Problem 1
Read the following derivation with regard to the deleted residuals and then proceed with solving Problems 2 and 3.

Solution:

Problem 2
Studentized deleted residuals. In the following, no assumption is made on the data or the model unless it is explicitly stated.

(a) Assume the observed response vector $Y \in \mathbb{R}^n$ has $Var(Y) = \sigma^2 I_n$. Show that, the ?? ith deleted residual $d_i = Y_i - \hat{Y}_{i(i)}$ has

$$Var(d_i) = \frac{\sigma^2}{1 - h_{ii}}$$

(b) Let

$$SSE_{(i)} = \sum_{j:j \neq i} (Y_j - \hat{Y}_{j(i)})^2, \quad MSE_{(i)} = \frac{SSE_{(i)}}{n - p - 1}$$

where $Y_{j(i)}$ is the fitted value of the jth case under the leave-i-out fit. So $SSE_{(i)}$ and $MSE_{(i)}$ are the SSE and MSE of the leave-i-out fit, respectively. Show that

$$SSE_{(i)} = SSE - \frac{e_i^2}{1 - h_{ii}}$$

here SSE, e_i, h_{ii} are from the regression fit using all n cases.

(c) The studentized deleted residuals are defined as:

$$t_i = \frac{d_i}{s\{d_i\}} = \frac{d_i}{\sqrt{MSE_{(i)}/(1 - h_{ii})}}, \quad i = 1, \dots, n$$

Show that:

$$t_i = e_i \sqrt{\frac{n - p - 1}{SSE(1 - h_{ii}) - e_i^2}}, \quad i = 1, \dots, n$$

(d) Under the Normality assumption, i.e., Y is an n dimensional Normal random vector with $Var(Y) = \sigma^2 I_n$, show that $SSE_{(i)}$ is independent with Y_i and $\hat{Y}_{i(i)}$. Therefore, $SSE_{(i)}$ is independent with d_i.

Solution:

(a)

$$Var\,(d_i) = Var\left(\frac{e_i}{1 - h_{ii}}\right) = \frac{1}{(1 - h_{ii})^2} Var(e_i) = \frac{\sigma^2}{(1 - h_{ii})^2}$$

(b) Observe,

$$\begin{aligned} SSE &= Y^T (I - H) Y \\ &= \tilde{Y}^T (I - H) \tilde{Y} + e_i Y^T (1 - H) Y e_i \\ &= SSE_{(i)} - \frac{e_i^2}{1 - h_{ii}} \end{aligned}$$

(c) Observe,

$$\begin{aligned} t_i &= \frac{d_i}{\sqrt{MSE_{(i)}/(1-h_{ii})}} \\ &= \frac{e_i}{(1-h_{ii})\sqrt{MSE_{(i)}/(1-h_{ii})}} \\ &= \frac{e_i}{\sqrt{MSE_{(i)}(1-h_{ii})}} \\ &= \frac{e_i}{\sqrt{\frac{SSE_{(i)}}{n-p-1}(1-h_{ii})}} \\ &= \frac{e_i}{\sqrt{\frac{SSE-\frac{e_i^2}{1-h_{ii}}}{n-p-1}(1-h_{ii})}} \\ &= \frac{e_i}{\sqrt{\frac{SSE(1-h_{ii})-e_i^2}{n-p-1}}} \\ &= e_i\sqrt{\frac{n-p-1}{SSE(1-h_{ii})-e_i^2}} \end{aligned}$$

(d) Observe,

$$Cov(SSE_{(i)}, Y_i) = Cov(SSE, Y_i) - Cov(\frac{e_i^2}{1-h_{ii}}, Y_i) = 0$$

$$Cov(SSE_{(i)}, \hat{Y}_{i(i)}) = Cov(SSE, \hat{Y}_{i(i)}) - Cov(\frac{e_i^2}{1-h_{ii}}, \hat{Y}_{i(i)}) = 0$$

Problem 3
Cook's distance. The Cook's distances are defined as

$$D_i := \frac{\sum_{j=1}^n (\hat{Y} - \hat{Y}_{j(i)})^2}{p \times MSE}, \quad i = 1, \ldots, n$$

where $\hat{Y}_{j(i)}$ is the fitted value of the jth case under the leave-i-out fit. Show that:

$$D_i = \frac{e_i^2}{p \times MSE} \frac{h_{ii}}{(1-h_{ii})^2}$$

Solution:

$$\begin{aligned} D_i &= \frac{\sum_{j=1}^n (\hat{Y} - \hat{Y}_{j(i)})^2}{p \times MSE} \\ &= \frac{(Y-\tilde{Y})^T H (Y-\tilde{Y})}{p \times MSE} \\ &= \frac{e_i^T/(1-h_{ii}) h_{ii} e_i^T/(1-h_{ii})}{p \times MSE} \\ &= \frac{e_i^2}{p \times MSE} \frac{h_{ii}}{(1-h_{ii})^2} \end{aligned}$$