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

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

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

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

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

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

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

Note that θ is replaced by $X(X'X)^{-1}\theta$. 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:

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

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

s^2 is rewritten as follows:

$$
\begin{aligned}
s^2 &= \frac{1}{n-k}e'e = \frac{1}{n-k}((I_n - X(X'X)^{-1}X')u)'(I_n - X(X'X)^{-1}X')u \\
&= \frac{1}{n-k}u'(I_n - X(X'X)^{-1}X')'(I_n - X(X'X)^{-1}X')u \\
&= \frac{1}{n-k}u'(I_n - X(X'X)^{-1}X')u
\end{aligned}
$$

Take the expectation of $u'(I_n - X(X'X)^{-1}X')u$ and note that $\mathrm{tr}(a) = a$ for a scalar a.

$$
\begin{aligned}
\mathrm{E}(s^2) &= \frac{1}{n-k}\mathrm{E}\Big(\mathrm{tr}\Big(u'(I_n - X(X'X)^{-1}X')u\Big)\Big) = \frac{1}{n-k}\mathrm{E}\Big(\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')uu'\Big)\Big) \\
&= \frac{1}{n-k}\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')\mathrm{E}(uu')\Big) = \frac{1}{n-k}\sigma^2\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')I_n\Big) \\
&= \frac{1}{n-k}\sigma^2\mathrm{tr}(I_n - X(X'X)^{-1}X') = \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}(X(X'X)^{-1}X')) \\
&= \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}((X'X)^{-1}X'X)) = \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}(I_k)) \\
&= \frac{1}{n-k}\sigma^2(n-k) = \sigma^2
\end{aligned}
$$

$\longrightarrow$ s^2 is an unbiased estimator of σ^2.

Note that we do not need normality assumption for unbiasedness of s^2.

[Review]

- $X'X \sim \chi^2(n)$ for $X \sim N(0, I_n)$.

- $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(n)$ for $X \sim N(\mu, \Sigma)$.

- $\dfrac{X'X}{\sigma^2} \sim \chi^2(n)$ for $X \sim N(0, \sigma^2 I_n)$.

- $\dfrac{X'AX}{\sigma^2} \sim \chi^2(G)$, where $X \sim N(0, \sigma^2 I_n)$ and A is a symmetric idempotent $n \times n$ matrix of rank $G \leq n$.

Remember that $G = \text{Rank}(A) = \text{tr}(A)$ when A is symmetric and idempotent.

[End of Review]

Under normality assumption for u, the distribution of s^2 is:

$$\frac{(n-k)s^2}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2(\mathrm{tr}(I_n - X(X'X)^{-1}X'))$$

Note that $\mathrm{tr}(I_n - X(X'X)^{-1}X') = n - k$, because

$$\mathrm{tr}(I_n) = n$$

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

Asymptotic Normality (without normality assumption on u): Using the central limit theorem, without normality assumption we can show that as $n \longrightarrow \infty$, under the condition of $\frac{1}{n}X'X \longrightarrow M$ we have the following result:

$$\frac{\hat{\beta}_j - \beta_j}{s\sqrt{a_{jj}}} \longrightarrow N(0,1),$$

where M denotes a $k \times k$ constant matrix.

Thus, we can construct the confidence interval and the testing procedure, using the t distribution under the normality assumption or the normal distribution without the normality assumption.

4 Properties of OLSE

1. Properties of $\hat{\beta}$: **BLUE (best linear unbiased estimator，最良線形不偏推定量)**, i.e., minimum variance within the class of linear unbiased estimators **(Gauss-Markov theorem，ガウス・マルコフの定理)**

   Proof:

   Consider another linear unbiased estimator, which is denoted by $\tilde{\beta} = Cy$.

   $$\tilde{\beta} = Cy = C(X\beta + u) = CX\beta + Cu,$$

   where C is a $k \times n$ matrix.

   Taking the expectation of $\tilde{\beta}$, we obtain:

   $$\mathrm{E}(\tilde{\beta}) = CX\beta + C\mathrm{E}(u) = CX\beta$$

   Because we have assumed that $\tilde{\beta} = Cy$ is unbiased, $\mathrm{E}(\tilde{\beta}) = \beta$ holds.

That is, we need the condition: $CX = I_k$.

Next, we obtain the variance of $\tilde{\beta} = Cy$.

$$\tilde{\beta} = C(X\beta + u) = \beta + Cu.$$

Therefore, we have:

$$\mathrm{V}(\tilde{\beta}) = \mathrm{E}((\tilde{\beta} - \beta)(\tilde{\beta} - \beta)') = \mathrm{E}(Cuu'C') = \sigma^2 CC'$$

Defining $C = D + (X'X)^{-1}X'$, $\mathrm{V}(\tilde{\beta})$ is rewritten as:

$$\mathrm{V}(\tilde{\beta}) = \sigma^2 CC' = \sigma^2 (D + (X'X)^{-1}X')(D + (X'X)^{-1}X')'.$$

Moreover, because $\hat{\beta}$ is unbiased, we have the following:

$$CX = I_k = (D + (X'X)^{-1}X')X = DX + I_k.$$

Therefore, we have the following condition:

$$DX = 0.$$

Accordingly, $\mathrm{V}(\tilde{\beta})$ is rewritten as:

$$\begin{aligned}\mathrm{V}(\tilde{\beta}) &= \sigma^2 CC' = \sigma^2 (D + (X'X)^{-1}X')(D + (X'X)^{-1}X')' \\ &= \sigma^2 (X'X)^{-1} + \sigma^2 DD' = \mathrm{V}(\hat{\beta}) + \sigma^2 DD'\end{aligned}$$

Thus, $\mathrm{V}(\tilde{\beta}) - \mathrm{V}(\hat{\beta})$ is a positive definite matrix.

