

# ความเที่ยง

การประเมินลักษณะมนุษย์

สันทัต พรประเสริฐมานิต

# โครงร่างการนำเสนอ

- สูตรความแปรปรวนของตัวแปรสุ่ม
- นิยามความเที่ยง
- รูปแบบของความเที่ยง
- ปัจจัยที่มีผลต่อความเที่ยง
- ความผิดพลาดในการวัด

# สูตรความแปรปรวนของตัวแปรสุ่ม

- สูตรของความแปรปรวน

$$\text{Var}(cX) = c^2\text{Var}(X)$$

$$\text{Var}(X + c) = \text{Var}(X)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Cov}(cX, Y) = c\text{Cov}(X, Y)$$

$$\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$$

$$\text{Cov}(X + Y, U + V) = \text{Cov}(X, U) + \text{Cov}(X, V) + \text{Cov}(Y, U) + \text{Cov}(Y, V)$$

# นิยามความเที่ยง

- ทฤษฎีการวัดแบบคลาสสิก (Classical test theory)

$$O = T + E$$

- $O$  = คะแนนที่สังเกตได้ (Observed score)
- $T$  = คะแนนที่แท้จริง (True score)
- $E$  = ความผิดพลาดในการวัด (Measurement error)
  - ~~แบบเป็นระบบ (Systemic error)~~
  - แบบสุ่ม (Random error)

ตามทฤษฎี

1.  $T$  และ  $E$  ไม่สัมพันธ์กัน
2.  $E_1$  และ  $E_2$  ไม่สัมพันธ์กัน
3. ค่าเฉลี่ยของ  $E$  เท่ากับ 0

# นิยามความเที่ยง

- สมการการวัด

$T$  และ  $E$  ไม่สัมพันธ์กัน

$$\text{Var}(O) = \text{Var}(T + E) = \text{Var}(T) + \text{Var}(E)$$

$$\sigma_O^2 = \sigma_T^2 + \sigma_E^2$$

- $\sigma_O^2$  = ความแปรปรวนของคะแนน (Observed variance)
- $\sigma_T^2$  = ความแปรปรวนที่แท้จริง (True variance)
- $\sigma_E^2$  = ความแปรปรวนจากคะแนนผิดพลาด (Error variance)

# นิยามความเที่ยง

- ความเที่ยง คือ สัดส่วนของคะแนนที่แท้จริงต่อคะแนนทั้งหมด

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_O^2}$$

- กล่าวคือ คะแนนที่สังเกตได้นั้นเป็นคะแนนส่วนที่ไม่ได้เกิดจากการสุ่มอยู่ใน สัดส่วนเท่าไร

# รูปแบบของความเที่ยง

- วัดสองครั้งด้วยเวลาที่ห่างกัน ความผิดพลาดแบบสุ่มทำให้คะแนนทั้งสองมีค่าไม่เหมือนกัน
- ความเที่ยงแบบวัดซ้ำ (Test-retest reliability)
- ความสัมพันธ์ระหว่างคะแนนวัดซ้ำสองครั้ง

$$O_1 = T + E_1 \qquad O_2 = T + E_2$$

$$Cor(O_1, O_2) = \frac{Cov(O_1, O_2)}{\sqrt{Var(O_1)Var(O_2)}} = \frac{Var(T)}{Var(O)} = \rho_{XX}$$

# รูปแบบของความเที่ยง

- วัดสองครั้งด้วยผู้ประเมินที่แตกต่างกัน ความผิดพลาดแบบสุ่มทำให้คะแนนทั้งสองมีค่าไม่เหมือนกัน
- ความเที่ยงระหว่างผู้ประเมิน (Interrater reliability)
- ความสัมพันธ์ระหว่างคะแนนวัดจากผู้ประเมินสองคน

$$O_1 = T + E_1 \qquad O_2 = T + E_2$$

$$Cor(O_1, O_2) = \frac{Cov(O_1, O_2)}{\sqrt{Var(O_1)Var(O_2)}} = \frac{Var(T)}{Var(O)} = \rho_{XX}$$



