1. $\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2$

$=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$

$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=\sum_{i=1}^{n}\hat{\varepsilon}_i(x_i'\hat{\beta}-\bar{x}'\hat{\beta})=\sum_{i=1}^{n}\hat{\varepsilon}_i(x_i'-\bar{x}')\hat{\beta}$, when $\bar{x}'=\frac{1}{n}\sum_{i=1}^{n}x_i'$

$=\sum_{i=1}^{n}\hat{\varepsilon}_i x_i'\hat{\beta}-n\cdot\hat{\varepsilon}_i\cdot\frac{1}{n}\sum_{i=1}^{n}x_i'\hat{\beta}=0$

$\therefore\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$

2. $\hat{\rho}_{y\hat{y}}^2=\dfrac{\left(\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{y})\right)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}$

$\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{y})=\sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\bar{y})(\hat{y}_i-\bar{y})=\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$

$\therefore\hat{\rho}_{y\hat{y}}^2=\dfrac{\left[\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2\right]^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}=\dfrac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=R^2$

3. $studr_i=\dfrac{\hat{\varepsilon}_{[-i]}}{\sqrt{\hat{\sigma}^2_{[-i]}/(1-h_{ii})}}=\dfrac{y_i-x_i^T\hat{\beta}_{[-i]}}{\sqrt{\hat{\sigma}^2_{[-i]}/(1-h_{ii})}}$

$y_i-x_i^T\hat{\beta}_{[-i]}=\hat{\varepsilon}_{[-i]}\sim N(0,\sigma^2/(1-h_{ii}))\Rightarrow\dfrac{y_i-x_i^T\hat{\beta}_{[-i]}}{\sigma/\sqrt{1-h_{ii}}}\sim N(0,1)$

$(n-p-1)\hat{\sigma}^2_{[-i]}/\sigma^2=Y'(I_n-H_{[-i]})Y/\sigma^2\sim\chi^2_{n-1-p}$

And we know $\hat{\beta}_{[-i]}$ and $\hat{\sigma}^2_{[-i]}$ are mutually independent under GM linear model

$\therefore\ studr_i=\dfrac{(y_i-x_i^T\hat{\beta}_{[-i]})/[\sigma/\sqrt{1-h_{ii}}]}{\sqrt{\hat{\sigma}^2_{[-i]}/\sigma^2}}=\dfrac{y_i-x_i^T\hat{\beta}_{[-i]}}{\sqrt{\hat{\sigma}^2_{[-i]}/(1-h_{ii})}}\sim t_{n-1-p}$

4. $Cook_i=\dfrac{(\hat{\beta}_{[-i]}-\hat{\beta})X^TX(\hat{\beta}_{[-i]}-\hat{\beta})}{p\hat{\sigma}^2}=\dfrac{(X\hat{\beta}_{[-i]}-X\hat{\beta})'(X\hat{\beta}_{[-i]}-X\hat{\beta})}{p\hat{\sigma}^2}$

$\hat{\beta}_{[-i]}-\hat{\beta}=-(1-h_{ii})^{-1}(X^TX)^{-1}x_i\hat{\varepsilon}_i$, $standr_i=\dfrac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$

$(\hat{\beta}_{[-i]}-\hat{\beta})X^TX(\hat{\beta}_{[-i]}-\hat{\beta})=(1-h_{ii})^{-2}\hat{\varepsilon}_i x_i'(X^TX)^{-1}x_i'\hat{\varepsilon}_i=h_{ii}(1-h_{ii})^{-2}\hat{\varepsilon}_i^2$

$Cook_i=\dfrac{h_{ii}(1-h_{ii})^{-2}\hat{\varepsilon}_i^2}{p\hat{\sigma}^2}$, $standr_i^2=\dfrac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2(1-h_{ii})}$

$standr_i^2\times\dfrac{h_{ii}}{p(1-h_{ii})}=\dfrac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2(1-h_{ii})}\dfrac{h_{ii}}{p(1-h_{ii})}=\dfrac{\hat{\varepsilon}_i^2\cdot h_{ii}}{p\hat{\sigma}^2(1-h_{ii})^2}=Cook_i$

