

資料整理來源：呂金河譯，多變量分析

陳順宇著，多變量分析

第六章 主成分分析(Principal Component Analysis)：

我們常需要對一組變數訂出一個總指標(或指數)，例如：表示一個國家或地區的物價指數、環境污染指數、國家競爭力等指標。為了訂出一個總指標，研究學者常會找許多相關項目(或稱變數)，然後做加權平均，但問題在於如何給定每個項目(或稱變數)的權重，一種方式是在收集資料前主觀的人為訂定，但如果能在收集資料後，透過資料本身傳達的訊息來訂定權重將會更客觀。

主成分分析由皮爾森(Pearson, 1901)提出，再由侯特齡(Hotelling, 1933)加以發展。

以兩變數為例，導出新軸與新變數(由 X_1 軸旋轉 θ) (照相角度的選取法)，希望能找出變異數最大的線性組合。

(資料來源：呂金河譯，多變量分析)

$$x_1^* = \cos\theta \times x_1 + \sin\theta \times x_2$$

$$x_2^* = -\sin\theta \times x_1 + \cos\theta \times x_2$$

→ 取 $\theta = 43.261^\circ$

$$x_1^* = 0.728 x_1 + 0.685 x_2$$

$$x_2^* = -0.685 x_1 + 0.728 x_2$$

觀察值	原始資料		均值修正資料		新變數($\theta = 43.261^\circ$)	
	X_1	X_2	X_1	X_2	X_1^*	X_2^*
1	16	8	8	5	9.253	-1.841
2	12	10	4	7	7.710	2.356
3	13	6	5	3	5.697	-1.242
4	11	2	3	-1	1.499	-2.784
5	10	8	2	5	4.883	2.271
6	9	-1	1	-4	-2.013	-3.598
7	8	4	0	1	0.685	0.728
8	7	6	-1	3	1.328	2.870
9	5	-3	-3	-6	-6.297	-2.313
10	3	-1	-5	-4	-6.382	0.514
11	2	-3	-6	-6	-8.481	-0.257
12	0	0	-8	-3	-7.882	3.298
均值	8	3	0	0	0	0
平方和			254	232	424.334	61.666
變異數	23.091	21.091	23.091	21.091	38.576	5.606

1. 觀察點投影至 X_1 、 X_2 軸得原變數的座標。
2. 新軸或新變數 X_1^* 、 X_2^* 稱為主成分，觀察點投影至 X_1^* 、 X_2^* 軸所得新的值稱為主成分計分(principal components scores)。
3. 新變數 X_1^* 、 X_2^* 為原變數 X_1 、 X_2 的線性組合，且修正後均值皆保持不變為 0。
4. 新變數 X_1^* 、 X_2^* 與原變數 X_1 、 X_2 的總平方和相同；總變異數也相同。
5. X_1^* 解釋總變異的百分比一定比任何一個原始變數 X_1 或 X_2 解釋總變異的百分比大。
6. 新變數 X_1^* 、 X_2^* 的相關係數為 0，即 X_1^* 、 X_2^* 不相關。

主成分是維度化簡的方法

1. 主成分分析基本上在找新的互相垂直的新軸，點在新軸的投影即為主成分計分。

$$Y_i = k_i'X, \quad k_i'k_i = 1 \text{ 且 } k_i'k_j = 0 \quad i \neq j。$$

λ_i ：S 或 R 矩陣的特徵值； k_i ：其相對應的特徵向量。

2. p 個主成分彼此不相關，故其解釋的變異數不重疊(互變異數=0)。

$$\text{Cov}(Y_i, Y_j) = \text{Cov}(k_i'X, k_j'X) = k_i'\Sigma k_j = k_i'\lambda_j k_j = \lambda_j k_i'k_j = 0, \quad i \neq j$$

