e.g. If random variables (X, Y) is from $Y = \underset{1}{\beta_0} + \underset{5}{\beta_1} X_1 + \varepsilon$

What is the value of β_0^*, β_1^* s.t.

$$(\beta_0^*, \beta_1^*) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} E_{XY}[(Y - \beta_0 - \beta_1 X_1)^2]$$

$$\begin{cases} \beta_1^* = \dfrac{Cov(Y, X)}{Var(X)} \\ \beta_0^* = E(Y) - \beta_1^* E(X_1) \end{cases}$$

Now plug-in $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

$$\begin{aligned} \beta_1^* &= \frac{Cov(\beta_0 + \beta_1 X_1 + \varepsilon, X_1)}{Var(X_1)} \\ &= \frac{\beta_1 Cov(X_1, X_1)}{Var(X_1)} \\ &= \frac{\beta_1 Var(X_1)}{Var(X_1)} \\ &= \beta_1 = 5 \end{aligned}$$

$$\begin{aligned} \beta_0^* &= E(Y) - \beta_1^* E(X) \\ &= E(\beta_0 + \beta_1 X_1 + \varepsilon) - \beta_1 E(X_1) \\ &= \beta_0 + \beta_1 E(X_1) + 0 - \beta_1 E(X_1) \\ &= \beta_0 = 1 \end{aligned}$$

Therefore, for the variables (X_1, Y) genereated from the simple linear regression model $Y = \beta_0 + \beta_1 X_1 + \varepsilon$, the optimal linear prediction is just

$$m^*(x_1) = E(Y \mid X_1 = x_1) = \beta_0 + \beta_1 x_1$$

1.5 Least Squares in Matrix Format

We have the simple linear regression model

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where $E(\varepsilon | X = x) = 0$ and $Var(\varepsilon | X = x) = \sigma^2$ and ε is uncorelated with X and uncorelated accross observations.

Our data consists of n paired observation of X and Y, sampled from the above model.

$$(X_1, Y_1) \cdots \quad \cdots (X_n, Y_n)$$

Hence, $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1 \cdots n$

If we denote

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ \vdots & \vdots \\ 1 & X_{n1} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

design matrix
(an $n \times 2$ matrix with the first column always 1)

Then we can write the set of equations $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \quad i = 1 \cdots n$ in the simpler form

$$Y = X\beta + \varepsilon$$

The training MSE is

$$S(\beta) = \hat{MSE}(\beta) = \frac{1}{n}(Y - X\beta)^T (Y - X\beta)$$

$$= \frac{1}{n}(Y^T - \beta^T X^T)(Y - X\beta)$$

$$= \frac{1}{n}(Y^T Y - \underline{Y^T X\beta - \beta^T X^T Y} + \beta^T X^T X\beta)$$

$\beta^T X^T Y = (Y^T X \beta)^T$

$= \underset{1\times n \;\; n\times 2 \;\; 2\times 1}{Y^T X \beta}$ scalar

$$= \frac{1}{n}(Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta)$$

$$\nabla_\beta S(\beta) = \frac{1}{n}(\nabla_\beta Y^T Y - 2\nabla_\beta \hat{\beta}^T X^T Y + \nabla_\beta \hat{\beta}^T X^T X \hat{\beta})$$

$$= \frac{1}{n}(0 - 2X^T Y + 2X^T X \hat{\beta})$$

$$= \frac{2}{n}(X^T X \hat{\beta} - X^T Y)$$

$$= 0$$

$$\Rightarrow \quad X^T X \hat{\beta} = X^T Y$$

This equation, for the two dimensional vector $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ corresponding to our pair of normal equation for $\hat{\beta}_0 \cdot \hat{\beta}_1$

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \end{bmatrix}$$

If this is correct, the equation we got above should in fact reproduce the least squares estimates we have already derived.