5. (1) $(n-p-1)\hat{\sigma}^2_{[-i]}=\sum_{j\neq i}(y_j-x_j'\hat{\beta}_{[-i]})^2=\sum_{j=1}^{n}(y_j-x_j'\hat{\beta}_{[-i]})^2-(y_i-x_i'\hat{\beta}_{[-i]})^2$

$\sum_{j=1}^{n}(y_j-x_j'\hat{\beta}_{[-i]})^2=\sum_{j=1}^{n}(y_j-x_j'\hat{\beta}+x_j'\hat{\beta}-x_j'\hat{\beta}_{[-i]})^2=\sum_{j=1}^{n}(y_j-x_j'\hat{\beta})^2+\sum_{j=1}^{n}(x_j'\hat{\beta}-x_j'\hat{\beta}_{[-i]})^2$

$-2\sum_{j=1}^{n}(y_j-x_j'\hat{\beta})(x_j'\hat{\beta}-x_j'\hat{\beta}_{[-i]})$

$=(n-p)\hat{\sigma}^2+\sum_{j=1}^{n}\left(\dfrac{x_j'(X^TX)^{-1}x_i\hat{\varepsilon}_i}{1-h_{ii}}\right)^2-2\sum_{j=1}^{n}\dfrac{\hat{\varepsilon}_j\cdot x_j'(X^TX)^{-1}x_i\hat{\varepsilon}_i}{1-h_{ii}}$

$=(n-p)\hat{\sigma}^2+\sum_{j=1}^{n}\left(\dfrac{h_{ij}}{1-h_{ii}}\right)^2\hat{\varepsilon}_i^2-2\sum_{j=1}^{n}\dfrac{\hat{\varepsilon}_j h_{ij}}{1-h_{ii}}x_i\hat{\varepsilon}_i$, since $H\hat{\varepsilon}=0$, then we know $\sum_{j=1}^{n}\hat{\varepsilon}_j h_{ij}=0$

$=(n-p)\hat{\sigma}^2+\sum_{j=1}^{n}\dfrac{(h_{ij})^2\hat{\varepsilon}_i^2}{(1-h_{ii})^2}=(n-p)\hat{\sigma}^2+\dfrac{h_{ii}}{(1-h_{ii})^2}\hat{\varepsilon}_i^2$

$\therefore(n-p-1)\hat{\sigma}^2_{[-i]}=(n-p)\hat{\sigma}^2+\dfrac{h_{ii}}{(1-h_{ii})^2}\hat{\varepsilon}_i^2-\dfrac{\hat{\varepsilon}_i^2}{(1-h_{ii})^2}=(n-p)\hat{\sigma}^2-\dfrac{\hat{\varepsilon}_i^2}{1-h_{ii}}$

(2) We know: $(n-p-1)\hat{\sigma}^2_{[-i]} = (n-p)\hat{\sigma}^2 - \frac{\hat{\varepsilon}_i^2}{1-h_{ii}}$

$$\hat{\sigma}_{[-i]} = \frac{n-p}{n-p-1}\hat{\sigma}^2 - \frac{1}{n-p-1}\frac{\hat{\varepsilon}_i^2}{1-h_{ii}} = \frac{n-p}{n-p-1}\hat{\sigma}^2 - \hat{\sigma}^2 \cdot standr_i^2 \frac{1}{n-p-1}$$

$$studr_i = \frac{y_i - x_i'\beta_{[-i]}}{\sqrt{\hat{\sigma}_{[-i]}/(1-h_{ii})}} = \frac{(y_i - x_i'\hat{\beta} + \frac{h_{ii}}{1-h_{ii}}\hat{\varepsilon}_i)\sqrt{1-h_{ii}}}{\sqrt{\frac{1}{n-p-1}[(n-p)\hat{\sigma}^2 - \hat{\sigma}^2 standr_i^2]}} = \frac{\hat{\sigma}\cdot\frac{\hat{\varepsilon}_i}{\hat{\sigma}(1-h_{ii})}\sqrt{1-h_{ii}}}{\sqrt{\frac{1}{n-p-1}[(n-p)\hat{\sigma}^2 - \hat{\sigma}^2 standr_i^2]}}$$

$$= standr_i\sqrt{\frac{n-p-1}{n-p-standr_i^2}}$$

6. $x \perp\!\!\!\perp y \Rightarrow cov(y, x) = 0 \quad \therefore \rho_{yx} = 0$

Conditional independence doesn't imply zero partial correlation.

