計量経済学

(99 年度作成)

1 計量経済学について

- 経済理論 (ミクロ，マクロ，財政，金融，国際経済, ⋯)
- データ (GNP，消費，投資，金利，為替レート，⋯)

計量経済学 $\Longrightarrow$ 経済理論が現実に成り立つものかどうかを，データを用いて，統計的に検証する。

1.1 例 1： マクロの消費関数

$$C = f(Y)$$

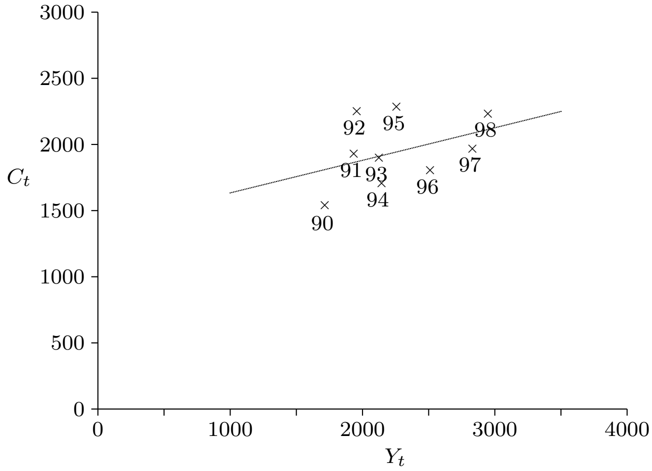
ただし，C は消費，Y は所得。

1. $Y \nearrow \Longrightarrow C \nearrow$
2. $\dfrac{\mathrm{d}C}{\mathrm{d}Y}$ = 限界消費性向 = 所得 1 円増加で消費が何円増加するか
3. すなわち，$\dfrac{\mathrm{d}C}{\mathrm{d}Y} > 0$

モデルの定式化

1. $C = a + bY$
2. $b = \dfrac{\mathrm{d}C}{\mathrm{d}Y}$ = 限界消費性向
3. a = 基礎消費 ($Y = 0$ のときに必要な消費)
4. 符号条件： $a > 0$，$b > 0$ (しかも，$1 > b$)

図 1： 消費 (C_t) と所得 (Y_t)



1. $\times \longrightarrow$ 実際のデータ
2. $(Y_t, C_t) \Longrightarrow t$ 期のデータ, i.e., $t = 1, 2, \cdots, 9$
3. $t = 1 \Longrightarrow 1990$ 年,
   $t = 2 \Longrightarrow 1991$ 年,
   $\cdots$,
   $t = 9 \Longrightarrow 1998$ 年,

1. 実際のデータを用いて，a, b を求める。
2. a, b を求める $\equiv$ 現実の経済構造を求める
3. その結果，もし $a > 0$，$1 > b > 0$ なら，経済理論は現実経済を説明していると言える。

1.2 例 2： 日本酒の需要関数

$$Q = f(Y, P_1, P_2)$$

ただし，Q は日本酒の需要量，Y は所得，P_1 は日本酒の価格，P_2 は洋酒の価格。

1. $Y \nearrow \Longrightarrow Q \nearrow$,
   $P_1 \nearrow \Longrightarrow Q \searrow$,
   $P_2 \nearrow \Longrightarrow Q \nearrow$
2. $\dfrac{\partial Q}{\partial Y} > 0$, $\dfrac{\partial Q}{\partial P_1} < 0$, $\dfrac{\partial Q}{\partial P_2} > 0$
3. 日本酒と洋酒は代替財
4. モデルの定式化 (A)

$$Q = a + b_1 Y + b_2 P_1 + b_3 P_2$$

5. Q, Y, P_1, P_2 を用いて，a, b_1, b_2, b_3 を求める (日本酒の需要構造を求める)。
6. 符号条件： $b_1 > 0$, $b_2 < 0$, $b_3 > 0$, a ?
7. t 期のデータ $(Q_t, Y_t, P_{1t}, P_{2t})$
8. T 組のデータ, i.e., $t = 1, 2, \cdots, T$
9. モデルの定式化 (B)
$$Q = a + b_1 Y + b_2 \frac{P_1}{P_2}$$
符号条件： $b_1 > 0$, $b_2 < 0$
10. モデルの定式化 (C)
$$\log(Q) = a + b_1 \log(Y) + b_2 \log(\frac{P_1}{P_2})$$
符号条件： $b_1 > 0$, $b_2 < 0$
11. モデル (A), (B), (C) のどれが最も現実的かを得られた結果から判断する。

