くり返しのないデータの分散均一性の検定

三浦 幹彦・山田 浩一
信州大学繊維学部繊維原料学研究室

通常の最小二乗法を用いた回帰分析や分散分析などの線形モデルによる解析では，誤差の独立性並びに誤差分散の均一性を仮定している。このため，この仮定が成立しないときは，最小二乗推定値は最良線形不偏推定値にならない。したがって，データ解析においては，常にモデルの仮定に対する妥当性の検討が必要である。

誤差の独立性の仮定は成立しているが，誤差分散の均一性に疑問があるとき，応答データにくり返しがあれば，分散均一性の検定によるモデル妥当性の検討がしばしば行われる。にもかかわらず，同一条件の下で一回しか応答データがない場合には，あまり検討されていないのが現状である。それは，くり返しがない場合，これまで提案されている検定統計量の計算が簡単でなかったこと，また，その分布のパーセント点が求めにくかったことなどが原因と思われる。最近，Cook & Weisberg (1983) は，対立仮説に特定の関数を想定して，有効スコアーに基づくスコアー検定統計量を提案した。これは，回帰効果を計算するパッケージプログラムを利用すれば，簡単に計算できる利点をもっている。また，その分布は漸近的に χ^2 分布になることが知られている。

そこで，この論文では，Cook & Weisberg と同じ対立仮説の下で，スコアー検定統計量と漸近的に同等な性質をもつ尤度比及びワルド検定統計量を導いた。次に，小標本での検定統計量の分布を検討するため，特別な場合ではあるが，スコアー検定統計量の正確な確率を求めて，データ数 n と χ^2 近似の程度について考察を加えた。さらに，この三つの検定統計量を製糸用水データの解析に適用して，分散均一性の診断を行うとともに，その特性を比較した。

1 検定統計量

$y_i(i=1,\ 2,\cdots,n)$ を $p+1$ 個の説明変数 $x_i^T=(1\quad x_{1i}\ x_{2i}\ \cdots\ x_{pi})$ に対する互いに独立な応答とし，y_i と x_i^T との間に次のような線形回帰モデルを仮定する。

$$Y=X\beta+\varepsilon$$

ただし，$Y=(y_1\ y_2\cdots y_n)^T$ を観測ベクトル，$\beta=(\beta_0\ \beta_1\cdots\beta_p)^T$ を未知係数ベクトル，$\varepsilon=(\varepsilon_1\ \varepsilon_2\cdots\varepsilon_n)^T$ を誤差ベクトルとし $X^T=(x_1\ x_2\cdots x_n)$ とする。また，

$$E[\varepsilon]=0,\quad E[\varepsilon\varepsilon^T]=\sigma^2\begin{pmatrix} w_1 & & & 0\\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n\end{pmatrix}$$

$\mathrm{rank}(X)=p+1<n$ が成立しているとする。ここで $w_i=1(i=1,\ 2,\cdots,n)$ のとき，誤差分散が一様な通常の回帰モデルである。

説明変数の一つの点に対して，一つしか応答がない場合の分散均一性に関する尤度比及びワルド検定統計量 (Rao, 1973) を求めるために，Cook & Weisberg と同様に w_i が

$$w_i=w(z_i,\ \lambda)\quad (i=1,\ 2,\cdots,n) \tag{1}$$

と表され，$q_z\times 1$ の既知ベクトル z_i と $q_\lambda\times 1$ の未知ベクトル λ に依存すると仮定する。ただし，w は λ に関して二回微分可能で，$w(z,\lambda^*)=1$ なる λ^* が存在する関数である。このとき，誤差分散 $\sigma^2 w_i(i=1,\ 2,\cdots,n)$ が一様かどうかは，複合帰無仮説

$$H_0:\lambda=\lambda^*,\ \beta,\ \sigma^2$$

を複合対立仮説

$$H_A:\lambda\neq\lambda^*,\ \beta,\ \sigma^2$$

に対して検定すればよい。そこで，誤差 ε_i が平均 0，分散 $\sigma^2 w_i$ で互いに独立に正規分布に従っていると仮定すれば，対数尤度関数は，

$$\log L(\lambda,\ \beta,\ \sigma^2)=-\frac{1}{2}n\log 2\pi-\frac{1}{2}n\log\sigma^2 -\frac{1}{2}\sum_{i=1}^{n}\log w_i-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{(y_i-x_i{}^T\beta)^2}{w_i} \tag{2}$$

