を得る。

(2) 式から，

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

を得る。

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

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

を得る。なぜなら，

$$\sum_{i=1}^{n} \hat{Y}_i \hat{u}_i = \sum_{i=1}^{n} (\hat{\alpha} + \hat{\beta} X_i) \hat{u}_i$$

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

である。

i	Y_i	X_i	$\hat{Y}_i$	$\hat{u}_i$	$X_i\hat{u}_i$	$\hat{Y}_i\hat{u}_i$
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	$\sum Y_i$	$\sum X_i$	$\sum \hat{Y}_i$	$\sum \hat{u}_i$	$\sum X_i\hat{u}_i$	$\sum \hat{Y}_i\hat{u}_i$
	35	**52**	**35.0**	**0.0**	**0.0**	**0.00**

9.4　決定係数 R^2 について

次の式

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

の両辺を二乗して，総和すると，

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

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

となる。まとめると，

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

を得る。さらに，

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

それぞれの項は，

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

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

3. $\sum_{i=1}^{n}\hat{u}_i^2 \Longrightarrow \hat{Y}_i$ **(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。**

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

または，

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

として書き換えられる。

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

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

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。

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

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

i	Y_i	X_i	$\hat{Y}_i$	$\hat{u}_i$	$\hat{u}_i^2$	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	$\sum Y_i$	$\sum X_i$	$\sum \hat{Y}_i$	$\sum \hat{u}_i$	$\sum \hat{u}_i^2$	$\sum Y_i^2$
	35	**52**	**35.0**	**0.0**	**2.30**	**317**

なので，

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

9.5　まとめ

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

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

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

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。

決定係数の計算には以下の公式を用いる。

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

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

9.6　最小 2 乗推定量の分布と性質

$$Y_i = \hat{\alpha} + \hat{\beta}X_i + \hat{u}_i$$

$\hat{\alpha}$**，**$\hat{\beta}$ **は推定量または推定値，** $\hat{u}_i$ **は残差**

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

α**，**β**，**σ^2 **は母数（パラメータ），** u_i **は誤差項**

仮定：

- $u_1, u_2, \cdots, n_n$ は互いに独立
- $u_i \sim N(0, \sigma^2)$

$\hat{\alpha}$ の分布と $\hat{\beta}$ の分布は，それぞれ，

$$\hat{\alpha} \sim N\Big(\alpha,\ \sigma^2(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2})\Big)$$
$$\hat{\beta} \sim N\Big(\beta,\ \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\Big)$$

であることがわかっている（証明略）。

$\mathbf{E}(\hat{\alpha}) = \alpha$，$\mathbf{E}(\hat{\beta}) = \beta$ であるので，$\hat{\alpha}$ と $\hat{\beta}$ は α と β の不偏推定量である（証明略）。

標準化して，

$$\frac{\hat{\alpha}-\alpha}{\sigma\sqrt{\frac{1}{n}+\frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}} \sim \mathbf{N}(0,\ 1)$$

$$\frac{\hat{\beta}-\beta}{\sigma/\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} \sim \mathbf{N}(0,\ 1)$$

一方，σ^2 の推定量 s^2

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

σ^2 を推定するためには $\hat{\alpha}$，$\hat{\beta}$ の 2 つを推定する必要がある。

そのため，$n-1$ でなく，$n-2$ で割る。

以下が分かっていて，

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

t 分布が得られる（証明略）。

結果的には，σ^2 を s^2 で置き換えると，

$$\frac{\hat{\alpha}-\alpha}{s\sqrt{\frac{1}{n}+\frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}} \sim t(n-2)$$

$$\frac{\hat{\beta}-\beta}{s/\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} \sim t(n-2)$$

となる。

$$Se(\hat{\alpha}) = s\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

$$Se(\hat{\beta}) = s/\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

として，$Se(\hat{\alpha})$，$Se(\hat{\beta})$ をそれぞれ α の標準誤差，β の標準誤差と呼ぶ。

よって，

$$\frac{\hat{\alpha} - \alpha}{Se(\hat{\alpha})} \sim t(n-2)$$

$$\frac{\hat{\beta} - \beta}{Se(\hat{\beta})} \sim t(n-2)$$

として，α や β に関する区間推定，仮説検定を行う。