The partial variance-covariance matrix of $Z:(X,Y)$ is define as:

$$\Sigma_{XY\cdot w} = \begin{bmatrix} \sigma_{11\cdot w} & \sigma_{12\cdot w} \\ \sigma_{21\cdot w} & \sigma_{22\cdot w} \end{bmatrix} = \Sigma_{XY} - \Sigma_{Xw}\Sigma_{ww}^{-1}\Sigma_{wY}$$

$$\rho_{XY\cdot w} = \frac{\sigma_{12\cdot w}}{\sqrt{\sigma_{11\cdot w}\sigma_{22\cdot w}}} \quad \text{(partial correlation)}$$

$$\Sigma_{XY|w} = cov(X, Y|w) = \begin{bmatrix} \sigma_{11|w} & \sigma_{12|w} \\ \sigma_{21|w} & \sigma_{22|w} \end{bmatrix} = E[(X - E(X|w))(Y - E(Y|w))|w]$$

$$\rho_{X,Y|w} = \frac{\sigma_{12|w}}{\sqrt{\sigma_{11|w}\sigma_{22|w}}} \quad \text{(conditional correlation)}$$

Only when $E(Z|w) = \alpha + BZ$, then we can get $\Sigma_{XY\cdot w} = \Sigma_{XY|w}$. Otherwise they're not equivalent. Therefore, when $\rho_{X,Y|w} = 0$, we can not get $\rho_{XY\cdot w} = 0$, which means conditionally independence doesn't lead to partial correlation equals 0.

For example, consider a random 3×1 vector (X_1, X_2, Y)

$X_1, X_2 | Y=y \sim H\left(\frac{X_1 - \mu_1(y)}{\sigma_1(y)}, \frac{X_2 - \mu_2(y)}{\sigma_2(y)}\right)$, $\sigma_1(y) > 0$, $\sigma_2(y) > 0$,

assume H has zero mean, unit variance and correlation coefficient ρ

The conditional covariance matrix is:

$$\begin{pmatrix} \sigma_1(y)^2 & \rho\sigma_1(y)\sigma_2(y) \\ \rho\sigma_1(y)\sigma_2(y) & \sigma_2(y)^2 \end{pmatrix}$$

(i) If $\mu_i(y) = a_i + b_i y$ and $\sigma_i(y) = \sigma_i$ for $i=1,2$, then $\sigma_{12|Y} = \rho\sigma_1\sigma_2 = \sigma_{12\cdot Y}$

(ii) If $\mu_i(y) = y^2$, $\sigma_i(y) = \sigma_i$ for $i=1,2$. $\sigma_{12|Y} = \rho\sigma_1\sigma_2$, $\sigma_{12\cdot Y} = (\rho + var(Y^2)(1 - \rho(Y, Y^2)^2))\sigma_1\sigma_2$

which shows that $\rho_{X_1,X_2\cdot Y}$ doesn't equal to $\rho_{X_1,X_2|Y}$ all the time.

Therefore conditionally independence $\not\Rightarrow$ partial correlation $= 0$

7. Let the best linear approximation be $(\alpha + \beta X)$

$$\Rightarrow \begin{cases} \beta = \dfrac{Cov(X, Y)}{var X} \\ \alpha = EY - EX \cdot \beta \end{cases}$$

$Cov(X,Y) = Cov(X, X^3 + \varepsilon) = Cov(X, X^3) = EX^4 - EXEX^3 = 3$

$var X = 1, \quad EX = 0, \quad EY = EX^3 = 0$

$\therefore$ The Best Linear Approximation is $3X$