# รูปแบบของความเที่ยง

- วัดสองครั้งด้วยแบบทดสอบที่ใกล้เคียงกัน (Alternate form) มีความผิดพลาดแบบสุ่มที่ทำให้คะแนนจากแบบทดสอบทั้งสองไม่เหมือนกัน
- ความเที่ยงจากแบบทดสอบใกล้เคียง (Alternate-form reliability)
- ความสัมพันธ์ระหว่างคะแนนแบบทดสอบทั้งสอง

$$O_1 = T + E_1 \qquad O_2 = T + E_2$$

$$Cor(O_1, O_2) = \frac{Cov(O_1, O_2)}{\sqrt{Var(O_1)Var(O_2)}} = \frac{Var(T)}{Var(O)} = \rho_{XX}$$

# รูปแบบของความเที่ยง

- หากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบใกล้เคียงเท่ากัน อาจเรียกว่าแบบทดสอบคู่ขนาน (Parallel form)
- วิธีการหนึ่งในการสร้างแบบทดสอบคู่ขนาน คือ การแบ่งครึ่งแบบทดสอบเป็นสองส่วน (Split half) – **ผมไม่เน้น ไม่มีใครใช้อีกแล้ว**
- Spearman-Brown formula เป็นสูตรที่เดิมใช้ในการแบ่งครึ่งแบบทดสอบ แต่สามารถประยุกต์ใช้ได้ว่าหากเพิ่มจำนวนข้อคำถามคู่ขนานเป็น  $m$  เท่าของข้อสอบเดิมแล้ว ความเที่ยงเป็นเท่าไร

# รูปแบบของความเที่ยง

- โดยส่วนใหญ่แล้ว แบบทดสอบมีข้อคำถามหลายข้อ แล้วใช้คะแนนรวม
- แบบทดสอบที่ดี ข้อคำถามควรสัมพันธ์กัน หากข้อคำถามไม่สัมพันธ์กัน แสดงว่ามีความผิดพลาดในการวัดในข้อคำถามแต่ละข้อสูง

$$O_1 = T + E_1$$

$$O_2 = T + E_2$$

⋮

$$O_p = T + E_p$$

$$Y = O_1 + O_2 + \cdots + O_p$$

$$T_Y = T + T + \cdots + T$$

$$\rho_{YY} = \frac{\sigma_{T_Y}^2}{\sigma_Y^2}$$

$$\text{Cov}(O_i, O_j) = \text{Cov}(T + E_i, T + E_j) = \text{Var}(T); \quad i \neq j$$

$\text{Var}(Y) =$  ผลรวมของสมาชิกทุกตัวใน

$$\begin{bmatrix} \text{Var}(O_1) & \text{Cov}(O_1, O_2) & \dots & \text{Cov}(O_1, O_p) \\ \text{Cov}(O_2, O_1) & \text{Var}(O_2) & \dots & \text{Cov}(O_2, O_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(O_p, O_1) & \text{Cov}(O_p, O_2) & \dots & \text{Var}(O_p) \end{bmatrix}$$

$\text{Var}(T)$  จึงเป็นค่าเฉลี่ยของค่าความแปรปรวนร่วมเหล่านี้

$$\begin{aligned} \text{Var}(T) &= \frac{\sum_{i=1}^p \sum_{j=1}^{i-1} \text{Cov}(O_i, O_j)}{\frac{p(p-1)}{2}} = \frac{\text{Var}(Y) - \sum_{i=1}^p \text{Var}(O_i)}{\frac{p(p-1)}{2}} \\ &= \text{Var}(Y) \frac{1 - \frac{\sum_{i=1}^p \text{Var}(O_i)}{\text{Var}(Y)}}{p(p-1)} \end{aligned}$$

$$\text{Var}(T_Y) = \text{Var}(pT) = p^2 \text{Var}(T)$$

$$\rho_{YY} = \frac{\text{Var}(T_Y)}{\text{Var}(Y)} = p^2 \text{Var}(Y) \frac{1 - \frac{\sum_{i=1}^p \text{Var}(O_i)}{\text{Var}(Y)}}{p(p-1)}}{\text{Var}(Y)}$$

