ダミー変数と数量化法への応用の訂正

青　山　博　次　郎

(1965年11月受付)

Errata on "Dummy Variable and Its Application to the Quantification Methods"

Hirojiro Aoyama

Instead of (30) (Vol. 16, No. 1, 1965, p. 6) read as follows:

$$D^2(\hat{\alpha}_i) \leqq \frac{(\sum_{j,k} \alpha_j \alpha_k \sigma_{jk})^2}{2n(1-\rho_i^2)(\sum_j \alpha_j \sigma_{ij})^2}$$

where ρ_i is the multiple correlation coefficient between $\hat{\alpha}_i$ and other $\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_p$.

Institute of Statistical Mathematics

前号の著者の正準相関係数を用いる方法は既に昭和28年統計数理研究所「工業統計」講座に於て，赤池弘次氏が相関分析法についての説明中に述べられているという御注意があったので，著者の不注意をおわびします．

なお6頁に述べた $\hat{\alpha}_i$, $\hat{\beta}_j$ などの標本分散については，次のように訂正します．(繁雑さをさけるために ^ 印は省略)．

$$2\sum_{j,k} \sigma_{jk}\alpha_j \varDelta\alpha_k + \sum_{j,k} \alpha_j\alpha_k \varDelta\sigma_{jk} = 0 \qquad j, k = 1, 2, \cdots, p \tag{1}$$

いま

$$f_i = 2\sum_j \alpha_j \sigma_{ij} \tag{2}$$

とおき，(1) 式を2乗ずると，

$$\sum_{j,k} f_j f_k \varDelta\alpha_j \varDelta\alpha_k = (\sum_{j,k} \alpha_j \alpha_k \varDelta\sigma_{jk})^2$$

α_j, σ_{ij} などは停留値に近いとして一定と考え，期待値をとると，

$$\sum_{i,j} f_j f_k \operatorname{cov}(\alpha_j, \alpha_k) = \frac{A}{n} \tag{3}$$

ただし

$$A = 2(\sum_{i,j} \alpha_j \alpha_k \sigma_{jk})^2$$

(3) 式の左辺は，$D(\alpha_j) = x_j$, $\rho_{jk} = \rho(\alpha_j, \alpha_k)$ とおくと，x_j についての正値二次形式であり，

$$\varDelta = \begin{vmatrix} f_1^2 & f_1 f_2 \rho_{12} & \cdots & f_1 f_p \rho_{1p} & 0 \\ f_1 f_2 \rho_{12} & f_2^2 & \cdots & f_2 f_p \rho_{2p} & 0 \\ \cdots & \cdots & \cdots & \cdots & \\ f_1 f_p \rho_{1p} & f_2 f_p \rho_{2p} & \cdots & f_p^2 & 0 \\ 0 & 0 & \cdots & 0 & -\dfrac{A}{n} \end{vmatrix} = -\frac{A}{n} f_1^2 \cdots f_p^2 R < 0$$

となるから，（3）は p 次元の超楕円体であることが分る．ここで R は相関行列式である．

そこで $D(\alpha_j)$ の値を評価するために，超平面

$$a_1x_1+a_2x_2+\cdots+a_px_p+b=0$$

が（3）に接する条件より

$$\begin{vmatrix} f_1^2 & f_1f_2\rho_{12} & \cdots & f_1f_p\rho_{1p} & 0 & a_1 \\ f_1f_2\rho_{12} & f_2^2 & \cdots & f_2f_p\rho_{2p} & 0 & a_2 \\ \cdots & \cdots & \cdots & \cdots & & \\ f_1f_p\rho_{1p} & f_2f_p\rho_{2p} & \cdots & f_p^2 & 0 & a_p \\ 0 & 0 & \cdots & 0 & -\dfrac{A}{n} & b \\ a_1 & a_2 & \cdots & a_p & b & 0 \end{vmatrix}=0 \tag{4}$$

特に $x_1=-\dfrac{b}{a_1}$ が接平面となるときは，（4）式において $a_2=\cdots=a_p=0$，$-\dfrac{b}{a_1}$ の代りに x_1 をおいて（4）を展開すると，

$$x_1^2=\frac{A}{nf_1^2}\cdot\frac{R_{11}}{R} \tag{5}$$

が得られる．

ここで R_{11} は R の (1, 1) 要素の余因数を示す．

いま α_1 と $\alpha_2,\cdots,\alpha_p$ の重相関係数を ρ_1 で示すと，

$$1-\rho_1^2=\frac{R}{R_{11}}$$

であるから，（5）式は

$$x_1^2=\frac{A}{nf_1^2(1-\rho_1^2)}$$

となる．

このときの x_1 の値は最大の $D(\hat{\alpha}_1)$（以下再び＾印を用いる）に対応しているので，結局一般には，

$$D^2(\hat{\alpha}_i)\leqq\frac{A}{nf_i^2(1-\rho_i^2)}=\frac{(\sum_{j,k}\alpha_i\alpha_k\sigma_{jk})^2}{2n(1-\rho_i^2)(\sum_j\alpha_j\sigma_{ij})^2} \tag{6}$$