3. 新變數變異數的加總等於原始資料的總變異。

$$\text{Var}(Y_i) = \text{Var}(k_i'X) = k_i'\Sigma k_i = k_i'\lambda_i k_i = \lambda_i k_i'k_i = \lambda_i$$

$$\sum_{i=1}^p \lambda_i = \text{tr}(S) \rightarrow \text{tr}(S) = \sum_{i=1}^p \text{Var}(x_i) \rightarrow \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{Var}(x_i) \rightarrow \sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(x_i)$$

4. p 個變數，希望只用 m ($m < p$) 個主成分來描述原始資料，故主成分分析是維度化簡的方法。

主成分分析的目的

1. 所得新軸或新變數稱為主成分，觀察點投影至新軸所得新座標稱為主成分計分。(照相角度的選取；每個人都清楚)
2. 新變數為原變數的線性組合。 $Y_i = k_i'X$ (橢圓長軸的想法；特徵向量 k_i)
3. 第一個新變數解釋原資料最大的變異量。 $\text{Var}(Y_1) = \lambda_1$ ， $\lambda_1 > \lambda_2 > \dots > \lambda_p$
4. 第二個新變數解釋最多第一個新變數未能解釋的總變異。 $\text{Var}(Y_2) = \lambda_2$
5. 第三個新變數解釋最多前兩個新變數未能解釋的總變異。 $\text{Var}(Y_3) = \lambda_3$
6. 第 p 個新變數解釋最多前 p-1 個新變數未能解釋的總變異。 $\text{Var}(Y_p) = \lambda_p$

7. p 個新變數彼此不相關。 $Cov(Y_i, Y_j) = 0, i \neq j$

8. 第 i 個新變數解釋原資料解釋原變數的總變異百分比為 $\frac{\lambda_i}{\sum \lambda_i}$ 。

若原始資料的總變異數被少數幾個新變數解釋的總變異之百分比夠大，則可用少數幾個主成分取代原來 p 個變數來解釋原始資料或做進一步的統計分析。因此，主成分分析可以約化變數的個數是資料維度簡化的統計方法。

主成分分析的解釋

1. 主成分，由特徵向量得對應的主成分(新變數)。
2. 主成分計分，將原始資料代入主成分的式子中，即得主成分計分。
3. 新變數變異數即為特徵值。原總變異數與新總變異數相同，即為特徵值加總，故新變數解釋總變異的比率為該特徵值除以總特徵值。
4. 負荷(loading)，新變數與舊變數間的相關係數，表示原始變數對新變數的影響力或重要性。負荷愈大表示原始變數對新變數的影響力愈高。負荷的大小常用來定義或解釋主成分的意義。公式如下：

$$l_{ij} = \frac{w_{ij} \sqrt{\lambda_i}}{s_j} = r_{y_i x_j}$$

l_{ij} 為第 j 個變數在第 i 個主成分的負荷， w_{ij} 為第 j 個變數在第 i 個主成分的權重， λ_i 為第 i 個主成分的特徵值， s_j 為第 j 個變數的標準差。

5. 共通性(communality，或稱共同性)，第 i 主成分解釋變數 X_j 的變異數比例， $h_j^2 = r_{y_i x_j}^2$ 。

用 R 或 S 的分析結果不同

若不想讓個別變數的變異數的大小影響權重，則應該利用標準化資料(用 R 矩陣)做主成分分析。主成分分析的結果受資料尺度的影響，變數的相對變異數愈大，對應的該變數的權重就愈大。

主成分分析是否為適當的分析方法?