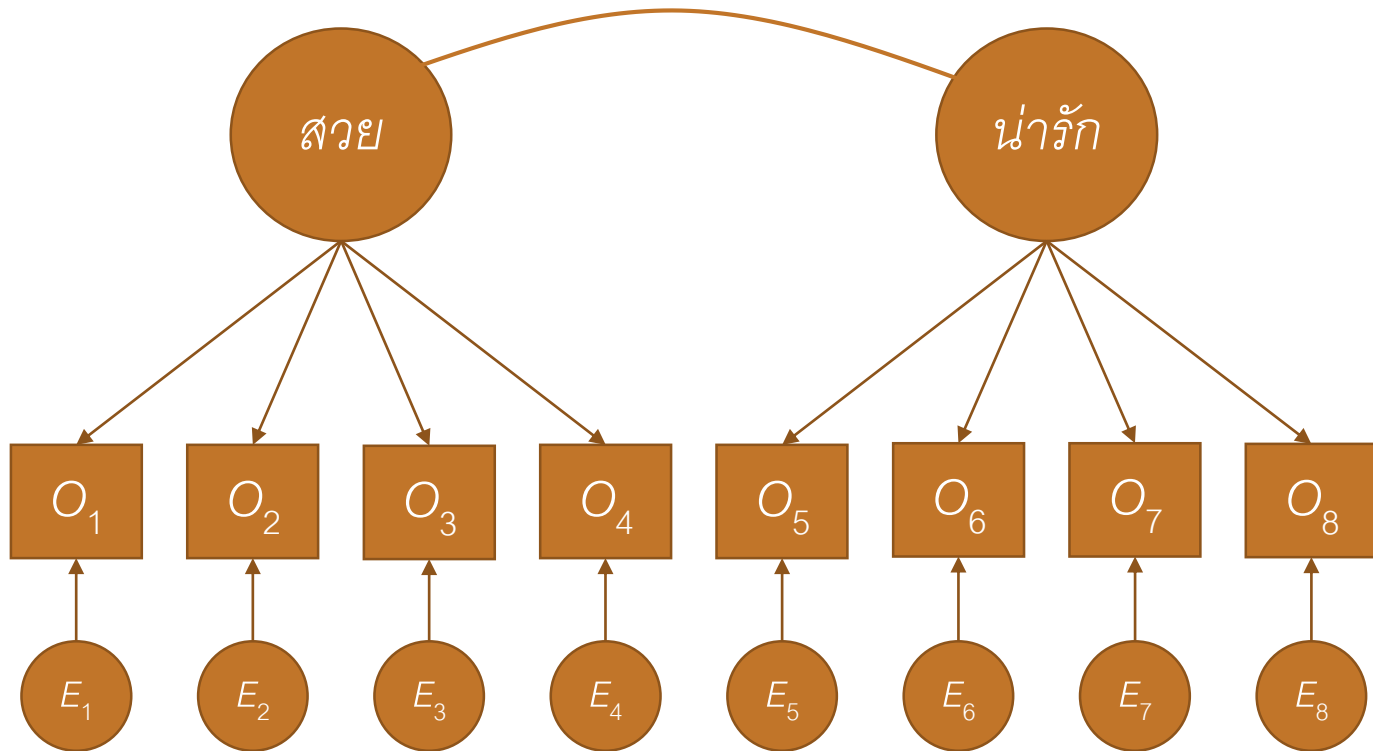
$$\rho_{YY} = \alpha = \left( \frac{p}{p-1} \right) \left( 1 - \frac{\sum_{i=1}^p \text{Var}(O_i)}{\text{Var}(Y)} \right)$$

# รูปแบบของความเที่ยง

- ค่าอัลฟา (alpha, Cronbach's alpha) จึงเป็นค่าที่วัดความสอดคล้องภายใน (Internal consistency) ระหว่างข้อคำถาม
- ยิ่งค่าอัลฟามีค่าสูง แสดงว่าข้อคำถามมีความสัมพันธ์กันสูง
- มักได้ยืนยันข้อเสนอแนะว่า ระดับ .7, .8, และ .9 เท่ากับพอใช้, ดี, และดีมากตามลำดับ
- แต่ค่าอัลฟาไม่ได้บอกว่าข้อคำถามทั้งหมดมาจากมิติเดียวกัน (Unidimensionality) กล่าวคือ ข้อคำถามทุกข้อที่มีคะแนนขึ้นลงนั้น อาจไม่ได้เกิดจากเหตุผลเดียวกัน

