ความเที่ยง

การประเมินลักษณะมนุษย์

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- สูตรความแปรปรวนของตัวแปรสุ่ม
- นิยามความเที่ยง
- รูปแบบของความเที่ยง
- ปัจจัยที่มีผลต่อความเที่ยง
- ความผิดพลาดในการวัด

ตัวแปรสุ่ม

- ตัวแปรสุ่ม (Random Variable)
  - เหตุการณ์ที่แตกต่างกัน ทำให้เกิดค่าที่แตกต่างกัน
  - สามารถแปลงค่าเป็นตัวเลขได้
  - ระดับการวัดแบบอันตรภาค (Interval) ขึ้นไป หรือเป็นตัวแปรดัมมี่ (Dummy Variable)
  - เช่น ส่วนสูง เกิดแผ่นดินไหวหรือไม่ ผมสีดำหรือไม่ สัดส่วนเพศหญิง ละติจูด ฯลฯ
- ให้ X, Y เป็นตัวแปรสุ่ม
  - $E(X)$ แทนค่าคาดหวัง (ค่าเฉลี่ย) ของตัวแปร X
  - $\text{Var}(X)$ แทนความแปรปรวนของตัวแปร X
  - $\text{Cov}(X, Y)$ แทนความแปรปรวนร่วมของตัวแปร X, Y

สูตรความแปรปรวนของตัวแปรสุ่ม

- ให้ X, Y, U, V เป็นตัวแปรสุ่ม และ c เป็นค่าคงที่

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

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

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

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

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

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

ความเที่ยง

- คะแนนของแต่ละคน ไม่มีทางที่จะเหมือนกันได้ตลอดเวลา เช่น ถามว่า

	ไม่จริง	ค่อนข้างไม่จริง	ค่อนข้างจริง	จริง
ฉันมีความสุข			X	

- ความเที่ยงแสดงถึงความคงเส้นคงวาของคะแนน โดยเชื่อว่าคะแนนเกิดจากคะแนนที่แท้จริงภายในแต่ละบุคคล บวกกับความผิดพลาดโดยสุ่ม ที่เกิดขึ้นได้เสมอ ถ้าเที่ยง ความผิดพลาดโดยสุ่มจะน้อย

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

- ทฤษฎีการวัดแบบคลาสสิก (Classical test theory)

$$O = T + E$$

- นักวิจัยอยากทราบว่า คะแนนที่สังเกตได้ (*O*) มีค่าความผิดพลาดในการวัด (*E*) มาเกี่ยวข้องมากน้อยเพียงใด
  - ถ้าเกี่ยวข้องน้อย แสดงว่าคะแนนที่สังเกตได้น่าเชื่อถือมาก
  - ถ้าเกี่ยวข้องมาก แสดงว่าคะแนนที่สังเกตได้ไม่ค่อยน่าเชื่อถือ เพราะไม่ได้สะท้อนค่าคะแนนที่แท้จริง (*T*)

นิยามความเที่ยง

- สมการการวัด

T และ *E* ไม่สัมพันธ์กัน

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

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

- σ_O^2 = ความแปรปรวนของคะแนน (Observed variance)
- σ_T^2 = ความแปรปรวนที่แท้จริง (True variance)
- σ_E^2 = ความแปรปรวนจากคะแนนผิดพลาด (Error variance)

นิยามความเที่ยง

- ความเที่ยง คือ สัดส่วนของคะแนนที่แท้จริงต่อคะแนนทั้งหมด

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_O^2} = 1 - \frac{\sigma_E^2}{\sigma_O^2}$$

- กล่าวคือ คะแนนที่สังเกตได้นั้นเป็นคะแนนส่วนที่ไม่ได้เกิดจากการสุ่มอยู่ในสัดส่วนเท่าไร
- ถ้าให้ σ_O^2 คือ ความแปรปรวนของค่าคะแนนที่เราใช้ (ซึ่งวัดได้) จะหาความเที่ยงของ O สามารถทำได้โดยการหา σ_E^2 ซึ่งเราไม่มีทางหาค่า *E* ได้ สไลด์ถัดไป จะถามว่าพอมีวิธีอะไรบ้าง ที่ประมาณค่า σ_E^2 ได้

รูปแบบของความเที่ยง

- 1. ให้ประมาณค่า *E* ด้วยการวัด *O* ตัวเดียวกันสองครั้ง ด้วยเวลาที่ห่างกัน เชื่อว่า *T* ไม่เปลี่ยน แต่ *E* ทำให้คะแนนทั้งสองมีค่าไม่เหมือนกัน
- ความสัมพันธ์ระหว่างคะแนนวัดซ้ำสองครั้งเป็นความเที่ยง เรียกว่า ความเที่ยงแบบวัดซ้ำ (Test-retest reliability)

