

การวิเคราะห์องค์ประกอบ เชิงสำรวจ (1)

การประเมินลักษณะมนุษย์

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- โมเดลการวิเคราะห์องค์ประกอบ
- ทฤษฎีการทดสอบแบบคลาสสิก
- รูปแบบการวิเคราะห์องค์ประกอบ
- การสกัดองค์ประกอบ
- การวิเคราะห์ส่วนประกอบ
- การเลือกจำนวนองค์ประกอบ

แนะนำ

- สมมติว่าคุณอยากรู้ว่าคุณหน้าตาดีหรือไม่ คุณจะสามารถรู้ได้อย่างไร
- คุณอาจจะถามเพื่อน 8 คน ให้เพื่อนให้คะแนนระหว่าง 1-10 จากหน้าตาแย่มากๆ ไปยังหน้าตาดีสุดๆ
- หลังจากนั้นคุณก็นำข้อมูลของเพื่อน 8 คน มาประเมินว่าหน้าตาของคุณจะดีขนาดไหน
- คุณอาจให้นำน้ำหนักของคะแนนจากเพื่อนทุกคนเท่ากัน แล้วใช้คะแนนรวมบอกระดับหน้าตาของคุณ
- คุณอาจเชื่อใจเพื่อนบางคนมากกว่า แล้วให้นำน้ำหนักคนดังกล่าวมากกว่า

แนะนำ

- หน้าตาของคุณ เป็นสิ่งที่เรียกว่าองค์ประกอบ (Factor) หรือตัวแปรแฝง (Latent variable) ที่เป็นสิ่งที่คุณสนใจ แต่ไม่สามารถวัดได้โดยตรง
- ข้อมูลจากเพื่อนทั้ง 8 คน เป็นสิ่งที่คุณวัดได้ (Manifest variables or Observed variables) และคุณเชื่อว่าสิ่งนี้เป็นการวัดหน้าตา (องค์ประกอบ) ทางอ้อม
- การวิเคราะห์องค์ประกอบ (Factor analysis) เป็นสถิติที่บอกว่าข้อมูลจากเพื่อนทั้ง 8 คนนั้น สอดคล้องกันหรือไม่ ถ้าไม่สอดคล้องกันแล้ว จะแบ่งออกได้เป็นกี่รูปแบบ โดยวิเคราะห์ผ่านเมทริกซ์สหสัมพันธ์

แนะนำ

- สมมติว่า คุณให้เพื่อนทั้ง 8 คน ประเมินหน้าตาของคน 200 คน ได้เมทริกซ์สหสัมพันธ์ดังต่อไปนี้

หากสหสัมพันธ์เป็นดังต่อไปนี้

	1	2	3	4	5	6	7	8
1	1							
2	.7	1						
3	.7	.7	1					
4	.7	.7	.7	1				
5	.7	.7	.7	.7	1			
6	.7	.7	.7	.7	.7	1		
7	.7	.7	.7	.7	.7	.7	1	
8	.7	.7	.7	.7	.7	.7	.7	1

เพื่อนของคุณทุกคนให้คะแนนสอดคล้องกันดี

เนื่องจากสหสัมพันธ์ไม่ใช่ 1 แสดงว่าแต่ละคนยังมีลักษณะการมองหน้าตาที่ไม่เหมือนกับคนอื่น

หากสหสัมพันธ์เป็นดังต่อไปนี้

	1	2	3	4	5	6	7	8
1	1							
2	.8	1						
3	.8	.8	1					
4	.8	.8	.8	1				
5	.2	.2	.2	.2	1			
6	.2	.2	.2	.2	.8	1		
7	.2	.2	.2	.2	.8	.8	1	
8	.2	.2	.2	.2	.8	.8	.8	1

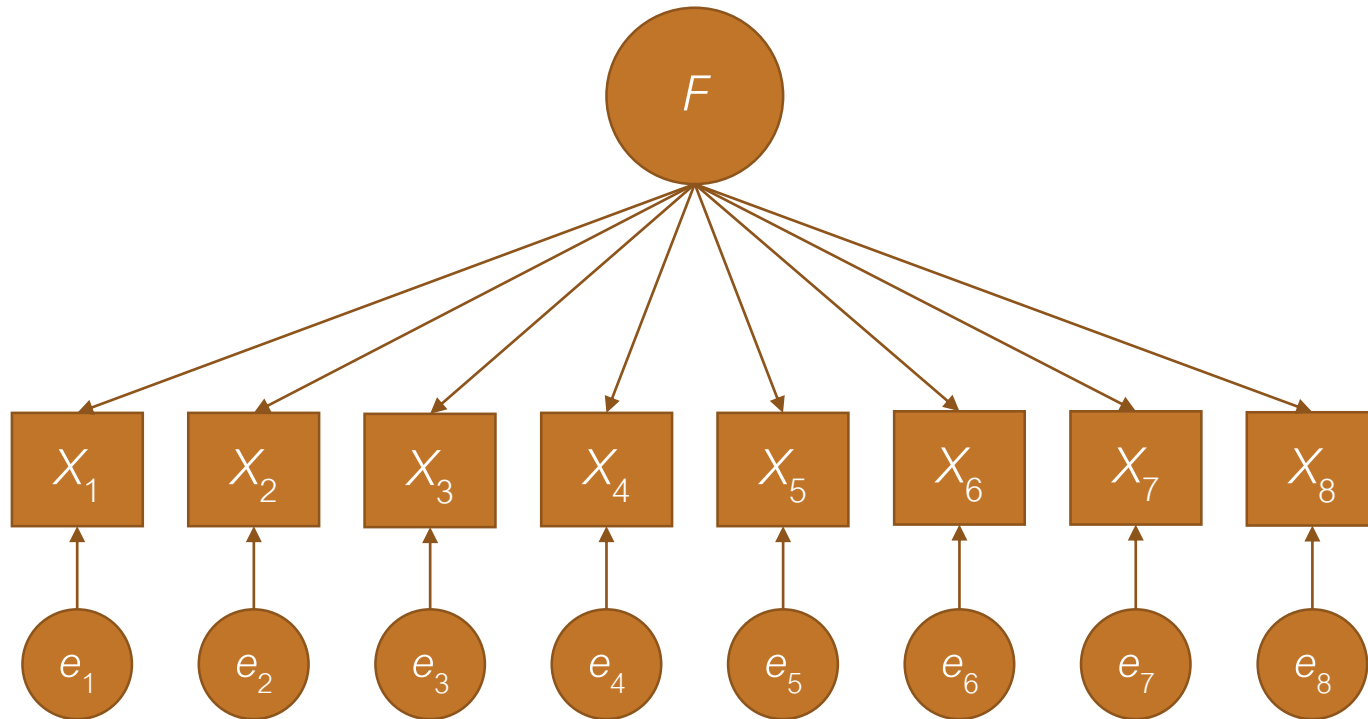
เพื่อน 4 คนแรก มองความหน้าตาดีในรูปแบบเดียวกัน และเพื่อน 4 คนหลังมองความหน้าตาดีในรูปแบบเดียวกัน เช่น เพื่อนกลุ่มแรกมองความหน้าตาดีแบบฝรั่ง ส่วนเพื่อนกลุ่มที่สองมองหน้าตาดีแบบเอเชีย

หากสหสัมพันธ์เป็นดังต่อไปนี้

	1	2	3	4	5	6	7	8
1	1							
2	0	1						
3	0	0	1					
4	0	0	0	1				
5	0	0	0	0	1			
6	0	0	0	0	0	1		
7	0	0	0	0	0	0	1	
8	0	0	0	0	0	0	0	1

ดูเหมือนว่าทั้ง 10 คน มองความหน้าตาดีไม่สอดคล้องกันเลย เพราะสหสัมพันธ์เท่ากับ 0 ทั้งหมด