となる。ただし，$x_i{}^T$ は行列 X の第 i 行を表す。したがって，$\hat\lambda,\ \hat\beta,\ \hat\sigma^2$ と $\hat\beta_0,\ \hat\sigma_0{}^2$ をそれぞれパラメータに制約がない場合と帰無仮説 H_0 の下での最尤推定量とすれば，尤度比検定統計量は，一般的に，(2)式から

$$\begin{aligned}-2\log\Lambda&=2\{\log L(\hat\lambda,\ \hat\beta,\ \hat\sigma^2)-\log L(\lambda^*,\ \hat\beta_0,\ \hat\sigma_0{}^2)\}\\&=n\log\frac{\hat\sigma_0{}^2}{\hat\sigma^2}-\sum_{i=1}^{n}\log\hat w_i\end{aligned} \tag{3}$$

で与えられる。また，$\theta^T=(\lambda^T,\ \beta^T,\ \sigma^2)$ としたとき，フイッシャー情報行列

$$I.(\lambda,\ \beta,\ \sigma^2)=E\left[\left(\frac{\partial\log L}{\partial\theta}\right)\left(\frac{\partial\log L}{\partial\theta^T}\right)\right]$$

を

$$\begin{bmatrix} I._{\lambda\lambda}(\lambda,\ \beta,\ \sigma^2) & I._{\lambda(\beta,\ \sigma^2)}(\lambda,\ \beta,\ \sigma^2)\\ I._{(\beta,\ \sigma^2)\lambda}(\lambda,\ \beta,\ \sigma^2) & I._{(\beta,\ \sigma^2)(\beta,\ \sigma^2)}(\lambda,\ \beta,\ \sigma^2)\end{bmatrix} \tag{4}$$

と分割すれば，ワルド検定統計量は定義より，

$$W_e=\hat{\lambda}^T\{I._{\lambda\lambda}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)$$
$$-I._{\lambda(\beta,\ \sigma^2)}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)I.^{-1}{}_{(\beta,\ \sigma^2)(\beta,\ \sigma^2)}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)I._{(\beta,\ \sigma^2)\lambda}(\hat{\lambda},\hat{\beta},\hat{\sigma}^2)\}\hat{\lambda}$$

であるが，(4)をさらに，

$$\begin{pmatrix} I._{\lambda\lambda}(\lambda,\ \beta,\ \sigma^2) & I._{\lambda\beta}(\lambda,\ \beta,\ \sigma^2) & I._{\lambda\sigma^2}(\lambda,\ \beta,\ \sigma^2) \\ I._{\beta\lambda}(\lambda,\ \beta,\ \sigma^2) & I._{\beta\beta}(\lambda,\ \beta,\ \sigma^2) & I._{\beta\sigma^2}(\lambda,\ \beta,\ \sigma^2) \\ I._{\sigma^2\lambda}(\lambda,\ \beta,\ \sigma^2) & I._{\sigma^2\beta}(\lambda,\ \beta,\ \sigma^2) & I._{\sigma^2\sigma^2}(\lambda,\beta,\sigma^2) \end{pmatrix}$$

と分割したとき，(1)，(2)式から

$$\frac{\partial^2 \log L(\lambda,\beta,\sigma^2)}{\partial\lambda_i\,\partial\beta_k}=-\frac{1}{\sigma^2}\sum_{j=1}^{n}\frac{x_{jk}\dfrac{\partial w_j}{\partial\lambda_i}}{w_j^2}(y_j-x_j^T\beta)$$

$$\frac{\partial^2 \log L(\lambda,\ \beta,\ \sigma^2)}{\partial\beta_i\,\partial\sigma^2}=-\frac{1}{\sigma^4}\sum_{j=1}^{n}\frac{x_{ij}}{w_j}(y_j-x_j^T\beta)$$

が得られ，$I._{\lambda\beta}(\lambda,\ \beta,\ \sigma^2)$，$I._{\beta\lambda}(\lambda,\ \beta,\ \sigma^2)$，$I._{\sigma^2\beta}(\lambda,\ \beta,\ \sigma^2)$，$I._{\beta\sigma^2}(\lambda,\ \beta,\ \sigma^2)$ は要素が全て 0 の行列となり，

$$W_e=\hat{\lambda}^T\{I._{\lambda\lambda}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)$$
$$-I._{\lambda\sigma^2}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)I.^{-1}{}_{\sigma^2\sigma^2}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)I._{\sigma^2\lambda}(\hat{\lambda},\hat{\beta},\hat{\sigma}^2)\}\hat{\lambda} \tag{5}$$

