Monday, 26 August 2019 5:04 PM

(Previous class not attended)

$DFBETA_{j(i)}$

difference in Betas

β_j when the i^{th} observation (y_i, x_i^T) is not present in the data and the estimate of β_j when the whole data is used.

Belsley, Kuh, Welsch (1980)

$\hat\beta_j - \hat\beta_{j(i)} \sim ?$

$$l^T(\hat\beta - \hat\beta_{(i)}) = \hat\beta_j - \hat\beta_{j(i)}$$

$$l^T = (0\ 0\ 0 \cdots \underset{j+1}{1}\ 0 \cdots 0)$$

$$\hat\beta - \hat\beta_{(i)} = \frac{(X^TX)^{-1}x_i e_i}{1-h_{ii}} \quad — (1)$$

Denote $R = (X^TX)^{-1}X^T$.

$$\Rightarrow RR^T = ((X^TX)^{-1}X^T)(X(X^TX)^{-1}) = (X^TX)^{-1} = (C_{ij})$$

$$\hat\beta \sim N(\beta, \sigma^2 C)$$

$$l^T(\hat\beta - \hat\beta_{(i)}) = \frac{r_{j,i}e_i}{(1-h_{ii})} = \hat\beta_j - \hat\beta_{j(i)}$$

r_{ji} is the j,i^{th} element of R.

$$V(l^T(\hat\beta - \hat\beta_{(i)})) = (\sigma^2 r_j^T r_j)(1-h_{ii}) = \sigma^2 C_{jj}(1-h_{ii}) \text{ (Check).}$$

$$\frac{l^T(\hat\beta - \hat\beta_{(i)})}{\sqrt{\sigma^2 C_{jj}(1-h_{ii})}} = \frac{\hat\beta_j - \hat\beta_{j(i)}}{\sqrt{\sigma^2 C_{jj}(1-h_{ii})}}$$

If we use the estimated value of σ^2 and $S_{(i)}^2$ then the estimate will be

$$\frac{\hat\beta_j - \hat\beta_{j(i)}}{\sqrt{S_{(i)}^2 r_j^T r_j}} = \frac{\hat\beta_j - \hat\beta_{j(i)}}{\sqrt{C_{jj} S_{(i)}^2}}$$

$$\frac{\hat\beta_j - \hat\beta_{j(i)}}{\sqrt{r_j^T r_j (S_{(i)}^2)}} = \left(\frac{r_{ji} e_i}{(1-h_{ii})}\right) \Big/ \sqrt{r_j^T r_j S_{(i)}^2}$$

$$\frac{r_{j,i}}{\sqrt{r_j^T r_j S_{(i)}^2}} \frac{e_i}{\sqrt{1-h_{ii}}} \frac{1}{\sqrt{1-h_{ii}}}$$

$$= \left(\frac{r_{ji}}{\sqrt{r_j^T r_j}}\right)\underbrace{\left(\frac{e_i}{\sqrt{S_{(i)}^2(1-h_{ii})}}\right)}_{\text{t-statistic}}\left(\frac{1}{\sqrt{1-h_{ii}}}\right)$$

$$= \frac{r_{ji}}{\sqrt{r_j^T r_j}}\, t_i \frac{1}{\sqrt{1-h_{ii}}}$$

We can consider there is an influence point if $|DFBeta_{j(i)}| > \frac{2}{\sqrt n}$

$$DFFits_{(i)} = \frac{\hat y_i - \hat y_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}} = \frac{x_i^T(\hat\beta - \hat\beta_{(i)})}{\sqrt{S_{(i)}^2 h_{ii}}}$$

$$= \frac{x_i^T(X^TX)^{-1}x_i e_i}{(1-h_{ii})\sqrt{S_{(i)}^2 h_{ii}}}$$

$$= \frac{h_{ii}}{(1-h_{ii})}\frac{e_i}{\sqrt{S_{(i)}^2 h_{ii}}} = \left(\frac{h_{ii}}{1-h_{ii}}\right)\left(\frac{e_i}{\sqrt{S_{(i)}^2(1-h_{ii})}}\right)$$

$$= \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} t_i$$

$|DFFits| > 2\frac{(K+1)}{n}$ then it is an influence point

General Variance of β and β.

$$\text{Cov Ratio} = \frac{|Var(\hat\beta_{(i)})|}{|Var(\hat\beta)|} = \frac{|\hat\sigma^2(X_{(i)}^T X_{(i)})^{-1}|}{|\hat\sigma^2(X^TX)^{-1}|}$$

$$= \frac{|S_{(i)}^2(X_{(i)}^T X_{(i)})^{-1}|}{|MSRes(X^TX)^{-1}|}$$

$$= \left[\frac{S_{(i)}^2}{MSRes}\right]^{k+1} \frac{|(X_{(i)}^T X_{(i)})^{-1}|}{|(X^TX)^{-1}|}$$

$$|A + bc^T| = |A|(1 + c^TA^{-1}b)$$

$$\Rightarrow \frac{|X^TX|}{|X^TX|(1 - x_i^T(X^TX)^{-1}x_i)} = \frac{1}{1-h_{ii}}$$

$$\therefore \left(\frac{S_{(i)}^2}{MSRes}\right)^{k+1}\frac{1}{1-h_{ii}}$$