回帰方程式における異常値の検出

井　上　勝　雄

§0　一般に，次節で定式化するような単一回帰方程式モデルを最小二乗推定する場合，個々の残差を計算すると，異常にその絶対値が大きな残差がでる場合がある．そのような大きな残差に対応する観測標本を「異常値」と通常呼んでいる．このような異常値の原因としてその標本に対してのみモデルの説明変数として考慮されていない何らかの要因が影響を与えることが考えられる．この場合には，モデルの説明変数の係数がその標本に対してのみ変化すると考えられる．「異常値」の原因として以上の他に，観測誤差や集計上の誤りが挙げられる．

ある標本が異常値であると判定する単純な方法は，残差の絶対値がある指定された値よりも大きくなれば，その対応する標本は異常値であると判定する仕方である．[1)]

本稿の目的は，異常値の判定に関する方法を検定方式の議論にのせることにある．つまり，異常値の判定を統計的仮説検定の枠組の中で処理する方式を考察する．

§1で，以下の考察の基礎となる回帰モデルを定式化し，周知の最小二乗推定量とその統計的特性を利用する範囲内でまとめておく．したがってこの§で示されることがらについては通常のテキストブックでその詳細が記されている．§2で，検定すべき標本が先験的にどれであるかが判明しているケースについて異常値の検定方法を述べる．さらにこの方法は一般に予測区間を求める仕方と同等であることを示す．§3で，モデル推定の結果，特定の標本が異常値と

1)　たとえば，佐和〔3〕pp. 94—95参照.

判定すべきか否かについての方式を明らかにする．この節が本稿の主要な部分であるといえる．§ 4 で異常値と判定すべき標本が 2 個以上ある場合について議論する．一面からすれば § 4 の方式の特殊ケースが § 3 の方式といえる．が，§ 3 の方式とは異なる意味が見い出されることをみる．つまり，異常値の判定といわゆるモデルの構造変化テストとの関係について考察する．§ 4 の検定方式に構造変化検定の意味が生じるならば，周知の Chow の構造変化テスト[1]との関係を見なければならない．これらの考察を § 5 でする．§ 6 は補論である．

§ 1 本節で，まずわれわれが以下でとり扱う回帰モデルを定式化しておく．

単一回帰モデル

$$y_i=x_i\beta+u_i \quad (i=1,2,\cdots\cdots)$$

は k 個の説明変量のベクトル x_i で変量 y_i を説明する．β は $k\times 1$ の回帰係数ベクトルである．確率的攪乱項 u_i は，互いに独立に，

$$u_i\sim N(0,\sigma^2)$$

とする．

いま上のモデルを推定するために，T_1 個の標本が得られたとしよう．ここでそれら標本を，

$$y_1=\begin{pmatrix}y_1\\y_2\\\vdots\\y_{T_1}\end{pmatrix} \quad X_1=\begin{pmatrix}x_1\\x_2\\\vdots\\x_{T_1}\end{pmatrix} \quad x_i=(x_{1i},x_{2i},\cdots\cdots x_{ki})$$

とベクトル，行列表示するならば，推定のために，

$$y_1=X_1\beta+u_1$$

$$u_1\sim N(0,\sigma^2 I_{T_1})$$

という確率モデルを定式化することができる．あるいは，T_1 次元ベクトル y_1 について，

$$y_1\sim N(X_1\beta,\ \sigma^2 I_{T_1})$$

---

1) Chow〔1〕, Fisher〔2〕.

に関して β, σ^2 を推定する問題でもある.

周知のように，最小二乗推定法では，β の推定量 $\hat{\beta}$ は，

$$\hat{\beta}=(X_1'X_1)^{-1}X_1'y_1$$

となり，

$$\hat{y}_1=X_1\hat{\beta}$$

と定義して攪乱項 u_1 の推定量として

$$\hat{u}_1=y_1-\hat{y}_1$$

が得られる．また，σ^2 の推定量は，

$$\hat{\sigma}^2=\frac{\hat{u}_1'\hat{u}_1}{T_1-k}$$

であり，$\hat{\beta}$, $\hat{\sigma}^2$ はそれぞれ不偏推定量となるし，また，特定の条件のもとに，一致性やその他統計学的に望ましい性質があることもよく知られている.

さらに，推定量 $\hat{\beta}$, $\hat{u}_1$ に関して，これらが互いに独立に，

(1) $\hat{\beta}\sim N(\beta, \sigma^2(X_1'X_1)^{-1})$

(2) $\frac{\hat{u}_1'\hat{u}_1}{\sigma^2}\sim\chi^2(T_1-k)$

であることも周知に属する.

