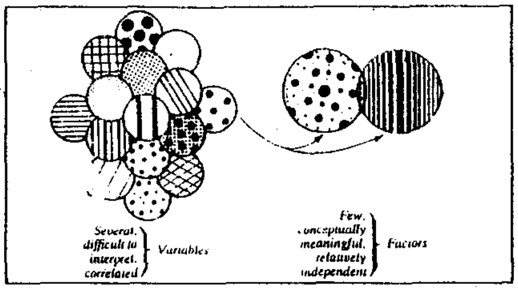
บทที่ 6
การวิเคราะห์ส่วนประกอบสำคัญและการวิเคราะห์ปัจจัย
[Principal Component Analysis and Factor Analysis]

		หน้า
6.1	บทนำ	150
6.2	เทอมต่าง ๆ ในการวิเคราะห์ปัจจัย	152
6.3	สรุปขั้นต่าง ๆ ในการวิเคราะห์ปัจจัย	155
6.4	Principal Components Technique	158
	6.4.1 คำจำกัดความ	158
	6.4.2 การตีความหมายของ Principal Components	160
	6.4.3 ตัวอย่าง Principal Components	160
	6.4.4 การหมุนแกนของตัวประกอบเริ่มต้น (Rotation of Initial Factors)	161
6.5	การหาค่าของปัจจัย (Factor Scores)	171
	6.5.1 สรุปวิธีวิเคราะห์ของกรณีตัวอย่าง	171
	6.5.2 การศึกษาผลการวิเคราะห์	172
	แบบฝึกหัดที่ 6	176

บทที่ 6
การวิเคราะห์ส่วนประกอบสำคัญและการวิเคราะห์ปัจจัย
(Principal Component Analysis and Factor Analysis) .

6.1 บทนำ

การวิเคราะห์ปัจจัยเป็นวิธีการวิเคราะห์ของตัวแปรพหุซึ่งมีจุดประสงค์ที่จะอธิบายถึงความสัมพันธ์ระหว่างกลุ่มตัวแปรซึ่งมีความสัมพันธ์กันและยากที่จะตีความหมาย (difficult-to-interpret) ให้อยู่ในเทอมของ factors 2-3 factors ซึ่งมีความหมายและมีความเป็นอิสระต่อกัน



รูปที่ 6.1 รูปแสดงวัตถุประสงค์โดยทั่วไปของการวิเคราะห์ปัจจัย

จากรูปที่ 6.1 จะเห็นว่าจากวงกลมซึ่งมีลวดลายต่าง ๆ กันหลาย ๆ วงซึ่งซ้อน ๆ กันอยู่ เราจะสร้างวงกลม 2 วงซึ่งถือได้ว่าไม่ซ้อนกัน ซึ่งมีลวดลายต่างกัน

ตัวอย่างที่ 6.1

เนื่องจากได้มีผู้ทำวิจัยได้ผลว่าคนผิวดำในอเมริกานั้น มีแนวโน้มที่จะมีระดับความดันโลหิตเฉลี่ยสูงกว่าคนผิวขาว และอัตราการตายเนื่องจากโรคความดันโลหิตสูงและโรคอื่น ๆ ที่เกี่ยวข้องนั้นในคนผิวดำสูงกว่าในคนผิวขาวมาก แต่สาเหตุของความแตกต่างข้างต้นยังมีผู้สงสัยและไม่เข้าใจ จึงได้มีผู้ทำการวิจัยเพิ่มเติมโดยศึกษาเกี่ยวกับเรื่องกรรมพันธุ์และปัจจัยของสิ่งแวดล้อมทางสังคม (socioenviromental factor) รวมทั้งบริการทางแพทย์ที่ได้รับด้วย เมื่อ ค.ศ. 1976 เจมส์ (James) และ คลีนบวม (Kleinbaum)[1] ได้ทำการศึกษาสิ่งแวดล้อมทางสังคมโดยเน้นการศึกษาว่าคนผิวดำที่อาศัยอยู่ในบริเวณ high stress นั้นมีแนวโน้มจะมีอัตราการตายซึ่งเกี่ยวข้องกับโรคความดันโลหิตสูง (hypertension-related mortality rate) สูงกว่า

---

[1]James, S.A., and Kleinbaum, D.G. (1976) "Socioecologic Stress and Hypertension-Related Mortality Rates in North Carolina." **American Journal Public Health, 66(4):354-358.**

คนผิวดำที่อาศัยในบริเวณ low-stress หรือคนผิวขาวที่อาศัยในบริเวณ high-stress หรือคนผิวขาวที่อาศัยในบริเวณ low-stress หน่วยตัวอย่างที่ใช้คือเมือง 86 เมืองในรัฐนอร์ธคาโรไลนา เจมส์และคลีนบวมคำนวณอัตราการตายซึ่งเกี่ยวข้องกับโรคความดันโลหิตสูงในช่วงเวลา 3 ปี คือตั้งแต่ปี ค.ศ.1959-1961 สำหรับแต่ละเมืองและสร้างความสัมพันธ์ระหว่างอัตราดังกล่าว ของเมืองเหล่านี้กับดรรชนีของ socioecologic stress สำหรับแต่ละเมืองซึ่งหาได้จากการวิเคราะห์ปัจจัย คำถามขั้นแรกของการศึกษาคือ "เมืองที่อยู่ในบริเวณ high-stress มีอัตราการตายสูงกว่าของเมืองในบริเวณ low-stress อย่างมีนัยสำคัญหรือไม่"

ดรรชนีของ socioecologic stress ถูกหาจากการใช้ตัวแปร 15 ตัว (จากค่าของปี 1960 ได้ค่าสำหรับแต่ละสีผิวแยกกันในแต่ละเมือง) ซึ่งเป็นสิ่งที่ใช้สะท้อนสถานภาพทางเศรษฐกิจและสังคมว่าดีหรือไม่ของเมืองระหว่างปี 1959-1961 ตัวแปร 15 ตัว และชื่อย่อของตัวแปรเหล่านั้นแสดงไว้ในตารางที่ 6.1

ตารางที่ 6.1 ตัวแปรที่ใช้ในการศึกษาของ เจมส์และคลีนบวม

⊙	1 Per capita income	PCI
⊙	2 Median years of education	MED
⊙ ☒	3 Percent unemployment	UEM
⊙	4 Percent families earning over $8 000/year	P8K t
⊙	5 Percent male white-collar jobs	WCM
⊙	6 Percent male blue-collar Jobs	BC
⊙ ☒	7 Percent families earning under $3,000/year	P3K-
☒	8 Percent females separated or divorced	PSDF
☒	9 Juvenile-delinquency-Index males	JDM
☒	10 Juvenile-delinquency-index females	JDF
☒	11 Percent of males in correction school	CSM
☒	12 Percent of females in correction school	CSF
☒	13 Percent of males in prisons	PM
☒	14 Homicide rate	HR
☒	15 Percent of children under 18 not with parent	PCBH

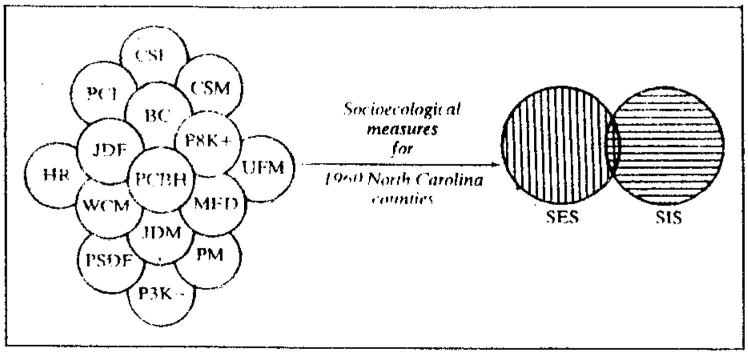
ในตารางที่ 6.1 ยังแสดงให้เห็น 2 factors ซึ่งเป็นผลที่ได้จากการทำการวิเคราะห์ปัจจัย และถูกใช้ร่วมกันในการกำหนดดรรชนีของ socioecologic stress 2 factors ดังกล่าวมีชื่อว่า

socioeconomic status (SES) factor → ⊙

และ socioinstability (SIS) factor → ☒

ดังที่เห็นในตารางที่ 6.1 **ตัวแปรกลุ่มหนึ่งซึ่งแสดงด้วยเครื่องหมาย** ⊙ (เช่น PCI, MED) **เราคาดว่าจะสัมพันธ์กับ SES** มากกว่ากับ SIS ส่วนตัวแปรอื่น ๆ (เช่น JDM, JDF, CSM) คาดว่าเป็นกลุ่มตัวแปรใน SIS ถูกแสดงด้วยเครื่องหมาย ☒ และตัวแปรบางตัว (เช่น UEM, P3K-) นั้นเราคาดว่าจะสัมพันธ์กับทั้ง SES และ SIS factor และเราคาดต่อไปอีกว่าแต่ละ factor ที่

ได้จากการวิเคราะห์ปัจจัยนั้นจะมีความสัมพันธ์อย่างสูงกับตัวแปรที่เราคิดว่าจะสัมพันธ์กับ factor นั้นและจะไม่มีความสัมพันธ์กับตัวแปรที่เราคิดว่าสัมพันธ์กับอีก factor หนึ่ง ในหัวข้อหลังของบทนี้เราจะได้เห็นว่าสิ่งที่กล่าวมาข้างต้นนั้นเกิดขึ้นจริงหรือไม่



รูปที่ 6.2 *แสดงจุดประสงค์ของการวิเคราะห์ปัจจัยในการศึกษาของเจมส์และคลีนบวม*

จากรูปที่ 6.2 วงกลม 15 วงด้านซ้ายแทนตัวแปร 15 ตัว ซึ่งมีความสัมพันธ์กันเราใช้ตัวแปร 15 ตัวเพื่อหา 2 factors ที่เป็นอิสระต่อกัน (SES และ SIS) ทางขวา

6.2 เทอมต่าง ๆ ในการวิเคราะห์ปัจจัย

Factor loadings คือสหสัมพันธ์ระหว่าง factors และตัวแปรซึ่งใช้ในการหา factors

Factor loading matrix นี้จะเป็นตัวชี้ว่าตัวแปรเหล่านั้นมีความสัมพันธ์อย่างสูง (i.e., "load high") กับ factor ใด ดังนั้นเราจึงจะแปลความหมายของ factor ได้

ตารางที่ 6.2

Factor loadings

VARIABLES	FACTOR SES	FACTOR SIS
PCI	0.88	0.22
MED	0.88	−0.07
UEM	0.03	−0.03
P8K+	0.65	0.03
WCM	0.56	0.23
BC	0.81	0.12
P3K−	0.88	−0.03
PSDF	0.18	0.75
JDM	0.34	0.73
JDF	0.44	0.62
CSM	0.38	0.60
CSF	0.50	0.35
PM	−0.06	0.71
HR	−0.41	0.26
PCBH	−0.12	0.75

