

This dataset contains information on new born babies and their parents.

Name	Variable	Data type
x₁	Head Circumference	Scale
x₂	Length of baby (inches)	Scale
x₃	Weight of baby (lbs)	Scale
x₄	Mother age	Scale
x₅	Mothers height (inches)	Scale
x₆	Father's age	Scale
x₇	Father's height (inches)	Scale

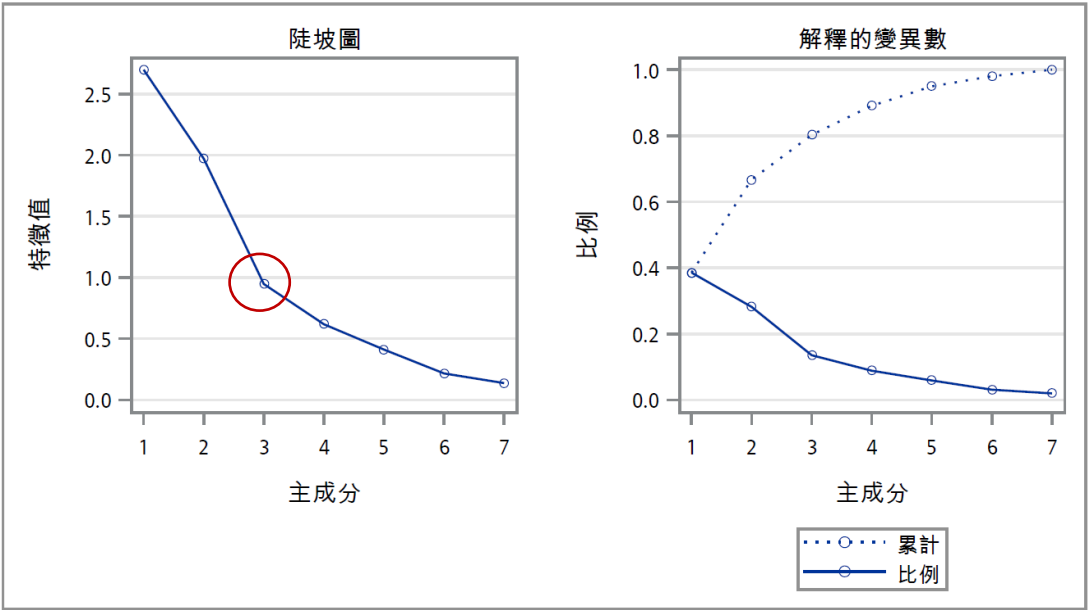
Correlation matrix:

相關矩陣							
	x1	x2	x3	x4	x5	x6	x7
x1	1.0000	0.5653	0.7364	0.1121	0.3813	0.3014	0.0405
x2	0.5653	1.0000	0.6970	-.0207	0.4147	0.0789	0.1871
x3	0.7364	0.6970	1.0000	0.0010	0.3679	0.1768	0.0248
x4	0.1121	-.0207	0.0010	1.0000	0.0468	0.8066	-.2036
x5	0.3813	0.4147	0.3679	0.0468	1.0000	-.0717	0.2728
x6	0.3014	0.0789	0.1768	0.8066	-.0717	1.0000	-.2986
x7	0.0405	0.1871	0.0248	-.2036	0.2728	-.2986	1.0000

PCA
Eigenvalues:

相關矩陣的特徵值				
	特徵值	差異	比例	累計
1	2.69864750	0.72700847	0.3855	0.3855
2	1.97163903	1.02448148	0.2817	0.6672
3	0.94715756	0.32895520	0.1353	0.8025
4	0.61820236	0.20693612	0.0883	0.8908
5	0.41126624	0.19522424	0.0588	0.9496
6	0.21604200	0.07899670	0.0309	0.9804
7	0.13704530		0.0196	1.0000

Scree plot:



由陡坡圖的肘點，取前 3 個主成分，可以解釋 80.25%原始資料的變異

Eigenvectors:

特徵向量							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
x1	0.523739	0.031800	-0.172328	0.093523	0.638979	-0.480218	-0.217576
x2	0.498277	-0.153514	-0.094940	0.178148	-0.739919	-0.373540	-0.019788
x3	0.532254	-0.059904	-0.309203	0.101569	0.043219	0.777518	-0.028139
x4	0.123959	0.604908	0.423698	-0.018037	-0.143058	0.117532	-0.636079
x5	0.362858	-0.207654	0.455600	-0.760297	0.029830	0.021526	0.195544
x6	0.200364	0.629402	0.138806	0.190116	0.030573	-0.034254	0.711469
x7	0.083342	-0.408232	0.677735	0.578515	0.141661	0.099363	0.049555

$$\text{Prin1} = 0.52x_1 + 0.50x_2 + 0.53x_3 + 0.12x_4 + 0.36x_5 + 0.20x_6 + 0.08x_7$$

$$\text{Prin2} = 0.03x_1 - 0.15x_2 - 0.06x_3 + 0.60x_4 - 0.21x_5 + 0.63x_6 - 0.41x_7$$

$$\text{Prin3} = -0.17x_1 - 0.10x_2 - 0.31x_3 + 0.42x_4 + 0.46x_5 + 0.14x_6 + 0.68x_7$$

Prin1 在 x_1, x_2, x_3 的權重大，可視為嬰兒發育狀態的指標

Prin2 在 x_4, x_6 的權重大，可視為雙親年齡的指標

Prin3 在 x_5, x_7 的權重大，可視為雙親身高的指標

PC scores:

obs	Prin1
1	-3.92934
37	4.13496

obs	Prin2
23	-2.21438
34	3.29775

obs	Prin3
17	-1.87296
4	1.85824

由主成分計分可知，

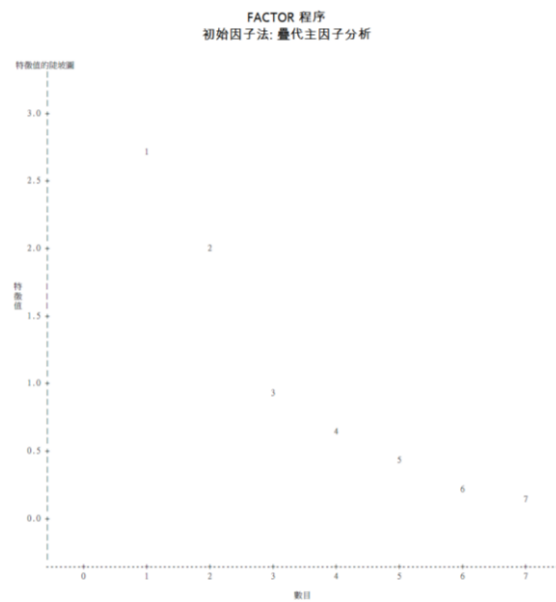
嬰兒發育狀態最好的為第 37 組觀察值，最差的為第 1 組觀察值

雙親年齡最高的為第 34 組觀察值，最低的為第 23 組觀察值

雙親身高最高的為第 4 組觀察值，最矮的為第 17 組觀察值

Factor Analysis – PC method

根據對陡坡圖的觀察，以及前述特徵值>1之主成分，取 4 個因素



Matrix of the factor loading:

L=

旋轉的因子模型			
	Factor1	Factor2	Factor3
x1	0.77509	0.16645	0.10916
x2	0.70917	0.00142	0.28840
x3	0.94837	0.00853	0.03399
x4	-0.02582	0.94918	-0.00543
x5	0.37000	0.04921	0.57599
x6	0.21376	0.85373	-0.29955
x7	0.03000	-0.20111	0.48369

