

[Review] Trace (トレース):

1. $A: n \times n$, $\text{tr}(A) = \sum_{i=1}^n a_{ii}$, where a_{ij} denotes an element in the i th row and the j th column of a matrix A .
2. a : scalar (1×1), $\text{tr}(a) = a$
3. $A: n \times k$, $B: k \times n$, $\text{tr}(AB) = \text{tr}(BA)$
4. $\text{tr}(X(X'X)^{-1}X') = \text{tr}((X'X)^{-1}X'X) = \text{tr}(I_k) = k$
5. When X is a square matrix of random variables, $E(\text{tr}(AX)) = \text{tr}(AE(X))$

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:

$$\begin{aligned} (I_n - X(X'X)^{-1}X')(I_n - X(X'X)^{-1}X') &= I_n - X(X'X)^{-1}X', \\ (I_n - X(X'X)^{-1}X')' &= I_n - X(X'X)^{-1}X'. \end{aligned}$$

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 $\text{tr}(a) = a$ for a scalar a .

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

→ 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 $X \sim N(\mu, \Sigma)$.
- $\frac{X'X}{\sigma^2} \sim \chi^2(n)$ for $X \sim N(0, \sigma^2 I_n)$.
- $\frac{X'AX}{\sigma^2} \sim \chi^2(G)$, 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:

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

Note as follows:

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

\implies The i th diagonal element of A , i.e., a_{ii} , is positive (choose d such that the i th 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, $\frac{U/n}{V/m} \sim F(n, m)$.

[End of Review]