

## 6.3 多重共線性について

簡単化のため、定数項がなく、説明変数が2つの回帰式

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

の場合を考える。 $W_i$ 、 $X_i$  が説明変数、 $Y_i$  は被説明変数、 $u_i$  は互いに独立な攪乱項とする。

$W_i$  と  $X_i$  の相関が大きいことを多重共線性 (**multicollinearity**) が強いと言う。

結論から言えば、 $W_i$  と  $X_i$  の相関が大きい場合は、 $\alpha$ 、 $\beta$  の推定値は不安定になる。

極端な場合、 $W_i$  と  $X_i$  の相関が 1 の場合 (完全相関の場合) は、すべての  $i$  について、 $W_i = \gamma X_i$  となる。

この場合，回帰式は

$$\begin{aligned} Y_i &= \alpha W_i + \beta X_i + u_i \\ &= (\alpha\gamma + \beta)X_i + u_i \end{aligned}$$

となり， $\alpha\gamma + \beta$  を推定することは可能だが， $\alpha$ ， $\beta$  を別々に推定することはできなくなる。

言い換えると， $Y_i = \alpha W_i + \beta X_i + u_i$  を推定した場合， $\alpha\gamma + \beta$  の推定値が一定値となる  $\hat{\alpha}$ ， $\hat{\beta}$  の組み合わせは無数に存在する。

この意味で， $\hat{\alpha}$ ， $\hat{\beta}$  は不安定であると言える。

より詳しく見るために，定数項なしで説明変数 2 個の場合を考える。

$\hat{\alpha}$ ,  $\hat{\beta}$  を  $\alpha$ ,  $\beta$  の推定量とする。最小二乗推定量は,

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (Y_i - \hat{\alpha}W_i - \hat{\beta}X_i)^2$$

を最小にする  $\hat{\alpha}$ ,  $\hat{\beta}$  である。

すなわち,

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (Y_i - \hat{\alpha}W_i - \hat{\beta}X_i)W_i = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}} = -2 \sum_{i=1}^n (Y_i - \hat{\alpha}W_i - \hat{\beta}X_i)X_i = 0$$

の連立方程式を解くことになる。

$$\sum_{i=1}^n Y_i W_i - \hat{\alpha} \sum_{i=1}^n W_i^2 - \hat{\beta} \sum_{i=1}^n X_i W_i = 0$$

$$\sum_{i=1}^n Y_i X_i - \hat{\alpha} \sum_{i=1}^n W_i X_i - \hat{\beta} \sum_{i=1}^n X_i^2 = 0$$

行列表示により，

$$\begin{pmatrix} \sum Y_i W_i \\ \sum Y_i X_i \end{pmatrix} = \begin{pmatrix} \sum W_i^2 & \sum X_i W_i \\ \sum W_i X_i & \sum X_i^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$$

$\hat{\alpha}$  ,  $\hat{\beta}$  について表すと，

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \sum W_i^2 & \sum X_i W_i \\ \sum W_i X_i & \sum X_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum Y_i W_i \\ \sum Y_i X_i \end{pmatrix}$$

逆行列を計算して，

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \frac{1}{(\sum W_i^2)(\sum X_i^2) - (\sum W_i X_i)^2} \begin{pmatrix} \sum X_i^2 & -\sum X_i W_i \\ -\sum W_i X_i & \sum W_i^2 \end{pmatrix} \begin{pmatrix} \sum Y_i W_i \\ \sum Y_i X_i \end{pmatrix}$$

完全な多重共線性の場合 ( $W_i = \gamma X_i$  の場合)，

$$(\sum W_i^2)(\sum X_i^2) - (\sum W_i X_i)^2 = (\sum \gamma^2 X_i^2)(\sum X_i^2) - (\sum \gamma X_i^2)^2 = 0$$