1. 若研究目標是為了找出不相關的新變數，以做進一步的統計分析，則是否採用主成分分析，應該視主成分是否可以解釋而定，如果主成分無法解釋或沒有特別意思，則不應該用主成分形成新變數。
2. 若研究目標是要簡化變數個數，則必須要求少數幾個主成分，可解釋大部分的總變異數，而不會嚴重損失原始資料的資訊。

3. 若原始變數彼此直交成不相關，則主成分分析完全無法減少變數個數只有在變數彼此高相關時，才可能簡化變數的個數，且變數間相關性愈強，資料愈可能化約。
4. 若原變數完全相關，則只需第一主成分，即可解釋 100% 的總變異。
5. 假若變數相關性不高，則資料做主成分分析就不合適。

應萃取的主成分個數

1. 對標準化資料，保留 R 的特徵值大於 1 的主成分，稱為特徵值大於 1 法則。SAS 及 SPSS 等軟體內建的方法。
2. Cattell(1966)，將每一主成分解釋變異數的百分比對主成分作圖並尋找肘點，稱此圖為陡坡圖(scree plot)。此方法適用於均值修正的資料及標準化資料。
3. 只保留具統計顯著性的主成分。但 Bartlett's 檢定法對樣本大小太敏感。

解釋主成分

負荷愈大表示該變數對主成分的影響愈大，一般取“ ≥ 0.5 ”為夠大的取捨標準，藉以對新變數命名。

主成分計分的應用

將主成分計分做為自變數，可做群集分析、迴歸分析或判別分析。此時自變數為不相關的新變數沒有共線性的問題，但可能會發生不易解釋主成分計分在模型上的意義。

補充內容:(源自: 陳順宇著 多變量分析)

若 p 個變數，變數之間的相關係數($r > 0$)皆相等時，若以 R 作主成分分析，則

$$\lambda_1 = 1 + (p-1)r$$

$$\lambda_2 = \lambda_3 = \cdots = \lambda_p = 1 - r$$

對於 λ_i 的特徵向量為

$$k_1 = \begin{bmatrix} \frac{1}{\sqrt{p}} \\ \frac{1}{\sqrt{p}} \\ \vdots \\ \frac{1}{\sqrt{p}} \end{bmatrix}, \quad k_2 = \begin{bmatrix} \frac{1}{\sqrt{1 \times 2}} \\ \frac{-1}{\sqrt{1 \times 2}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad k_3 = \begin{bmatrix} \frac{1}{\sqrt{2 \times 3}} \\ \frac{1}{\sqrt{2 \times 3}} \\ \frac{-2}{\sqrt{2 \times 3}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad k_p = \begin{bmatrix} \frac{1}{\sqrt{(p-1)p}} \\ \frac{1}{\sqrt{(p-1)p}} \\ \vdots \\ \frac{-(p-1)}{\sqrt{(p-1)p}} \end{bmatrix}$$

SPSS 報表說明，以建立烏龜體型指標為例：

資料檔源自：陳順宇著，多變量分析

想要建立烏龜的體型指標，量測 24 隻公烏龜身體長、寬、高。因為體型=長 X 寬 X 高，所以改成 $\ln \text{體型} = \ln \text{長} + \ln \text{寬} + \ln \text{高}$ ，故將資料先做對數後再做主成分分析。

烏龜.sav - SPSS 資料編輯程式

檔案(F) 編輯(E) 檢視(V) 資料(D) 轉換(T) 分析(A) 統計圖(G) 公用程式(U) 視窗(W) 輔助說明(H)

1: 長 93

	長	寬	高	ln長	ln寬	ln高	
1	93	74	37	4.53	4.30	3.61	
2	94	78	35	4.54	4.36	3.56	
3	96	80	35	4.56	4.38	3.56	
4	101	84	39	4.62	4.43	3.66	
5	102	85	38	4.62	4.44	3.64	
6	103	81	37	4.63	4.39	3.61	
7	104	83	39	4.64	4.42	3.66	
8	106	83	39	4.66	4.42	3.66	
9	107	82	38	4.67	4.41	3.64	
10	112	89	40	4.72	4.49	3.69	
11	113	88	40	4.73	4.48	3.69	
12	114	86	40	4.74	4.45	3.69	
13	116	90	43	4.75	4.50	3.76	
14	117	90	41	4.76	4.50	3.71	
15	117	91	41	4.76	4.51	3.71	
16	119	93	41	4.78	4.53	3.71	
17	120	89	40	4.79	4.49	3.69	
18	120	93	44	4.79	4.53	3.78	
19	121	95	42	4.80	4.55	3.74	
20	125	93	45	4.83	4.53	3.81	
21	127	96	45	4.84	4.56	3.81	
22	128	95	45	4.85	4.55	3.81	
23	131	95	46	4.88	4.55	3.83	
24	135	106	47	4.91	4.66	3.85	
25							

