X is independent of Y, i.e., $\mathrm{E}\Big((X-\mu_x)(Y-\mu_y)'\Big) = 0$ in the case of normal random variables.

$$\frac{(X-\mu_x)'\Sigma_x^{-1}(X-\mu_x)/n}{(Y-\mu_y)'\Sigma_y^{-1}(Y-\mu_y)/m} \sim F(n,m)$$

4. If $X \sim N(0,\sigma^2 I_n)$ and A is a symmetric idempotent $n\times n$ matrix of rank G, then $X'AX/\sigma^2 \sim \chi^2(G)$.

   Note that $X'AX = (AX)'(AX)$ and $\mathrm{rank}(A) = \mathrm{tr}(A)$ because A is idempotent.

5. If $X \sim N(0,\sigma^2 I_n)$, A and B are symmetric idempotent $n\times n$ matrices of rank G and K, and $AB = 0$, then

$$\frac{X'AX}{G\sigma^2}\Big/\frac{X'BX}{K\sigma^2} = \frac{X'AX/G}{X'BX/K} \sim F(G,K).$$

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.

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:

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

for $i = 1, 2, \cdots, n$, where x_i and β 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 given by:

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

$x_{i,j}$ denotes the ith observation of the jth 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 $i = 1, 2, \cdots, n$, we have:

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

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:

$$y = X\beta + u, \tag{18}$$

where y, X and u are denoted by:

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

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

In (18), replacing β 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 ith element of e is given by e_i.

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:

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

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:

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

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:

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

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

In the fifth equality, $\mathrm{E}(uu') = \sigma^2 I_n$ is used, which implies that $\mathrm{E}(u_i^2) = \sigma^2$ for all i and $\mathrm{E}(u_i u_j) = 0$ 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: $k \times 1$, u: $n \times 1$, $\hat{\beta}$: $k \times 1$

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