回帰分析における変数選択の問題

——問題の所在と性質——

1. 分析の目的

回帰分析は，各種の統計的手法の中でも，実際に応用されることがもっとも多いが，それだけに現実の場においては，必ずしも数理統計の論理によって割り切ることのできない，いろいろな問題が生ずることが多い．そのなかでも，もっとも重要な問題が，説明変数選択の問題であるといえよう．

これについては他の諸論文でも述べられているように，いくつかの手法や考え方が提案されている．しかしながら，問題の性質上これについて特定の最適な手続きなどというものは存在しえないのであって，現実には形式的な手法と，理論的，経験的な知識とを組み合わせて，常識によって判断を下すほかはない．パッケージ化された「変数選択プログラム」などに頼ることは危険であるといわねばならない．しかしながら逆にアメリカなどでも一部の「データ解析派」の人々がやっているように，計算結果の直観的な「もっともらしさ」のみを手がかりにして，確率計算にもとづく理論的基礎づけを無視するのも正しくない．そういうやり方をとると結局自分のもっている先入感を正当化するだけに終わってしまい，データが本当に示していることを見落してしまう危険性がある．

変数選択の問題は，より一般的にはモデル選択の問題の一種であると考えることができる．すなわち与えられたデータに対して，いろいろな確率モデルが考えられる場合，どれがいちばん適当であるかを定める問題の特別の場合と考えることができる．この場合モデルを包括的なものとすればするほど，それがデータについてのいわば「真の構造」に近いものを含むようになることは確かであるが，他方より包括的なモデルはそれだけ多くの未知母数を含むことになり，データからそれらの値を推定するときの推定誤差がそれだけ大きくなるので，「推定されたモデル」はかえって真の構造からかけ離れたものとなってしまう可能性がある．データに対する「あてはまりのよさ」が同程度ならば，より単純なモデルのほうが望ましいことは，直観的にも自明であろう．そこでモデルは現実のデータの構造を近似するうえでの精密さと，単純さとのバランスのうえで定められねばならない．赤池弘次氏はこの問題に対する一つの手法として，情報量基準とよばれるものを提案された．それはモデルに含まれる尤度関数を$L(\theta_1\cdots\theta_p)$（$\theta_j$は実母数）とするとき，

$$AIC=2\max_{\theta_1\cdots\theta_p}\log L(\theta_1\cdots\theta_p)-2p$$

という値を基準として，この値を最大にするようなモデルを最適なモデルとして選択しようという考え方である注1)．ここで$2\max\log L$の値はモデルのあてはまりのよさをあらわし，$-2p$は母数の数を増すことに対するペナルティーを意味すると考えられる，AICの意味やその性質についての議論にはここではこれ以上立ち入らないが，それはいろいろな複雑な構造をもった問題に対して少なくとも一つの便利な手法として役立ち，かつ多くの場合統計的に好ましい性質を示すことが知られている．

注1)　このいわゆる「赤池情報量」については雑誌「数理科学」の特集「情報量基準」を参照

回帰分析の場合についていえば，一般に説明変数の数を増せば，重相関係数(決定係数)あるいは残差平方和を基準とする「あてはまり」はよくなる．しかしながら説明変数の数をむやみに増すと各変数の回帰係数の推定値が不安定になり，推定誤差が増大する．説明変数の間にいわゆる多重共線関係 multi-collearty が生じて，回帰係数の推定値が一見不合理な値になることが起りやすい．

このような問題を処理するときの困難の一つは「真の構造」は知ることができないという想定から出発しなければならないことである．すなわち考えられるいくつかのモデルのうち，一つが「正しいモデル」であって，他は間違ったものであるという想定はこのような場合には非現実的であって，すべてのモデルはいずれも完全に正しくはないものと想定しなければならない．ただその中であるものは充分な近似として役立ち得るものと考えられるので，近似としてもっともよいものはどれであるかを知ることが問題となるのである．このような問題については，古典的な仮説検定や多重決定の理論は直接には適用できない．それらは「正しいモデル」が存在するという前提で論じられているからである．またモデルが現実の構造に対する一つの近似をあらわすものであるにすぎない以上，そこに含まれている母数も，現実の構造の中に含まれている数量を直接表現していると考えることはできない．したがって推定量のよさについても，その「真値」からの誤差の大きさを基準とする伝統的な理論は意味を失ってしまう．

