บทที่ 2

วิธีการวิเคราะห์ตัวประกอบ

การวิเคราะห์ตัวประกอบ เป็นสาขาหนึ่งของสถิติ-วิทยาศาสตร์ Spearman เป็นผู้ที่นำเอาเทคนิคการวิเคราะห์ตัวประกอบออกมาใช้ เป็นที่แพร่หลายในลักษณะที่เป็นคณิตศาสตร์ โดยนำมาใช้อธิบายทฤษฎีที่เกี่ยวกับความสามารถและพฤติกรรมของมนุษย์ และได้มีการพัฒนาวิธีการวิเคราะห์นี้เรื่อยมา ในปี ค.ศ. 1930 Thurstone ได้นำคำว่า "ตัวประกอบพหุคูณ" (Multiple Factor)มาใช้อย่างแพร่หลาย และได้เสนอหลัก"ความสอดคล้อง" (Principle of Parsimony) ในความหมายที่ว่า ตัวประกอบน่าจะสัมพันธ์กับตัวแปรในลักษณะที่ตัวประกอบแต่ละตัวมีตัวแปรเด่นชัดที่ไม่ซ้ำกันอยู่หลายตัว และตัวประกอบร่วม (Common Factor) น่าจะมีคุณสมบัติดังกล่าวนี้ ซึ่งได้ตั้งชื่อว่า "โครงสร้างอย่างง่าย" (Simple Structure) วิธีการนี้ได้ถูกนำไปใช้ในทุกสาขาและเป็นกระบวนการจัดข้อมูลที่ได้รับความนิยมมากขึ้นทั้งในสาขาวิทยาศาสตร์และสังคม ซึ่งปัจจุบันทิศทางของการวิเคราะห์ตัวประกอบมุ่งไปสู่การทดสอบสมมุติฐานเกี่ยวกับจำนวนตัวประกอบ โดยคาดว่าผลการวิเคราะห์ที่ได้รับจากกลุ่มตัวอย่างน่าจะสรุปอ้างอิงไปยังประชากรตัวประกอบได้ ผู้นำในเรื่องนี้คือ Darrel Bock, Bargmann และ Jöreskog

<u>สมการพื้นฐานของการวิเคราะห์ตัวประกอบ</u>

ในการวิเคราะห์ตัวประกอบนั้นมักจะเริ่มด้วยเมตริก สหสัมพันธ์ (correlation matrix)ระหว่างตัวแปรและลงท้ายด้วยเมตริก น้ำหนักตัวประกอบ (loading factor) ซึ่งแสดงถึงขอบเขตที่ ตัวแปรสัมพันธ์กับตัวประกอบที่สมมุติขึ้น นั่นคือ สหสัมพันธ์ระหว่างตัวแปรกับตัวประกอบนั่นเอง ดังนั้น จึงมีข้อตกลงเบื้องต้นว่า "คะแนนมาตรฐานของข้อมูลตัวแปรสามารถบรรยายให้เป็นผลบวกเชิงเส้นตรงของคะแนนตัวประกอบร่วม คะแนนตัวประกอบเฉพาะและคะแนนตัวประกอบคลาดเคลื่อน"

(1) $Z_{ik} = a_{i1}F_{1k} + a_{i2}F_{2k} + \dots a_{im}F_{mk} + a_{is}S_{ik} + a_{ie}F_{ik}$

Z_{ik} = คะแนนมาตรฐานสำหรับจังหวัดที่ k บนตัวแปรที่ i

a_{im} = น้ำหนักตัวประกอบสำหรับข้อมูลตัวแปรที่i บนตัวประกอบร่วม m

a_{is} = น้ำหนักตัวประกอบสำหรับตัวแปรที่ i บนตัวประกอบเฉพาะ s

a_{ic} = น้ำหนักตัวประกอบสำหรับตัวแปรที่ i บนตัวประกอบคลาดเคลื่อน e

F_{mk} = คะแนนมาตรฐานสำหรับจังหวัดที่ k บนตัวประกอบร่วม m

S_{ik} = คะแนนมาตรฐาน สำหรับจังหวัดที่ k บนตัวประกอบเฉพาะ i

E_{ik} = คะแนนมาตรฐาน สำหรับจังหวัดที่ k บนตัวประกอบคลาดเคลื่อนi

โดย Z, F, S, Fและ E เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ย (mean) = 0 และคะแนนเบี่ยงเบนมาตรฐาน (Standard error) = 1

จาก (1) อาจแทนทุกค่าของ i และ k ด้วย รูปเมตริก คือใช้ค่าของทุกจังหวัดและทุกตัวแปรพร้อมกับ ลักษณะสมการเมตริก เขียนได้ดังนี้

(2) $Z = AF$

สมการ (2) กล่าวว่า เมตริกของคะแนนตัวแปร Z อาจได้จากการคูณเมตริกน้ำหนัก ตัวประกอบ A ด้วยคะแนนเมตริก คะแนนตัวประกอบ F

