This lecture F - test.

① last lecture : overall F-test

for response $y \in \mathbb{R}$ (one obs. with) and predictors $x_1, \dots, x_p \in \mathbb{R}$

$$H_0: E[y|x] = \beta_0 \quad \text{v.s} \quad H_A: E[y|x] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

② General setting : H_0: simpler model, H_A: more complex model] H_0 model is a submodel of the H_A model

For instance

$$H_0: E[y|x] = \beta_0 + \beta_1 x_1 \iff \beta_2 = \beta_3 = 0$$

$$H_A: E[y|x] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta x_3$$

Step 1 : Fit H_0 model, get RSS_{H_0} and df_{H_0}

$$RSS_{H_0} \vee RSS_{H_A}$$

Step 2 : Fit H_A model, get RSS_{H_A} and df_{H_A}

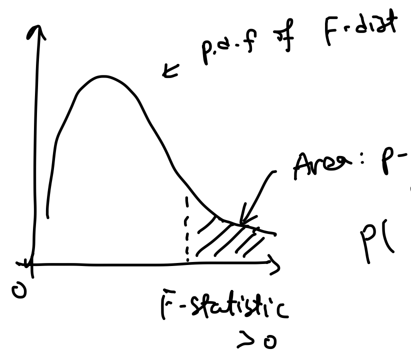
Step 3 :

$$F\text{-statistic} = \frac{(RSS_{H_0} - RSS_{H_A}) / (df_{H_0} - df_{H_A})}{RSS_{H_A} / df_{H_A}}$$

$$df_{H_0} - df_{H_A} = (n - \#\beta \text{ under } H_0) - (n - \#\beta \text{ in } H_A) = \#\beta \text{ under } H_A - \#\beta \text{ under } H_0$$

$$RSS_{H_A} / df_{H_A} = \hat{\sigma}^2_{HA}$$ estimated σ^2 under the more complex model (H_A)

under H_0, normally distributed statistical errors: F-distribution with d.f.s $(df_{H_0} - df_{H_A}, df_{H_A})$



Area: p-value = $P(\text{F-dist}(df_{H_0} - df_{H_A}, df_{H_A}) > \text{F-statistic computed from data})$

★ Reject H_0 if p-value $< \alpha$ (e.g. 5%)
Accept H_0 otherwise.

In R, "anova" function can be used to compare two nested models by F-test.

Examples.

(3) Analysis of variance Tables (Anova)

(3.1) Type I Anova : use anova (model) in R

(Sequential tests)

Example 2