で与えられる。

また，有効スコアーベクトルを

$$U._{\lambda}=\frac{\partial \log L(\lambda,\ \beta,\ \sigma^2)}{\partial\lambda}$$

で表し，情報行列の逆行列 $I.^{-1}(\lambda,\ \beta,\ \sigma^2)$ を

$$\begin{bmatrix} I.^{\lambda\lambda}(\lambda,\ \beta,\ \sigma^2) & I.^{\lambda(\beta,\ \sigma^2)}(\lambda,\ \beta,\ \sigma^2) \\ I.^{(\beta,\ \sigma^2)\lambda}(\lambda,\ \beta,\ \sigma^2) & I.^{(\beta,\ \sigma^2)(\beta,\ \sigma^2)}(\lambda,\ \beta,\ \sigma^2) \end{bmatrix}$$

と分割したとき，スコアー検定統計量 (Cox & Hinkley, 1974) は，

$$W_u=U^T._{\lambda}(\lambda^*,\ \hat{\beta}_0,\ \hat{\sigma}_0^2)\,I.^{\lambda\lambda}(\lambda^*,\ \hat{\beta}_0,\ \hat{\sigma}_0^2)\,U._{\lambda}(\lambda^*,\ \hat{\beta}_0,\ \hat{\sigma}_0^2) \tag{6}$$

で与えられる。

さらに，Cook & Weisberg と同様に w_i として具体的に，$q_z=q_\lambda=q$ として

(i) $w_i=\exp\left(\sum_{j=1}^{q}\lambda_j z_{ij}\right)$

(ii) $w_i=\prod_{j=1}^{q} z_{ij}{}^{\lambda_j}$

(iii) $w_i=\exp(\lambda x_i^T\beta)$

を考える。ここで，実際には z_{ij} として行列 X の任意の列を選ぶことになる。したがって，(i)，(ii)は誤差分散が説明変数の任意の組に影響を受けると考えることに相当する。また，(iii)は誤差分散が平均応答の関数となっていることを示している。(3)式から，尤度比検定統計量として，

$$n\log\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}-\sum_{j=1}^{q}\lambda_j\left(\sum_{i=1}^{n}z_{ij}\right) \tag{7}$$

$$n\log\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}-\sum_{j=1}^{q}\lambda_j\left(\sum_{i=1}^{n}\log z_{ij}\right) \tag{8}$$

$$n\log\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}-\lambda\sum_{i=1}^{n}x_i^T\beta \tag{9}$$

が求められる。

$$Z=\begin{pmatrix} z_{11} & z_{12} & \cdots\cdots & z_{1q} \\ z_{21} & z_{22} & \cdots\cdots & z_{2q} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots\cdots & z_{nq} \end{pmatrix} \tag{10}$$

としたとき，(i)の関数に対して，(1)，(2)式から

$$I._{\lambda\lambda}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)=\frac{1}{2}Z^TZ$$

$$I._{\lambda\sigma^2}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)=\frac{1}{2\hat{\sigma}^2}Z^T1$$

$$I._{\sigma^2\sigma^2}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma})=\frac{n}{2\hat{\sigma}^4}$$

が求まるから，この結果を(5)式へ代入すれば，w_i として (i) の関数を仮定したときのワルド検定統計量

$$\frac{1}{2}\lambda^T\bar{Z}^T\bar{Z}\lambda \tag{11}$$

が求まる。ただし，$\bar{Z}=Z-11^TZ/n$ であり，Z の各要素から，その要素の属する列の平均値を引いたものを要素にもつ行列である。同様にして，w_i として (ii) の関数を仮定すれば，(10)の行列の要素 z_{ij} を $\log z_{ij}$ に変えたものを新たに Z と考えれば，形式的に(11)式と同じ式が得られる。(iii) の関数に対しては，

$$I._{\lambda\lambda}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)=\frac{1}{2}\beta^TX^TX\beta$$

$$I._{\lambda\sigma^2}(\hat{\lambda},\ \hat{\beta},\ \hat{\sigma}^2)=\frac{1}{2\hat{\sigma}^2}1^TX\beta$$

から

$$\frac{1}{2}\lambda^2\beta^T\overline{X}^T\overline{X}\beta \tag{12}$$

が求められる。ただし，$\overline{X}$は$\overline{Z}$と同様に $\overline{X}=X-11^T/n$ を表すものとする。以上のようにして，対立仮説に特定の関数を想定した場合の分散均一性の尤度比検定及びワルド検定が導かれた。

