計量経済基礎

Tue., 8:50-10:20

場所：　文法経研究講義棟　**3** 階　**32** 番

1　最小二乗法について

経済理論に基づいた線型モデルの係数の値をデータから求める時に用いられる手法 $\Longrightarrow$ 最小二乗法

1.1　最小二乗法と回帰直線

$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ のように n 組のデータがあり，X_i と Y_i との間に以下の線型関係を想定する。

$$Y_i = \alpha + \beta X_i,$$

X_i は説明変数，Y_i は被説明変数，α, β はパラメータとそれぞれ呼ばれる。

上の式は回帰モデル (または，回帰式) と呼ばれる。目的は，切片 α と傾き β をデータ $\{(X_i, Y_i), i = 1, 2, \cdots, n\}$ から推定すること，

データについて：

1. タイム・シリーズ (時系列)・データ： i が時間を表す (第 i 期)。

2. クロス・セクション (横断面)・データ： i が個人や企業を表す (第 i 番目の家計，第 i 番目の企業)。

1.2 切片 α と傾き β の推定

次のような関数 $S(\alpha, \beta)$ を定義する。

$$S(\alpha, \beta) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

このとき,

$$\min_{\alpha, \beta} S(\alpha, \beta)$$

となるような α, β を求める (最小自乗法)。このときの解を $\widehat{\alpha}, \widehat{\beta}$ とする。

最小化のためには,

$$\frac{\partial S(\alpha,\beta)}{\partial \alpha} = 0$$

$$\frac{\partial S(\alpha,\beta)}{\partial \beta} = 0$$

を満たす α, β が $\widehat{\alpha}, \widehat{\beta}$ となる。　　すなわち，$\widehat{\alpha}, \widehat{\beta}$ は,

$$\sum_{i=1}^{n}(Y_i - \widehat{\alpha} - \widehat{\beta}X_i) = 0, \tag{1}$$

$$\sum_{i=1}^{n} X_i(Y_i - \widehat{\alpha} - \widehat{\beta}X_i) = 0, \tag{2}$$

を満たす。　　さらに,

$$\sum_{i=1}^{n} Y_i = n\widehat{\alpha} + \widehat{\beta}\sum_{i=1}^{n} X_i, \tag{3}$$

$$\sum_{i=1}^{n} X_iY_i = \widehat{\alpha}\sum_{i=1}^{n} X_i + \widehat{\beta}\sum_{i=1}^{n} X_i^2,$$

行列表示によって，

$$\begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix},$$

逆行列の公式：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$\widehat{\alpha}, \widehat{\beta}$ について，まとめて，

$$\begin{aligned} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} &= \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \\ &= \frac{1}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \begin{pmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \end{aligned}$$

さらに，$\widehat{\beta}$ について解くと，

$$\widehat{\beta} = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

連立方程式の (3) 式から，

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{X}$$

となる。ただし，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

とする。

数値例：　以下の数値例を使って，回帰式 $Y_i = \alpha + \beta X_i$ の α，β の推定値 $\widehat{\alpha}$，$\widehat{\beta}$ を求める。

i	Y_i	X_i
1	6	10
2	9	12
3	10	14
4	10	16

$\widehat{\alpha}$，$\widehat{\beta}$ を求めるための公式は

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$
$$\widehat{\alpha} = \overline{Y} - \widehat{\beta\overline{X}}$$

なので，必要なものは $\overline{X}$，$\overline{Y}$，$\displaystyle\sum_{i=1}^{n} X_i^2$，$\displaystyle\sum_{i=1}^{n} X_i Y_i$ である。

i	Y_i	X_i	X_iY_i	X_i^2
1	6	10	60	100
2	9	12	108	144
3	10	14	140	196
4	10	16	160	256
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$
	35	52	468	696
平均	$\overline{Y}$	$\overline{X}$		
	8.75	13		