SPSS的點法：

Analyze → Data Reduction → Factor Analysis

Variables : LN長, LN寬, LN高

Descriptives -> Correlation Matrix : Coefficients, ...

Extraction -> Method : Principal Components;

Analyze : Correlation Matrix (用 R 做分析)

Covariance Matrix (用 S 做分析)

(兩者選擇一個使用)

Extract : Eigenvalues over : 1 (只列出 $\lambda > 1$ 的主成分)

Number of factors: 3 (自行選擇主成分的個數)

Display : Unrotated factors solution (主成分分析未旋轉)

Scree plot (陡坡圖)

以R作分析：

SPSS的程式：

```
FACTOR
/VARIABLES ln長 ln寬 ln高 /MISSING LISTWISE /ANALYSIS ln長 ln寬 ln高
/PRINT INITIAL CORRELATION EXTRACTION
/PLOT EIGEN
/CRITERIA FACTORS(3) ITERATE(25)
/EXTRACTION PC
/ROTATION NOROTATE
/METHOD=CORRELATION .
```

報表：

Factor Analysis

Correlation Matrix ^a

		LN長	LN寬	LN高
Correlation	LN長	1.000	.951	.942
	LN寬	.951	1.000	.911
	LN高	.942	.911	1.000
Sig. (1-tailed)	LN長	.000	.000	.000
	LN寬	.000	.000	.000
	LN高	.000	.000	.000

a. Determinant = 1.043E-02

R的行列式

Communalities

	Initial	Extraction
LN長	1.000	1.000
LN寬	1.000	1.000
LN高	1.000	1.000

Extraction Method: Principal Component Analysis.

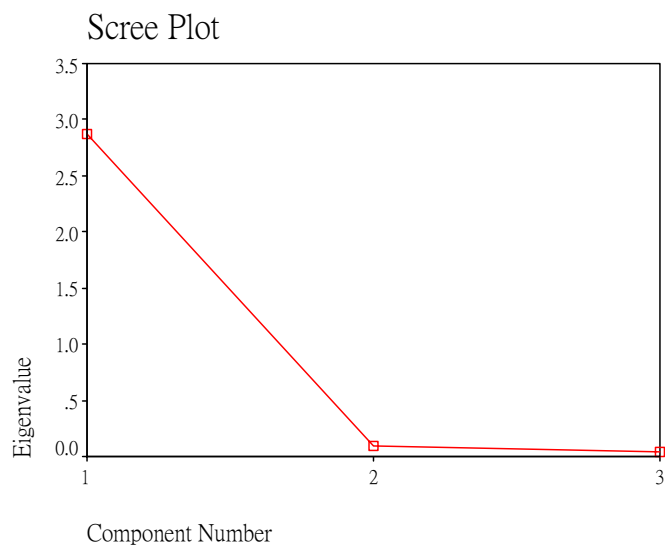
特徵值

第 i 主成分解釋變異百分比

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.870	95.661	95.661	2.870	95.661	95.661
2	8.961E-02	2.987	98.649	8.961E-02	2.987	98.649
3	4.054E-02	1.351	100.000	4.054E-02	1.351	100.000

Extraction Method: Principal Component Analysis.