モデル選択の問題，それについてのいろいろな手法を比較吟味するには，古典的な数理統計学の推測理論では不充分である．すなわち単に「正しいモデル」のもとでの特定の手法の統計的性質のみでなく，「真の構造」を必ずしも表現しないモデルをあたかも「正しいモデル」であるかのように考えて議論を進めたとき，どういうことが起こるかを追求しなければならない．そのためにはデータからの推測という，狭い意味の「データ解析」の枠を越えて，「モデル」を前提として得られた結論が，具体的にどのような目的にどのように利用されるかを考えなければならない．そうしてその目的との関連において，モデルの適切性とは何であるか，それをどのようにして測るかを論じなければならないのである．

回帰分析の場合には，データは管理された実験のもとで得られたものではないのがほとんどであるから，モデルが現実の構造に対する近似にすぎないことは，実は最初から明白であるといってよい．このような場合，モデルを前提として推定された回帰式は，もしそれを特定の目的に利用しようとするのでなければ，実は一定の仮説的想定のもとでの試算結果という以上の意味をもたない．

したがってそれは統計的推測の種々の形式に従って行なわれた計算の結果であっても，単なる「あてはめ」という記述的意味しかない．もちろんこのような計算も無意味ではないのは，すべての記述統計的手続きが場合によって重要な意味をもつのと同様である．しかしその場合にはモデル選択はそもそも重要な問題にはならない．

単なる記述を離れて，回帰分析の結果の利用目的には，1) 構造分析　2) 予測　3) 制御 の3種類が考えられる．1) 構造分析とは，回帰分析の結果を用いて，当面のデータにおいて観測された対象より広い，より一般的な対象の構造についてなんらかの判断を下そうとするものである．2)予測は，被説明変数(従属変数)の未来の値について説明変数(独立変数)の値を前提として判断を下すものである．3) 制御とは 被説明変数の値を望ましい水準に保つように，独立変数の値を定めることである．これらの目的には，いずれも一定のモデルを前提とした計算結果が用いられるが，同時にこれらの目的との関連において「正しくないモデル」を用いたとき，どのような危険が生ずるかを考えることができる．

上記の三つの目的は，一応それぞれ別個の種類のものであるが，なかでも 2) 予測 がもっとも基

本的であると考えることができる．構造分析は，より一般的な条件のもとにおける被説明変数の変化を予測する問題，また制御は，独立変数の異なる水準に対応する従属変数の値を予測して，その中でもっとも望ましい場合を選択する問題と考えられるからである．そこで以下においては予測を中心として問題を論じよう．

2. 平均予測2乗誤差の基準

つぎに回帰分析の問題についてより具体的に考えよう．Yを被説明変数，$x_1, x_2, \cdots$を考えられる説明変数とし，それらについてのデータを Y_i, x_{1i}, $x_{2i}, \cdots$, $i=1, 2, \cdots, n$としよう．

これについて一つの「モデル」M_αを

$$Y_i=\beta_0^\alpha+\beta_1^\alpha x_{\alpha 1i}+\cdots+\beta_{p_\alpha}^\alpha x_{\alpha p i}+u_i \quad (1)$$
$$i=1, 2, \cdots, n$$

とあらわそう，ここで $x_{\alpha 1}, \cdots, x_{\alpha p}$ が，このモデルにとり上げられた説明変数である．このモデルはつぎのような想定を意味する．

$$E(Y|x_1, x_2, \cdots)=\beta_0^\alpha+\beta_1^\alpha x_{\alpha 1}+\cdots+\beta^\alpha{}_{p_\alpha} x_{\alpha p}$$

ここでさらに，

i) u_i は互いに独立

ii) $E(u_i^2)=\sigma^2$

iii) u_i は正規分布に従う

ことを仮定すればふつうの最小2乗法による推定・検定の手法が応用できて，最小2乗推定量がもっともよい推定量になることはいうまでもない．

ところで「真の構造」のもとでのYの条件付期待値を，

$$E(Y|x_1, x_2, \cdots)=\eta(x_1, x_2, \cdots)$$

とし，$\eta_i=\eta(x_{1i}, x_{2i}, \cdots)$とあらわせば，モデル$M_\alpha$の「偏り」をつぎのように定義できる．すなわち β_j^α および ξ_i^α を，

$$\xi_i^\alpha=\eta_i-\beta_0^\alpha-\beta_1^\alpha x_{\alpha 1i}-\cdots-\beta_p^{\alpha_p} x_{\alpha p i}$$
$$i=1, \cdots\cdots, n \quad (2)$$

かつ $\sum_i \xi_i^\alpha=0 \quad \sum_i \xi_i^\alpha x_{\alpha 1i}=\cdots=\sum \xi_i^\alpha x_{\alpha p i}=0$

