Lesson 11. ANOVA for Simple Linear Regression – Part 1

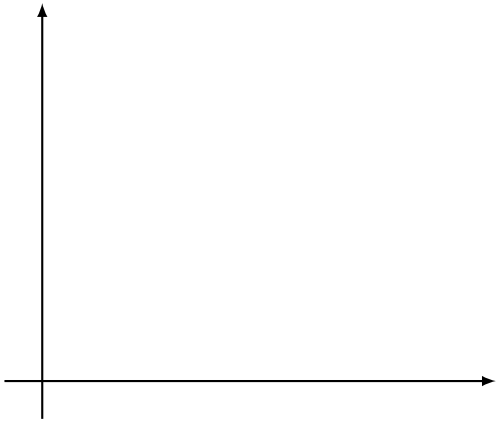
Note. In Part 2 of this lesson, you can run the R code that generates the outputs in here.

1 Overview

- Our main question: is the overall model effective?
  - For simple linear regression, this is just another test of whether $\beta_1 = 0$
  - 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
- Our approach: compare the amount of variability in the response explained by the model to the amount of variability not explained by the model

2 Sum of squares

- The **ANOVA sum of squares identity**:



  - Sum of squares total:
  - Sum of squares model:
  - Sum of squares error:

3 The ANOVA table for simple linear regression

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

4 The ANOVA test for simple linear regression

1. State the hypotheses

2. Calculate the test statistic

3. Calculate the p-value

   - If the conditions for simple linear regression hold, then the sampling distribution of the test statistic under H_0 is

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

   - Provide your **decision**: e.g., reject H_0, fail to reject H_0
   - State your **conclusion** in terms of evidence: e.g.,
     - We see significant evidence that the model is effective overall
     - We do not see significant evidence that the model is effective overall

Example 1. Let's continue looking at the `PorschePrice` 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 = PorschePrice)
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	5565.685	5565.68453	108.2543	3.981734e-11
Residuals	28	1439.565	51.41304	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 test and the t-test for slope:

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