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

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \end{bmatrix}$$

Then we show that the above equations give

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$$

Note that

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 \end{bmatrix}^{-1} = \frac{\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & -\sum_{i=1}^{n} x_{i1} \\ -\sum_{i=1}^{n} x_{i1} & n \end{bmatrix}}{n\sum_{i=1}^{n} x_{i1}^2 - (\sum_{i=1}^{n} x_{i1})^2}$$

$$\text{Thus, } \hat{\beta}_1 = \frac{-\sum_{i=1}^{n} x_{i1} \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} x_{i1} y_i}{n\sum_{i=1}^{n} x_{i1}^2 - (\sum_{i=1}^{n} x_{i1})^2}$$

$$= \frac{-n\overline{x_{i1}}\,\bar{y} + \sum_{i=1}^{n} x_{i1} y_i}{\sum_{i=1}^{n} x_{i1}^2 - n\bar{x}_1^2}$$

$$= \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_1 = \frac{\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & -\sum_{i=1}^{n} x_{i1} \\ -\sum_{i=1}^{n} x_{i1} & n \end{bmatrix}}{n\sum_{i=1}^{n} x_{i1}^2 - (\sum_{i=1}^{n} x_{i1})^2} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \end{bmatrix}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} x_{i1}^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_{i1} \sum_{i=1}^{n} x_{i1} y_i}{n \sum_{i=1}^{n} x_{i1}^2 - (\sum_{i=1}^{n} x_{i1})^2}$$

$$= \frac{\bar{y} \sum_{i=1}^{n} x_{i1}^2 - \bar{x}_1 \sum_{i=1}^{n} x_{i1} y_i}{\sum_{i=1}^{n} x_{i1}^2 - n \bar{x}_1^2}$$

$$= \bar{y} + \frac{n \bar{y} \bar{x}_1^2 - \bar{x}_1 \sum_{i=1}^{n} x_{i1} y_i}{S_{xx}}$$

$$= \bar{y} - \bar{x}_1 \frac{S_{xy}}{S_{xx}}$$

$$= \bar{y} - \bar{x}_1 \hat{\beta}_1$$

Therefore, the least squares in the matrix format gives the same estimation results for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Note: The resulting line from least squares is

$$m(x_1) = \hat{\beta}_0 + \hat{\beta}_1 x_1$$
$$= (\bar{y} - \bar{x}_1 \hat{\beta}_1) + \hat{\beta}_1 x_1$$
$$= \bar{y} + \hat{\beta}_1 (x_1 - \bar{x}_1)$$

If $x_1 = \bar{x}_1$, we see that

$$m(\bar{x}_1) = \bar{y}$$

Thus the line passes through $(\bar{x}_1, \bar{y})$.
In comparison, the optimal line prediction (true regression line) goes through $(E(X), E(Y))$

1.6 Bias, Variance and Standard Error of Parametric Estimators

Estimators vs Estimates

An estimator is a random variable and an estimate is a number (realized value of estimator)

e.g. We can potentially draw n samples

$$X_1, X_2 \cdots X_n$$

from a distribution, if X_i's have not been actually observed yet, then they can be viewed as random variables from that distribution.

Therefore, the sample mean

$$\bar{X} = \frac{X_1 + \cdots X_n}{n}$$ big X

is also a random variable, called an estimator of the population mean μ.

But if the drawing have been conducted and the values of X_i's have been realized

$X_i = x_i$, $i = 1 \cdots n$, then the realized value

$$\bar{x} = \frac{x_1 + \cdots x_n}{n}$$ small x

is estimate of μ.

<Statements to be proved>

$$E(\hat{\beta}_1) = \beta_1 \qquad Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

$$E(\hat{\beta}_0) = \beta_0 \qquad Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}_1}{S_{xx}}\right)$$

proof

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)y_i}{S_{xx}}$$

$$= \sum_{i=1}^{n} \underbrace{\boxed{\frac{x_{i1} - \bar{x}_1}{S_{xx}}}}_{c_i} y_i$$

$$= \sum_{i=1}^{n} c_i y_i \qquad \text{where } c_i = \frac{x_{i1} - \bar{x}_1}{S_{xx}} \qquad \sum_{i=1}^{n} c_i = 0$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}_1\hat{\beta}_0 = \bar{y} - \bar{x}_1\sum_{i=1}^{n} c_i y_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i - \bar{x}_1\sum_{i=1}^{n} c_i y_i$$

$$= \sum_{i=1}^{n} \underbrace{\boxed{\left(\frac{1}{n} - \bar{x}_1 c_i\right)}}_{b_i} y_i$$

