ダミー変数について*

遠藤　薫

1. 序

計量経済分析においてダミー変数はしばしば次のように用いられる。地域別の時系列をプールしたデータで推定を行なう場合は，定数項あるいは係数における地域差を表わすために(1)，また時系列を二つの時期に分けたときは，前期と後期での構造の変化を表わすために用いられる。また季節調整をしていない四半期データで季節要因を推定したり，異常値を処理するために用いられたりする(2)。

本稿ではこのようなダミー変数は広く線形回帰モデルの中でどのように位置づけられるかを相等性のテストあるいは共分散分析との関係で述べ，最後に地域あるいは時期によって説明変数の変動に著しい違いがある場合に生じる一つの問題点について考察する。

2. 回帰関係の分散分析

回帰分析，共分散分析，実験計画法における分散分析はいずれも線形モデルの分散分析としてとらえることができるが，本稿の基礎となる回帰関係の分散分析について最初に述べる(3)。

---

* 本稿は長谷部亮一教授，久次智雄教授の有益な御教示に負うものであり，ここに深く感謝致したい．

(1) 経済企画庁『全国地域計量モデルの研究』(大蔵省印刷局，1968)．

(2) しかしながらダミー変数の意味するところのものについては注意深く検討されねばならない．たとえば経済企画庁，前掲書，p. 67.

(3) 本節は J. Johnston, *Econometric Methods, 2nd edition* (McGraw-Hill, 1972) [竹内・関谷・栗山・美添・舟岡訳『計量経済学の方法，上・下』，東洋経済新報社]，第5章に拠っている．

線形回帰モデルを

$$Y_i=\beta_1+\beta_2X_{2i}+\cdots+\beta_kX_{ki}+u_i \qquad i=1,\ \ldots,\ n \tag{1}$$

とする。$X_2,\ X_3,\ \ldots,\ X_k$ は説明変数，Y は被説明変数，u は攪乱項である。攪乱項は平均が 0，分散が σ^2 の正規分布に従うとし，共分散は 0 とする。また説明変数は確定変数であり，定数項に関する n 個の 1 と，各説明変数についての n 個の観測値を成分とする $n\times k$ 型行列の階数は k とする。

最小二乗法による β の推定値を $\hat{\beta}$ とすると

$$Y_i=\hat{\beta}_1+\hat{\beta}_2X_{2i}+\cdots+\hat{\beta}_kX_{ki}+e_i \qquad i=1,\ \ldots,\ n \tag{2}$$

となる。ただし

$$e_i=Y_i-(\hat{\beta}_1+\hat{\beta}_2X_{2i}+\cdots+\hat{\beta}_kX_{ki}) \qquad i=1,\ \ldots,\ n \tag{3}$$

とする。

一方 Y と $X_2,\ X_3,\ \ldots,\ X_k$ について各々観測値の平均からの偏差をとりそれを小文字で表わすと，

$$y_i=\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i}+\cdots+\hat{\beta}_kx_{ki}+e_i \qquad i=1,\ \ldots,\ n \tag{4}$$

となる。ただし $y_i=Y_i-\bar{Y},\ x_{2i}=X_{2i}-\bar{X}_2,\ \cdots$ であり，$\bar{Y}=\sum_{i=1}^{n}Y_i/n,\ \bar{X}_2=\sum_{i=1}^{n}X_{2i}/n,\ \cdots$ である。(2) と (4) の e_i は同じ値となる。(4) 式を行列で表わして

$$\boldsymbol{y}=\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{e} \tag{5}$$

とする。ただし $\boldsymbol{y}$ は $n\times1$，$\boldsymbol{X}$ は $n\times(k-1)$，$\hat{\boldsymbol{\beta}}$ は $(k-1)\times1$，$\boldsymbol{e}$ は $n\times1$ の行列である。このとき

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{6}$$

$$\boldsymbol{e}'\boldsymbol{e}=\boldsymbol{y}'\boldsymbol{y}-\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \tag{7}$$

となっている。ここで

$$\boldsymbol{Z}=\boldsymbol{X}\boldsymbol{W} \tag{8}$$

ただし

$$\boldsymbol{Z}=\begin{pmatrix} z_{21} & z_{31} & \cdots & z_{k1} \\ z_{22} & z_{32} & \cdots & z_{k2} \\ \vdots & \vdots & & \vdots \\ z_{2n} & z_{3n} & & z_{kn} \end{pmatrix} \qquad \boldsymbol{W}=\begin{pmatrix} w_{22} & w_{32} & \cdots & w_{k2} \\ 0 & w_{33} & \cdots & w_{k3} \\ \vdots & & & \\ 0 & 0 & \cdots & w_{kk} \end{pmatrix} \tag{9}$$