よって，

$$\widehat{\beta} = \frac{468 - 4 \times 13 \times 8.75}{696 - 4 \times 13^2} = \frac{13}{20} = 0.65$$

$$\widehat{\alpha} = 8.75 - 0.65 \times 13 = 0.3$$

となる。

注意事項：

1. α, β は真の値で未知

2. $\widehat{\alpha}, \widehat{\beta}$ は α, β の推定値でデータから計算される

回帰直線は

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i,$$

として与えられる。

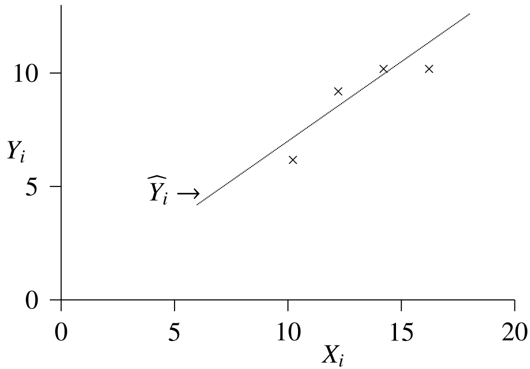
上の数値例では,

$$\widehat{Y}_i = 0.3 + 0.65 X_i$$

となる。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$
1	6	10	60	100	6.8
2	9	12	108	144	8.1
3	10	14	140	196	9.4
4	10	16	160	256	10.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$
	35	52	468	696	35.0
平均	$\overline{Y}$	$\overline{X}$			
	8.75	13			

図 2：Y_i，X_i，$\widehat{Y}_i$



$\widehat{Y}_i$ を実績値 Y_i の予測値または理論値と呼ぶ。

$$\widehat{u}_i = Y_i - \widehat{Y}_i,$$

$\widehat{u}_i$ を残差と呼ぶ。

$$Y_i = \widehat{Y}_i + \widehat{u}_i = \widehat{\alpha} + \widehat{\beta} X_i + \widehat{u}_i,$$

さらに，$\overline{Y}$ を両辺から引いて，

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + \widehat{u}_i,$$

1.3　残差 $\widehat{u}_i$ の性質について

$\widehat{u}_i = Y_i - \widehat{\alpha} - \widehat{\beta} X_i$ に注意して，(1) 式から，

$$\sum_{i=1}^{n} \widehat{u}_i = 0,$$

を得る。　　　(2) 式から，

$$\sum_{i=1}^{n} X_i \widehat{u}_i = 0,$$

を得る。　$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i$ から，

$$\sum_{i=1}^{n} \widehat{Y}_i \widehat{u}_i = 0,$$

を得る。なぜなら，

$$\begin{aligned}
\sum_{i=1}^{n} \widehat{Y}_i \widehat{u}_i &= \sum_{i=1}^{n} (\widehat{\alpha} + \widehat{\beta} X_i) \widehat{u}_i \\
&= \widehat{\alpha} \sum_{i=1}^{n} \widehat{u}_i + \widehat{\beta} \sum_{i=1}^{n} X_i \widehat{u}_i \\
&= 0
\end{aligned}$$

である。

i	Y_i	X_i	$\widehat{Y}_i$	$\widehat{u}_i$	$X_i\widehat{u}_i$	$\widehat{Y}_i\widehat{u}_i$
1	6	10	6.8	−0.8	−8.0	−5.44
2	9	12	8.1	0.9	10.8	7.29
3	10	14	9.4	0.6	8.4	5.64
4	10	16	10.7	−0.7	−11.2	−7.49
合計	$\sum Y_i$	$\sum X_i$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$	$\sum X_i\widehat{u}_i$	$\sum \widehat{Y}_i\widehat{u}_i$
	35	52	35.0	0.0	0.0	0.00

1.4　決定係数 R^2 について

次の式

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + \widehat{u}_i,$$

の両辺を二乗して，総和すると，

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n}\Big((\widehat{Y}_i - \overline{Y}) + \widehat{u}_i\Big)^2 \\ &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + 2\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})\widehat{u}_i + \sum_{i=1}^{n}\widehat{u}_i^2 \\ &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n}\widehat{u}_i^2\end{aligned}$$

