共分散分析の計算方法 : 拡大観測データ行列による方法(村上義弘教授還暦記念号)

共分散分析の計算方法

——拡大観測データ行列による方法——

今川　　正

ここで r 個の回帰をいっしょに研究するものについて考察する。これは分散分析と回帰分析とを結合したものであり共分散分析とよばれているものである。

r 個の処理レベル $k=P,\ Q,\ \cdots, S$（r 個）によってえられる反応 Y_{ik} ($i=1, 2, \cdots, n$) のデータを表1に示す。また，そこに利用できる共変量 covariate の測定値 X_{ik} も示しておいた。

この分析の目的は，共変量の効果を除去したあとの反応の差を調べることである。このためつぎのモデルを用いる。

$$Y_{ik}=\alpha_k+\beta_k X_{ik}+\varepsilon_{ik}$$

$$\varepsilon_{ik} \sim N(0, \sigma^2)\ \text{独立}$$

$$k=P,\ Q,\ \cdots,\ S$$

$$i=1,\ 2,\ \cdots n$$

われわれのモデルのパラメータ空間をつぎのようにあらわしておく。

$$\Pi=\{\alpha_P,\ \cdots,\ \beta_S,\ \sigma^2 \mid \infty<\alpha_P,\ \cdots,\ \beta_S<\infty, 0<\sigma^2<\infty\}$$

平均の仮説の検定

ここで r 個の回帰直線が平行であるという帰無仮説の検定について考える。$\alpha=0.05$。ここで用いる尤度関数をつぎのようにあらわしておく。

$$L(\Pi)=\left(\frac{1}{2\pi\sigma^2}\right)^{\pi n/2}\exp-\frac{1}{2\sigma^2}SS$$

$$SS=\Sigma\Sigma\varepsilon_{ik}^2$$

$$=\Sigma\Sigma(Y_{ik}-\alpha_k-\beta_k X_{ik})^2$$

ここで仮説を

$$H_0(\beta):\beta_P=\beta_Q=\ \cdots\ =\beta_S=\beta$$

$H_1(\beta):H_0(\beta)$ が成立しない

とあらわしておき，帰無仮説，対立仮説，パラメータ空間をつぎのようにあらわす。

$H_0:H_0(\beta)$

$H_1:H_1(\beta)$

ここでは H_0 は $H_0(\beta)$ に等しく，H_1 は $H_1(\beta)$ に等しい。

$$\omega=\{\Pi \mid H_0\}$$

$$\bar{\omega}=\{\Pi \mid H_1\}$$

$$\Omega=\omega+\bar{\omega}$$

この ω——これを平行 p-ケースとよぶ——においては，つぎの最尤推定量（これに～の記号をつける）をえる。

$$\tilde{\alpha}_k=\Sigma Y_{ik}/n-\tilde{\beta}\Sigma X_{ik}/n \qquad k=P,\ Q,\ \cdots,\ S$$

$$\tilde{\beta}=\frac{\Sigma\Sigma XY-\Sigma[(\Sigma X_{ik})(\sum Y_{ik})]/n}{\Sigma\Sigma X^2-\Sigma(\Sigma X_{ik})^2/n}$$

$$\tilde{\varepsilon}_{ik}=Y_{ik}-\tilde{\alpha}_k-\tilde{\beta}_k X_{ik}$$

$$SS_p=\Sigma\Sigma(\tilde{\varepsilon}_{ik})^2$$

$$=\Sigma\Sigma(Y_{ik}-\tilde{\alpha}_k-\tilde{\beta}_k X_{ik})^2$$

$$DF_p=rn-(r+1)$$

$$\tilde{L}=\left(\frac{n}{2\pi SS_p}\right)^{rn/2}\exp-\frac{n}{2}$$

ここに，シングル・サム Σ はすべての i に関する合計 $\sum_{j=1}^{n}$ をあらわしている。ダブル・サムのうしろにつづく変数についての添字は省略しておいた。

