Lecture Summary: Feb. 6, 2023

- F-test for lack-of-fit

Regression model: $Y_{ij} = \beta_0 + \beta_1 x_j + \epsilon_{ij}$, $j = 1, \ldots, c, i = 1, \ldots, n_j$, where x_j is the jth value of x, c is the number of different x values, and $Y_{ij}, i = 1, \ldots, n_j$ are the Y values corresponding to the same x_j.

Full model: $Y_{ij} = \mu_j + \epsilon_{ij}$, $j = 1, \ldots, c, i = 1, \ldots, n_j$.

F-statistic:

$$F = \frac{\mathrm{SSE}(R) - \mathrm{SSE}(F)}{\mathrm{df}_R - \mathrm{df}_F} \left\{ \frac{\mathrm{SSE}(F)}{\mathrm{df}_F} \right\}^{-1},$$

where

$$\mathrm{SSE}(R) = \sum_j \sum_i (Y_{ij} - \hat{Y}_{ij})^2$$

with $\hat{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_j$,

$$\mathrm{SSE}(F) = \sum_j \sum_i (Y_{ij} - \hat{\mu}_j)^2$$

with $\hat{\mu}_j = \bar{Y}_{\cdot j} = n_j^{-1} \sum_{i=1}^{n_j} Y_{ij}$, $\mathrm{df}_R = n - 2$ with $n = \sum_{j=1}^{c} n_j$, and $\mathrm{df}_F = n - c$.

Under H_0: The assumed model is correct, $F \sim F_{c-2, n-c}$.