We find that both estimates are linear combination of y_i's. The corresponding estimators of β_0, β_1 are formed by replacing the observed y_i by the random variable Y_i in the previous formulas

i.e. $$\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$$

$$\hat{\beta}_0 = \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x}_1 c_i\right) Y_i$$

In this section, we are going to treat x_i's as non-random variable. This is appropriate in designed or controlled experiments, where we get to choose their values. But in randomized experiments or in observational studies, x_i's are not necessarily fixed.

$$
\begin{aligned}
E(\hat{\beta}_1) &= E(\hat{\beta}_1 \mid x_{11} \cdots\cdots x_{n1}) \\
&= E\left(\sum_{i=1}^{n} c_i Y_i \mid x_{11} \cdots\cdots x_{n1}\right) \\
&= \sum_{i=1}^{n} c_i \, E(Y_i \mid x_{11} \cdots\cdots x_{n1}) \\
&= \sum_{i=1}^{n} c_i \, E(\beta_0 + \beta_1 x_{i1} + \varepsilon_i \mid x_{11} \cdots\cdots x_{n1})
\end{aligned}
$$

if SLR is true

$$
\begin{aligned}
&= \sum_{i=1}^{n} c_i (\beta_0 + \beta_1 x_{i1}) \\
&= \beta_0 \underbrace{\sum_{i=1}^{n} c_i}_{0} + \beta_1 \underbrace{\sum_{i=1}^{n} c_i x_{i1}}_{1} \\
&= \beta_1
\end{aligned}
$$

$$\sum_{i=1}^{n} c_i x_{i1} = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1) x_{i1}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_{i1}^2 - n(\bar{x}_1)^2}{S_{xx}} = 1$$

Thus if we assume the modle is correct, the β_1 is an unbiased estimator. Similarly,

$$E(\hat{\beta}_0) = \beta_0$$ Homework

The variance of $\hat{\beta}_1$ is

$$
\begin{aligned}
Var(\hat{\beta}_1) &= Var(\hat{\beta}_1 \mid x_{11} \cdots\cdots x_{n1}) \\
&= Var\left(\sum_{i=1}^{n} c_i Y_i \mid x_1 \cdots\cdots x_{n1}\right) \\
&= \sum_{i=1}^{n} c_i^2 \underbrace{Var(Y_i \mid x_1 \cdots\cdots x_{n1})}
\end{aligned}
$$

Y_i's uncorrelated

$$
\begin{aligned}
&= Var(\beta_0 + \beta_1 x_{i1} + \varepsilon \mid x_1 \cdots\cdots x_{n1}) \\
&= Var(\beta_0 + \beta_1 x_{i1} \mid x_1 \cdots\cdots x_{n1}) + Var(\varepsilon \mid x_1 \cdots x_{n1}) \\
&= \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma^2 \sum_{i=1}^{n} c_i \\
&= \sigma^2 \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2}{S_{xx}^2}
\end{aligned}
$$

The variance of $\hat{\beta}_0$ is

$$\begin{aligned}
Var(\hat{\beta}_0) &= Var(\hat{\beta}_0 \mid x_{11} \cdots x_{n1}) \\
&= Var(\bar{Y} - \hat{\beta}_1 \bar{x}_1 \mid x_{11} \cdots x_{n1}) \\
&= Var(\bar{Y} \mid x_{11} \cdots) + \bar{x}_1^2 Var(\hat{\beta}_1 \mid x_{11} \cdots) - 2\bar{x}_1 \underbrace{Cov(\bar{Y}, \hat{\beta}_1 \mid x_{11} \cdots)}_{0} \\
&= \frac{\sigma^2}{n} + \bar{x}_1^2 Var(\hat{\beta}_1) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_1^2}{S_{xx}} \right)
\end{aligned}$$

$$\begin{aligned}
&= Var\left(\frac{Y_1 + Y_2 + \cdots Y_n}{n} \mid x_{11} \cdots x_{n1} \right) \\
&= \frac{1}{n^2} Var(Y_1) + \frac{1}{n^2} Var(Y_2) \cdots + \frac{1}{n^2} Var(Y_n) \\
&= \frac{1}{n} \sigma^2
\end{aligned}$$