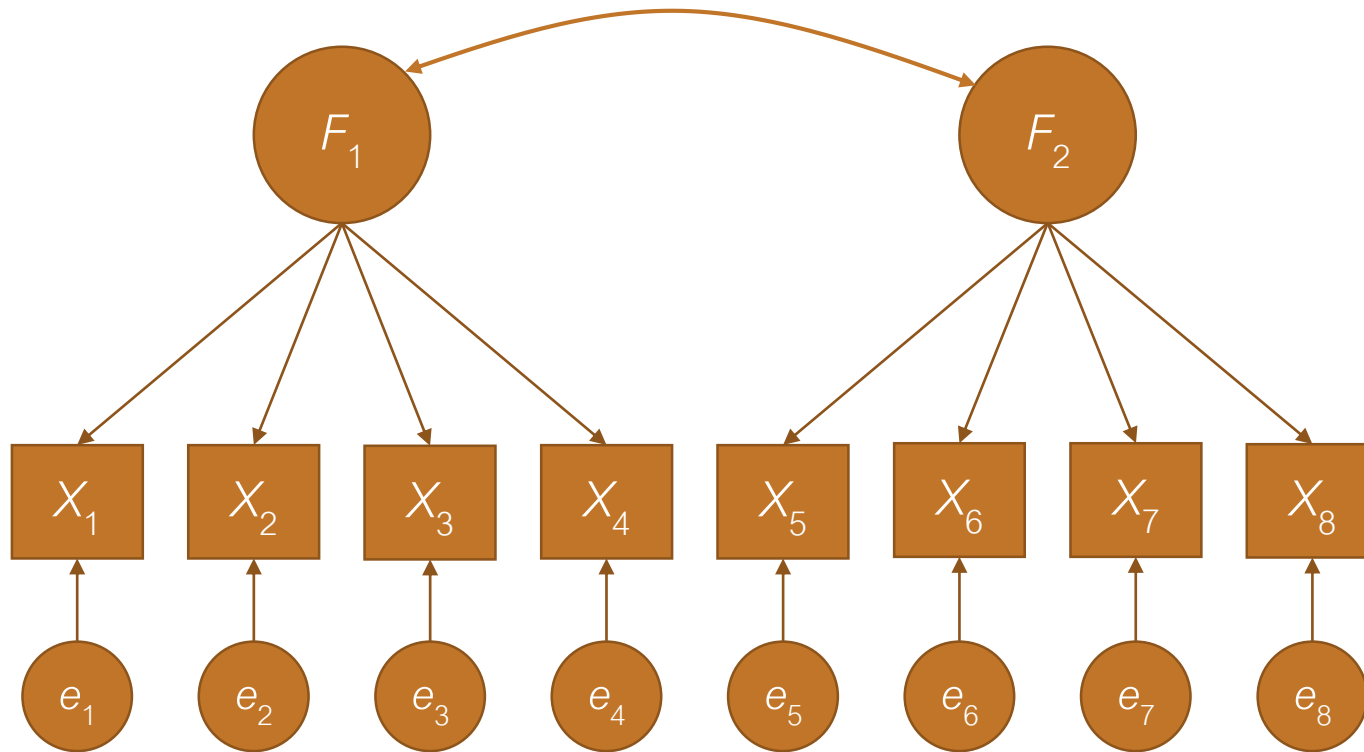
ให้วงกลมแทนองค์ประกอบหรือตัวแปรแฝง ตัวอย่างแรกสามารถแสดงเป็นภาพได้ดังนี้



กล่าวคือ คะแนนจากเพื่อนแต่ละคนสามารถแบ่งออกเป็น 2 ส่วน คือ

1. ส่วนที่ถูกอธิบายด้วย F ร่วมกับคะแนนจากเพื่อนคนอื่น
2. ส่วนที่ไม่ได้ถูกอธิบายด้วย F ซึ่งเป็นคะแนนจำเพาะของเพื่อนแต่ละคน

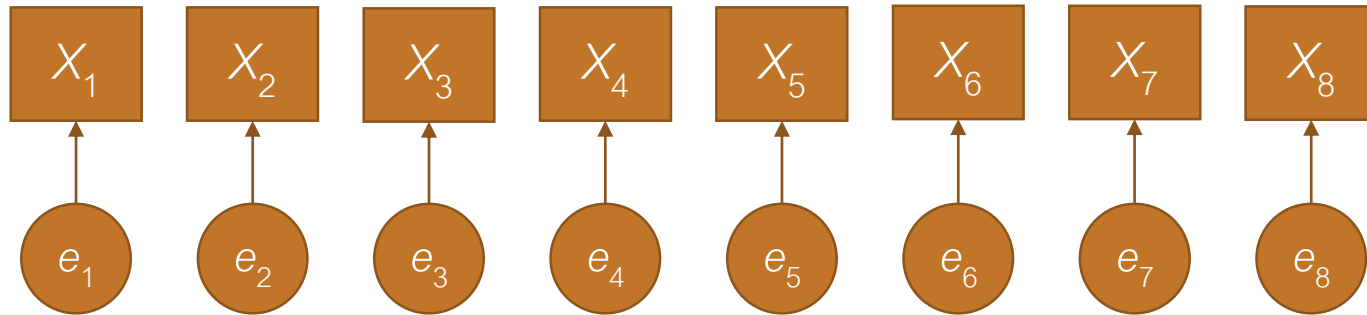
ตัวอย่างที่ 2 สามารถแสดงเป็นภาพได้ดังนี้



ในที่นี้ ส่วนที่เป็นองค์ประกอบถูกแบ่งออกเป็นสองส่วน คือ องค์ประกอบที่ 1 และองค์ประกอบที่ 2 ซึ่งอาจแทนความสวยแบบฝรั่งและความสวยแบบเอเชีย

เช่นเดิม e จะเป็นส่วนที่จำเพาะในคะแนนของเพื่อนแต่ละคน

ตัวอย่างสุดท้ายสามารถแสดงเป็นภาพได้ดังนี้



ในที่นี้ ส่วนที่อธิบายร่วมกันด้วย F ไม่มีเลย เนื่องจากว่าคะแนนไม่มีความสัมพันธ์กัน

แนะนำ

- ในความเป็นจริงแล้ว เมทริกซ์สหสัมพันธ์ไม่ได้สามารถสังเกตง่ายขนาดนี้ ทำให้นักวิจัยไม่ทราบว่า
 - ข้อคำถามเหล่านี้มีองค์ประกอบร่วมกันหรือไม่
 - ถ้ามีแล้ว มีกี่องค์ประกอบ
 - แล้วองค์ประกอบอะไร อธิบายคะแนนของข้อใดหรือเพื่อนคนใดเป็นหลัก
 - แล้วองค์ประกอบเหล่านี้ มีความหมายว่าอย่างไร
- ตามกันใน [#การวิเคราะห์องค์ประกอบ](#) [#ถ่ายนิดเดียวเอง](#) [#ยากโคตร](#)

ตัวอย่างแนะนำขั้นตอนการวิเคราะห์ องค์ประกอบ

สถิติชั้นกลางสำหรับจิตวิทยา

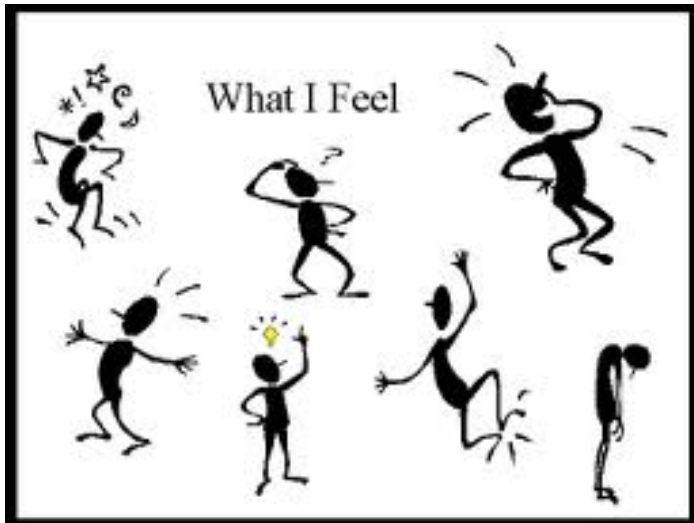
สันหัด พรประเสริฐมานิต

ตัวอย่างแนะนำขั้นตอนการวิเคราะห์องค์ประกอบ

- สำหรับสภานิติบัญญัตินี้ ผมจะแนะนำขั้นตอนการวิเคราะห์จากตัวอย่างก่อน เพื่อเป็นการแนะนำคำศัพท์พื้นฐาน พร้อมทั้งเข้าใจภาพรวมของการวิเคราะห์
- สมมติว่า ท่านกำลังสนใจว่าอารมณ์ทางบวกและทางลบนั้น
 - เป็นปัจจัยเดียวกันที่อยู่กันคนละขั้ว คือ มีอารมณ์บวกแล้วจะต้องไม่มีอารมณ์ลบ
 - เป็นคนละปัจจัย คือ มีอารมณ์บวกแล้วอาจมีอารมณ์ลบไปด้วยได้
- ด้วยเหตุนี้ ท่านจึงใช้การวิเคราะห์องค์ประกอบเพื่อตอบคำถามดังกล่าว

ตัวอย่างแนะนำขั้นตอนการวิเคราะห์องค์ประกอบ

ท่านต้องการจับกลุ่มคำที่แสดงถึง
ความดีใจหรือเสียใจ รวมจำนวน 10 คำ เพื่อดูว่าอยู่ใน
องค์ประกอบเดียวกันหรือไม่



ท่านจึงเก็บข้อมูล 200 คน ให้ประเมินความรู้สึกของตน
ในช่วงหนึ่งวันที่ผ่านมา

- | | |
|--------------|------------|
| 1. ดีใจ | 6. หดหู่ |
| 2. ร่าเริง | 7. เศร้า |
| 3. สบายใจ | 8. ทุกข์ใจ |
| 4. มีความสุข | 9. หมดหวัง |
| 5. ยินดี | 10. กลุ้ม |

การประเมินจะเป็นมาตร 5 ระดับ: ไม่ตรงกับฉันอย่างยิ่ง, ไม่ตรงกับฉัน, ตรงและไม่ตรงพอกัน, ตรงกับฉัน, ตรงกับฉันอย่างยิ่ง

```
> round(cor(dat), 2)
```