とおいたとき，

$$\sum_{i=1}^{n} z_{ji}^2 = 1,\ j=2,\ \ldots,\ k \tag{10}$$

$$\sum_{i=1}^{n} z_{ji} z_{li} = 0,\ j, l=2,\ \ldots,\ k;\ j \neq l \tag{11}$$

となるように w の値を $x_2, x_3, \ldots, x_k$ から決定することができる。しかも w_{22} は x_{2i} だけの関数として，w_{32} と w_{33} は x_{2i} と x_{3i} だけの関数としてというように表わすことができる。この結果任意の j について z_{ji} は $x_{2i}, x_{3i}, \ldots, x_{ji}$ までだけによって決定され，$x_{j+1,i}, x_{j+2,i}, \ldots, x_{ki}$ とは無関係であることがわかる。さらに

$$\boldsymbol{\beta}^* = \boldsymbol{W}^{-1}\boldsymbol{\beta} \text{ および } \hat{\boldsymbol{\beta}}^* = \boldsymbol{W}^{-1}\hat{\boldsymbol{\beta}} \tag{12}$$

とおくと

$$\boldsymbol{y} = \boldsymbol{Z}\hat{\boldsymbol{\beta}}^* + \boldsymbol{e} \tag{13}$$

を導くことができ，

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{Z}'\boldsymbol{y} \tag{14}$$

は $z_2, z_3, \ldots, z_k$ を説明変数とする回帰モデルにおける係数の最小2乗推定量であることを示すことができる。また

$$\mathrm{Var}\,[\hat{\boldsymbol{\beta}}^*] = \sigma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \sigma^2\boldsymbol{I}_{k-1} \tag{15}$$

を導くこともできるので，$\hat{\beta}_i^*$ は平均 β_i^*，分散 σ^2 で独立に正規分布に従うことになる。したがって $\sum_{i=2}^{k}(\hat{\beta}_i^* - \beta_i^*)^2/\sigma^2$ は自由度 $k-1$ の x^2 分布に従う，一方 $\sum_{i=1}^{n} e_i^2/\sigma^2$ はこれと独立に自由度 $n-k$ の x^2 分布に従うことから，

$$F = \frac{\sum_{i=1}^{k}(\hat{\beta}_i^* - \beta_i^*)/(k-1)}{\sum_{i=1}^{n} e_i^2/(n-k)} \tag{16}$$

は自由度 $k-1$, $n-k$ の F 分布に従い，仮説 $\beta_2^* = \beta_3^* = \ldots = \beta_k^* = 0$ のもとでは

$$F = \frac{\sum_{i=2}^{k}\hat{\beta}_i^{*2}/(k-1)}{\sum_{i=1}^{n} e_i^2/(n-k)} \tag{17}$$

が自由度 $k-1$, $n-k$ の F 分布に従う。

このとき β_i^* についての上記の仮説は(12)より $\boldsymbol{\beta}=\boldsymbol{0}$ あるいは $\beta_2=\beta_3=\cdots=\beta_k=0$ という仮説と同じである。したがって線形回帰モデル(1)において係数 $\beta_2, \beta_3, \cdots, \beta_k$ がともに 0 と有意に異なるかどうかは仮説 $\beta_2=\beta_3=\cdots=\beta_k=0$ のもとで(17)を用いて検定されることになる。(8)と(12)より

$$\sum_{i=2}^{k}\hat{\beta}_i^{*2}=\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \tag{18}$$

であるから分散分析表は表1のようになる。

表1　仮説 $\beta_2=\beta_3=\cdots=\beta_k=0$ を検定するための分散分析表

要　因	平　方　和	自由度	平　均　平　方
$X_2, X_3, \cdots, X_k$	$\sum_{i=2}^{k}\hat{\beta}_i^{*2}=\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$	$k-1$	$\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}/(k-1)$
残　差	$\sum_{i=1}^{n}e_i^2=\boldsymbol{e}'\boldsymbol{e}$	$n-k$	$\boldsymbol{e}'\boldsymbol{e}/(n-k)$
総　計	$\sum_{i=1}^{n}y_i^2=\boldsymbol{y}'\boldsymbol{y}$	$n-1$	$\boldsymbol{y}'\boldsymbol{y}/(n-1)$

