Lesson 22. Testing Subsets of Predictors – Part 1

1   Overview

- We want to test a subset of predictors <u>simultaneously</u>

- Consider the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \qquad \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

To test...	Hypotheses	Use...
single β_i		
whole model		
subset		

2   The nested F-test

- Suppose

  - we have a multiple linear regression model with k predictors, and
  - we want to test a subset of ℓ predictors

- Formal steps:

  1. State the hypotheses:

  $$H_0 : \beta_i = 0 \quad \text{for all predictors in the subset}$$
 $$H_A : \beta_i \neq 0 \quad \text{for at least one predictor in the subset}$$

  2. Calculate the test statistic:

  - The **full model** is the model with all k predictors
  - The **reduced model** is the model <u>without</u> the ℓ predictors being tested
  - The SSE terms can be found in the ANOVA tables (see Lesson 16)

3. Calculate the p-value:

- o If the conditions for multiple linear regression hold, then the the test statistic F follows $F(\ell, n-(k+1))$

$$\Rightarrow p\text{-value} = \boxed{}$$

$f_{F(\ell, n-(k+1))}(x)$

x
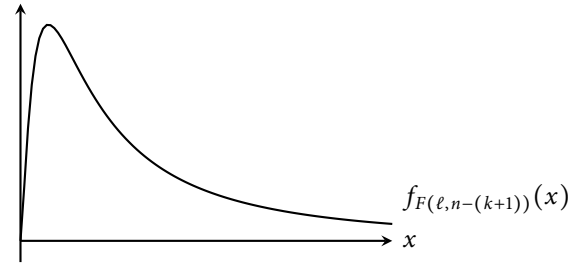
4. State your conclusion, based on the given significance level α:

If we reject H_0 (p-value $\leq \alpha$):

We see significant evidence that including <u>subset of predictors</u> improves the model.

If we fail to reject H_0 (p-value $> \alpha$):

We do not see significant evidence that including <u>subset of predictors</u> improves the model.