第7章　ダミー変数

7.1　異常値ダミー

データに異常値が含まれている場合，経済構造がある時期から変化した場合，ダミー変数を使う。

ダミー変数とは，0 と 1 から成る変数のことである。

例えば，データが 20 期間あるとして，9 期目のデータが，回帰直線から離れている場合 (異常値の場合) を考える。

$$D_i = \begin{cases} 0, & i \neq 9 \text{ のとき} \\ 1, & i = 9 \text{ のとき} \end{cases}$$

という変数を作り,

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

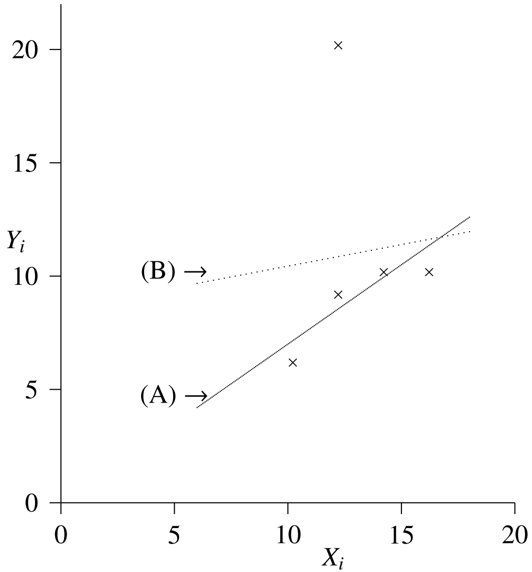
を推定する。δ の推定値 $\hat{\delta}$ の有意性を調べることによって，異常値かどうかの検定ができる。

数値例：　今までと同様に，以下の数値例をとりあげる。

i	Y_i	X_i	D_i
1	6	10	0
2	9	12	0
3	10	14	0
4	10	16	0
5	20	12	1

第 5 期目が異常値である。

図 3： 異常値



(A) は $i = 1, 2, 3, 4$ のデータを使って，推定した回帰直線である。(B) は $i = 1, 2, 3, 4, 5$ のデータを使って，推定した回帰直線である。(A), (B) の推定結果は以下のとおりである。

$$\text{(A):}\quad Y_i = \underset{(0.095)}{0.3} + \underset{(2.708)}{0.65} X_i,$$
$$R^2 = 0.786, \qquad s^2 = 1.072^2,$$

$$\text{(B):}\quad Y_i = \underset{(0.49)}{8.54} + \underset{(0.14)}{0.19} X_i,$$
$$R^2 = 0.007, \qquad s^2 = 6.09^2,$$

ただし，係数の推定値の下の括弧内は t 値を表すものとする。

　このように，結果が大幅に変わる。第 5 期は異常値なので，ダミー変数を用いて，

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

として推定を行う。$i = 1, 2, 3, 4$ について，$D_i = 0$ とし，$i = 5$ について，$D_i = 1$ とする変数である。この回帰式の意味は,

$$Y_i = \begin{cases} \alpha + \beta X_i + u_i, & i = 1, 2, 3, 4 \text{ のとき}, \\ (\alpha + \gamma) + \beta X_i + u_i, & i = 5 \text{ のとき}, \end{cases}$$

となる。推定結果は,

$$Y_i = \underset{(0.095)}{0.3} + \underset{(2.708)}{0.65} X_i + \underset{(9.73)}{11.9} D_i,$$

$$R^2 = 0.979, \quad s^2 = 1.072^2,$$

となる。この場合，$\hat{Y}_5 = Y_5$，すなわち，$\hat{u}_5 = 0$ となることに注意。

7.2　構造変化ダミー

次に，9 期目以前と以降とで，経済構造が変化している場合を考える。

$$D_i = \begin{cases} 0, & i = 1, 2, \cdots, 8 \text{ のとき} \\ 1, & i = 9, 10, \cdots, 20 \text{ のとき} \end{cases}$$

という変数を作り,

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

を推定する (定数項だけが変化したと考えた場合)。または,

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

を推定する (定数項も係数も変化)。

δ や γ の推定値の有意性を調べることによって，構造変化の検定を行うことができる。

上の例でデータを示すと,

i	Y_i	X_i	D_i	D_iX_i
1	Y_1	X_1	0	0
2	Y_2	X_2	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
8	Y_8	X_8	0	0
9	Y_9	X_9	1	X_9
10	Y_{10}	X_{10}	1	X_{10}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
20	Y_{20}	X_{20}	1	X_{20}

