Finally, replacing σ^2 by its consistent estimator s^2, it is known as follows:

$$\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \longrightarrow N(0, 1), \tag{16}$$

where s^2 is defined as:

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2, \tag{17}$$

which is a consistent and unbiased estimator of σ^2. $\longrightarrow$ Proved later.

Thus, using (16), in large sample we can construct the confidence interval and test the hypothesis.

[Review] Confidence Interval (信頼区間，区間推定)):

Suppose $X_1, X_2, \cdots, X_n$ are iid with mean μ and variance σ^2. $\longrightarrow$ No N assumption

From CLT, $\dfrac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$.

Replacing σ^2 by $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$, we have: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \longrightarrow N(0,1)$.

That is, for large n,

$$P\Big(-1.96 < \frac{\overline{X} - \mu}{S/\sqrt{n}} < 1.96\Big) = 0.95, \text{ i.e., } P\Big(\overline{X} - 1.96\frac{S}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{S}{\sqrt{n}}\Big) = 0.95.$$

Note that 1.96 is obtained from the normal distribution table.

Then, replacing the estimators $\overline{X}$ and S^2 by the estimates $\overline{x}$ and s^2, we obtain the 95% confidence interval of μ as follows:

$$(\overline{x} - 1.96\frac{s}{\sqrt{n}},\ \overline{x} + 1.96\frac{s}{\sqrt{n}}).$$

[End of Review]

Going back to OLS, we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \longrightarrow N(0, 1).$$

Therefore,

$$P\Big(-2.576 < \frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < 2.576\Big) = 0.99,$$

i.e.,

$$P\Big(\hat{\beta}_2 - 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < \beta_2 < \hat{\beta}_2 + 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big) = 0.99.$$

Note that 2.576 is 0.005 value of $N(0, 1)$, which comes from the statistical table.

Thus, the 99% confidence interval of β_2 is:

$$\Big(\hat{\beta}_2 - 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}},\ \hat{\beta}_2 + 2.576\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big),$$

where $\hat{\beta}_2$ and s^2 should be replaced by the observed data.

[Review] Testing the Hypothesis (仮説検定):

Suppose that $X_1, X_2, \cdots, X_n$ are iid with mean μ and variance σ^2.

From CLT, $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \longrightarrow N(0,1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$, which is known as the unbiased estimator of σ^2.

- The null hypothesis $H_0: \ \mu = \mu_0$, where μ_0 is a fixed number.
- The alternative hypothesis $H_1: \ \mu \neq \mu_0$

Under the null hypothesis, in large sample we have the following disribution:

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim N(0,1).$$

Replacing $\overline{X}$ and S^2 by $\overline{x}$ and s^2, compare $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}$ and $N(0,1)$.

H_0 is rejected at significance level 0.05 when $\left|\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| > 1.96$.

[End of Review]

In the case of OLS, the hypotheses are as follows:

- The null hypothesis $H_0 : \beta_2 = \beta_2^*$
- The alternative hypothesis $H_1 : \beta_2 \neq \beta_2^*$

Under H_0, in large sample,

$$\frac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \sim N(0,1).$$

Replacing $\hat{\beta}_2$ and s^2 by the observed data, compare $\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$ and $N(0,1)$.

H_0 is rejected at significance level 0.05 when $\left|\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right| > 1.96$.

Exact Distribution of $\hat{\beta}_2$: We have shown asymptotic normality of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$, which is one of the large sample properties.

Now, we discuss the small sample properties of $\hat{\beta}_2$.

In order to obtain the distribution of $\hat{\beta}_2$ in small sample, the distribution of the error term has to be assumed.

Therefore, the extra assumption is that $u_i \sim N(0, \sigma^2)$.

Writing (13), again, $\hat{\beta}_2$ is represented as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i.$$

First, we obtain the distribution of the second term in the above equation.

[Review]   Content of Special Lectures in Economics (Statistical Analysis)

Note that the **moment-generating function (積率母関数, MGF)** is given by $M(\theta) \equiv \mathrm{E}(\exp(\theta X)) = \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2)$ when $X \sim N(\mu, \sigma^2)$.

$X_1, X_2, \cdots, X_n$ are mutually independently distributed as $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, \cdots, n$.

MGF of X_i is $M_i(\theta) \equiv \mathrm{E}(\exp(\theta X_i)) = \exp(\mu_i\theta + \frac{1}{2}\sigma_i^2\theta^2)$.

Consider the distribution of $Y = \sum_{i=1}^n (a_i + b_i X_i)$, where a_i and b_i are constant.

$$
\begin{aligned}
M_y(\theta) &\equiv \mathrm{E}(\exp(\theta Y)) = \mathrm{E}(\exp(\theta \textstyle\sum_{i=1}^n (a_i + b_i X_i))) \\
&= \textstyle\prod_{i=1}^n \exp(\theta a_i)\mathrm{E}(\exp(\theta b_i X_i)) = \prod_{i=1}^n \exp(\theta a_i) M_i(\theta b_i) \\
&= \textstyle\prod_{i=1}^n \exp(\theta a_i)\exp(\mu_i \theta b_i + \frac{1}{2}\sigma_i^2(\theta b_i)^2) = \exp(\theta \sum_{i=1}^n (a_i + b_i\mu_i) + \frac{1}{2}\theta^2 \sum_{i=1}^n b_i^2\sigma_i^2),
\end{aligned}
$$

which implies that $Y \sim N(\sum_{i=1}^n (a_i + b_i\mu_i), \sum_{i=1}^n b_i^2\sigma_i^2)$.

