The sum of squared residuals is written as follows:

$$S(\hat{\beta}) = \sum_{i=1}^{n} e_i^2 = e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) = (y' - \hat{\beta}'X')(y - X\hat{\beta})$$
$$= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} = y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}.$$

In the last equality, note that $\hat{\beta}'X'y = y'X\hat{\beta}$ because both are scalars.

To minimize $S(\hat{\beta})$ with respect to $\hat{\beta}$, we set the first derivative of $S(\hat{\beta})$ equal to zero, i.e.,

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0.$$

Solving the equation above with respect to $\hat{\beta}$, the **ordinary least squares estimator (OLS,** 最小自乗推定量) of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{19}$$

Thus, the ordinary least squares estimator is derived in the matrix form.

(*) Remark

The second order condition for minimization:

$$\frac{\partial^2 S(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}'} = 2X'X$$

is a positive definite matrix.

Set $c = Xd$.

For any $d \neq 0$, we have $c'c = d'X'Xd > 0$.

Now, in order to obtain the properties of $\hat{\beta}$ such as mean, variance, distribution and so on, (19) is rewritten as follows:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u. \qquad (20)\end{aligned}$$

Taking the expectation on both sides of (20), we have the following:

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'\mathrm{E}(u) = \beta,$$

because of $\mathrm{E}(u) = 0$ by the assumption of the error term u_i.

Thus, unbiasedness of $\hat{\beta}$ is shown.

The variance of $\hat{\beta}$ is obtained as:

$$\begin{aligned} V(\hat{\beta}) &= E((\hat{\beta}-\beta)(\hat{\beta}-\beta)') = E\Big((X'X)^{-1}X'u((X'X)^{-1}X'u)'\Big) \\ &= E((X'X)^{-1}X'uu'X(X'X)^{-1}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}. \end{aligned}$$

The first equality is the definition of variance in the case of vector.

In the fifth equality, $E(uu') = \sigma^2 I_n$ is used, which implies that $E(u_i^2) = \sigma^2$ for all i and $E(u_i u_j) = 0$ for $i \neq j$.

Remember that $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2.

Under normality assumption on the error term u, it is known that the distribution of $\hat{\beta}$ is given by:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}).$$

Proof:

First, when $X \sim N(\mu, \Sigma)$, the moment-generating function, i.e., $\phi(\theta)$, is given by:

$$\phi(\theta) \equiv \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp\Big(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta\Big)$$

θ_u: $n \times 1$, u: $n \times 1$, θ_β: $k \times 1$, $\hat{\beta}$: $k \times 1$

The moment-generating function of u, i.e., $\phi_u(\theta_u)$, is:

$$\phi_u(\theta_u) \equiv \mathrm{E}\Big(\exp(\theta_u' u)\Big) = \exp\Big(\frac{\sigma^2}{2}\theta_u'\theta_u\Big),$$

which is $N(0, \sigma^2 I_n)$.

The moment-generating function of $\hat{\beta}$, i.e., $\phi_\beta(\theta_\beta)$, is:

$$\begin{aligned}\phi_\beta(\theta_\beta) &\equiv \mathrm{E}\Big(\exp(\theta'_\beta\hat{\beta})\Big) = \mathrm{E}\Big(\exp(\theta'_\beta\beta + \theta'_\beta(X'X)^{-1}X'u)\Big) \\ &= \exp(\theta'_\beta\beta)\mathrm{E}\Big(\exp(\theta'_\beta(X'X)^{-1}X'u)\Big) = \exp(\theta'_\beta\beta)\phi_u\Big(\theta'_\beta(X'X)^{-1}X'\Big) \\ &= \exp(\theta'_\beta\beta)\exp\Big(\frac{\sigma^2}{2}\theta'_\beta(X'X)^{-1}\theta_\beta\Big) = \exp\Big(\theta'_\beta\beta + \frac{\sigma^2}{2}\theta'_\beta(X'X)^{-1}\theta_\beta\Big),\end{aligned}$$

which is equivalent to the normal distribution with mean β and variance $\sigma^2(X'X)^{-1}$.

Note that $\quad \theta_u = X(X'X)^{-1}\theta_\beta$. QED

Taking the jth element of $\hat{\beta}$, its distribution is given by:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 a_{jj}), \qquad \text{i.e.,} \qquad \frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{a_{jj}}} \sim N(0, 1),$$

where a_{jj} denotes the jth diagonal element of $(X'X)^{-1}$.

Replacing σ^2 by its estimator s^2, we have the following t distribution:

$$\frac{\hat{\beta}_j - \beta_j}{s \sqrt{a_{jj}}} \sim t(n - k),$$

where $t(n - k)$ denotes the t distribution with $n - k$ degrees of freedom.

[Review] Trace (トレース):

1. A: $n \times n$, $\quad \mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$, where a_{ij} denotes an element in the ith row and the jth column of a matrix A.

2. a: scalar (1×1), $\quad \mathrm{tr}(a) = a$

3. A: $n \times k$, B: $k \times n$, $\quad \mathrm{tr}(AB) = \mathrm{tr}(BA)$

4. $\mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}((X'X)^{-1}X'X) = \mathrm{tr}(I_k) = k$

5. When X is a square matrix of random variables, $\mathrm{E}(\mathrm{tr}(AX)) = \mathrm{tr}(A\mathrm{E}(X))$

End of Review

s^2 is taken as follows:

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-k}e'e = \frac{1}{n-k}(y - X\hat{\beta})'(y - X\hat{\beta}),$$

which leads to an unbiased estimator of σ^2.

Proof:

Substitute $y = X\beta + u$ and $\hat{\beta} = \beta + (X'X)^{-1}X'u$ into $e = y - X\hat{\beta}$.

$$\begin{aligned} e = y - X\hat{\beta} &= X\beta + u - X(\beta + (X'X)^{-1}X'u) \\ &= u - X(X'X)^{-1}X'u = (I_n - X(X'X)^{-1}X')u \end{aligned}$$

$I_n - X(X'X)^{-1}X'$ is idempotent and symmetric, because we have:

$$(I_n - X(X'X)^{-1}X')(I_n - X(X'X)^{-1}X') = I_n - X(X'X)^{-1}X,'$$

$$(I_n - X(X'X)^{-1}X')' = I_n - X(X'X)^{-1}X'.$$