Lesson 12. ANOVA for Simple Linear Regression – Part 1

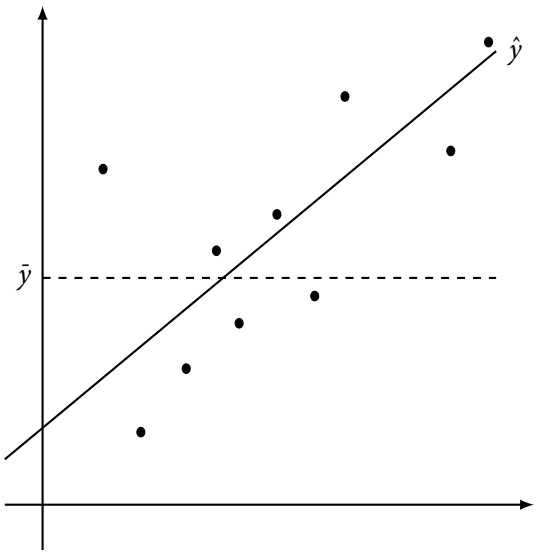
Note. In Part 2 of this lesson, you can run the R code that generates the outputs here in Part 1.

1 Overview

- Main question: is the overall model effective?
  - For simple linear regression, this is just another test of whether $\beta_1 = 0$
  - *Looking ahead...* when we have more than one predictor, we will test whether all the predictor coefficients equal 0, versus at least one of them not being 0
- Approach: compare the amount of variability in the response explained by the model to the amount of variability not explained by the model

2 Sums of squares

- The ANOVA sum of squares identity:



  - Total sum of squares:
  - Model sum of squares:
  - Error sum of squares:

3 The ANOVA table for simple linear regression

Source	df	Sum of Squares	Mean Square	F-Statistic
Model				
Error				
Total				

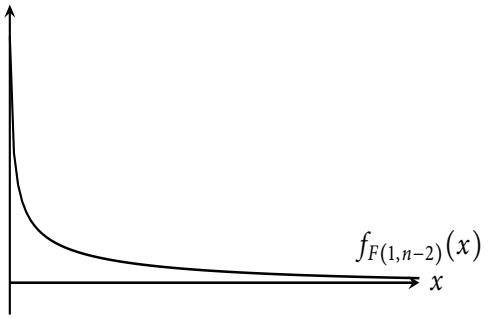
4 The ANOVA F-test for simple linear regression

- Question: **Is the model as a whole effective?**
- Formal steps:
  1. State the hypotheses:

  2. Calculate the test statistic:

  3. Calculate the p-value:
     - If the conditions for simple linear regression hold, then the test statistic F follows

     $\Rightarrow$ p-value =



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

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

     We reject H_0 because the p-value is less than the significance level α. We see significant evidence that the model is effective overall.

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

     We fail to reject H_0 because the p-value is greater than the significance level α. We do not see significant evidence that the model is effective overall.

Example 1. Let's continue looking at the `AccordPrice` data and use the ANOVA table that R outputs to test the effectiveness of the simple linear model predicting `Price` from `Mileage`.

Suppose we run the following R code:

```
fit <- lm(Price ~ Mileage, data = AccordPrice)
anova(fit)
```

The output is below:

A anova: 2 × 5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
Mileage	1	687.6644	687.664433	72.25284	3.055011e-09
Residuals	28	266.4892	9.517473	NA	NA

Use the output above to conduct the appropriate test.

5 Other things to note

- For simple linear regression, there is an equivalence between this ANOVA F-test and the t-test for slope:

- When we have more than one predictor, these tests will have different purposes