Lecture Summary: March 1, 2023

• Qualitative predictor

If x is a qualitative predictor with k categories, one can define $k-1$ indicator variables as follows:

$$x_1 = \begin{cases} 1, & \text{if category 1,} \\ 0, & \text{otherwise;} \end{cases}$$
$$x_2 = \begin{cases} 1, & \text{if category 2,} \\ 0, & \text{otherwise;} \end{cases}$$
$$\ldots$$
$$x_{k-1} = \begin{cases} 1, & \text{if category } k-1, \\ 0, & \text{otherwise;} \end{cases}$$

Note: one only needs $k-1$, not k, indicator variables (otherwise, there will be multicollinearity).

• Interaction

Examples:

Model 1 (no interaction): $\mathrm{E}(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, x_1 quantitative, x_2 indicator corresponding to a qualitative predictor.

If $x_2 = 0$, $\mathrm{E}(Y) = \beta_0 + \beta_1 x_1$;

If $x_2 = 1$, $\mathrm{E}(Y) = (\beta_0 + \beta_2) + \beta_1 x_1$.

Model 2 (interaction):

$\mathrm{E}(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, x_1, x_2 as above.

If $x_2 = 0$, $\mathrm{E}(Y) = \beta_0 + \beta_1 x_1$;

If $x_2 = 1$, $\mathrm{E}(Y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1$.

• Regression variable selection: Introduction

Reasons for variable selection.

Interpretation of relationship.

Trade-off between model fit and model complexity.

- Adjusted R^2

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

cannot be directly used in variable selection, because, if so, "larger model always wins".