ところで，e_i を仮説が成立するときの残差として，

$$F=\left[\frac{e_1^2}{\hat{\sigma}_0^2}\quad\frac{e_2^2}{\hat{\sigma}_0^2}\quad\cdots\cdots\quad\frac{e_n^2}{\hat{\sigma}_0^2}\right]^T$$

とおけば

$$U._\lambda(\lambda^*,\ \beta_0,\ \hat{\sigma}_0^2)=\frac{1}{2}\overline{Z}^T(F-1)$$

$$I.^{\lambda\lambda}(\lambda^*,\ \beta_0,\ \hat{\sigma}_0^2)=2(\overline{Z}^T\overline{Z})^{-1}$$

となり，(6)式からスコアー検定統計量

$$W_u=\frac{1}{2}F^T\overline{Z}(\overline{Z}^T\overline{Z})^{-1}\overline{Z}^TF \tag{13}$$

が得られる。これが，Cook & Weisberg により与えられた検定統計量であり，(13)式の右辺は形式的な回帰モデル

$$F=\nu_0 1+Z\nu+\varepsilon$$

に対する回帰平方和の 1/2 となっている。ここで Z はワルド検定統計量の場合と同様に，w_i が (i) のときは(10)の行列式，(ii) のときは要素を $\log z_{ij}$ としたものを表している。

2　検定統計量の分布

前節で与えた検定統計量は，データ数 n が大きいときいずれも帰無仮説 $H_0:\lambda=\lambda^*$ の下で，漸近的に自由度が q_λ の χ^2 分布に従うことが知られている。しかし，実際のデータ解析においては，データ数 n が小さいか中程度の場合が多い。そこで，こうした n に対する統計量の χ^2 近似の程度を知ることが重要である。尤度比及びワルド検定統計量の正確な分布を直接計算するのは困難でシミュレーション等によらなければならない。スコアー検定統計量も同様であるが，この場合，$q_\lambda=1$ であれば正確な確率を計算することができる。

(10)式において $q=1$ であるから，(13)式は，

$$W_u=\frac{1}{2}\frac{(F^T\bar{Z})^2}{\bar{Z}^T\bar{Z}}$$

$$=\frac{1}{2}\left\{n\frac{e^TAe}{e^Te}-1^TZ\right\}\Big/(\bar{Z}^T\bar{Z})$$

$$=\frac{1}{2}\left\{n\frac{\varepsilon^TQAQ\varepsilon}{\varepsilon^TQ\varepsilon}-1^TZ\right\}\Big/(\bar{Z}^T\bar{Z})$$

$$=\frac{(nT-c_2)^2}{2c_1}$$

となる。ただし，$Q=I-X(X^TX)^{-1}X^T$，$e=(e_1\,e_2\cdots\cdots e_n)^T$，$c_1=\bar{Z}^T\bar{Z}$，$c_2=1^TZ$

$$T=\frac{\varepsilon^TQAQ\varepsilon}{\varepsilon^TQ\varepsilon}$$

また，Aは $\dfrac{dw(z_i,\ \lambda^*)}{d\lambda}$ を対角要素とする対角行列である。したがって，

$$P(W_u\leq c)=P\left\{\frac{1}{n}(c_2-\sqrt{2cc_1})\leq T\leq\frac{1}{n}(c_2+\sqrt{2cc_1})\right\}$$

が得られる。ここで，Tは正規分布に従う変量 ε の二次形式の比となっているから Box (1954) の方法により，互いに独立な自由度１の χ^2 変量の線形結合の分布に変換して，Davies (1980) のプログラムを用いて W_u の正確な分布を計算することができる。そこで，w_i として前節 (i) の関数を仮定し，Zとして１を除いた説明変数行列 X_0 の第一列を用いることとし，X_0 に 25×10，50×10，100×10，200×10 の標準正規乱数を与え，それぞれ５種類の異なった X_0 の値に対して W_u の正確な確率を計算した。表１(a)に自由度１の χ^2分布の0.50，0.90，0.95，0.99 に対する W_u の正確な確率を示した。データ数25では，X_0 の影響を少し受けるが，200 になればかなり安定し χ^2 分布の確率によく近似していることが知られた。また，w_i，Z ともに上と同じ条件で，X_0 の列の数を３に減らした場合の値を表１(b)に示した。これも，表１(a)と同様な傾向を示している。さらに，次節で用いる製糸用水の煮沸後 pH，アルカリ度，電導度に関する209個のデータ及び異常値と思われる一個を除いた208個のデータをX_0とし，Zにこの３変数の各々を選び，w_iを前節に示した関数(i)，(ii)，(iii) と変えたときの正確な W_u の確率を計算し表２に示した。q_λ が２以上の場合については，Cook & Weisberg がシミュレーションにより検討しているが，同様な結果が得られている。

