

数値例： $\hat{\alpha} = 0.5$, $\hat{\beta} = 0.7$ なので , $\hat{Y}_i = 0.5 + 0.7X_i$, $\hat{u}_i = Y_i - 0.5 - 0.7X_i$ により , \hat{Y}_i , \hat{u}_i を計算する。

i	X_i	Y_i	X_i^2	X_iY_i	\hat{Y}_i	\hat{u}_i
1	5	4	25	20	4.0	0.0
2	1	1	1	1	1.2	-0.2
3	3	1	9	3	2.6	-1.6
4	2	3	4	6	1.9	1.1
5	4	4	16	16	3.3	0.7
合計	$\sum X_i$	$\sum Y_i$	$\sum X_i^2$	$\sum X_iY_i$	$\sum \hat{Y}_i$	$\sum \hat{u}_i$
	15	13	55	46	13	0.0
平均	\bar{X}	\bar{Y}				
	3	2.6				

誤差項 (または, 攪乱項) の母分散 σ^2 の不偏推定量 s^2 は,

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{5-2} (0.0^2 + (-0.2)^2 + (-1.6)^2 + 1.1^2 + 0.7^2) = 1.43333$$

によって与えられる。

s は「回帰式の標準誤差 (Standard Error of Regression)」と呼ばれ, この例では, $s = \sqrt{1.43333} = 1.197$ となる。

$\hat{\alpha}$, $\hat{\beta}$ の分散の不偏推定量: $\hat{\alpha}$ と $\hat{\beta}$ の分散は,

$$\mathbf{V}(\hat{\alpha}) = \sigma_{\hat{\alpha}}^2 = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} \quad \mathbf{V}(\hat{\beta}) = \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

によって, 与えられる。

σ^2 をその不偏分散 s^2 に置き換えることによって, $\hat{\alpha}, \hat{\beta}$ の分散の不偏推定量を次のように得ることができる。

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

さらに, 平方根をとって, $\hat{\alpha}, \hat{\beta}$ の標準誤差はそれぞれ,

$$s_{\hat{\alpha}} = s \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}} \quad s_{\hat{\beta}} = \frac{s}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

として与えられる。

数値例： $s^2 = 1.43333$ なので，

$$s_{\hat{\beta}}^2 = \frac{s^2}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{s^2}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = 1.433333 \times 0.1 = 0.1433333$$

$$s_{\hat{\alpha}}^2 = \frac{s^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{s^2 \sum_{i=1}^n X_i^2}{n(\sum_{i=1}^n X_i^2 - n\bar{X}^2)} = 1.433333 \times 1.1 = 1.5766667$$

$\hat{\alpha}$ ， $\hat{\beta}$ の標準誤差はそれぞれ，平方根をとって，

$$s_{\hat{\beta}} = \sqrt{0.1433333} = 0.3786, \quad s_{\hat{\alpha}} = \sqrt{1.5766667} = 1.25565$$

となる。

5.3.8 $\hat{\alpha}$, $\hat{\beta}$ の分布 (σ^2 が未知の場合)

(* 復習) t 分布について :

$Z \sim N(0, 1)$, $U \sim \chi^2(k)$, Z と U は独立のとき,

$$T = \frac{Z}{\sqrt{U/k}} \sim t(k) \text{ となる。}$$

(* 復習) t 分布について (その 2):

X_1, X_2, \dots, X_n は互いに独立で, $X_i \sim N(\mu, \sigma^2)$

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$ なので, $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ となる。

一方, $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ となる。ただし, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

\bar{X} と S^2 は独立 ($\text{Cov}(\bar{X}, S^2) = 0$ を示せばよいが, ここでは証明略)。

$$\frac{Z}{\sqrt{U/(n-1)}} = \frac{\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \text{ となる。}$$

$\hat{\beta}$ について: $\hat{\beta}$ は次のように書き換えられる。

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^n (X_i - \bar{X})u_i}{\sum_{j=1}^n (X_j - \bar{X})^2} = \beta + \sum_{i=1}^n \omega_i u_i$$

$u_i \sim N(0, \sigma^2)$ で, かつ, それぞれ独立に分布する。また, $\hat{\beta}$ の平均, 分散はそれぞれ,

$$\mathbf{E}(\hat{\beta}) = \beta \quad \mathbf{V}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

となるので，

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)$$

を得る。変形すると，

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

となる。

さらに，

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

となり（証明略）， $\hat{\beta}$ とは独立なので（ $\text{Cov}(\hat{\beta}, s^2) = 0$ を示せばよいが，ここでは証明略），

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

結果的には， σ を s で置き換えることによって，

$$\frac{\hat{\beta} - \beta}{\sigma / \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \sim N(0, 1) \quad \Rightarrow \quad \frac{\hat{\beta} - \beta}{s / \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \sim t(n-2)$$