(17)は $X_2, X_3, \cdots, X_k$ によって説明される平方和が残差平方和に比べて，自由度で調整したうえでどれだけ大きいかみたものであり，この比が1近くの小さな値であるときは説明変数の効果が誤差と区別できないことになり，仮説 $\boldsymbol{\beta}=\boldsymbol{0}$ は棄却されない，すなわち説明変数 $X_2, X_3, \cdots, X_k$ は全体として Y への影響を持たないのではないかという仮説を退けることはできない。正確には有意水準あるいは危険率を定めて F 表から結論を導く。

次に説明変数としてははじめは $X_2, X_3, \cdots, X_r$ を用いたが，さらに $X_{r+1}, X_{r+2}, \cdots, X_k$ という $k-r$ 個の説明変数をつけ加えたとき，そのことに意味があるかどうかを調べたい。このときは仮説として $\beta_{r+1}=\beta_{r+2}=\cdots=\beta_k=0$ をおき，

$$F=\frac{\sum_{i=r+1}^{k}\hat{\beta}_i^{*2}/(k-r)}{\sum_{i=1}^{n}e_i^2/(n-k)} \tag{19}$$

が自由度 $k-r$, $n-1$ の F 分布に従うことを用いて検定を行なう。なおこのときの仮説は W^{-1} が上三角形行列であるため $\beta_{r+1}^*=\beta_{r+2}^*=\cdots=\beta_k^*=0$ と同じである。分散分析表は表2となる。

表2　$\beta_{r+1}=\beta_{r+2}=\cdots=\beta_k=0$ を検定するための分散分析表

要　　因	平　方　和	自由度	平 均 平 方
$X_2,\ X_3,\ \cdots,\ X_r$	$\sum_{i=2}^{r}\hat{\beta}_i^{*2}$	$r-1$	$\sum_{i=2}^{r}\hat{\beta}_i^{*2}/(r-1)$
$X_{r+1},\ X_{r+2},\ \cdots,\ X_k$	$\sum_{i=r+1}^{k}\hat{\beta}_i^{*2}$	$k-r$	$\sum_{i=r+1}^{k}\hat{\beta}_i^{*2}/(k-r)$
$X_2,\ X_3,\ \cdots,\ X_k$	$\sum_{i=2}^{k}\hat{\beta}_i^{*2}=\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$	$k-1$	$\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}/(k-1)$
残　　差	$\sum_{i=1}^{n}e_i^2=\boldsymbol{e}'\boldsymbol{e}$	$n-k$	$\boldsymbol{e}'\boldsymbol{e}/(n-k)$
総　　計	$\sum_{i=1}^{n}y_i^2=\boldsymbol{y}'\boldsymbol{y}$	$n-1$	$\boldsymbol{y}'\boldsymbol{y}/(n-1)$

先にみたように任意の j について z_{ji} は $x_{2i},\ x_{3i},\ \cdots,\ x_{ji}$ までによって決定され，$x_{j+1,i},\ x_{j+2,i},\ \cdots,\ x_{ki}$ とは無関係であるから，

$$\hat{\beta}_j^*=\sum_{i=1}^{n}z_{ji}y_i$$

は $x_{2i},\ x_{3i},\ \cdots,\ x_{ji}$ と y_i のみによって決定される。したがって $\sum_{i=2}^{r}\hat{\beta}_i^{*2}$ は

$$y_i=\hat{\hat{\beta}}_2x_{2i}+\hat{\hat{\beta}}_3x_{3i}+\cdots+\hat{\hat{\beta}}_rx_{ri}+s_i \qquad i=1,\ \cdots,\ n \tag{20}$$

ただし，

$$s_i=y_i-(\hat{\hat{\beta}}_2x_{2i}+\hat{\hat{\beta}}_3x_{3i}+\cdots+\hat{\hat{\beta}}_rx_{ri}) \qquad i=1,\ \cdots,\ n, \tag{21}$$

における，説明変数 $X_2,\ X_3,\ \cdots,\ X_r$ によって説明される平方和と考えることができる。なお，$\hat{\hat{\beta}}_2,\ \hat{\hat{\beta}}_3,\ \cdots,\ \hat{\hat{\beta}}_r$ は $X_2,\ X_3,\ \cdots,\ X_r$ を説明変数とする回帰モデルにおける係数の最小二乗推定量である。(20) を行列で表わして

$$\boldsymbol{y}=\boldsymbol{X}_1\hat{\hat{\boldsymbol{\beta}}}+\boldsymbol{s} \tag{22}$$