表1　標準正規乱数を説明変数行列X_0として用いたときのスコア検定統計量 W_u の真の確率
$w_i=\exp\lambda z_i$ として z はX_0の第1列を与えた

(a)

X_0	1	2	3	4	5	χ^2
25×10	0.523	0.511	0.556	0.533	0.470	0.500
	0.906	0.907	0.941	0.926	0.887	0.900
	0.951	0.953	0.975	0.967	0.946	0.950
	0.988	0.990	0.997	0.996	0.992	0.990
50×10	0.525	0.517	0.502	0.520	0.508	0.500
	0.917	0.914	0.902	0.911	0.907	0.900
	0.960	0.960	0.951	0.955	0.955	0.950
	0.993	0.993	0.990	0.990	0.991	0.990
100×10	0.512	0.506	0.513	0.511	0.506	0.500
	0.908	0.903	0.908	0.905	0.903	0.900
	0.955	0.952	0.955	0.952	0.951	0.950
	0.991	0.990	0.991	0.990	0.990	0.990
200×10	0.505	0.507	0.506	0.505	0.508	0.500
	0.903	0.904	0.903	0.903	0.904	0.900
	0.951	0.952	0.951	0.952	0.952	0.950
	0.990	0.990	0.990	0.990	0.990	0.990

(b)

X_0	1	2	3	4	5	χ^2
25×3	0.554	0.499	0.529	0.516	0.527	0.500
	0.930	0.905	0.923	0.917	0.922	0.900
	0.968	0.955	0.965	0.962	0.963	0.950
	0.994	0.992	0.994	0.994	0.993	0.990
50×3	0.512	0.518	0.518	0.515	0.525	0.500
	0.909	0.912	0.909	0.914	0.916	0.900
	0.955	0.957	0.955	0.958	0.960	0.950
	0.991	0.992	0.990	0.992	0.992	0.990
100×3	0.510	0.507	0.516	0.509	0.510	0.500
	0.906	0.905	0.910	0.906	0.906	0.900
	0.953	0.953	0.955	0.954	0.953	0.950
	0.990	0.990	0.991	0.991	0.990	0.990
200×3	0.505	0.507	0.506	0.505	0.507	0.500
	0.903	0.904	0.903	0.903	0.904	0.900
	0.951	0.952	0.951	0.951	0.952	0.950
	0.990	0.990	0.990	0.990	0.990	0.990

表2 製糸用水データを説明変数行列X_0として用いたときのスコアー検定統計量 W_u の真の確率

w	X_0	z: x_1	x_2	x_3	χ^2
$\exp(\Sigma\lambda z)$	209×3	0.505	0.512	0.508	0.500
		0.908	0.912	0.910	0.900
		0.955	0.958	0.957	0.950
		0.990	0.991	0.991	0.990
	208×3	0.506	0.509	0.503	0.500
		0.908	0.909	0.905	0.900
		0.955	0.955	0.954	0.950
		0.990	0.990	0.991	0.990
Πz^{λ}	209×3	0.505	0.508	0.509	0.500
		0.911	0.905	0.904	0.900
		0.958	0.953	0.952	0.950
		0.991	0.990	0.990	0.990
	208×3	0.505	0.507	0.508	0.500
		0.911	0.904	0.904	0.900
		0.958	0.952	0.952	0.950
		0.991	0.990	0.990	0.990
$\exp(\lambda x^T\beta)$	209×3		0.511		0.500
			0.907		0.900
			0.954		0.950
			0.990		0.990
	208×3		0.510		0.500
			0.907		0.900
			0.954		0.950
			0.990		0.990

3 製糸用水データの解析への適用

尾藤（1973）は製糸用水の水質と繭層セリシン溶解度との関係を明らかにするため，全国各地に散らばる製糸工場周辺の井水及び河川の水質を調査し，繭層セリシン溶解度の測定を行った。ここでは，そのうち，関東，信越，山梨，東海，山陰，四国地方の井水データ 209 個を対象として誤差分散の均一性に関する診断を行った。尾藤の実験の目的に応じて繭層セリシン溶解度（y）を応答変数とした。また，水質調査は22項目にわたり行われたが変数増加及び減少法により繭層セリシン溶解度に影響すると思われた煮沸後 pH(x_1)，アルカリ度（x_2），電導度（x_3）の 3 変数を説明変数として選択した。誤差が独立で同一分散をもつ通常の線形回帰モデル

PLOT OF E*PRY　　LEGEND: A = 1 OBS, B = 2 OBS, ETC.