Ω——これを個別 s-ケースとよぶ——においては，つぎの最尤推定量（これにへの記号をつける）をえる。

$$\hat{\alpha}_k=\Sigma Y_{ik}/n-\hat{\beta}_k\Sigma X_{ik}/n$$

$$\hat{\beta}_k=\frac{\Sigma X_{ik}Y_{ik}-(\Sigma_{ik})(\Sigma Y_{ik})/n}{\Sigma X_{ik}^2-(\Sigma X_{ik})^2/n} \qquad k=P,\ Q,\ \cdots,\ S$$

$$\hat{\varepsilon}_{ik}=Y_{ik}-\hat{\alpha}_k-\hat{\beta}_k X_{ik}$$

$$SS_s=\Sigma\Sigma(\hat{\varepsilon}_{ik})^2$$

$$=\Sigma\Sigma(Y_{ik}-\hat{\alpha}_k-\hat{\beta}_k X_{ik})^2$$

$$DF_s = rn - 2r$$

$$\hat{L} = \left(\frac{n}{2\pi SS_s}\right)^{\pi n/2} \exp - \frac{n}{2}$$

ここで，

$$SS_\beta = SS_p - SS_s$$

$$DF_\beta = DF_p - DF_s = r - 1$$

とおくと，尤度比基準 $\lambda = \tilde{L}/\hat{L} < h$ よりつぎのものをえる。

$$\lambda^{-2/rn} - 1 = \frac{SS_\beta}{SS_s}$$

$$\frac{SS_\beta/DF_\beta}{SS_s/DF_s} = \frac{v^2}{s^2} = w^2 \overset{H_0}{\sim} F(DF_\beta,\ DF_s)$$

$$C = F(DF_\beta, DF_s)_{1-\alpha}$$

$$R = \{w^2 \mid w^2 > c\}$$

分散分析表はつぎのようにあらわされる。

$ANOVA$ β効果（s-ケース対 p-ケース）

	SS	DF	$MS = SS/DF$	$TS = v^2/s^2$
$D(\beta)$	SS_β	DF_β	v^2	w^2
$E(s)$	SS_s	DF_s	s^2	
$T(p)$	SS_p	SS_p		

一体の仮説の検定

つぎに r 個の回帰直線が合致し一本のものである，一体のものであるという仮説の検定について考える。ここで追加的に用いる仮説を

$$H_0(\alpha) : \alpha_P = \alpha_Q = \cdots = \alpha_S = \alpha$$

$H_1(\alpha)$: $H_0(\alpha)$ が成立しない

とあらわしておき，ここで用いる帰無仮説，対立仮説，パラメータ空間をつぎのようにあらわす。

$$H_0 : H_0(\beta),\ H_0(\alpha)$$

$$H_1 : H_0(\beta),\ H_1(\alpha)$$

ここに，帰無仮説，対立仮説の双方に $H_0(\beta)$ が含まれている点に注目して

おこう。

$$\omega = \{\Pi \mid H_0\}$$

$$\bar{\omega} = \{\Pi \mid H_1\}$$

$$\Omega = \omega + \bar{\omega}$$

この Ω——これは平行 p-ケースである——においては，まえのパラグラフで述べたものと同じ最尤推定量をえることができる。(それに～の記号をつけることも同じようにする)。

ω においては——これを一体 u-ケースとよぶ——つぎの最尤推定量をえる。(ここではこれに一の記号をつけておく)。SS，DF には添字は u をつけておく。

$$\bar{\alpha} = \Sigma\Sigma Y/rn - \bar{\beta}\Sigma\Sigma X/rn$$

$$\bar{\beta} = \frac{\Sigma\Sigma XY - (\Sigma\Sigma X)(\Sigma\Sigma Y)/rn}{\Sigma X^2 - (\Sigma\Sigma X)^2/rn}$$

$$\bar{\varepsilon}_{tk} = Y_{tk} - \bar{\alpha} - \bar{\beta}X_{tk}$$

$$SS_u = \Sigma\Sigma(\bar{\varepsilon}_{tk})^2$$

$$= \sum\sum(Y_{tk} - \bar{\alpha} - \bar{\beta}X_{tk})^2$$

$$DF_u = rn - 2$$

$$\bar{L} = \left(\frac{n}{2\pi SS_u}\right)^{rn/2} \exp - \frac{n}{2}$$

これを用い，尤度比基準よりえる検定統計量の形はうえの平行の仮説の検定のものと同じである。

$$SS_\alpha = SS_u - SS_p$$

$$DF_\alpha = DF_u - DF_p = r - 1$$

$$\frac{SS_\alpha/DF_\alpha}{SS_p/DF_p} = \frac{v^2}{s^2} = w^2 \overset{H_0}{\sim} F DF_\alpha, DF_p)$$

$$C = F(DF_\alpha,\ DF_p)_{1-\alpha}$$

$$R = \{w^2 \mid w^2 > c\}$$

分散分析表はつぎのようにあらわされる。

$ANOVA$ α効果（p-ケース対 u-ケース）

	SS	DF	$MS=SS/DF$	$TS=v^2/s^2$
$D(\alpha)$	SS_α	DF_α	v^2	w^2
$E(_p)$	SS_p	DF_p	s^2	
$T(u)$	SS_u	DF_u		