とする。ただし $\boldsymbol{X}_1$ は $n\times(r-1)$, $\hat{\hat{\boldsymbol{\beta}}}$ は $(r-1)\times1$, $\boldsymbol{s}$ は $n\times1$ の行列とする。このとき表2の最初の二つの平方和は次のように表わすことができる。

$$\sum_{i=2}^{r} \hat{\beta}_i{}^{*2} = \hat{\hat{\beta}}' X_1' y \tag{23}$$

$$\sum_{i=r+1}^{k} \hat{\beta}_i{}^{*2} = \sum_{i=2}^{k} \hat{\beta}_i{}^{*2} - \sum_{i=2}^{r} \hat{\beta}_i{}^{*2}$$

$$= \hat{\beta}' X' y - \hat{\hat{\beta}}' X_1' y \tag{24}$$

ここで

$$\hat{\beta}' X' y - \hat{\hat{\beta}}' X_1' y = y'y - e'e - (y'y - s's)$$

$$= s's - e'e \tag{25}$$

であるから (19) は

$$F = \frac{(s's - e'e)/(k-r)}{e'e/(n-k)} \tag{26}$$

と表わすことができる。すなわち説明変数を $k-r$ 個追加したことによる，説明される平方和の増分は，説明変数が追加されたことによる残差平方和の減少分に等しい。$y'y$ はどちらでも同じだからである。

説明変数を一つだけ追加したときは $X_r = X_{k-1}$ の場合であるから，仮説 $\beta_k = 0$ (すなわち $\beta_k{}^* = 0$) を検定するためには (19) の分子を $\hat{\beta}_k{}^{*2}/[k-(k-1)] = \hat{\beta}_k{}^{*2}$ とすればよい。

これは t 検定において t が自由度 $n-k$ の t 分布に従うことと同じことであり，$t^2 = F$ の関係にある。

以上いくつかの説明変数を線形回帰モデルの後のほうに追加した場合を述べたが，並べ変えを行なうことにより，途中の説明変数に関してもまったく同様のことがいえる。次に本節の方法を用いて，観測値が追加された場合，それがはじめの観測値と同じ構造から得られたものかどうか，あるいは相異なる二つの組の観測値が同じ構造から得られたものであるかどうかの検定についてそれを説明変数の追加に帰せしめて述べる。

3. 相等性のテスト

地域Ⅰと地域Ⅱ，あるいは前期と後期に同じ回帰直線をあてはめてよいかどうか，別々に回帰直線をあてはめたとして，それが統計的に有意に異なる

ものなのかどうかを知りたい。このときまったく違う回帰直線があてはまると考えてよいのか，あるいは一部分だけがちがう回帰直線があてはまると考えてよいのかが問題となる。二つの級（地域，期間等）の構造について，それがまったく同じ構造のものであるか，あるいは一部分が同じで残りの部分は違うのかということに関するものである。具体的には回帰直線の切片と勾配に関してどれは同じでどれは異なるかを調べることである。このことについて，二つの級で観測値の数に違いがあることも考慮しながら，どのような仮説をたててどのように検定を行うかを述べる。(4)

(1)　すべての係数（定数項も含む）についての相等性

二つの組について同じ定式化をした線形回帰モデルを考え，最初の級にあてはめられた回帰直線を

$$\boldsymbol{y}_1=\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1+\boldsymbol{e}_1 \tag{27}$$

二つ目の級にあてはめられた回帰直線を

$$\boldsymbol{y}_2=\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2+\boldsymbol{e}_2 \tag{28}$$

とする。ただし $\boldsymbol{y}_1$ は $n\times 1$, $\boldsymbol{X}_1$ は $n\times k$, $\hat{\boldsymbol{\beta}}_1$ は $k\times 1$, $\boldsymbol{e}_1$ は $n\times 1$ の行列とし，$\boldsymbol{y}_2$ は $m\times 1$, $\boldsymbol{X}_2$ は $m\times k$, $\hat{\boldsymbol{\beta}}_2$ は $k\times 1$, $\boldsymbol{e}_2$ は $m\times 1$ の行列とする。ただしここでは2節でと違って観測値はそのままの形で用い，偏差は用いない。したがって $\hat{\boldsymbol{\beta}}_1$ と $\hat{\boldsymbol{\beta}}_2$ の内容は定数項と係数のすべてを含むものである。また $k<n, m$ とする。このとき最初の級と二つ目の級とでは観測値が得られた構造に違いがあるかどうかは，仮説 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$ を検定することによって行なわれる。もしこの仮説が棄却されたときはある有意水準のもとで，二つの級の構造に違いがあることになり，棄却されないときは，構造が同じであるという仮説が受容されることになる。このとき2節での方法を直接用いるな

(4)　小宮隆太郎「計量経済学と共分散分析」森嶋・篠原・内田編『新しい経済分析』(創文社，1960)，第9章；Chow, G. C., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28 (Jul. 1960), pp. 591–605.；内田・栗林・矢島・渡部『経済予測と計量モデル』(日本経済新聞社，1966)，第Ⅶ章；Fisher, F. M., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, 38 (Mar. 1970), pp. 361–6.；Johnston, *op. cit.*, 第6章.

