Lesson Learning Objectives

LO 1. The significance of the model as a whole is assessed using an F-test.

- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$
- H_A : At least one $\beta_i \neq 0$
- $df = n - k - 1$ degrees of freedom.
- Usually reported at the bottom of the regression output.

LO 2. Note that the p-values associated with each predictor are conditional on other variables being included in the model, so they can be used to assess if a given predictor is significant, given that all others are in the model.

- $H_0 : \beta_1 = 0$, given all other variables are included in the model
- $H_A : \beta_1 \neq 0$, given all other variables are included in the model
- These p-values are calculated based on a t distribution with $n - k - 1$ degrees of freedom
- The same degrees of freedom can be used to construct a confidence interval for the slope parameter of each predictor:

$$b_i \pm t^{\star}_{n-k-1} SE_{b_i}$$

LO 3. Stepwise model selection (backward or forward) can be done based on p-values (drop variables that are not significant) or based on adjusted R^2 (choose the model with higher adjusted R^2).

LO 4. The general idea behind **backward**-selection is to start with the full model and eliminate one variable at a time until the ideal model is reached.

- p-value method:
1. Start with the full model.
2. Drop the variable with the highest p-value and refit the model.
3. Repeat until all remaining variables are significant.
- adjusted R^2 method:
1. Start with the full model.
2. Refit all possible models omitting one variable at a time, and choose the model with the highest