Unconditional-on-X properties

- the Law of Total Expectation

$$\begin{aligned}
E(\hat{\beta}_1) &= E_X [\underbrace{E_{Y|X}(\hat{\beta}_1 \mid x_{11} \cdots x_{n1})}_{\beta_1}] \\
&= E_X(\beta_1) \\
&= \beta_1 \qquad \text{unconditionally unbiased}
\end{aligned}$$

- the Law of Total Variance

$$\begin{aligned}
Var(\hat{\beta}_1) &= E_X [Var_{Y|X}(\hat{\beta}_1 \mid x_{11} \cdots x_{n1})] + Var_X [\underbrace{E_{Y|X}(\hat{\beta}_1 \mid x_{11} \cdots x_{n1})}_{\beta_1}] \\
&= E_X \left[\frac{\sigma^2}{S_{xx}} \right] + \underbrace{Var(\beta_1)}_{0} \\
&= \sigma^2 E_X \left[\frac{1}{S_{xx}} \right]
\end{aligned}$$

1.7 Bias, Variance and Estimators in Matrix Format

$Y = X\beta + \varepsilon$

where $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 I$

$$Var(\varepsilon) = \begin{bmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1, \varepsilon_2) & \cdots \\ Cov(\varepsilon_2, \varepsilon_1) & \cdots & \\ \cdots & & Var(\varepsilon_n) \end{bmatrix}$$

Note : $Var(\varepsilon_i) = \sigma^2$ $\quad Cov(\varepsilon_i, \varepsilon_j) = 0$

$$= \begin{bmatrix} \sigma^2 & \ddots & 0 \\ 0 & & \sigma^2 \end{bmatrix}$$

This implies

We have that $\hat{\beta} = \underbrace{(X^T X)^{-1} X^T}_{A} Y = AY$

$$\Rightarrow E(\hat{\beta} | X) = E(AY | X)$$
$$= A E(Y | X)$$
$$= (X^T X)^{-1} X^T E(Y | X)$$
$$= (X^T X)^{-1} X^T X\beta$$
$$= \beta$$

i.e. $E(\hat{\beta}_0 | X) = \beta_0$, $\quad E(\hat{\beta}_1 | X) = \beta_1$

$$Var(\hat{\beta} | X) = Var(AY | X)$$
$$= A \, Var(Y | X) A^T$$
$$= A \sigma^2 I A^T$$
$$= \sigma^2 A A^T$$

$$AA^T = [(X^TX)^{-1}X^T][(X^TX)^{-1}X^T]^T$$

$$= \underline{(X^TX)^{-1}X^T \quad X}[(X^TX)^{-1}]^T$$

cancle out

$$= [(X^TX)^{-1}]^T$$

$$= (X^TX)^{-1}$$

$$\underset{n\times 2}{(X^TX)^{-1}} = \frac{1}{n\sum_{i=1}^{n} x_{i1}^2 - (\sum_{i=1}^{n} x_{i1})^2}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & -\sum_{i=1}^{n} x_{i1} \\ -\sum_{i=1}^{n} x_{i1} & n \end{bmatrix}$$

$$= \frac{1}{n S_{xx}}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & -\sum_{i=1}^{n} x_{i1} \\ -\sum_{i=1}^{n} x_{i1} & n \end{bmatrix}$$

$$\text{Also, } Var(\hat{\beta}|X) = \begin{bmatrix} Var(\hat{\beta}_0|X) & Cov(\hat{\beta}_0, \hat{\beta}_1|X) \\ Cov(\hat{\beta}_1, \hat{\beta}_0|X) & Var(\hat{\beta}_1|X) \end{bmatrix}$$

?