§ 2 この節で，§ 1 で考えた T_1 個の標本 $(y_1\ X_1)$ とは別に新たに 1 標本 $(y_0\ x_0)$ (y_0 はスカラー，x_0 は k 次行ベクトル) が得られ，これが先の標本と同種の構造から生起したかどうかを検定する方式を考える.

したがって，上述の検定を行なおうとするとき，その帰無仮説として

$$H_0: y_0=x_0\beta+u_0\ \sim N(x_0\ \beta,\ \sigma^2)$$

とすることは自然である．もし帰無仮説 H_0 が受容されるならば，追加された標本 $(y_0\ x_0)$ は先に得られた T_1 個の標本と同種の構造 (あるいは母集団) から生起したと判断される．また帰無仮説 H_0 が棄却されるならば，標本 $(y_0\ x_0)$ は $(y_1\ X_1)$ とは別な母集団から生起したと考えられる．換言すれば，(T_1+1) 個の標本

$$\begin{pmatrix} y_1 & X_1 \\ y_0 & x_0 \end{pmatrix}$$

全体を問題にしたとき，$(y_0\ x_0)$ は異常値であるといえる．

さて，帰無仮説 H_0 は別の表現をすれば，観測されないスカラーの確率変数 u_0 について，

$$H_0 : u_0 = y_0 - x_0\beta \sim N(0, \sigma^2)$$

とできる．

帰無仮説 H_0 の検定は以下のように定式化できる．T_1 個の標本による最小二乗推定量 $\hat{\beta}$ を用いて，

$$\hat{y}_0 = x_0\hat{\beta}$$

とすると，(1)より，

(3) $\hat{y}_0 \sim N(x_0\beta,\ \sigma_2 x_0(X_1'X_1)^{-1}x_0')$

である．また確率変数 u_0 の推定量 $\hat{u}_0$ を

$$\hat{u}_0 = y - \hat{y}_0$$

と定義すると，帰無仮説 H_0 が真であるとき，(3)を利用して，

(4) $\hat{u}_0 = y_0 - \hat{y}_0 \sim N(0,\ \sigma^2(1+x_0(X_1'X_1)^{-1}x_0'))$

が導出できる．したがって，

(4′) $$\frac{\hat{u}_0^2}{\sigma^2(1+x_0(X_1'X_1)^{-1}x_0')} \sim \chi^2(1)$$

が直ちに導ける．

一方，§ 1でも述べたように，$\hat{\beta}$ と $\hat{u}_1$ とは統計的に独立である．したがって u_0 と $\hat{\beta}$ とに依存する $\hat{u}_0$ は $\hat{u}_1$ と独立である．ゆえに，(4′)と(2)から，帰無仮説 H_0 が真のとき，

(5) $$\hat{F} = \frac{\dfrac{\hat{u}_0^2}{\sigma^2(1+x_0(X_1'X_1)^{-1}x_0')}}{\dfrac{\hat{u}_1'\hat{u}_1}{\sigma^2(T_1-k)}}$$

$$= \frac{\hat{u}_0^2(T_1-k)}{\hat{u}_1'\hat{u}_1(1+x_0(X_1'X_1)^{-1}x_0')} \sim F(1, T_1-k)$$

となる.[1] いま，§1における T_1 個の標本での σ^2 の推定量 $\hat{\sigma}^2$ や係数 $\hat{\beta}$ の分散・共分散行列 $V(\hat{\beta})$ の推定量 $\hat{V}(\hat{\beta})$ は，

$$\hat{\sigma}^2=\frac{\hat{u}_1'\hat{u}_1}{T_1-k}$$

$$\hat{V}(\hat{\beta})=\hat{\sigma}^2(X_1'X_1)^{-1}$$

であるから，$\hat{F}$ は，

$$(5')\quad \hat{F}=\frac{\hat{u}_0^2}{\hat{\sigma}^2(1+x_0(X_1'X_1)^{-1}x_0')}=\frac{\hat{u}_0^2}{\hat{\sigma}^2+x_0\hat{V}(\hat{\beta})x_0'}$$

と表現することも出来る.

帰無仮説 H_0 を有意水準 α で検定する場合，自由度 1，T_1-k の F 分布の α %点 F_α を得て

$$(6)\quad F_\alpha<\hat{F}$$

ならば H_0 を棄却することになる．つまり，(6)が成立するならば $(y_0\ x_0)$ は異常値であると判断することになる.

