Thursday, 22 August 2019 12:10 PM

$$e_i = y_i - \hat{y}_i \sim N(0, \sigma^2(1-h_{ii}))$$

$$\Rightarrow \frac{e_i}{\sqrt{\sigma^2(1-h_{ii})}} = \frac{e_i}{\sqrt{V(e_i)}} \sim N(0,1)$$

$$e_{(i)} = \frac{e_i}{1-h_{ii}} \sim N\left(0, \frac{\sigma^2}{1-h_{ii}}\right)$$

$$\Rightarrow \frac{e_i}{\sqrt{V(e_{(i)})}} = \frac{e_i}{\sqrt{\sigma^2(1-h_{ii})}} = \frac{e_i}{\sqrt{V(e_i)}}$$

(Even if we leave one out we get the same expression but there are some differences)

$$(\underset{n\times n}{A} + \underset{n\times 1}{u}\,\underset{(n\times 1)^T = 1\times n}{v^T})^{-1} = A^{-1} - \frac{A^{-1}u v^T A^{-1}}{1+v^T A^{-1} u}$$

X matrix when i^{th} row is removed $\equiv X_{(i)}$.

$$[X^T_{(i)} X_{(i)}]^{-1} = (X^T X - x_i x_i^T)^{-1} \quad — ①$$

$$X^T Y = X_{(i)}^T Y_{(i)} + x_i y_i$$

$$\Rightarrow X_{(i)}^T Y_{(i)} = X^T Y - x_i y_i$$

$$\begin{aligned}
e_{(i)} &= y_i - \hat{y}_{(i)} \\
&= y_i - x_i^T \hat{\beta}_{(i)} \\
&= y_i - \underset{1\times d}{x_i^T} \underset{d\times n-1\ \ n-1\times d}{[X_{(i)}^T X_{(i)}]^{-1}} \underset{d\times}{X_{(i)}^T Y_{(i)}} \\
&= y_i - x_i^T\left[(X^TX)^{-1} + \frac{(X^TX)^{-1} x_i x_i^T (X^TX)^{-1}}{1 - x_i^T (X^TX)^{-1} x_i}\right] X_{(i)}^T Y_{(i)} \\
&= y_i - x_i^T\left[(X^TX)^{-1} + \frac{(X^TX)^{-1} x_i x_i^T (X^TX)^{-1}}{1 - h_{ii}}\right] X_{(i)}^T Y_{(i)} \\
&= \frac{\left[(1-h_{ii}) y_i - x_i^T (X^TX)^{-1} X_{(i)}^T Y_{(i)} (1-h_{ii}) - (x_i^T (X^TX)^{-1} x_i) x_i^T (X^TX)^{-1} X_{(i)}^T Y_{(i)}\right]}{(1-h_{ii})} \\
&= \left[\frac{(1-h_{ii}) y_i - x_i^T (X^TX)^{-1} X_{(i)}^T Y_{(i)}}{1-h_{ii}}\right] \\
&= \frac{(1-h_{ii}) y_i - x_i^T (X^TX)^{-1} [X^T Y - x_i y_i]}{1-h_{ii}} \\
&= \frac{(1-\cancel{h_{ii}}) y_i - x_i^T \hat{\beta} + \cancel{h_{ii} y_i}}{1-h_{ii}} \\
&= \left(\frac{y_i - \hat{y}_i}{1-h_{ii}}\right) = \frac{e_i}{1-h_{ii}}
\end{aligned}$$

When all data are used then

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n-k-1} = MS_{Res}$$

When the i^{th} observation is removed then

$$\hat{\sigma}^2 = S_{(i)}^2 = \frac{(n-k-1) MS_{RES} - \frac{e_i^2}{1-h_{ii}}}{n-k-2}$$

For proof see C.8 Montgomery

Test that i^{th} observation is an outlier

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$

$$y_i - \hat{y}_i = e_{(i)} \sim N\left(0, \frac{\sigma^2}{1-h_{ii}}\right)$$

$$\frac{e_i}{\sqrt{\sigma^2/(1-h_{ii})}} \sim N(0,1)$$

$$\Rightarrow \frac{e_{(i)}}{\sqrt{S_{(i)}^2/(1-h_{ii})}} \sim t_{n-k-2}$$