	EMO01	EMO02	EMO03	EMO04	EMO05	EMO06	EMO07	EMO08	EMO09	EMO10
EMO01	1.00	0.54	0.54	0.64	0.59	-0.32	-0.29	-0.26	-0.33	-0.33
EMO02	0.54	1.00	0.56	0.58	0.55	-0.29	-0.27	-0.27	-0.29	-0.32
EMO03	0.54	0.56	1.00	0.69	0.72	-0.29	-0.35	-0.22	-0.36	-0.32
EMO04	0.64	0.58	0.69	1.00	0.72	-0.32	-0.32	-0.29	-0.37	-0.34
EMO05	0.59	0.55	0.72	0.72	1.00	-0.33	-0.39	-0.30	-0.39	-0.37
EMO06	-0.32	-0.29	-0.29	-0.32	-0.33	1.00	0.45	0.57	0.61	0.58
EMO07	-0.29	-0.27	-0.35	-0.32	-0.39	0.45	1.00	0.46	0.51	0.48
EMO08	-0.26	-0.27	-0.22	-0.29	-0.30	0.57	0.46	1.00	0.67	0.56
EMO09	-0.33	-0.29	-0.36	-0.37	-0.39	0.61	0.51	0.67	1.00	0.67
EMO10	-0.33	-0.32	-0.32	-0.34	-0.37	0.58	0.48	0.56	0.67	1.00

สังเกตว่า อารมณ์ทางบวกด้วยกันหรืออารมณ์ทางลบด้วยกัน สัมพันธ์ด้วยกันเอง สูงกว่าความสัมพันธ์อารมณ์ทางบวกกับทางลบ

น่าจะมีสององค์ประกอบ

สกัดองค์ประกอบ คือ วิธีการดึงข้อมูล
ส่วนที่มีความสัมพันธ์ร่วมกันระหว่าง
ข้อคำถามออกมา ตกให้ได้ 2 องค์ประกอบ

การหมุนแกน เป็นการเปลี่ยน
มิติในการมององค์ประกอบ
เพื่อให้แปลความหมายได้ง่ายขึ้น

เลือกวิธีการสกัดองค์ประกอบ
แบบ Maximum Likelihood

เลือกวิธีการ Quartimin
ในการหมุนแกน

```
> result <- fa(dat[,1:10], fm="ml", nfactors=2, rotate="quartimin")
```

เลือกจำนวนองค์ประกอบเท่ากับ 2

ใส่ตัวแปรที่ใช้ในการวิเคราะห์องค์ประกอบลงไป

นำหน้าองค์ประกอบหลังจากหมุนแกน

ส่วนส่วนของความแปรปรวนที่องค์ประกอบทั้งหมดสามารถอธิบายข้อความได้

ในตารางนี้ บอกว่าองค์ประกอบแรกแสดงถึงอารมณ์ทางบวก และองค์ประกอบที่ 2 แสดงถึงอารมณ์ทางลบ

จากหน้าหน้าองค์ประกอบนี้ บอกได้ค่อนข้างชัดเจนว่า อารมณ์ทางบวกและอารมณ์ทางลบสามารถมองเป็นคนละเรื่องกันได้

	ML1	ML2	h2	u2	com
EMO01	0.69	-0.05	0.51	0.49	1.0
EMO02	0.67	-0.03	0.46	0.54	1.0
EMO03	0.84	0.04	0.67	0.33	1.0
EMO04	0.86	0.02	0.73	0.27	1.0
EMO05	0.84	-0.02	0.72	0.28	1.0
EMO06	-0.02	0.72	0.53	0.47	1.0
EMO07	-0.14	0.53	0.38	0.62	1.1
EMO08	0.08	0.81	0.60	0.40	1.0
EMO09	-0.01	0.86	0.74	0.26	1.0
EMO10	-0.04	0.75	0.59	0.41	1.0

	ML1	ML2
SS Loadings	3.14	2.80
Proportion Var	0.31	0.28
Cumulative Var	0.31	0.59
Proportion Explained	0.53	0.47
Cumulative Proportion	0.53	1.00

สัดส่วนของความแปรปรวนของข้อคำถามทั้งหมดที่แต่ละองค์ประกอบสามารถอธิบายได้

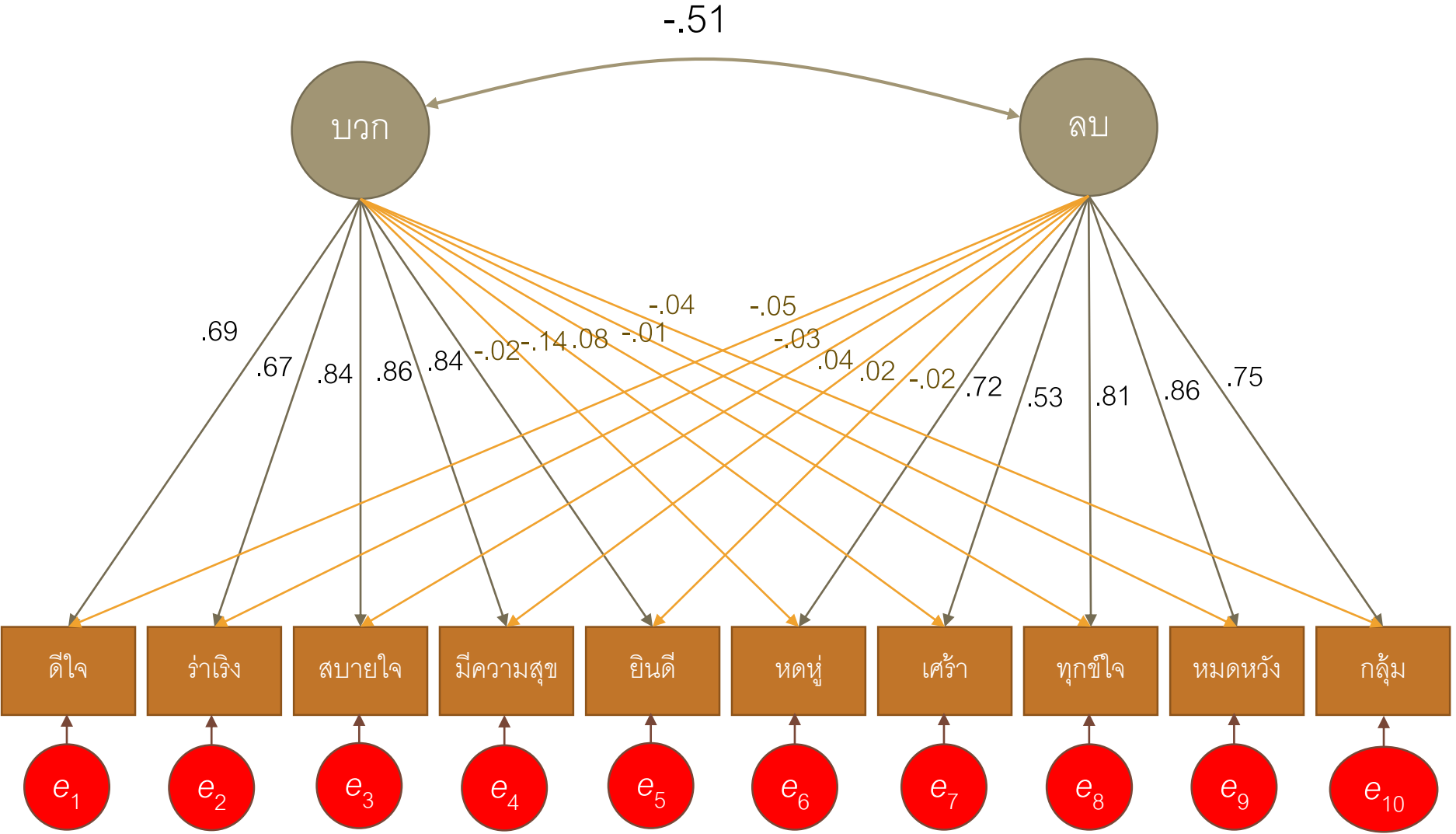
with factor correlations of

	ML1	ML2
ML1	1.00	-0.51
ML2	-0.51	1.00

ความสัมพันธ์ระหว่างองค์ประกอบทั้งสอง

ดัชนีที่บอกว่า องค์ประกอบ 2 องค์ประกอบที่คัดเลือกมา
เพียงพอในการอธิบายข้อคำถามทั้งหมดหรือไม่

Tucker Lewis Index of factoring reliability = 1.012
RMSEA index = 0 and the 90 % confidence intervals are 0 0.036
BIC = -118.29
Fit based upon off diagonal values = 1
Measures of factor score adequacy



ตัวอย่างแนะนำขั้นตอนการวิเคราะห์องค์ประกอบ

- จากตัวอย่างที่ผ่านมา ค่อนข้างมั่นใจได้ว่าอารมณ์ทางบวกและอารมณ์ทางลบ อยู่คนละองค์ประกอบ (มิติ) กัน
- แต่ทำไม ผมถึงเลือกสกัดองค์ประกอบเพียงแค่ 2 องค์ประกอบ?
- ทำไมผมถึงใช้วิธีการหมุนแกนแบบ Quartimin?
- แล้วผมแน่ใจได้อย่างไร ว่าการวิเคราะห์แบบนี้ถูกต้องแล้ว?