が得られる。

$\hat{\alpha}$ について： また， $\hat{\alpha}$ の平均，分散はそれぞれ，

$$\mathbf{E}(\hat{\alpha}) = \alpha \quad \mathbf{V}(\hat{\alpha}) = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}$$

となるので，

$$\hat{\alpha} \sim N\left(\alpha, \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}\right)$$

を得る。変形すると，

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

となる。

さらに,

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

となり (証明略), $\hat{\alpha}$ とは独立なので ($\text{Cov}(\hat{\alpha}, s^2) = 0$ を示せばよいが, ここでは証明略),

$$\frac{Z}{\sqrt{U/(n-2)}} = \frac{\frac{\hat{\alpha} - \alpha}{\sigma \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}}}}{\sqrt{\frac{(n-2)s^2}{\sigma^2} / (n-2)}} = \frac{\hat{\alpha} - \alpha}{s \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}}} \sim t(n-2)$$

となる。

結果的に見ると， σ を s で置き換えると，

$$\frac{\hat{\alpha} - \alpha}{\sigma \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}}} \sim N(0, 1) \quad \Rightarrow \quad \frac{\hat{\alpha} - \alpha}{s \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}}} \sim t(n-2)$$

となる。

まとめ：

$$\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \sim t(n-2) \quad \frac{\hat{\alpha} - \alpha}{s_{\hat{\alpha}}} \sim t(n-2)$$

ただし，

$$s_{\hat{\beta}} = \frac{s}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \quad s_{\hat{\alpha}} = s \sqrt{\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}}$$

とする。 $s_{\hat{\beta}}$ は $\hat{\beta}$ の標準誤差 (Standard Error)， $s_{\hat{\alpha}}$ は $\hat{\alpha}$ の標準誤差とそれぞれ呼ばれる。

5.3.9 α, β の区間推定 (信頼区間)

(* 復習) 区間推定 (Interval Estimation) について :

X_1, X_2, \dots, X_n は互いに独立で, $X_i \sim N(\mu, \sigma^2)$ とする。

$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$ となる。

$t_{\alpha/2}(n-1), t_{1-\alpha/2}(n-1)$ をそれぞれ自由度 $n-1$ の t 分布の上側から $100 \times \frac{\alpha}{2}$ % 点,

$100 \times (1 - \frac{\alpha}{2})$ % 点の値とする ($\alpha = 0.05, 0.01$ がよく用いられ, t 分布表から得られる)。

確率変数 $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ が $t_{\alpha/2}(n-1)$ と $t_{1-\alpha/2}(n-1)$ の間に入る確率が $1 - \alpha$ という意味である。

t 分布はゼロを中心として左右対称なので, $t_{1-\alpha/2}(n-1) = -t_{\alpha/2}(n-1)$ となる。

$$\text{Prob}\left(-t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\right) = 1 - \alpha$$

$$\text{Prob}\left(\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

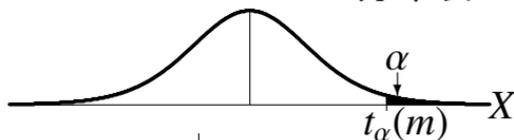
推定量 \bar{X} , S^2 をそれぞれ推定値 \bar{x} , s^2 で置き換えて,

信頼係数 $1 - \alpha$ の μ の信頼区間 (**Confidence Interval**) が,

$$\left(\bar{x} - t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} \right)$$

となる。ただし, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ とする。

t 分布表： $X \sim t(m)$



$$\alpha = \mathbf{Prob}(X > t_{\alpha}(m)) = \int_{t_{\alpha}(m)}^{\infty} f(x)dx$$

m (自由度)	α	.10	.05	.025	.010	.005
1		3.0777	6.3138	12.7062	31.8205	63.6567
2		1.8856	2.9200	4.3027	6.9646	9.9248
3		1.6377	2.3534	3.1824	4.5407	5.8409
4		1.5332	2.1318	2.7764	3.7470	4.6041
5		1.4759	2.0150	2.5706	3.3649	4.0322
6		1.4398	1.9432	2.4469	3.1427	3.7074
7		1.4149	1.8946	2.3646	2.9980	3.4995
8		1.3968	1.8595	2.3060	2.8965	3.3554
9		1.3830	1.8331	2.2622	2.8214	3.2498
10		1.3722	1.8125	2.2281	2.7638	3.1693
11		1.3634	1.7959	2.2010	2.7181	3.1058
12		1.3562	1.7823	2.1788	2.6810	3.0545
∞		1.2816	1.6449	1.9600	2.3263	2.5758

回帰分析に応用する。

$\hat{\alpha}$, $\hat{\beta}$ の分布は, 以下のように得られた。

$$\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \sim t(n-2) \quad \frac{\hat{\alpha} - \alpha}{s_{\hat{\alpha}}} \sim t(n-2)$$

