)

Residuals:
   Min      1Q  Median      3Q     Max
-8.249 -2.470 -0.838  2.550 10.515

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
β̂_0 (Intercept)  74.7866     2.8267  26.458  < 2e-16 ***
β̂_1 CouponRate    3.0661     0.3068   9.994 1.64e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.175 on 33 degrees of freedom
Multiple R-squared:  0.7516,    Adjusted R-squared:  0.7441
F-statistic: 99.87 on 1 and 33 DF,  p-value: 1.645e-11
```

First level model assessment

- R^2 value=0.7516
- Hypothesis Testing
  - Coefficients
  - Full model and reduced model

First level model assessment- Hypothesis test on coefficients

- Inorder to check if linear model is good we can check if the estimate $\hat{\beta}_1$ is significant
- Hypothesis Testing
- Null Hypothesis H_0: $\hat{\beta}_1 = 0 \Rightarrow \hat{y}_i = \hat{\beta}_0 + \epsilon_i$ → Reduced Model
- Alternate Hypothesis H_1: $\hat{\beta}_1 \neq 0 \Rightarrow \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i$ → Full Model
- The confidence interval is computed to check if $\hat{\beta}_1$ is significant

First level model assessment- Hypothesis test on coefficients

- Test on $\hat{\beta}_1$ is a two sided test
- At $\alpha = 0.05$ i.e 95% confidence level

```
> alpha=0.05
> n=35
> p=1
> qt(p = 1-(alpha/2),df = n-p-1)
[1] 2.034515
```

- $\hat{\beta}_1 = 3.0661$ and the standard deviation associated is $s_{\hat{\beta}_1} = 0.3068$
- Confidence interval for $\hat{\beta}_1$ is,

```
> 3.0661-(2.034515*0.3068)
[1] 2.441911
> 3.0661+(2.034515*0.3068)
[1] 3.690289
```


First level model assessment- Hypothesis Test on models

- Computing F statistic

$$F_o = \frac{SST-SSE}{SSE/(n-2)} = \frac{SSR}{SSE/(n-2)}$$

$$\text{SSE} = \sum(y_i - \hat{y}_i)^2$$

$$\text{SSR} = \sum(\hat{y}_i - \bar{y})^2$$

```
> SSE<-sum((bonds$BidPrice-bondsmod$fitted.values)^2)
> SSE
[1] 575.3418
> SSR<-sum((bondsmod$fitted.values-mean(bonds$BidPrice))^2)
> SSR
[1] 1741.263
> n=35
> (SSR/SSE)*(n-2)
[1] 99.87401
```

- This F statistic is returned by the summary command

First level model assessment- Hypothesis Test on models

- The F statistic from table for 1 and 33 degrees of freedom is 4.17 at 5% significance level
- The observed value of F statistic is 99.87 which is greater than the theoretical

First level model assessment- Hypothesis test on coefficients

- Conclusion:
- Reject the null hypothesis since the confidence interval does not include 0
- Therefore $\hat{\beta}_1$ is significant