上の検定方式は本質的には区間予測の方法と同等であるといえる．一般に区間予測というのは，モデル推定のための標本 $(y_1\ X_1)$ とは別に，先決変数の k 次ベクトル x_0 のもとに y_0 を予測するということである．通常，予測区間の導出のために，(4)より

$$\frac{y_0-x_0\hat{\beta}}{\sqrt{\sigma^2(1+x_0(X_1'X_1)^{-1}x_0')}}\sim N(0,1)$$

を得て，上式と(2)より

$$\frac{y_0-x_0\hat{\beta}}{\sqrt{\hat{\sigma}^2(1+x_0(X_1'X_1)^{-1}x_0')}}\sim t(T_1-k)$$

---

1) (5)の代りに，

$$\frac{\hat{u}_0}{\sqrt{\hat{u}_1'\hat{u}_1(1+x_0(X_1'X_1)^{-1}x_0')/(T_1-k)}}\sim t(T_1-k)$$

としても同等である．これも，帰無仮説が真のとき(4)，(2)から導出できる．以下の考察参照.

を導出することができる．上式と(5)は同等であって，たとえば $1-\alpha$ の信頼係数による信頼区間は，自由度 T_1-k の t 分布の $\frac{\alpha}{2}\%$ 点 $t_{\frac{\alpha}{2}}$ を用いて，

$$y_0 : x_0\hat{\beta} \pm t_{\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2(1+x_0(X_1'X_1)^{-1}x_0')}$$

とできるのである．

以上の考察からいえることは，予測区間内に y_0 の標本が入る場合は，$(y_0\ x_0)$ は異常値とは見做されないし，逆にこの予測区間の外に落ちる場合は，標本 (y_0x_0) は異常値であると判断することになる．

§ 3 T_1 個の標本で，つまり $(y_1\ X_1)$ の資料よりモデルの推定を行ない，新たに1標本 $(y_0\ x_0)$ が得られ，これが先の標本 $(y_1\ X_1)$ と同一母集団から得られたか否かを検討する場合は§ 2の検定方法が応用できる．

しかし，本来T個 $(T=T_1+1)$ の標本

$$\begin{pmatrix} y_1 & X_1 \\ y_0 & x_0 \end{pmatrix}$$

よりモデル推定を行なっている場合には，$(y_0\ x_0)$ が他とは異なる母集団より生起した標本であるとか，この標本が異常値であるということは先験的には不明であろう．T個の標本でモデル推定を行ない，その結果によって資料の解析を行なう段階で，つまり，たとえば被説明変数と誤差項との比を検討したり，あるいは，標本散布図と推定方程式とのグラフ化によって $(y_0\ x_0)$ が異常値ではないかとの見当がつくものである．

この場合に§ 2の方法を応用しようとするとT個の標本で推定した結果，見当のつけた $(y_0\ x_0)$ を除いて T_1 個 $(T_1=T-1)$ の標本によって改めてモデル推定を行なう必要がある．

本節では，上述の方法ではなく，モデル推定の当初からT個の標本を用いて係数推定を行なう場合，その結果よりある特定の $(y_0\ x_0)$ について先の帰無仮説 H_0 を検定する方式を明らかにする．

いま，

$$X=\begin{pmatrix}X_1\\x_0\end{pmatrix},\quad y=\begin{pmatrix}y_1\\y_0\end{pmatrix},\quad u=\begin{pmatrix}u_1\\u_0\end{pmatrix}$$

と定義すれば，T個の標本を用いて最小二乗法による β の推定量を $\tilde{\beta}$，u の推定量を $\tilde{u}$ とすれば，

（7） $\tilde{\beta}=(X'X)^{-1}X'y$

（8） $\tilde{u}=\begin{pmatrix}\tilde{u}_1\\\tilde{u}_0\end{pmatrix}=y-X\tilde{\beta}=\begin{pmatrix}y_1-X_1\tilde{\beta}\\y_0-x_0\tilde{\beta}\end{pmatrix}$

である．

§1での $\hat{\beta}$，$\hat{u}_1$，§2での $\hat{u}_0$ 及び上の $\tilde{\beta}$，$\tilde{u}$ との間に以下の事実がある．

