General F-Test

- Tests for k linear restrictions on the slope coefficients of a multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \epsilon$$

- $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$ v.s. H_A : at least one $\beta_i \neq 0$
- If the null hypothesis is true, it implies the following model: $Y = \beta_0 + \epsilon$
- Since the restricted and unrestricted models have the same dependent variable, the SST is the same for both models.
- For the restricted model, because there are no regressors, the SSR is zero and the SSE is equal to the SST.
- The F-statistic can be calculated as follows:

$$F = \frac{n-k-1}{k}\frac{SSR}{SSE} = \frac{n-k-1}{k}\frac{SST-SSE}{SSE} = \frac{n-k-1}{k}\frac{SSE_r-SSE}{SSE}$$

Checking for Multicollinearity

- One of the assumptions of the multiple linear regression model is that the independent variables are not perfectly correlated with each other. (LR5)

- If the independent variables are highly correlated with each other, it can lead to unreliable estimates of the coefficients and inflated standard errors.

- To check for multicollinearity, there are 3 rules of thumb:

  1. If the R^2 is high but only few independent variables are significant with logical signs.
  2. If the correlation was strong between the independent variables or if it is stronger than the correlation between the independent variables and the dependent variable.
  3. If one or several variance inflation factors (VIF) are greater than 5.