When these conditions hold, we can rewrite $\widetilde{\beta}_2$ as:

$$\widetilde{\beta}_2 = \beta_2 + \sum_{i=1}^n (\omega_i + d_i) u_i.$$

The variance of $\widetilde{\beta}_2$ is derived as:

$$\begin{aligned}
\mathrm{V}(\widetilde{\beta}_2) &= \mathrm{V}\Big(\beta_2 + \sum_{i=1}^n (\omega_i + d_i) u_i\Big) = \mathrm{V}\Big(\sum_{i=1}^n (\omega_i + d_i) u_i\Big) = \sum_{i=1}^n \mathrm{V}\Big((\omega_i + d_i) u_i\Big) \\
&= \sum_{i=1}^n (\omega_i + d_i)^2 \mathrm{V}(u_i) = \sigma^2 (\sum_{i=1}^n \omega_i^2 + 2\sum_{i=1}^n \omega_i d_i + \sum_{i=1}^n d_i^2) \\
&= \sigma^2 (\sum_{i=1}^n \omega_i^2 + \sum_{i=1}^n d_i^2).
\end{aligned}$$

From unbiasedness of $\widetilde{\beta}_2$, using $\sum_{i=1}^n d_i = 0$ and $\sum_{i=1}^n d_i x_i = 0$, we obtain:

$$\sum_{i=1}^n \omega_i d_i = \frac{\sum_{i=1}^n (x_i - \overline{x}) d_i}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n x_i d_i - \overline{X} \sum_{i=1}^n d_i}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0,$$

which is utilized to obtain the variance of $\widetilde{\beta}_2$ in the third line of the above equation.
From (15), the variance of $\hat{\beta}_2$ is given by: $\mathrm{V}(\hat{\beta}_2) = \sigma^2 \sum_{i=1}^n \omega_i^2$.

Therefore, we have:

$$\mathrm{V}(\widetilde{\beta}_2) \geq \mathrm{V}(\hat{\beta}_2),$$

because of $\sum_{i=1}^n d_i^2 \geq 0$.

When $\sum_{i=1}^n d_i^2 = 0$, i.e., when $d_1 = d_2 = \cdots = d_n = 0$, we have the equality: $\mathrm{V}(\widetilde{\beta}_2) = \mathrm{V}(\hat{\beta}_2)$.

Thus, in the case of $d_1 = d_2 = \cdots = d_n = 0$, $\hat{\beta}_2$ is equivalent to $\widetilde{\beta}_2$.

As shown above, the least squares estimator $\hat{\beta}_2$ gives us the **minimum variance linear unbiased estimator (最小分散線形不偏推定量)**, or equivalently the **best linear unbiased estimator (最良線形不偏推定量，BLUE)**, which is called the **Gauss-Markov theorem (ガウス・マルコフ定理)**.

Asymptotic Properties of $\hat{\beta}_2$: We assume that as n goes to infinity we have the following:

$$\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 \longrightarrow m < \infty,$$

where m is a constant value. From (12), we obtain:

$$n \sum_{i=1}^n \omega_i^2 = \frac{1}{(1/n)\sum_{i=1}^n (x_i - \overline{x})} \longrightarrow \frac{1}{m}.$$

Note that $f(x_n) \longrightarrow f(m)$ when $x_n \longrightarrow m$, called **Slutsky's theorem (スルツキー定理)**, where m is a constant value and $f(\cdot)$ is a function.

We show both consistency of $\hat{\beta}_2$ and asymptotic normality of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$.
● First, we prove that $\hat{\beta}_2$ is a consistent estimator of β_2.
Chebyshev's inequality is given by:

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

where $\mu = \mathrm{E}(X)$ and $\sigma^2 = \mathrm{V}(X)$.
Replace X, $\mathrm{E}(X)$ and $\mathrm{V}(X)$ by:

$$\hat{\beta}_2, \qquad \mathrm{E}(\hat{\beta}_2) = \beta_2, \quad \text{and} \quad \mathrm{V}(\hat{\beta}_2) = \sigma^2 \sum_{i=1}^n \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})},$$

respectively.

Then, when $n \longrightarrow \infty$, we obtain the following result:

$$P(|\hat{\beta}_2 - \beta_2| > \epsilon) \leq \frac{\sigma^2 \sum_{i=1}^n \omega_i^2}{\epsilon^2} = \frac{\sigma^2 n \sum_{i=1}^n \omega_i^2}{n\epsilon^2} \longrightarrow 0,$$

where $\sum_{i=1}^n \omega_i^2 \longrightarrow 0$ because $n \sum_{i=1}^n \omega_i^2 \longrightarrow \dfrac{1}{m}$ from the assumption.

Thus, we obtain the result that $\hat{\beta}_2 \longrightarrow \beta_2$ as $n \longrightarrow \infty$.

Therefore, we can conclude that $\hat{\beta}_2$ is a consistent estimator of β_2.

● Next, we want to show that $\sqrt{n}(\hat\beta_2 - \beta_2)$ is asymptotically normal.

Note that $\hat\beta_2 = \beta_2 + \sum_{i=1}^n \omega_i u_i$ as in (13).

From the central limit theorem, asymptotic normality is shown as follows:

$$\frac{\sum_{i=1}^n \omega_i u_i - \mathrm{E}(\sum_{i=1}^n \omega_i u_i)}{\sqrt{\mathrm{V}(\sum_{i=1}^n \omega_i u_i)}} = \frac{\sum_{i=1}^n \omega_i u_i}{\sigma\sqrt{\sum_{i=1}^n \omega_i^2}} = \frac{\hat\beta_2 - \beta_2}{\sigma/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} \longrightarrow N(0,1),$$

where $\mathrm{E}(\sum_{i=1}^n \omega_i u_i) = 0$, $\mathrm{V}(\sum_{i=1}^n \omega_i u_i) = \sigma^2 \sum_{i=1}^n \omega_i^2$, and $\sum_{i=1}^n \omega_i u_i = \hat\beta_2 - \beta_2$ are substituted in the first and second equalities.

Moreover, we can rewrite as follows:

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

or equivalently,

$$\sqrt{n}(\hat\beta_2 - \beta_2) \longrightarrow N(0, \frac{\sigma^2}{m}).$$

Thus, the asymptotic normality of $\sqrt{n}(\hat\beta_2 - \beta_2)$ is shown.

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.

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.

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.: $A : k \times k,\quad B : n \times k,\quad C : k \times n.$

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, $\mathrm{rank}(A) = \mathrm{tr}(A)$.

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

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