例（効用関数の特定化 — より厳密に）： 効用関数をコブ・ダグラス型を仮定すると，効用関数と予算制約式は次のように書かれる。

$$\begin{aligned} &\text{効用関数：} && u(Q_1, Q_2) = Q_1^{\alpha} Q_2^{\beta} \\ &\text{予算制約式：} && Y = P_1 Q_1 + P_2 Q_2 \end{aligned}$$

効用関数は，通常，それぞれの財について，(i) 需要量が増えれば効用が増える（一階微分は正），(ii) 限界効用は逓減する（二階微分は負）という仮定を置くことが多い。すなわち，

$$\begin{aligned} \frac{\partial u(Q_1, Q_2)}{\partial Q_1} &= \alpha Q_1^{\alpha-1} Q_2^{\beta} > 0 \quad \longrightarrow \quad \text{条件 (i)} \\ \frac{\partial^2 u(Q_1, Q_2)}{\partial Q_1^2} &= \alpha(\alpha-1) Q_1^{\alpha-2} Q_2^{\beta} < 0 \quad \longrightarrow \quad \text{条件 (ii)} \\ \frac{\partial u(Q_1, Q_2)}{\partial Q_2} &= \beta Q_1^{\alpha} Q_2^{\beta-1} > 0 \quad \longrightarrow \quad \text{条件 (i)} \\ \frac{\partial^2 u(Q_1, Q_2)}{\partial Q_2^2} &= \beta(\beta-1) Q_1^{\alpha} Q_2^{\beta-2} < 0 \quad \longrightarrow \quad \text{条件 (ii)} \end{aligned}$$

が導かれ，$0 < \alpha < 1$，$0 < \beta < 1$ が得られる。また，条件 (ii) について，需要量 Q_1，Q_2 をともに t 倍 $(t > 1)$ すれば，効用は t 倍より小さくなる。すなわち，

$$u(tQ_1, tQ_2) < tu(Q_1, Q_2) \quad \longrightarrow \quad \text{条件 (ii)}$$

から $\alpha + \beta < 1$ が得られる。ただし，需要量は正を仮定している（すなわち，$Q_1 > 0$，$Q_2 > 0$）。
予算制約式のもとで効用最大化問題を解く。

$$L = Q_1^{\alpha} Q_2^{\beta} + \lambda(Y - P_1 Q_1 - P_2 Q_2)$$

のラグランジェ関数を設定して，Q_1，Q_2，λ についてそれぞれ微分してゼロと置く（ラグランジェ未定乗数法と呼ばれる）。λ はラグランジェ乗数と呼ばれる。Q_1，Q_2，λ について，

$$\begin{aligned} \frac{\partial L}{\partial Q_1} &= \alpha Q_1^{\alpha-1} Q_2^{\beta} - \lambda P_1 = 0 \\ \frac{\partial L}{\partial Q_2} &= \beta Q_1^{\alpha} Q_2^{\beta-1} - \lambda P_2 = 0 \\ \frac{\partial L}{\partial \lambda} &= Y - P_1 Q_1 - P_2 Q_2 = 0 \end{aligned}$$

最初の 2 つの式から λ を消去して，$\beta Q_1 P_1 = \alpha Q_2 P_2$ が得られ，最後の 3 つ目の式を用いて，需要関数 Q_1，Q_2 は

$$\begin{aligned} Q_1 &= \frac{\alpha}{\alpha+\beta} Y P_1^{-1} \\ Q_2 &= \frac{\beta}{\alpha+\beta} Y P_2^{-1} \end{aligned}$$

として得られる。需要関数の推定の際には，対数を取って，

$$\begin{aligned} \log Q_1 &= a_1 + a2 \log Y + a_3 \log P_1 \\ \log Q_2 &= b_2 + b2 \log Y + b_3 \log P_2 \end{aligned}$$