ら，

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} X_1 & 0 \\ X_2 & X_2 \end{pmatrix}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\lambda} \end{pmatrix}+\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \tag{29}$$

に関して仮説 $\lambda=0$ を検定することになる。ただし $\lambda=\beta_2-\beta_1$ とする。このとき(27)の $\hat{\beta}_1$ と(29)の $\hat{\beta}_1$ はいずれも $(X_1'X_1)^{-1}X'y$ であり，また(29)の $\hat{\lambda}$ は(28)の $\hat{\beta}_2$ と(27)の $\hat{\beta}_1$ の差であるこも確かめるこができる，したがって(27)，(28)での残差 e_1, e_2 と(29)での残差 e_1, e_2 は同じものである。いま，

$$V_1=\begin{pmatrix} X_1 \\ 0 \end{pmatrix} \qquad V_2=\begin{pmatrix} 0 \\ X_2 \end{pmatrix} \qquad V=V_1+V_2 \tag{30}$$

とすると

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=V\hat{\beta}_1+V_2\hat{\lambda}+\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \tag{31}$$

となり，分散分析表は表2と同じように作ることができる。それを表3とする。

表3　仮説 $\lambda=0$ を検定するための分散分析表

要　因	平方和	自由度	平均平方
V	A	$k-1$	$A/(k-1)$
V_2	B	k	B/k
V, V_2	C	$2k-1$	$C/(2k-1)$
残　差	$e_1'e_1+e_2'e_2$	$n+m-2k$	$e_1'e_1+e_2'e_2/(n+m-2k)$
総　計	D	$n+m-1$	$D/(n+m-1)$

このとき仮説 $\lambda=0$ のもとでは

$$F=\frac{B/k}{(e_1'e+e_2'e_2)/(n+m-2k)} \tag{32}$$

が自由度 k, $n+m-2k$ の F 分布に従うことを用いて検定を行なう。ここで B は(22)～(26)と同じようにして次のように求められる。まず最小二乗法により

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = V\hat{\beta} + s \tag{33}$$

となる推定値 $\hat{\beta}$ と残差 s を求める。ただし，$\hat{\beta}$ は $k\times 1$，s は $n\times 1$ の行列である。このとき

$$D = A + s's \tag{34}$$

となる。次に(31)より

$$D = C + (e_1'e_1 + e_2'e_2) \tag{35}$$

また

$$C = A + B \tag{36}$$

である，したがって(34)～(36)より

$$B = s's - (e_1'e_1 + e_2'e_2) \tag{37}$$

となる。以上により(32)の F 値を求めるためにはまず(33)のように二つの級の観測値をあわせたものに共通の回帰直線をあてはめて残差平方和 $s's$ を求め，次に(31)のように説明変数の数（定数項も含めて）を2倍にして観察値 $V_2' = (0'X_2')$ を加えたときの回帰曲線を求めそのときの残差平方和 $e_1'e_1 + e_2'e_2$ を求めるとよい。しかし $e_1'e_1 + e_2'e_2$ については $e_1'e_1$ は(27)から，$e_2'e_2$ は(28)から計算されるものと同じなので，一般にはこれら(27)，(28)から求められる。なお(31)は，残差に関して

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \tag{38}$$

と同じであり[(5)]，これは(27)，(28)と同じである。

以上は最初の級の観察値の数 n も，二つ目の級の観察値の数 m もともに各々の級に共通に定式化された線形回帰モデルの係数（定数項も含めて）k よりも大きい場合であったが，二つ目の級では m が k より小さいことも生じる。この場合は二つ目の級の観測値に回帰直線を完全にあてはめることができ $e_2'e_2 = 0$ となる。このことから仮説，$\beta_1 = \beta_2$ あるいは $\lambda = 0$ に関する

(5) Fisher, *op. cit.*, p. 365.

