線型回帰推定値のための
尸化法に関する注意

附属技術員養成所　遠藤健児

さきに"相関を利用する推定法"（本講究録，第四巻第2号）に於て regression type 或は ratio type の推定値の平均自乗誤差の主要項である。(2,9) の V_1 及 (2,10) の V_2 を基にして尸化の効果について述べ単に $V_1 \geq V_2$ となることを幾何学的に示した。此処では更に $V_1 - V_2$ の組成を直接に求めて，尸化のための対照を明らかにしよう。

先づ，V_1，V_2 は夫々尸化しない場合，尸化して比例割当法に依った場合の誤差の主要項であって

$$(2,9)\qquad N^2 V_1 = N\left(\frac{N}{n}-1\right)\sigma^2_{X^{(1)}}(1-\rho^2)$$

$$(2.10)\qquad N^2 V_1 = \left(\frac{N}{n}-1\right)\sum_h N_h \sigma^2_{hX^{(1)}}(1-\rho_h^2)$$

に依って与えられたものである。此処で ρ は変数 $X^{(1)}$ の他に対する重相関係数で，特に変数が2個の場合は相関係数である。$N\sigma^2_{X^{(1)}}(1-\rho^2)$ は残差平方和（N_X 回帰平面のまわりの分散）で，分散行列式を $V = |C_{\mu\nu}|$　$(\mu,\nu = 1, 2, ----, k)$

$$C_{\mu\nu} = \frac{1}{N}\sum(X_i^{(\mu)} - \overline{X}^{(\mu)})(X_i^{(\nu)} - \overline{X}^{(\nu)})$$ の余因子を $V_{\mu\nu}$，

$\beta_\mu \equiv \dfrac{\overline{V}_{1\mu}}{\overline{V}_{11}}$ （$X^{(1)\mu}$ の $X^{(\mu)}$ に対する偏回帰係数）とすれば $N\dfrac{\overline{V}}{\overline{V}_{11}}$ で表わされ，層 h での対応する量に添字 h をつけて表せば（2,9）及（2,10）は夫々

$$N^2\overline{V}_1 = \left(\frac{N}{n}-1\right)N\frac{\overline{V}}{\overline{V}_{11}}$$

$$N^2\overline{V}_2 = \left(\frac{N}{n}-1\right)\sum N_h\frac{\overline{V}_h}{\overline{V}_{h,11}}$$

となる。従って

(1) $$N^2(\overline{V}_1-\overline{V}_2) = \left(\frac{N}{n}-1\right)\left\{N\frac{\overline{V}}{\overline{V}_{11}} - \sum_h N_h\frac{\overline{V}_h}{\overline{V}_{h,11}}\right\}$$

処で

(2) $$N\frac{\overline{V}}{\overline{V}_{11}} = \sum_{i=1}^{N}\left\{\sum_\mu \beta_\mu(X_i^{(\mu)}-\overline{X}^{(\mu)})\right\}^2$$

$$= \sum_h\sum_{i=1}^{N_h}\left\{\sum_\mu \beta_\mu(X_{hi}^{(\mu)}-\overline{X}^{(\mu)})\right\}^2$$

(3) $$N\frac{\overline{V}_h}{\overline{V}_{h,11}} = \sum_{i=1}^{N_h}\left\{\sum_\mu \beta_{h,\mu}(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)})\right\}^2 \qquad (h=1,2\cdots)$$

であって，一方

(4) $$\sum_i\left\{\sum_\mu \beta_\mu(X_{hi}^{(\mu)}-\overline{X}^{(\mu)})\right\}^2 = \sum_i\left\{\sum_\mu \beta_\mu(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)})\right\}^2 + N_h\left\{\sum_\mu \beta_\mu(\overline{X}_h^{(\mu)}-\overline{X}^{(\mu)})\right\}^2$$

$$\sum_{i=1}^{N_h}\left\{\sum_{\mu}\beta_\mu\left(\overline{X}_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}^2$$

$$=\sum_i\left\{\sum_\mu\beta_{h,\mu}\left(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}+\sum_i\left\{\sum_\mu(\beta_\mu-\beta_{h,\mu})\left(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}^2$$

$$+2\sum_i\sum_{\mu,\nu}\beta_{h,\mu}(\beta_\gamma-\beta_{h,\nu})\left(X_{hi}^{(\mu)}-X_h^{(\mu)}\right)\left(X_{hi}^{(\nu)}-\overline{X}_h^{(\nu)}\right)$$