Component Matrix ^a

	Component	1	2	3
LN長		.986	-2.592E-02	-.163
LN寬		.976	-.197	9.770E-02
LN高		.972	.224	6.710E-02

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

	K1	K1*	K2	K2*	K3	K3*
	0.9860	0.5821	-0.0259	-0.0866	-0.1630	-0.8088
	0.9760	0.5762	-0.1970	-0.6579	0.0977	0.4848
	0.9720	0.5738	0.2240	0.7481	0.0671	0.3329
K _i =	1.6940	1.0000	0.2994	1.0000	0.2015	1.0000

$$||K_i|| = \sqrt{\lambda_i} ; K_i^* = K_i / ||K_i||$$

第一主成分 $Y_1 = 0.5821 * \ln \text{長(標)} + 0.5762 * \ln \text{寬(標)} + 0.5738 * \ln \text{高(標)}$

第二主成分 $Y_2 = -0.0866 * \ln \text{長(標)} - 0.6579 * \ln \text{寬(標)} + 0.7481 * \ln \text{高(標)}$

第三主成分 $Y_3 = -0.8088 * \ln \text{長(標)} + 0.4848 * \ln \text{寬(標)} + 0.3329 * \ln \text{高(標)}$

第一主成分將解釋 95.661% 變異，若僅用第一主成分，則只損失 4.34%的變異未解釋，而變量將減少2個。

以S作分析：

SPSS的程式：

FACTOR

```

/VARIABLES ln長 ln寬 ln高 /MISSING LISTWISE /ANALYSIS ln長 ln寬 ln高
/PRINT INITIAL CORRELATION SIG DET EXTRACTION
/PLOT EIGEN
/CRITERIA FACTORS(3) ITERATE(25)
/EXTRACTION PC
/ROTATION NOROTATE
/METHOD=COVARIANCE .

```

Factor Analysis

Correlation Matrix

		LN長	LN寬	LN高
Correlation	LN長	1.000	.951	.942
	LN寬	.951	1.000	.911
	LN高	.942	.911	1.000
Sig. (1-tailed)	LN長		.000	.000
	LN寬	.000		.000
	LN高	.000	.000	

Covariance Matrix ^a

a. Determinant = 5.017E-09

S的行列式

Communalities

	Raw		Rescaled	
	Initial	Extraction	Initial	Extraction
LN長	1.107E-02	1.107E-02	1.000	1.000
LN寬	6.417E-03	6.417E-03	1.000	1.000
LN高	6.773E-03	6.773E-03	1.000	1.000

Extraction Method: Principal Component Analysis.

特徵值

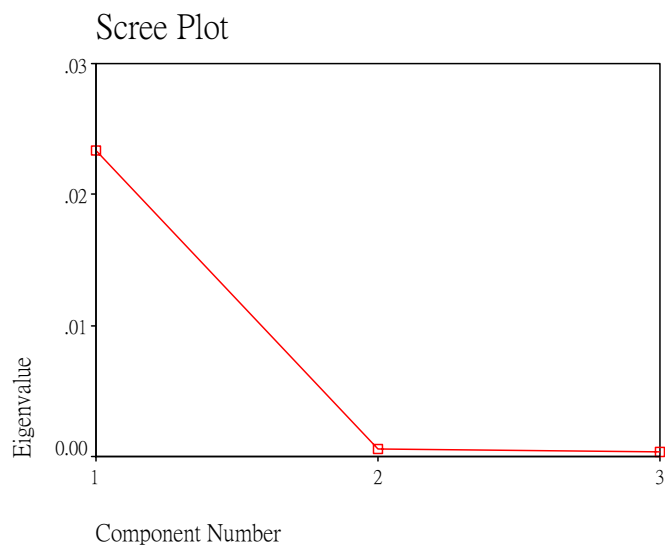
第 i 主成分解釋變異百分比

Total Variance Explained

Component		Initial Eigenvalues ^a			Extraction Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Raw	1	2.330E-02	96.051	96.051	2.330E-02	96.051	96.051
	2	5.983E-04	2.466	98.517	5.983E-04	2.466	98.517
	3	3.598E-04	1.483	100.000	3.598E-04	1.483	100.000
Rescaled	1	2.330E-02	96.051	96.051	2.867	95.567	95.567
	2	5.983E-04	2.466	98.517	8.920E-02	2.973	98.540
	3	3.598E-04	1.483	100.000	4.379E-02	1.460	100.000

Extraction Method: Principal Component Analysis.

a. When analyzing a covariance matrix, the initial eigenvalues are the same across the raw and rescaled solution.