透過每個因子所解釋的變異數		
Factor1	Factor2	Factor3
2.1872709	1.7004409	0.7517223

$$\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ x_3 - \mu_3 \\ x_4 - \mu_4 \\ x_5 - \mu_5 \\ x_6 - \mu_6 \\ x_7 - \mu_7 \end{bmatrix} = \begin{bmatrix} 0.78 & 0.17 & 0.11 \\ 0.71 & 0.00 & 0.29 \\ 0.95 & 0.01 & 0.03 \\ -0.03 & 0.95 & -0.01 \\ 0.37 & 0.05 & 0.58 \\ 0.21 & 0.85 & -0.30 \\ 0.03 & -0.20 & 0.48 \end{bmatrix} * \begin{bmatrix} \text{Factor1} \\ \text{Factor2} \\ \text{Factor3} \end{bmatrix} + \begin{bmatrix} 0.36 \\ 0.41 \\ 0.10 \\ 0.10 \\ 0.53 \\ 0.14 \\ 0.72 \end{bmatrix}$$

觀察各 common factor 在各變數的 loading，可以發現:

Factor1 在 x₁, x₂, x₃ 的負荷量最大，為嬰兒發育程度的因素

Factor2 在 x₄, x₆ 的負荷量最大，可視為雙親年齡的因素

Factor3 在 x₅, x₇ 的負荷量最大，可視為雙親身高的因素

總變異為 7；

Factor1 解釋了 2.187/7=31.24% 的變異，Factor2 解釋了 1.700/7=24.29% 的變異，Factor3 解釋了 0.752/7=10.74% 的變異

3 個因素共解釋了 66.27% 的變異

Communalities:

$h^2=$

最終公因子變異數估計值: 總計 = 4.639434						
x1	x2	x3	x4	x5	x6	x7
0.64039304	0.58610582	0.90064183	0.90163993	0.47108050	0.86427472	0.27529826

x5, x7 的共通性小，顯示對因素的貢獻不大

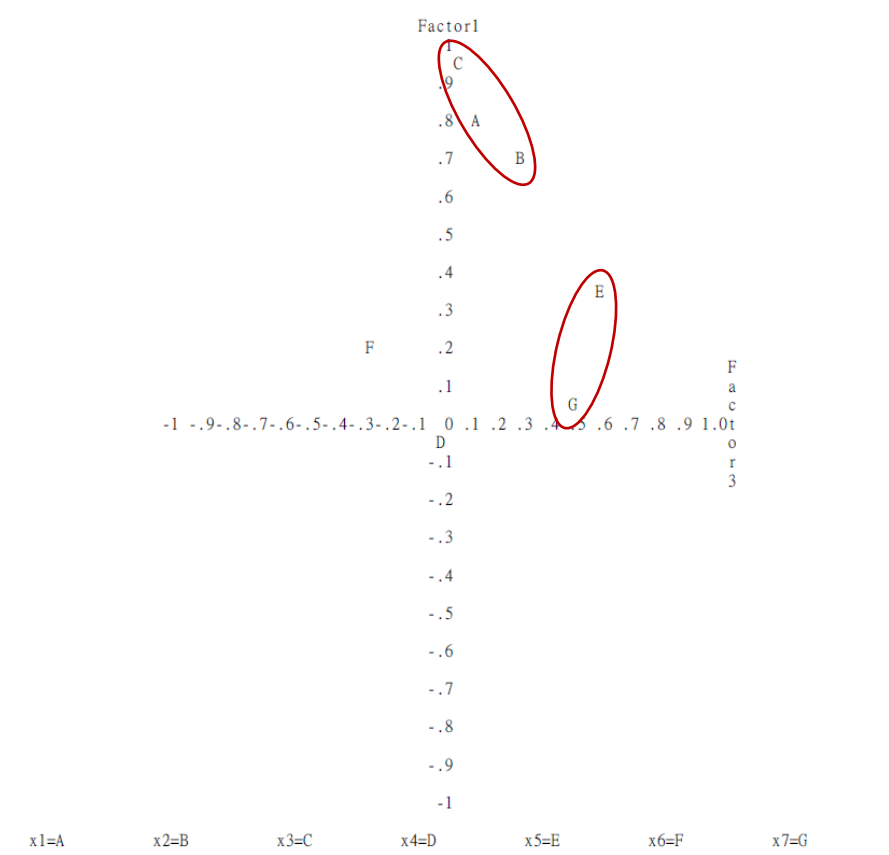
Specific variance:

