Lecture Summary: Jan. 27, 2023

- Analysis of Variance

1. Decomposition of SS:

$$\text{SSTO} = \text{SSR} + \text{SSE},$$

where

$$\begin{aligned}\text{SSTO} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2,\\ \text{SSR} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2,\\ \text{SSE} &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.\end{aligned}$$

2. Degrees of freedom

The d.f. of SSTO is $n-1$; the d.f. of SSR is 1; the d.f. of SSE is $n-2$, so that $n-1 = 1 + n - 2$.

3. MS: SS divided by its d.f.

$$\text{MSR} = \frac{\text{SSR}}{1} = \text{SSR}, \quad \text{MSE} = \frac{\text{SSE}}{n-2}.$$

4. F-statistic

$$F = \frac{\text{MSR}}{\text{MSE}}.$$

The distribution of F under the null hypothesis $\text{H}_0 : \beta_1 = 0$ is $F_{1,n-2}$.

5. ANOVA table

Source	SS	d.f	MS	F
Regression	SSR	1	MSR	F
Error	SSE	$n-2$	MSE	
Total	SSTO	$n-1$		

- Coefficient of determination:

R^2: A measure of goodness of fit, which is the proportion of variation in Y explained by the regression (i.e., by x).

$$R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSTO}}.$$

- Properties of R^2:

1. $0 \leq R^2 \leq 1$.
2. $R^2 \approx 1$ if there is a strong linear association between x and Y.
3. $R^2 \approx 0$ if there is a weak or no linear association between x and Y.
4. Both R^2 is a measure of linear association only.