โมเดลการวิเคราะห์องค์ประกอบ

สถิติขั้นกลางสำหรับจิตวิทยา

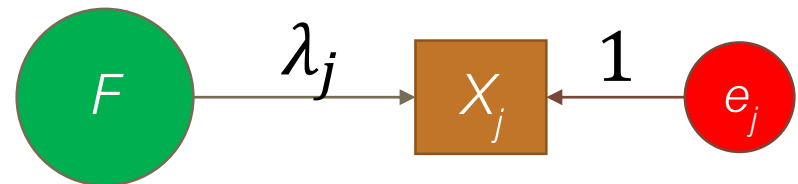
สันหัต พรประเสริฐมานิต

โมเดลการวิเคราะห์องค์ประกอบ

- จากรูปองค์ประกอบเดียว คะแนนแต่ละข้อถูกอธิบายด้วยองค์ประกอบและคะแนนจำเพาะในแต่ละข้อ คือ

$$X_j = \mu_j + \lambda_j F + e_j$$

- X_{ij} คือ คะแนนที่สังเกตได้จากข้อมูลข้อที่ j
- μ_j คือ ค่าเฉลี่ยของข้อคำถามระหว่างบุคคลในข้อที่ j
- λ_j คือ คะแนนสังเกตได้ของข้อที่ j ที่เปลี่ยนแปลงไปเมื่อคะแนนองค์ประกอบเพิ่มขึ้น 1 หน่วย เรียกว่า น้ำหนักองค์ประกอบ (Factor loading) ของข้อที่ j

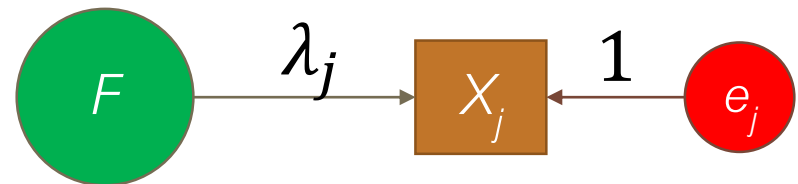


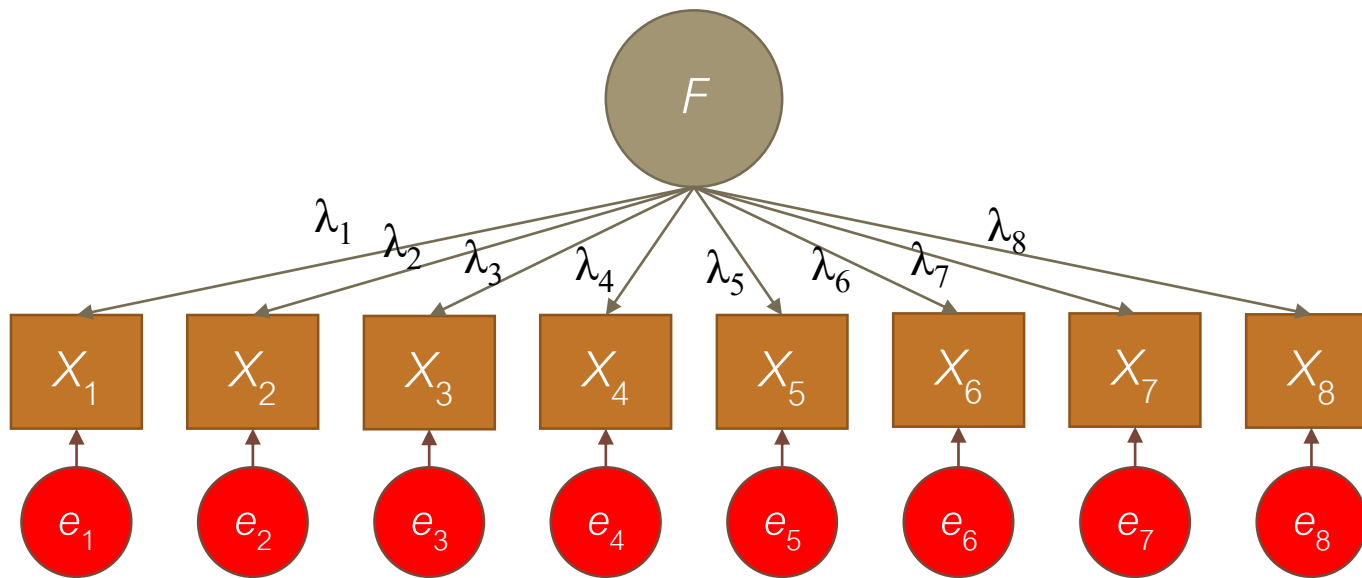
โมเดลการวิเคราะห์องค์ประกอบ

- จากรูปองค์ประกอบเดียว คะแนนแต่ละข้อถูกอธิบายด้วยองค์ประกอบและคะแนนจำเพาะในแต่ละข้อ คือ

$$X_j = \mu_j + \lambda_j F + e_j$$

- F คือ คะแนนขององค์ประกอบ (ที่วัดไม่ได้)
- e_j คือ ความแตกต่างระหว่างคะแนนที่สังเกตได้และคะแนนจากองค์ประกอบในข้อที่ j ซึ่งอาจแสดงถึงคะแนนจำเพาะของแต่ละข้อคำถาม
- สังเกตว่าโมเดลนี้จะคล้ายกับการวิเคราะห์ถดถอย แต่ตัวแปรทำนายจะเปลี่ยนจากสังเกตได้เป็นตัวแปรแฝง





$$X_1 = \mu_1 + \lambda_1 F + e_1$$

$$X_2 = \mu_2 + \lambda_2 F + e_2$$

$$X_3 = \mu_3 + \lambda_3 F + e_3$$

$$X_4 = \mu_4 + \lambda_4 F + e_4$$

$$X_5 = \mu_5 + \lambda_5 F + e_5$$

$$X_6 = \mu_6 + \lambda_6 F + e_6$$

$$X_7 = \mu_7 + \lambda_7 F + e_7$$

$$X_8 = \mu_8 + \lambda_8 F + e_8$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \\ \mu_7 \\ \mu_8 \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} [F] + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

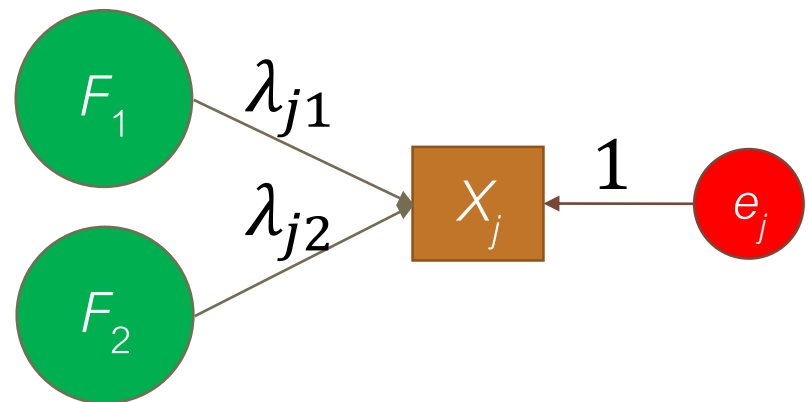
$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\lambda}F + \mathbf{e}$$

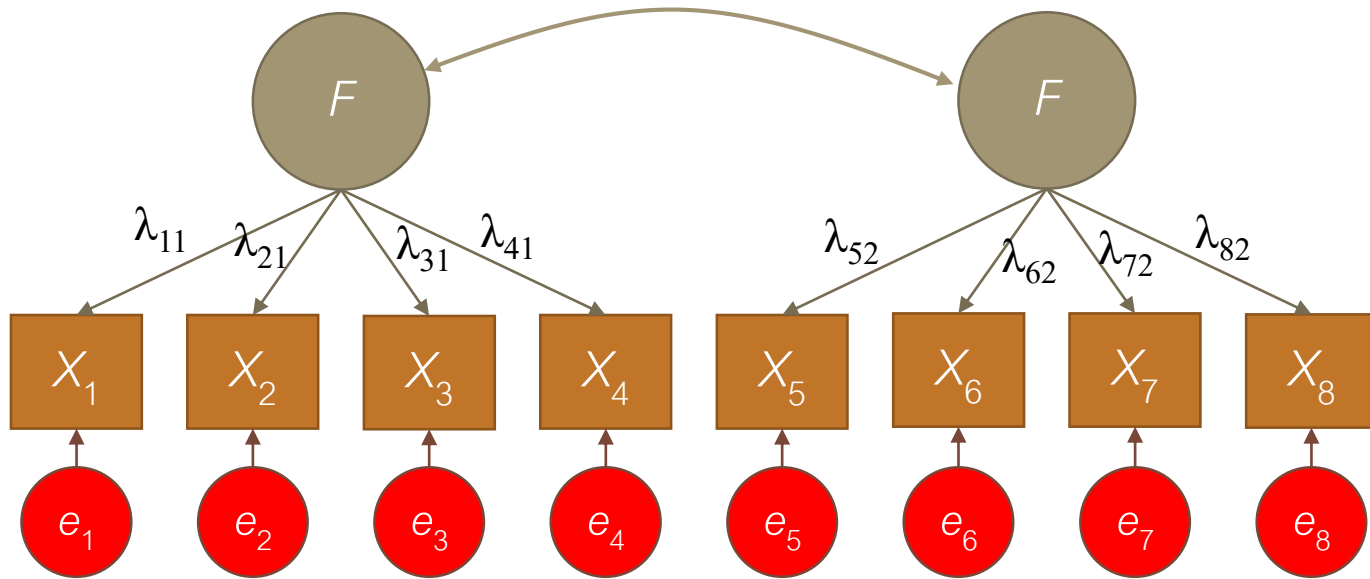
โมเดลการวิเคราะห์องค์ประกอบ

- ถ้ามีสององค์ประกอบ คะแนนแต่ละข้อถูกอธิบายด้วยองค์ประกอบทั้งสอง และคะแนนจำเพาะในแต่ละข้อ คือ