を満たすように定めれば，

$$Y_i=\beta_0^\alpha+\beta_1^\alpha x_{\alpha 1i}+\cdots+\beta_p^\alpha x_{\alpha p i}+\xi_i^\alpha+v_i^\alpha$$

とあらわされ，

$$E(v_i^\alpha|x_1, x_2, \cdots)=0$$

また β_j^α の最小2乗推定量を $\hat{\beta}_j^\alpha$ とすれば，

$$E(\hat{\beta}_j^\alpha)=0$$

となる．したがって ξ_i^α がモデルの偏りをあらわす項と考えることができる．

そこでさらに，

$$E(v_i^{\alpha 2})=\tau^2 \qquad i=1, 2, \cdots, n$$
$$E(v_i^\alpha v_{i'}^\alpha)=0 \qquad i \neq i'$$

と仮定すれば，最小2乗推定量の分散共分散は，

$$E\{(\hat{\beta}_j^\alpha-\beta_j^\alpha)(\hat{\beta}_k^\alpha-\beta_k^\alpha)\}=m^{jk}\tau^2$$

という形になり，m^{jk} は $x_{\alpha 1}, \cdots, x_{\alpha p}$ のモーメント行列の逆行列の要素となる．

モデル M_α のもとでの σ^2 の推定量の分散を，

$$\hat{\sigma}_\alpha{}^2=\sum(Y_i-\hat{Y}_i^\alpha)^2/(n-p-1)=Q_\alpha/(n-p-1)$$

ただし $\hat{Y}_i^\alpha=\hat{\beta}_0^\alpha+\hat{\beta}_1^\alpha x_{\alpha 1i}+\cdots+\hat{\beta}_p x_{\alpha p i}$

とおけば，

$$E(Q_\alpha)=\sum_{i=1}^{n} \xi_i^{\alpha 2}+(n-p-1)\tau^2$$

となるから $\lambda_\alpha=\sum \xi_i^{\alpha 2}$ とおけば，

$$E(\hat{\sigma}_\alpha{}^2)=\lambda_\alpha/(n-p-1)+\tau^2$$

となり $\hat{\sigma}_\alpha{}^2$ は τ^2 の過大な推定量，したがって $\hat{\beta}_j^\alpha$ の分散の推定量 $m^{jj}\hat{\sigma}_\alpha{}^2$ も，その過大な推定量になる．しかしこのことは，正しくないモデル M_α のもとでの計算が「あてはまりのよさ」を過小に評価しているということを意味するものではない．そもそも母数 β_j^α, $j=1, \cdots, p_\alpha$ は与えられたモデル M_α に対応して条件式(2)から定められたものであるから，それ自体は実はモデル M_α のもとでの最小2乗推定量の期待値という以上の意味をもつものではない．したがってその分散が小さいこともとくに有利な点とはならない．

モデル M_α を採用するということは，実は x_1, x_2, … の値が与えられたときの Y の値を，

$$\hat{Y}_0=\hat{\beta}_0^\alpha+\hat{\beta}_1^\alpha x_{\alpha 10}+\cdots+\hat{\beta}_p^\alpha x_{\alpha p 0} \quad (3)$$

という形で予測することを意味すると解釈することができる．いま，

$$\xi_0^\alpha=\eta(x_{10}, x_{20}, \cdots)-\beta_0^\alpha-\beta_1^\alpha x_{\alpha 10}-\cdots-\beta_p^\alpha x_{\alpha p 0}$$

と定義すれば，予測の偏りは，

$$E(Y_0-\hat{Y}_0)=\xi_0^{\alpha}$$

その分散は，

$$V(\hat{Y}_0)=(\sum_j\sum_{k=0}^{p_\alpha}m^{jk}x_{\alpha j0}x_{\alpha k0})\tau^2$$

（ただし $x_{\alpha 0}\equiv 1$ とする）

$$=c_0^{\alpha 2}\tau^2$$

となるから，予測の平均2乗誤差は，

$$E(Y_0-\hat{Y}_0)^2=(1+c_0^{\alpha 2})\tau^2+\xi_0^{\alpha 2}$$

となる．考えられるいくつかのモデルの中で，この値を小さくするようなものがもっともよいと考えられる．

しかしながら一般には ξ_0^{α} の値は x_{10}，x_{20}，… の値に応じて変化するが，それは未知の関数 η によって定められるから，それを一般に求めることはできない．しかしつぎのような場合にはそれを推定することができる．