このための数値計算の方法を節を改めて述べる。

観測データ行列による方法

予備の計算

処理レベル k ごとに観測値の合計をもとめる。

ΣY_{ik}

ΣX_{ik} $\qquad k=P,\ Q,\ \cdots,\ S$

観測値の総合計をもとめる。

$\Sigma\Sigma Y_{ik}$

$\Sigma\Sigma X_{ik}$

以下シングル・サム Σ は $i=1,\ 2,\ \cdots,\ n$ のすべての成分に関する合計をあらわすことにする。またこれらの平均，すなわち処理レベルごとの平均，総平均をつぎの記号であらわす。

$\bar{Y}_{\cdot k}=\Sigma Y_{ik}/n$

$\bar{Y}_{\cdot\cdot}=\Sigma\Sigma Y_{ik}/rn$ 等。

この２種類の平均からの偏差をもとめ，積和をつくり，表２，分散分析表の所定の行へ記入する。

S行　（r 個の行）

処理レベル k の１つづつについてつぎの偏差をつくる。

$y_{ik}=Y_{ik}-\bar{Y}_{\cdot k}=Y_{ik}-\Sigma Y_{ik}/n$

$x_{ik}=X_{ik}-\bar{X}_{\cdot k}=X_{ik}-\Sigma X_{ik}/n$

そして次式によって，三種の積和をもとめる。

$$\Sigma y_{ik}^2=\Sigma(Y_{ik}-\bar{Y}_{\cdot k})^2$$

$$=\Sigma Y_{ik}^2-n(\bar{Y}_{\cdot k})^2$$

$$=\Sigma Y_{ik}^2-(\Sigma Y_{ik})^2/n$$

$$\Sigma x_{ik}y_{ik}=\Sigma(Y_{ik}-\bar{Y}_{\cdot k})(X_{ik}-\bar{X}_{ik})$$
$$=\sum X_{ik}Y_{ik}-n\bar{X}_{\cdot k}\bar{Y}_{\cdot k}$$
$$=\Sigma X_{ik}Y_{ik}-(\Sigma X_{ik})(\Sigma X_{ik})/n$$

$$\Sigma x_{ik}^2=\Sigma(X_{ik}-\bar{X}_{\cdot k})^2$$

$$=\Sigma X_{ik}^2-n(\bar{X}_{\cdot k})^2$$

$$=\Sigma X_{ik}^2-(\Sigma X_{ik})^2/n$$

Σ行

表 1　観測データ行列

	1	2		n	
P	Y_{1P}	Y_{2P}	………	Y_{nP}	ΣY_{iP}
	X_{1P}	X_{2P}	………	X_{nP}	ΣX_{iP}
Q	Y_{1Q}	Y_{2Q}	………	Y_{nQ}	ΣY_{iQ}
	X_{1Q}	X_{2Q}	………	X_{nQ}	ΣX_{iQ}
…			………		
S	Y_{1S}	Y_{2S}		Y_{nS}	ΣY_{iS}
	X_{1S}	X_{2S}		X_{nS}	ΣX_{iS}
					$\Sigma\Sigma Y_{ik}$
					$\Sigma\Sigma X_{ik}$