となる。まとめると，

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

を得る。さらに，

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

それぞれの項は，

1. $\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \Longrightarrow y$ の全変動

2. $\sum_{i=1}^{n}(\widehat{Y_i} - \overline{Y})^2 \Longrightarrow \widehat{Y_i}$ (回帰直線) で説明される部分

3. $\sum_{i=1}^{n}\widehat{u_i^2} \Longrightarrow \widehat{Y_i}$ (回帰直線) で説明されない部分

となる。

回帰式の当てはまりの良さを示す指標として，決定係数 R^2 を以下の通りに定義する。

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

または,

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

として書き換えられる。

または，$Y_i = \widehat{Y}_i + \widehat{u}_i$ と

$$\begin{aligned}
\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y} - \widehat{u}_i) \\
&= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y}) - \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})\widehat{u}_i \\
&= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y})
\end{aligned}$$

を用いて，

$$\begin{aligned}
R^2 &= \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \\
&= \frac{\left(\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2\right)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2} \\
&= \left(\frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}}\right)^2
\end{aligned}$$

と書き換えられる。　　すなわち，R^2 は Y_i と $\widehat{Y}_i$ の相関係数の二乗と解釈される。

$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n}\widehat{u}_i^2$ から，明らかに，

$$0 \leq R^2 \leq 1,$$

となる。R^2 が 1 に近づけば回帰式の当てはまりは良いと言える。しかし，t 分布のような数表は存在しない。したがって,「どの値よりも大きくなるべき」というような基準はない。

慣習的には，メドとして 0.9 以上を判断基準にする。

数値例：　決定係数の計算には以下の公式を用いる。

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

計算に必要なものは，$\widehat{u}_i = Y_i - (\widehat{\alpha} + \widehat{\beta} X_i)$，$\overline{Y}$，$\sum_{i=1}^{n} Y_i^2$ である。

i	Y_i	X_i	$\widehat{Y}_i$	$\widehat{u}_i$	$\widehat{u}_i$	Y_i^2
1	6	10	6.8	−0.8	0.64	36
2	9	12	8.1	0.9	0.81	81
3	10	14	9.4	0.6	0.36	100
4	10	16	10.7	−0.7	0.49	100
合計	$\sum Y_i$	$\sum X_i$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$	$\sum \widehat{u}_i^2$	$\sum Y_i^2$
	35	52	35.0	0.0	2.30	317

$\sum \widehat{u}_i^2 = 2.30$，$\overline{X} = 13$，$\overline{Y} = 8.75$，$\sum_{i=1}^{n} Y_i^2 = 317$ なので，

$$R^2 = 1 - \frac{2.30}{317 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

1.5　まとめ

$\widehat{\alpha}$, $\widehat{\beta}$ を求めるための公式は

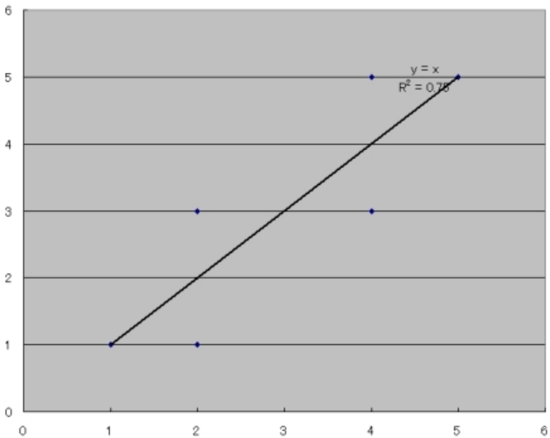
$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