補助定理1 $\tilde{u}'\tilde{u}-\hat{u}_1'\hat{u}_1=\tilde{u}_0\hat{u}_0$

補助定理2 $\hat{u}_0=(1+x_0(X_1'X_1)^{-1}x_0')\tilde{u}_0$

これらを示すために，まず，（7）で $\tilde{\beta}$ を得る正規方程式を明示しておこう．

$$X'X\tilde{\beta}=X'y$$

がその正規方程式であるが，これは

$$(X_1'X_1+x_0'x_0)\tilde{\beta}=X_1'y_1+x_0'y_0$$

であり，また（1）の $\hat{\beta}$ を得る正規方程式は

$$X_1'X_1\hat{\beta}=X_1'y_1$$

であるから，結局，

（9） $(X_1'X_1+x_0'x_0)\tilde{\beta}=X_1'X_1\hat{\beta}+x_0'y_0$

が成立する．

さて，（9）を利用して，

$$\begin{aligned}\tilde{u}'\tilde{u}-\hat{u}_1'\hat{u}_1&=y'y-\tilde{\beta}'X'X\tilde{\beta}-y_1'y_1+\hat{\beta}'X_1'X_1\hat{\beta}\\&=y_0^2-\tilde{\beta}'(X_1'X_1+x_0'x_0)\tilde{\beta}+\hat{\beta}'X_1'X_1\hat{\beta}\\&=y_0^2-\tilde{\beta}(X_1'X_1\hat{\beta}+x_0'y_0)+\hat{\beta}'(X_1'X_1\tilde{\beta}+x_0'x_0\tilde{\beta}-x_0'y_0)\\&=y_0^2-\tilde{\beta}'x_0'y_0+\hat{\beta}'x_0'x_0\tilde{\beta}-\hat{\beta}'x_0'y_0\\&=(y_0-x_0\hat{\beta})(y_0-x_0\hat{\beta})=\tilde{u}_0\hat{u}_0\end{aligned}$$

が示されるから補助定理1が証明された．

また，(9)より，左より $(X_1'X_1)^{-1}$ をかけて，

$$\tilde{\beta}+(X_1'X_1)^{-1}x_0'x_0\tilde{\beta}=\hat{\beta}+(X_1'X_1)^{-1}x_0'y_0$$

が得られる．さらに上式に左より x_0 をかけて，

$$x_0\tilde{\beta}-x_0\hat{\beta}=x_0(X_1'X_1)^{-1}x_0'(y_0-x_0\tilde{\beta})$$

である．一方，

$$\hat{u}_0-\tilde{u}_0=y_0-x_0\hat{\beta}-(y_0-x_0\tilde{\beta})=x_0\tilde{\beta}-x_0\hat{\beta}$$

だから

$$\hat{u}_0-\tilde{u}_0=x_0(X_1'X_1)^{-1}x_0'\tilde{u}_0$$

が得られ，これより補助定理2が証明された．

補助定理1，2より(5)の $\hat{F}$ は，T個の標本による推定量 $\tilde{u}$ で表現でき，これを $\tilde{F}$ と記すと，

$$\hat{F}=\frac{(T_1-k)\tilde{u}_0^2}{\dfrac{\tilde{u}'\tilde{u}}{1+x_0(X_1'X_1)^{-1}x_0'}-\tilde{u}_0^2}=\tilde{F}$$

と変形される．したがって帰無仮説 H_0 の検定統計量として，上の形式を用いることが出来る．しかし，ここで不都合なことは，われわれはいまT個の標本でモデル推定をしたケースを想定しているので，その際に $(X'X)^{-1}$ は導出しているが，$(X_1'X_1)^{-1}$ の情報は得られていないと思われる．したがって上の検定統計量では改めて $(X_1'X_1)^{-1}$ の計算を要する．この不都合さは次の補助定理によって解決されるであろう．

補助定理3 $\dfrac{1}{1+x_0(X_1'X_1)^{-1}x_0'}=1-x_0(X'X)^{-1}x_0'$

これを示すために，まず，

$$I=(X'X)(X'X)^{-1}$$

において，$X'X=X_1'X_1+x_0'x_0$ を代入する．ゆえに，

$$I-X_1'X_1(X'X)^{-1}-x_0'x_0(X'X)^{-1}=0$$

が得られ，上式の左から $x_0(X_1'X_1)^{-1}$ をかけ，右から x_0' をかけて

$$x_0(X_1'X_1)^{-1}x_0'-x_0(X'X)^{-1}x_0'-x_0(X_1'X_1)^{-1}x_0'x_0(X'X)^{-1}x_0'=0$$

が導けこれより，

$$(1+x_0(X_1'X_1)^{-1}x_0')(1-x_0(X'X)^{-1}x_0')=1$$

となり，補助定理3が証明された．したがって，補助定理3を援用して先の $\tilde{F}$ は

$$(10)\quad \tilde{F}=\frac{(T_1-k)\tilde{u}_0^2}{(1-x_0(X'X)^{-1}x_0')\tilde{u}'\tilde{u}-\tilde{u}_0^2}$$