$t_{\alpha/2}(n-2)$, $t_{1-\alpha/2}(n-2)$ をそれぞれ自由度 $n-2$ の t 分布の上側から $100 \times \frac{\alpha}{2}$ %点, $100 \times (1 - \frac{\alpha}{2})$ %点の値とする。ただし, この α と切片の α は異なるものなので注意すること。このとき,

$$\mathbf{Prob}\left(t_{1-\alpha/2}(n-2) < \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} < t_{\alpha/2}(n-2)\right) = 1 - \alpha,$$

すなわち, t 分布はゼロを中心として左右対称なので, $t_{1-\alpha/2}(n-2) = -t_{\alpha/2}(n-2)$ となり,

$$\mathbf{Prob}\left(-t_{\alpha/2}(n-2) < \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} < t_{\alpha/2}(n-2)\right) = 1 - \alpha,$$

を得る。ただし、自由度と α が決まれば、 $t_{\alpha/2}(n-2)$ は t 分布表から得られる。

書き直して、

$$\text{Prob}(\hat{\beta} - t_{\alpha/2}(n-2)s_{\hat{\beta}} < \beta < \hat{\beta} + t_{\alpha/2}(n-2)s_{\hat{\beta}}) = 1 - \alpha,$$

と表される。

したがって、 $\hat{\beta}$ 、 $s_{\hat{\beta}}$ を推定値で置き換えて、信頼係数 $1 - \alpha$ の β の信頼区間は、

$$(\hat{\beta} - t_{\alpha/2}(n-2)s_{\hat{\beta}}, \hat{\beta} + t_{\alpha/2}(n-2)s_{\hat{\beta}})$$

となる。

同様に，信頼係数 $1 - \alpha$ (この α は確率) の α (この α は切片) の信頼区間は，

$$\left(\hat{\alpha} - t_{\alpha/2}(n-2)s_{\hat{\alpha}}, \hat{\alpha} + t_{\alpha/2}(n-2)s_{\hat{\alpha}} \right)$$

となる。

数値例： 今までと同様に，以下の数値例をとりあげる。

i	X_i	Y_i
1	5	4
2	1	1
3	3	1
4	2	3
5	4	4

回帰モデル $Y_i = \alpha + \beta X_i + u_i$ を推定した結果，以下の推定値を得た。

$$\hat{\beta} = 0.7, \quad \hat{\alpha} = 0.5$$

$$s_{\hat{\beta}} = \sqrt{0.1433333} = 0.3786, \quad s_{\hat{\alpha}} = \sqrt{1.5766667} = 1.25565$$

$t_{0.025}(3) = 3.1824$ なので、信頼係数 **0.95** の傾き β の信頼区間は、

$$(0.7 - 3.1824 \times 0.3786, 0.7 + 3.1824 \times 0.3786) = (-0.505, 1.905)$$

となり、信頼係数 **0.95** の切片 α の信頼区間は、

$$(0.5 - 3.1824 \times 1.25565, 0.5 + 3.1824 \times 1.25565) = (-3.496, 4.496)$$

となる。

同様にして、 $t_{0.05}(3) = 2.3534$ なので、信頼係数 **0.90** の傾き β の信頼区間は、

$$(0.7 - 2.3534 \times 0.3786, 0.7 + 2.3534 \times 0.3786) = (-0.191, 1.591)$$

となり，信頼係数 **0.90** の切片 α の信頼区間は，

$$(0.5 - 2.3534 \times 1.25565, 0.5 + 2.3534 \times 1.25565) = (-2.455, 3.455)$$

となる。

5.3.10 α, β の仮説検定

「帰無仮説 $H_0 : \beta = \beta_0$, 対立仮説 $H_1 : \beta \neq \beta_0$ 」の検定を行う。帰無仮説 : $H_0 : \beta = \beta_0$ が正しいとき ,

$$\frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}} \sim t(n-2),$$

となる。

よって , 検定の手順は ,

1. 検定統計値 $\frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}}$ を計算する。ただし ,

$$s_{\hat{\beta}}^2 = \frac{s^2}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad s^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2, \quad \hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta}X_i,$$

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X},$$

である。

2. 有意水準 α を決めて (この α は, 回帰式の定数項の α とは異なることに注意), 自由度 $n-2$ の t 分布表から $100 \times \alpha$ % 点の値を求める。通常, $\alpha = 0.01, 0.05$ とす場合が多い。
3. 自由度 $n-2$ の t 分布表から得られた $100 \times \alpha$ % 点の値と検定統計値の大小を比較する。すなわち,

$$\left| \frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}} \right| > t_{\alpha/2}(n-2),$$