โดยมีข้อสมมุติว่า

1. mean ของ 2=0 และ covariance = R

2. ตัวประกอบแต่ละตัวไม่มีความสัมพันธ์กัน นั่นคือ เมตริก F เป็น orthogonal matrix

<u>เทคนิคการวิเคราะห์ตัวประกอบ</u>

ในการวิเคราะห์ตัวประกอบมี 3 ขั้นตอนคือ

1. <u>การสะกัดตัวประกอบ</u> หมายถึง การใช้สัมประสิทธิ์ในเมตริกสหสัมพันธ์ (R) หาค่าคอลัมน์สัมประสิทธิ์ระหว่างตัวแปรกับโครงสร้างตัวประกอบที่สมมุติขึ้นแทนตัวแปร วิธีการก็คือ พยายาม "สะกัด" ตัวประกอบจากเมตริกจนกระทั่งไม่มีความแปรปรวนเหลืออยู่เลย สัมประสิทธิ์ที่ได้จากการสะกัดตัวประกอบ ก็คือ น้ำหนัก หรือตัวถ่วงของตัวแปรต่อตัวประกอบนั้น ซึ่งแสดงถึงขอบเขตที่ตัวแปรสัมพันธ์ตัวประกอบที่สมมุติขึ้น นั่นก็คือสหสัมพันธ์ระหว่างตัวแปรกับตัวประกอบนั้นเอง เช่น ตัวแปรตัวหนึ่งมีน้ำหนักตัวประกอบที่สะกัดได้ เป็น 0.7 บนตัวประกอบตัวหนึ่ง สหสัมพันธ์ระหว่างตัวแปรกับโครงสร้างตัวประกอบที่สมมุติขึ้นจะอยู่ประมาณ

๐.7 ตัวแปรตัวหนึ่งอาจมีน้ำหนักที่เป็นลบที่สูงมากก็ได้ แสดงว่า ตัวแปรนั้นมีสหสัมพันธ์เป็นลบกับโครงสร้างตัวประกอบนั้น

การสะกัดตัวประกอบมีอยู่หลายวิธี เช่น วิธีตัวประกอบสองตัว วิธีตัวประกอบคู่ และ วิธีตัวประกอบหลายตัว ในที่นี้จะกล่าวเฉพาะวิธีตัวประกอบหลายตัวซึ่งเป็นวิธีวิเคราะห์ตัวประกอบที่ให้ตัวประกอบทั่วไปหลายตัว โดยมีจุดมุ่งหมาย เพื่อลดความซ้ำซ้อนของข้อมูลให้เหลือจำนวนตัวประกอบร่วมไม่กี่ตัวที่สามารถอธิบายตัวแปรหรือข้อมูลทั้งหมดได้ โดยใช้เทคนิคแกน หรือตัวประกอบสำคัญ (The Principal Factor Technique) อันเป็นวิธีที่ใช้แพร่หลายกันมาก

หลักการของเทคนิคแกนหรือตัวประกอบสำคัญนี้ ได้จากทฤษฎีหนึ่งที่กล่าวว่า เมตริกสมมาตรสามารถลดลงเหลือเฉพาะเทอมในแนวทะแยงได้ คือ

$$BRB' = D$$

โดย R = เมตริกสมมาตรที่จะลดเหลือในแนวทะแยง (symmetric matrix)

B = othogonal matrix

B' = transpose ของ matrix B

D = เมตริกสมมาตรที่เหลือแต่ค่าในแนวทะแยงหรือ Diagonal matrix

ขั้นตอนในการสะกัดตัวประกอบสำคัญมีดังนี้

1. หาเมตริก R(Correlation matrix) โดยค่าตามแนวทะแยงให้ใช้ตัว R^2 เมตริก R นี้มีคุณสมบัติ Symmetric matrix

2. ทำ element ของ matrix R ที่มีค่ามากที่สุดให้เป็น ๐ โดยใช้ matrix B ซึ่งมีคุณสมบัติ orthogonal matrix

$$B = \begin{bmatrix} \cos\zeta_{ij} & -\sin\zeta_{ij} \\ \sin\zeta_{ij} & \cos\zeta_{ij} \end{bmatrix}$$

ζ_{ij} นี้หาได้จาก

$$\tan 2\zeta_{ij} = -2r_{ij}/h_i^2 - h_j^2$$

3. matrix B ที่หาได้จากข้อ ๒ นั้นเอาไปคูณกับ R จะได้ D_1

$$B_1RB_1' = D_1$$

4. แล้วทำ element ของ D_1 ที่มีค่ามากที่สุดเป็น 0 โดยการหา matrix B ตามวิธีในข้อ 2 แล้วเอามาคูณ matrix D_1

5. ทำเช่นนี้ไปเรื่อย ๆ จนกระทั่งค่า element ของ D ที่อยู่นอกแนวทะแยงมีค่าน้อยที่สุด

6. นำ Matrix B ทั้งหมดข้างต้นมาคูณกัน

$$B = B_k B_{k-1} \cdots B_1$$

7. สามารถหา A ได้จาก

$$A = B' \sqrt{D}$$

ซึ่ง matrix D ที่ใช้นั้น เป็น matrix ที่ได้ในข้อ 5 และจาก matrixนี้จะทำให้ทราบจำนวนตัวประกอบ (Factor) ด้วยเพราะ จำนวนตัวประกอบต้องเท่ากับจำนวน eigen ที่มีค่าเป็นบวก