Component Matrix ^a

	Raw			Rescaled		
	Component			Component		
	1	2	3	1	2	3
LN長	.104	-.004	-.014	.991	-.037	-.128
LN寬	.078	-.015	.012	.972	-.181	.147
LN高	.080	.019	.006	.969	.234	.075

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

$$K1 = \sqrt{\lambda_i} \quad K1^*$$

	K1	K1*	K2	K2*	K3	K3*
	0.104	0.681	-0.004	-0.163	-0.014	-0.722
	0.078	0.511	-0.015	-0.611	0.012	0.619
	0.080	0.524	0.019	0.774	0.006	0.309
K =	0.153	1.000	0.025	1.000	0.019	1.000

$$||K|| = \sqrt{\lambda_i} ; \quad K_i^* = K_i / ||K||$$

第一主成分 $Y1 = 0.681 * \ln長 + 0.511 * \ln寬 + 0.524 * \ln高$

第二主成分 $Y2 = -0.163 * \ln長 - 0.611 * \ln寬 + 0.774 * \ln高$

第三主成分 $Y3 = -0.722 * \ln長 + 0.619 * \ln寬 + 0.309 * \ln高$

第一主成分將解釋 96.051% 變異，若僅用第一主成分，則只損失 3.85%的變異未解釋，而變量將減少2個。

<< 主成分分析 SAS 程式 >>

以 S 做主成分分析

```
data aa;
input 變數1 變數2...變數p;
cards;
資料...;

proc princomp COV out=bb;
var 變數1 變數2...變數p;
proc corr; var 變數1 變數2...變數p; with 主成分1 主成分2...主成分p;
proc print; var 變數1 變數2...變數p 主成分1 主成分2...主成分p;
run;
```

(以 R 做主成分分析時，將 ``COV'' 去掉 ; proc princomp out=bb;)

```
data aa;
input 變數1 變數2...變數p;
cards;
資料...;

proc princomp out=bb;
var 變數1 變數2...變數p;
proc corr; var 變數1 變數2...變數p; with 主成分1 主成分2...主成分p;
proc print; var 變數1 變數2...變數p 主成分1 主成分2...主成分p;
run;
```

資料檔源自：陳順宇著 多變量分析

例 1: 收集 15 筆資料，訂定體型指標。

x1:身高, x2:體重 (以 S 做主成分分析)

```
data aa;
input x1 x2;
cards;
173 66
155 49
175 72
171 68
166 63
167 64
163 61
155 52
159 55
168 65
166 61
169 73
159 57
154 49
160 60
;
proc princomp cov out=bb;
var x1 x2;
proc corr; var x1 x2; with prin1 prin2;
proc print; var x1 x2 prin1 prin2;
run;
```

資料檔源自：陳順宇著 多變量分析

例 2: 收集 15 位大學生，訂定成績指標。

x1:國文, x2:英文, x3:統計, x4:會計, x5:經濟

x1	x2	x3	x4	x5
83	76	79	71	76
65	59	71	61	69
85	82	80	69	74
81	78	67	72	69
76	73	65	60	67
77	74	68	55	70
73	71	78	71	75
65	62	68	61	67
69	65	58	57	67
78	75	57	56	64
76	71	64	64	66
79	83	68	64	72
69	67	64	66	68
64	59	66	56	67
70	70	75	69	76