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

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

รูปแบบของความเที่ยง

- 2. ให้ประมาณค่า *E* ด้วยการวัด *O* ตัวเดียวกันสองครั้ง ด้วยผู้ประเมิน (Rater) คนละคนกัน เชื่อว่า *T* ไม่เปลี่ยน แต่ *E* ทำให้คะแนนทั้งสองมีค่าไม่เหมือนกัน
- ความสัมพันธ์ระหว่างคะแนนวัดจากผู้ประเมินสองคน เรียกว่า ความเที่ยงระหว่างผู้ประเมิน (Interrater reliability)

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

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

รูปแบบของความเที่ยง

- ให้ประมาณค่า *E* ด้วยการวัด *O* ตัวเดียวกันสองครั้ง ด้วยแบบวัดคู่ขนาน (Parallel form) คนละแบบวัด เชื่อว่า *T* ไม่เปลี่ยน แต่ *E* ทำให้คะแนนทั้งสองมีค่าไม่เหมือนกัน
- ความสัมพันธ์ระหว่างคะแนนแบบทดสอบทั้งสอง เรียกว่า ความเที่ยงจากแบบทดสอบใกล้เคียง (Parallel-form reliability)

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

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

รูปแบบของความเที่ยง

- ตัวอย่างแบบทดสอบคู่ขนาน
  - การทดสอบความจำกลับด้าน (Backward Memory Span)
  - การทดสอบผลสัมฤทธิ์เข้ามหาวิทยาลัย
  - การสอบ TOEFL, TOEIC
- แบบทดสอบคู่ขนาน (Parallel form) อาจถูกเรียกว่าแบบทดสอบใกล้เคียง (Equivalent form, Alternate form) บางตำราจะบอกว่าแบบทดสอบคู่ขนานมีลักษณะทางสถิติเหนือกว่า

รูปแบบของความเที่ยง

- วิธีการหนึ่งในการสร้างแบบทดสอบคู่ขนาน คือ การแบ่งครึ่งแบบทดสอบเป็นสองส่วน (Split half)
- ทำให้สามารถประมาณค่าความเที่ยงได้ โดยเก็บข้อมูลครั้งเดียว โดยหาความสัมพันธ์ระหว่างแต่ละครึ่งแบบทดสอบ แล้วมาปรับค่าให้เป็นความสัมพันธ์ของเต็มฉบับ ด้วย Spearman-Brown formula
- วิธีนี้ไม่มีใครใช้แล้ว เพราะ ไม่แน่ใจว่าจะแบ่งครึ่งอย่างไร (ข้อคู่ คี่?) ให้เป็นแบบทดสอบคู่ขนานจริงๆ

รูปแบบของความเที่ยง

- 4. ใช้ความสัมพันธ์ระหว่างข้อคำถามย่อย ที่นำมารวมกันเป็น *O* ในการประมาณค่าความเที่ยง
- ให้ $$Y = O_1 + O_2 + \cdots + O_p$$
- ถ้าให้ทุกข้อย่อย O_1, O_2, ... O_p วัด *T* ตัวเดียวกันแล้ว สิ่งที่ทำให้ข้อย่อยแต่ละข้อแตกต่างกัน คือ ค่า *E* ของแต่ละข้อย่อยที่แตกต่างกัน

$$O_1 = T + E_1 \quad O_2 = T + E_2 \quad \cdots \quad O_p = T + E_p$$

รูปแบบของความเที่ยง

- จุดมุ่งหมายในวิธีนี้ คือ ต้องการหาความเที่ยงของคะแนนรวม Y ผ่านข้อมูลของข้อคำถามย่อย O_1, O_2, ... O_p ซึ่งสามารถหาได้ดังนี้

$$O_1 = T + E_1$$
$$O_2 = T + E_2$$
$$\vdots$$
$$O_p = T + E_p$$

$$Y = O_1 + O_2 + \cdots + O_p$$
$$T_Y = T + T + \cdots + T$$
$$\rho_{YY} = \frac{\sigma^2_{T_Y}}{\sigma^2_Y}$$