<u>ตัวอย่าง</u> 1. เมื่อเราหา เมตริกสหสัมพันธ์ของตัวแปร 4 ตัว ได้ดังตารางข้างล่าง

ตารางที่ 2.1 เมตริกสหสัมพันธ์ของตัวแปร 4 ตัว

ตัวแปร	1	2	3	4
1	(.65)	.50	.16	.17
2	.50	(.40)	.22	.24
3	.16	.22	(.65)	.73
4	.17	.24	.73	(.82)

2. element ที่มีค่ามากที่สุดคือ .73 ซึ่งเป็น correlation ระหว่างตัวแปรที่3 กับ ตัวแปรที่ 4 เราก็ทำการหา matrix B ที่เทอมในแนวทะแยง (diagonal)=1 ส่วนค่าที่เทอม b_{33} และ $b_{44} = \cos\zeta_{34}$ $b_{34} = \sin\zeta_{34}$ และ $b_{43} = -\sin\zeta_{34}$ ส่วนค่า element อื่นใน matrix B= 0 ซึ่ง ζ นั้นก็สามารถหาได้จากสูตร

$\tan 2\xi_{ij} = -2\gamma_{ij}/h_i^2 - h_j^2$ matrix B ที่ได้นี้คือ B_1

3. เอา matrix B_1 นี้คูณกับ matrix R ในข้อ 1 ก็จะได้ D_1 ดังตารางที่ 2.2 แล้วทำไปเรื่อยจนได้ D น้อยสุดคือ

ตารางที่ 2.2 คำตอบตัวประกอบสำคัญของปัญหาตัวแปร 4 ตัว

B_i				D_i			
1	2	3	4	1	2	3	4
1	0	0	0	.6500	.5000	.0065	.2334
0	1	0	0	.5000	.4000	.0047	.3255
0	0	.7469	-.6650	.0065	.0047	.0001	.0000
0	0	.6650	.7469	.2334	.3255	.0000	1.4699
.7882	.6154	0	0	1.0404	0	.0080	.3843
-.6154	.7882	0	0	0	.0096	-.0002	.1130
0	0	1	0	.0080	-.0002	.0001	0
0	0	0	1	.3843	.1130	0	1.4699
.8625	0	0	-.5060	.8149	-.0572	.0069	0
0	1	0	0	-.0572	.0096	-.0002	.0974
0	0	1	0	.0069	-.0002	.0001	.0040
.5060	0	0	.8625	0	.0974	.0040	1.6954
1	0	0	0	.8149	-.0571	.0069	-.0033
0	.9983	0	-.0575	-.0571	.0040	-.0005	0
0	0	1	0	.0069	-.0005	.0001	.0040
0	.0575	0	.9983	-.0033	0	.0040	1.7010
.9976	-.0699	0	0	.8189	0	.0069	-.0033
.0699	.9976	0	0	0	0	0	-.0002
0	0	1	0	.0069	0	.0001	.0040
0	0	0	1	-.0033	-.0002	.0040	1.7010
1	0	.0084	0	.8190	0	0	-.0032
0	1	0	0	0	0	0	-.0002
-.0084	0	1	0	0	0	0	.0040
0	0	0	1	-.0032	-.0002	.0040	1.7010
1	0	0	0	.8190	0	0	-.0032
0	1	0	0	0	0	0	-.0002
0	0	1	-.0023	0	0	0	.0002
0	0	.0023	1	-.0032	-.0002	.0002	1.7010
1	0	0	.0036	.8190	0	0	0
0	1	0	0	0	0	0	-.0002
0	0	1	0	0	0	0	.0002
-.0036	0	0	1	0	-.0002	.0002	1.7010
.7240	.4771	-.3250	-.3777	1 .655	0	0	.470
-.5883	.8042	-.0564	-.0634	2 .432	0	0	.462
-.0069	-.0048	.7484	-.6632	2-.294	0	0	.751
.3601	.3545	.5755	.6430	4-.342	0	0	.839

$$D_8 = \begin{bmatrix} .8190 & 0 & 0 & 0 \\ 0 & 0 & 0 & -.0002 \\ 0 & 0 & 0 & .0002 \\ 0 & -.0002 & .0002 & 1.7010 \end{bmatrix}$$

4. นำเอา matrix B ทั้ง 8 มาคูณกับ ก็จะได้ matrix B คือ rowสุดท้าย

$$B = B_8 \ldots B_3 B_2 B_1$$

5. เอา B ที่ได้นั้นไปคูณกับ $\sqrt{D_8}$ ก็จะได้ matrix A

$$A = \begin{bmatrix} .470 & .655 \\ .462 & .432 \\ .751 & -.294 \\ .839 & -.342 \end{bmatrix}$$