検定は，

$$F=\frac{(s's-e_1'e_1)/m}{e_1'e_1/(n-k)} \tag{39}$$

を用いて行なわれる。[6]

ところで $(n+m)\times k$ 型行列 V_2 の一列目は n 個の0と m 個の1から構成されている。このように0または1の値をとる変数はダミー変数と呼ばれる。また V_2 の二列目は V の二列目のうち最初の n 個に0をかけ，残りの m 個に1をかけたものである。このようなときはダミー変数の乗法的使用と呼ばれる。[7] あるいは前者を定数項ダミー，後者を係数ダミーと呼ぶ。[8]

(2) 一部の係数（定数項も含む）についての相等性

(1)では二つの級の観測値に同一の線形回帰モデルを想定したとき，各々の級の定数項およびすべての係数が，全体として有意に異なるかどうかについての仮説検定であった。しかしさらに細かく考えて，その中の一部の係数あるいは定数項だけが異なるかどうかを調べたい。このとき残りの係数あるいは定数項については相異なることを前提としてである。まず最初の級の観測値だけにあてはめられた回帰直線を

$$y_1=X_1\hat{\beta}_1+W_1\hat{\alpha}_1+e_1, \tag{40}$$

二番目の組については

$$y_2=X_2\hat{\beta}_2+W_2\hat{\alpha}_2+e_2 \tag{41}$$

とする。ただし X_1 は $n\times k_1$, $\hat{\beta}_1$ は $k_1\times 1$, W_1 は $n\times k_2$, $\hat{\alpha}_1$ は $k_2\times 1$, e_1 は $n\times 1$ の行列，X_2 は $m\times k_1$, $\hat{\beta}_2$ は $k_2\times 1$, W_2 は $m\times k_2$, $\hat{\alpha}_2$ は $k_2\times 1$, e_2 は $m\times 1$ の行列とする。ただし $k_1+k_2<n, m$ とする。このとき仮説 $\beta_1=\beta_2$ の検定は

$$\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{pmatrix}X_1\\X_2\end{pmatrix}\hat{\beta}_1+\begin{pmatrix}0\\X_2\end{pmatrix}\hat{\lambda}+\begin{pmatrix}W_1 & 0\\0 & W_2\end{pmatrix}\begin{pmatrix}\hat{\alpha}_1\\\hat{\alpha}_2\end{pmatrix}+\begin{pmatrix}e_1\\e_2\end{pmatrix} \tag{42}$$

---

(6) Chow, *op. cit.*, pp. 598–9.
(7) 森口親司『計量経済学』(岩波書店，1974)，pp. 122–5；pp. 145–153.
(8) 経済企画庁，前掲書，pp. 66–7.

における仮説 $\boldsymbol{\lambda}=\boldsymbol{0}$ の検定と同じである。したがってここでの $\boldsymbol{e}_1'\boldsymbol{e}_1$ と

$$\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix}\hat{\boldsymbol{\beta}}+\begin{pmatrix}\boldsymbol{W}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{W}_2\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\alpha}}_1\\ \hat{\boldsymbol{\alpha}}_2\end{pmatrix}+\boldsymbol{s} \tag{43}$$

における $\boldsymbol{s}'\boldsymbol{s}$（ただし $\hat{\boldsymbol{\beta}}$ は $k_1\times 1$，$\boldsymbol{s}$ は $(n+m)\times 1$ の行列）を用いて，

$$F=\frac{[\boldsymbol{s}'\boldsymbol{s}-(\boldsymbol{e}_1'\boldsymbol{e}_1+\boldsymbol{e}_2'\boldsymbol{e}_2)]/k_1}{(\boldsymbol{e}_1'\boldsymbol{e}_1+\boldsymbol{e}_2'\boldsymbol{e}_2)/[n+m-2(k_1+k_2)]} \tag{44}$$

が自由度 k_1, $n+m-2(k_1+k_2)$ の F 分布に従うことを用いて検定を行なう。

いまの場合各々の級における観測値の数 n, m は(40)あるいは(41)における定数項および係数の数 k_1+k_2 より大きい場合であった。次に $m\leqq k_1+k_2$（ただし $m>k_2$）のときは，

$$F=\frac{(\boldsymbol{s}'\boldsymbol{s}-\boldsymbol{e}_1'\boldsymbol{e}_1)/(m-k_2)}{\boldsymbol{e}_1'\boldsymbol{e}_1/(n-k_1-k_2)} \tag{45}$$