Factor weights คือ weights ที่กำหนดให้แก่ตัวแปรแต่ละตัว เพื่อจะใช้ในการหา factor scores

Factor weights ไม่ใช่สหสัมพันธ์ แต่เป็นเลขจำนวนซึ่งถูกกำหนดให้ตัวแปรตัวหนึ่ง (ซึ่งถูกทำให้เป็นตัวแปรมาตรฐาน (standardized variable) แล้ว) เพื่อใช้ในการหา scores สำหรับแต่ละ factor (factor scores)

ตารางที่ 6.3

Factor weights

VARIABLE	FACTOR	
	SES	SIS
PCI	0.19	−0.00
MED	0 22	−0.10
UEM	0.01	−0.01
PGK −	0.15	−0.05
WCM	0.11	0.03
BC	0.18	−0.03
P3K −	−0.21	0.07
PSDF	−0.02	0.24
JDM	0.01	0.22
JDF	0.05	0.17
CSM	0.04	0.17
CSF	0.09	0.07
PM	0.08	0.25
HR	−0.12	0.13
PCBH	−0.10	0.27

Factor loadings ค่าสูงมักจะสมนัยกับ factor weights ที่มีค่าสูงด้วย ดังนั้นข้อมูลในตารางที่ 6.2 และ 6.3 จึงให้ข่าวสารข้อมูลที่คล้ายคลึงกัน แต่การวัดนั้นวัดในเสกลที่ต่างกัน และถูกใช้เพื่อวัตถุประสงค์ต่างกัน นั่นคือ factor weights ถูกใช้เพื่อคำนวน factor scores ส่วน factor loadings นั้นถูกใช้เพื่ออธิบายสหสัมพันธ์

Factor Cosines คือสหสัมพันธ์ระหว่าง factor กับ factor ค่าสหสัมพันธ์นี้สำคัญเพราะใช้วัดความสัมพันธ์ระหว่าง factor ตามที่กล่าวมาข้างต้นว่าในการทำการวิเคราะห์ปัจจัยนั้นเราต้องการให้ factor ที่เราได้มานั้นเป็นอิสระต่อกัน นั่นคือต้องการให้ค่าของ factor cosines เข้าใกล้ศูนย์ หรืออาจพิจารณาได้ว่า ถ้า factor ทั้งสองมีความสัมพันธ์กันสูง ก็แสดงว่า factor ทั้งสองให้ข่าวสารอย่างเดียวกัน จึงอาจใช้ factor ใด factor หนึ่งไม่ต้องใช้ทั้งสอง factor ความเป็นอิสระต่อกันที่ต้องการนั้น ไม่จำเป็นต้องเป็นอิสระกันอย่างสมบูรณ์ แต่ถ้าเป็นอิสระต่อกันอย่างมีนัยสำคัญก็ใช้ได้ เช่น corr (F_1, F_2) จากตารางที่ 6.4 มีค่า 0.17 แต่ไม่มีนัยสำคัญ

ตารางที่ 6.4

Factor Cosines

FACTOR	FACTOR	
	SES	SIS
SES	1.0	0.17
SIS	0 17	1.0

Standardized factor scores คือ standardized values ของ factor ซึ่งเป็น weighted sum ของoriginal standardized variabies

ตารางที่ 6.5

Standardized factor scores

COUNTY	FACTOR SES	FACTOR SIS
1	0.87	-0.57
2	1 04	-0.52
3	-0.72	-0.08
85	-1.04	0 94
86	0.25	-0.29

Factor scores ของแต่ละหน่วยตัวอย่าง คือค่าของ factor ซึ่งคำนวณจากค่าของตัวแปร (มักจะเป็น standardized) ต่าง ๆ ของหน่วยตัวอย่างนั้น ๆ

$(X_1, \ldots, X_p)$ = original variables [มักจะอยู่ในรูป standardized form(dimensionless)]

$$\text{Factor } F = \sum_{j=1}^{p} W_j X_j = W_1X_1 + W_2X_2 + \ldots + W_pX_p$$

$$F_i = \sum_{j=1}^{p} W_{ij} X_j = W_{i1}X_1 + W_{i2}X_2 + \ldots + W_{ip}X_p, \; i = 1, \ldots, k$$

โดยที่ k = # factors

p = # ตัวแปร

n = # หน่วยตัวอย่าง

$\boxed{W_{i1},\ldots,W_{ip} \text{ คือ factor weights ของ } F_i}$ (ซึ่งจะต้องประมาณค่าจากข้อมูล)

เมื่อที่ h : observed ค่า $x_{1h}, x_{2h}, \ldots, x_{ph}$

$\therefore$ Factor score F_i ของ sampling unit ที่ h คือ $\sum_{j=1}^{p} W_{ij}\, x_{jh}$, $h = 1,\ldots,n$

จากตัวอย่างที่ 6.1 p = 15,n = 86

Factor score F_1 ของ sampling unit ที่ h = $\sum_{j=1}^{p} W_{1j}\, x_{jh}$

$$= W_{11}x_{1h} + W_{12}x_{2h} + \ldots + W_{1p}x_{ph}, \; h = 1, \ldots, n$$

Factor score F_2 ของ sampling unit ที่ h = $\sum_{j=1}^{p} W_{2j}\, x_{jh}$

$$= W_{21}x_{1h} + W_{22}x_{2h} + \ldots + W_{2p}x_{ph}, \; h = 1, \ldots, n$$

โดยการใช้ factor weights จากตารางที่ 6.3

F_1 : SES factor (ค่าประมาณ) = 0.19 (PCI) + 0.22 (MED) + ...-0.10 (PCBH)

F_2 : SIS factor (ค่าประมาณ) = -0.00 (PCI) -0.10 (MED) + ... + 0.27 (PCBH)

สรุปความแตกต่างระหว่าง factor และ factor score

factor เป็นนิพจน์ทั่วไปสำหรับ F ซึ่งจะมีค่า ๆ หนึ่งซึ่งสมนัยกับหน่วยตัวอย่างหนึ่ง ๆ **factor score** เป็นค่าค่าหนึ่งของ **factor** ซึ่งได้จากการแทนสูตรหรือนิพจน์ทั่วไปด้วยค่าของตัวแปรชุดนั้นจากหน่วยตัวอย่างที่เลือกมา

6.3 สรุปขั้นต่าง ๆ ในการทำการวิเคราะห์ปัจจัย

(n × p) data matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} \cdots x_{1p} \\ \cdots \\ x_{n1} \cdots x_{np} \end{bmatrix} \quad \begin{array}{l} \rightarrow j = 1 \rightarrow \bar{x}_1, s_1 \\ \\ \rightarrow j = p \rightarrow \bar{x}_p, s_p \end{array}$$

เป็น matrix ของ **ตัวอย่างสุ่มขนาด n** จาก $N_p(\underline{\mu}, \Sigma)$

นั่นคือ ถ้า $\mathbf{X}' = [X_1 \ldots X_p]$

$\mathbf{X} \sim N_p(\underline{\mu}, \Sigma)$

เราประมาณ Σ ด้วย **S** (sample variance-covariance matrix) ดังนั้นข่าวสารทั้งหมดที่เราต้องการในการวิเคราะห์ส่วนประกอบ (principal component analysis) จะอยู่ในเมตริกซ์ **S** นั่นเอง อย่างก็ดีเรามีความจำเป็นต้องเลือกวัดความสัมพันธ์กันหรือการขึ้นต่อกัน เราควรจะใช้ค่า variance และ covariance ของข้อมูลในการวิเคราะห์ของเรา และวิเคราะห์กับหน่วยเดิมของข้อมูล หรือรูปแบบของความสัมพันธ์จะเด่นชัด ถ้าได้จากการที่เราแปลงรูป x_{jh} ให้อยู่ในรูป **standard score** (ไม่มีหน่วย)

$$\boxed{z_{jh} = \frac{x_{jh} - \bar{x}_j}{s_j}} \qquad , j = 1, \ldots, p \text{ และ } h = 1, \ldots, n$$

และใช้เมตริกซ์ **R** ในการวิเคราะห์ Components ที่ได้มาจากเมตริกซ์ **S** และ **R** นั้นโดยทั่วไปแล้วจะไม่เหมือนกัน และไม่สามารถจะเปลี่ยนคำตอบจากรูปหนึ่งไปเป็นอีกรูปหนึ่งโดยการ

หาค่าคงที่มาคูณเข้าเท่านั้นได้ การประยุกต์ใช้เทคนิคต่าง ๆ นั้นจะเกี่ยวข้องกับ correlation matrix ดังเช่นเราจะใช้ในการวิเคราะห์ปัจจัย

เหตุที่ต้องใช้ standardized variables และ correlation matrix ทั้งนี้เนื่องจาก ถ้าข้อมูลมีหน่วยต่าง ๆ กัน (อายุมีหน่วยเป็นปี น้ำหนักเป็น กิโลกรัม ส่วนค่าที่วัดทางเคมี ชีววิทยา และฟิสิกส์ก็มีหน่วยที่ต่าง ๆ กัน) linear combination ของตัวแปรดั้งเดิม (original variables) จะมีความหมายเพียงเล็กน้อย ดังนั้นจึงมีความจำเป็นที่จะต้องใช้ค่าที่ไม่ขึ้นอยู่กับหน่วยของการวัด คือใช้ standardized variables และ correlation matrix

ตารางที่ 6.6

Correlation matrices สำหรับการศึกษาของเจมส์และคลีนบวม

Correlation matrix—whites

	1(PCI)	2(MED)	3(UEM)	4(PK8+)	5(WCM)	6(BC)	7(PK3−)	8(PSDF)	9(JDM)	10(JDF)	11(CSM)	12(CSF)	13(PM)	14(HR)	15(PCBH)
(PCI)1	1.000														
(MED)2	0.646	1.000													
(UEM)3	−0.188	−0.227	1.000												
(PK8+)4	0.931	0.673	−0.201	1.000											
(WCM)5	0.734	0.865	−0.040	0.785	1.000										
(BC)6	0.286	−0.281	0.278	0.086	−0.170	1.000									
(PK3−)7	−0.905	−0.492	0.129	−0.799	−0.567	−0.481	1.000								
(PSDF)8	0.649	0.225	0.207	0.582	0.484	0.447	−0.650	1.000							
(JDM)9	0.629	0.396	0.025	0.574	0.534	0.303	−0.596	0.566	1.000						
(JDF)10	0.442	0.172	0.084	0.396	0.274	0.439	−0.477	0.483	0.625	1.000					
(CSM)11	0.237	0.034	0.175	0.199	0.245	0.348	−0.270	0.480	0.369	0.397	1.000				
(CSF)12	0.250	0.076	0.241	0.219	0.269	0.395	−0.258	0.451	0.401	0.567	0.694	1.000			
(PM)13	0.158	0.038	0.138	0.189	0.225	0.005	−0.206	0.310	0.308	0.296	0.486	0.192	1.000		
(HR)14	−0.137	−0.176	0.209	−0.135	−0.138	0.137	0.100	0.127	−0.009	0.052	0.071	0.050	−0.005	1.000	
(PCBH)15	−0.158	−0.175	0.265	−0.103	0.003	0.111	0.116	0.402	−0.007	0.071	0.363	0.351	0.235	0.306	1.000

