

『固有値と主成分分析』について

統計数理研究所(名誉教授) 鈴木義一郎

固有値と固有ベクトルの計算

<pre>itr=:3 :0 NB.Iteration m=.m +/ .*m=.m%>./ ,m=.m+/. *m=.>{:y (v%>./ v=(>{.y)+/ .*m);m) mev=:3 :0 NB. igenvalue & igenvector v=.v%+/*:v.>{itr^:20(1;y) (+/(y+/. *v)%v*#v),v) red=:3 : 'y-({.lv)**/^}.lv=.mev y' evs=:4 : 'mev^2(red^(i.x))y'</pre>	<pre>]M2=:2 2 \$ 6 2 2 3 6 2 2 3 red M2 0.4 _0.8 _0.8 1.6 mev red M2 2 _0.447214 0.894427 M3=:3 3 \$ 18 0 9 0 6 _3 9 _3 6 2 evs M3 22.9373 0.873575 _0.0848832 0.479229 7.06275 0.264951 0.908916 _0.321982</pre>
<p>NB. Principal Component Analysis NB. 左引数は主成分の個数で右引数は行列 NB. 出力は固有値と固有ベクトル</p>	

美女のプロポーション

「STYLE」という変数に、第2回から 25 回までの、「ミスユニバース日本代表」の身長・体重・バスト・ウエスト・ヒップのデータを入力する。(毎日新聞『数学雑学事典』)	
STYLE=:165 53 86 56 92	STYLE=:STYLE,164 52 87 58 90
STYLE=:STYLE,160 47 84 52 92	STYLE=:STYLE,167 53 86 59 88
STYLE=:STYLE,166 55 86 64 89	STYLE=:STYLE,169 58 89 60 90
STYLE=:STYLE,164 56 90 60 95	STYLE=:STYLE,169 51 84 60 90
STYLE=:STYLE,168 55 87 56 87	STYLE=:STYLE,166 50 86 59 87
STYLE=:STYLE,164 54 87 57 92	STYLE=:STYLE,168 53 88 60 88
STYLE=:STYLE,168 54 94 58 97	STYLE=:STYLE,165 54 88 62 90
STYLE=:STYLE,169 55 88 57 92	STYLE=:STYLE,167 50 88 58 89
STYLE=:STYLE,169 53 86 58 93	STYLE=:STYLE,170 55 88 60 90
STYLE=:STYLE,166 56 84 57 90	STYLE=:STYLE,168 57 84 62 92
STYLE=:STYLE,165 53 85 55 90	STYLE=:STYLE,168 56 85 62 94
STYLE=:STYLE,163 49 84 59 90	STYLE=:STYLE,163 52 83 60 88

相関行列を導出するプログラム

stand=:3 :0]R=:corm STYLE
NB. stadadyzation of data	1 0.542494 0.326565 0.372425 0.0115011
s=.y-(k=. \$y)\$(+/%#)y	0.542494 1 0.3131 0.449056 0.241926
s%k\$%:((<0 1) :(:s)+/ .*s){.k	0.326565 0.3131 1 0.0605705 0.424285
)	0.372425 0.449056 0.0605705 1 0.0649716
corm=:3 :'(:(:s)+/ .*s)%#s=.stand y'	0.0115011 0.241926 0.424285 0.0649716 1
NB. 相関行列を出力する関数	

相関行列による主成分分析

2 evs R	最大と次に大きな固有値の合計は「3.456」であることから、第1主成分と第2主成分だけで、5次元データの7割程度の情報を吸収していることが分かる。
2.14898 0.521033 0.570928 0.419977 0.399578 0.25792	
1.30735 0.236935 0.104455 0.483214 0.488062 0.679155	

T=: :0 1}.5 evs R [X=:stand STYLE	1(2)列目の数値が第1(2)主成分
6j2 ": X+/ .*T	
_0.74 0.94 0.08 _0.62 _0.10	_0.89 0.33 0.05 0.53 _0.41
_4.12 2.03 0.25 _0.29 0.16	_0.26 _0.95 _0.57 0.13 _0.13
0.86 _1.62 1.02 0.87 _0.30	2.18 _0.38 _0.52 _0.23 _0.81
1.37 1.79 1.28 0.56 _0.70	_0.26 _1.11 _0.22 _0.17 1.48
0.00 _0.63 _1.62 _0.68 _0.89	_1.24 _1.01 _0.79 0.88 0.20
_0.40 1.00 0.41 _0.22 _0.61	0.47 _0.84 _0.93 0.73 _0.06
2.39 3.24 _0.42 0.88 0.44	0.57 _0.41 0.62 1.16 _0.49
1.08 0.68 _0.78 _0.73 0.18	_0.62 0.05 _1.14 0.89 0.47
0.54 0.44 _0.19 _0.72 1.15	1.55 _0.56 _0.81 _0.03 0.27
_0.27 _0.44 0.22 _1.41 _0.64	1.38 _1.08 1.28 _0.90 0.29
_1.28 0.37 _0.36 _0.92 _0.41	1.55 _0.28 1.44 _0.62 0.77
_2.15 _0.25 0.75 0.67 0.52	_1.71 _1.32 0.94 0.24 _0.38
0j3 ": }.1 evs R	最大固有値に対応する固有ベクトルで、全ての数値が正数なので「ボリューム」を表す。
0.521 0.571 0.420 0.400 0.258	
0j3 ": }.2 evs R	第2主成分はヒップ、次いでバストが大きな数値、ウェストが負で、「体型の複雑さ」を表す。
_0.237 _0.104 0.483 _0.488 0.679	

第2主成分の値が群を抜いて大きいのは7行目のデータで、これに対応する原データの数値は、「168 54 94 58 97」となっていて、正に「ゴールデン・プロポーション」という感じである。実際、このデータは第8回の代表である「児島明子」のスタイルで、彼女だけが唯一ミス・ユニバースの栄冠に輝いた。逆に、第2主成分の最も小さな「_1.62」に対応する3行目のデータは「166 55 86 64 89」で、一般女性ならともかくも、いやしくも日本代表である。ちとビヤ樽すぎる感じは否めない。

主成分分析の総括プログラム

<pre>pca=:4 :0 NB. Principal Component Analysis s=.y-(k=. \$y)\$(+/%#)y s=.s%k\$%:((<0 1) : (:s)+/ .*s)%{.k ev=.x evs((:s)+/ .*s)%{.k t=. 'e-value & vector :',7j3":ev t,'pc-values :',6j2":s+/.* . :ev)</pre>	<pre>2 pca 10{.STYLE e-value & vector : 2.374 0.489 0.550 0.442 0.487 0.160 1.542 _0.182 _0.291 0.530 _0.214 0.745 pc-values : _0.78 0.07 _4.01 1.03 1.01 _1.65 1.18 1.07 0.08 _1.56 _0.41 0.13 1.88 2.48 0.88 _0.14 0.35 _0.08 _0.19 _1.35</pre>
---	---