Lecture Summary: March 6, 2023

• Forward stepwise selection:

1. Choose the first predictor, $x_{(1)}$, that has the largest $|t|$ for the slope under a simple linear regression with the predictor.

2. Choose the second predictor, $x_{(2)}$, that has the largest $|t|$ for the coefficient under a linear regression with $x_{(1)}$ and a new predictor.

3. And so on.

4. Stopping rule: The p-value needs to be at most 0.10 to add a predictor.

5. Every time a new predictor is added, need to check to make sure that all of the existing predictors have p-value at most 0.15; otherwise, drop the existing predictor(s).

• Conditional residual plots

$e(Y|x_2)$—residual of fitting Y against x_2;

$e(x_1|x_2)$—resisual of fitting x_1 against x_2.

Similarly define $e(Y|x_1), e(x_2|x_1)$.

Sometimes, plotting $e(Y|x_2)$ against $e(x_1|x_2)$, or plotting $e(Y|x_1)$ against $e(x_2|x_1)$ can reveal some parttern of departure from model assumptions.

For example, a linear pattern in the plot of $e(Y|x_1)$ against $e(x_2|x_1)$ may suggest that an important predictor, x_2, is missing in the simple linear regression of $Y = \beta_0 + \beta_1 x_1 + \epsilon$.

• Studentized residual: Let $\hat{\epsilon}_i$ denotes the residual. The studentized residual is defined as

$$r_i = \frac{\hat{\epsilon}_i}{\text{s.e.}(\hat{\epsilon}_i)} = \frac{\hat{\epsilon}_i}{\sqrt{\text{MSE}(1 - h_{ii})}},$$

where h_{ii} is the ith diagonal element of $H = X(X'X)^{-1}X'$, called the hat matrix; h_{ii} is called the leverage for the ith case.