Correlation matrix—nonwhites

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000														
2	0.720	1.000													
3	0.010	0.053	1.000												
4	0.633	0.446	0.118	1.000											
5	0.519	0.587	0.311	0.433	1.000										
6	0.802	0.702	0.043	0.436	0.377	1.000									
7	−0.930	−0.693	0.088	−0.568	−0.420	−0.732	1.000								
8	0.425	0.289	0.111	0.197	0.473	0.295	−0.269	1.000							
9	0.547	0.337	−0.021	0.251	0.437	0.464	−0.387	0.579	1.000						
10	0.590	0.376	−0.044	0.402	0.441	0.496	−0.463	0.519	0.881	1.000					
11	0.537	0.415	−0.166	0.247	0.301	0.458	−0.404	0.575	0.581	0.499	1.000				
12	0.533	0.437	−0.225	0.257	0.235	0.496	−0.456	0.370	0.427	0.492	0.570	1.000			
13	0.276	−0.011	−0.189	0.117	0.132	0.083	−0.213	0.432	0.463	0.397	0.445	0.226	1.000		
14	−0.144	−0.292	0.022	−0.070	−0.126	−0.170	0.142	0.006	−0.073	−0.118	−0.142	−0.220	0.200	1.000	
15	0.178	−0.024	0.224	0.104	0.206	0.209	0.017	0.559	0.433	0.366	0.359	0.245	0.283	0.182	1.000

ขั้นที่ 1 เตรียมเมตริกซ์ **R** (Correlation matrix)

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{cccc} X_1 & X_2 & & \ldots X_p \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ \\ X_p \end{array} & \begin{bmatrix} 1 & r_{12} & & \ldots r_{1p} \\ r_{12} & 1 & & \ldots r_{2p} \\ & & & \\ r_{1p} & r_{2p} & & \ldots 1 \end{bmatrix} \end{array}$$

ขั้นที่ 2 ใช้เมตริกซ์ **R** ในการหาชุดของ initial factors (a set of initial factors) ซึ่งเราจะใช้วิธีของ principal components (method of principal components) ในการหา

จุดสำคัญที่จะต้องกล่าวในขั้นนี้คือ การหา initial factors ตามวิธีของ principal component จะทำให้เราบรรลุ**วัตถุประสงค์** 2 ข้อแรกของการวิเคราะห์ปัจจัย จากทั้งหมด 3 ข้อ คือ

1) parsimony

2) approximately independence

และ 3) conceptual meaningfulness

1) **Parsimony** จะบรรลุได้โดยการแทนข่าวสารที่อยู่ในตัวแปรดั้งเดิมในรูปของ factor ซึ่งมีจำนวนตัว (จำนวน factor) น้อยที่สุดเท่าที่จะทำได้ (a much smaller number of factors)

2) **Approximately independence** จะบรรลุได้โดยการหา factors เพื่อให้ได้ factors ซึ่งจะเป็นอิสระต่อกันอย่างมีนัยสำคัญทางสถิติ (จริง ๆ แล้ว initial factors ที่ได้จากวิธีของ principal components นั้นมีความเป็นอิสระต่อกันอย่าง**สมบูรณ์** ไม่ใช่เพียงประมาณว่าเป็นอิสระต่อกัน เทอม "approximate" นั้นจะมีความหมายหลังจากขั้นที่ 3 ทั้งนี้ขึ้นอยู่กับวิธีการที่จะใช้ในขั้นที่ 3 ซึ่งคือวิธีที่ใช้ในการหมุนแกนนั้นเอง)

โชคไม่ดีที่ถึงแม้ว่าชุดของ initial factors นั้นมีคุณลักษณะตามที่ต้องการแต่เรามักพบปัญหาว่าเป็นการยากในการตีความหมาย factors เหล่านั้นทันที ดังนั้นเราต้องจัดกระทำบางอย่างกับ set ของ initial factors นั้นเพื่อทำให้การตีความหมายของ factors ง่ายขึ้น เราจึงปฏิบัติการในขั้นที่ 3 ต่อไป

ขั้นที่ 3 ในขั้นนี้เราทำการหมุนแกน "Rotation" เพื่อทำให้เราบรรลุวัตถุประสงค์ข้อที่ 3 ของการวิเคราะห์ปัจจัย นั่นคือเพื่อให้ได้คุณลักษณะ "conceptual meaningfulness"

ขั้นที่ 4 หลังจากทำการหมุนแกนแล้ว เราจะได้ชุดของ factor weights ซึ่งจะถูกใช้ในการคำนวณ factor scores ในขั้นนี้เราจะได้ชุดของ factor scores ของแต่ละหน่วยตัวอย่างสำหรับแต่ละ factor ที่หาได้ ซึ่งตัวแปรที่ถูกสร้างขึ้นใหม่ (newly constructed) นี้จะถูกนำไปใช้ในการวิเคราะห์ต่อไป

6.4 Principal Components Technique

จากขั้นที่ 2 ในข้อ 6.3 เราได้กล่าวถึง principal components analysis หรืออาจเรียกว่า principal-axis method ซึ่งคือวิธีที่ใช้ในการหา initial factors

"principal components" คือชื่อที่ใช้เรียก set ของ initial factors ที่หาได้จากการใช้ principal components method นั่นเอง

6.4.1 คำจำกัดความ

1) Total Variation ใน $X_1, X_2, \dots, X_p$

Total Variation $= S_1^2 + S_2^2 + \dots + S_p^2 \Rightarrow$ วัดปริมาณ "uncertainty" ของข้อมูลทั้งในตัวแปร p ตัว n observations

โดยที่ S_j^2 คือ sanple variance ของ

$X_j, j = 1, \dots, p$

หมายเหตุ

ถ้า X_j เป็น standardized form, $S_j^2 = 1$, $\forall_j = 1, \dots, p$ ดังนั้น

Total Variation $= p =$ trace **S**

$=$ จำนวนตัวแปรนั่นเอง

2) Principal Components

จุดประสงค์ของ principal components analysis คือการหา factor (i.e. principal components) โดยที่จะใช้ factor จำนวนน้อยตัวเท่าที่จะเป็นไปได้ (parsimony) เพื่ออธิบาย total variation ของข้อมูลให้ได้มากเท่าที่จะมากได้

First principal component : PC (1)

PC(1) คือ weighted linear combination ของตัวแปรซึ่งสมนัยกับส่วนที่ใหญ่ที่สุดของ total variation ของข้อมูล

นั่นคือถ้า PC(1) เป็น linear combination ของ X' S

$PC(1) = W_{(1)1}X_1 + W_{(1)2}X_2 + \dots + W_{(1)p}X_p$

โดยที่ $W_{(1)1}, W_{(1)2}, \dots, W_{(1)p}$ เป็น weights ซึ่งถูกเลือกมาเพื่อทำให้ปริมาณ

$$\left[\frac{\text{Variance ของ PC(1)}}{\text{Total Variation}}\right] \text{ มีค่ามากที่สุด}$$

หรืออาจกล่าวได้ว่าไม่มี linear combination (ของ X' S) อื่นที่จะมี variance มากเท่ากับ PC(1)

หมายเหตุ

ถ้า X'S อยู่ในรูป standardized (i.e. ใช้ matrix **R** ในการวิเคราะห์)

proportion ของ total variation ซึ่งเนื่องมาจาก PC(1) คือ

$$\left[\frac{\text{Variance ของ PC(1)}}{p}\right]$$

ดังนั้นการเลือก weights ต้องสอดคล้องกับข้อจำกัดว่า $\sum_{j=1}^{p} W_{(1)j}^2 = 1$

ทั้งนี้เพื่อทำให้ Var [PC(1)] มีค่าไม่เกิน Total Variation

ถ้า $PC(1) = \mathbf{W}_1' \mathbf{X}$

โดยที่ $\mathbf{W}_1' = [\, W_{(1)1}\ W_{(1)2} \cdots W_{(1)p} \,]$

$\mathbf{X}' = [\, X_1\ X_2 \ldots X_p \,]$

$\text{Var}[\, PC(1) \,] = \mathbf{W}_1' \mathbf{S} \mathbf{W}_1$

ถ้า Var [PC(1)] มีค่ามากที่สุด PC(1) คือ largest cheracteristic root ของ **S**

$\mathbf{W}_1$ คือ characteristic vector ที่สมนัยกับ PC(1)

Second principal component : PC(2)

PC(2) คือ weighted linear combination ของตัวแปรซึ่ง uncorrelated กับ PC(1) ซึ่งสมนัยกับส่วนที่ใหญ่ที่สุดของ total variation ที่เหลืออยู่เมื่อหักส่วนที่สมนัยกับ PC(1) ออกแล้ว หรือกล่าวว่า

$$PC(2) = W_{(2)1}X_1 + W_{(2)2}X_2 + \ldots + W_{(2)p}X_p$$

เป็น linear combination ของ X's ซึ่งมี variance มากที่สุดในบรรดา linear combination อื่น ๆ ซึ่ง uncorrelated กับ PC(1)

โดยทั่ว ๆ ไป principal component ที่ i : PC (i) คือ

$$PC(i) = W_{(i)1}X_1 + W_{(i)2}X_2 + \ldots + W_{(i)p}X_p$$

ซึ่งมี variance สูงสุดในบรรดา linear combination ที่ uncorrelated กับ (i-1) principal components ซึ่งหามาก่อนแล้ว

อันที่จริงแล้วเราอาจหา principal component จำนวนมากเท่าใดก็ได้แต่อย่างไรก็ดี ในทางปฏิบัติส่วนมาก **ส่วนใหญ่ของ Total variation มักจะเนื่องมาจาก principal component 2-3 ตัวแรก** นอกจากนั้น components เหล่านี้ยังถูกเลือกให้ uncorrelated กันทุกคู่ด้วย (mutually uncorrelated) ดังนี้เราจึงบรรลุวัตถุประสงค์ที่ต้องการคือ parsimony และ independence จากการใช้วิธีการของ principal components

6.4.2 การตีความหมายของ Principal Components

ดังที่กล่าวมาแล้วข้างต้นว่า principal components นั้นมักจะยากแก่การตีความหมาย ดังนั้นเราจึงต้องจัดการกับ principal components ดังกล่าวต่อไปอีก

ในกรณีใด ๆ เรามักจะพบว่า first principal components [PC(1)] นั้นเป็นตัวแทนของข่าวสารซึ่งรวมอยู่ในตัวแปรทั้งหมดทุกตัว ดังนั้นจะเห็นว่า "a general index" มักจะมีค่า (ค่าสมบูรณ์) ของ factor loadings ของตัวแปรเกือบทุกตัวสูง ดังนั้นผลที่ตามมาคือเป็นการยากที่จะให้ชื่อ (label) แก่ factor ลักษณะนี้

