＜正準相関分析（統計）＞

(1) 正準相関分析とは

q個の変量$(x_1, x_2, \ldots, x_q)$があるとき、この内のr個の変量の組$(x_1, x_2, \ldots, x_r)$と、q−r個の変量の組$(x_{r+1}, x_{r+2}, \ldots, x_q)$との関係を知りたい場合、これらの各組の変量の線形結合

$y = a_1x_1 + a_2x_2 + \ldots + a_rx_r$　　　$z = b_1x_{r+1} + b_2x_{r+2} + \ldots + b_{q-r}x_q$　　を考え

yとzの間の相関係数$r_{y,z}$を最大にするように係数$a_1, a_2, \ldots, a_r; b_1, b_2, \ldots, b_{q-r}$を推定するのが正準相関分析である。
これによって、第1の組と第2の組の関係の程度を知ろうとするものである。ここで、$r \leqq q-r$とする。

(2) 係数の求め方

q個の変量$(x_1, x_2, \ldots, x_q)$がN個の標本について測定されているとする。
それらの測定値から、標本の分散共分散行列Σを求める。q×q行列であるΣの小行列を考える。

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$s = q - r$ として、Σ_{11}は$r \times r$, Σ_{12}は$r \times s$, Σ_{21}は$s \times r$, Σ_{22}は$s \times s$行列である。
$T = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$としたとき、固有方程式 $Ta = \lambda a$ の固有値の最大値が求める相関係数の自乗に対応する。

係数ベクトル　$a=\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix}$　と　$b=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{pmatrix}$　は　$\begin{cases} Ta = \lambda a \\ a^T\Sigma_{11}a = 1 \end{cases}$　と　$b = \dfrac{1}{\sqrt{\lambda}}\Sigma_{22}^{-1}\Sigma_{21}a$　より求める。

(3) 計算例1

右の表はある集団の身長、座高、体重、胸囲のデータに対する分散共分散行列である。
身長、座高を第1組の変量、体重、胸囲を第2組の変量として、第1組と第2組の正準相関係数を求める。

	x_1	x_2	x_3	x_4
身長 x_1	26.76	11.67	16.91	6.38
座高 x_2	11.67	8.81	10.57	4.69
体重 x_3	16.91	10.57	32.45	17.39
胸囲 x_4	6.38	4.69	17.39	15.07

$$\Sigma = \begin{pmatrix} 26.76 & 11.67 & 16.91 & 6.38 \\ 11.67 & 8.81 & 10.57 & 4.69 \\ 16.91 & 10.57 & 32.45 & 17.39 \\ 6.38 & 4.69 & 17.39 & 15.07 \end{pmatrix}$$

小行列　$\Sigma_{11}=\begin{pmatrix} 26.76 & 11.67 \\ 11.67 & 8.81 \end{pmatrix}$　$\Sigma_{12}=\begin{pmatrix} 16.91 & 6.38 \\ 10.57 & 4.69 \end{pmatrix}$

$\Sigma_{21}=\begin{pmatrix} 16.91 & 10.57 \\ 6.38 & 4.69 \end{pmatrix}$　$\Sigma_{22}=\begin{pmatrix} 32.45 & 17.39 \\ 17.39 & 15.07 \end{pmatrix}$

$T=\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$　よって　$T=\begin{pmatrix} 0.191532 & 0.104693 \\ 0.423092 & 0.270868 \end{pmatrix}$

Tは非対称行列であるので固有値は行列式を解いて求める。

$\det(T-\lambda E)=0$　　ただし　$E=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

区間指定法より　$\lambda = 0.0170312423897269$　　$\lambda = 0.445367788744155$

大きい方の値をλに定めると正準相関係数は　$\sqrt{\lambda}=0.6674$

次に係数ベクトルaを計算する。$A=T-\lambda E$　として　$Au=0$　の解を求める。

今の場合、2次元であるので手計算で求めることもできるが、ここでは一般的な方法を用いる。
線形方程式の解法に特異値分解を利用する。

$U=0$　$V=0$　$w=svd(A,U,V)$　$u=V_{*,2}$　$u=\begin{pmatrix} -0.381284169721717 \\ -0.924457885422381 \end{pmatrix}$

符号をかえて　$u=-u$　$a=\dfrac{1}{\sqrt{u^T\Sigma_{11}u}}u$　従って　$a=\begin{pmatrix} 0.0860 \\ 0.2086 \end{pmatrix}$　$b=\dfrac{1}{\sqrt{\lambda}}\Sigma_{22}^{-1}\Sigma_{21}a$　従って　$b=\begin{pmatrix} 0.2296 \\ -0.1131 \end{pmatrix}$

以上の計算より、正準変量は

$y = a_1x_1+a_2x_2 = 0.0860x_1+0.2086x_2$

$z = b_1x_3+b_2x_4 = 0.2296x_3-0.1131x_4$

正準相関係数は　0.6674