となる．

いま，T個の標本による β の推定量を $\tilde{\beta}$，u の推定量を $\tilde{u}$ としているが，σ^2 の推定量 $\tilde{\sigma}^2$ 及び $\tilde{\beta}$ の分散・共分散行列 $V(\tilde{\beta})$ の推定量を $\tilde{V}(\tilde{\beta})$ と表現すると

$$\tilde{\sigma}^2=\frac{\tilde{u}'\tilde{u}}{T-k}$$

$$\tilde{V}(\tilde{\beta})=\tilde{\sigma}^2(X'X)^{-1}$$

である．ゆえに，これらを用いて，次のようにまとめることができる．

定　理　帰無仮説 H_0 のもとに，

$$\tilde{F}=\frac{(T_1-k)\tilde{u}_0^2}{(1-x_0(X'X)^{-1}x_0')\tilde{u}'\tilde{u}-\tilde{u}_0^2}$$

$$=\frac{(T-k-1)\tilde{u}_0^2}{(T-k)(\tilde{\sigma}^2-x_0\tilde{V}(\tilde{\beta})x_0')-\tilde{u}_0^2}\sim F(1,T-k-1)$$

以上で示されたことを以下のように要約できる．一般に，T個の標本を用いてモデル推定を行なう．したがって，この推定から回帰係数ベクトル β の推定量 $\tilde{\beta}$ と，σ^2 の推定量 $\tilde{\sigma}^2$，及び係数推定量 $\tilde{\beta}$ の分散・共分散行列の推定量 $\tilde{V}(\tilde{\beta})$ の導出が可能である．モデル推定の結果，ある特定の標本（y_0 x_0）が異常値であるのかどうかを検定しようとすると，定理に述べた統計量 $\tilde{F}$ が計算可能となる．いま，自由度1，T−k−1 の F 分布の α %点を F_α とするなら

$$F_\alpha\leq\tilde{F}$$

であれば，有意水準 α で帰無仮説 H_0 が棄却される．つまり標本（y_0 x_0）は他

の標本とは異なる母集団から生起したと判断され，これは異常値であると見做し得るのである．反対に，

$$\tilde{F}<F_{\alpha}$$

ならば，$(y_0\ x_0)$ は異常値であるとは判断できないといえる．

§ 4 本節では，異常値かどうかを検定したいと思う標本の数が § 3 とは異なって2個以上ある場合について，先の検定方式を拡張することが可能なことを示そう．

まず T_1 個の標本 $(y_1\ X_1)$ によってモデル推定が可能であることは § 1 と全く同じである．これら T_1 個の標本とは別に $(y_2\ X_2)$ という T_2 個の標本が得られたとする．そうして，帰無仮説として，

$$H_0 : y_2 \sim N(X_2\beta,\ \sigma^2 I_{T_2})$$

を検定する方式を考える．この場合，§ 2 の議論と同様にして

$$\hat{y}_2 = X_2\hat{\beta} \sim N(X_2\beta_2,\ \sigma^2 X_2(X_1'X_1)^{-1}X_2')$$

が得られ，H_0 が真であるとき，

$$\hat{u}_2 = y_2 - \hat{y}_2 \sim N(0,\ \sigma^2(I+X_2(X_1'X_1)^{-1}X_2'))$$

が導ける．したがって

$$\frac{1}{\sigma^2}\hat{u}_2'[I+X_2(X_1'X_1)^{-1}X_2']^{-1}\hat{u}_2 \sim \chi^2(T_2)$$

が成り立つ．

結局，T_1 個の標本でモデル推定を行ない，新たに得られた T_2 個の標本について帰無仮説 H_0 を検定する方式として，§ 2 と同様に，

$$(11)\quad \hat{F} = \frac{\tilde{u}_2'[I+X_2(X_1'X_1)^{-1}X_2']^{-1}\hat{u}_2/T_2}{\hat{u}_1'\hat{u}_1/(T_1-k)} \sim F(T_2,\ T_1-k)$$

が帰無仮説 H_0 の真なるときに成立するから，(6) を H_0 の棄却域にとることができる．

次に，§ 3 と同様に，T個 $(T=T_1+T_2)$ の標本全体

$$\begin{pmatrix} y_1 & X_1 \\ y_2 & X_2 \end{pmatrix}$$

でモデル推定をして，その結果より H_0 を検定する場合を考える．

$$y=\begin{pmatrix}y_1\\y_2\end{pmatrix}\quad X=\begin{pmatrix}X_1\\X_2\end{pmatrix}$$

として，

$$\tilde{\beta}=(X'X)^{-1}X'y,$$

$$\tilde{u}=\begin{pmatrix}\tilde{u}_1\\\tilde{u}_2\end{pmatrix}=\begin{pmatrix}y_1-X_1\tilde{\beta}\\y_2-X_2\tilde{\beta}\end{pmatrix}$$