$$X_j = \mu_j + \lambda_{j1}F_1 + \lambda_{j2}F_2 + e_j$$

- λ_{j1} และ λ_{j2} คือ น้ำหนักองค์ประกอบ (Factor loading) ของข้อที่ j สำหรับองค์ประกอบที่ 1 และ 2 ตามลำดับ





$$\begin{aligned}
 X_1 &= \mu_1 + \lambda_{11}F_1 + e_1 \\
 X_2 &= \mu_2 + \lambda_{21}F_1 + e_2 \\
 X_3 &= \mu_3 + \lambda_{31}F_1 + e_3 \\
 X_4 &= \mu_4 + \lambda_{41}F_1 + e_4 \\
 X_5 &= \mu_5 + \lambda_{52}F_2 + e_5 \\
 X_6 &= \mu_6 + \lambda_{62}F_2 + e_6 \\
 X_7 &= \mu_7 + \lambda_{72}F_2 + e_7 \\
 X_8 &= \mu_8 + \lambda_{82}F_2 + e_8
 \end{aligned}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \\ \mu_7 \\ \mu_8 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}$$

โมเดลการวิเคราะห์องค์ประกอบ

- ดังนั้น โมเดลการวิเคราะห์องค์ประกอบในรูปของเมทริกซ์ คือ

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}$$

- ให้ J เป็นจำนวนข้อคำถาม และ K เป็นจำนวนองค์ประกอบ
- \mathbf{x} คือ เวกเตอร์ความยาว J แสดงคะแนนทุกข้อ
- $\boldsymbol{\mu}$ คือ เวกเตอร์ความยาว J แสดงค่าเฉลี่ยของคะแนนทุกข้อ
- $\boldsymbol{\Lambda}$ คือ เมทริกซ์ขนาด $J \times K$ ที่แสดงน้ำหนักองค์ประกอบของแต่ละข้อที่มีต่อแต่ละองค์ประกอบ

โมเดลการวิเคราะห์องค์ประกอบ

- ดังนั้น โมเดลการวิเคราะห์องค์ประกอบในรูปของเมทริกซ์ คือ

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}$$

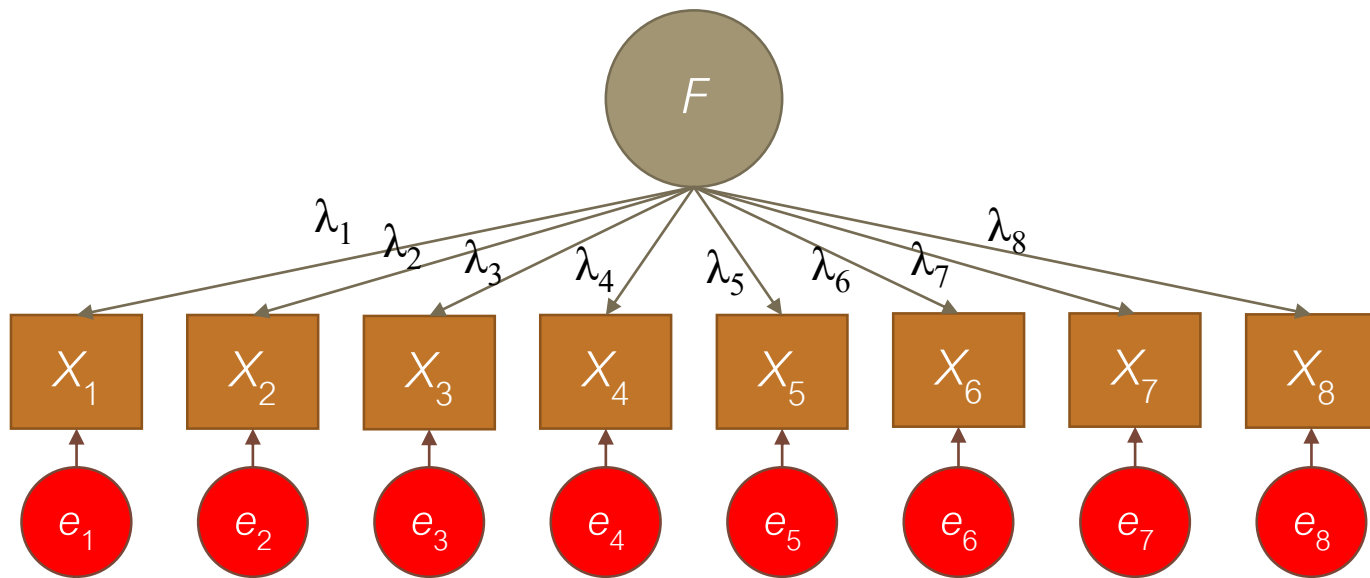
- \mathbf{F} คือ เวกเตอร์ความยาว K แสดงคะแนนองค์ประกอบ
- \mathbf{e} คือ เวกเตอร์ความยาว J แสดงค่าเบี่ยงเบนของทุกข้อ
- โดยที่ค่าเฉลี่ยและเมทริกซ์ความแปรปรวนร่วมของ \mathbf{x} จะมีค่าเท่ากับ $\boldsymbol{\mu}$ และ $\boldsymbol{\Sigma}$ ตามลำดับ
- โดยที่ค่าเฉลี่ยขององค์ประกอบ \mathbf{F} เท่ากับ 0 เสมอ และ $\boldsymbol{\Phi}$ แทนเมทริกซ์ความแปรปรวนร่วมระหว่างองค์ประกอบ ส่วนใหญ่แล้วกำหนดให้องค์ประกอบมีความแปรปรวนเท่ากับ 1 ทำให้ $\boldsymbol{\Phi}$ เป็นเมทริกซ์สหสัมพันธ์

โมเดลการวิเคราะห์องค์ประกอบ

- ดังนั้น โมเดลการวิเคราะห์องค์ประกอบในรูปของเมทริกซ์ คือ

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}$$

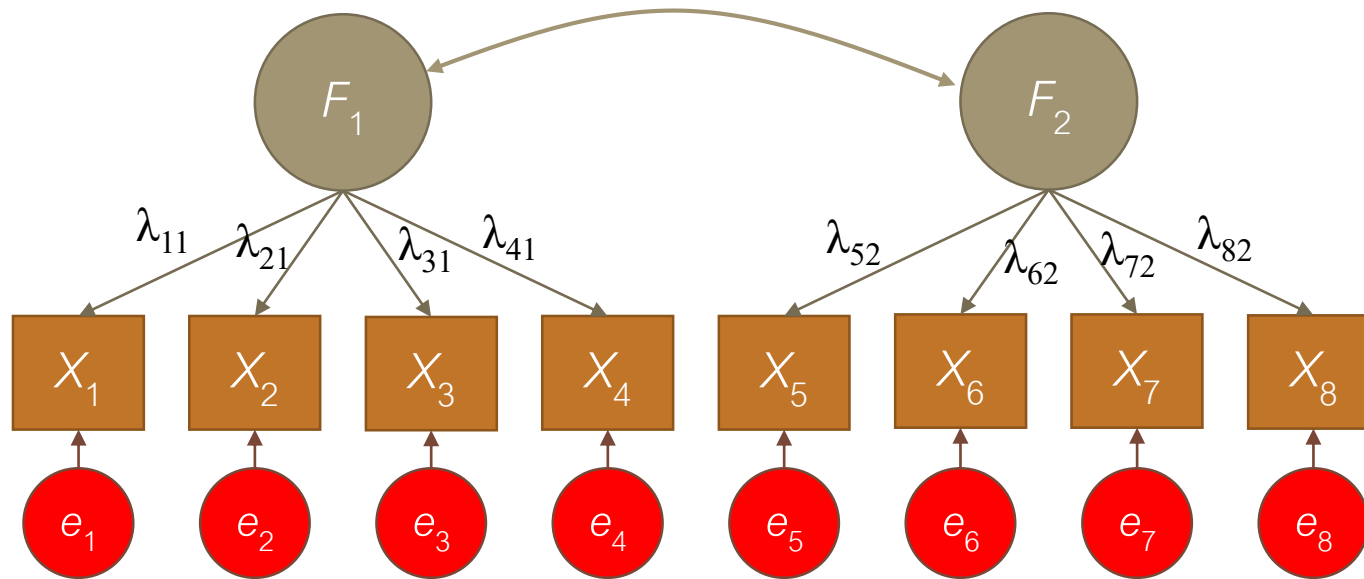
- ค่าเฉลี่ย \mathbf{e} มีค่าเท่ากับ 0 เสมอ และ $\mathbf{0}$ เป็นเมทริกซ์ความแปรปรวนระหว่างความผิดพลาดในการทำนาย โดยส่วนใหญ่แล้วจะให้คะแนนจำเพาะของแต่ละข้อคำถามไม่มีความสัมพันธ์กัน หรือหน่วยนอกแนวทแยงมีค่าเท่ากับ 0
- เนื่องจากค่าองค์ประกอบเป็นค่าไม่มีอยู่จริง ในสมการเราไม่รู้ทั้ง \mathbf{e} และ \mathbf{F} ทำให้ไม่สามารถหาค่าของ $\boldsymbol{\Lambda}$ ได้โดยตรง ซึ่งต่างกับการวิเคราะห์ถดถอย