となる。

また、 $\hat{\alpha}$ 、 $\hat{\beta}$  の分散共分散行列と前出の逆行列との関係から、

$$\begin{aligned} \mathbf{V} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} &= \begin{pmatrix} \mathbf{V}(\hat{\alpha}) & \mathbf{Cov}(\hat{\alpha}, \hat{\beta}) \\ \mathbf{Cov}(\hat{\alpha}, \hat{\beta}) & \mathbf{V}(\hat{\beta}) \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} \sum W_i^2 & \sum X_i W_i \\ \sum W_i X_i & \sum X_i^2 \end{pmatrix}^{-1} \\ &= \frac{\sigma^2}{(\sum W_i^2)(\sum X_i^2) - (\sum W_i X_i)^2} \begin{pmatrix} \sum X_i^2 & -\sum X_i W_i \\ -\sum W_i X_i & \sum W_i^2 \end{pmatrix} \end{aligned}$$

となるので、完全な多重共線性の場合には、 $(\sum W_i^2)(\sum X_i^2) - (\sum W_i X_i)^2 = 0$  となり（すなわち、逆行列が存在しない）、推定値の分散が無限大となる。

推定値の分散が無限大という意味は、どこにパラメータがあるか分からない（ $\alpha$  や  $\beta$  の信頼区間が  $-\infty$  から  $+\infty$  の範囲になる）ということの意味する。

簡単化のため， $\bar{W} = \frac{1}{n} \sum W_i = 0$ ， $\bar{X} = \frac{1}{n} \sum X_i = 0$  とする。 $W_i$  と  $X_i$  との相関係数を  $r$  とすると，

$$r = \frac{\frac{1}{n} \sum (W_i - \bar{W})(X_i - \bar{X})}{\sqrt{\frac{1}{n} \sum (W_i - \bar{W})^2} \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2}} = \frac{\sum W_i X_i}{\sqrt{(\sum W_i^2)(\sum X_i^2)}}$$

となる。さらに， $r$  を用いて， $\mathbf{V}(\hat{\alpha})$ ， $\mathbf{V}(\hat{\beta})$  を求めると，

$$\mathbf{V}(\hat{\alpha}) = \frac{\sigma^2 \sum X_i^2}{(\sum W_i^2)(\sum X_i^2) - (\sum W_i X_i)^2} = \frac{\sigma^2}{(\sum W_i^2) \left( 1 - \frac{(\sum W_i X_i)^2}{(\sum W_i^2)(\sum X_i^2)} \right)} = \frac{\sigma^2}{(\sum W_i^2)(1 - r^2)}$$

$$\mathbf{V}(\hat{\beta}) = \frac{\sigma^2 \sum W_i^2}{(\sum W_i^2)(\sum X_i^2) - (\sum W_i X_i)^2} = \frac{\sigma^2}{(\sum X_i^2) \left( 1 - \frac{(\sum W_i X_i)^2}{(\sum W_i^2)(\sum X_i^2)} \right)} = \frac{\sigma^2}{(\sum X_i^2)(1 - r^2)}$$

が得られる。

これは、 $r$  が 1 または  $-1$  に近づくにつれて (または、 $r^2$  が 1 に近づくにつれて)、 $V(\hat{\alpha})$ 、 $V(\hat{\beta})$  は大きくなるということを意味する。

⇒ 係数の推定値の有意性が低くなる。

⇒ 本来は  $W_i$  や  $X_i$  が  $Y_i$  に影響を与えているにもかかわらず、統計的に有意な推定値は得られなくなるので、回帰分析によって理論モデルを立証しようという試みは成功しなくなる。

一般化： 重回帰モデル

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

において、説明変数  $X_{1i}$ 、 $X_{2i}$ 、 $\cdots$ 、 $X_{ki}$  間に線形関係があれば、多重共線性が起こる。



ダミー変数の注意点： 例えば，四半期データの場合，ダミー変数を 4 つでなく，3 つにした理由は，多重共線性を回避するためである。

すなわち，下記の回帰式は推定可能か？

$$Y_i = \alpha + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta X_i + u_i$$