表 2　分散分析表

	(1)	(2)	(3)	(4)	(5)	(6)
S	Σy_{iP}^2	$\Sigma x_{iP}y_{iP}$	Σx_{iP}^2	$\hat{\beta}_P$	$\hat{\beta}_P\Sigma x_{iP}y_{iP}$	$\Sigma y_{iP}^2-\hat{\beta}_P\Sigma x_{iP}y_{iP}$
	Σy_{iQ}^2	$\Sigma x_{iQ}y_{iQ}$	Σx_{iQ}^2	$\hat{\beta}_Q$	$\hat{\beta}_Q\Sigma x_{iQ}y_{iQ}$	$\Sigma y_{iQ}^2-\hat{\beta}_Q\Sigma x_{iQ}y_{iQ}$
	⋮	⋮	⋮	⋮	⋮	⋮
	Σy_{iS}^2	$\Sigma x_{iS}y_{iS}$	Σx_{iS}^2	$\hat{\beta}_S$	$\hat{\beta}_S\Sigma x_{iS}y_{iS}$	$\Sigma y_{iS}^2-\hat{\beta}_S\Sigma x_{iS}x_{iS}y_{iS}$
Σ	$\Sigma\Sigma y_{ik}^2$	$\Sigma\Sigma x_{ik}y_{ik}$	$\Sigma\Sigma x_{ik}^2$		$\Sigma\hat{\beta}_k\Sigma x_{ik}y_{ik}$	SS_s
A						
E	$\Sigma\Sigma y_{ik}^2$	$\Sigma\Sigma x_{ik}y_{ik}$	$\Sigma\Sigma x_{ik}^2$	$\tilde{\beta}$	$\tilde{\beta}\Sigma\Sigma x_{ik}y_{ik}$	SS_p
T	$\Sigma\Sigma y_T^2$	$\Sigma\Sigma x_Ty_T$	$\Sigma\Sigma x_T^2$	$\bar{\beta}$	$\bar{\beta}\Sigma\Sigma x_Ty_T$	SS_u

こうしてえたものをすべての $k=P,\ Q,\ \cdots,\ S$ について合計してつぎのものをえる。

$$\Sigma\Sigma y_{ik}^2$$

$$\Sigma\Sigma x_{ik}y_{ik}$$

$$\Sigma\Sigma x_{ik}^2$$

A行

このつくり方についてはあとで述べる。

E行

うえでえたΣ行をそのまま転記してE行をつくる。

T行

観測値の総平均からの偏差をつぎにあらわしておく。混乱のおそれがないので，添字 i，j を省略しておいた。

$$y_T=Y_{ik}-\bar{Y}_{..}=Y_{ik}-\Sigma\Sigma Y_{ik}/rn$$

$$x_T=X_{ik}-\bar{X}_{..}=X_{ik}-\Sigma\Sigma X_{ik}/rn$$

$$\Sigma\Sigma y_T^2=\Sigma\Sigma(Y_{ik}-\bar{Y}_{..})^2$$

$$=\Sigma\Sigma Y_{ik}^2-rn\bar{Y}_{..}{}^2$$

$$=\Sigma\Sigma Y_{ik}^2-(\Sigma\Sigma Y_{ik})^2/rn$$

$$\Sigma\Sigma x_Ty_T=\Sigma\Sigma(X_{ik}-\bar{X}_{..})(Y_{ik}-\bar{Y}_{..})$$

$$=\Sigma\Sigma X_{ik}Y_{ik}-rn\bar{X}_{..}\bar{Y}_{..}$$

$$=\Sigma\Sigma X_{ik}Y_{ik}-(\Sigma\Sigma X_{ik})(\Sigma Y_{ik})/rn$$

$$\Sigma\Sigma x_T^2=\Sigma\Sigma(X_{ik}-\bar{X}_{..})^2$$

$$=\Sigma\Sigma X_{ik}^2-rn\bar{X}_{..}$$

$$=\Sigma\Sigma X_{ik}^2-(\Sigma\Sigma X_{ik})^2/rn$$

A行の成分はT行の成分より，同じ列にあるE行の成分を差引くことによってえることができる。これはここで当面必要でないので，空欄のままにしておいた。

こうして表2，分散分析表の左半分（縦の分割線より左のもの）がえられ

る。

表の右半分はつぎのようにしてつくる。

列 (4) はつぎの関係式によりもとめられる。

列(4)＝列(2)/列(3)

例

$$\hat{\beta}_p = \Sigma x_{ip}^2 / \Sigma x_{ip} y_{iP}$$

列 (5) はつぎの関係式によってもとめられる。

列(5)＝列(4)×列(2)

例

$$\hat{\beta}_p \times \Sigma x_{ip} y_{ip}$$

列 (6) はつぎの関係式によってもとめられる。

列(6)＝列(1)－列(5)