This verifies that

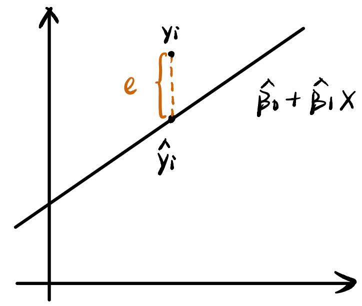
$$Var(\hat{\beta}_0|X) = \frac{\sigma^2}{n}\frac{\sum_{i=1}^{n} x_{i1}^2}{S_{xx}} = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}_1^2}{S_{xx}}\right)$$

?

$$Var(\hat{\beta}_1|X) = \frac{\sigma^2}{S_{xx}}$$

$$Cov(\hat{\beta}_1, \hat{\beta}_0) = -\frac{\sigma^2\sum_{i=1}^{n} x_{i1}}{n S_{xx}}$$

1.8 Residuals and Additional Properties



The residual is defined as

$$e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1}$$
$$= y_i - \hat{y}_i$$

Recall that

$$\frac{\partial S}{\partial \beta_0} = \frac{1}{n}(-2)\sum_{i=1}^{n}(\underline{y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1}}) = 0$$

$$\frac{\partial S}{\partial \beta_1} = \frac{1}{n}(-2)\sum_{i=1}^{n}(\underline{y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1}})x_{i1} = 0$$

$$\Rightarrow \sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} e_i x_{i1} = 0$$

Sum of squared residuals is basically the least square

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

② $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$

③ $\sum_{i=1}^{n} x_{i1} e_i = 0$

} combined to prove ④

④ $\sum_{i=1}^{n} \hat{y}_i e_i = 0$

Estimation of σ^2

① σ^2 captures the additional randomness in y

② σ^2 is in $Var(\hat{\beta}_1 | X)$ and $Var(\hat{\beta}_0 | X)$

$$Var(\varepsilon) = \sigma^2$$

$$Var(\hat{\beta}_0 | X) = \frac{\sigma^2}{n}\frac{\sum_{i=1}^{n} x_{i1}^2}{S_{xx}} = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}_1^2}{S_{xx}}\right)$$

$$Var(\hat{\beta}_1 | X) = \frac{\sigma^2}{S_{xx}}$$

Under SLR, $E[(Y-\beta_0-\beta_1X_1)^2] = E(\varepsilon^2)$

$$= Var(\varepsilon)+[E(\varepsilon)]^2$$
$$= \sigma^2+0$$
$$= \sigma^2$$

Replace Y with y_i, $i=1\cdots n$

X_1 with x_{i1}, $i=1\cdots n$

β_0 with $\hat{\beta}_0$

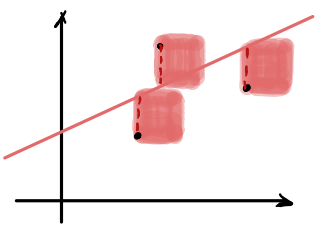
β_1 with $\hat{\beta}_1$

E with $\frac{1}{n}\sum_{i=1}^{n}$

Let $SS_{Res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \overbrace{\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1})^2}^{S(\hat{\beta}_0,\hat{\beta}_1)}$

"Residual sum of squares"

Measure aggregate mis-fit of the least square line



$$SS_{Res} = \underbrace{\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2}_{SS_T} - \hat{\beta}_1 S_{xy}$$

SS_T total sum of squares

The unbiased estimator of σ^2 is $\hat{\sigma}^2 = \frac{SS_{Res}}{n-2} = MS_{Res}$

$\downarrow$

"Residual mean square"

MSE vs MS_{Res}

$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_{i1})^2$ General case (not specify β_0, β_1)

$MS_{Res} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_{i1})^2$

Statistical Properties of $\hat{\sigma}^2$ in Matrix Format

Recall that $X\hat{\beta} = \underbrace{X(X^TX)^{-1}X^T}_{\text{hat matrix}}Y = HY$