となる。

7.3 季節ダミー

季節性のあるデータを扱う場合,

$$D_{1i} = \begin{cases} 1, & i\text{ が第一四半期のとき} \\ 0, & \text{他} \end{cases}$$

$$D_{2i} = \begin{cases} 1, & i \text{ が第二四半期のとき} \\ 0, & \text{他} \end{cases}$$

$$D_{3i} = \begin{cases} 1, & i \text{ が第三四半期のとき} \\ 0, & \text{他} \end{cases}$$

という 3 つのダミー変数を作り，

7.4　地域差ダミー

関西と関東とで賃金格差があるかどうかを調べたい。

$$w_i = \alpha + \beta D_i + \cdots + u_i$$

添え字 i は個人を表すものとする。

$$D_i = \begin{cases} 1, & i \text{ 番目の人が関東に住んでいるとき} \\ 0, & i \text{ 番目の人が関西に住んでいるとき} \end{cases}$$

7.5　男女別ダミー

男女間で賃金格差があるかどうかを調べたい。

$$w_i = \alpha + \beta D_i + \cdots + u_i$$

添え字 i は個人を表すものとする。

$$D_i = \begin{cases} 1, & i \text{ 番目の人が女性のとき} \\ 0, & i \text{ 番目の人が男性のとき} \end{cases}$$

第8章　関数型について

線型：

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

この場合，

$$\beta = \frac{\mathrm{d}Y_i}{\mathrm{d}X_i}$$

なので，β は，X_i が一単位上昇 (下落) したとき，Y_i は何単位上昇 (下落) するのかを表す。すなわち，β は限界係数と呼ばれる。

成長率：

$$100 \times \frac{Y_i - Y_{i-1}}{Y_{i-1}} = \alpha + \beta X_i + u_i,$$

として，成長率を被説明変数として用いる場合もある。$100 \times \dfrac{Y_i - Y_{i-1}}{Y_{i-1}}$ という変数をあらかじめ作っておき，これをこれまでの Y_i として扱う。

注意：

$Y_i = \alpha + \beta X_i + u_i$ と $100 \times \dfrac{Y_i - Y_{i-1}}{Y_{i-1}} = \alpha + \beta X_i + u_i$ では，得られる決定係数の大きさが全く異なる。単純に，R^2 や $\overline{R}^2$ による比較はこの場合出来ない。

$\Longrightarrow s^2$ で比較すればよい。

対数線型：

$$\log(Y_i) = \alpha + \beta \log(X_i) + u_i,$$

この場合，

$$\beta = \frac{\mathrm{d}\log(Y_i)}{\mathrm{d}\log(X_i)} = \frac{\dfrac{\mathrm{d}Y_i}{Y_i}}{\dfrac{\mathrm{d}X_i}{X_i}} = \frac{100\dfrac{\mathrm{d}Y_i}{Y_i}}{100\dfrac{\mathrm{d}X_i}{X_i}}$$

となる。

2 つ目の等号では，$\dfrac{\mathrm{d}\log(Y_i)}{\mathrm{d}Y_i} = \dfrac{1}{Y_i}$ が利用される。

3 つ目の等号の分子 $100\dfrac{\mathrm{d}Y_i}{Y_i}$ や分母 $100\dfrac{\mathrm{d}X_i}{X_i}$ は上昇率を表す。

したがって，β は，X_i が 1%上昇 (下落) したとき，Y_i は何%上昇 (下落) するのかを表す。β は弾力性と呼ばれる。

例： コブ＝ダグラス型生産関数：

$$Q_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3}$$

ただし，Q_i は生産量，K_i は資本，L_i は労働である。この場合，対数変換によって，

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

として，$\log(Q_i)$, $\log(K_i)$, $\log(L_i)$ のデータをあらかじめ変換しておき，最小二乗法で $\beta_1', \beta_2, \beta_3$ を推定する。また，生産関数には一次同次の制約 $\beta_2 + \beta_3 = 1$ を置く場合が多い。この場合は,

$$\begin{aligned}\log(Q_i) &= \beta_1' + \beta_2 \log(K_i) + \beta_3 \log(L_i)\\ &= \beta_1' + \beta_2 \log(K_i) + (1 - \beta_2)\log(L_i) + u_i\\ &= \beta_1' + \beta_2(\log(K_i) - \log(L_i)) + \log(L_i) + u_i,\end{aligned}$$

となるので,

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

を最小二乗法で推定し，β_1', β_2 を求めることになる。この場合も同様に，各変数をあらかじめ，$\log(Q_i) - \log(L_i)$，$\log(K_i) - \log(L_i)$ としてデータを作っておく必要がある。