อย่างไรก็ดีถ้าวัตถุประสงค์ของการวิเคราะห์ปัจจัย คือการลดรูปของข้อมูล (data reduction) โดยการสร้างดรรชนีหรือตัวแปรใหม่ โดยที่ไม่เน้นเรื่องการตีความหมายเป็นหลักแล้ว factors ที่เราจะได้ในขั้นสุดท้ายก็คือ principal components นั่นเอง แต่เราต้องการหา factors ที่มีความหมายเราจึงต้องดำเนินการในขั้นต่อไปคือทำการหมุนแกนเพื่อให้บรรลุวัตถุประสงค์นี้

6.4.3 ตัวอย่าง Principal Components

Principal components สำหรับการศึกษาของเจมส์และคลีนบวมของ Whites และ Nonwhites แสดงในตารางที่ 6.7

ตารางที่ 6.7

Principal Components สำหรับการศึกษาของเจมส์และคลีนบวม

	PRINCIPAL COMPONENTS			
	WHITES		NONWHITES	
VARIABLES	SES	SIS	SES	SIS
PCI	0.90	−0.31	0.90	−0.26
MED	0.60	−0.61	0.72	−0.50
UEM	0.00	0.58	0.01	−0.04
P8K +	0.85	−0.37	0.58	−0.31
WCM	0.76	−0.38	0.63	−0.10
BC	0.37	0.55	0.78	−0.31
P3K −	−0.86	0.16	−0.78	0.42
PSDF	0.79	0.32	0.64	0.49
JDM	0.79	0.05	0.76	0.39
JDF	0.65	0.30	0.79	0.26
CSM	0.51	0.55	0.72	0.27
CSF	0.54	0.55	0.66	0.02
PM	0.35	0.25	0.41	0.58
HR	−0.06	0.42	−0.18	0.41
PCBH	0.09	0.63	0.38	0.63
Variance of principal component	5.70	2.85	6.18	2.13
Proportion of total variation explained by principal component	5.7/15 ‖ 0.38	2.85/15 ‖ 0.19	6.18/15 ‖ 0.41	2.13/15 ‖ 0.14

จากตารางที่ 6.7 จะเห็นได้ว่า สำหรับคนผิวขาว (whites) นั้น principal components 2 ตัวแรกนั้นอธิบายได้ถึง 57% (38 + 19%) ของ Total variation ส่วนสำหรับคนผิวดำ (nonwhites) principal components 2 ตัวแรกนั้นอธิบายได้ถึง 55% (41 + 14%) ของ Total variation proportion .38, .19, .41 และ .14 นั้นคำนวณได้จาก variances ในแถวข้างบน (5.7, 2.85, 6.18 และ 2.13 ตามลำดับ) หารด้วยจำนวนตัวแปรคือ p = 15 นั่นเอง ทั้งนี้เพราะตัวแปรได้ถูก standardized มาล่วงหน้าก่อนการวิเคราะห์

โปรดสังเกตว่า PC(1) ของทั้ง whites และ nonwhites นั้นมี loadings ค่าสูงทั้งสิ้น เราจะได้เห็นในตอนหลังว่าเทคนิคการหมุนแกนนั้นทำให้เราได้ factor ที่มีความหมายมากกว่า ทั้งนี้โดยการลดจำนวน loadings ที่มีค่าสูงในแต่ละ factor ลง

6.4.4 การหมุนแกนของตัวประกอบเริ่มต้น (Rotation of Initial factors)

การหมุนแกนคือวิธีการที่จะเปลี่ยนรูปของ initial factors เพื่อทำให้การตีความหมายของ factor ดีขึ้น ตัวอย่างเช่นการหมุน principal components 2 ตัวแรกซึ่งได้มาจากการศึกษาของเจมส์และคลีนบวมโดยหวังว่า factors ที่ได้จากการหมุนแกนนั้นจะมีความหมายดีกว่าในเทอมของเซตย่อยของตัวแปรเริ่มต้น (original variables)

จุดประสงค์แรกของการหา factor ที่มีความหมายโดยการหมุนแกนนั้น เราอาจกล่าวว่าทำเพื่อให้ได้ simple structure โครงสร้างของ factor ที่ถือว่า simple นั้นคือ ตัวแปรเริ่มต้นแต่ละตัวมีความสัมพันธ์อย่างสูงกับ factor ใดเพียง factor เดียว และจะถือได้ว่าแต่ละ factor เสมือนเป็นตัวแทนของลักษณะร่วม ๆ กันของกลุ่มของตัวแปรซึ่งมีจำนวนตัวแปรค่อนข้างน้อย ดังนี้เราจะบรรลุถึง simple structure ถ้าสำหรับแต่ละ factor นั้น factor loadings สำหรับตัวแปรส่วนใหญ่เข้าใกล้ศูนย์ และ factor loadings ที่เหลืออยู่มีค่าค่อนข้างใหญ่ ถ้าเป็นดังนั้น factor นั้นจะอธิบายความแปรปรวนร่วมกันของตัวแปรในเซตย่อยของตัวแปรซึ่งมีความสัมพันธ์กันอย่างสูงกับ factor และไม่อธิบายความแปรปรวนของตัวแปรอื่น

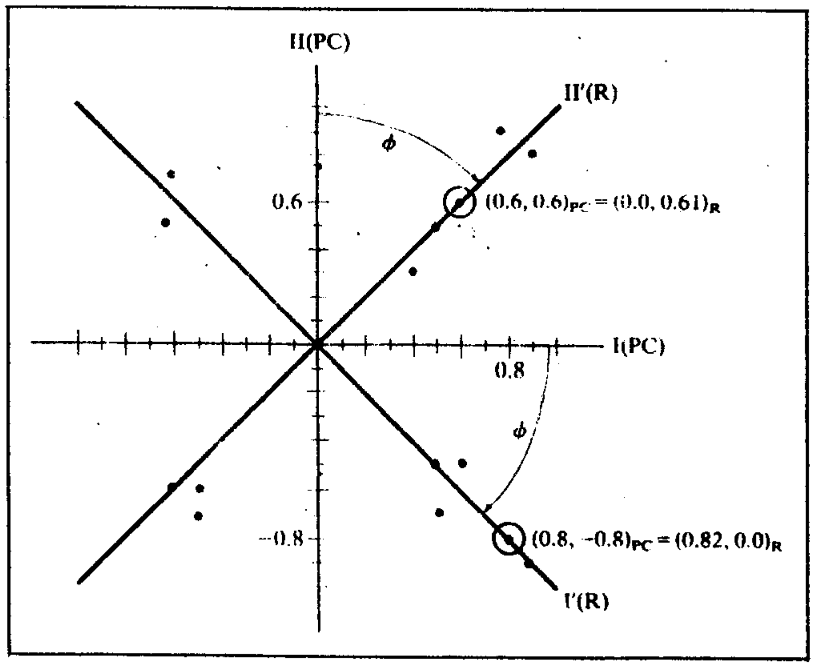
อย่างไรก็ดี เป็นเรื่องสำคัญที่ควรจะต้องยอมรับว่า simple structure ที่ได้จากการหมุนแกนไม่ได้ประกันว่าตัวแปรที่จับกลุ่มกัน (hang together) ในแต่ละ factor นั้นจะให้ความหมายที่เด่นชัด ทั้งนี้เพราะเป็นไปได้ที่ simple structure ที่ค่อนข้างดีจะให้ factors ที่ยากต่อการตีความหมายแต่ถ้าไม่มี simple structure แล้วการตีความหมายของ factors จะยิ่งเป็นไปไม่ได้

วิธีการ 2 วิธีที่ดีที่สุดในการอธิบายว่าการหมุนแกนนั้นจะทำให้ได้ simple structure ได้อย่างไร คือ

1) geometrically โดยการหมุนแกน coordinate
2) numerically โดยการปรับโครงสร้างของ factor loadings

6.4.4.1 การใช้วิธีทางกราฟฟิคแสดงจุดประสงค์ของการหมุนแกน

ในรูปที่ 6.3 แสดงว่าการหมุนแกนจะช่วยให้ได้ simple structure (ในการหมุนแกนที่แสดงในรูปนี้ แสดงเพียงการหมุนของ 2 factors ถ้ามี factor มากกว่า 2 factors ที่จะถูกหมุน การหมุนจะทำการหมุนไปทีละคู่จนกว่าจะได้ simple structure สำหรับทุก factor ที่พิจารณาอยู่) ในรูปแสดงผลที่เราต้องการจะได้จากการหมุนแกน ในรูปจะมีจุดเท่ากับจำนวนตัวแปร และแต่ละจุดสมนัยกับตัวแปรแต่ละตัว coordinates ของจุดแต่ละจุดคือ factor loadings 2 ตัว ของตัวแปรนั้น ๆ จาก 2 factors ที่เราจะทำการหมุน เมื่อหมุนแกนทั้ง 2 ไปแล้ว เราจะกำหนด factor ใหม่ 2 factors ได้



รูปที่ 6.3 ภาพแสดงจุดประสงค์ของการหมุนแกน

จุดประสงค์ในที่นี้คือ การหมุนแกนเพื่อให้ได้

1) แต่ละจุดนั้นใกล้กับแกนใดแกนหนึ่ง (แกนใหม่) เท่านั้น (i.e. ได้ simple structure)

และ 2) จุดต่าง ๆ ที่ใกล้กับแกนเดียวกันจะเป็นตัวกำหนด factor ที่มีความหมาย (meaningful factor)

เนื่องจากว่า initial factors นั้นคือ principal components 2 แกนก่อนการหมุนแกน คือ 2 principal components ในรูปที่ 6.3 เราจะเรียก I(PC) และ II(PC) สำหรับ first principal component และ second principal component ตามลำดับ จุดที่วงกลมล้อมรอบไว้ 2 จุดคือ (0.6, 0.6)

และ (0.8, −0.8) บนแกน I(PC) และ II(PC) เมื่อทำการหมุนแกนทั้ง 2 ไปตามเข็มนาฬิกาเป็นมุม ϕ จุดทั้ง 2 จะมี coordinates (0.0, 0.61) และ (.82, 0.0) บนแกน I′(R) และ II′(R) ตามลำดับ ในทำนองเดียวกันจะเห็นได้ว่าแต่ละ coordinates ของจุดแต่ละจุดที่เหลืออยู่จะมีตัวหนึ่งมีค่าสูงและอีกตัวเข้าใกล้ศูนย์ นั่นคือทำให้เราได้ simple structure จุดสำคัญของความสำเร็จในการหมุนแกนนี้คือ จะเห็นว่าจุดต่าง ๆ (i.e. variables) จะจับกลุ่มกันเป็น 2 กลุ่มย่อย กลุ่มย่อยกลุ่มหนึ่งอยู่ใกล้กับแกนใหม่แกนหนึ่ง และอีกกลุ่มย่อยหนึ่งนั้นอยู่ใกล้กับแกนใหม่อีกแกนหนึ่ง เนื่องจากแกนใหม่ 2 แกนนี้แทน factor ใหม่ 2 factors เราจึงจะทำการแปลความหมายของ factor ใหม่ในเทอมของกลุ่มย่อยของตัวแปรที่อยู่ใกล้กับ factor