ならば, 有意水準 α で帰無仮説 $H_0: \beta = \beta_0$ を棄却する。帰無仮説 $H_0: \beta = \beta_0$ を中心とした分布を考えると, データから得られた検定統計量は, 分布の端にあり, 確率的に起こりにくいと考える。

となる。

定数項の推定量 $\hat{\alpha}$ についても同様。

t 値について： 特に「帰無仮説： $H_0 : \beta = 0$ ，対立仮説： $H_1 : \beta \neq 0$ 」の検定は特別の意味を持つ。

帰無仮説のもとでは（「帰無仮説が正しいとき」と同じ意味），

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

の分布に従う。

このときの t 統計量の値（すなわち， $\frac{\hat{\beta}}{s_{\hat{\beta}}}$ ）を t 値と呼ぶ。

$H_0 : \beta = 0$ という帰無仮説は，回帰式

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

において，「 X_i が Y_i に何の影響も与えない」という仮説であることを意味する。

有意水準 α のもとで (この α は，回帰式の定数項の α とは異なることに注意)，

$$\left| \frac{\hat{\beta}}{s_{\hat{\beta}}} \right| > t_{\alpha/2}(n-2)$$

ならば，有意水準 α で帰無仮説 $H_0 : \beta = 0$ を棄却する。

1. 回帰モデル

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

において，経済理論から $\beta > 0$ という符号条件が想定されたとする。

2. 有意水準 α を決める。(例えば， $\alpha = 0.05, 0.01$)

3. 実際のデータから， $\hat{\beta} > 0$ が得られた場合

(a) t 値が，

$$\frac{\hat{\beta}}{s_{\hat{\beta}}} > t_{\alpha/2}(n-2)$$

となった場合，帰無仮説 $H_0 : \beta = 0$ が棄却され， $\beta > 0$ が統計的にも証明され，
経済理論は現実経済をサポートする結果。

(b) t 値が，

$$\frac{\hat{\beta}}{s_{\hat{\beta}}} < t_{\alpha/2}(n-2)$$

となった場合，帰無仮説 $H_0 : \beta = 0$ を棄却できず， $\hat{\beta} > 0$ にもかかわらず， $\beta < 0$ の可能性もあるため，経済理論が現実経済を積極的には支持しない。

4. 実際のデータから， $\hat{\beta} < 0$ が得られた場合

(a) t 値が，

$$-t_{\alpha/2}(n-2) < \frac{\hat{\beta}}{s_{\hat{\beta}}}$$

となった場合，帰無仮説 $H_0 : \beta = 0$ を棄却できず， $\hat{\beta} < 0$ にもかかわらず， $\beta > 0$ の可能性もあるため，得られた推定結果と経済理論は完全に矛盾しているとは言えない。

(b) t 値が，

$$-t_{\alpha/2}(n-2) > \frac{\hat{\beta}}{s_{\hat{\beta}}}$$

となった場合，帰無仮説 $H_0 : \beta = 0$ は棄却され，統計的には $\beta < 0$ となり，得られた推定結果と経済理論は完全に矛盾している。すなわち，この場合，経済理論の立て直しが必要。

数値例： 今までと同様に，以下の数値例をとりあげる。

i	X_i	Y_i
1	5	4
2	1	1
3	3	1
4	2	3
5	4	4

回帰モデル $Y_i = \alpha + \beta X_i + u_i$ を推定した結果，以下の推定値を得た。

$$\hat{\beta} = 0.7, \quad s_{\hat{\beta}} = \sqrt{0.1433333} = 0.3786,$$

帰無仮説 $H_0 : \beta = 0$ ，対立仮説 $H_0 : \beta \neq 0$ の検定を行う。 t 値は $0.7/0.3786 = 1.849$ ，有意水

準 5 % の $t_{\alpha/2}(n-2)$ の値は **3.1824** となり ($\alpha = 0.05, n = 5$) ,

$$\frac{\hat{\beta}}{s_{\hat{\beta}}} = \frac{0.7}{\sqrt{0.1433333}} = 1.849 < t_{\alpha/2}(n-2) = 3.1824$$

を得る。このように , 有意水準 5 % で帰無仮説 $H_0 : \beta = 0$ は棄却されない。よって , β の符号は統計学的に確定できない。

また , α についても同様に , t 値を計算できる。

$$\hat{\alpha} = 0.5, \quad s_{\hat{\alpha}} = \sqrt{1.5766667} = 1.25565,$$

なので , t 値は ,

$$\frac{\hat{\alpha}}{s_{\hat{\alpha}}} = \frac{0.5}{\sqrt{1.5766667}} = 0.398 < t_{\alpha/2}(n-2) = 3.1824,$$

となり，有意水準 5 % で $H_0: \alpha = 0$ を棄却できない。

しかし，定数項については，経済学的意味が無い場合が多い。