[End of Review]

Substitute $a_i = 0$, $\mu_i = 0$, $b_i = \omega_i$ and $\sigma_i^2 = \sigma^2$.

Then, using the moment-generating function, $\sum_{i=1}^n \omega_i u_i$ is distributed as:

$$\sum_{i=1}^{n} \omega_i u_i \sim N(0,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2).$$

Therefore, $\hat{\beta}_2$ is distributed as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i \sim N(\beta_2,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2),$$

or equivalently,

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{\sum_{i=1}^n \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} \sim N(0, 1),$$

for any n.

Moreover, replacing σ^2 by its estimator s^2 defined in (17), it is known that we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \sim t(n-2),$$

where $t(n-2)$ denotes t distribution with $n-2$ degrees of freedom.

Thus, under normality assumption on the error term u_i, the $t(n-2)$ distribution is used for the confidence interval and the testing hypothesis in small sample.

Or, taking the square on both sides,

$$\Big(\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big)^2 \sim F(1, n-2),$$

which will be proved later.

Before going to **multiple regression model (**重回帰モデル**)**,

2 Some Formulas of Matrix Algebra

1. Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lk} \end{pmatrix} = [a_{ij}]$,

   which is a $l \times k$ matrix, where a_{ij} denotes ith row and jth column of A.

   The **transposed matrix (転置行列)** of A, denoted by A', is defined as:

   $$A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{l1} \\ a_{12} & a_{22} & \cdots & a_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{lk} \end{pmatrix} = [a_{ji}],$$

   where the ith row of A' is the ith column of A.

2. $(Ax)' = x'A'$,

where A and x are a $l \times k$ matrix and a $k \times 1$ vector, respectively.

3. $a' = a$,

where a denotes a scalar.

4. $\dfrac{\partial a'x}{\partial x} = a$,

where a and x are $k \times 1$ vectors.

5. $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$,

where A and x are a $k \times k$ matrix and a $k \times 1$ vector, respectively.

Especially, when A is symmetric,

$$\frac{\partial x'Ax}{\partial x} = 2Ax.$$

6. Let A and B be $k \times k$ matrices, and I_k be a $k \times k$ **identity matrix (単位行列)** (one in the diagonal elements and zero in the other elements).

   When $AB = I_k$, B is called the **inverse matrix (逆行列)** of A, denoted by $B = A^{-1}$.

   That is, $AA^{-1} = A^{-1}A = I_k$.

7. Let A be a $k \times k$ matrix and x be a $k \times 1$ vector.

   If A is a **positive definite matrix (正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax > 0.$$

   If A is a **positive semidefinite matrix (非負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \geq 0.$$

If A is a **negative definite matrix (負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax < 0.$$

If A is a **negative semidefinite matrix (非正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \leq 0.$$

Trace, Rank and etc.: $A : k \times k, \qquad B : n \times k, \qquad C : k \times n.$

1. The **trace (トレース)** of A is: $\mathrm{tr}(A) = \sum_{i=1}^{k} a_{ii}$, where $A = [a_{ij}]$.
2. The **rank (ランク, 階数)** of A is the maximum number of linearly independent column (or row) vectors of A, which is denoted by $\mathrm{rank}(A)$.

3. If A is an **idempotent matrix (べき等行列)**, $A = A^2$.

4. If A is an idempotent and symmetric matrix, $A = A^2 = A'A$.

5. A is idempotent if and only if the eigen values of A consist of 1 and 0.

6. If A is idempotent, $\text{rank}(A) = \text{tr}(A)$.

7. $\text{tr}(BC) = \text{tr}(CB)$

Distributions in Matrix Form:

1. Let X, μ and Σ be $k \times 1$, $k \times 1$ and $k \times k$ matrices.

   When $X \sim N(\mu, \Sigma)$, the density function of X is given by:

$$f(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big).$$

$\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) = \mathrm{E}\Big((X-\mu)(X-\mu)'\Big) = \Sigma$

The moment-generating function: $\phi(\theta) = \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta)$

(*) In the univariate case, when $X \sim N(\mu, \sigma^2)$, the density function of X is:

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big).$$

2. If $X \sim N(\mu, \Sigma)$, then $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(k)$.

   Note that $X'X \sim \chi^2(k)$ when $X \sim N(0, I_k)$.

3. X: $n \times 1$, Y: $m \times 1$, $X \sim N(\mu_x, \Sigma_x)$, $Y \sim N(\mu_y, \Sigma_y)$

   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 $\text{rank}(A) = \text{tr}(A)$ 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

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

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:

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