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

$Var(Y) =$ ผลรวมของสมาชิกทุกตัวใน

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

$Var(T)$ จึงเป็นค่าเฉลี่ยของค่าความแปรปรวนร่วมเหล่านี้

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

$$= Var(Y) \frac{1 - \frac{\sum_{i=1}^{p} Var(O_i)}{Var(Y)}}{p(p-1)}$$

$$Var(T_Y) = Var(pT) = p^2 Var(T)$$

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

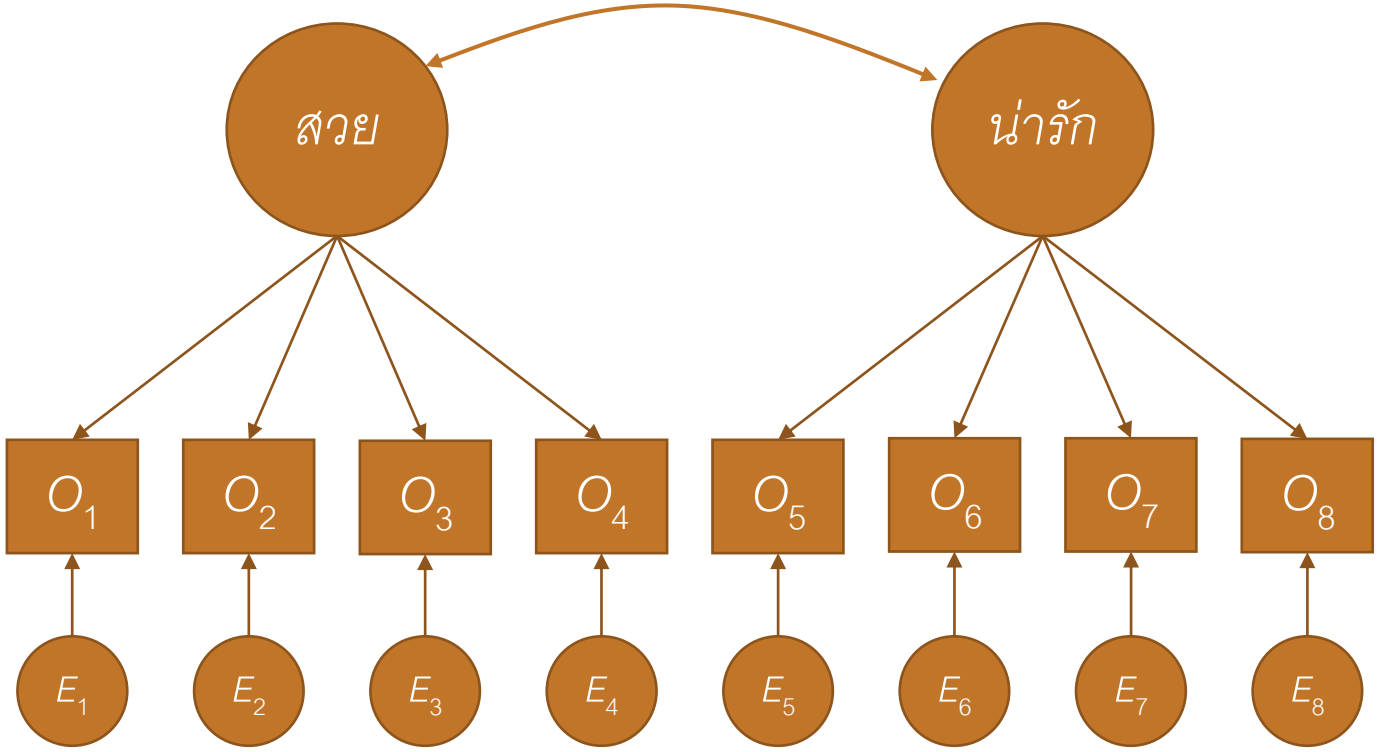
$$\rho_{YY} = \alpha = \left(\frac{p}{p-1}\right)\left(1 - \frac{\sum_{i=1}^{p} Var(O_i)}{Var(Y)}\right)$$

รูปแบบของความเที่ยง

- ค่าอัลฟ่า (alpha, Cronbach's alpha) จึงเป็นค่าที่วัดความสอดคล้องภายใน (Internal consistency) ระหว่างข้อคำถาม
- ยิ่งค่าอัลฟ่ามีค่าสูง แสดงว่าข้อคำถามมีความสัมพันธ์กันสูง เพราะแสดงว่า $Cov(O_i, O_j)$ มีค่าสูง
- มักได้ยินข้อแนะนำว่า ระดับ .7, .8, และ .9 เท่ากับพอใช้, ดี, และดีมาก ตามลำดับ
- แต่ค่าอัลฟ่าไม่ได้บอกว่าข้อคำถามทั้งหมดมาจากมิติเดียวกัน (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