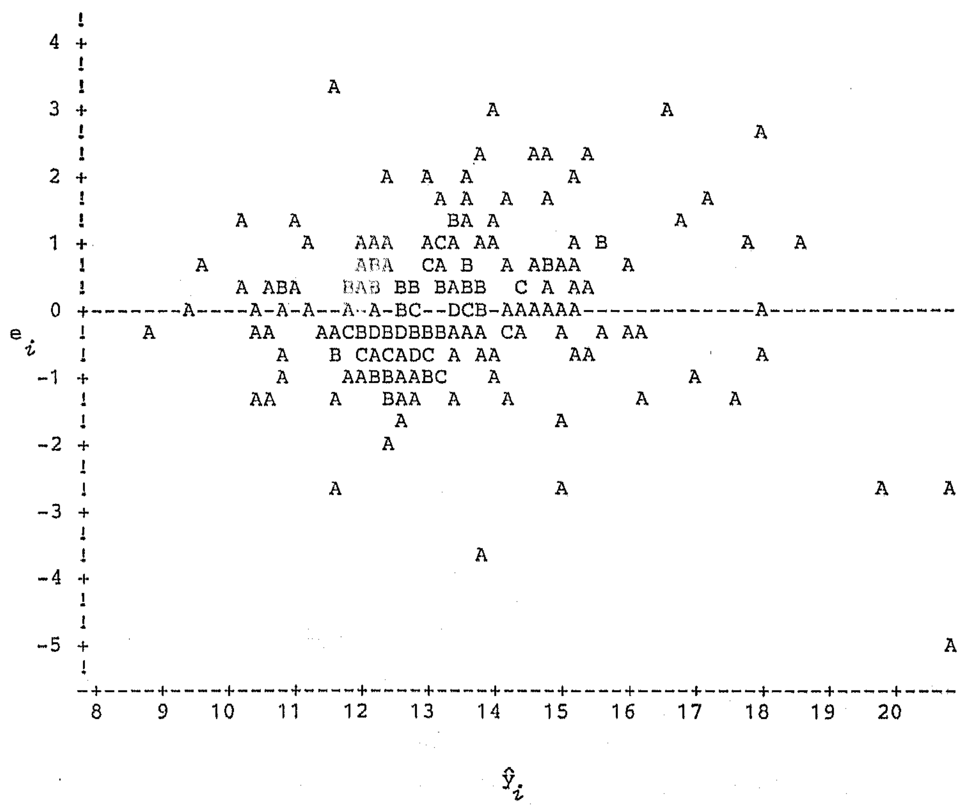

図１　製糸用水データにおける推定値 $\hat{y}_i$ に対する残差の散布図（n＝209）

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon$$

により回帰分析を行った結果を表４に，また，このときの推定値 $\hat{y}$ に対する残差プロットを図１に示した。また，異常値と思われるデータを除いた208個のデータに対して同じ回帰分析を行った結果を表４，図２に示した。図１，２の残差プロットからは，いくぶんばらつきに均一性を欠く様子が見られるが，顕著な傾向はない。説明変数 x_1，x_2，x_3 に対する残差プロットにおいても同様であった。そこで，前節で与えた漸近的に同等な３種類の検定統計量を用いて誤差分散均一性の検定を行った。検定統計量の値は表３に示すように，場合によっては，尤度比，ワルド，スコアー間でかなりばらついていたが，いずれの結果からも誤差分散は煮沸後 pH のみによる場合均一であり，アルカリ度や電導度，またはそれらとの組合せの影響を受けて不均一分散となっていることが知られた。したがって，OLS（普通最小二乗法）により表４に求めた回帰係数の推定値は最良なものでなく，最良推定値を求めるには WLS（重みつき最小二乗法）あるいは GLS（一般化最小二乗法）など別の推定法を用いなければならない。