# รูปแบบของความเที่ยง

- ลองดูมาตรวัดนี้



# รูปแบบของความเที่ยง

- ได้คะแนนความสัมพันธ์แบบนี้

	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

$$Y = O_1 + O_2 + \dots + O_8$$

$$\alpha_Y = .87$$



# รูปแบบของความเที่ยง

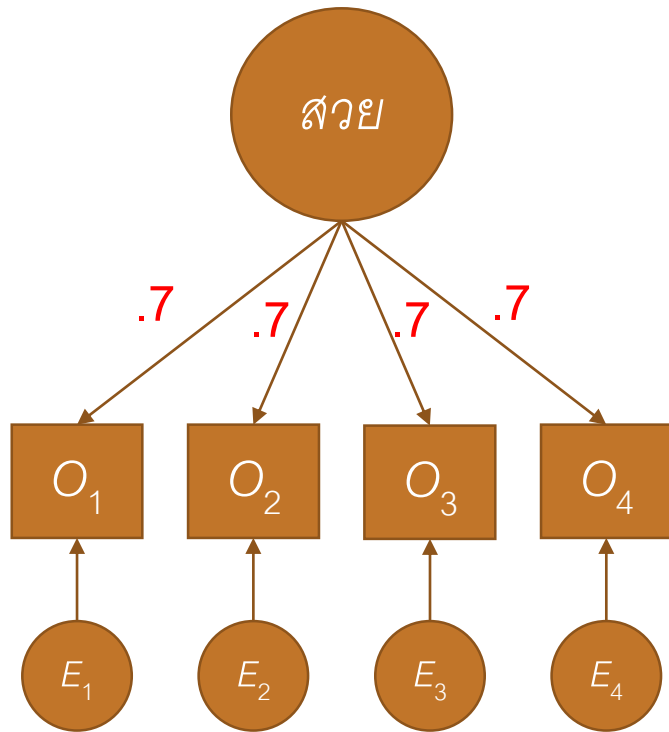
- ค่าอัลฟ่าจะบอกความเที่ยงก็ต่อเมื่อมั่นใจแล้วว่ามาตรวัดนั้นมีมิติเดียว
- นอกจากนี้ ค่าอัลฟ่าใช้ค่าเฉลี่ยของความแปรปรวนร่วมในการประมาณค่า  $Var(T)$

$$\begin{bmatrix} Var(O_1) & Cov(O_1, O_2) & \dots & Cov(O_1, O_p) \\ Cov(O_2, O_1) & Var(O_2) & \dots & Cov(O_2, O_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(O_p, O_1) & Cov(O_p, O_2) & \dots & Var(O_p) \end{bmatrix}$$

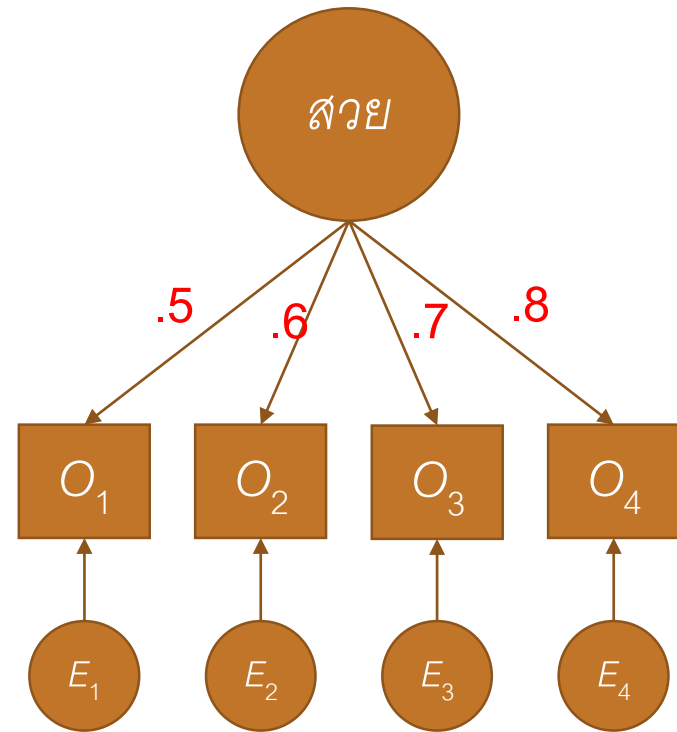
การประมาณนี้จะถูกต้องก็ต่อเมื่อ  
ข้อคำถามทุกข้อมีความสามารถ  
ในการวัดภาวะสันนิษฐาน  
เท่าเทียมกัน หรือที่เรียกว่า  
ความไวใกล้เคียง (Tau-equivalence)