$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & & & & & & \\ \sigma_{21} & \sigma_2^2 & & & & & & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & & & & & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & & & & & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 & & & & \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_6^2 & & & \\ \sigma_{71} & \sigma_{72} & \sigma_{73} & \sigma_{74} & \sigma_{75} & \sigma_{76} & \sigma_7^2 & & \\ \sigma_{81} & \sigma_{82} & \sigma_{83} & \sigma_{84} & \sigma_{85} & \sigma_{86} & \sigma_{87} & \sigma_8^2 & \end{bmatrix}$$

$$\Theta = \begin{bmatrix} \theta_{11} & & & & & & & & \\ 0 & \theta_{22} & & & & & & & \\ 0 & 0 & \theta_{33} & & & & & & \\ 0 & 0 & 0 & \theta_{44} & & & & & \\ 0 & 0 & 0 & 0 & \theta_{55} & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{77} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{88} & \end{bmatrix}$$

$$\Phi = [1]$$



$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & & & & & & \\ \sigma_{21} & \sigma_2^2 & & & & & & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & & & & & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & & & & & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 & & & & \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_6^2 & & & \\ \sigma_{71} & \sigma_{72} & \sigma_{73} & \sigma_{74} & \sigma_{75} & \sigma_{76} & \sigma_7^2 & & \\ \sigma_{81} & \sigma_{82} & \sigma_{83} & \sigma_{84} & \sigma_{85} & \sigma_{86} & \sigma_{87} & \sigma_8^2 & \end{bmatrix}$$

$$\Theta = \begin{bmatrix} \theta_{11} & & & & & & & & \\ 0 & \theta_{22} & & & & & & & \\ 0 & 0 & \theta_{33} & & & & & & \\ 0 & 0 & 0 & \theta_{44} & & & & & \\ 0 & 0 & 0 & 0 & \theta_{55} & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{77} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{88} & \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 1 & & & & & & & & \\ \phi_{21} & 1 & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \\ & & & & & & & & \end{bmatrix}$$

โมเดลการวิเคราะห์องค์ประกอบ

- นักสถิติจึงหลีกเลี่ยงสมการดังกล่าว โดยใช้หลักการความแปรปรวนร่วมของค่าในทางซ้ายของสมการต้องเท่ากับค่าทางขวาของสมการ

$$\text{Cov}(\mathbf{x}) = \text{Cov}(\boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e})$$

$$\text{Cov}(\mathbf{x}) = \text{Cov}(\boldsymbol{\Lambda}\mathbf{F} + \mathbf{e})$$

เนื่องจาก $\boldsymbol{\mu}$ เป็นค่าคงที่

$$\text{Cov}(\mathbf{x}) = \text{Cov}(\boldsymbol{\Lambda}\mathbf{F}) + \text{Cov}(\mathbf{e})$$

เนื่องจาก \mathbf{F} และ \mathbf{e} ไม่สัมพันธ์กัน

$$\text{Cov}(\mathbf{x}) = \boldsymbol{\Lambda}\text{Cov}(\mathbf{F})\boldsymbol{\Lambda}' + \text{Cov}(\mathbf{e})$$

เนื่องจาก $\boldsymbol{\Lambda}$ เป็นค่าคงที่

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

โมเดลการวิเคราะห์องค์ประกอบ

$$\Sigma = \Lambda\Phi\Lambda' + \Theta$$

- หลักการวิเคราะห์ จะใช้การหาค่าของ Λ , Φ , และ Θ เพื่อให้ค่าทางขวาใกล้เคียงกับทางซ้ายหรือ Σ มากที่สุด ตามลักษณะของโมเดลที่ผู้วิจัยกำหนด (เช่น จำนวนองค์ประกอบ, ความสัมพันธ์ระหว่างองค์ประกอบ)

ทฤษฎีการทดสอบแบบคลาสสิก

สถิติขั้นกลางสำหรับจิตวิทยา

สันหัต พรประเสริฐมานิต

ทฤษฎีการทดสอบแบบคลาสสิก

- ในทฤษฎีการทดสอบแบบคลาสสิก (Classical test theory) คะแนนของแต่ละข้อคำถามสามารถแบ่งได้เป็นสองส่วน

$$X_j = T_j + d_j$$

- X_j คือ คะแนนที่สังเกตได้จากข้อมูลข้อที่ j
- T_j คือ คะแนนที่แท้จริงที่ได้ของข้อที่ j
- d_j คือ คะแนนความผิดพลาดในการวัดของข้อที่ j (ไม่ใช่ e เพราะจะซ้ำกับในสมการการวิเคราะห์องค์ประกอบ)

ทฤษฎีการทดสอบแบบคลาสสิก

- คะแนนจริง (T_j) สามารถแบ่งออกเพิ่มเติมได้เป็นสองส่วน

$$X_j = T_j + \varepsilon_j = \mu_j + C_j + U_j + d_j$$

- C_j คือ คะแนนส่วนที่สัมพันธ์กับคะแนนอื่นหรือคะแนนร่วม (Common score)
- U_j คือ คะแนนจริงที่จำเพาะภายในข้อคำถามดังกล่าว (Unique score)
- เมื่อเปรียบเทียบกับโมเดลวิเคราะห์องค์ประกอบ

โมเดลทฤษฎีการทดสอบแบบคลาสสิก

$$X_j = \mu_j + C_j + U_j + d_j$$

โมเดลการวิเคราะห์องค์ประกอบ

$$X_j = \mu_j + \lambda_j F + e_j$$

ทฤษฎีการทดสอบแบบคลาสสิก

- ดังนั้นคะแนนจำเพาะในโมเดลการวิเคราะห์องค์ประกอบ จึงประกอบด้วยสองส่วน คือ คะแนนจริงจำเพาะในข้อคำถามนั้น และความผิดพลาดในการวัด
- เช่น คะแนนการประเมินความหน้าตาดีจากเพื่อน
 - คะแนนรวม (C_j) คือ คะแนนประเมินที่สัมพันธ์กับคะแนนของคนอื่นไม่ว่าจะด้วยเหตุผลอะไรก็ตาม เช่น เพราะวัดหน้าตาเหมือนกัน หรือเพราะชอบคนหน้าแปลกเหมือนกัน
 - คะแนนจำเพาะ (U_j) คือ คะแนนประเมินที่ไม่สัมพันธ์กับคะแนนคนอื่น เช่น ชอบคนหน้าตาเหมือนแฟนเก่าของตนเอง
 - คะแนนจริง (T_j) คือ คะแนนประเมินจริงทั้งส่วนของคะแนนรวมและคะแนนจำเพาะ
 - ความผิดพลาดในการวัด (d_j) คือ ค่าความผิดพลาดในการวัดที่เกิดขึ้นเฉพาะเวลาที่วัด ไม่สามารถสร้างซ้ำได้ เช่น วันดังกล่าวเพื่อนของคุณอารมณ์ไม่ดี เห็นภาพไม่ชัด

รูปแบบการวิเคราะห์องค์ประกอบ

สถิติขั้นกลางสำหรับจิตวิทยา

สันหัต พรประเสริฐมานิต

รูปแบบการวิเคราะห์องค์ประกอบ

- รูปแบบการวิเคราะห์องค์ประกอบ มีสองรูปแบบ คือ
 - การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis; EFA)
 - การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory factor analysis; CFA)

