Mean and Variance of $\hat{\beta}_2$: $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2, but they are not necessarily normal.

Remember that we do not need normality assumption to obtain mean and variance but the normality assumption is required to test a hypothesis.

From (16), the expectation of $\hat{\beta}_2$ is derived as follows:

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \mathrm{E}(\sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) = \beta_2. \tag{17}$$

It is shown from (17) that the ordinary least squares estimator $\hat{\beta}_2$ is an **unbiased estimator (不偏推定量)** of β_2.

From (16), the variance of $\hat{\beta}_2$ is computed as:

$$
\begin{aligned}
\mathrm{V}(\hat{\beta}_2) &= \mathrm{V}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sum_{i=1}^{n} \mathrm{V}(\omega_i u_i) = \sum_{i=1}^{n} \omega_i^2 \mathrm{V}(u_i) \\
&= \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \qquad (18)
\end{aligned}
$$

The third equality holds because $u_1, u_2, \cdots, u_n$ are mutually independent.

The last equality comes from (15).

Thus, $\mathrm{E}(\hat{\beta}_2)$ and $\mathrm{V}(\hat{\beta}_2)$ are given by (17) and (18).

Gauss-Markov Theorem (ガウス・マルコフ定理): $\hat{\beta}_2$ has minimum variance within a class of the linear unbiased estimators.

$\longrightarrow$ **best linear unbiased estimator (BLUE, 最良線型不偏推定量)**

(Proof is omitted.)

Distribution of $\hat{\beta}_2$: 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 (16), 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.

It is well known that sum of normal random variables results in a normal distribution.

Therefore, $\sum_{i=1}^{n} \omega_i u_i$ 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 $s^2 = \dfrac{1}{n-2} \displaystyle\sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$, 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,

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

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

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we can obtain: $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$.

That is,

$$P\Big(-t_{\alpha/2}(n-1) < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\Big) = 1-\alpha$$

i.e.,

$$P\Big(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\Big) = 1-\alpha.$$

Note that $t_{\alpha/2}(n-1)$ is obtained from the t distribution table, given α and $n-1$.

Then, replacing $\overline{X}$ by $\overline{x}$, we obtain the $100(1-\alpha)\%$ confidence interval of μ as follows:

$$(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}).$$

[End of Review]

In the case of OLS,

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

where $t_{\alpha/2}(n-2)$ denotes $100 \times \alpha/2\%$ point from the $t(n-2)$ distribution.

Rewriting,

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

Replacing $\hat{\beta}_2$ and s^2 by observed data, the $100(1-\alpha)\%$ confidence interval of β_2 is given by:

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

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

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we obtain: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, 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, we have the disribution: $\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$.

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

H_0 is rejected when $\left|\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| > t_{\alpha/2}(n-1)$.

$t_{\alpha/2}(n-1)$ is obtained from the significance level α and the degrees of freedom $n-1$.

[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,

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

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 $t(n-2)$.

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

(*) $\hat{\beta}_2$ = Coefficient, $\dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ = Standard Error,

s = Standard Error of Regression

3　多重回帰

n 組のデータ $(Y_i, X_{1i}, X_{2i}, \cdots, X_{ki})$, $i = 1, 2, \cdots, n$ を用いて，k 変数の多重回帰モデルを考える。

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

ただし，X_{ji} は j 番目の説明変数の第 i 番目の観測値を表す。u_i は誤差項 (または，攪乱項) で，同じ仮定を用いる (すなわち，$u_1, u_2, \cdots, u_n$ は互いに独立に，平均ゼロ，分散 σ^2 の正規分布に従う)。

$\beta_1, \beta_2, \cdots, \beta_k$ は推定されるべきパラメータである。

すべての i について，$X_{1i} = 1$ とすれば，β_1 は定数項として表される。

次のような関数 $S(\beta_1, \beta_2, \cdots, \beta_k)$ を定義する。

$$S(\beta_1, \beta_2, \cdots, \beta_k) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2$$

このとき,

$$\min_{\beta_1,\beta_2,\cdots,\beta_k} S(\beta_1,\beta_2,\cdots,\beta_k)$$

となるような $\beta_1,\beta_2,\cdots,\beta_k$ を求める。$\Longrightarrow$ 最小自乗法

このときの解を $\widehat{\beta}_1,\widehat{\beta}_2,\cdots,\widehat{\beta}_k$ とする。

最小化のためには,

$$\frac{\partial S(\beta_1,\beta_2,\cdots,\beta_k)}{\partial\beta_1}=0,\qquad \frac{\partial S(\beta_1,\beta_2,\cdots,\beta_k)}{\partial\beta_2}=0,\qquad \cdots,\qquad \frac{\partial S(\beta_1,\beta_2,\cdots,\beta_k)}{\partial\beta_k}=0$$

を満たす $\beta_1,\beta_2,\cdots,\beta_k$ が $\widehat{\beta}_1,\widehat{\beta}_2,\cdots,\widehat{\beta}_k$ となる。