6.4.4.2 วิธีการหมุนแกน

การหมุนแกนนั้นทำได้ 2 วิธีคือ

วิธีที่ 1 คือวิธีที่เรียกว่า "Orthogonal rotation" คือการหมุนแกน 2 แกนไปโดยที่แกนใหม่ทั้ง 2 แกนยังคงตั้งฉากกัน

วิธีที่ 2 คือวิธีที่เรียกว่า "Oblique rotation" คือการหมุนแกน 2 แกนไม่ตั้งฉากกัน

ถ้าให้**แกนเก่า**คือ I และ II

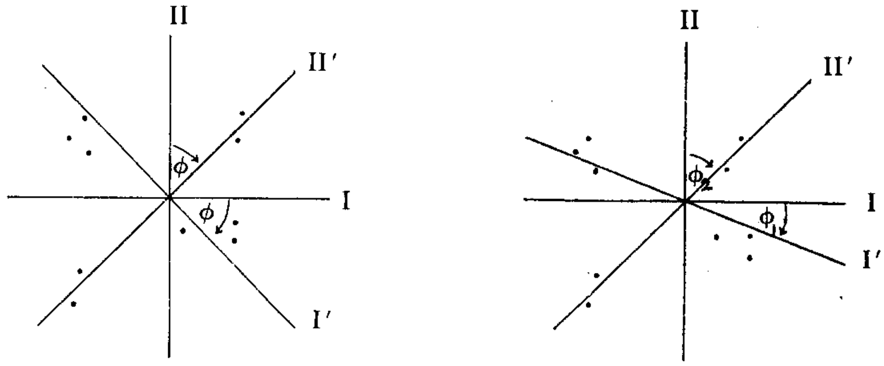
แกนใหม่ คือ I′ และ II′

วิธีแรก นั้นมุมระหว่าง I และ I′ จะเท่ากับมุมระหว่าง II กับ II′ คือเท่ากับ ϕ

วิธีที่ 2 นั้นมุมระหว่าง I และ I′ (คือ ϕ_1) จะไม่เท่ากับมุมระหว่าง II กับ II′ (คือ ϕ_2)

ความแตกต่างที่สำคัญทางสถิติระหว่าง "orthogonal rotation" และ "oblique rotation" คือ factors ที่ได้จากการทำ orthogonal rotation กับ principal components ยังคงเป็นอิสระต่อกันทางสถิติ (i.e. factor cosines จะมีค่าเป็นศูนย์หมด) แต่ factors ที่ได้จากการทำ oblique rotation จะยังคงมีความสัมพันธ์กันอยู่ (i.e. factor cosine บางตัวหรือทุกตัวมีค่า**ไม่เป็นศูนย์**) จุดประสงค์ที่ต้องการคือการที่จะหา factors ที่ไม่มีความสัมพันธ์กันทางสถิติ (statistically uncorrelated) เพราะมีประโยชน์ที่เกี่ยวข้องกับการหาสิ่งที่เป็นดรรชนี (index หรือคือ factor นั่นเอง) เพียง 2-3 ตัว ซึ่งไม่มีความสัมพันธ์กันเพื่อแทนชุดของตัวแปรที่มีความสัมพันธ์กันอย่างซับซ้อน คุณสมบัติซึ่งเป็นที่ต้องการของ orthogonal rotation อีกข้อหนึ่งคือขนาดของ total variation ซึ่งเนื่องมาจาก factors ภายใต้การพิจารณาของเรานั้นไม่ถูกกระทบกระเทือนจากการหมุนแกน ตัวอย่างเช่นในการศึกษาของเจมส์และคลีนบวมเมื่อใช้วิธี orthogonal rotation กับ 2 principal components แล้วจะเห็นว่า factor 2 factors ที่ได้จากการหมุนแกนนั้นคิดเป็น 57% ของ total variation ซึ่งคือค่าเดียวกับเปอร์เซ็นต์ที่คิดจาก 2 principal components

แต่อย่างไรก็ดี โชคไม่ดีที่การหมุนแกนแบบ orthogonal นั้นอาจให้ factors ที่ไม่ใช่ชุดของ factors ที่ดีที่สุด ดังนั้นถ้าการหมุนแกนแบบ oblique จะทำให้บรรจุวัตถุประสงค์ของการหมุนแกนที่ต้องการคือเพื่อให้ได้ simple structure และ meaningful structure เราก็จะใช้วิธี oblique ในการหมุน ซึ่งมักจะทำให้ได้ factors ที่มีความหมายมากกว่าเสมอ อย่างไรก็ดีก่อนที่จะตัดสินใจว่าจะใช้วิธีใดในการหมุนแทนเราอาจใช้วิธีทางกราฟฟิคช่วยได้



a) Orthogonal rotation b) Oblique rotation

รูปที่ 6.4 การหมุนแกนแบบ Orthogonal และ Oblique

6.4.4.3 Computer Algorithms สำหรับ Orthogonal Rotation

Algorithm 3 วิธีสำหรับการหมุนแกนแบบ orthogonal ที่เรามักจะพบใน Factor-analysis computer program ส่วนมากคือ

1) Varimax เป็นวิธีที่นิยมใช้มากที่สุดระหว่าง 3 วิธีนี้ วิธีนี้พยายามจะให้ได้ simple structure with respect to (w.r.t.) columns ของ factor loading matrix นั่นคือเลือกการหมุนแกนที่จะทำให้ความแปรปรวนของกำลังสองของสมาชิกใน column ของ factor loading matrix มีค่าน้อยที่สุด

ถ้า k = จำนวน factor

p = จำนวนตัวแปร

l_{ij} $(i = 1, \ldots, p : j = 1, \ldots k)$ เป็นสมาชิกใน row ที่ i และ column ที่ j ของ factor loading matrix

ให้ $d_j = \sum_{i=1}^{p} l_{ij}^2, \; j = 1, \ldots, k$

$$p^{-1} d_j = \frac{1}{p} \left[\sum_{i=1}^{p} l_{ij}^2 \right]$$

ทำให้จำนวนต่อไปนี้มีค่ามากที่สุด

$$\sum_{j=1}^{k} \sum_{i=1}^{p} \left(l_{ij}^2 - p^{-1} d_j \right)^2$$

วิธีการดังกล่าวพยายามที่จะทำให้สมาชิกในคอลัมน์ของ factor loading matrix มีขนาดใหญ่ หรือให้มีค่าเป็นศูนย์ นั่นคือพยายามที่จะหา factor ซึ่งจะมีความสัมพันธ์กับตัวแปรสูงหรือไม่มีความสัมพันธ์กันเลย

2) Quartimax วิธีนี้ทำให้ cross products ของ factor loadings มีค่าน้อยที่สุด

ถ้า $\mathbf{L} = (l_{ij})$ เป็น factor loading matrix

วิธีนี้จะทำให้จำนวนต่อไปนี้มีค่าน้อยที่สุด

$$\sum_{1 \le s < t \le k} \sum_{i=1}^{p} (l_{is}\, l_{it})^2$$

ซึ่งจะ equivalent กับการทำให้ $\sum_{j=1}^{k} \sum_{i=1}^{p} l_{ij}^4$ มีค่ามากที่สุด อันเป็นที่มาของชื่อว่า "quartimax"

3) Equamax เป็นวิธีหมุนแกนแบบ orthogonal อีกวิธีหนึ่ง

6.4.4.4 Computer Algorithms สำหรับ Oblique Rotation

สำหรับการหมุนแกนวิธีนี้มี Computer algorithm จำนวนมาก แต่ที่มักจะหาใช้ได้สะดวกคือ Oblimin, quartimin, biquartimin และ Covarimin algorithm ซึ่งแต่ละวิธีออกแบบเพื่อให้ได้ simple structure ซึ่งมีลักษณะต่าง ๆ กัน และไม่มีวิธีใดเพียงวิธีเดียวที่จะให้ผลลัพธ์ที่ดีเลิศเลยดังนั้นเราอาจจะต้องลองใช้วิธีต่าง ๆ กันสำหรับข้อมูลชุดเดียวกัน

สรุป

จากวิธี Varimax และ quartimax นั้น จริง ๆ แล้วเราทำให้จำนวนต่อไปนี้มีค่าน้อยที่สุด

$$G = \sum_{1 \le i < j \le k} \left[\sum_{s=1}^{p} l_{si}^2\, l_{sj}^2 - \delta p^{-1} d_i\, d_j \right]$$

โดยที่ $d_t = \sum_{s=1}^{p} l_{st}^2$ และ δ เป็นตัวกำหนดชนิดของการหมุนแกน

สำหรับ orthogonal rotations เราเรียกวิธีการต่าง ๆ ดังนี้คือ

$\delta = 0$ คือ Quartimax

$\delta = 1$ คือ Varimax

$\delta = \frac{k}{2}$ คือ Equamax

ใช้ค่าต่ำของ δ และพยายาม simplify rows ของ factor loading matrix

ใช้ค่าสูงของ δ และพยายาม simplify columns ของ factor loading matrix

สำหรับ oblique rotations นั้นจะลดค่าของ G ลง (แต่ไม่ใช่ทำให้มีค่าต่ำที่สุด) เราเรียกวิธีการต่าง ๆ ดังนี้คือ

$\delta = 0$ คือ Direct quartimin

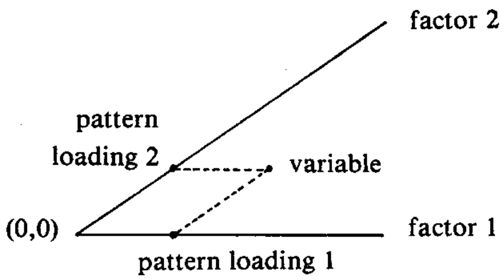
$\delta = \frac{1}{2}$ คือ Biquartimin

$\delta = 1$ คือ Covarimin

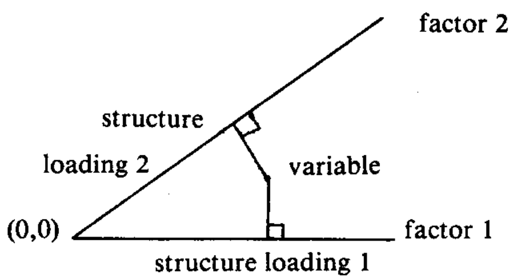
ผลลัพธ์จากการใช้ Computer algorithm สำหรับ oblique rotation นั้นมีสิ่งที่ต้องพิจารณาเพิ่มเติม (ใน orthogonal rotation จะไม่พบปัญหานี้) นั่นคือจะมีวิธีแสดงค่าของ factor loadings (coordinates ของจุดหรือตัวแปรบนแกนใหม่) 2 วิธีคือ

