Lecture Summary: March 8, 2023

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.

• Identifying outlying Y observations

Deleted (jackknife) residual:

Fit the regression with the ith case deleted; let $\hat{Y}_{i(-i)}$ denote the predicted value for Y_i, under this regression.

$d_i = Y_i - \hat{Y}_{i(-i)}$ is called the deleted residual.

Studentized deleted residual:

$$t_i = \frac{d_i}{\text{s.e.}(d_i)},$$

where

$$\text{s.e.}(d_i) = \sqrt{\frac{\text{MSE}_{(i)}}{1 - h_{ii}}},$$

$\text{MSE}_{(i)}$ is the MSE of fitting regression without the ith case. Another expression:

$$\text{MSE}_{(i)} = \frac{(1 - h_{ii})\text{SSE} - \hat{\epsilon}_i^2}{(n - p - 1)(1 - h_{ii})} = \frac{n - p}{n - p - 1}\text{MSE} - \frac{\hat{\epsilon}_i^2}{(n - p - 1)(1 - h_{ii})}.$$

Under the null hypothesis that there are no outliers, $t_i \sim t_{n-p-1}$.

• Bonferroni method for obtaining critical value for studentized deleted residual: Reject H_0: no outliers, if

$$\max_{1 \le i \le n} |t_i| > t_{n-p-1}\left(1 - \frac{\alpha}{2n}\right),$$

where p is the number of β's (= number of predictors +1).

Example (Body fat): $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$, $i = 1, \ldots, 20$. So, $p = 3, n = 20$. Suppose $\alpha = 0.10$. $t_{n-p-1}(1 - \alpha/2n) = t_{16}(1 - 0.0025) = t_{16}(0.9975) = 3.25$; $\max_{1 \leq i \leq 20} |t_i| = 1.825$. So, do not reject H_0 and conclude that there is no outlier.

• Identifying outlying x observations

Recall h_{ii} is the ith diagonal element of the hat matrix $H = P_X$, which is called the leverage for the ith case.

A proporty: $\sum_{i=1}^{n} h_{ii} = p$, the number of β's.

If $h_{ii} > 2\bar{h} = 2p/n$, case i is considered outlying in x.