表3　製糸用水データに関する尤度比，ワルド，スコアー検定統計量の値

x_1：煮沸後 pH　　x_2：アルカリ度　　x_3：電導度

w	z	尤度比	ワルド	スコアー	$d.f.$	χ^2
$\exp(\lambda z)$	x_1	0.813 (n=209)	1.169	0.496	1	3.841
		2.414 (n=208)	3.155	1.694		
	x_2	46.561	24.714	92.880	1	3.841
		25.868	15.798	46.728		
	x_3	38.096	27.236	64.376	1	3.841
		15.681	17.362	14.527		
	x_1, x_2	47.061	27.188	99.517	2	5.991
		26.627	18.595	48.306		
	x_1, x_3	38.772	28.873	65.110	2	5.991
		16.644	18.900	14.740		
	x_2, x_3	47.509	25.363	97.012	2	5.991
		25.983	16.144	47.090		
	x_1, x_2, x_3	47.980	27.697	103.243	3	7.815
		26.668	18.819	48.674		
Πz^{λ}	x_1	0.955	1.436	0.594	1	3.841
		2.561	3.326	1.769		
	x_2	17.682	8.759	30.626	1	3.841
		8.616	4.370	15.482		
	x_3	28.173	24.429	28.772	1	3.841
		13.775	14.134	12.357		
	x_1, x_2	18.126	11.334	35.786	2	5.991
		9.995	8.125	16.103		
	x_1, x_3	28.276	24.692	29.652	2	5.991
		13.908	14.025	12.400		
	x_2, x_3	28.480	23.058	36.255	2	5.991
		14.055	13.048	17.299		
	x_1, x_2, x_3	28.550	23.203	41.302	3	7.815
		14.249	13.112	17.929		
$\exp(\lambda x^T \beta)$		46.249	28.176	69.350	1	3.841
		26.437	17.962	41.482		

