Unbiased estimator of σ^2

Now, prove that $\widehat{\sigma^2}$ is an unbiased estimator of σ^2, i.e., $E\left(\widehat{\sigma^2}\right) = \sigma^2$.

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-p}$$

Consider the numerator.

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \left\|Y - X\hat{\beta}\right\|_2^2$$

$$\begin{aligned} Y - X\hat{\beta} &= Y - X(X^TX)^{-1}X^TY = (X\beta + \varepsilon) - X(X^TX)^{-1}X^T(X\beta + \varepsilon) \\ &= X\beta + \varepsilon - X(X^TX)^{-1}X^TX\beta - X(X^TX)^{-1}X^T\varepsilon \\ &= X\beta + \varepsilon - X\beta - X(X^TX)^{-1}X^T\varepsilon = (I_n - H)\varepsilon \end{aligned}$$

where $H = X(X^TX)^{-1}X^T$

We can see that H has the following properties

$$H^T = H, HH = H$$

So

$$\begin{aligned} \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 &= \left\|Y - X\hat{\beta}\right\|_2^2 = \|(I_N - H)\varepsilon\|_2^2 = [(I_n - H)\varepsilon]^T[(I_n - H)\varepsilon] \\ &= \varepsilon^T(I_n - H)^T(I_n - H)\varepsilon = \varepsilon^T\left[\left({I_n}^T - H^T\right)(I_n - H)\right]\varepsilon \\ &= \varepsilon^T\left({I_n}^TI_n - {I_n}^TH - H^TI_n + H^TH\right)\varepsilon = \varepsilon^T(I_n - H - H + H)\varepsilon = \varepsilon^T(I_n - H)\varepsilon \end{aligned}$$

Let δ_{ij} be the element in $(I_n - H)$ in row i column j.
δ_{ij} is deterministic because both I_n and H are deterministic.

Note that

$$\varepsilon^T(I_n - H)\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]\begin{bmatrix}\delta_{11} & \dots & \delta_{n1} \\ \vdots & \ddots & \vdots \\ \delta_{1n} & \dots & \delta_{nn}\end{bmatrix}\begin{bmatrix}\varepsilon_1 \\ \vdots \\ \varepsilon_n\end{bmatrix} = \begin{bmatrix}\varepsilon_1\delta_{11} + \dots + \varepsilon_n\delta_{1n}, \\ \varepsilon_1\delta_{21} + \dots + \varepsilon_n\delta_{2n}, \\ \dots, \\ \varepsilon_1\delta_{n1} + \dots + \varepsilon_n\delta_{nn}\end{bmatrix}\begin{bmatrix}\varepsilon_1 \\ \vdots \\ \varepsilon_n\end{bmatrix}$$

$$= (\varepsilon_1\varepsilon_1\delta_{11} + \dots + \varepsilon_1\varepsilon_n\delta_{1n}) + \dots + (\varepsilon_1\varepsilon_n\delta_{n1} + \dots + \varepsilon_n\varepsilon_n\delta_{nn})$$

So

$$\begin{aligned} E[\varepsilon^T(I_n - H)\varepsilon] &= E[(\varepsilon_1\varepsilon_1\delta_{11} + \dots + \varepsilon_1\varepsilon_n\delta_{1n}) + \dots + (\varepsilon_1\varepsilon_n\delta_{n1} + \dots + \varepsilon_n\varepsilon_n\delta_{nn})] \\ &= [\delta_{11}E(\varepsilon_1\varepsilon_1) + \dots + \delta_{1n}E(\varepsilon_1\varepsilon_n)] + \dots + [\delta_{n1}E(\varepsilon_n\varepsilon_1) + \dots + \delta_{nn}E(\varepsilon_n\varepsilon_n)] \end{aligned}$$

Given that $E(\varepsilon_i\varepsilon_j) = \sigma^2$ for $i = j$ and $E(\varepsilon_i\varepsilon_j) = 0$ for $i \neq j$, we have

$$E[\varepsilon^T(I_n - H)\varepsilon] = \delta_{11}\sigma^2 + \delta_{22}\sigma^2 + \cdots + \delta_{nn}\sigma^2 = \sigma^2 \times \mathrm{Trace}(I_n - H)$$
$$= \sigma^2 \times [\mathrm{Trace}(I_n) - \mathrm{Trace}(H)]$$
$$\text{because } \mathrm{Trace}(A - B) = \mathrm{Trace}(A) - \mathrm{Trace}(B)$$

Note that

$$\mathrm{Trace}(H) = \mathrm{Trace}[X(X^TX)^{-1}X^T] = \mathrm{Trace}([X(X^TX)^{-1}]^T(X^T)^T) = \mathrm{Trace}[(X^TX)^{-1}X^TX]$$
$$= \mathrm{Trace}(I_k) = \mathrm{k}$$

The equation above used the property for trace of a product that

$$\mathrm{Trace}(AB) = \mathrm{Trace}(A^TB^T).$$

So putting everything together

$$E\left[\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2\right] = E[\varepsilon^T(I_N - H)\varepsilon] = \sigma^2 \times [\mathrm{Trace}(I_n) - \mathrm{Trace}(H)] = \sigma^2(\mathrm{n} - \mathrm{p})$$

Therefore,

$$E\left[\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n - p}\right] = E\left(\frac{RSS}{n - p}\right) = \sigma^2$$

And therefore, an unbiased estimator of σ^2 is

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n - p} = \frac{\left\|Y - X\hat{\beta}\right\|_2^2}{n - p} = \frac{RSS}{n - p}$$