すなわち，$\widehat{\beta}_1,\widehat{\beta}_2,\cdots,\widehat{\beta}_k$ は,

$$\sum_{i=1}^{n}(Y_i-\widehat{\beta}_1X_{1i}-\widehat{\beta}_2X_{2i}-\cdots-\widehat{\beta}_kX_{ki})X_{1i}=0,$$
$$\sum_{i=1}^{n}(Y_i-\widehat{\beta}_1X_{1i}-\widehat{\beta}_2X_{2i}-\cdots-\widehat{\beta}_kX_{ki})X_{2i}=0,$$

$$\vdots$$

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})X_{ki} = 0,$$

を満たす。

さらに,

$$\sum_{i=1}^{n} X_{1i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} X_{1i}X_{2i} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{1i}X_{ki},$$

$$\sum_{i=1}^{n} X_{2i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{2i}X_{ki},$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ki}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{ki} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}X_{ki} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{ki}^2,$$

行列表示によって，

$$\begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix},$$

が得られ，$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ についてまとめると，

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix},$$

を解くことになる。$\Longrightarrow$ コンピュータによって計算

3.1 推定量の性質

$\beta_1, \beta_2, \cdots, \beta_k$ の最小二乗推定量は $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

誤差項 (または，攪乱項) u_i の分散 σ^2 の推定量 s^2 は,

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})^2$$

として表される。

このとき,

$$\mathrm{E}(\widehat{\beta}_j) = \beta_j, \qquad \mathrm{E}(s^2) = \sigma^2,$$

を証明することが出来る。(証明略)

分布について： $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ の分散は以下のように表される。

$$
\begin{aligned}
\mathrm{V}\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix}
&= \begin{pmatrix}
\mathrm{V}(\widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_k) \\
\mathrm{Cov}(\widehat{\beta}_2, \widehat{\beta}_1) & \mathrm{V}(\widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_2, \widehat{\beta}_k) \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_2) & \cdots & \mathrm{V}(\widehat{\beta}_k)
\end{pmatrix} \\
&= \sigma^2 \begin{pmatrix}
\sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\
\sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\
\vdots & \vdots & \ddots & \vdots \\
\sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2
\end{pmatrix}^{-1}
\end{aligned}
$$

$\widehat{\beta}_j$ の分散 (すなわち，上の逆行列の j 番目の対角要素) を，

$$
\mathrm{V}(\widehat{\beta}_j) = \sigma^2_{\widehat{\beta}_j},
$$

として，その推定量を $s^2_{\widehat{\beta}_j}$ とする。

このとき，

$$\widehat{\beta}_j \sim N(\beta_j, \sigma^2_{\widehat{\beta}_j}),$$

となり，標準化すると，

$$\frac{\widehat{\beta}_j - \beta_j}{\sigma_{\widehat{\beta}_j}} \sim N(0, 1),$$

が得られる。さらに，

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k),$$

となり (証明略)，しかも，$\widehat{\beta}_j$ と s^2 の独立性から (証明略)，

$$\frac{\widehat{\beta}_j - \beta_j}{s_{\widehat{\beta}_j}} \sim t(n-k)$$

となる。

よって，通常の区間推定や仮説検定を行うことが出来る。

決定係数について：　また，決定係数 R^2 についても同様に表される。

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

ただし，$\widehat{Y}_i = \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki}$，$Y_i = \widehat{Y}_i + \widehat{u}_i$ である。

R^2 は，説明変数を増やすことによって，必ず大きくなる。なぜなら，説明変数が増えることによって，$\sum_{i=1}^{n} \widehat{u}_i^2$ が必ず減少するからである。

R^2 を基準にすると，被説明変数にとって意味のない変数でも，説明変数が多いほど，よりよいモデルということになる。この点を改善するために，自由度修正済み決定係数 $\overline{R}^2$ を用いる。

$$\overline{R}^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{u}_i^2/(n-k)}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n-1)},$$

$\sum_{i=1}^{n} \widehat{u}_i^2/(n-k)$ は u_i の分散 σ^2 の不偏推定量であり，$\sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n-1)$ は Y_i の分散の不偏推定量である。

R^2 と $\overline{R}^2$ との関係は，

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k},$$

となる。さらに，

$$\frac{1-\overline{R}^2}{1-R^2} = \frac{n-1}{n-k} \geq 1,$$

という関係から，$\overline{R}^2 \leq R^2$ という結果を得る。($k = 1$ のときのみに，等号が成り立つ。)

数値例：　今までと同じ数値例で，$\overline{R}^2$ を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

まず R^2 は,

$$R^2 = 1 - \frac{\sum \widehat{u}_i^2}{\sum Y_i^2 - n\overline{Y}^2} = 1 - \frac{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2}{35 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

となり，$\overline{R}^2$ は，

$$\overline{R}^2 = 1 - \frac{\sum \widehat{u}_i^2/(n-k)}{(\sum Y_i^2 - n\overline{Y}^2)/(n-1)} = 1 - \frac{2.30/(4-2)}{10.75/(4-1)} = 0.679$$

となる。

注意： R^2 や $\overline{R}^2$ を比較する場合，被説明変数が同じことが必要である。被説明変数が異なる場合 (例えば，被説明変数を上昇率とするかそのままの値を用いるかによって，被説明変数が異なる)，誤差項 u_i の標準誤差で比較すべきである (標準誤差の小さいモデルを採用する)。 $\Longrightarrow$ 関数型の選択