が自由度 $m-k_2$, $n-k_1-k_2$ の F 分布に従うことが知られており，これをもちいて検定を行なう。

さてここでも（$\boldsymbol{0}'\boldsymbol{X}_2'$）はダミー変数を用いて表現することが出来る。一般によく用いられるのは二つの級の観測値にそれぞれあてはめられた回帰直線に勾配の差があるからどうかということであろう。このとき定数項については何ら仮定をおかない。そこでもし勾配に差がないという仮説が受容されたなら，勾配は同じであるとして切片に有意な差があるかを調べることになる。このことについては次節で述べる。なお本節では二つの組だけに限ったが二つ以上の級に関しても同じである。

4. ダミー変数

計測された回帰直線にダミー変数の利用が見出されるときは，前節(2)での一部の係数に関する相等性のテストで，構造の一部に違いはないという仮説が受容され，それを前提とした上で残りの部分に違いがあるかどうかが検定され，そこで同じという仮説が棄却されたものであると考えることができる。この後半部分ははじめからダミー変数を意識的に用いて検定を行なうこ

とになる。

(1) 切片の差の検定

最も簡単な場合として，説明変数は定数項の他に一つだけであるとし，その説明変数の勾配は二つの級で差がないと仮定する。このとき定数項あるいは切片に有意な差があるかどうかは

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} 1 \\ 1 \end{pmatrix}\hat{\alpha}_1+\begin{pmatrix} 0 \\ 1 \end{pmatrix}\hat{\lambda}+\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\hat{\beta}+\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \tag{46}$$

において仮説 $\lambda=0$ を検定するとよい。ただし X_1 は $n\times 1$, X_2 は $m\times 1$ の行列である。このときは0あるいは1の値をとるダミー変数を一つ追加したとき，そのことに意味があるかどうかの検定であり，2節の方法をそのまま適用できる。しかもこのときは F 検定も t 検定でもどちらでもよく，t 検定のほうが容易である。もし仮説が棄却されたときは，両方の組の説明変数を X, 被説明変数を Y で表わしダミー変数を D で表わしたとき，

$$\hat{Y}_i=\hat{\alpha}_1+\hat{\lambda}D+\hat{\beta}X_i \qquad i=1, \ldots, n+m \tag{47}$$

が推定された回帰線となる。

(2) 異常値について

被説明変数 Y の中に他と比べて大きく違った観測値があるとする。このときこの観測値は他の観測値と同じ構造から得られたものかどうか検定したい。このとき前節(1)のすべての係数の相等性のテストのところで二つ目の級の観測値の数は1，すなわち $m=1(<k)$ として検定できるであろう。またもしすべての係数の勾配は同じと仮定されるなら(46)のように切片だけについて差があるかどうか検定されることになる。このときは問題の観測値が i 期に得られたとするとダミー変数は i 期についてのみ1で，あとはすべて0であるようにするとよい。

5. 一つの問題点

前節ではダミー変数の使用を主として仮説検定の場面で考えてきたが，そこで測定されたことになる回帰直線を予測に用いる場合にはいろいろな問題

点が出てくることになろう。[9] 本節では一つの問題点について考察する。もっとも簡単な場合として前節(1)のモデルをとりあげる。もし二つの級で勾配に差がないと仮定せず，各々の級で別々に回帰直線を求めると

$$Y_{1j}=\hat{\alpha}_1+\hat{\beta}_1 X_{1j}+e_{1j} \qquad j=1, \ldots, n \tag{48}$$

$$Y_{2j}=\hat{\alpha}_2+\hat{\beta}_2 X_{2j}+e_{2j} \qquad j=1, \ldots, m \tag{49}$$

となる。このとき

$$\hat{\beta}_1=\frac{\sum_{j=1}^{n}(X_{1j}-\bar{X}_1)(Y_{1j}-\bar{Y}_1)}{\sum_{j=1}^{n}(X_{1j}-\bar{X}_1)^2} \tag{50}$$

および

$$\hat{\beta}_2=\frac{\sum_{j=1}^{m}(X_{2j}-\bar{X}_2)(Y_{2j}-\bar{Y}_2)}{\sum_{j=1}^{m}(X_{2j}-\bar{X}_2)^2} \tag{51}$$

となっている。ただし $\bar{X}_1=\sum_{j=1}^{n}X_{1j}/n$ であり他も同様である。一方二つの級では勾配には差はないが，切片については差がないという仮説が棄却され，

$$Y_{ij}=\hat{\alpha}_1+\hat{\lambda}D+\hat{\beta}X_{ij}+e_{ij} \qquad i=1, 2 \qquad j=1, \ldots, n_i \tag{52}$$