1) วิธีที่เรียกว่า pattern loadings (รูปที่ 6.5) ซึ่งจะฉาย (project) จุดแต่ละจุดไปยังแกนใหม่ โดยใช้เส้นที่ขนานกับแกนใหม่ทั้ง 2

2) วิธีที่เรียกว่า structure loadings (รูปที่ 6.6) ซึ่งจะฉาย (project) จุดแต่ละจุดไปยังแกนใหม่โดยใช้เส้นที่ตั้งฉากกับแกนใหม่ทั้ง 2



รูปที่ 6.5 Pattern loadings

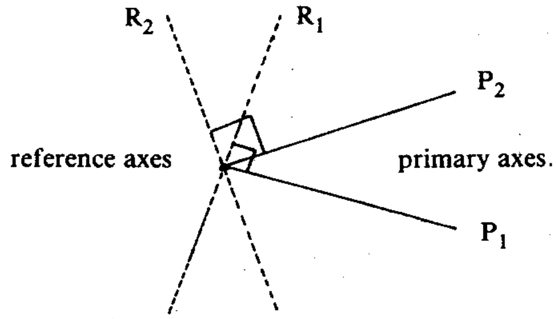


รูปที่ 6.6 structure loadings

Computer package บาง package เช่น SPSS จะให้ทั้ง 2 factor loading matrices ซึ่งเรียกว่า Pattern matrix และ Structure matrix สำหรับข้อมูลชุดเดียวกัน แต่บาง package (เช่น BMDP) จะให้ผู้ใช้เลือกระบุว่าจะเอา matrix ใดใน print out

ความแตกต่างระหว่าง pattern loading และ structure loading คือ pattern loading นั้นไม่ใช่ correlation coefficient ระหว่างตัวแปรและ factor **ในขณะที่ structure loading ยังคงแทน correlation ดังกล่าว อย่างไรก็ดี pattern matrix นั้นมักจะมีประโยชน์มากกว่า structure matrix** เมื่อใช้ในการตีความหมายของ factor ที่ได้จากการหมุนแกน

ลักษณะเพิ่มเติมของการหมุนแกนแบบ oblique นั้นคือเราอาจเลือกใช้ค่าบนแกนที่ถูกปรับแล้วคือปรับแกนที่ได้จากการหมุนแกนอีกที (มักจะถูกเรียกว่า reference) หรือค่าที่ได้บนแกนใหม่จากการหมุนแกน (original หรือ primary) แกนที่ถูกปรับแล้ว 2 แกน (new adjusted axes) จะตั้งฉากกับ primary axes ดูรูปที่ 6.7



รูปที่ 6.7 แสดงความสัมพันธ์ระหว่าง reference axes และ primary axes

อย่างไรก็ดี reference pattern loadings คือ correlation coefficients (เช่นเดียวกับ primary structure loadings ในขณะที่ reference structure loadings (ซึ่งเหมือนกับ primary pattern loadings) ไม่ใช่ correlation coefficients บาง computer program จะให้หรือแสดง reference structure loadings และ primary pattern loadings เท่านั้น เพราะใช้ประโยชน์ในการแปลความหมายได้ดีกว่า

6.4.4.5 การหมุนแกนในการศึกษาของเจมส์และคลีนบวม

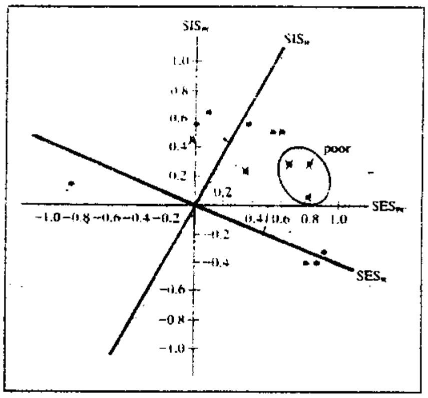
รูป 6.8 และ 6.9 แสดงสิ่งที่ได้จากการทำการหมุนแกนแบบ Oblique ต่อ principal components 2 components

สัญลักษณ์สำหรับจุดในรูปคือ ● = SES - type variable
✖ = SIS - type variable

จุดที่อยู่ในวงกลมคือตัวแปรซึ่งไม่รวมอยู่ใน factor ใด factor หนึ่งอย่างเด่นชัด ตัวแปรเหล่านี้เราจะเห็นได้เมื่อเราพิจารณา factor - loading matrices แต่จากรูปจะเห็นได้ว่าผลของการหมุนแกนนี้ยังห่างไปจากสิ่งที่เราต้องการเพราะจุด (●) ทั้งหมดยังไม่เกาะอยู่กับแกนใดแกนหนึ่งเพียงแกนเดียว และทำนองเดียวกับจุด (✖) ก็ไม่เกาะอยู่กับแกนใดแกนหนึ่งเพียงเพียงแกนเดียวเช่นกัน

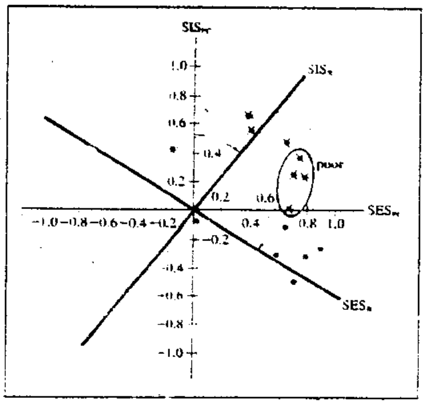
แต่อย่างไรก็ดียังพอประมาณได้ว่าตัวแปรในแต่ละกลุ่ม (SES และ SIS) จับกลุ่มกันอยู่กับแกนใดแกนหนึ่ง

(*: SIS-type variable; •: SES-type variable)



รูปที่ 6.8 **Oblique rotation : Whites**

(*: SIS-type variable; •: SES-type variable)



รูปที่ 6.9 **Oblique rotation : Nonwhites**

6.4.4.6 Examination of Factor-loading Matrices

จากตารางที่ 6.8 คือ Idealized factor loadings สำหรับการศึกษาของเจมส์และคลีนบวม

ตารางที่ 6.8

Idealized factor loadings for J-K study

VARIABLE	SES	SIS
PCI	High(+)	0
MED	High(+)	0
UEM	High(−)	High(+)
P8K+	High(+)	0
WCM	High(+)	0
BC	High(+)	?
P3K−	High(−)	?
PSDF	0	High(+)
JDM	0	High(+)
JDF	0	High(+)
CSM	0	High(+)
CSF	0	High(+)
PM	0	High(+)
HR	0	High(−)
PCBH	0	High(+)

โปรดสังเกตว่า SES factor นั้นเราคาดว่าจะประกอบด้วยตัวแปรที่เป็น SES variables ซึ่งมี factor loading สูงและตัวแปรที่เป็น SIS variables ซึ่งมี factor loadings ใกล้ศูนย์ และในทางตรงกันข้าม SIS factor นั้นคาดว่าควรจะประกอบด้วยตัวแปรที่เป็น SIS variables ซึ่งมี factor loadings สูงและตัวแปรที่เป็น SES variables ซึ่งมี factor loadings ใกล้ศูนย์ **ตัวแปรบางตัวเช่น BC และ P3K มีค่าของ factor loading ไม่เด่นชัดใน SIS factor จึงใส่เครื่องหมาย ?** ไว้

6.4.4.6.1 การเปรียบเทียบ Factor Loadings ของ Initial Factors กับ Rotated Factors (factors ที่ได้จากการหมุนแกน)

ตารางที่ 6.9 และ 6.10 แสดงว่าการหมุนแกนช่วยปรับปรุง factor-loading matrix โดยคำนึงถึง simple structure ในตารางทั้ง 2 แสดง factor loadings สำหรับ principal components และ factor loadings ของ oblique (reference)-factor solution สำหรับ whites และ nonwhites จากการศึกษาของเจมส์และคลีนบวม **factor loadings ที่ถือว่า "poor" เมื่อเทียบกับ idealized loadings (ตารางที่ 6.8) จะถูกวงรอบไว้** ให้พิจารณาจำนวน loadings ที่ถูกวงรอบไว้ว่ามีจำนวนน้อยลงอย่างมากสำหรับ factor ที่เป็นผลจากการหมุนแกนเมื่อเทียบกับ principal components ดังนั้นถึงแม้ว่า factors ที่ได้รับจากการหมุนแกนจะไม่อยู่ในรูป simple structure ที่สมบูรณ์ แต่เราก็ได้ factor ที่มีความหมายมากกว่าก่อนทำการหมุนแกนคือได้ SES factor และ SIS factor

ตารางที่ 6.9

Principal Components และ Oblique - rotated factor loadings สำหรับ Whites ใน J - K study

VARIABLE	PRINCIPAL COMPONENTS		OBLIQUE FACTORS	
	SES	SIS	SES	SIS
PCI	0.90	-0.31	0.95	0.04
MED	0.69	−0.61	0.82	−0.34
UEM	0.00	0.58	−0.27	0.54
P8K-	0.85	−0.37	0.93	−0.03
WCM	0.76	−0.38	0.86	−0.06
BC	0.37	0.55	0.06	0.65
P3K-	−0.86	0.16	−0.84	−0.17
PSDF	0.79	0.32	0.55	0.59
JDM	0.79	0.05	0.67	0.34
JDF	0.65	0.30	0.44	0.53
CSM	0.51	0.55	0.19	0.71
CSF	0.54	0.55	0.21	0.71
PM	0.35	0.25	0.19	0.37
HR	−0.06	0.42	−0.25	0.37
PCBH	0.09	0.63	-0.22	0.62

ตารางที่ 6.10

Principal Components และ Oblique - rotated factor loadings สำหรับ nonwhites ใน J - K study

VARIABLE	PRINCIPAL COMPONENTS		OBLIQUE FACTORS	
	SES	SIS	SES	SIS
PCI	0.90	−0.26	0.88	0.22
MED	0.72	−0.50	0.88	−0.07
UEM	0.01	−0.04	0.03	−0.03
P8K-	0.58	−0.31	0.65	0.03
WCM	0.63	−0.10	0.56	0.23
BC	0.78	-0.31	0.81	0.12
P3K-	−0.78	0.42	−0.88	−0.03
PSDF	0.64	0.49	0.18	0.75
JDM	0.76	0.39	0.34	0.73
JDF	0.79	0.26	0.44	0.62
CSM	0.72	0.27	0.38	0.60
CSF	0.66	0.02	0.50	0.35
PM	0.41	0.58	−0.06	0.71
HR	−0.19	0.41	−0.41	0.26
PCBH	0.36	0.63	−0.12	0.75

6.4.4.6.2 การเปรียบเทียบการหมุนแกน

ตารางที่ 6.11 และ 6.12 แสดงว่ามีความแตกต่างกันเล็กน้อยในผลลัพธ์จากวิธีการหมุนแกน 3 วิธีเหล่านั้น ตามที่เราคาดไว้