で，最後の項を 2K とおけば…

$$\sum_\mu\beta_{h,\mu}C_{h,\mu\nu}=\overline{V}_h\delta_{\nu 1},\quad \beta_1=1 \quad \text{だから}$$

$$K=N_h\sum_{\mu,\nu}\beta_{h,\mu}\beta_\nu C_{h,\mu\nu}-N_h\sum_{\mu,\nu}\beta_{h,\mu}\beta_{h,\nu}C_{h,\mu,\nu})$$

$$=N_h\left\{\beta_1\frac{\overline{V}_h}{\overline{V}_{h,11}}-\frac{\overline{V}_{h,11}\overline{V}_h}{\overline{V}_{h,11}^2}\right\}=0$$

故に

$$\sum_{i=1}^{N_h}\left\{\sum_\mu\beta_\mu\left(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}^2$$

$$\text{(5)}\qquad =\sum_i\left\{\sum_\mu\beta_{h,\mu}\left(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}^2+\sum_i\left\{\sum_\mu(\beta_\mu-\beta_{h,\mu})\left(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}^2$$

(2) (3) (4) (5) を順次に (1) に代入すれば

$$\text{(6)}\qquad N^2(\overline{V}_1-\overline{V}_2)=\left(\frac{N}{n}-1\right)\left[\sum_h N_h\left\{\sum_\mu\beta_\mu\left(\overline{X}_h^{(\mu)}-\overline{X}^{(\mu)}\right)\right\}^2\right.$$
$$\left.+\sum_h\sum_{i=1}^{N_h}\left\{\sum_\mu(\beta_\mu-\beta_{h,\mu})\left(X_{hi}^{(\mu)}-\overline{X}_h^{(\mu)}\right)\right\}^2\right]\geqq 0$$

扨て $(x^{(1)},x^{(2)},\cdots,x^{(k)})$ を k 次元空间の点とすれば

$$\text{(7)}\qquad \sum_\mu\beta_\mu\left(x^{(\mu)}-\overline{X}^{(\mu)}\right)=0$$

(8) $\sum_{\mu} \beta_{h,\mu} (x^{(\mu)} - \overline{X}_h^{(\mu)}) = 0 \qquad (h = 1, 2, \cdots)$

は夫々毋集団，及び各戸の最小自乗法的回帰平面であること，及び $\beta_1 = \beta_{h.1} = 1 \quad (h = 1, 2, \cdots)$ なることに注意すれば(6)の〔 〕内の二つの項は模型的に下図の如く示される。

即ち第一項は戸 h に於ける平均点 $M_h = (\overline{X}_h^{(1)}, \overline{X}_h^{(2)}, \cdots, \overline{X}_h^{(k)})$ からその戸の回帰平面(8)迄の $x^{(1)}$ 軸に平行な距離の平方と各の戸についての加重和，第二項は戸 h に於て $(X_{hi}^{(1)}, X_{hi}^{(2)}, \cdots X_{hi}^{(k)})$ を通る $x^{(1)}$ 軸に平行な直線が，平均点 M_h を通る(7)と平行な平面と(8)とを切る二点間の距離の平方の，すべての点についての和を，すべての戸について加えたものとなる。

$\sigma^2_{x^{(1)}} (1 - \rho^2)$ は $x^{(1)}$ の分散と云う量の拡張と考えればこの結果は $\sigma^2_{x^{(1)}} (1 - \rho^2)$ の組成を示すもので

$\frac{1}{N} \sum N_h \frac{V_h}{V_{h.11}}$ は戸内分散，〔 〕内の第一項の $\frac{1}{N}$ の

$$\frac{1}{N} \sum_h N_h \left\{ \sum_{\mu} \beta_{\mu} (\overline{X}_h^{(\mu)} - \overline{X}^{(\mu)}) \right\}^2$$ は

戸間分散に対応する量である。

此処ですべての戸について

$$\sum_{\mu} \beta_{\mu} (\overline{X}_h^{(\mu)} - \overline{X}^{(\mu)}) = 0$$

であることはすべての平均点 M_h が毋回帰平面上にあることを意味し，又

$$\sum_{\mu} (\beta_{\mu} - \beta_{h,\mu}) (X_{hi}^{(\mu)} - \overline{X}_h^{(\mu)}) = 0 \qquad (i = 1, 2, \cdots, N_h)$$

であることは一般に $N_h \geq k$ である限り M_h を通って(7)と平行な平面と(8)とが $(k + 1)$ 個以上の点を共有することを意味し，従ってこの平面が一致することを示す。

故に（6）は母回帰平面と，各層での回帰平面とが一致するとき，又そのときに限って0となる。　逆に云えば層化して多くの効果が期待出来るためには（i）M_hが母回帰平面を距るように，（ii）層内の回帰平面と母回帰平面との喰い違いが大きいように層化しなければならない。　此処で更に（iii）層毎の線型回帰性が著しければそれだけ精度が上ることは云うまでもない。