- H is symmetric
- H is idempotent

$$H^TH = HH = [X(X^TX)^{-1}X^T][X(X^TX)^{-1}X^T]$$
$$= X\underline{(X^TX)^{-1}X^TX}(X^TX)^{-1}X^T$$
$$= X(X^TX)^{-1}X^T$$
$$= H$$

- $I_n - H$ is idempotent

$$(I_n-H)^T(I_n-H) = I_n^TI_n - I_n^TH - H^TI_n + H^TH$$
$$= I_n - H - H + H$$
$$= I_n - H$$

$\Rightarrow$
$$SS_{Res} = (Y - X\hat{\beta})^T(Y - X\hat{\beta})$$
$$= (Y - HY)^T(Y - HY)$$
$$= Y^T(I_n - H)^T(I_n - H)Y$$
$$= Y^T(I_n - H)Y$$

If V is a random vector
$E(V) = \mu \quad Var(V) = \Sigma$

For a constant matrix A
$E(V^TAV) = trace(A\Sigma) + \mu^TA\mu$

$$\boxed{\text{Now we have } E(Y|X) = \overset{\mu}{X\beta} \text{ and } Var(Y|X) = \overset{\Sigma}{\sigma^2 I_n}}$$

Assumption: SLR

$$E(SS_{Res}|X) = E[\underbrace{Y^T}_{V^T}\underbrace{(I_n-H)}_{A}\underbrace{Y}_{V}|X]$$
$$= trace((I_n-H)\sigma^2 I_n) + (X\beta)^T(I_n-H)(X\beta)$$
$$= \sigma^2 trace(I_n-H) + \underbrace{\beta^TX^T(I_n-H)X\beta}_{0}$$
$$= \sigma^2 trace(I_n) - \sigma^2 trace(H) + 0$$
$$= \sigma^2(n - trace(H))$$

Since $X^T(I_n - H)X = X^T X - X^T H X$

$$= X^T X - \underline{X^T X (X^T X)^{-1}} X^T X$$

$$= X^T X - X^T X$$

$$= 0$$

trace(ABC) = trace(CAB) = trace(BCA)

We know that $\text{trace}(H) = \text{trace}(X(X^T X)^{-1} X^T)$

$$= \text{trace}(X^T X (X^T X)^{-1})$$

$$\underset{n \times p}{X}$$

$$= \text{trace}(I_p)$$

$$= p$$

where $p = 2$ for SLR

$\Rightarrow E(SS_{Res} \mid X) = \sigma^2 (n-2)$

$$\Rightarrow E(\hat{\sigma}^2 \mid X) = E\left(\frac{SS_{Res}}{n-2} \,\middle|\, X\right)$$

$$= E\left(\frac{Y^T (I_n - H) Y}{n-2} \,\middle|\, X\right)$$

$$= \frac{\sigma^2 (n-2)}{n-2}$$

$$= \sigma^2$$

1.3 Use $\hat{\sigma}^2$ to estimate $Var(\hat{\beta}_0|X)$ and $Var(\hat{\beta}_1|X)$

Recall $Var(\hat{\beta}|X) = \begin{pmatrix} Var(\hat{\beta}_0|X) & Cov(\hat{\beta}_0, \hat{\beta}_1|X) \\ Cov(\hat{\beta}_1, \hat{\beta}_0|X) & Var(\hat{\beta}_1|X) \end{pmatrix}$

The diagonal entries can be written as

$$Var(\hat{\beta}_0|X) = \frac{\sigma^2}{n} \frac{\sum_{i=1}^{n} x_i^2}{S_{xx}} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

$$Var(\hat{\beta}_1|X) = \frac{\sigma^2}{S_{xx}}$$

The standard error (se) of the estimators are the square root of the variance

$$se(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0|X)} = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

$$se(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1|X)} = \sqrt{\frac{\sigma^2}{S_{xx}}}$$

Replace σ^2 with known $\hat{\sigma}^2$, then we get estimated standard error.

$$ese(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

$$ese(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$