データにおいて与えられたのと同じ n 組の値 x_{1i}，x_{2i}，$\cdots(i=1,\cdots,n)$ を考え，これらの値に対応する，データの値とは独立な Y の値 $Y_i^0(i=1,\cdots,n)$ を予測することを考える．そうすると，

$$E(Y_i^0)=\beta_0^{\alpha}+\beta_1^{\alpha}x_{\alpha 1i}+\cdots+\beta_{p\alpha}^{\alpha}x_{\alpha pi}+\xi_i^{\alpha}$$

であるから，

$$E(Y_i^0-\hat{Y}_i)^2=(1+\sum\sum m^{jk}x_{\alpha ji}x_{\alpha ki})\tau^2+\xi_i^{\alpha 2}$$

$$=(1+c_i^{\alpha 2})\tau^2+\xi_i^{\alpha 2}$$

という形になる．したがって n 個の値の平均2乗誤差の和は，

$$\sum_i E(Y_i^0-\hat{Y}_i)^2=\sum_i(1+c_i^{\alpha 2})\tau^2+\sum_i\xi_i^{\alpha 2}$$

となるが，ここで $\sum c_i^{\alpha 2}=p+1$ となることを用いれば，

$$\sum E(Y_i^0-\hat{Y}_i)^2=(n+p+1)\tau^2+\lambda_\alpha$$

$$=E(Q_\alpha)+2(p+1)\tau^2$$

となることが示される．したがって τ^2 がなんらかの形で推定できれば，上記の平均2乗誤差の和を，

$$Q_\alpha+2(p+1)\hat{\tau}^2$$

という形で推定することができる．そうしてこの値を最小にするモデルが望ましいモデルであると考えられる．Mallows の C_p 統計量はこの考え方にもとづいて導かれたものである．

この基準によれば，二つのモデル M_α および M_β において，M_β に含まれる説明変数の組が M_α に含まれる変数の組の部分集合になっているとき，M_β に含まれる変数の数を q とすれば，

$$Q_\alpha+2(p+1)\hat{\tau}^2\gtrless Q_\beta+2(q+1)\hat{\tau}^2 \qquad (4)$$

に応じてモデル M_α をとるか M_β をとるかが定められることになる．ここで $\hat{\tau}^2$ を，

$$\hat{\tau}^2=\hat{\sigma}_\alpha^2=Q_\alpha/(n-p-1)$$

で求めることにすれば，(4)は，

$$F=(Q_\beta-Q_\alpha)/(p-q)\hat{\sigma}_\alpha^2\gtrless 2$$

と同値になる．すなわちふつうの分散分析法における F 検定統計量を用いて，棄却限界として（有意水準と無関係に）2という値を用いることに対応する．

3. 因子分析型のモデル

前節の議論にはなおいくつかの問題点が残されている．とくに予測するケースとして，データとまったく同じ説明変数の値の組のくり返しを想定することは不自然なように思われるかもしれない．とくにここで議論の前提として，単にモデルの中に含まれている変数だけでなく，すべての説明変数の値がそのままくり返されることが要求されている点に注意しなければならない．しかしこれはやむを得ないところであって，もしある変数の影響が実際には大きいにもかかわらず，観測されたデータの中ではその値がまったく変化しなかったとすれば，データからその影響を推定することはできないから，その変数はモデルから除かざるをえない．そうして予測時においてその変数の値が大きく変わったために，モデルを用いた予測に大きな偏りが生じたとしても，それはさけられないことである．このような場合は，回帰式の係数自体が変化した場合と同じく，むしろ「構造変化」が生じたものと考えるほうが実際的である．そうして予測は「構造変化」は起こらないという前提で行なわざるをえないことは自明であろう．

説明変数についてこれと違った想定は，それらが同時確率分布に従う確率変数であるとみなすことである．そうして予測は，同じ分布に従う説明変数の組に対して行なうと考えるのである．このときもし任意の説明変数の組に対して，Yの条件付期待値が説明変数の線形関数になるならば，任意のモデルM_αに対して，

$$Y_i=\beta_0^\alpha+\beta_1^\alpha x_{\alpha_1 i}+\cdots+\beta_p^\alpha x_{\alpha_p i}+u_i^\alpha \qquad i=1,\ \cdots,\ n$$