$Var(T)$  จึงเป็นค่าเฉลี่ยของค่าความแปรปรวนร่วมเหล่านี้

# รูปแบบของความเที่ยง



ความไวใกล้เคียงกัน



ความไวไม่ใกล้เคียงกัน

# รูปแบบของความเที่ยง

- ข้อตกลงเบื้องต้นที่ว่าความไวใกล้เคียงกัน แทบเป็นไปไม่ได้ในความเป็นจริง
  - เช่น คำถามว่า “คุณเป็นคนสวย” กับ “คุณเป็นคนมีเสน่ห์” กับ “คุณเป็นคนหน้าตาดี” สามประโยคนี้อันสะท้อนถึงความสวยของคนแตกต่างกัน
  - ทำให้ความไวไม่เท่ากัน
- ดังนั้น ค่าอัลฟาโดยส่วนใหญ่จึงไม่ได้วัดความเที่ยง
- ค่าสัมประสิทธิ์โอเมก้า (Coefficient omega) สำหรับมาตรวัดมิติเดียว จะสะท้อนให้เห็นถึงความเที่ยงของคะแนนรวมได้ดีกว่า ซึ่งจะกล่าวในบทการวิเคราะห์องค์ประกอบ

# มาตรวัดมิติเดียว

- ลักษณะของข้อคำถามว่าสอดคล้องกัน (Homogeneous) หรือแตกต่างกัน (Heterogeneous) เป็นคุณสมบัติที่ต้องคิดตั้งแต่ต้น
- มาตรวัดที่ดี ไม่จำเป็น ต้องมีความเที่ยงสูง
- เช่น มาตรวัดความสุขในชีวิต แต่ถามเรื่องงานเพียงอย่างเดียว จะทำให้ความเที่ยงสูง แต่ไม่ครอบคลุมนิยามความสุขในชีวิตทั้งหมด
- หลักการที่ดี คือ ครอบคลุมส่วนย่อยทั้งหมดของนิยาม และข้อคำถามในแต่ละส่วนย่อยมีความสอดคล้องกัน

# ปัจจัยที่มีต่อความเที่ยง

- การจำกัดพิสัย (Restriction of range)
- กลุ่มตัวอย่าง
- ลักษณะของภาวะสันนิษฐาน
- ผู้ดำเนินการทดสอบ
- ปัจจัยอื่นๆ ที่ก่อให้เกิดความผิดพลาดแบบสุ่ม

# ค่าความผิดพลาดในการวัด

- ความผิดพลาดในการวัดมาตรฐาน (Standard error of measurement; SEM) คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนความผิดพลาดในการวัด

$$\begin{aligned} SEM &= \sqrt{\sigma_E^2} = \sqrt{\sigma_X^2 - \sigma_T^2} = \sqrt{\sigma_X^2 - \rho_{XX}\sigma_X^2} \\ &= \sigma_X \sqrt{1 - \rho_{XX}} \end{aligned}$$

- ค่าสังเกตได้ (Observed score) จะเป็นการประมาณค่าแบบจุดที่ดีที่สุดของคะแนนที่แท้จริง (True score)

# ค่าความผิดพลาดในการวัด

- แต่การประมาณค่าที่แท้จริงที่ดี ควรเป็นการประมาณค่าแบบช่วงผ่านช่วงเชื่อมั่น (Confidence interval)

$$CI_{1-\alpha}(T) = (O - z_{\alpha/2}SEM, O + z_{\alpha/2}SEM)$$

# ค่าความผิดพลาดในการวัด

- ทฤษฎีความสามารถในการแผ่ขยาย (Generalizability theory)
  - คำนึงถึงในกรณีที่คะแนนเฉลี่ยหนึ่ง ถูกคำนวณจากหลายข้อคำถาม หลายผู้ประเมิน หลายช่วงเวลา
  - คำนวณความเที่ยงทั้งมาตรฐานแบบอิงกลุ่มและอิงเกณฑ์
- ทฤษฎีการตอบสนองของข้อคำถาม
  - ความเที่ยง หรือความผิดพลาดในการวัดมาตรฐาน ขึ้นอยู่กับช่วงของคะแนนในภาวะสันนิษฐาน
  - เช่น ข้อสอบง่าย จะจำแนกคนได้ดี (ความผิดพลาดน้อย) ในช่วงคนที่มีความสามารถน้อยและน้อยมาก



# คาบต่อไป

- ใบงานที่ 2
  - เขียนข้อความถาม
  - เก็บข้อมูลนำร่อง
- อ่าน Cohen et al บทที่ 6
- ทำแบบฝึกหัดในสัมมนาการใช้ R เบื้องต้น
  - <http://www.sunthud.com/teaching.html>