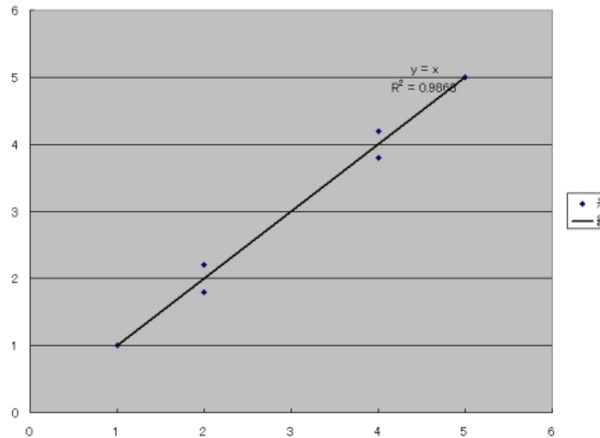
$$\widehat{\alpha} = \overline{Y} - \widehat{\beta X}$$

なので，必要なものは $\overline{X}$, $\overline{Y}$, $\displaystyle\sum_{i=1}^{n} X_i^2$, $\displaystyle\sum_{i=1}^{n} X_i Y_i$ である。
決定係数の計算には以下の公式を用いる。

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

計算に必要なものは，$\sum \widehat{u}_i^2$, $\overline{Y}$, $\displaystyle\sum_{i=1}^{n} Y_i^2$ である。


y = x
R^2 = 0.75
0
1
2
3
4
5
6


y = x
R^2 = 0.9868
0
1
2
3
4
5
6

2 Regression Analysis (回帰分析)

2.1 Setup of the Model

When (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available, suppose that there is a linear relationship between y and x, i.e.,

$$y_i = \beta_1 + \beta_2 x_i + u_i, \tag{4}$$

for $i = 1, 2, \cdots, n$. x_i and y_i denote the ith observations.

$\longrightarrow$ **Single (or simple) regression model (単回帰モデル)**

y_i is called the **dependent variable (従属変数)** or the **explained variable (被説明変数)**, while x_i is known as the **independent variable (独立変数)** or the **explanatory (or explaining) variable (説明変数)**.

β_1 = **Intercept (切片)**, β_2 = **Slope (傾き)**

β_1 and β_2 are unknown **parameters (パラメータ，母数)** to be estimated.

β_1 and β_2 are called the **regression coefficients (回帰係数)**.

u_i is the unobserved **error term (誤差項)** assumed to be a random variable with mean zero and variance σ^2.
σ^2 is also a parameter to be estimated.

x_i is assumed to be **nonstochastic (非確率的)**, but y_i is **stochastic (確率的)** because y_i depends on the error u_i.

The error terms $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed, which is called ***iid***.

It is assumed that u_i has a distribution with mean zero, i.e., $\mathrm{E}(u_i) = 0$ is assumed.

Taking the expectation on both sides of (4), the expectation of y_i is represented as:

$$\begin{aligned} \mathrm{E}(y_i) &= \mathrm{E}(\beta_1 + \beta_2 x_i + u_i) = \beta_1 + \beta_2 x_i + \mathrm{E}(u_i) \\ &= \beta_1 + \beta_2 x_i, \end{aligned} \tag{5}$$

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

Using $\mathrm{E}(y_i)$ we can rewrite (4) as $y_i = \mathrm{E}(y_i) + u_i$.

(5) represents the true regression line.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be estimates of β_1 and β_2.

Replacing β_1 and β_2 by $\hat{\beta}_1$ and $\hat{\beta}_2$, (4) turns out to be:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + e_i, \tag{6}$$

for $i = 1, 2, \cdots, n$, where e_i is called the **residual** (**残差**).

The residual e_i is taken as the experimental value (or realization) of u_i.

We define $\hat{y}_i$ as follows:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i, \tag{7}$$

for $i = 1, 2, \cdots, n$, which is interpreted as the **predicted value** (**予測値**) of y_i.

