Expectation and Covariance matrix of the estimators.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + \varepsilon)$$

$$\Rightarrow \hat{\beta} = \beta + (X^T X)^{-1} X^T \varepsilon$$

Since $E[\varepsilon] = 0$, $E[\hat{\beta}] = \beta$ i.e the least square estimator is unbiased.

$$Cov[\hat{\beta}] = E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] = E[(X^T X)^{-1} X^T \varepsilon \varepsilon^T X (X^T X)^{-1}]$$

Note

$$(X^T X)^T = X^T X \Rightarrow X^T X \text{ symmetric}$$

A symmetric $\Rightarrow A = P \Lambda P^T$ and $A^{-1} = P \Lambda^{-1} P^T$

$\Rightarrow (A^{-1})^T = P \Lambda^{-1} P^T = A^{-1} \Rightarrow A^{-1}$ is symmetric

We get

$$Cov[\hat{\beta}] = (X^T X)^{-1} X^T E[\varepsilon \varepsilon^T] X (X^T X)^{-1}$$

$$= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

Maximum likelihood estimation

Assuming $\varepsilon \sim N(0, \sigma^2 I)$ we have for $Y = X\beta + \varepsilon$ that $Y \sim N(X\beta, \sigma^2 I)$ and the likelihood ~~estimator~~ for n indepent variables can be written

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2}$$

where $x_i = \begin{bmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ik} \end{bmatrix}$. This shows that the ML estimates for β are in the case of equal variances found by minimizing $\sum_{i=1}^{n}(y_i - x_i^T\beta)^2$ with respect to β and therefore equals the least squares estimates.

However, $\hat{\underline{\beta}}$ is then $N(\underline{\beta}, \sigma^2(\underline{X}^T\underline{X})^{-1})$ since $\hat{\underline{\beta}}$ is a linear transformation of $\underline{Y}$.

Partitioning of variation

A measure of the total variance in the response is given by $\sum_{i=1}^{n}(y_i - \bar{y})^2$. We can split up this as follows:

We write $\quad y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y}$

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

Now
$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \sum_{i=1}^{n}(y_i-\hat{y}_i)\hat{y}_i - \sum_{i=1}^{n}(y_i-\hat{y}_i)\bar{y} = 0$$

This follows because both the two last expressions are zero since the normal equations have to be fulfilled.

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

which can be written as: $\quad SS_T = SS_E + SSR$

SS_T: total sum of squares; SS_E: error sum of squares; SSR: regression sum of squares

This partitioning is only valid if the normal equations are fulfilled. In models with no constant term the partitioning is normally not valid.

An estimate for the error variance is given by $\frac{SS_E}{n-k-1}$, but to explain this we need the help of idempotent matrices.

Expressed by idempotent matrices

Let $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$ be a vector of random variables. We have

$$\bar{\underline{Y}} = \begin{bmatrix} \bar{Y} \\ \bar{Y} \\ \vdots \\ \bar{Y} \end{bmatrix} = \frac{1}{m} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \frac{1}{m} \underline{1}\underline{1}^T \underline{Y}$$

$\underline{I}_{m\times m} - \frac{1}{m}\underline{1}\underline{1}^T$ is symmetric and idempotent and $\underline{Y} - \bar{\underline{Y}} = (\underline{I}_{m\times m} - \frac{1}{m}\underline{1}\underline{1}^T)\underline{Y}$

Hence $(\underline{y} - \bar{\underline{y}})^T(\underline{y} - \bar{\underline{y}}) = \underline{y}^T(\underline{I} - \frac{1}{m}\underline{1}\underline{1}^T)(\underline{I} - \frac{1}{m}\underline{1}\underline{1}^T)\underline{y}$

$= \underline{y}^T(\underline{I} - \frac{1}{m}\underline{1}\underline{1}^T)\underline{y} = SS_T$

Further $\hat{\underline{Y}} = \underline{X}\hat{\underline{\beta}} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y} = \underline{H}\underline{Y}$ where

$\underline{H} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T$ is the hat matrix.

$\underline{H}^T = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T \Rightarrow \underline{H}$. symmetric

and $\underline{H}\underline{H} = \underline{X}(\underline{X}^T\underline{X})\underline{X}^T\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T = \underline{H} \Rightarrow$ idempotent

We get. $\underline{Y} - \hat{\underline{Y}} = (\underline{I}_{m\times m} - \underline{H}_{m\times m})\underline{Y}$ and $(\underline{y} - \hat{\underline{y}})^T(\underline{y} - \hat{\underline{y}})$

$= \underline{y}^T(\underline{I} - \underline{H})(\underline{I} - \underline{H})\underline{y} = \underline{y}^T(\underline{I} - \underline{H})\underline{y} = SS_E$

Note: $(\underline{I} - \underline{H})(\underline{I} - \underline{H}) = \underline{I} - \underline{H} - \underline{H} + \underline{H}^2 = \underline{I} - \underline{H}$

Finally: $\hat{\underline{Y}} - \bar{\underline{Y}} = (\underline{H} - \frac{1}{m}\underline{1}\underline{1}^T)\underline{Y}$

We have: $\underline{H}\underline{X} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{X} = \underline{X} \Rightarrow \underline{H}\underline{1} = \underline{1}$

Therefore $(\underline{H} - \frac{1}{m}\underline{1}\underline{1}^T)(\underline{H} - \frac{1}{m}\underline{1}\underline{1}^T) = \underline{H}^2 - \frac{1}{m}\underline{1}\underline{1}^T\underline{H} - \underline{H}\frac{1}{m}\underline{1}\underline{1}^T + \frac{1}{m}\underline{1}\underline{1}^T$

$= \underline{H}^2 - \frac{1}{m}\underline{1}\underline{1}^T - \frac{1}{m}\underline{1}\underline{1}^T + \frac{1}{m}\underline{1}\underline{1}^T = \underline{H} - \frac{1}{m}\underline{1}\underline{1}^T$

$\uparrow$

$\underline{1}^T\underline{H} = (\underline{H}\underline{1})^T = \underline{1}^T$

We get $(\hat{y} - \bar{y})^T (\hat{y} - \bar{y}) = y^T (H - \frac{1}{n} 1 1^T)(H - \frac{1}{n} 1 1^T) y$

$= y^T (H - \frac{1}{n} 1 1^T) y = SS_R$

Also. H, $\frac{1}{n} 1 1^T$, $I - H$, $I - \frac{1}{n} 1 1^T$ and $H - \frac{1}{n} 1 1^T$

are all idempotent matrices and also projection matrices.