2. <u>การหมุนแกน</u> ตัวประกอบที่ยังไม่ได้หมุนแกน มักจะให้ประโยชน์ต่องานวิทยาศาสตร์น้อยมาก เพราะวิธีการสกัดตัวประกอบ เป็นวิธีที่มุ่งสะกัดความแปรปรวนสำหรับตัวประกอบแต่ละตัวให้ได้มากที่สุด เท่าที่จะเป็นได้ ดังจะเห็นได้จากขนาดของตัวประกอบจะลดลง เรื่อย ๆ ตัวประกอบที่ยังไม่ได้หมุนแกน ซึ่งได้รับจากการสะกัดตัวประกอบที่มุ่งความแปรปรวนสูงสุดมักจะโครงสร้างตัวประกอบที่ซับซ้อนมาก เพราะไป เกี่ยวข้องกับตัวแปรจำนวนมาก ดังตารางที่ 2.3 กล่าวคือ ตัวแปรทั้ง 4 นี้มี เพียงตัวแปรที่ 3 ตัว เดียว เท่านั้นที่ เกี่ยวข้องกับ

ตารางที่ 2.3 เมตริกตัวประกอบ

ตัวแปร	ตัวประกอบ	
	I	II
1	.470	.655
2	.462	.432
3	.751	-.292
4	.839	-.342

ตัวประกอบที่ 1 เพียงตัวเดียว ตัวประกอบที่ยุ่งยากซับซ้อนนี้ยากที่จะแปลความหมายและนำไปใช้บรรยายลักษณะทางวิทยาศาสตร์ได้ ดังนั้นจึงได้ทำการหมุนแกน ซึ่งจะได้เมตริกตัวประกอบที่เหมือนเมตริกที่ยังไม่ได้หมุนแกนทุกประการในเชิงคณิตศาสตร์ แต่จะให้ความสัมพันธ์อีกแบบหนึ่ง

ในที่นี้จะกล่าวเฉพาะการหมุนแกนแบบ Varimax ของ Kaiser เท่านั้นเพราะเป็นวิธีที่มีชื่อเสียง และให้ประโยชน์มาก วิธีหมุนแกนนี้ตั้งอยู่บนแนวความคิดที่ว่าตัวประกอบที่แปลความหมายได้ดีที่สุด จะต้องมีน้ำหนักไม่สูงก็ต่ำ และมีน้ำหนักขนาดปานกลางเพียง 2-3 ตัว ซึ่งความแปรปรวนของน้ำหนักตัวประกอบ เขียนในรูปสมการได้ดังนี้

$$\sigma_j^2 = 1/n \sum_{i=1}^{n} (a_{ij})^2 - 1/n^2 \left(\sum_{i=1}^{n} a_{ij}^2\right)^2$$

ความแปรปรวนควรจะมีค่าสูงสำหรับตัวประกอบทั้งหมด ดังนั้นคำตอบ othogonal จึงค้นหา V เป็นค่าสูงสุด โดยตั้งต้นว่า

$$V = \sum_{j=1}^{m} \sigma_j^2$$

ในการหมุนแกนนั้นจะทำการหมุนแกนตัวประกอบทีละคู่โดยพยายามให้ได้ค่า $\sigma_i^2 + \sigma_j^2$ สำหรับตัวประกอบที่หมุนแกนแล้วมีค่ามากที่สุด เมตริกที่นำมาเปลี่ยนรูป เพื่อให้ได้ผลดังกล่าวคือ

$$\Lambda = \begin{bmatrix} \cos\zeta & -\sin\zeta \\ \sin\zeta & \cos\zeta \end{bmatrix}$$ เมื่อเศษ + ส่วน +

$$\Lambda = \begin{bmatrix} \cos\zeta & \sin\zeta \\ -\sin\zeta & \cos\zeta \end{bmatrix}$$ เมื่อเศษ - ส่วน +

$$\Lambda = \begin{bmatrix} \cos(45^\circ-\zeta) & -\sin(45^\circ-\zeta) \\ \sin(45^\circ-\zeta) & \cos(45^\circ-\zeta) \end{bmatrix}$$ เมื่อเศษ + ส่วน -

$$\Lambda = \begin{bmatrix} \cos(45^\circ-\zeta) & \sin(45^\circ-\zeta) \\ -\sin(45^\circ-\zeta) & \cos(45^\circ-\zeta) \end{bmatrix}$$ เมื่อเศษ - ส่วน -

ซึ่ง ζ นี้มีคุณสมบัติ เป็น orthogonal matrix หาได้จาก

$$\tan 4\zeta = \frac{2\{n\Sigma(x^2-y^2)\ 2xy - \Sigma(x^2-y^2)\Sigma(2xy)}{n\{\Sigma|(x^2-y^2)^2 - (2xy)^2|\}-\{|\Sigma(x^2-y^2)|^2-|\Sigma(2xy)^2|\}}$$

ตัวอย่าง

matrix ที่ได้จากการสะกัดตัวประกอบ (ตารางที่ 2.3)