資料檔源自：陳順宇著 多變量分析

例 3: 收集威斯康辛州 14 個地區，訂定社經指標。

x1:人口數(千人), x2:平均受教育平均年限, x3:就業人數(千人),
x4:健康服務人員數(千人), x5:平均房價(萬元)

x1	x2	x3	x4	x5
5.935	14.2	2.265	2.27	2.91
1.523	13.1	0.597	0.75	2.62
2.599	12.7	1.237	1.11	1.72
4.009	15.2	1.649	0.81	3.02
4.687	14.7	2.312	2.50	2.22
8.044	15.6	3.641	4.51	2.36
2.766	13.3	1.244	1.03	1.97
6.538	17.0	2.618	2.39	1.85
6.451	12.9	3.147	5.52	2.01
3.314	12.2	1.606	2.18	1.82
3.777	13.0	2.119	2.83	1.80
1.530	13.8	0.798	0.84	4.25
2.768	13.6	1.336	1.75	2.64
6.585	14.9	2.763	1.91	3.17

資料檔源自：陳順宇著 多變量分析

例 4: 收集 24 隻公烏龜身體長、寬、高，訂定體型指標。

x1:長, x2:寬, x3:高

	x1	x2	x3
93	74	37	
94	78	35	
96	80	35	
101	84	39	
102	85	38	
103	81	37	
104	83	39	
106	83	39	
107	82	38	
112	89	40	
113	88	40	
114	86	40	
116	90	43	
117	90	41	
117	91	41	
119	93	41	
120	89	40	
120	93	44	
121	95	42	
125	93	45	
127	96	45	
128	95	45	
131	95	46	
135	106	47	

作業： <<請於 3/13(二)繳交>>

1. 若用 x_1, x_2 的相關矩陣 R 做主成分分析，則下列何種線性組合不可能是第一主成分？

- (1) $0.7 x_1 + 0.3 x_2$
- (2) $0.8 x_1 - 0.6 x_2$
- (3) $0.6 x_1 + 0.8 x_2$
- (4) $0.2 x_1 + 0.8 x_2$
- (5) $-0.8 x_1 - 0.6 x_2$

2. 設有 4 個變數 x_1, x_2, x_3, x_4 ，其兩兩變數之間的相關係數皆為 0.2，如果用相關矩陣 R 做主成分分析，

- (1) 第一主成分 y_1 如何表成 x_1, x_2, x_3, x_4 的線性組合
- (2) y_1 與 x_1 的相關係數
- (3) $\text{Var}(y_1)$
- (4) y_1 解釋了 x_1, x_2, x_3, x_4 多少變異數比例
- (5) $\text{Var}(y_2)$
- (6) 第二主成分 y_2 如何表成 x_1, x_2, x_3, x_4 的線性組合
- (7) y_2 解釋了 x_1, x_2, x_3, x_4 多少變異數比例
- (8) 如果選用 3 個主成分， x_1, x_2, x_3, x_4 共被解釋了多少變異數比例

3. x_1, x_2 的相關矩陣 R 為

$$\begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}$$

標準差分別為 6 和 5，平均數分別為 60 和 70。

(1) 若以 R 做主成分分析

- (a) 求第一主成分 y_1 的線性組合
- (b) 若第三筆資料為 (66, 60)，求第一主成分得點
- (c) 求第一主成分的變異數，及解釋 x_1 、 x_2 變異數的比例

(2) 若以 S 做主成分分析

- (a) 求第一主成分 y_1 的線性組合
- (b) 若第三筆資料為 (66, 60)，求第一主成分得點
- (c) 求第一主成分的變異數，及解釋 x_1 、 x_2 變異數的比例

4. 若 x_1, x_2, x_3 的相關矩陣 R 為

$$\begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- (1) y_1 解釋了 x_1, x_2, x_3 多少變異數比例
- (2) 求第一主成分 y_1 的線性組合
- (3) 求第二主成分 y_2 的線性組合
- (4) 第一主成分和第二主成分共解釋了 x_1, x_2, x_3 多少變異數比例

5. 試將前頁“例 2”或“例 3”擇一例子，使用 SPSS 或 SAS 做主成分分析，請列出報表及其說明。