ただし， $j = 1, 2, 3, 4$  について，

$$D_{ji} = \begin{cases} 1, & i \text{ 番目のデータ (第 } i \text{ 期) が第 } j \text{ 四半期の場合} \\ 0, & \text{それ以外の場合} \end{cases}$$

$i$	定数項	$D_{1i}$	$D_{2i}$	$D_{3i}$	$D_{4i}$
1 (第1四半期)	1	1	0	0	0
2 (第2四半期)	1	0	1	0	0
3 (第3四半期)	1	0	0	1	0
4 (第4四半期)	1	0	0	0	1
5 (第1四半期)	1	1	0	0	0
6 (第2四半期)	1	0	1	0	0
7 (第3四半期)	1	0	0	1	0
8 (第4四半期)	1	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮

この場合， $D_{1i} + D_{2i} + D_{3i} + D_{4i} = 1$ （すなわち，定数項）となる。

定数項とダミー変数  $D_{ji}$ ， $j = 1, 2, 3, 4$  が線形結合の関係にある。

$D_{4i}$  を消去して ,

$$\begin{aligned} Y_i &= \alpha + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta X_i + u_i \\ &= \alpha + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (1 - D_{1i} - D_{2i} - D_{3i}) + \beta X_i + u_i \\ &= (\alpha + \alpha_4) + (\alpha_1 - \alpha_4) D_{1i} + (\alpha_2 - \alpha_4) D_{2i} + (\alpha_3 - \alpha_4) D_{3i} + \beta X_i + u_i \\ &= \alpha^* + \alpha_1^* D_{1i} + \alpha_2^* D_{2i} + \alpha_3^* D_{3i} + \beta X_i + u_i \end{aligned}$$

$\alpha^*$  ,  $\alpha_1^*$  ,  $\alpha_2^*$  ,  $\alpha_3^*$  (4つのパラメータ) を推定することはできるが ,  $\alpha$  ,  $\alpha_1$  ,  $\alpha_2$  ,  $\alpha_3$  ,  $\alpha_4$  (5つのパラメータ) を推定することは出来なくなる。

多重共線性の症状 : 多重共線性が起こっていると考えられるケースは ,

1. 推定値の符号が理論と合わない。

2. 決定係数 ( $R^2$  や  $\bar{R}^2$ ) は大きいのに、個々の  $t$  値は小さい。
3. 観測値の数 (データ数) を少し増やすと、推定値が大きく変わる。
4. 説明変数を増減すると、推定値が大きく変動する。

等である。

## 6.4 $F$ 検定について

複数の線形制約の検定を行う場合に  $F$  検定が用いられる。

## 6.4.1 いくつかの例

例 1：コブ＝ダグラス型生産関数：  $Q_i$  は生産量， $K_i$  は資本， $L_i$  は労働とする。生産関数を推定する。

$$\log(Q_i) = \beta_1' + \beta_2 \log(K_i) + \beta_3 \log(L_i) + u_i,$$

において，一次同次の制約  $\beta_2 + \beta_3 = 1$  を検定したい。

すなわち，帰無仮説，対立仮説は以下のように表される。

$$\text{帰無仮説 } H_0 : \beta_2 + \beta_3 = 1 ,$$

$$\text{対立仮説 } H_1 : \beta_2 + \beta_3 \neq 1 ,$$

例 2：構造変化の検定：  $n_0$  期以前と  $n_0 + 1$  期以降とで経済構造が変化したと考えて推定を行う。

しかも，定数項，傾き共に変化したと想定した場合，回帰式は以下のようなになる。

$$Y_i = \alpha + \beta X_i + \gamma D_i + \delta D_i X_i + u_i,$$

ただし，

$$D_i = \begin{cases} 0, & i = 1, 2, \dots, n_0 \text{ のとき,} \\ 1, & i = n_0 + 1, n_0 + 2, \dots, n \text{ のとき,} \end{cases}$$