รูปแบบการวิเคราะห์องค์ประกอบ

- EFA เป็นรูปแบบการวิเคราะห์ที่ผู้วิจัยยังไม่มีโครงสร้างขององค์ประกอบ ผู้วิจัยยังไม่ทราบว่าองค์ประกอบมีกี่องค์ประกอบ องค์ประกอบใดควรมีอิทธิพลต่อข้อคำถามใด
- ผู้วิจัยจะใช้ EFA เพื่อค้นหาว่ามีองค์ประกอบภายใต้คำถามเหล่านี้อยู่ในรูปแบบใดได้บ้าง
- ผู้วิจัยจำเป็นต้องคัดเลือกรูปแบบองค์ประกอบที่เหมาะสม หลังจากนั้นแล้วควรมีการทำซ้ำเพื่อพิสูจน์ว่าโครงสร้างองค์ประกอบดังกล่าวใช้ได้จริง

รูปแบบการวิเคราะห์องค์ประกอบ

- CFA เป็นรูปแบบการวิเคราะห์ที่ผู้วิจัยมีโครงสร้างขององค์ประกอบชัดเจน
- นักวิจัยจะใช้ CFA เพื่อทดสอบว่าโมเดลที่ตั้งขึ้นมานั้น เหมาะสมกับข้อมูลที่มี มากน้อยเพียงใด
- EFA และ CFA จะมีโมเดลวิเคราะห์แบบเดียวกัน ก็คือ โมเดลวิเคราะห์ องค์ประกอบ เพียงแต่ข้อจำกัดที่ใส่ในโมเดล (model constraints) จะ แตกต่างกัน โดย CFA จะมีข้อจำกัดมากกว่า
- ต่อไปจะกล่าวถึง EFA ซึ่งกระบวนการจะประกอบไปด้วยสองส่วนใหญ่ๆ คือ การสกัดองค์ประกอบและการหมุนแกน

การสกัดองค์ประกอบ

สถิติขั้นกลางสำหรับจิตวิทยา

สันหัด พรประเสริฐมานิต

โมเดลการวิเคราะห์องค์ประกอบ

- ลองนึกภาพว่า ท่านถูกให้หาค่า $a + b + c = 10$ และ $b + c - a = 5$ ท่านจะพบว่า a, b, c จำนวนมากที่ทำให้สมการทั้งสองเป็นจริง
- เช่นกันกับค่า $\mathbf{C} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}'$ ที่ค่าได้ในที่แตกต่างกัน ทำให้ได้ผลลัพธ์เท่ากัน เรียกปรากฏการณ์นี้ว่า ความกำหนัดไม่ได้ขององค์ประกอบ (Factor Indeterminacy)

$$\Lambda_A = \begin{array}{r} \\ \text{EMO01} \\ \text{EMO02} \\ \text{EMO03} \\ \text{EMO04} \\ \text{EMO05} \\ \text{EMO06} \\ \text{EMO07} \\ \text{EMO08} \\ \text{EMO09} \\ \text{EMO10} \end{array} \begin{array}{cc} & \begin{array}{c} [,1] \\ [,2] \end{array} \\ \begin{array}{c} 0.686 \\ 0.666 \\ 0.837 \\ 0.865 \\ 0.836 \\ -0.018 \\ -0.144 \\ 0.080 \\ -0.006 \\ -0.042 \end{array} & \begin{array}{c} -0.049 \\ -0.030 \\ 0.036 \\ 0.018 \\ -0.024 \\ 0.721 \\ 0.533 \\ 0.811 \\ 0.857 \\ 0.746 \end{array} \end{array}$$

$$\Phi_A = \begin{array}{r} \\ \text{ML1} \\ \text{ML2} \end{array} \begin{array}{cc} & \begin{array}{c} [,1] \\ [,2] \end{array} \\ \begin{array}{c} 1.000 \\ -0.506 \end{array} & \begin{array}{c} -0.506 \\ 1.000 \end{array} \end{array}$$

$$\Lambda_A \Phi_A \Lambda'_A = \Lambda_B \Phi_B \Lambda'_B$$

ยังมีชุดผลลัพธ์ที่ C, D, E, ... ที่ทำให้สมการนี้เป็นจริง

$$\Lambda_B = \begin{array}{r} \\ \text{EMO01} \\ \text{EMO02} \\ \text{EMO03} \\ \text{EMO04} \\ \text{EMO05} \\ \text{EMO06} \\ \text{EMO07} \\ \text{EMO08} \\ \text{EMO09} \\ \text{EMO10} \end{array} \begin{array}{cc} & \begin{array}{c} [,1] \\ [,2] \end{array} \\ \begin{array}{c} 0.672 \\ 0.648 \\ 0.796 \\ 0.827 \\ 0.811 \\ -0.198 \\ -0.272 \\ -0.127 \\ -0.221 \\ -0.227 \end{array} & \begin{array}{c} -0.235 \\ -0.211 \\ -0.194 \\ -0.220 \\ -0.252 \\ 0.703 \\ 0.555 \\ 0.764 \\ 0.831 \\ 0.733 \end{array} \end{array}$$

$$\Phi_B = \begin{array}{r} \\ \text{ML1} \\ \text{ML2} \end{array} \begin{array}{cc} & \begin{array}{c} \text{ML1} \\ \text{ML2} \end{array} \\ \begin{array}{c} 1 \\ 0 \end{array} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

โมเดลการวิเคราะห์องค์ประกอบ

- นักสถิติจึงกำหนดว่า ให้หาค่า Λ และ Φ ชุดแรกออกมาก่อน โดยกำหนดข้อจำกัด (Model Constraints) เพิ่มเพื่อให้ค่าผลลัพธ์ออกมาเหมือนกัน และให้ค่า C หนึ่ง โดยข้อจำกัดเป็นดังต่อไปนี้
 - ให้ Φ เป็น Identity Matrix
 - หาก $l_k = \lambda_{1k}^2 + \lambda_{2k}^2 + \dots + \lambda_{jk}^2$ กำหนดให้ $l_1 \geq l_2 \geq \dots \geq l_k$ โดยค่า l_k จะเรียกว่าค่า Eigenvalue ขององค์ประกอบ k
- เมื่อสกัดองค์ประกอบมาแล้ว ได้ค่า Λ และ Φ ชุดแรกแล้ว ค่อยเลือกชุดของค่า Λ และ Φ ที่สามารถแปลความหมายได้ดีที่สุด

โมเดลการวิเคราะห์องค์ประกอบ

- จากสมการ

$$\Sigma = \Lambda\Phi\Lambda' + \Theta \rightarrow \Lambda\Lambda' + \Theta = \Sigma_M$$

- วิธีการที่นิยมมากที่สุดในการหาค่าของ Λ คือการสุ่มค่าไปเรื่อยๆ จนกว่าค่าด้านขวาของสมการจะใกล้เคียงด้านซ้ายของสมการ โดยให้เป็นไปตามข้อจำกัดเพื่อหลีกเลี่ยง Factor Indeterminacy
- ค่าใดที่ทำให้ใกล้เคียงมากที่สุด จะเป็นคำตอบ
- วิธีการนี้เรียกว่า วิธีการความเป็นไปได้สูงสุด (Maximum Likelihood)

การสกัดองค์ประกอบ

- สมการความเป็นไปได้ (Likelihood function) คือ

$$F_M = \text{tr}\{\Sigma\Sigma_M^{-1}\} - \log|\Sigma\Sigma_M^{-1}| - J$$

- ยิ่งค่า F_M น้อย แสดงว่า Σ ใกล้เคียงกับ Σ_M มาก
- โดยผู้วิจัยจะต้องกำหนดจำนวนองค์ประกอบก่อน แล้วคอมพิวเตอร์จะปรับค่า λ_{ik} และ θ_{kk} ที่ให้ F_M มีค่าน้อยที่สุด

การสกัดองค์ประกอบ

- วิธีการอื่นที่ใช้ในการสกัดองค์ประกอบ ที่แนะนำ คือ วิธีองค์ประกอบหลัก (Principal Axis Factoring)
- วิธีความแปรปรวนสูงสุด มีข้อได้เปรียบที่สามารถประมาณค่า Standard Error ของค่าที่ประมาณการทั้งหมด (λ_{ik} และ θ_{kk}) แต่มีข้อตกลงเบื้องต้นก่อนการใช้สถิติว่า การกระจายของตัวแปรที่เก็บข้อมูลมาต้องเป็น Multivariate Normal Distribution เพื่อให้การประมาณค่า SE ถูกต้อง
- ถ้า Multivariate Normal Distribution ไม่เป็นจริง สามารถแก้ไข โดยใช้วิธีการปรับแก้ของ Satorra & Bentler

การวิเคราะห์ส่วนประกอบ

สถิติชั้นกลางสำหรับจิตวิทยา

สันหัด พรประเสริฐมานิต

การวิเคราะห์ส่วนประกอบ

- การวิเคราะห์ส่วนประกอบ (Principal Component Analysis; PCA) เป็นสถิติที่ใกล้เคียงกับ EFA จนกระทั่งนักวิจัยหลายคนสับสนและเชื่อว่า PCA คือ EFA
- การวิเคราะห์ส่วนประกอบ คือ วิธีการที่ต้องการลดจำนวนตัวแปรให้น้อยลง โดยให้ส่วนประกอบที่เหลืออยู่ยังรักษาความแปรปรวนของข้อคำถามให้มากที่สุดเท่าที่เป็นไปได้
- PCA ไม่ค้ำประกันว่าตัวแปรแต่ละตัวจะมีความผิดพลาดในการวัดหรือข้อมูลจำเพาะ กล่าวคือ