สำหรับ Whites จำนวน poor loadings ของทั้ง 3 วิธีเท่ากันแต่สำหรับ Nonwhites จำนวน poor loadings ของวิธี oblique มีน้อยกว่าอีก 2 วิธีอยู่ 1 ตัว

ตารางที่ 6.11

Factor loadings สำหรับการหมุนแกน 3 วิธี : Whites

VARIABLE	REFERENCE OBLIQUE		ORTHOGONAL VARIMAX		ORTHOGONAL QUARTIMAX	
	SES	SIS	SES	SIS	SES	SIS
PCI	0.95	0.04	0.95	0.10	0.95	0.10
MED	0.82	−0.34	0.80	−0.30	0.80	−0.29
UEM	−0.27	0.54	−0.24	0.52	−0.24	0.52
P8K+	0.93	−0.03	0.93	0.02	0.93	0.03
WCM	0.86	−0.06	0.85	−0.01	0.85	−0.02
BC	0.06	0.65	0.10	0.66	0.10	0.66
P3K−	−0.84	−0.17	−0.85	−0.22	−0.85	−0.22
PSDF	0.55	0.59	0.59	0.62	0.58	0.62
JDM	0.67	0.34	0.69	0.38	0.69	0.38
JDF	0.44	0.53	0.47	0.55	0.46	0.55
CSM	0.19	0.71	0.23	0.72	0.22	0.72
CSF	0.21	0.71	0.25	0.73	0.25	0.73
PM	0.19	0.37	0.21	0.38	0.20	0.38
HR	−0.25	0.37	−0.23	0.36	−0.23	0.36
PCBH	−0.22	0.62	−0.19	0.60	−0.19	0.60

ตารางที่ 6.12

Factor loadings สำหรับการหมุนแกน 3 วิธี : Nonwhites

VARIABLE	REFERENCE OBLIQUE		ORTHOGONAL VARIMAX		ORTHOGONAL QUARTIMAX	
	SES	SIS	SES	SIS	SES	SIS
PCI	0.88	0.22	0.88	0.34	0.92	0.19
MED	0.88	−0.07	0.87	0.05	0.87	−0.09
UEM	0.03	−0.03	0.03	−0.03	0.03	−0.03
P8K+	0.65	0.03	0.65	0.12	0.66	[illegible]
WCM	0.56	0.23	0.56	0.31	0.61	0.21
BC	0.81	0.12	0.80	0.23	0.83	0.16
P3K−	−0.88	−0.03	−0.87	−0.15	−0.88	−0.01
PSDF	0.18	0.75	0.21	0.78	0.33	0.73
JDM	0.34	0.73	0.36	0.76	0.48	0.77
JDF	0.44	0.62	0.46	0.69	0.57	0.60
CSM	0.38	0.60	0.40	0.68	0.50	0.58
CSF	0.50	0.35	0.51	0.42	0.51	0.33
PM	−0.06	0.71	−0.04	0.70	0.08	0.70
HR	−0.41	0.26	−0.40	0.21	−0.36	0.27
PCBH	−0.12	0.75	−0.09	0.73	0.03	0.74

6.5 การหาค่าของปัจจัย (Factor scores)

ตารางที่ 6.13

Factor weights จาก (reference) Oblique rotation ในการศึกษาของเจมส์และคลีนบวม

VARIABLE	WHITES		NONWHITES	
	SES	SIS	SES	SIS
PCI	0.19	−0.07	0.19	−0.00
MED	0.18	0.14	0.22	−0.10
UEM	−0.07	0.18	0.01	0.01
P6K+	0.19	−0.04	0.15	−0.05
WCM	0.17	0.05	0.11	0.07
BC	−0.01	0.20	0.18	−0.05
P3K−	−0.16	−0.02	−0.21	0.07
PSDF	0.09	0.16	−0.02	0.24
JDM	0.12	0.08	0.01	0.22
JDF	0.07	0.15	0.05	0.17
CSM	0.01	0.21	0.04	0.17
CSF	0.02	0.21	0.09	0.07
PM	0.02	[illegible]	−0.08	0.25
HR	−0.06	0.13	−0.12	0.13
PCBH	−0.07	0.20	−0.10	0.27

จากตารางที่ 6.13 แสดง factor weights ซึ่งสัมพันธ์กับ factor loadings ที่ได้จาก (reference) oblique rolation ในการศึกษาของเจมส์และคลีนบวมดังที่ได้กล่าวมาแล้วว่า factor scores คือ ค่าที่เป็นตัวเลขของ factor หนึ่ง (F) ซึ่งได้มาจากการแทนค่า (standardized) X's ในนิพจน์

$$F = W_1X_1 + W_2X_2 + \dots + W_pX_p$$

การคำนวณค่า factor scores ซึ่งจะใช้ในการวิเคราะห์เพื่อตอบปัญหาว่า "Do high-stress counties have higher mortality rates than low-stress counties?"

6.5.1 สรุปวิธีวิเคราะห์กรณีตัวอย่าง

จากการศึกษาของเจมส์และคลีนบวมเราใช้แสดงการใช้ประโยชน์ของ factor scores เพื่อตอบปัญหาข้างต้นที่อยู่ในความสนใจของนักวิจัย

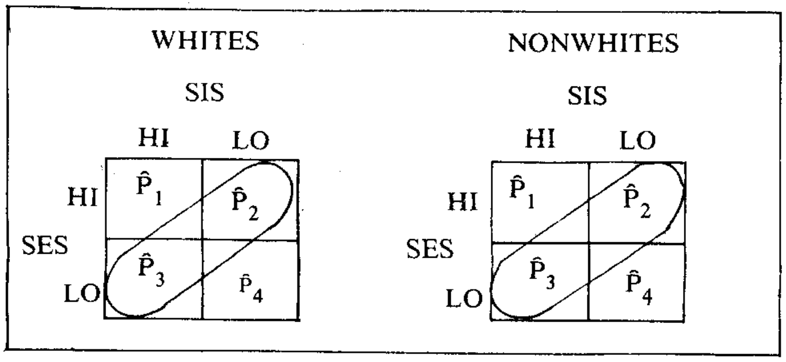
ขั้นตอนที่ทำเพื่อใช้ในการวิเคราะห์คือ

1) factor scores สำหรับ counties ถูกใช้ในการเรียงลำดับเป็นขั้นแรก ซึ่งจะเรียงแยกกันสำหรับแต่ละกลุ่มสีผิว การเรียงคะแนน 86 คะแนนนี้ทำให้เราได้ 4 กลุ่ม แต่ละกลุ่มมีลักษณะเป็น combination ของ factor (SES และ SIS) และสีผิว (Whites และ Nonwhites)

2) สำหรับแต่ละกลุ่มในข้อ 1) factor scores ของ counties จะถูกแบ่งเป็น 2 กลุ่มโดยใช้ median เป็นจุดแบ่ง กลุ่มหนึ่งเรียก "high" scores on the factor อีกกลุ่มหนึ่งเรียก "low" scores on the factor นั่นคือจะได้ 4 กลุ่มดังนี้

SES - W	SIS - W	SES - NW	SIS - NW
HI	HI	HI	HI
LO	LO	LO	LO

3) สำหรับแต่ละสีผิวจะจัด 2 กลุ่มของ HI-LO ในข้อ 2 ในลักษณะตารางแบบ 2 ทาง ดังแสดงในรูปที่ 6.10



รูปที่ 6.10 แสดง diagram สำหรับวิธีการวิเคราะห์ในการศึกษาของเจมส์และคลีนบวม

จากรูปที่ 6.10 นั้นใน 4 เซล เราสนใจเพียง 2 เซลคือ

"HI-SES, LO - SIS" คือกลุ่ม low stress → "healthy" county ซึ่งคาดว่ามี low mortality rate และ "LO - SES, HI - SIS" คือกลุ่ม high stress → "unhealthy" county ซึ่งคาดว่ามี high mortality rate

ส่วนอีก 2 เซลนั้น (HI-SES, HI-SIS) และ (LO-SES, LO-SIS) เราไม่สนใจในการวิเคราะห์

4) สำหรับแต่ละเซลในตารางแบบ 2 ทาง 2 ตาราง (ตารางหนึ่งสำหรับ Whites และอีกตารางหนึ่งสำหรับ Nonwhites) จะคำนวณอัตราการตายที่เกี่ยวข้องกับโรคความดันโลหิตสูงในช่วงเวลา 3 ปี อย่างหยาบ ๆ สำหรับ age-sex specific groups (ในที่นี้จะแสดงผลสำหรับชายอายุ 45-54 ปี)

การเปรียบเทียบเบื้องต้นนั้นจะทำระหว่าง $\hat{P}_2$ และ $\hat{P}_3$ (the "off-diagonal") rates ในแต่ละตารางในรูปที่ 6.10

6.5.2 การศึกษาผลการวิเคราะห์

เจมส์และคลีนบวมคาดไว้ว่าอัตราการตายที่เกี่ยวข้องกับโรคความดันโลหิตสูงสำหรับ Nonwhite males ใน high-stress counties จะสูงกว่าค่าเดียวกันของ low-stress counties อย่าง

มีนัยสำคัญ ไม่คาดว่าจะมี socioecologic stress ต่อ rates เหล่านี้สำหรับ white males ดำเนินการตามข้อ **2** ใน **6.5.1** ได้ตารางที่ **6.14** และ **6.15**

ตารางที่ 6.14

Standardized factor scores in the four stress groups: White males. 45-54 years of age.

	N=25		N=18		N=18		N=25	
	HI SES	HI SIS	HI SES	LO SIS	LO SES	HI SIS	LO SES	LO SIS
	1.51	0.03	0.25	0.61	0.13	0.71	−1.00	−0.31
	1.09	1.37	0.63	0.79	−0.83	0.53	−0.64	0.38
	0.79	1.54	0.45	−0.25	−1.08	0.40	−0.41	−1.70
	0.15	2.75	2.58	−0.25	−0.35	0.99	−0.59	−0.39
	0.23	0.91	0.59	−1.84	−0.14	−0.04	−0.57	−1.52
	0.84	0.73	0.59	−0.15	−0.84	0.06	−0.24	−0.76
	1.07	0.11	3.09	−0.42	−0.35	−0.03	−0.80	−0.10
	0.37	0.76	0.25	−1.64	−0.24	0.12	−1.05	−1.12
	2.73	0.35	1.75	−1.31	−0.84	0.13	−0.75	−0.43
	0.60	1.39	0.78	−0.79	−1.10	1.92	−0.63	−1.65
	2.29	0.72	0.39	−1.29	−0.36	0.07	−1.13	−1.84
	0.33	0.55	−0.11	−0.30	−0.54	1.41	−1.05	−1.53
	0.40	0.48	0.01	−0.73	−0.45	0.91	−1.31	−0.52
	0.58	0.70	2.77	−0.27	−0.50	0.56	−1.64	−0.87
	0.34	0.61	−0.07	−1.32	−0.16	2.24	−0.36	−0.74
	0.45	0.30	−0.08	−0.68	−1.44	2.70	−0.57	−0.71
	0.30	0.23	0.36	−0.36	−1.40	0.30	−0.92	−0.37
	0.47	0.22	0.14	−0.48	−1.19	0.97	−1.18	−0.27
	2.07	1.37					−0.65	−1.66
	0.23	0.00					−0.33	−0.84
	0.08	1.20					−1.11	−1.10
	−0.06	1.54					−0.25	−0.32
	0.03	0.69					−1.27	−0.55
	0.74	0.58					−0.61	−0.24
	0.12	0.59					−0.86	−0.39
[illegible]	17.66	19.78	13.95	−13.40	−11.94	13.73	−20.07	20.3[illegible]
[illegible]an	0.71	0.79	0.76	−0.74	−0.66	0.76	−0.90	−0.8[illegible]

ตารางที่ 6.15

Standardized factor scores in the four stress groups: Nonwhite males, 45-54 years of age.