として，$a_1, a_2, a_3, b_1, b_2, b_3$ を推定することになる。ただし，$a_1 = \log \dfrac{\alpha}{\alpha+\beta} < 0$，$b_1 = \log \dfrac{\beta}{\alpha+\beta} < 0$，$a_2 = b_2 = 1$，$a_3 = b_3 = -1$ となる。したがって，推定した結果，$a_1 < 0$，$b_1 < 0$，$a_2 = b_2 = 1$，$a_3 = b_3 = -1$ が成り立つかどうかを統計的に検定することになる。
また，二階の条件を求めるために，下記のように二階微分を求める。

$$\begin{pmatrix} \frac{\partial L^2}{\partial Q_1^2} & \frac{\partial L^2}{\partial Q_1 \partial Q_2} \\ \frac{\partial L^2}{\partial Q_2 \partial Q_1} & \frac{\partial L^2}{\partial Q_2^2} \end{pmatrix} = \begin{pmatrix} \alpha(\alpha-1) Q_1^{\alpha-2} Q_2^{\beta} & \alpha\beta Q_1^{\alpha-1} Q_2^{\beta-1} \\ \beta\alpha Q_1^{\alpha-1} Q_2^{\beta-1} & \beta(\beta-1) Q_1^{\alpha} Q_2^{\beta-2} \end{pmatrix}$$

上記の最適解 Q_1，Q_2 のもとで，この行列が負値定符号行列であれば，効用関数が最大になっている。負値定符号行列のためには，

$$\begin{aligned} &\alpha(\alpha-1) Q_1^{\alpha-2} Q_2^{\beta} < 0 \\ &\beta(\beta-1) Q_1^{\alpha} Q_2^{\beta-2} < 0 \end{aligned}$$

$$\begin{vmatrix} \alpha(\alpha-1)Q_1^{\alpha-2}Q_2^{\beta} & \alpha\beta Q_1^{\alpha-1}Q_2^{\beta-1} \\ \beta\alpha Q_1^{\alpha-1}Q_2^{\beta-1} & \beta(\beta-1)Q_1^{\alpha}Q_2^{\beta-2} \end{vmatrix}$$
$$= \alpha\beta(1-\alpha-\beta)Q_1^{2\alpha-2}Q_2^{2\beta-2} > 0$$

という条件となる。したがって，α，β は $0<\alpha<1$，$0<\beta<1$，$\alpha+\beta<1$ でなければならない。

2 準備： 和記号 $\sum$ について

データとして，$X_1, X_2, \cdots, X_n$ と $Y_1, Y_2, \cdots, Y_n$ のを考える。

1. $\displaystyle\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$
2. $\displaystyle\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2$
3. $\displaystyle\sum_{i=1}^{n} X_iY_i = X_1Y_1 + X_2Y_2 + \cdots + X_nY_n$
4. 定数 c について，
   $\displaystyle\sum_{i=1}^{n} c = c + c + \cdots + c = nc$
5. $\displaystyle\sum_{i=1}^{n} cX_i = c\sum_{i=1}^{n} X_i$
6. $\displaystyle\sum_{i=1}^{n}(X_i + Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$
7. $\displaystyle\sum_{i=1}^{n}(X_i + Y_i)^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{i=1}^{n} X_iY_i + \sum_{i=1}^{n} Y_i^2$
8. $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} X_iY_j = \sum_{i=1}^{n} X_i \sum_{j=1}^{m} Y_j$

$(X_1, X_2, \cdots, X_n) = (1, 3, -3, 2)$,
$(Y_1, Y_2, \cdots, Y_n) = (2, 1, 2, -1)$ の例をとる。

1. $\displaystyle\sum_{i=1}^{n} X_i = 1 + 3 - 3 + 2 = 3$
2. $\displaystyle\sum_{i=1}^{n} X_i^2 = 1^2 + 3^2 + (-3)^2 + 2^2 = 23$
3. $\displaystyle\sum_{i=1}^{n} X_iY_i = 1\times 2 + 3\times 1 + (-3)2 + 2(-1) = -3$
4. $\displaystyle\sum_{i=1}^{n}(X_i+Y_i) = (1+2)+(3+1)+(-3+2)+(2-1) = 7$
5. $\displaystyle\sum_{i=1}^{n}(X_i+Y_i)^2 = (1+2)^2+(3+1)^2+(-3+2)^2+(2-1)^2 = 27$