とあらわすと，u_i^αの$x_{\alpha_1 i}\cdots x_{\alpha_p i}$を与えたときの条件付期待値は0になるから，$x$に関する$Y$の条件付分布を考えれば，ふつうの線形モデルと同様の関係が成立することになる．u_i^αの分散σ_α^2がxの値には無関係であると仮定すれば，β_i^αの最小2乗推定量の条件付分散共分散はふつうの場合と同じく$m^{jk}\sigma_\alpha^2$とあらわされるから，そのxの分布に関する期待値は$E(m^{jk})\sigma_\alpha^2$となる．xが多変量正規分布に従うと仮定し，その分散共分散行列を$\sum$，$x_{\alpha_1}\cdots x_{\alpha_p}$の分散共分散行列の逆行列の要素を$\nu_\alpha^{jk}$とあらわせば，この値は，

$$E(m^{jk})\sigma_\alpha^2=\nu^{jk}\sigma_\alpha^2/(n-p-2)$$

となることが知られている．このことから予測の誤差分散の期待値は，

$$E(Y_0-\hat{Y})^2=E(1+\sum\sum\nu^{jk}x_{j_0}x_{k_0})\sigma_\alpha^2=\{1+1/n+p/(n-p-2)\}\sigma_\alpha^2$$

となる．σ_α^2をその推定量$\hat{\sigma}_\alpha^2=Q_\alpha/(n-p-1)$でおきかえれば，結局

$$\left\{1+\frac{1}{n}+\frac{p}{n-p-2}\right\}\frac{Q_\alpha}{(n-p-1)}$$

をモデル選択の基準として用いればよいことが示される．もちろんここでも説明変数がまったくランダムに変動し，かつその分布が多変量正規分布に従うという想定は，一般には非現実的であるといわざるを得ない場合が多いであろう．

説明変数についての第3の想定は，因子分析モデルともいうべきものである．すなわち説明変数および被説明変数に影響を与える，直接には観測されない「真の」構造変数ともいうべき変数ζ_1, $\cdots$, ζ_rが存在して，

$$x_{ji}=\gamma_{j0}+\gamma_{j1}\zeta_{1i}+\cdots+\gamma_{jr}\zeta_{ri}+w_{ji} \qquad j=1,\ 2,\ \cdots,\ i=1,\ \cdots,\ n$$

$$Y_i=\delta_0+\delta_1\zeta_{1i}+\cdots+\delta_r\zeta_{ri}+v_i \qquad i=1,\ 2,\ \cdots,\ n$$

という関係が成り立っているものと考えるのである．ただしここでw_{ji}, v_iはすべて互いに独立であるとする．ここにrも未知であるが，それは考えられる説明変数の数よりはかなり小さいものと考えてよい．このようなモデルはRagnar Frischの考えたものであり，変数誤差モデルの拡張とみなすこともできる．

一般性を失うことなく$\sum\zeta_{ji}=0$, $\sum\zeta_{ji}^2=n$, $\sum\zeta_{ji}\zeta_{j'i}=0$と仮定することができる．また$E(w_{ji}^2)=\sigma_j^2$, $E(v_i^2)=\tau^2$とあらわす．いまモデルM_αを想定したときの関係を，

$$Y_i=\beta_0^\alpha+\beta_1^\alpha x_{\alpha 1 i}+\cdots+\beta_p^\alpha x_{\alpha_p i}+\xi_i^\alpha+u_i^\alpha$$

とあらわせば，

$$u_i^\alpha=v_i-\beta_0^\alpha-\beta_1^\alpha x_{1i}-\cdots-\beta_p^\alpha x_{\alpha_p i}$$

$$\xi_i^\alpha=\phi_1^\alpha\zeta_{1i}+\cdots+\phi_r^\alpha\zeta_{ri}$$

$$\phi_k^\alpha=\delta_k-\beta_1^\alpha\gamma_{\alpha_1 k}-\cdots-\beta_p^\alpha\gamma_{\alpha p k},\ k=1,\ 2,\cdots,\ r$$

であるから，

$$E(\sum_i x_{\alpha ji}(\xi_i^\alpha+u_i^\alpha))=0,\ j=1,2,\cdots,p$$

となるようにβ_j^αを定める．すなわち，

$$\sum_k\gamma_{\alpha jk}\phi_k^\alpha-\beta_j^\alpha\sigma_{\alpha j}^2=0,\ j=1,2,\cdots,p$$

を満たすようにβ_j^αを定めれば，最小2乗推定量の期待値がβ_j^αになる．またその分散共分散は，

$$E(m^{jk})\sigma_u^2 \quad \text{ただし} \quad \sigma_u^2=\tau^2+\beta_1^\alpha\sigma_{\alpha_1}^2+\cdots+\beta_p^\alpha\sigma_{\alpha_p}^2$$