ตัวแปร	ตัวประกอบ	
	I	II
1	.470	.655
2	.462	.432
3	.751	-.292
4	.839	-.342

เราจะทำการหมุนแกนตัวประกอบดังกล่าวด้วยวิธี Varimax ของKaiser ดังนี้

x	y	x^2-y^2	$2xy$	$(x^2-y^2)^2$	$(2xy)^2$	$(x^2-y^2)2xy$	$(x^2-y^2)-(2xy)^2$
.470	.655	-.2081	.6157	.0433	.3791	-.1281	-.5872
.462	.432	.0268	.3992	.0007	.1594	.0107	-.1326
.751	-.292	.4787	-.0962	.2292	.0093	-.0461	.4694
.839	-.342	.5869	.5739	.3445	.3294	-.3368	.2575
รวม		1.1418	.3448	.6177	.8772	-.5003	.0071

$$\tan 4\zeta = \frac{2|4(-.5003)-(1.1418)(.3448)|}{4|\ (.0071)-\{(1.1418)^2-(.3448)^2\}|}$$

$$= \frac{2|(-2.0012)-(.3937)|}{4|.0071 - (1.3037-.1189)|}$$

$$= \frac{-4.7898}{-4.7108} = 1.0168$$

ค่าสมบูรณ์ของ $\tan 4\zeta$ คือ 1.0168 ซึ่งเราก็จะสามารถหามุมที่จะใช้หมุนได้ แต่เนื่องจากเศษ และส่วนเป็น - เมตริก เปลี่ยนรูปจึงเป็น

$$\Lambda = \begin{bmatrix} \cos(45^{\circ}-\zeta) & \sin(45^{\circ}-\zeta) \\ -\sin(45^{\circ}-\zeta) & \cos(45^{\circ}-\zeta) \end{bmatrix}$$

เมื่อ Λ ที่ได้นี้ไปคูณ matrix ตัวประกอบข้างต้นก็จะได้เมตริกตัวประกอบใหม่ที่หมุนแกนแล้ว

๓. <u>การแปลผล</u> เมื่อทำการหมุนแกนเรียบร้อยแล้ว ก็จะพิจารณาว่าตัวแปรต่าง ๆ ที่มีค่าตัวถ่วงสูงบนตัวประกอบใด ถ้าตัวแปรข้อมูลสัมพันธ์อย่างสมบูรณ์กับตัวประกอบก็จะได้รับการพิจารณาว่า เป็นอันเดียวกับตัวประกอบที่ตัวแปรเหล่านี้วัด เนื่องจากตัวแปรข้อมูลไม่มีความเที่ยงอย่างสมบูรณ์ ก็ไม่สามารถสัมพันธ์อย่างสมบูรณ์กับตัวประกอบได้ แต่ตัวแปรข้อมูลที่มีความเที่ยง ๐.๘๑ และตัวถ่วงตัวประกอบ .๙๐ แสดงความซับซ้อนทั้งหมดในความแปรปรวนที่แท้จริง ระหว่างตัวแปรข้อมูลและตัวประกอบ โดยทั่วไปตัวถ่วงตัวประกอบสำหรับตัวแปรขอ้มูลที่เท่ากับกรณฑ์สองของสัมประสิทธิ์ของสหสัมพันธ์ของมันแสดงว่าเป็นตัวประกอบบริสุทธิ์ หรือชี้ตัวแปรข้อมูลที่บริสุทธิ์ ความแปรปรวนที่แท้จริงทั้งหมดของมันจะอยู่บนตัวประกอบนั้น ซึ่งทำให้เหมือนตัวประกอบทุกประการ ยกเว้นความแปรปรวนของความคลาดเคลื่อน

ในทางปฏิบัติ สหสัมพันธ์ของตัวแปรข้อมูลกับตัวประกอบที่มีค่าต่ำกว่ากรณฑ์สองของความเที่ยงของตัวแปรข้อมูล แสดงว่าตัวแปรนั้นมีความแปรปรวนที่แท้จริงแต่ไม่ได้ร่วมกับตัวประกอบนี้ ความแปรปรวนที่แท้จริงที่เพิ่มเข้ามาอาจเนื่องมาจากความแปรปรวนจำเพาะ หรือเนื่องมากจากผลรวมของความแปรปรวนจำเพาะกับความแปรปรวนร่วมบางส่วนที่เกี่ยวข้องกับตัวประกอบอื่น

โดยปกติ สหสัมพันธ์ของตัวแปรข้อมูลกับตัวประกอบที่ได้กำหนดอย่างชัดเจนว่าอยู่ใต้ระดับที่ยอมให้มีการอ้างอิงได้อย่างง่ายระหว่างตัวประกอบกับตัวแปรตัวใดตัวหนึ่ง อาจเป็นการวิเคราะห์ที่ใช้เวลาที่ให้ได้ค่าตัวถ่วงสูงในตัวแปรให้เข้าไปสู่ส่วนใดส่วนหนึ่ง เพื่อว่าส่วนที่ร่วมกันกับตัวแปรอื่นทั้งหมดที่ใช้หาตัวประกอบจะได้รับการแสดงตัวและตั้งชื่อ ขบวนการพยายามที่จะแสดงธรรมชาติของตัวประกอบ ทำได้หลายวิธี