二次式：

$$Y_i = \alpha + \beta X_i + \gamma X_i^2 + u_i,$$

$\Longrightarrow$ 平均費用と生産量との関係等

逆数：

$$Y_i = \alpha + \beta \frac{1}{X_i} + u_i,$$

$\Longrightarrow$ 賃金上昇率と失業率との関係 (フィリップス曲線)

遅れのある変数：　習慣的効果を考慮に入れたモデル：

$$Y_i = \alpha + \beta X_i + \gamma Y_{i-1} + u_i,$$

ラグ付き内生変数が説明変数に用いられる。

X_i の Y_i への効果は，短期効果，長期効果の 2 つある。β は短期効果を表す係数である。長期効果とは，$Y_i = Y_{i-1}$ となるときの，X_i から Y_i への影響を示す効果である。

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

として，Y_i について解くと，

$$Y_i = \frac{\alpha}{1-\gamma} + \frac{\beta}{1-\gamma} X_i + \frac{1}{1-\gamma} u_i,$$

となり，$\dfrac{\beta}{1-\gamma}$ が X_i の Y_i への長期効果を表す係数となる。

問題点：

1. 最小二乗法の仮定の一つに，説明変数は確率変数ではないという仮定がある。ラグ付き内生変数を説明変数に加えることによって，この仮定が満たされなくなる。最小二乗推定量は最小分散線型不偏推定量ではなくなる。

2. Y_i と X_i とは，経済理論的に考えると，相関が高いはず。Y_i と Y_{i-1} は相関が高い。当然，Y_{i-1} と X_i も高い相関を示す。

   $\Longrightarrow$ 多重共線性の可能性が高い。

3. DW 統計量は意味をなさない。(DW については，後述)

遅れのある変数の解釈 (部分調整モデル)：　X_i が与えられたときの Y の最適水準を Y_i^* とする。

$$Y_i^* = \alpha + \beta X_i,$$

現実の水準 Y_i は，最適水準 Y_i^* と前期の水準 Y_{i-1} との差の一定割合と前期の水準 Y_{i-1} との和で与えられるとする。調整関数を考える。

$$Y_i - Y_{i-1} = \lambda(Y_i^* - Y_{i-1}) + u_i,$$

ただし，u_i は互いに独立で同一な分布の誤差項，$0 < \lambda < 1$ とする。

よって，

$$Y_i = \lambda\alpha + \lambda\beta X_i + (1 - \lambda)Y_{i-1} + u_i,$$

を得る。

Y_{i-1} と u_i との相関はない。

しかし，Y_{i-1} が説明変数の一つに入っている (説明変数間が確率変数でないという仮定に反する)。

推定量は不偏推定量ではないが，一致推定量である (証明略)。

第9章　雑多なこと

9.1　系列相関：DW について

9.1.1　DW について

最小自乗法の仮定の一つに,「撹乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」というものがあった。ダービン・ワトソン比 (DW) とは，誤差項の系列相関，すなわち，u_i と u_{i-1} との間の相関の有無を検定するために考案された。

$\Longrightarrow$ 時系列データのときのみ有効

$u_1, u_2, \cdots, u_n$ の系列について，それぞれの符号が，＋＋＋----＋＋---＋＋のように，プラスが連続で続いた後で，マイナスが連続で続くというような場

合，$u_1, u_2, \cdots, u_n$ は正の系列相関があると言う。また，+ - + - + - + - + のように交互にプラス，マイナスになる場合，$u_1, u_2, \cdots, u_n$ 負の系列相関があると言う。

特徴： $u_1, u_2, \cdots, u_i$ から u_{i+1} の符号が予想できる。$\Longrightarrow$「$u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」という仮定に反する。

すなわち，ダービン・ワトソン比とは，回帰式が

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

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときに，H_0： $\rho = 0, H_1$： $\rho \neq 0$ の検定である。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

ダービン・ワトソン比の定義は次の通りである。

$$DW = \frac{\sum_{i=2}^{n}(\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^{n}\hat{u}_i^2}$$

DW は近似的に，次のように表される。

$$DW = \frac{\sum_{i=2}^{n}(\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^{n}\hat{u}_i^2} = \frac{\sum_{i=2}^{n}\hat{u}_i^2 - 2\sum_{i=2}^{n}\hat{u}_i\hat{u}_{i-1} + \sum_{i=2}^{n}\hat{u}_{i-1}^2}{\sum_{i=1}^{n}\hat{u}_i^2}$$