とすることは§ 3と同様である．また，§ 3の補助定理1，2に対応して，

補助定理 1′ $\tilde{u}'\tilde{u}-\hat{u}_1'\hat{u}_1=\hat{u}_2'\tilde{u}_2$

補助定理 2′ $\hat{u}_2=[I+X_2(X_1'X_1)^{-1}X_2']\tilde{u}_2$

がほぼ同様に証明することができる．もちろんこれらの証明に関して，(9)と同様の関係

(12) $X'X\tilde{\beta}=(X_1'X_1+X_2'X_2)\tilde{\beta}=X_1'X_1\hat{\beta}+X_2'y_2$

を利用する．

さらに，補助定理3と同様の関係

補助定理 3′

$$[I+X_2(X_1'X_1)^{-1}X_2']^{-1}=I-X_2(X'X)^{-1}X_2'$$

が証明できる．これは以下のように示される．

$$I=(X'X)(X'X)^{-1}$$

において，$X'X=X_1'X_1+X_2'X_2$ を代入し，

$$I-X_1'X_1(X'X)^{-1}-X_2'X_2(X'X)^{-1}=O$$

を得て上式の左から $X_2(X_1'X_1)^{-1}$ を，右から X_2' をかけて，さらに両辺に I を加えるならば，

$$(I+X_2(X_1'X_1)^{-1}X_2')(I-X_2(X'X)^{-1}X_2')=I$$

が導出できる．

以上の補助定理1′, 2′, 3′を利用して，(11)の $\hat{F}$ を変形し，いまそれを $\tilde{F}$ と表わせば，帰無仮説 H_0 のもとに，

$$(13)\quad \tilde{F}=\frac{\tilde{u}_2'[I-X_2(X'X)^{-1}X_2']^{-1}\tilde{u}_2/T_2}{(\tilde{u}'\tilde{u}-\tilde{u}_2'[I-X_2(X'X)^{-1}X_2']^{-1}\tilde{u}_2)/T_1-k}\sim F(T_2, T_1-k)$$

が明らかとなる.

上で示されたことは，T個の標本全体でモデル推定をした結果，それら標本のうち T_2 個の標本 $(y_2\ X_2)$ について，帰無仮説 H_0 を検定しようとするとき，(13)の $\tilde{F}$ をその検定統計量とすればよい.

本節での検定すべき帰無仮説 H_0 は§ 3での帰無仮説の拡張と考えられる. すなわち，本節の議論で $T_2=1$ としたケースを§ 3でとり扱ったことになる. しかし，形式的には§ 4は§ 3の拡張であるといえるが，§ 4は§ 3と異なった意味が生じてくる. つまり§ 4の検定方式は異常値の検出としての意味よりも，むしろ，いわゆる構造変化テストとしての意味がある. 本節の帰無仮説 H_0 では，検定すべき T_2 個の標本$(y_2\ X_2)$を一括して，それらが他の T_1 個の標本 $(y_1\ X_1)$ の生起した構造（母集団）と同一の構造から生起したかを検定することになる.したがってこの帰無仮説の中に，T_2 個の標本 $(y_2\ X_2)$ は同一の構造を前提していることになる. つまり，$(y_2\ X_2)$という T_2 個の標本のそれぞれは同一の構造から生起したと想定している. このことから本節の検定方式は異常値の検出というより構造変化テストとしての意味が強くなる. というのは，複数個の異常値である標本が本来同一の構造をもっていると前提するのはそれほど意味がない. 検定しようとしている標本 $(y_2\ X_2)$ はその T_2 個の標本間には異種性がなく同一構造から生起するが，別の T_1 個の標本 $(y_1\ X_1)$ の生起した構造とは異なるかもしれないということを想定している. このことは，T_1 個の標本 $(y_1\ X_1)$ と T_2 個の標本 $(y_2\ X_2)$ との間で，モデルの構造上の違いを検定することになるのである.

§ 5 前節で示した検定方式が構造変化テストとしての意味で理解できるならば，構造変化テストとして周知の Chow 検定との関係を明らかにしなければならない.

本節では，先の検定方式が，Chow の構造変化テストで，検定すべき二つの

構造で一方の構造を推定するには標本が不足している場合（われわれの用いている記号では $k>T_2$ のケース）と同等であることを示そう．

$k>T_2$ の場合，Chow の構造変化テストは以下のように要約できる．

T_1 個の標本（y_1X_1）は，

$$y_1=X_1\beta_1+u_1,\quad u_1\sim N(0,\sigma^2 I_{T_1})$$

から得られ，T_2 個の標本（y_2X_2）は，

$$y_2=X_2\beta_2+u_2,\quad u_2\sim N(0,\sigma^2 I_{T_2})$$

から生起したものとする．そうして，帰無仮説

$$H_C:\beta_1=\beta_2$$

を検定する方式を考える．この検定のための統計量は，

$$(14)\quad F_C=\frac{[(X_1\tilde{\beta}-X_1\tilde{\beta})'(X_1\tilde{\beta}-X_1\hat{\beta})+(y_2-X_2\tilde{\beta})'(y_2-X_2\tilde{\beta})]/T_2}{\hat{u}_1'\hat{u}_1/(T_1-k)}$$