例

$$\Sigma y_{ip}^2 - \hat{\beta}_p \Sigma x_{iP} y_{ip}$$

とくに SS の推定値はつぎのようにしてもとめられる。

$$SS_s = \Sigma\Sigma y_{ik}^2 - \Sigma\hat{\beta}_k \Sigma x_{ik} y_{ik}$$

$$SS_p = \Sigma\Sigma y_{ik}^2 - \tilde{\beta}\Sigma\Sigma x_{ik} y_{ik}$$

$$SS_u = \Sigma\Sigma y_T^2 - \bar{\beta}\Sigma\Sigma x_T y_T$$

この結果の検定における利用法は拡大観測データ行列によるものと同じである。

なお自由度は，紙幅の都合で省いておいた。

拡大観測データ行列による方法

共分散分析のための拡大観測データ行列による計算方法を表 3 に述べておいた。

はじめに拡大観測行列 D を 3 種類準備する。個別 s-ケース，平均 p-ケース，一体 u-ケースのものがそれである。つぎにこのそれぞれのケースのものについて交叉積内積の行列 $D'D$ をもとめる。それについてガウス・ショルダ

Y_{1P} … Y_{nP} | Y_{1Q} … Y_{nQ} | … | Y_{1S} … Y_{nS}

$$SS[Y]=\Sigma\Sigma Y^2$$

$$SS_s=SS[Y]-\Sigma(\Sigma Y_{ik})^2/n-\Sigma\left[\frac{\{\Sigma X_{ik}Y_{ik}-(\Sigma X_{ik})(\Sigma Y_{ik})/n\}^2}{\Sigma X_{ik}^2-(\Sigma X_{ik})^2/n}\right]$$

$$=SS[Y]-\Sigma(\Sigma Y_{ik})^2/n-\Sigma\hat{\beta}_k\{\Sigma X_{ik}Y_{ik}-(\Sigma X_{ik})(\Sigma Y_{ik})/n\}$$

$$SS_p=SS[Y]-\Sigma(\Sigma Y_{ik})^2/n-\frac{\{\Sigma\Sigma XY-\Sigma(\Sigma X_{ik})(\Sigma Y_{ik})/n\}^2}{\Sigma X^2-\Sigma(\Sigma X_{ik})^2/n}$$

$$=SS[Y]-\Sigma(\Sigma Y_{ik})^2/n-\tilde{\beta}\{\Sigma\Sigma XY-\Sigma(\Sigma X_{ik})(\Sigma Y_{ik})/n\}$$

$$SS_u=SS[Y]-(\Sigma\Sigma Y)^2/rn-\frac{\{\Sigma\Sigma XY-(\Sigma\Sigma X)(\Sigma\Sigma Y)/rn\}^2}{\Sigma\Sigma X^2-(\Sigma\Sigma X)^2/rn}$$

$$=SS[Y]-(\Sigma\Sigma Y)^2/rn-\bar{\beta}\{\Sigma\Sigma XY-(\Sigma\Sigma X)(\Sigma\Sigma Y)/rn\}$$

$\Sigma\Sigma Y$ | $\Sigma\Sigma XY$ | $SS[Y]$

3行

注） ここに平行 p-ケース，一体 u-ケースのDを記してあるが，実際上は省くことができる。すなわち個別 s-ケースのD よりつくった $D'D$ を利用して平行 p-ケース，一体 u-ケースの $D'D$ をつくって掃き出しのステップへ移ればよい。ダブル・サム $\Sigma\Sigma$ をもとめるには最初に $i=P, Q, \cdots, S$ について合計する。この場合変数の添字は省略しておいた。

対称行列においては，左下の成分の記載を省いておいたものがある。自明のゼロ成分も記載を省いた。

$\Sigma\Sigma Y/n$ | $(\Sigma\Sigma Y)/rn$ | $^2/rn$

$\bar{\alpha}$ | $\bar{\beta}$ | SS_u

$-2=DF_u$

ンの掃出し法を用いて，ガウス階段行列をもとめる。ただし表3においては掃き出しのはじめのラウンド（Yの平均が算出される段階）と最後のラウンド（その最後の縦列成分のみ）を記し，中間ラウンドの記載は省いておいた。

計算の手順は2元分散分析についてと基本的に同じである。そこに述べたことを参照されたい。

文　　献

観測データ行列による方法は，主としてつぎのものによった。

Dunn, O. J. and Clark, V. A., 1974, *Applied Statistics: Analysis of Variance and Regression.*

ダン，クラーク共著，中村慶一訳，1975。『応用統計学——分散分析と回帰分析——』