第 7 回　重回帰モデル：推定②

[2] 最小二乗法による重回帰モデルの推定

（問題設定）重回帰モデルで表される母集団からランダム標本をとり，n 組の$(x_{1i},\cdots,x_{ki},y_i)$を観察したとする。このとき，次の関係が成り立つ。

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i \quad (i = 1,\cdots,n) \quad ・・・ ①$$

このとき，$\beta_0, \beta_1, \cdots, \beta_k$ の値を推定したい。

※ x_{ji} ：j 番目の説明変数x_jについての i 番目のデータ

β_j の推定量を $\hat{\beta}_j$ (j = 1, ..., k) として，フィットと残差を次のように定義する。

フィット： $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$

残　　差： $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki}$

⇒ 残差平方和： $\mathrm{RSS} = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki})^2$

最小二乗法：RSS を β_j (j = 1, ..., k) について最小化するように $\hat{\beta}_j$ を推定する方法

（例）説明変数が 2 個の場合，最小二乗推定量 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ は次のように求められる。

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_0} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = 0$$

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_1} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} x_{1i} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = 0$$

$$\frac{\partial \mathrm{RSS}}{\partial \hat{\beta}_2} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} x_{2i} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = 0$$

$$\Rightarrow \quad \hat{\beta}_1 = \frac{S_{22} S_{1y} - S_{12} S_{2y}}{S_{11} S_{22} - S_{12}^2}, \qquad \hat{\beta}_2 = \frac{S_{11} S_{2y} - S_{12} S_{1y}}{S_{11} S_{22} - S_{12}^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$S_{jk} = \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k), \quad S_{jy} = \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)(y_i - \bar{y}) \quad (j, k = 1, 2)$$

・重回帰モデルの係数の意味（説明変数が従属変数に及ぼす影響の測り方）

重回帰モデルでは他の説明変数の値を一定にすることにより，「条件一定化」ができる。誤差項 u の平均が 0 であり，説明変数が確率変数でなければ，

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \quad \Rightarrow \quad \Delta E(y) = \beta_1 \Delta x_1 + \cdots + \beta_k \Delta x_k$$

$\boxed{\beta_k \text{ の意味}}$ $\Delta x_j = 0\ (j \neq k)$ とするとき， x_k の 1 単位の増加が E(y)に及ぼす効果

つまり，<u>他の説明変数を一定として</u>， x_k の変化がyの平均に及ぼす効果

⇒ 同じ $x_j\ (j \neq k)$ の値をもつデータがなくても，実験をしたかのような効果が測れる。

（例）log(wage) = 0.284 + 0.092 educ + 0.0041 exper + 0.022 tenure

educの係数：<u>労働市場の参加年数exper，現在の職場での在職年数tenureが同じ人について</u>，教育が 1 年増えると賃金は平均的に 9.2%増える。

（<u>exper と tenure が同じ人のデータを集めなくても</u>，この効果が測れる。）

・フィットと残差の性質と重決定係数

1) $\sum_{i=1}^{n} \hat{u}_i = 0$ （残差の合計は 0）

2) $\sum_{i=1}^{n} x_{ji}\hat{u}_i = 0 \ \ (j = 1, \cdots, k)$ （残差と説明変数 x_j の標本共分散は 0）

3) フィットの式は標本平均の点 $(\bar{x}_1, \cdots, \bar{x}_k, \bar{y})$ を通る。

また， $TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$ ， $ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ ， $RSS = \sum_{i=1}^{n}\hat{u}_i^2$ とすると $TSS = ESS + RSS$ が成り立ち，重回帰モデルのあてはまりのよさを表す尺度が得られる。

重決定係数 $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$ （$0 \leq R^2 \leq 1$）

[3] 最小二乗推定量の性質

単回帰モデルのときと同じように，最小二乗推定量 $\hat{\beta}_j$ はやはり確率変数である。$\hat{\beta}_j$ の平均と分散を求めるため，次の 5 つの仮定をおく。

(M1) $E(u_i) = 0$

(M2) すべての説明変数は確率変数ではない。

(M3) 説明変数 $x_{1i}, \cdots, x_{ki}$ は定数を含まず，厳密な線形関係をもたない。

(M4) $\mathrm{Var}(u_i) = \sigma^2$

(M5) $i \neq j$ について u_i と u_j は確率的に独立である。

仮定(M3)については，次の点を理解しておけばよい。

・$x_{1i}, \cdots, x_{ki}$ が厳密な線形関係をもつ ⇒ $a_1 x_{1i} + \cdots + a_k x_{ki} = b$ とかける（a_j, b：定数）

・(M3)が成り立たない状況：**完全な多重共線性** ⇒ 最小二乗推定量は<u>計算できない</u>。

（例）ある説明変数が他の説明変数の定数倍になっているとき

$$\log(\mathrm{wage}) = \beta_0 + \beta_1 \log(\mathrm{educ}) + \beta_2 \log(\mathrm{educ}^2) + u$$

上記の仮定が成り立てば，最小二乗推定量の平均と分散について，次の定理が得られる。

定理 3.1（最小二乗推定量の不偏性）

仮定(M1)～(M3)が成り立てば， $E(\hat{\beta}_j) = \beta_j$ (j = 0, 1, …, k)

定理 3.2（最小二乗推定量の分散）

仮定(M1)～(M5)が成り立てば， $\mathrm{Var}(\hat{\beta}_j) = \dfrac{\sigma^2}{S_{jj}(1 - R_j^2)}$ （$j = 1, \cdots, k$）である。

$S_{jj} = \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)^2$ ， R_j^2 は x_j を他のすべての説明変数に回帰したときの重決定係数

定理 3.2 より，重回帰モデルでは，三つの要因が分散（推定精度）を決定している。

1）誤差分散 σ^2 ：誤差 u の変動が大きいほど，$\hat{\beta}_j$ の推定精度は下がる。

2）x_j の変動 S_{jj} ：説明変数 x_j の変動が大きいほど，$\hat{\beta}_j$ の推定精度は上がる。

3）説明変数間の関係の強さ R_j^2 ： R_j^2 が大きいほど，$\hat{\beta}_j$ の推定精度は下がる。

多重共線性：定数項，説明変数 $x_{1i},\cdots,x_{ki}$ の間に高い相関関係があり（R_j^2 が 1 に近く），最小二乗推定量の推定精度が低くなる問題。

※「完全な多重共線性」と区別すること。

※ R_j^2 が高いこと自体は問題ではない（$\hat{\beta}_j$ の推定精度が問題)。

$\mathrm{Var}(\hat{\beta}_j)$ を計算するには，次の方法で誤差分散を推定する必要がある。

> **定理 3.3（誤差分散 σ^2 の不偏推定量）**
>
> 仮定(M1)～(M5)の下で，$\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\hat{u}_i^2}{n-k-1}$ は σ^2 の不偏推定量である。

定理 3.2 と定理 3.3 より，$\hat{\beta}_j$ の推定精度を表す標準誤差を次のように計算できる。

$\hat{\beta}_j$ の標準誤差： $$\mathrm{se}(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{S_{jj}(1-R_j^2)}} \qquad (j = 1,\cdots,k)$$

重回帰モデルの推定でも最小二乗推定量が利用されるのは，次の定理によっている。

> **定理 3.4（ガウス＝マルコフの定理）**
>
> 仮定(M1)～(M5)の下で，β_j（$j = 1,\cdots,k$）の最小二乗推定量 $\hat{\beta}_j$ は最良線形不偏推定量（BLUE）である。