である．ここに $\tilde{\beta}$ は T 個（$T=T_1+T_2$）の標本による係数推定量，$\hat{\beta}$ は T_1 個のみの標本による係数推定量，$\hat{u}_1$ は T_1 個のみの標本でモデル推定をしたときの u_1 の推定量である．そうして F_C が，帰無仮説 H_C のもとに，

$$F_C\sim F(T_2,T_1-k)$$

であることが示されるので，これを検定統計量とできるのである．[1)]

さて，以下で(14)の F_C と § 4 に示した(11)の $\hat{F}$ が同値であることを証明する．

F_C と $\hat{F}$ の分母は同一であるから，分子が同値であることを示せば十分である．一方，

$$\tilde{u}_2=y_2-X_2\tilde{\beta}$$

とできるから，結局

$$(15)\quad (\tilde{\beta}-\hat{\beta})'X_1'X_1(\tilde{\beta}-\hat{\beta})+\tilde{u}_2'\tilde{u}_2=\hat{u}_2'[I+X_2(X_1'X_1)^{-1}X_2']^{-1}\hat{u}_2$$

を証明すればよい．§ 4 の(12)より

---

1) Chow テストに関して証明しなければならない数学的事実は，最近では，Fisher〔2〕の証明があげられる．

$$X_1'X_1(\tilde{\beta}-\hat{\beta})=X_2'(y_2-X_2\tilde{\beta})=X_2'\tilde{u}_2$$

$$\therefore\quad (\tilde{\beta}-\hat{\beta})=(X_1'X_1)^{-1}X_2'\tilde{u}_2$$

が得られる．これを(15)の左辺に代入するならば，

$$(\tilde{\beta}-\hat{\beta})'X_1'X_1(\tilde{\beta}-\hat{\beta})+\tilde{u}_2'\tilde{u}_2=\tilde{u}_2'[I+X_2(X_1'X_1)^{-1}X_2']\tilde{u}_2$$

さらに上式右辺に補助定理2′を適用すれば，(15)の右辺に等しくなる．ゆえに

(16)　$F_c=\hat{F}$

が証明された．先述のように $\hat{F}$ と(13)の $\tilde{F}$ の同値性は示されているから，結局，先の帰無仮説 H_0 の検定方式と，$k>T_2$ の場合の Chow テストとの同等性が示されたことになる．

さて，以上の考察の結果，Chow テストあるいは§2の検定方式と，§3の検定方式との相違を考えてみよう．

いずれの検定方式も，その検定統計量，つまり F_c あるいは $\hat{F}$ と $\tilde{F}$ の同値性が以上で示されているので，統計学的には本質的に同等である．しかし，その統計量の定式化から考えられる適用範囲には相違がある．

Chow テストの F_c，あるいは§2の検定方式による $\hat{F}$ をその検定統計量とする場合，異常値となる標本がどれであるかをモデル推定に先立って判明しているケースを想定している．つまり，T個の標本全体のうち，検定しようとする標本を先験的に決めて，その上でモデル推定を行なう場合を考慮している．したがって，T個ある標本の中で，特定の標本について検定すべき何らかの情報が得られていなければならない．この意味で，Chow テストはまさに，構造変化検定に利用され得る．というのは，構造変化テストを要するとき，検定に先立って，モデル設定の段階でその構造変化に関する仮説が経済的観点等から存在するのが普通であって，統計的検定は，その仮説の検証として有意味であるからである．

一方，§3の検定方式による $\tilde{F}$ をその検定統計量とする場合，モデル推定に先達って，検定すべき標本がどれであるかという情報は全く不要である．$\tilde{F}$ の定式化からわかるように異常値となるかもしれない標本について先験的に何ら

情報は要らない．モデル推定の結果§3にも述べたように，被説明変数の期待値の推定量や誤差項の分析を通じて検定すべき標本がどれかについての情報が得られる．そうして，その標本が異常値であると判定する良否を検定することになる．したがって Chow テストの前提とする特定標本についての経済的観点等からの仮説なり情報というものを，§3の検定方式による検定の場合には統計的帰無仮説が棄却された段階で考慮しなければならないものとなるのである．

