Hypothesis Testing in Linear Regression Models

The key test is the F-**test**. Compare two nested models

- H_0: reduced model with p_0 coefficients;
- H_a: full model with p_a coefficients.

Nested: if the reduced model is a special case of the full model, e.g.,

$$H_0 : Y \sim X_1 + X_2, \quad H_a : Y \sim X_1 + X_2 + X_3.$$

Note that $\text{RSS}_a < \text{RSS}_0$ and $p_a > p_0$.

F-test

Test statistic:

$$F = \frac{(\text{RSS}_0 - \text{RSS}_a)/(p_a - p_0)}{\text{RSS}_a/(n - p_a)},$$

which $\sim F_{p_a - p_0, n - p_a}$ under the null.

- Numerator: variation (per dim) in the data not explained by the reduced model, but explained by the full model, i.e., evidence supporting H_a.
- Denominator: variation (per dim) in the data not explained by either model, which is used to estimate the error variance.

Reject H_0, if F-stat is large, i.e., the variation missed by the reduced model, when being compared with the error variance, is significantly large.

Special Cases of the F-test

- The so-called t-test for each regression parameter (see the `R` output) is a special case of F-test. For example, the test for the j-th coef β_j compares
  - $H_0 : Y \sim 1 + X_1 + \cdots + X_{j-1} + \qquad X_{j+1} + \cdots + X_p$
  - $H_a : Y \sim 1 + X_1 + \cdots + X_{j+1} + X_j + X_{j+1} + \cdots + X_p$
- The overall F-test (at the bottom of the `R` output) compares
  - $H_0 : Y \sim 1$
  - $H_a : Y \sim 1 + X_1 + \cdots + X_{j+1} + X_j + X_{j+1} + \cdots + X_p$