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{P}_i \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$$

การวิเคราะห์ส่วนประกอบ

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{P}_i$$

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$$

- \mathbf{P}_i คือคะแนนส่วนประกอบ
- จะเห็นว่า ไม่มี \mathbf{e}_i และ Θ ในสมการ แสดงว่าโมเดลนี้ไม่สนใจความผิดพลาดในการวัด

การวิเคราะห์ส่วนประกอบ

- เนื่องจากในโมเดลไม่มี **Ⓜ** PCA จึงไม่เหมาะสมกับการวิเคราะห์ผลที่คาดหวังว่าจะหาตัวแปรแฝงหรือองค์ประกอบภายใต้ข้อคำถาม
- PCA เหมาะสมกับกรณีที่ผู้วิจัยต้องการลดจำนวนตัวแปร เช่น ตัวแปรต้นในการวิเคราะห์ถดถอยมีจำนวนมาก และต้องการลดตัวแปรลง
- ห้ามแปลความหมายราวกับว่ามีองค์ประกอบหรือตัวแปรแฝงภายใต้ข้อคำถามเหล่านั้น

การเลือกจำนวนองค์ประกอบ

สถิติชั้นกลางสำหรับจิตวิทยา

สันหัต พรประเสริฐมานิต

การเลือกจำนวนองค์ประกอบ

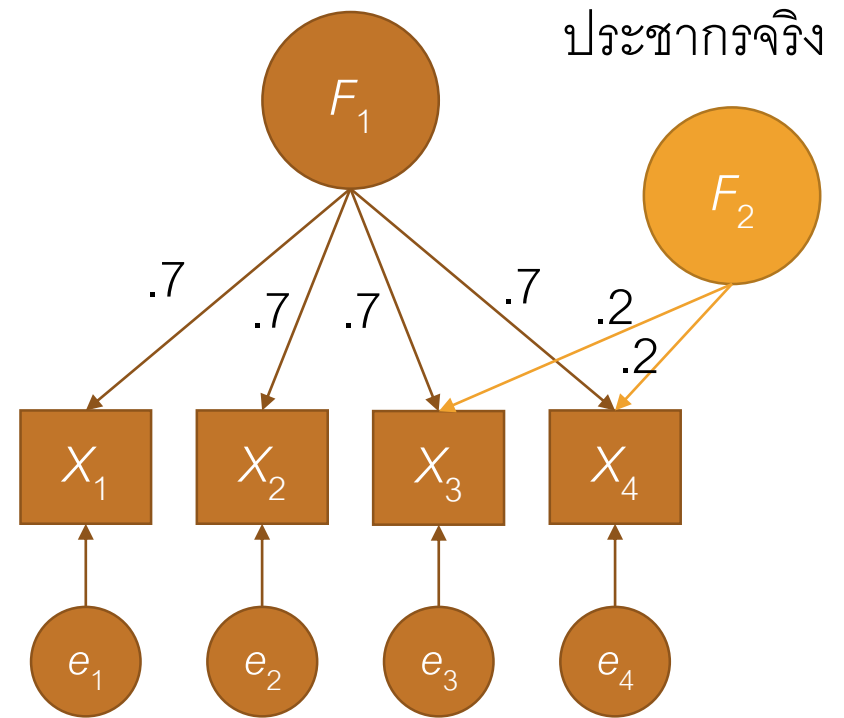
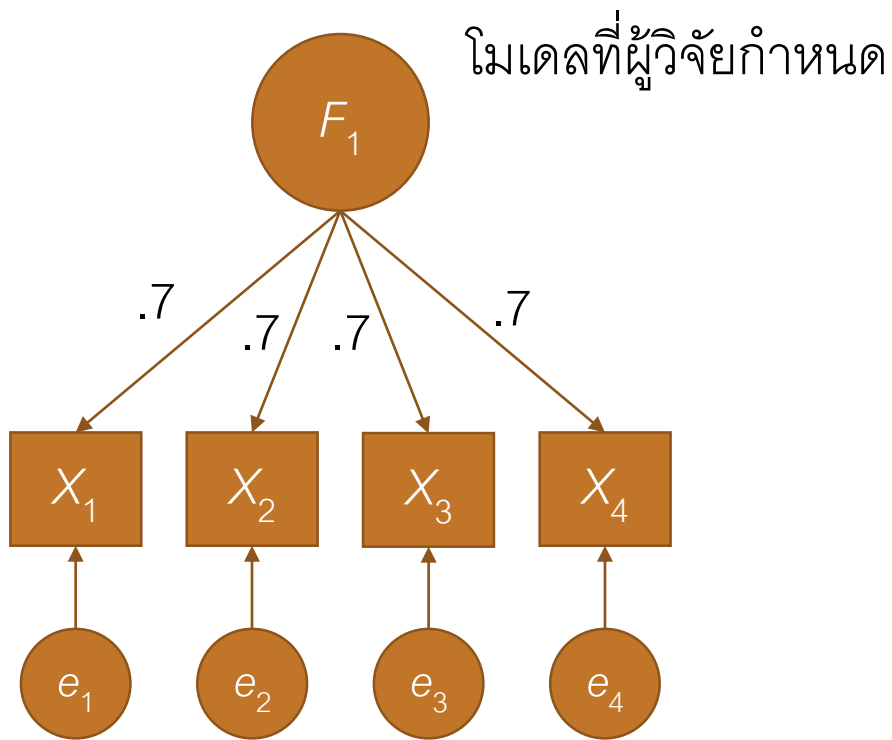
- โมเดลการวิเคราะห์องค์ประกอบ คือ การอธิบายว่าทำไมคนถึงตอบข้อความได้คะแนนต่างๆ
- ก่อนอื่น ต้องเข้าใจคำว่าโมเดล (Model) ก่อน ว่าโมเดล คือ รูปแบบจำลองที่สามารถอธิบายด้วยสมการคณิตศาสตร์ (หรือวิธีอื่น) ที่มุ่งหวังที่จะอธิบายปรากฏการณ์ที่เกิดขึ้น ด้วยความซับซ้อนที่น้อยที่สุด
- ถ้าอธิบายปรากฏการณ์ทั้งหมด (ซึ่งจริงๆ แล้วอาจเป็นไปได้เลยด้วยซ้ำ) จะทำให้การอธิบายซับซ้อนมาก
- ดังนั้นสิ่งที่โมเดลแสดงออกมา คือ สิ่งที่ซับซ้อนน้อยที่สุดที่ทำให้อธิบายสิ่งที่เกิดขึ้นได้มากที่สุด โมเดลจึงเป็นการ “ประมาณ” สิ่งที่เกิดขึ้นในธรรมชาติ

การเลือกจำนวนองค์ประกอบ

- ด้วยเหตุนี้ จำนวนองค์ประกอบที่แท้จริงจึงมีจำนวนเยอะนับไม่ถ้วน แต่การวิเคราะห์องค์ประกอบเป็นเพียงหาองค์ประกอบที่สำคัญออกมาอธิบายกระบวนการเกิดข้อมูลต่างๆ
- นอกจากนี้ องค์ประกอบที่สกัดออกมา จะต้องเป็นสิ่งที่เกิดขึ้น ไม่ว่าจะวิเคราะห์ด้วยกลุ่มตัวอย่างใดก็ตาม
- ผมอยากให้เห็นคำว่า “องค์ประกอบที่สำคัญ” และ “องค์ประกอบที่สามารถเกิดซ้ำในกลุ่มตัวอย่างอื่นได้”
- คำว่า “สำคัญ” เป็นสิ่งค่อนข้างจะเป็นอัตนัย (Subjective) กล่าวคือ ผู้วิจัยแต่ละคนอาจให้คำนิยามคำว่าสำคัญแตกต่างกัน

การเลือกจำนวนองค์ประกอบ

- อย่างไรก็ตาม นักวิจัยก็ควรจะเป็นคนตัดสินใจจำนวนองค์ประกอบเอง เพราะเป็นคนรู้จักข้อคำถามที่ดีที่สุด
- นอกจากนี้ นักวิจัยต้องควรเลือกเฉพาะองค์ประกอบที่สามารถเกิดซ้ำในกลุ่มตัวอย่างอื่นได้ หากหลักฐานการเกิดขึ้นขององค์ประกอบดังกล่าวไม่ชัดเจน ก็ควรทบทวนการสร้างข้อคำถาม หรือเก็บข้อมูลใหม่



ในการสร้างโมเดล ผู้วิจัยไม่ใช่เรียงกับองค์ประกอบเล็กน้อย ด้านซ้ายจึงเป็นโมเดลที่โอเค

คาบต่อไป

- ส่ง Project 1 (ส่งใน 28 กุมภาพันธ์ เวลา 7:00 น.)