となる．

いま特定の$\zeta_1^0\cdots\zeta_r^0$に対応するYの値の予測量を，このモデルを用いて，

$$\hat{Y}_0=\hat{\beta}_0^\alpha+\hat{\beta}_1^\alpha x_{\alpha_1}^0+\cdots+\hat{\beta}_p^\alpha x_{\alpha_p}^0$$

で与えれば，その偏りは，

$$E(Y_0-\hat{Y}_0)=\xi_0^\alpha=\phi_1^\alpha\zeta_1^0+\cdots+\phi_r^\alpha\zeta_r^0$$

となり，分散は，

$$V(Y_0-\hat{Y}_0)^2=\left(1+\frac{1}{n}+\sum\sum E(m^{jk})E(x_{\alpha j}^0x_{\alpha k}^0)\right)\sigma_u^2$$

となる．

このような関係から平均2乗誤差について一般的な関係を導くことは困難であるが，$p>r$ ならば偏りは小さくすることができる，また他方分散は p が増加すれば増加するから，p の値は r より小さくならない範囲で，あまり大きくならないことが望ましい．

したがって一つの方法として，説明変数，被説明変数のすべてについて因子分析法を適用して因子の数を定め，つぎに因子荷重の値を参照しながら，因子数とほぼ同数の説明変数を選び出すことが考えられる．しかしながらこのような方法について形式的な基準を定めることは困難である．

4. その他の問題

ところで回帰分析のモデルには，このほかにもいくつかの問題があり，それは説明変数選択の問題と理論的な関連をもっている．

一つは関係式の非線形性である．この問題を処理するには，一般に多項式のような一次式より一般的な関係式を用いるのと，変数変換によって線形関係式に帰着させるのと二つの方法がある．どちらによっても形式的には「あてはまり」はよくなる場合が少なくないが，変数変換を行なうことは，誤差分散についての仮定を変えることを意味するという点に注意しなければならない．すなわち，被説明変数を $Y=\phi(Y')$ によって Y' に変換するとき，Y' の分散が一定であると仮定することは，Y についてはその分散がほぼ，

$$\{\phi'(E(Y'))\}^2$$

に比例すると想定することを意味する．したがって変換の妥当性はこの点からも検討しなければならない．また予測誤差についても，Y' の予測量を $\hat{Y}'$，Y の予測量を $\hat{Y}=\phi(\hat{Y}')$ とすれば，

$$E(Y-\hat{Y})^2 \sim \{\phi'(E(Y'))\}^2 E(Y'-\hat{Y}')^2$$

という関係が成立するから，$\hat{Y}'$ の誤差の評価から $\hat{Y}$ の誤差についての評価を導くことができる．

誤差項に関する仮定，すなわち誤差の分散一定の仮定と，独立性の仮定については，単に「あてはまり」のよさだけからはチェックできないことに注意しよう．しかもこのような仮定が正しくなければ予測誤差の評価についての議論も妥当性を失うから，それについては一般に充分な事前および事後の検討が必要である．

またこれまでの議論において，回帰係数の推定はすべて最小2乗法によるものとしたが，このことは誤差項がほぼ正規分布に従うと想定することを意味する．もし誤差分布が正規分布からいちじるしく離れているならば，最小2乗推定量の効率は低くなり，他の推定法を採用することが必要になる．この問題は「ロバスト」な推定量を求める問題として論ぜられている．実際回帰分析が応用される多くの場合には「誤差」は狭い意味の観測誤差とは違って，多くの雑多な要因の影響の合成物であるから，それが厳密に正規分布に従うというような保証はないといってよい．

この問題については，この号の小柳氏の論文，および私の別稿を参照していただきたいが，つぎのことだけを注意しておこう．一つは非正規性の問題は現実のデータ解析において無視することはできないが，非線形性，分散の不均一性等の問題と比べてとくに重要な問題というわけではない．またこれらの問題と切り離して考えるのも正しくないということである．そのことは変数変換を考えても明らかであり，変数を変換することは関係式の形，分散の均一性，分布の形にすべて影響を与えるのである．第2に非正規性の問題を，形式的に処理すること，たとえば特定のロバストな推定方式のパッケージ化されたプログラムなどにたよることは適当でないということである．それよりもデータを注意深く眺めて検討すること，とくに残差をプロットしてみるということが大切であることを強調しておきたい．

たけうち・けい　1933年生
東京大学経済学部卒業，現在同教授