$\Psi=$

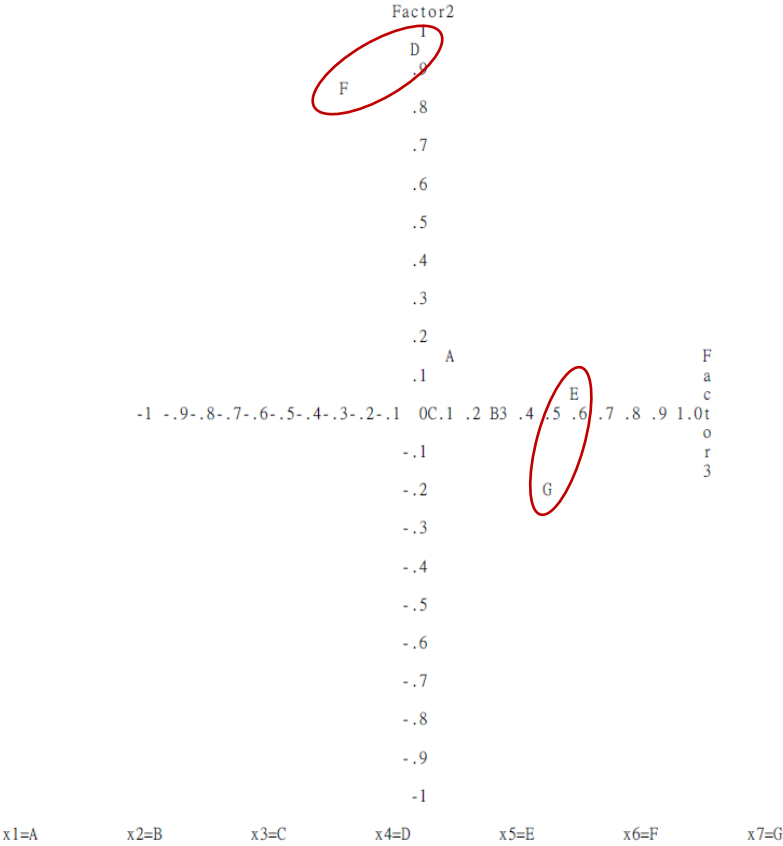
	x1	x2	x3	x4	x5	x6	x7
x1	0.35961	0	0	0	0	0	0
x2	0	0.41389	0	0	0	0	0
x3	0	0	0.09936	0	0	0	0
x4	0	0	0	0.09836	0	0	0
x5	0	0	0	0	0.52892	0	0
x6	0	0	0	0	0	0.13573	0
x7	0	0	0	0	0	0	0.72470

因子模型圖:

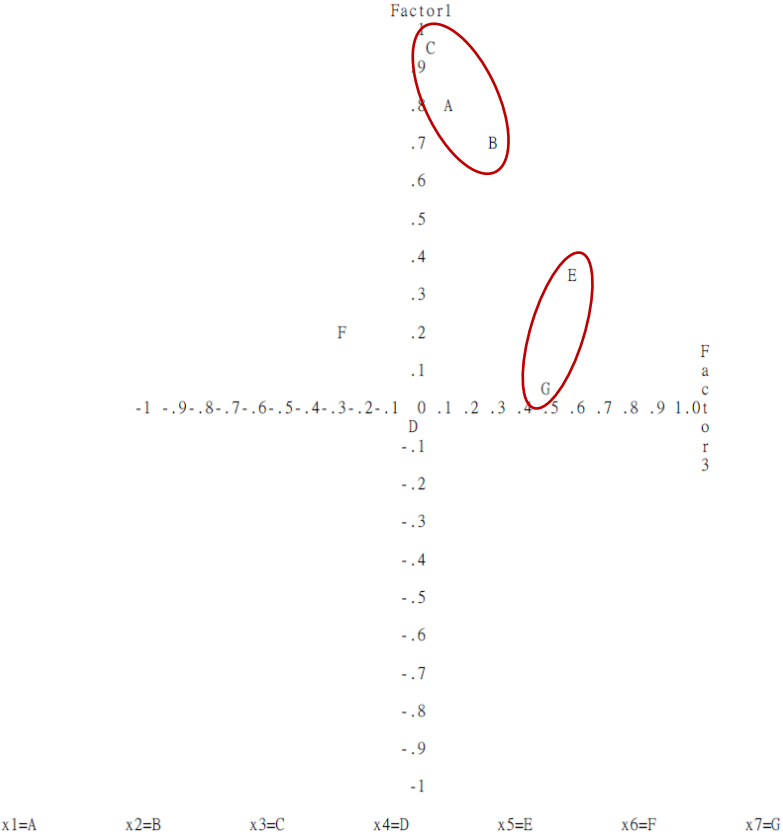
Factor1 和 Factor3 之因子模型的標繪圖



Factor2 和 Factor3 之因子模型的標繪圖



Factor1 和 Factor3 之因子模型的標繪圖



圖與前面因素模型的
結論一致

Factor scores:

obs	Factor1
4	-2.21736
37	2.06502

obs	Factor2
23	-1.30963
4	2.38930

obs	Factor3
1	-1.86785
31	1.75401

由共同因素計分可知，

嬰兒發育狀態最好的為第 37 組觀察值，最差的為第 4 組觀察值

雙親年齡最高的為第 4 組觀察值，最低的為第 23 組觀察值

雙親身高最高的為第 31 組觀察值，最矮的為第 1 組觀察值

Factor Analysis – ML method

根據 42 觀測值的顯著性檢定			
檢定	自由度	卡方	Pr > 卡方
H0: 沒有公因子	21	123.7851	<.0001
HA: 至少一個公因子			
H0: 1 個因子就足夠	14	58.8915	<.0001
HA: 需要更多個因子			

根據 42 觀測值的顯著性檢定			
檢定	自由度	卡方	Pr > 卡方
H0: 沒有公因子	21	123.7851	<.0001
HA: 至少一個公因子			
H0: 2 個因子就足夠	8	9.7611	0.2822
HA: 需要更多個因子			

Specific variance:

$\Psi =$

具有對角唯一性的殘差相關							
	x1	x2	x3	x4	x5	x6	x7
x1	0.32933	0	0	0	0	0	0
x2	0	0.41815	0	0	0	0	0
x3	0	0	0.17527	0	0	0	0
x4	0	0	0	0.33018	0	0	0
x5	0	0	0	0	0.77567	0	0
x6	0	0	0	0	0	0.00000	0
x7	0	0	0	0	0	0	0.88785

Communalities:

$h^2 =$

最終公因子變異數估計值和變數加權		
總公因子變異數: 已加權 = 10.577224 未加權 = 4.083534		
變數	公因子變異數	加權
x1	0.67066977	3.03636360
x2	0.58184659	2.39139627
x3	0.82472502	5.70528168
x4	0.66981526	3.02855021
x5	0.22433229	1.28929365

x5, x7 的共通性小，顯示對因素的貢獻不大

x6	1.00000000	Infty
x7	0.11214513	1.12633882

Matrix of the factor loading:

L=

旋轉的因子模型		
	Factor1	Factor2
x1	0.80198	0.16583
x2	0.76096	-0.05282
x3	0.90790	0.02087
x4	0.00217	0.81842
x5	0.44887	-0.15116
x6	0.17208	0.98508
x7	0.09804	-0.32021

透過每個因子所解釋的變異數		
因子	加權	未加權
Factor1	8.31106193	2.28722530
Factor2	2.26616240	1.79630876

$$\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ x_3 - \mu_3 \\ x_4 - \mu_4 \\ x_5 - \mu_5 \\ x_6 - \mu_6 \\ x_7 - \mu_7 \end{bmatrix} = \begin{bmatrix} 0.80 & 0.17 \\ 0.76 & -0.05 \\ 0.91 & 0.02 \\ 0.00 & 0.82 \\ 0.45 & -0.15 \\ 0.17 & 0.99 \\ 0.10 & -0.32 \end{bmatrix} * \begin{bmatrix} Factor1 \\ Factor2 \end{bmatrix} + \begin{bmatrix} 0.33 \\ 0.42 \\ 0.18 \\ 0.33 \\ 0.78 \\ 0.00 \\ 0.89 \end{bmatrix}$$