๑. ตัวถ่วงค่าสูงมากเท่าใด ดีกรีของการซับซ้อนของความแปรปรวนที่แท้จริงระหว่างตัวแปรข้อมูลกับตัวประกอบก็ยิ่งสูงขึ้น และตัวประกอบก็ยิ่งเหมือนตัวแปรมากขึ้น กำลังสองของความสัมพันธ์ระหว่างตัวแปรกับตัวประกอบนี้เป็นการชี้ที่ดีให้เห็นขอบเขตของความซับซ้อน การหาค่ามีด้วยสัมประสิทธิ์สหสัมพันธ์ สำหรับตัวแปรข้อมูลให้สัดส่วนของความแปรปรวนที่แท้จริงของตัวแปรข้อมูลที่ร่วมกับตัวประกอบ ในบางครั้งเรื่องเหล่านี้ช่วยมากในการหาค่าสหสัมพันธ์ของตัวแปรข้อมูลกับตัวประกอบด้วยกรณฑ์สองของสัมประสิทธิ์สหสัมพันธ์ เพื่อให้ได้ค่าของสหสัมพันธ์

ระหว่างสอง เรื่องนี้ ถ้าตัวแปรข้อมูลมีความ เที่ยงสมบูรณ์

2. ยิ่งตัวแปรที่ใช้สามารถชี้ตัวประกอบได้บริสุทธิ์มาก เท่าใด ก็ยิ่งง่ายที่จะทำการอ้างอิง เกี่ยวกับธรรมชาติของตัวแปรมาก เพียงนั้น ถ้าตัวแปร เป็นแบบที่ยุ่งยากซับซ้อนที่มีค่าตัวถ่วงสูงปานกลางบนตัวประกอบที่กำหนดก็ เป็นการยากที่จะตัดสินจากความรู้ เหล่านี้อย่าง เดียวว่าส่วนใดที่ควรรับผิดชอบตัวแปรข้อมูลของตัวประกอบที่ยุ่งยากซับซ้อนมากนี้ให้ประโยชน์ในขั้นการแปลตัวประกอบ เพียง เล็กน้อยกว่าในขั้นหมุนแกน

3. ยิ่งมีตัวแปรที่มีค่าตัวสูงบนตัวประกอบจำนวนมากขึ้นและถ้าสิ่งอื่น เท่ากันหมดก็ยิ่งง่ายที่จะแยกว่าตัวประกอบอะไรน่าจะแสดงตัวบ้าง คง เป็นไปได้ที่จะได้รับความประทับใจแบบผิด ๆ ในลักษณะตัวประกอบ ทั้ง ๆ ที่ตัวแปรมีค่าตัวถ่วงสูงกว่ากันอยู่ก็ตาม กลุ่มตัวแปรนี้ลำ เอียงไปในทางตัวประกอบมาก ถ้าตัวแปรข้อมูลที่รวมนี้พลาดที่จะ เป็นตัวแทนส่วนใดส่วนหนึ่งของตัวประกอบที่สำคัญ เช่นแสดงภาพที่ให้ว่าตัวประกอบตัวใดควร เป็นอย่างไร สถานการณ์เช่นนี้ เปรียบ เทียบได้ว่า การแสดงลักษณะหน้าตาที่สำคัญ ๆแสดงออกได้ แต่ไม่อาจบอกลักษณะทั้งหมดได้

สหสัมพันธ์ระหว่างตัวประกอบกับตัวแปรที่จะสรุปได้ว่าสามารถแปลความหมายตามจุดประสงค์ได้ว่า "สำคัญ"นั้นระดับตัดสินที่นิยมใช้กันมาก สำหรับตัวถ่วงประกอบคือ .30 นั่นคือไม่มีตัวแปรใดกับตัวถ่วงประกอบที่มีค่าต่ำกว่า .30 ในรายการตัวแปรที่ใช้หาตัวประกอบ กำลังสองของค่ามี $(.30)^2$ ให้ค่า .09 ซึ่งชี้ให้ เห็นว่าตัวแปรข้อมูลสัมพันธ์กับตัวประกอบน้อยกว่า .30 นั่น มีค่าน้อยกว่า ร้อยละ 10 ของความแปรปรวนที่ร่วมกับตัวประกอบ อีกร้อยละ 90 เป็นของสิ่งอื่น เช่น อยู่ในตัวประกอบ เฉพาะ ในกรณีออโธกอนอลซึ่งตัวถ่วง เช่นนี้สามารถแปลความหมาย เป็นค่าสหสัมพันธ์ได้ ค่า เช่นนี้ดู เหมือนมีค่าต่ำมาก นั่นคือ สัดส่วนของตัวแปรจากความแปรปรวนร้อยละ 90 ไม่ได้ เกี่ยวข้องกับตัวประกอบ จะ เป็นความผิดพลาดอย่างมากที่ให้ตัวประกอบนี้ ดังนั้นตัวถ่วง เช่นนี้ไม่สามารถจะ เป็นพื้นฐานสำหรับการแปลผลตัวประกอบได้ ซึ่งตารางข้างล่างนี้จะให้ความคิดอย่างคร่าว ๆ เกี่ยวกับค่าสหสัมพันธ์ของตัวแปรตัวประกอบ (ตัวประกอบแบบออโธกอนอล)

ตารางที่ 2.4สเกลของสหสัมพันธ์ตัวแปร-ตัวประกอบ

ตัวถ่วงประกอบแบบออโธกอนอล	เปอร์เซนต์ความแปรปรวน	การลำดับ
.71	50	ดีเลิศ
.63	40	ดีมาก
.55	30	ดี
.45	20	ปานกลาง
.32	10	เลว