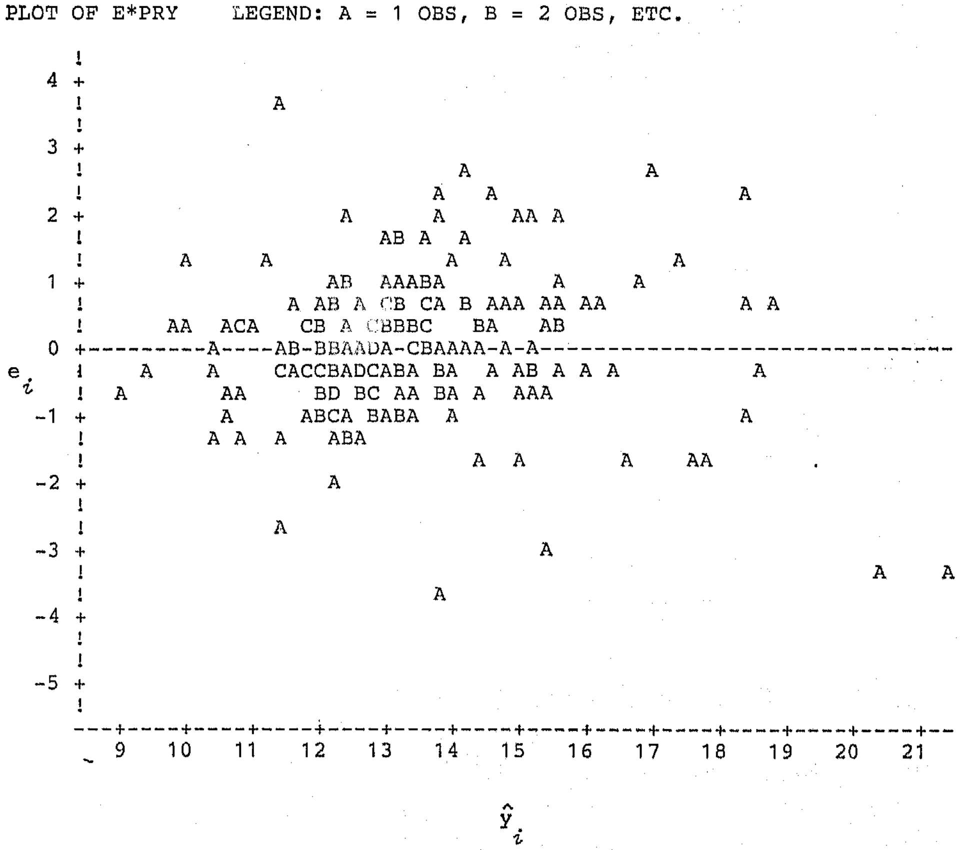

図２　製糸用水データにおける推定値 $\hat{y}_i$ に対する残差の散布図（n =208）

表４　製糸用水データの重回帰分析

$y=\beta_0+\beta_1$(煮沸後 pH) $+\beta_2$(アルカリ度) $+\beta_3$(電導度) $+\ \varepsilon$

X_0		自由度	推定値	標準誤差	t 値	R^2
209×3	β_0	1	1.604	1.520	1.056	0.741
	β_1	1	1.079	0.175	6.148	
	β_2	1	0.051	0.304×10^{-2}	16.883	
	β_3	1	−0.00282	0.945×10^{-3}	−2.986	
208×3	β_0	1	2.625	1.434	1.830	0.773
	β_1	1	0.928	0.166	5.575	
	β_2	1	0.054	0.289×10^{-2}	18.728	
	β_3	1	−0.00178	0.904×10^{-3}	−1.963	

```
DATA WATER;
  SET INSAS.WATER;
PROC NLIN DATA=WATER;
  PARAMS V=1.2073 B0=1.6043 B1=1.0761 B2=0.05134 B3=-0.0028205
         LAMDA1=0.0000 LAMDA2=0.0000 LAMDA3=0.0000;
  BOUNDS V>0;
  RES=Y-B0-B1*X1-B2*X2-B3*X3;
  RESSQ=RES**2;
  W=EXP(LAMDA1*X1+LAMDA2*X2+LAMDA3*X3);
  F=0.9189385+0.5*LOG(V)+0.5*(LAMDA1*X1+LAMDA2*X2+LAMDA3*X3)
 +0.5*RESSQ/(V*W);
  REP:F=F+1;
    IF F<0 THEN GOTO REP;
  FS=SQRT(F);
  FVW=FS*V*W;
  YIM=0;
  MODEL YIM=FS;
  DER.V=-0.5/FS*(-0.5/V+0.5*RESSQ/(V**2*W));
  DER.B0=-0.5*RES/FVW;
  DER.B1=-0.5*X1*RES/FVW;
  DER.B2=-0.5*X2*RES/FVW;
  DER.B3=-0.5*X3*RES/FVW;
  DER.LAMDA1=-0.25*X1*(-1/FS+RESSQ/FVW);
  DER.LAMDA2=-0.25*X2*(-1/FS+RESSQ/FVW);
  DER.LAMDA3=-0.25*X3*(-1/FS+RESSQ/FVW);
```

図3 SAS NLIN プロシジャーを用いた最尤推定のプログラム

ところで，尤度比検定統計量(7), (8), (9)の計算には，帰無仮説の下での最尤推定値 $\hat{\sigma}_0^2$ とパラメータに制約がない場合の最尤推定値 σ^2, $\hat{\lambda}$ が必要であり，このうち $\hat{\sigma}^2$ と $\hat{\lambda}$ は(2)式により得られる正規方程式からくり返し計算で求めなければならない。ワルド検定統計量(11), (12)の計算においても σ^2, λ の推定が必要である。これに対して，スコアー検定統計量(13)は帰無仮説の下での最尤推定値だけが必要で，あとは回帰平方和を計算すれば簡単に求まる利点がある。しかし，(2)式の対数尤度関数に基づく σ^2, λ, β の最尤推定も SAS NLIN プロシジャーや BMDP 等のパッケージプログラムを利用すれば，それほど大変とは思われない。例として表3の結果の計算に用いた SAS プログラムを図3に示した。分散均一性が否定された後のデータ処理を考えれば，尤度比検定，ワルド検定の方がスコアー検定より便利な面もあるのでどれが最良かの判定は困難である。

文献

尾藤省三：蚕試報, **25**, 371-583 (1973)

Box, G. E. P. : Ann. Math. Stat., **25**, 290-302 (1954)

Cook, R. D. & Weisberg, G. S. : Biometrika, **70**, 1-10 (1983)

Cox, D. R. & Hinkley, D. V. : Theoretical Statistics (1974) Chapman and Hall

Davies, R. B. : Applied Statistics, **29**, 323-333 (1980)

Rao, C. R. : Linear Statistical Inference and Its Applications (1973)Wiley

Summary

Test for heteroscedasticity in regression model without replication

Mikihiko MIURA and Kouichi YAMADA
(Received September 10, 1984)

Cook & Weisberg (1983) proposed a score test for homoscedasticity in regression model without replication based on a specific alternative model for heteroscedasticity. In this paper likelihood ratio and Wald test statistics asymptotically equivalent to the score test are given under the same alternative model. These tests are applied to the analysis of the quality of water for raw silk reeling.