§6 統計学の推論の中で，いわゆる平均値の差の検定について，その方式が確立されている．つまり一次元確率変数 X_i $(i=1,\cdots\cdots,N_x)$，および Y_i $(i=1,\cdots\cdots,N_y)$ は $N(\mu_x,\ \sigma^2)$，$N(\mu_y,\ \sigma^2)$ に従って分布し，その実現値が x_i $(i=1,\cdots\cdots,N_x)$，y_i $(i=1,\cdots\cdots,N_y)$ と得られたとき，帰無仮説 $\mu_x=\mu_y$ を検定する方法である．この検定方式の回帰方程式への応用が，Chow の構造変化テストであるといえる．形式的に言えば，回帰方程式で，実質的な説明変数がなく，被説明変数が定数項のみで説明されるような回帰分析をするなら，平均値の差の検定方式と Chow の構造変化テストとは同等になる．

これと同様な関係として，われわれが考察した異常値の検出方法と，統計学の中で，いわゆる Thompson の棄却検定法とを見ることが出来る．このことを補論的にこの§で考察する．

いわゆる Thompson の棄却検定法というのは次のように要約できる．
確率変数 Y_i $(i=1,\cdots\cdots,n)$ 及び Y_0 は互いに独立に未知の分散 σ^2 をもつ正規分布にしたがい

$$E(Y_i)=\mu \quad (i=1,\cdots\cdots,n)$$
$$E(Y_0)=\mu_0$$

であるとする．Y_i $(i=1,\cdots\cdots,n)$ 及び Y_0 の実現値が y_i $(i=1,\cdots\cdots,n)$ 及び y_0 と得られたとき，帰無仮説 $H_0:\mu=\mu_0$ を検定する場合，

$$\hat{t}=\sqrt{\frac{n-1}{n+1}}\cdot\frac{y_0-\bar{y}}{S}$$

を検定統計量とできる．ここで，

$$\bar{y}=\frac{1}{n}\Sigma y_i,\quad S^2=\frac{1}{n}\Sigma(y_i-\bar{y})^2$$

である[1]．というのは $\hat{t}$ が自由度 $n-1$ の t 分布をすることが証明されるからである．したがって，自由度 $n-1$ の t 分布の $\frac{\alpha}{2}$ %点を $t_{\frac{\alpha}{2}}$ とするなら，

$$|\hat{t}|\geq t_{\frac{\alpha}{2}}$$

が有意水準 α の棄却域とできる．

以上の棄却検定方式をわれわれが考察した異常値検出方式の特殊ケースで扱えることを示そう．

§ 1で，係数ベクトル β を，一次元であるとし，つまり β をスカラーとみなし，同様に，行列 X_1 をその要素が1であるベクトルとみなすならば，(1)，(2)は

$$\hat{\beta}=\bar{y}$$

$$\hat{u}_1'\hat{u}_1=T_1\cdot S^2=nS^2$$

である．ここで $n=T_1$ としている．さらに，§ 2において，考察中のケースでは，

$$\hat{u}_0=y_0-\bar{y},\quad (X_1'X_1)^{-1}=\frac{1}{n}$$

と出来るから，(5)に対応して，

$$\hat{F}=\frac{(y_0-\bar{y})^2(n-1)}{nS^2\left(1+\frac{1}{n}\right)}=\frac{(n-1)(y_0-\bar{y})^2}{S^2(n+1)}$$

が自由度 $1, n-1$ の F 分布をするといえる．これと先述の $\hat{t}$ が自由度 $n-1$ の t 分布をするということと同等である[2]．

さらに，

---

1) $\bar{y}$ は標本平均，S^2 は標本分散である．

2) p. 5 の脚註参照．

$$\tilde{y}=\frac{1}{n+1}\sum_{i=0}^{n} y_i$$

$$\tilde{S}^2=\frac{1}{n+1}\sum_{i=0}^{n}(y_i-\tilde{y})^2$$

とするならば，Thompson の棄却検定方式は

$$\tilde{F}=\frac{(n-1)(y_0-\tilde{y})^2}{\left(1-\frac{1}{n+1}\right)(n+1)\tilde{S}^2-(y_0-\tilde{y})^2}$$

$$=\frac{(n-1)(y_0-\tilde{y})^2}{n\tilde{S}^2-(y_0-\tilde{y})^2}$$

をその検定統計量とできる．これは§ 3の定理を援用して，$\tilde{F}$が自由度$1, n-1$のF分布をすることが明らかであるからである．

参 考 文 献

〔1〕 Chow, G. C., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," Econometrica, Vol. 28, 1960, pp. 591-605.

〔2〕 Fisher, F. M., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions : An Expository Note," Econometrica, Vol. 38, 1970, pp. 391-66.

〔3〕 佐和隆光，「計量経済学の基礎」，昭和45年7月，東洋経済新報社.

〔4〕 Thompson, W. R., "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of Deviation to sample Standard deviation," Annales of Mathematical Statistics, Vol. 6., 1940