總變異為 7；

Factor1 解釋了 2.287/7=32.67%的變異，Factor2 解釋了 1.796/7=25.66%的變異
2 個因素共解釋了 58.32%的變異

觀察各 common factor 在各變數的 loading，可以發現:

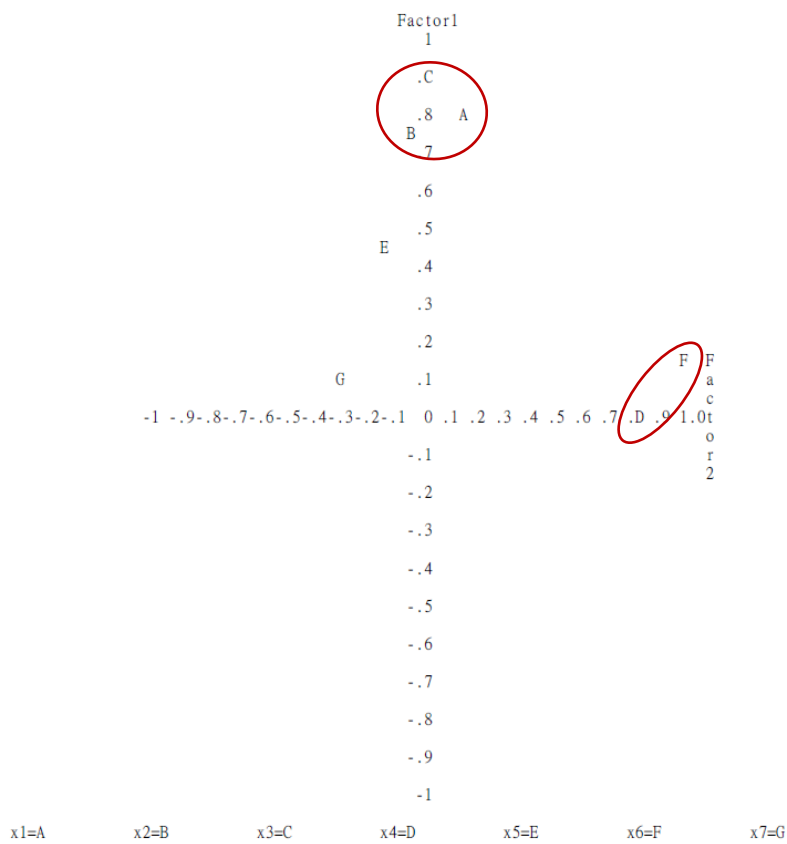
Factor1 在 x₁, x₂, x₃ 的負荷量最大，可視為嬰兒發育的指標

Factor2 在 x₄, x₆ 的負荷量最大，可是為雙親年齡的指標

2 個因素在 x₅, x₇ 的負荷量皆小，與共通性的結論相呼應

因子模型圖：

Factor1 和 Factor2 之因子模型的標繪圖



圖與前面因素模型的結論一致

Factor scores:

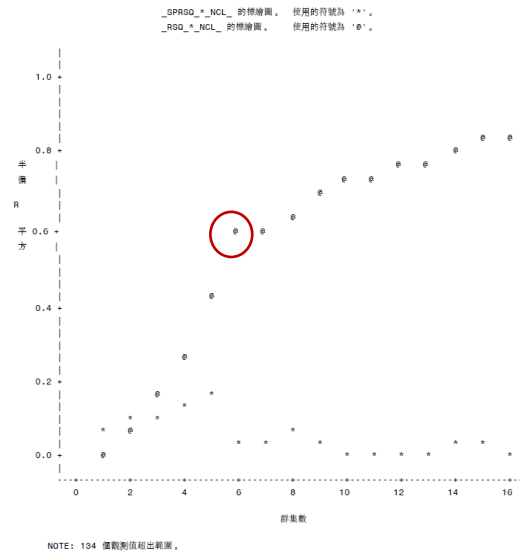
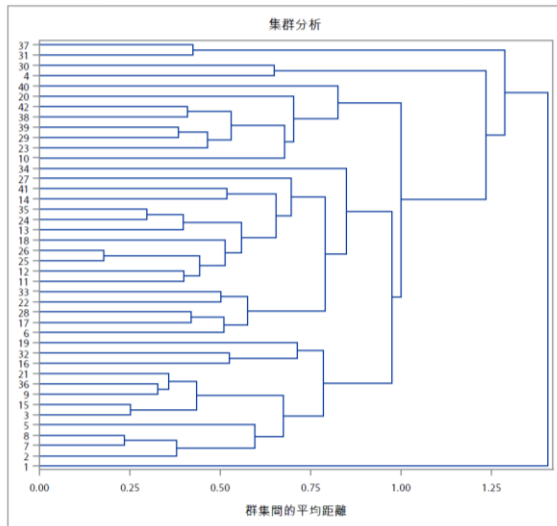
obs	Factor1
4	-2.00292
37	2.16200

obs	Factor2
23	-1.71827
34	2.45746

由共同因素計分可知，
嬰兒發育狀態最好的為第 37 組觀察值，最差的為第 4 組觀察值
雙親年齡最高的為第 34 組觀察值，最低的為第 23 組觀察值

Cluster Analysis—Centroid method

集群歷史								
群 集 數	聯結的集群		次數	新集群 RMS 標準差	半偏 R 平方	R 平方	範數 RMS 距離	繫結
...
21	16	32	2	0.5245	0.0067	.904	0.5245	
20	CL26	CL31	5	0.4943	0.0103	.894	0.5317	
19	CL23	CL33	8	0.5043	0.0170	.877	0.5589	
18	CL24	CL25	5	0.5421	0.0112	.866	0.576	
17	CL35	5	4	0.4844	0.0116	.854	0.5959	
16	4	30	2	0.6502	0.0103	.844	0.6502	
15	CL19	CL22	10	0.5624	0.0194	.824	0.6538	
14	CL17	CL28	9	0.5788	0.0330	.791	0.6745	
13	10	CL20	6	0.5618	0.0147	.777	0.6769	
12	CL15	27	11	0.5890	0.0152	.761	0.6962	
11	CL13	20	7	0.6056	0.0152	.746	0.7032	
10	CL21	19	3	0.6557	0.0143	.732	0.7123	
9	CL14	CL10	12	0.6740	0.0355	.696	0.7848	
8	CL18	CL12	16	0.6856	0.0587	.638	0.7907	
7	CL11	40	8	0.6672	0.0223	.615	0.825	
6	CL8	34	17	0.7069	0.0230	.592	0.8495	
5	CL9	CL6	29	0.8482	0.1745	.418	0.9756	
4	CL5	CL7	37	0.8980	0.1406	.277	1.001	
3	CL4	CL16	39	0.9369	0.0952	.182	1.2353	
2	CL3	CL29	41	0.9752	0.1099	.072	1.2874	
1	1	CL2	42	1.0000	0.0722	.000	1.407	



觀察上頁表格中 SP Rsq, Rsq 的變動在第 6 群時皆為最大，再本頁右上的標繪圖，可發現轉折點發生在分第 6 群時，因此可知最適當分群數為 6 群

Obs	id	CLUSTER	Obs	id	CLUSTER
1	25	1	22	9	2
2	26	1	23	36	2
3	24	1	24	21	2
4	35	1	25	2	2
5	13	1	26	16	2
6	11	1	27	32	2
7	12	1	28	5	2
8	17	1	29	19	2
9	28	1	30	29	3
10	22	1	31	39	3
11	33	1	32	38	3
12	6	1	33	42	3
13	18	1	34	23	3
14	14	1	35	10	3
15	41	1	36	20	3
16	27	1	37	40	3
17	34	1	38	31	4
18	7	2	39	37	4
19	8	2	40	4	5
20	3	2	41	30	5
21	15	2	42	1	6