การแปลผลตัวประกอบ ถ้าตัวแปรข้อมูลหลายตัว มีตัวอย่างที่จัดว่า "ดีมาก" ถึง "ดีเลิศ" ก็จะทำให้สามารถตัดสินใจได้ว่า ควรพูดอะไรเกี่ยวกับตัวประกอบถึงแม้ว่าผู้วิจัยจะต้องหลีกเลี่ยงประโยคที่จำเพาะเจาะจง นอกจากนี้การแปลผลดังกล่าวควรเลยไปถึงการกำหนดตัวแปรใดว่าเหมือนตัวประกอบและตัวใดไม่เหมือนโดยดูจากค่าตัวถ่วง ถ้าตัวแปรใดมีค่าตัวสูงก็จะได้รับการพิจารณาว่าเหมือนตัวประกอบ และผลที่ได้แสดงและตั้งชื่อสิ่งที่ปรากฏว่าเป็นสิ่งร่วมหรือสิ่งแทนโดยตัวประกอบวิธีแปลความหมายตัวประกอบที่ดีที่สุดจะทำได้โดยคำนึงถึงเรื่องอื่น ๆ ร่วมด้วย

4. คะแนนตัวประกอบ เมื่อตัวประกอบได้ถูกแยกออกแล้ว นักวิจัยมักกระตือรือล้นที่จะหาความเกี่ยวข้องกับตัวแปรนี้กับตัวแปรอื่น โดยปกติแล้วมักจะได้ตัวแปรหลายตัวที่สัมพันธ์กับตัวประกอบ ถ้าไม่มีตัวแปรใดที่มีความแปรปรวนเป็นร้อยละ 80 ของตัวประกอบ ก็จะต้องหาหรือประมาณคะแนนตัวประกอบโดยการใช้คะแนนของตัวแปรเหล่านี้ที่สัมพันธ์กับตัวประกอบในทางใดทางหนึ่ง ซึ่งวิธีการประมาณคะแนนตัวประกอบมีหลายวิธี ในที่นี้จะพูดเฉพาะวิธีการประมาณคะแนนโดยวิธีความถดถอยพหุคูณ โดยมีสมการพื้นฐานต่อไปนี้

$$Z_{fi} = B_1 Z_{1i} + B_2 Z_{2i} + \dots B_n Z_{ni}$$

Z_{fi} = คะแนนมาตรฐานใน Factor f สำหรับ จังหวัด i

Z_{1i} = คะแนนมาตรฐานใน ตัวแปร 1 สำหรับ จังหวัด i

Z_{2i} = คะแนนมาตรฐานใน ตัวแปร 2 สำหรับ จังหวัด i

Z_{ni} = คะแนนมาตรฐานใน ตัวแปร n สำหรับ จังหวัด i

B_i = สัมประสิทธิ์ความถดถอยมาตรฐานสำหรับตัวแปร i

<u>จุดมุ่งหมายของการวิเคราะห์ตัวประกอบ</u>

ในการวิเคราะห์ตัวประกอบมีหลักใหญ่ ๆ ที่ต้องการคำตอบอยู่ 2 ลักษณะคือ "มีความง่ายเชิงสถิติและมีความหมายในเนื้อหา" ซึ่งสามารถจำแนกจุดมุ่งหมายในการวิเคราะห์ตัวประกอบ ได้ดังนี้คือ

1. ช่วยให้ได้การบรรยายเกี่ยวกับปริเขต (Domain)ที่ต้องการศึกษา

2. ช่วยตรวจสอบทฤษฎีที่เกี่ยวกับความสหสัมพันธ์ระหว่างตัวแปร

3. ช่วยสร้างความสัมพันธ์เชิงหน้าที่ (Functional Relations)ระหว่างตัวแปร

4. วิเคราะห์บุคคลหรือวัตถุและจัดให้เป็นประเภทต่าง ๆ

5. วิเคราะห์โครงสร้างเชิงตัวประกอบ (Factorial Structure)ของตัวแปรที่เป็นเกณฑ์ และช่วยบ่งชี้ตัวแปรที่จะเป็นประโยชน์ในสมการถดถอยได้

6. เป็นการพิสูจน์ข้อค้นพบของเขากับของคนอื่นโดยใช้ข้อมูลของกลุ่มตัวอย่างใหม่จากประชากรกลุ่มเดียวกัน