が得られる（$\hat{\beta}_i$ についても同様である）．

7 頁の例に対しては

$$\hat{\sigma}_{12}=-\frac{42}{400},\quad \hat{\sigma}_{13}=-\frac{42}{400},\quad \hat{\sigma}_{14}=\frac{26}{400},\quad \hat{\sigma}_{15}=-\frac{26}{400}$$

$$\hat{\sigma}_{23}=-\frac{49}{400},\quad \hat{\sigma}_{24}=-\frac{3}{400},\quad \hat{\sigma}_{25}=\frac{3}{400},\quad \hat{\sigma}_{34}=-\frac{23}{400}$$

$$\hat{\sigma}_{35}=\frac{23}{400},\quad \hat{\sigma}_{45}=-\frac{99}{400}$$

これらを用いて $\hat{\beta}_i$ についての評価を行ってみると，

$$D^2(\hat{\beta}_1)\leqq\frac{1.435}{n(1-\rho_1^2)},\qquad D^2(\hat{\beta}_2)\leqq\frac{22.951}{n(1-\rho_2^2)}$$

$$D^2(\hat{\beta}_3)\leqq\frac{0.918}{n(1-\rho_3^2)},\qquad D^2(\hat{\beta}_4)\leqq\frac{2.550}{n(1-\rho_4^2)}$$

$\hat{\beta}_1$ の $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$ に対する重相関係数 ρ_1 などは計算できないので，n, ρ_i の 2, 3 の値につ

いて評価してみると次表のようになる.

$\max D(\hat{\beta}_1)$ の表

ρ_1 \ n	10	20	40	60	80	100
0	0.376	0.268	0.188	0.154	0.133	0.119
0.8	0.632	0.446	0.316	0.258	0.224	0.199
0.9	0.866	0.615	0.434	0.354	0.308	0.275

$\max D(\hat{\beta}_2)$ の表

ρ_2 \ n	10	20	40	60	80	100
0	1.515	1.070	0.757	0.618	0.535	0.479
0.8	2.525	1.785	1,261	1.030	0.892	0.798
0.9	3.472	2.455	1.740	1.420	1.227	1.099

$\hat{\beta}_1=-0.46809$, $\hat{\beta}_2=-0.01520$ としたので, n が小さいときは ρ_i の影響をうけてかなり誤差は大きくなる.

また 10 頁の例については, 同様にして

$$D^2(\hat{\beta}_i)\leqq\frac{4.0815}{n\bar{l}(1-\rho_i^2)} \qquad (i=1, 2, 3, 5)$$

$$D^2(\hat{\alpha}_i)\leqq\frac{0.9524}{n\bar{l}(1-\kappa_i^2)} \qquad (i=1, 3)$$

ここで ρ_i は $\hat{\beta}_i$ の他の$\hat{\beta}_j$ に関する重相関係数, κ_i は $\hat{\alpha}_i$ の他の $\hat{\alpha}_j$ に関する重相関係数を示す.

以上の計算では ρ_i が不明なので, 7 頁の例について母集団の大きさを例示した数値をすべて 50 倍した $N=1000$ とおき, これより大きさ $n=20$ のサンプルを 100 組抽出し, その平均と標準偏差 (例示したものと数値を合せるため すべて 10 倍してある) を求めたところ, 次のようになった.

$$\bar{\beta}_1=-0.45221, \quad s_{\beta_1}=0.58368, \quad \rho_1=0.68427$$
$$\bar{\beta}_2=-0.03640, \quad s_{\beta_2}=0.39405, \quad \rho_2=0.59311$$
$$\bar{\beta}_3=0.83482, \quad s_{\beta_3}=0.27896, \quad \rho_3=0.20734$$
$$\bar{\beta}_4=-0.28546, \quad s_{\beta_4}=0.52689, \quad \rho_4=0.77924$$

これらの s_{β_i} と, 上述の計算で求めた $D(\hat{\beta}_i)$ に ρ_i の値を代入したものとを比較してみると,

$$D(\hat{\beta}_1)\leqq 0.367, \qquad D(\hat{\beta}_2)\leqq 1.330$$
$$D(\hat{\beta}_3)\leqq 0.219, \qquad D(\hat{\beta}_4)\leqq 0.569$$

となり, かなり良い評価が得られていることが分る.

最後に標本誤差の評価について御注意を戴いた赤池弘次氏に厚く感謝の意を表します.

統計数理研究所