$\Longrightarrow \mathrm{V}(\tilde{\beta}_i) - \mathrm{V}(\hat{\beta}_i) > 0$

$\Longrightarrow \hat{\beta}$ is a minimum variance (i.e., best) linear unbiased estimator of β.

Note as follows:

$\Longrightarrow$ A is positive definite when $d'Ad > 0$ except $d = 0$.

$\Longrightarrow$ The ith diagonal element of A, i.e., a_{ii}, is positive (choose d such that the ith element of d is one and the other elements are zeros).

[Review] F **Distribution:**

Suppose that $U \sim \chi(n)$, $V \sim \chi(m)$, and U is independent of V.

Then, $\dfrac{U/n}{V/m} \sim F(n, m)$.

[End of Review]

F Distribution ($H_0: \beta = 0$): Final Result in this Section:

$$\frac{(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)/k}{e'e/(n-k)} \sim F(k, n-k).$$

Consider the numerator and the denominator, separately.

1. If $u \sim N(0, \sigma^2 I_n)$, then $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$.

   Therefore, $\dfrac{(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2(k)$.

2. **Proof:**

   Using $\hat{\beta} - \beta = (X'X)^{-1}X'u$, we obtain:

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

Note that $X(X'X)^{-1}X'$ is symmetric and idempotent, i.e., $A'A = A$.

$$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2\Big(\mathrm{tr}(X(X'X)^{-1}X')\Big)$$

The degree of freedom is given by:

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

Therefore, we obtain:

$$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(k)$$

3. (*) Formula:

   Suppose that $X \sim N(0, I_k)$.

   If A is symmetric and idempotent, i.e., $A'A = A$, then $X'AX \sim \chi^2(\mathrm{tr}(A))$.

   Here, $X = \dfrac{1}{\sigma}u \sim N(0, I_n)$ from $u \sim N(0, \sigma^2 I_n)$, and $A = X(X'X)^{-1}X'$.

4. **Sum of Residuals:** e is rewritten as:

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

Therefore, the sum of residuals is given by:

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

Note that $I_n - X(X'X)^{-1}X'$ is symmetric and idempotent.

We obtain the following result:

$$\frac{e'e}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2\Big(\mathrm{tr}(I_n - X(X'X)^{-1}X')\Big),$$

where the trace is:

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

Therefore, we have the following result:

$$\frac{e'e}{\sigma^2} = \frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k),$$

where

$$s^2 = \frac{1}{n-k}e'e.$$

5. We show that $\hat{\beta}$ is independent of e.

Proof:

Because $u \sim N(0, \sigma^2 I_n)$, we show that $\mathrm{Cov}(e, \hat{\beta}) = 0$.

$$\mathrm{Cov}(e, \hat{\beta}) = \mathrm{E}(e(\hat{\beta} - \beta)') = \mathrm{E}\Big((I_n - X(X'X)^{-1}X')u((X'X)^{-1}X'u)'\Big)$$

$$= \mathrm{E}\Big((I_n - X(X'X)^{-1}X')uu'X(X'X)^{-1}\Big) = (I_n - X(X'X)^{-1}X')\mathrm{E}(uu')X(X'X)^{-1}$$

$$= (I_n - X(X'X)^{-1}X')(\sigma^2 I_n)X(X'X)^{-1} = \sigma^2(I_n - X(X'X)^{-1}X')X(X'X)^{-1}$$

$$= \sigma^2(X(X'X)^{-1} - X(X'X)^{-1}X'X(X'X)^{-1}) = \sigma^2(X(X'X)^{-1} - X(X'X)^{-1}) = 0.$$

$\hat{\beta}$ is independent of e, because of normality assumption on u

[Review]

- Suppose that X is independent of Y. Then, $\text{Cov}(X, Y) = 0$. However, $\text{Cov}(X, Y) = 0$ does not mean in general that X is independent of Y.

- In the case where X and Y are normal, $\text{Cov}(X, Y) = 0$ indicates that X is independent of Y.

[End of Review]

[Review] Formulas — F Distribution:

- $\dfrac{U/n}{V/m} \sim F(n,m)$ when U
$sim \chi^2(n)$, $V \sim \chi^2(m)$, and U is independent of V.

- When $X \sim N(0, I_n)$, A and B are $n \times n$ symmetric idempotent matrices, $\text{Rank}(A) = \text{tr}(A) = G$, $\text{Rank}(B) = \text{tr}(B) = K$ and $AB = 0$, then $\dfrac{X'AX/G}{X'BX/K} \sim F(G, K)$.

Note that the covariance of AX and BX is zero, which implies that AX is independent of BX under normality of X.

[End of Review]

6. Therefore, we obtain the following distribution:

$$\frac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{\sigma^2} = \frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(k),$$

$$\frac{e'e}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2(n-k)$$

$\hat\beta$ is independent of e, because $X(X'X)^{-1}X'(I_n - X(X'X)^{-1}X') = 0$.

Accordingly, we can derive:

$$\frac{\dfrac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)}{\sigma^2}\Big/ k}{\dfrac{e'e}{\sigma^2}\Big/(n-k)} = \frac{(\hat\beta-\beta)'X'X(\hat\beta-\beta)/k}{s^2} \sim F(k, n-k)$$

Under the null hypothesis $H_0: \ \beta = 0$, $\dfrac{\hat\beta' X' X\hat\beta/k}{s^2} \sim F(k, n-k)$.

Given data, $\dfrac{\hat\beta' X' X\hat\beta/k}{s^2}$ is compared with $F(k, n-k)$.

If $\dfrac{\hat\beta' X' X\hat\beta/k}{s^2}$ is in the tail of the F distribution, the null hypothesis is rejected.