7. เป็นการลดจำนวนข้อมูลให้น้อยลงเพื่อให้ได้ลักษณะที่ร่วมกันที่ซ่อนอยู่

8. ในการทดสอบเป็นการหาความตรงเชิงโครงสร้าง (Construct Validity) ของแบบวัด

9. ช่วยในการสร้างแบบวัดลักษณะต่าง ๆ

<u>ประโยชน์ของวิธีวิเคราะห์ตัวประกอบ</u>

ในการวิเคราะห์ตัวประกอบนี้ได้ให้ประโยชน์ในแง่ต่าง ๆ คือ ช่วยแยกข้อมูลที่รวบรวมได้ออกเป็นกลุ่ม ๆ ตามความสัมพันธ์ ทำให้ได้แบบแผนของความสัมพันธ์ ซึ่งเป็นการช่วยลดความซ้ำซ้อนของข้อมูลลงเหลือเพียงไม่กี่มิติที่สามารถใช้อธิบายข้อมูลเหล่านี้ได้ เช่น ข้อมูลตัวแปร 20 ตัวของจังหวัด 72 จังหวัด เมื่อใช้วิธีวิเคราะห์ตัวประกอบจะได้มิติเพียง 5 มิติเท่านั้น นอกจากนี้วิธีวิเคราะห์ตัวประกอบยังสามารถใช้ ตรวจสอบสมมุติฐานได้ด้วยว่า สมมุติฐานที่ตั้งไว้นั้นเป็นจริงหรือไม่ และมีความสัมพันธ์กันอย่างไร

<u>การหาค่า correlation ระหว่างตัวประกอบกับผลิตภาพของที่ดิน</u>

เมื่อได้ตัวประกอบที่ได้จากการวิเคราะห์ตัวประกอบแล้วก็ทำการหา correlation ระหว่างตัวประกอบกับผลิตภาพของที่ดิน ได้ดังนี้

1. เมื่อเราทราบว่าตัวประกอบแต่ละตัวมีตัวแปรใดที่เกี่ยวข้องแล้วก็ดู factor score coefficient ระหว่างตัวแปรที่เกี่ยวข้องกับตัวประกอบมีค่าเท่าใด เช่น ในกรณีภาคเหนือ เราทราบว่าตัวประกอบที่ 3 มีตัวแปรที่เกี่ยวข้อง 2 ตัว คือ ความเป็นเมือง และแรงงานในภาคอุตสาหกรรม ก็ดูค่า factor score coefficient ระหว่างตัวประกอบที่ 2 กับความเป็นเมืองและแรงงานในภาคอุตสาหกรรม ซึ่งจะได้ว่า factor score coefficient ของตัวประกอบที่ 2 กับความเป็นเมือง .54458 และแรงงานในภาคเกษตร .41369

2. นำ factor score coefficient ที่ได้คูณกับมูลค่าของตัวแปรที่เกี่ยวข้องกับตัวประกอบนั้นทุกตัว แล้วหาผลรวมของค่าต่าง ๆ นี้ ดังในตาราง

3. คำนวณหาค่า correlation ระหว่างผลิตภาพของที่ดินกับตัวประกอบนั้น โดยใช้วิธีการหาสัมประสิทธิ์สหสัมพันธ์แบบเปียรสัน โปรดัก-โมเมนท์ (Pearson Product Moment Correlation Coefficient) จากสูตร

$$r_{XY} = \frac{N\Sigma XY - \Sigma X\Sigma}{\sqrt{N\Sigma X^2-(\Sigma X)^2 \quad N\Sigma^2 Y-(\Sigma)^2}}$$

เช่น

$$r_{F_3V_2} = \frac{17(377.171) - 169.991(33.3)}{\sqrt{17(1829.345)-28,896.94 \quad 499.14}}$$

$$= .7166$$

ตาราง ค่า

จังหวัด	$.54458V_{15}$	$.41369V_{17}$	V	V_2	VV_2	V_1^2	V_2^2
กำแพงเพชร	1.95504	4.79880	6.754	1.0	6.754	45.617	1.0
เชียงราย	2.38526	4.38511	13.524	1.5	20.286	182.899	2.25
เชียงใหม่	5.10816	10.34225	15.450	4.6	71.07	238.703	21.16
ตาก	7.66769	7.36368	15.031	1.7	25.553	225.931	2.89
นครสวรรค์	4.02989	7.52916	11.559	0.9	10.403	133.610	0.81
น่าน	3.15312	3.55773	6.711	4.3	28.857	45.038	18.49
พะเยา	2.87538	3.97142	6.847	1.1	7.532	46.881	1.21
แพร่	2.45606	6.82589	9.282	2.2	20.42	86.156	4.84
พิจิตร	4.40565	6.28809	10.694	0.9	9.625	114.362	0.81
พิษณุโลก	5.77255	6.5363	12.039	0.8	9.847	151.511	0.64
เพชรบูรณ์	2.5323	4.17827	6.711	0.9	6.04	45.038	0.81
แม่ฮ่องสอน	2.23278	5.46071	7.693	4.3	33.08	59.182	18.49
ลำปาง	3.62690	7.32231	10.949	2.3	25.183	119.881	5.29
ลำพูน	2.04218	7.32231	9.364	3.5	32.774	87.684	12.25

ตาราง ค่า (ต่อ)

จังหวัด	$.54458V_{15}$	$.41369V_{17}$	V	V_2	VV_2	V_1^2	V_2^2
สุโขทัย	3.26748	4.88154	8.149	1.2	9.779	66.406	1.44
อุตรดิตถ์	4.62348	5.41934	10.043	1.1	11.047	100.862	1.21
อุทัยธานี	3.70859	5.21249	8.921	1.0	8.921	79.584	1.0
Σ			169.991	33.3	337.171	1,829.345	94.59
$(\Sigma)^2$			28,896.94	1,108.89			