2.1 重要な公式・性質

1. 標本平均 $\displaystyle\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. 標本分散 $\displaystyle s^2 = s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
3. 標本共分散 $\displaystyle s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$
4. $\displaystyle\sum_{i=1}^{n} X_i = n\overline{X}$
5. $\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}) = 0$
6. $\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$
7. $\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$
   $\displaystyle= \sum_{i=1}^{n} X_iY_i - n\overline{XY}$
   $\displaystyle= \sum_{i=1}^{n}(X_i - \overline{X})Y_i$
   $\displaystyle= \sum_{i=1}^{n}(Y_i - \overline{Y})X_i$

3 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

3.1 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)。

3.2 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，

$$\begin{aligned} \sum_{i=1}^{n} Y_i &= n\widehat{\alpha} + \widehat{\beta} \sum_{i=1}^{n} X_i, \\ \sum_{i=1}^{n} X_i Y_i &= \widehat{\alpha} \sum_{i=1}^{n} X_i + \widehat{\beta} \sum_{i=1}^{n} X_i^2, \end{aligned} \tag{3}$$

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，

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

さらに，$\widehat{\beta}$ について解くと，

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

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

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

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

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

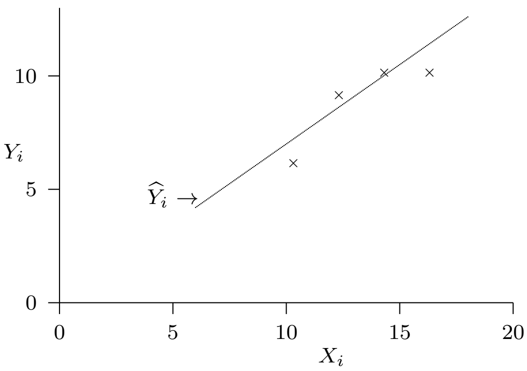
として与えられる。
上の数値例では，

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

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，

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

3.3　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

3.4　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，

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

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。
または，$\widehat{Y}_i=\widehat{\alpha}+\widehat{\beta}X_i$ と $\overline{Y}=\widehat{\alpha}+\widehat{\beta}\overline{X}$ を用いて，

$$\sum_{i=1}^n(\widehat{Y}_i-\overline{Y})^2=\widehat{\beta}^2\sum_{i=1}^n(X_i-\overline{X})^2$$

$$= \widehat{\beta} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

を利用すると，

$$R^2 = \frac{\widehat{\beta}^2 \sum_{i=1}^{n} (X_i - \overline{X})^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$
$$= \frac{\widehat{\beta} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$
$$= \left(\frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2 \sum_{i=1}^{n} (X_i - \overline{X})^2}} \right)^2$$

としても書き換えられる。

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

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

計算に必要なものは，$\widehat{\beta}$，$\overline{X}$，$\overline{Y}$，$\sum_{i=1}^{n} X_i^2$，$\sum_{i=1}^{n} Y_i^2$ である。

i	Y_i	X_i	Y_i^2	X_i^2
1	6	10	36	100
2	9	12	81	144
3	10	14	100	196
4	10	16	100	256
合計	$\sum Y_i$	$\sum X_i$	$\sum Y_i^2$	$\sum X_i^2$
	35	52	317	696
平均	$\overline{Y}$	$\overline{X}$		
	8.75	13		

$\widehat{\beta} = 0.65$，$\overline{X} = 13$，$\overline{Y} = 8.75$，$\sum_{i=1}^{n} X_i^2 = 696$，$\sum_{i=1}^{n} Y_i^2 = 317$ なので，

$$R^2 = \frac{0.65^2 (696 - 4 \times 13^2)}{317 - 4 \times 8.75^2} = \frac{8.45}{10.75} = 0.786$$

3.5 まとめ

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

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

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

決定係数の計算には以下の公式を用いる。

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

計算に必要なものは，$\widehat{\beta}$，$\overline{X}$，$\overline{Y}$，$\sum_{i=1}^{n} X_i^2$，$\sum_{i=1}^{n} Y_i^2$ である。