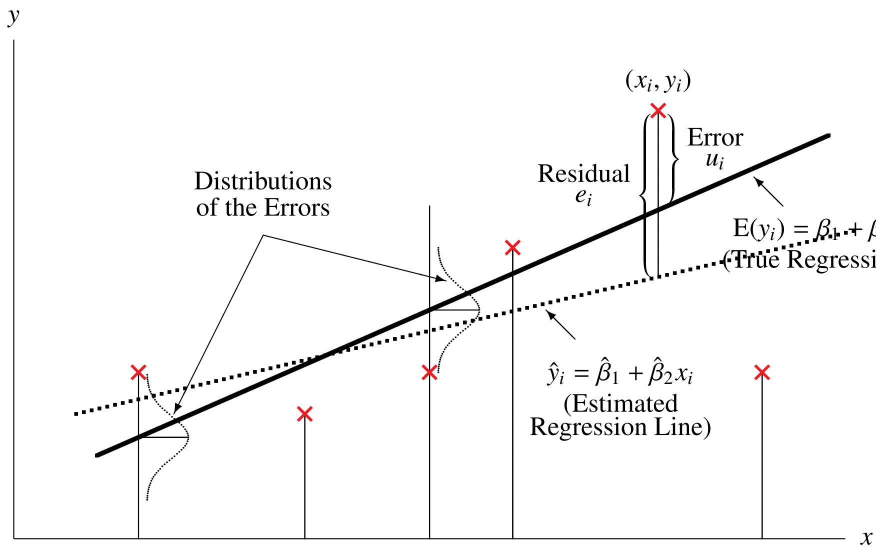
(7) indicates the estimated regression line, which is different from (5).

Moreover, using $\hat{y}_i$ we can rewrite (6) as $y_i = \hat{y}_i + e_i$.

(5) and (7) are displayed in Figure 1.

Consider the case of $n = 6$ for simplicity. $\times$ indicates the observed data series.

Figure 1. True and Estimated Regression Lines (回帰直線)



The true regression line (5) is represented by the solid line, while the estimated regression line (7) is drawn with the dotted line.

Based on the observed data, β_1 and β_2 are estimated as: $\hat{\beta}_1$ and $\hat{\beta}_2$.

In the next section, we consider how to obtain the estimates of β_1 and β_2, i.e., $\hat{\beta}_1$ and $\hat{\beta}_2$.

2.2 Ordinary Least Squares Estimation

Suppose that (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available.
For the regression model (4), we consider estimating β_1 and β_2.
Replacing β_1 and β_2 by their estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, remember that the residual e_i is given by:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i.$$

The sum of squared residuals is defined as follows:

$$S(\hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2.$$

It might be plausible to choose the $\hat{\beta}_1$ and $\hat{\beta}_2$ which minimize the sum of squared residuals, i.e., $S(\hat{\beta}_1, \hat{\beta}_2)$.

This method is called the **ordinary least squares estimation (最小二乗法，OLS)**.

To minimize $S(\hat{\beta}_1, \hat{\beta}_2)$ with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$, we set the partial derivatives equal to zero:

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0,$$

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0.$$

The second order condition for minimization is:

$$\begin{pmatrix} \frac{\partial^2 S(\hat\beta_1,\hat\beta_2)}{\partial\hat\beta_1^2} & \frac{\partial^2 S(\hat\beta_1,\hat\beta_2)}{\partial\hat\beta_1\partial\hat\beta_2} \\ \frac{\partial^2 S(\hat\beta_1,\hat\beta_2)}{\partial\hat\beta_2\partial\hat\beta_1} & \frac{\partial^2 S(\hat\beta_1,\hat\beta_2)}{\partial\hat\beta_2^2} \end{pmatrix} = \begin{pmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{pmatrix}$$

should be a positive definite matrix.

The diagonal elements $2n$ and $2\sum_{i=1}^n x_i^2$ are positive.

The determinant:

$$\begin{vmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{vmatrix} = 4n\sum_{i=1}^n x_i^2 - 4(\sum_{i=1}^n x_i)^2 = 4n\sum_{i=1}^n (x_i - \overline{x})^2$$

is positive. $\Longrightarrow$ The second-order condition is satisfied.