	$N = 23$		$N = 20$		$N = 20$		$N = 23$	
	HI SES	HI SIS	HI SES	LO SIS	LO SES	HI SIS	LO SES	LO SIS
	1.40	1.63	0.87	−0.57	−0.72	−0.06	−1.02	−0.74
	0.51	1.52	1.04	−0.52	−1.30	0.33	−0.48	−0.61
	1.37	0.12	1.55	−1.27	−0.60	0.51	−0.24	−1.02
	1.40	0.51	1.11	−0.18	−1.05	0.25	−0.20	−1.28
	−0.09	0.42	0.49	−1.34	−1.03	0.02	−0.80	−1.29
	1.24	0.22	0.92	−1.06	−0.91	0.39	−0.30	−0.25
	1.21	0.46	1.70	−1.28	−1.62	0.04	−0.64	−0.58
	1.53	2.19	−0.04	−0.30	−0.99	0.28	−0.29	−1.44
	1.93	2.68	1.28	−1.24	−1.46	0.11	−0.61	−0.65
	0.81	0.38	1.23	−1.30	−1.10	0.82	−1.67	−0.40
	2.47	1.06	0.96	−0.26	−0.37	0.72	−1.35	−0.26
	0.16	2.61	0.57	−0.78	−1.05	0.56	−0.21	−0.69
	0.15	0.08	0.43	−0.52	−0.45	−0.13	−1.67	−0.28
	−0.10	1.46	1.07	−0.45	−1.33	0.37	−0.33	−0.83
	1.61	2.38	0.82	−0.74	−1.29	0.35	−1.21	−0.60
	−0.06	0.99	0.05	−0.87	−0.93	−0.13	−0.57	−0.82
	0.94	3.29	0.23	−0.44	−0.79	0.36	−0.62	−0.84
	−0.02	0.22	1.50	−1.75	−0.32	0.26	−0.87	−1.07
	1.05	−0.14	0.17	−0.83	−0.86	−0.03	−0.55	−0.25
	0.53	0.90	0.25	−0.25	−1.04	0.94	−0.79	−0.67
	1.25	0.05					−1.04	−0.36
	0.75	1.72					−1.24	−0.66
	−0.07	1.34					−0.21	−0.65
Total	20.00	26.09	16.13	−15.93	−19.21	5.96	−16.91	−16.14
Mean	0.87	1.13	0.81	−0.80	−0.96	0.30	−0.74	−0.70

ดำเนินการตามข้อ 3 ใน 6.5.1 ได้ตารางที่ 6.16 และ **6.17**

ตารางที่ 6.16

อัตรา (อย่างหยาบ) การตายที่เกี่ยวข้องกับโรคความดันโลหิตสูง ใน

four stress groups : White males, 45-54 years of age.

SES	SIS HIGH	SIS LOW	MARGINALS
High	P_1: No. counties = 25; No. deaths = 75; PAR = 79,852; Mortality rate = 926.7×10^{-6}	P_2: No. counties = 18; No. deaths = 30; PAR = 42,946; Mortality rate = $\boxed{698.5 \times 10^{-6}}$	846.91×10^{-6}
Low	P_3: No. counties = 18; No. deaths = 24; PAR = 23,940; Mortality rate = $\boxed{1{,}002.5 \times 10^{-6}}$	P_4: No. counties = 25; No. deaths = 32; PAR = 24,298; Mortality rate = $1{,}316.9 \times 10^{-6}$	$1{,}160.91 \times 10^{-6}$
Marginals	944.19×10^{-6}	922.01×10^{-6}	

ตารางที่ 6.17

อัตรา (อย่างหยาบ) การตายที่เกี่ยวข้องกับโรคความดันโลหิตสูง ใน

four stress groups : Nonwhite males, 45-54 years of age.

SES	SIS HIGH	SIS LOW	MARGINALS
High	P_1: No. counties = 23; No. deaths = 156; PAR = 22,190; Mortality rate = $7{,}030.1 \times 10^{-6}$	P_2: No. counties = 20; No. deaths = 18; PAR = 4,540; Mortality rate = $\boxed{3{,}964.7 \times 10^{-6}}$	$6{,}509.53 \times 10^{-6}$
Low	P_3: No. counties = 20; No. deaths = 113; PAR = 14,391; Mortality rate = $\boxed{7{,}852.1 \times 10^{-6}}$	P_4: No. counties = 23; No. deaths = 50; PAR = 9,336; Mortality rate = $5{,}355.6 \times 10^{-6}$	$6{,}863.81 \times 10^{-6}$
Marginals	$7{,}253.54 \times 10^{-6}$	$4{,}900.54 \times 10^{-6}$	

จากตารางที่ 6.17 และ 6.18 ให้ข้อมูลที่จำเป็นสำหรับการทดสอบสมมติฐาน อัตราที่สนใจคือ (P_2 : ของ low stress group, P_3 : ของ high stress group)

Whites : $\hat{P}_3$ = 1002.5 deaths/million

$\hat{P}_2$ = 698.5 deaths/million

จากการใช้ one-tailed Z -test ยอมรับว่า $P_2 = P_3$ ($P > .10$)

Nonwhites : $\hat{P}_3$ = 7852.1 deaths/million

$\hat{P}_2$ = 3964.7 deaths/million

จากการใช้ one-tailed Z-test ไม่ยอมรับว่า $P_2 = P_3$ $(P < .005)$

หมายเหตุ $P = Pr[\, Z > z_c \,]$

ดังนั้นผลการศึกษาของเจมส์และคลีนบวมให้หลักฐานสนับสนุนทฤษฎีที่ว่า socioecologic stress เป็นปัจจัยที่ใช้ในการค้นหาเกี่ยวกับ cerebrovascular desease สำหรับประชากร ซึ่งมีที่อยู่ตามที่เขาทั้งสองใช้ในการศึกษานี้ แต่อย่างไรก็ดี ควรจะมีนักวิจัยอื่นทำการศึกษาอย่างเป็นอิสระจากงานศึกษานี้เพื่อช่วยยืนยันการค้นพบของเจมส์และคลีนบวม

แบบฝึกหัดที่ 6

6.1 จากผลแอปเปิลพันธุ์โจนาธาน ได้วัดค่าต่าง ๆ ต่อไปนี้

X_1 = ไนโตรเจนรวม

X_2 = โปรตีนไนโตรเจน

X_3 = ฟอสฟอรัส

X_4 = โปรแตสเซียม

X_5 = แคลเซียม

X_6 = แมกนีเซียม

X_7 = น้ำหนักเฉลี่ย

X_8 = เปอร์เซ็นต์ของส่วนที่ขม

Correlation matrix จากค่าที่วัดได้ (37 df) คือ

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1							
X_2	0.709	1						
X_3	0.803	0.795	1					
X_4	0.491	0.616	0.664	1				
X_5	-0.026	-0.375	-0.289	-0.548	1			
X_6	0.684	0.539	0.761	0.731	-0.099	1		
X_7	-0.024	0.163	0.202	0.603	-0.608	0.309	1	
X_8	-0.056	0.324	0.302	0.459	− 0.583	0.247	0.562	1

1. จงหา Initital factor loadings
2. จงหมุนแกนโดยวิธี Varimax และ Equamax เพื่อหา final factor loadings
3. จงแปลความหมายของ factors โดยเฉพาะอย่างยิ่งจงพิจารณาว่าความขมมีความเกี่ยวพันกับแร่ธาตุอย่างไรหรือไม่

6.2 ค่าในตารางต่อไปนี้คือสหสัมพันธ์ระหว่างความชุกชมของสัตว์พันธุ์ต่าง ๆ ในอเมริกาเหนือ

ตัวแปร	1	2	3	4	5	6	7
1	1	-0.002	0.81	0.69	0.34	-0.17	-0.65
2		1	0.11	0.50	0.38	-0.14	0.40
3			1	0.80	0.42	-0.85	-0.65
4				1	-0.18	0.95	0.70
5					1	-0.47	0.05
6						1	0.61
7							1

โดยที่ตัวแปรตัวที่

1 คือ กบ

2 คือ กิ้งก่า

3 คือ เต่า

4 คือ งู

5 คือ นกกินแมลง

6 คือ นกกินเมล็ดพืช

7 คือ สัตว์เลี้ยงลูกด้วยนมหรือสัตว์พวกจิงโจ้

จงทำการวิเคราะห์ปัจจัยสำหรับข้อมูลนี้

6.3 ตัวแปรทางโสตทัศน์ 13 ตัว คือ

ตัวแปรตัวที่

1 คือ Verbal Communication

2 คือ Experiential evaluation

3 คือ Induction (Visual)

4 คือ Auditory induction

5 คือ Memory span

6 คือ Temporal tracking

7 คือ Sound pattern recognition

8 คือ Relation perception

9 คือ Spatial scanning

10 คือ Flexibility of closure

11 คือ Perceptual speed

12 คือ Masking

13 คือ Tampo

ค่าในตารางต่อไปนี้ คือสหสัมพันธ์ของตัวแปรทั้ง 13 ตัว ดังกล่าวข้างต้นจงทำการวิเคราะห์ปัจจัย โดยใช้การหมุนแกนแบบ Varimax เพื่อหา main factors

ตัวแปร	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	.69	.54	.37	.46	.39	.37	.40	.53	.37	.62	.52	.29
2		1	.31	.28	.31	.21	.23	.20	.32	.36	.49	.29	.18
3			1	.43	.33	.42	.31	.43	.46	.38	.39	.39	.26
4				1	.30	.50	.52	.41	.41	.37	.37	.31	.27
5					1	.49	.33	.25	28	.33	.28	.36	.26
6						1	.60	.50	.36	.36	.48	.37	.19
7							1	.45	.33	.22	.37	.37	.29
8								1	.31	.36	.33	.28	.11
9									1	.55	.50	.28	.25
10										1	.38	.30	.21
11											1	.40	.27
12												1	.44
13													1