という回帰直線が計測されたとする。ただし D は最初の n 個が0で残りの m 個が1の値をとるダミー変数とし，$n_1=n$, $n_2=m$ とする。このとき，

$$\begin{aligned}\hat{\beta}&=\frac{\sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)(Y_{1j}-\bar{Y}_1)+\sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)(Y_{2j}-\bar{Y}_2)}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{1j}-\bar{X}_i)^2}\\&=\frac{\hat{\beta}_1\sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)^2+\hat{\beta}_2\sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)^2}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2}\end{aligned} \tag{53}$$

となる。ここで $\bar{X}=\sum_{i=1}^{2}\sum_{j=1}^{n_i}X_{ij}/(n_1+n_2)$ とすると，

(9) たとえば，Blumenthal, T., "A Test of the Klein-Shinkai Econometric Model of Japan," *International Economic Review*, 6 (May, 1965), pp. 211–28.

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2=\sum_{i=1}^{2}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2+\sum_{i=1}^{2}\sum_{j=1}^{n_i}(\bar{X}_i-\bar{X})^2 \qquad (54)$$

となり，右辺の第1項は級内変動，第2項は級間変動と呼ばれる。したがって(53)の $\hat{\beta}$ の分母は級内変動そのものであり，また分子は $\hat{\beta}_1$ に変数 X_1 の変動 $\sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)^2$ をかけたものと $\hat{\beta}_2$ に変数 X_2 の変動 $\sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)^2$ をかけたものの和である。

このように二つの級について勾配は同じであるとの仮定のもとに計測される(52)のような回帰直線の勾配 $\hat{\beta}$ は二つの級に別々にあてはめられた回帰直線の勾配 $\hat{\beta}_1$ と $\hat{\beta}_2$ から導くことができるが，そのとき二つの級の勾配のうち説明変数の変動が大きいほうの級の勾配が強くきいてくることになる。このことは説明変数のバラツキが大きいほどより精度の高い推定値が得られるという最小二乗法の基本的性質に基づくものであり，まったく当然のことといえる。

級が二つの場合について考えたが，それ以上の場合でも同様であろう。たとえば級が三つのときは，第三番目の組にあてはめられた直線を

$$Y_{3j}=\hat{\alpha}_3+\hat{\beta}_3X_{3j}+e_{3j} \qquad j=1,\ \ldots,\ n_3 \qquad (55)$$

とし，勾配は同じとして三つの級の観測値全体にあてはめられた回帰直線を

$$Y_{ij}=\hat{\alpha}_1+\hat{\lambda}_2D_2+\hat{\lambda}_3D_3+\hat{\beta}X_{ij}+e_{ij}$$
$$i=1,\ \ldots,\ 3 \qquad j=1,\ \ldots,\ n_i \qquad (56)$$

とする。ただし n_3 は三つ目の級の観測値の数であり，D_2 は最初の n_1 個と最後の n_3 個が0，中間の n_2 個が1の値をとるダミー変数，D_3 は最初の n_1+n_2 個が0，最後の n_3 個が1の値をとるダミー変数である。λ_2 は α_2 と α_1 の差，λ_3 は α_3 と α_1 の差を表わしている。(55)の $\hat{\beta}_3$ は

$$\hat{\beta}_3=\frac{\sum_{j=1}^{n_3}(X_{3j}-\bar{X}_3)(Y_{3j}-\bar{Y}_3)}{\sum_{j=1}^{n_3}(X_{3j}-\bar{X}_3)^2} \qquad (57)$$

であり，(56)の $\hat{\beta}$ はこれら $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ を用いると次のようになる。

$$\hat{\beta}=\frac{\sum_{i=1}^{3}\sum_{j=1}^{n_i}\hat{\beta}_i(X_{ij}-\bar{X}_i)^2}{\sum_{i=1}^{3}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2} \tag{58}$$

$\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ の加重平均が $\hat{\beta}$ であるといえる。

ダミー変数を(52)あるいは(56)のように用いた回帰直線によって予測を行なうときは，以上のように各級における説明変数の変動の違いを考慮に入れる必要があると思われる。

計量経済分析におけるダミー変数の使用について仮説検定との関係で基礎的事項について述べ，5節で一つの特徴をとりあげたが説明変数が一つの場合であった。二つ以上の説明変数がある場合についてみてみることがまだ残されているが，たとえば地域経済モデルのシミュレーション等において，これらのことが有用になると考える。