

### 3 Multiple Regression Model (重回帰モデル)

Up to now, only one independent variable, i.e.,  $x_i$ , is taken into the regression model.

In this section, we extend it to more independent variables, which is called the **multiple regression model** (重回帰モデル).

We consider the following regression model:

$$\begin{aligned}y_i &= \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k} + u_i \\ &= (x_{i,1}, x_{i,2}, \cdots, x_{i,k}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + u_i \\ &= x_i \beta + u_i,\end{aligned}$$

for  $i = 1, 2, \cdots, n$ ,

where  $x_i$  and  $\beta$  denote a  $1 \times k$  vector of the independent variables and a  $k \times 1$  vector

of the unknown parameters to be estimated, which are represented as:

$$x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,k}), \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

$x_{i,j}$  denotes the  $i$ th observation of the  $j$ th independent variable.

The case of  $k = 2$  and  $x_{i,1} = 1$  for all  $i$  is exactly equivalent to (1).

Therefore, the matrix form above is a generalization of (1).

Writing all the equations for  $i = 1, 2, \dots, n$ , we have:

$$y_1 = \beta_1 x_{1,1} + \beta_2 x_{1,2} + \dots + \beta_k x_{1,k} + u_1 = x_1 \beta + u_1,$$

$$y_2 = \beta_1 x_{2,1} + \beta_2 x_{2,2} + \dots + \beta_k x_{2,k} + u_2 = x_2 \beta + u_2,$$

$\vdots$

$$y_n = \beta_1 x_{n,1} + \beta_2 x_{n,2} + \dots + \beta_k x_{n,k} + u_n = x_n \beta + u_n,$$

which is rewritten as:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$
$$= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Again, the above equation is compactly rewritten as:

$$y = X\beta + u, \quad (18)$$

where  $y$ ,  $X$  and  $u$  are denoted by:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Utilizing the matrix form (18), we derive the ordinary least squares estimator of  $\beta$ , denoted by  $\hat{\beta}$ .

In (18), replacing  $\beta$  by  $\hat{\beta}$ , we have the following equation:

$$y = X\hat{\beta} + e,$$

where  $e$  denotes a  $n \times 1$  vector of the residuals.

The  $i$ th element of  $e$  is given by  $e_i$ .

The sum of squared residuals is written as follows:

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

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, 最小自乘推定量)** of  $\beta$  is given by:

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

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:

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

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

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

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

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

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

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

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  $u_1, u_2, \dots, u_n$  are assumed to be mutually independently and identically distributed with mean zero and variance  $\sigma^2$ .

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:

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

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

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

$$\phi_u(\theta_u) \equiv \text{E}(\exp(\theta_u' u)) = \exp\left(\frac{\sigma^2}{2} \theta_u' \theta_u\right),$$

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

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

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

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

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

QED

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

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

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

Replacing  $\sigma^2$  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.

$s^2$  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  $\sigma^2$ .

**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  $\text{tr}(a) = a$  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  $\sigma^2$ .

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

## 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 ( $1 \times 1$ ),  $\text{tr}(a) = a$
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 vector of random variables,  $E(\text{tr}(X)) = \text{tr}(E(X))$