The first two equations yield the following two equations:

$$\overline{y} = \hat\beta_1 + \hat\beta_2\overline{x}, \tag{8}$$

$$\sum_{i=1}^n x_i y_i = n\overline{x}\hat\beta_1 + \hat\beta_2\sum_{i=1}^n x_i^2, \tag{9}$$

where $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Multiplying (8) by $n\overline{x}$ and subtracting (9), we can derive $\hat{\beta}_2$ as follows:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{10}$$

From (8), $\hat{\beta}_1$ is directly obtained as follows:

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2\overline{x}. \tag{11}$$

When the observed values are taken for y_i and x_i for $i = 1, 2, \cdots, n$, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimate**s (or simply the **least squares estimate**s, 最小二乗推定値) of β_1 and β_2.

When y_i for $i = 1, 2, \cdots, n$ are regarded as the random sample, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimator**s (or the **least squares estimator**s, 最小二乗推定量) of β_1 and β_2.

2.3 Properties of Least Squares Estimator

Equation (10) is rewritten as:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (x_i - \overline{x}) y_i}{\sum_{i=1}^n (x_i - \overline{x})^2} - \frac{\overline{y} \sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$
$$= \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} y_i = \sum_{i=1}^n \omega_i y_i. \qquad (12)$$

In the third equality, $\sum_{i=1}^n (x_i - \overline{x}) = 0$ is utilized because of $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

In the fourth equality, ω_i is defined as: $\omega_i = \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}$.

ω_i is nonstochastic because x_i is assumed to be nonstochastic.

ω_i has the following properties:

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0, \qquad (13)$$

$$\sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i (x_i - \overline{x}) = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = 1, \tag{14}$$

$$\sum_{i=1}^{n} \omega_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \right)^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{\left(\sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^2} = \frac{1}{\sum_{i=1}^{n} (x_i - \overline{x})^2}. \tag{15}$$

The first equality of (14) comes from (13).

From now on, we focus only on $\hat{\beta}_2$, because usually β_2 is more important than β_1 in the regression model (4).

In order to obtain the properties of the least squares estimator $\hat{\beta}_2$, we rewrite (12) as:

$$\begin{aligned} \hat{\beta}_2 &= \sum_{i=1}^{n} \omega_i y_i = \sum_{i=1}^{n} \omega_i (\beta_1 + \beta_2 x_i + u_i) \\ &= \beta_1 \sum_{i=1}^{n} \omega_i + \beta_2 \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n} \omega_i u_i = \beta_2 + \sum_{i=1}^{n} \omega_i u_i. \end{aligned} \tag{16}$$

In the fourth equality of (16), (13) and (14) are utilized.

[Review] Random Variables:

Let X_1, X_2, $\cdots$, X_n be n random variavles, which are mutually independently and identically distributed.

mutually independent $\Longrightarrow$ $f(x_i, x_j) = f_i(x_i)f_j(x_j)$ for $i \neq j$.

$f(x_i, x_j)$ denotes a joint distribution of X_i and X_j.

$f_i(x)$ indicates a marginal distribution of X_i.

identical $\Longrightarrow$ $f_i(x) = f_j(x)$ for $i \neq j$.

[End of Review]

[Review] Mean and Variance:

Let X and Y be random variables (continuous type), which are independently distributed.

Definition and Formulas:

- $\mathrm{E}(g(X)) = \displaystyle\int g(x)f(x)\mathrm{d}x$ for a function $g(\cdot)$ and a density function $f(\cdot)$.
- $\mathrm{V}(X) = \mathrm{E}((X-\mu)^2) = \displaystyle\int (x-\mu)^2 f(x)\mathrm{d}x$ for $\mu = \mathrm{E}(X)$.
- $\mathrm{E}(aX+b) = a\mathrm{E}(X)+b$ and $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$.
- $\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$ and $\mathrm{V}(X \pm Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

[End of Review]

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.

⟶ **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,

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

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

Rewriting,

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

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

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

[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