從左側分群表格可知，
 編號 25、26、24、35、13、11、12、17、28、22、
 33、6、18、14、41、27、34 分為第一群；
 編號 7、8、3、15、9、36、21、2、16、32、5、19
 分為第二群；
 編號 29、39、38、42、23、10、20、40 分為第三群；
 編號 31、37 分為第四群；
 編號 4、30 分為第五群；
 編號 1 單獨為第六群

Cluster1

變數	N	平均值	標準差	最小值	最大值
x1	17	13.4705882	0.5144958	13.0000000	14.0000000
x2	17	19.9411765	0.6586528	19.0000000	21.0000000
x3	17	7.5176471	0.6454023	6.6000000	8.6000000
x4	17	28.3529412	3.3530057	23.0000000	37.0000000
x5	17	63.8235294	1.8109146	60.0000000	67.0000000
x6	17	34.6470588	4.9363597	29.0000000	46.0000000
x7	17	70.0588235	2.9679662	66.0000000	76.0000000

Cluster2

變數	N	平均值	標準差	最小值	最大值
x1	12	12.7500000	0.4522670	12.0000000	13.0000000
x2	12	19.3333333	0.8876254	18.0000000	21.0000000
x3	12	6.1250000	0.8874120	4.2000000	7.5000000
x4	12	21.4166667	2.5746433	19.0000000	27.0000000
x5	12	63.3333333	1.7752507	61.0000000	67.0000000
x6	12	23.0000000	2.2156468	20.0000000	27.0000000
x7	12	71.9166667	2.7784343	69.0000000	78.0000000

Cluster3

變數	N	平均值	標準差	最小值	最大值
x1	8	13.6250000	0.5175492	13.0000000	14.0000000
x2	8	20.8750000	0.6408699	20.0000000	22.0000000
x3	8	8.5625000	0.8568005	7.5000000	10.0000000
x4	8	20.6250000	1.9955307	18.0000000	24.0000000
x5	8	67.0000000	2.4494897	63.0000000	71.0000000
x6	8	23.2500000	3.5355339	19.0000000	30.0000000
x7	8	71.5000000	2.2038927	67.0000000	74.0000000

Cluster4

變數	N	平均值	標準差	最小值	最大值
x1	2	15.0000000	0	15.0000000	15.0000000
x2	2	22.0000000	0	22.0000000	22.0000000
x3	2	9.5000000	0.7071068	9.0000000	10.0000000
x4	2	33.5000000	2.1213203	32.0000000	35.0000000
x5	2	67.5000000	0.7071068	67.0000000	68.0000000
x6	2	34.5000000	4.9497475	31.0000000	38.0000000
x7	2	72.0000000	1.4142136	71.0000000	73.0000000

Cluster5

變數	N	平均值	標準差	最小值	最大值
x1	2	12.0000000	0	12.0000000	12.0000000
x2	2	19.0000000	1.4142136	18.0000000	20.0000000
x3	2	5.2500000	1.0606602	4.5000000	6.0000000
x4	2	39.0000000	2.8284271	37.0000000	41.0000000
x5	2	65.5000000	0.7071068	65.0000000	66.0000000
x6	2	34.0000000	4.2426407	31.0000000	37.0000000
x7	2	68.0000000	0	68.0000000	68.0000000

Cluster6

變數	N	平均值	標準差	最小值	最大值
x1	1	12.0000000	.	12.0000000	12.0000000
x2	1	17.0000000	.	17.0000000	17.0000000
x3	1	5.8000000	.	5.8000000	5.8000000
x4	1	24.0000000	.	24.0000000	24.0000000
x5	1	58.0000000	.	58.0000000	58.0000000
x6	1	26.0000000	.	26.0000000	26.0000000
x7	1	66.0000000	.	66.0000000	66.0000000

從上方表格中可發現各群的敘述統計量皆無太大區別，因此沒有顯而易見的群集特徵，且每群內的觀察值數量非平均分配，結果顯示分群的適合性、必要性不明顯