とする。

構造変化が  $n_0 + 1$  期で起こったかどうかを検定したい。

すなわち，帰無仮説，対立仮説は以下のように表される。

$$\text{帰無仮説 } H_0 : \gamma = \delta = 0 ,$$

$$\text{対立仮説 } H_1 : \gamma \neq 0 , \text{ または } , \delta \neq 0 ,$$

例 3：多重回帰モデルの係数の同時検定： 2つの説明変数が含まれる場合を考える。

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

のモデルにおいて， $X_i$  と  $Z_i$  のどちらも， $Y_i$  に影響を与えていないという仮説を検定したい。

この場合，帰無仮説，対立仮説は以下のように表される。

$$\text{帰無仮説 } H_0 : \beta = \gamma = 0 ,$$

$$\text{対立仮説 } H_1 : \beta \neq 0 , \text{ または } , \gamma \neq 0 ,$$

## 6.4.2 検定の方法

---

(\* 復習)  $F$  分布 :

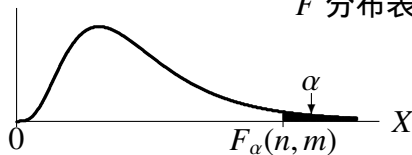
$U \sim \chi^2(n)$  ,  $V \sim \chi^2(m)$  ,  $U$  と  $V$  は独立とする。

このとき ,  $F = \frac{U/n}{V/m} \sim F(n, m)$  となる。

---



F 分布表 ( $\alpha = 0.05$ )



$$\alpha = \mathbf{Prob}(X > F_{\alpha}(n, m)) = \int_{F_{\alpha}(n, m)}^{\infty} f(x) \mathbf{d}x$$

$m \setminus n$	1	2	3	4	5	6	7	8	9	10	11	12	13
1	161	200	216	225	230	234	237	239	241	242	243	244	245
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45

例 1：ゼロ制約の検定：

多重回帰モデル

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

において，パラメータ  $\beta_1, \beta_2, \dots, \beta_k$  に何らかの制約が妥当かどうかを検定する。

制約の数を  $G$  個とする。

全く制約の無い場合に得られた残差を  $\hat{u}_i$  とする。

制約を含めて推定されたときの残差を  $\tilde{u}_i$  とする。

すなわち，

帰無仮説  $H_0 : \beta_{k-G+1} = \cdots = \beta_k = 0$

対立仮説  $H_1 : H_0$  でない

を検定する場合 ,

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{k-G} X_{k-G,i} \\ + \beta_{k-G+1} X_{k-G+1,i} + \cdots + \beta_k X_{ki} + u_i$$

の推定によって得られた残差を  $\hat{u}_i$  (制約なし残差) とおき ,

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{k-G} X_{k-G,i} + u_i$$

の推定によって得られた残差を  $\tilde{u}_i$  (制約付き残差) とする。

1.  $H_0$  が真のとき ,  $\frac{\sum \tilde{u}_i^2 - \sum \hat{u}_i^2}{\sigma^2} \sim \chi^2(G)$  となる。(証明略)
2. また ,  $\frac{\sum \hat{u}_i^2}{\sigma^2} \sim \chi^2(n-k)$  となる。(証明略)
3. さらに ,  $\frac{\sum \tilde{u}_i^2 - \sum \hat{u}_i^2}{\sigma^2}$  と  $\frac{\sum \hat{u}_i^2}{\sigma^2}$  とは独立に分布する。(証明略)
4. したがって , この場合 ,

$$\frac{(\sum \tilde{u}_i^2 - \sum \hat{u}_i^2)/G}{\sum \hat{u}_i^2/(n-k)} \sim F(G, n-k)$$

となる。(証明略)

$\frac{(\sum \tilde{u}_i^2 - \sum \hat{u}_i^2)/G}{\sum \hat{u}_i^2/(n-k)}$  を  $F$  統計量と呼ぶ。