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

$$\alpha_Y = .87$$

รูปแบบของความเที่ยง

- ค่าอัลฟ่าจะบอกความเที่ยงก็ต่อเมื่อมั่นใจแล้วว่ามาตรวัดนั้นมีมิติเดียว
- นอกจากนี้ ค่าอัลฟ่าใช้ค่าเฉลี่ยของความแปรปรวนร่วมในการประมาณค่า Var(T)

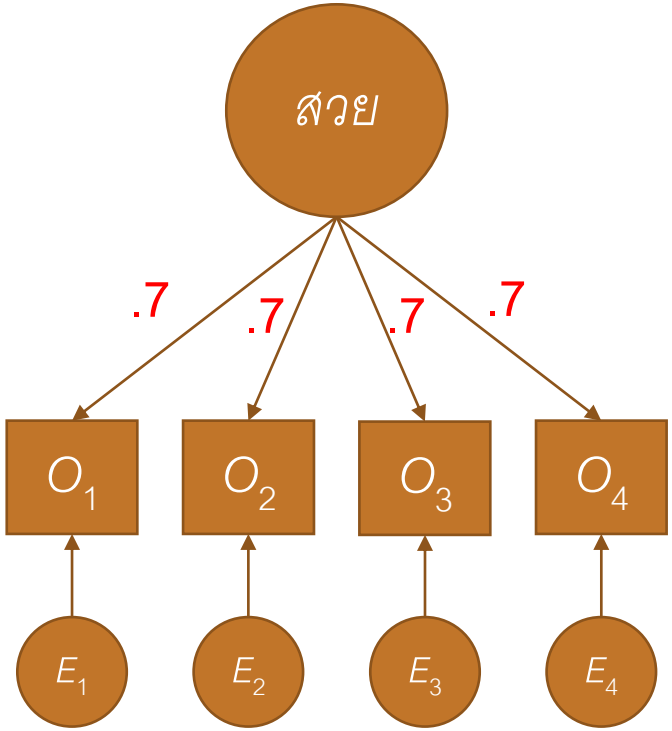
$$\begin{bmatrix} Var(O_1) & Cov(O_1,O_2) & \dots & Cov(O_1,O_p) \\ Cov(O_2,O_1) & Var(O_2) & \dots & Cov(O_2,O_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(O_p,O_1) & Cov(O_p,O_2) & \dots & Var(O_p) \end{bmatrix}$$

$Var(T)$ จึงเป็นค่าเฉลี่ยของค่าความแปรปรวนร่วมเหล่านี้

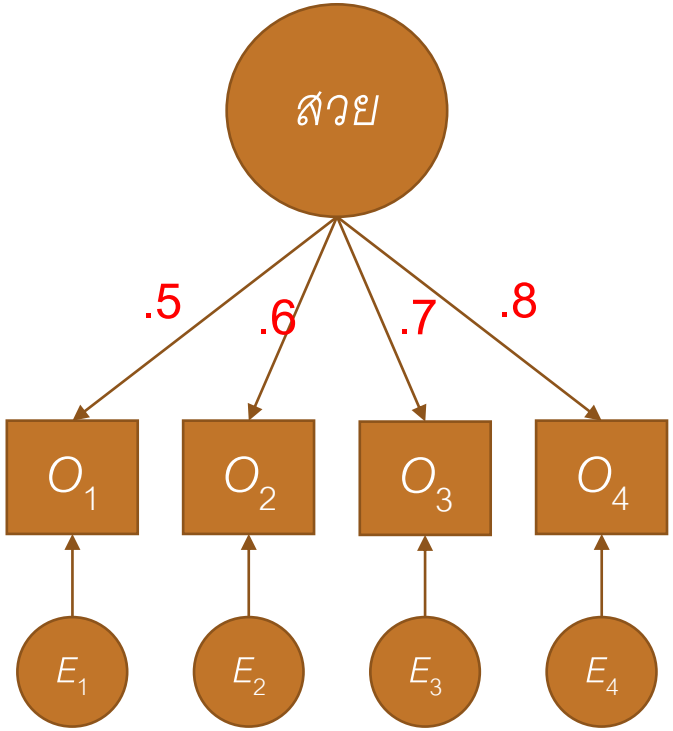
การประมาณนี้จะถูกต้องก็ต่อเมื่อ $Cov(O_p, O_{p'})$ มีค่าใกล้เคียงกัน (ถึงจะโอเคที่นำมาเฉลี่ยกัน เพราะประมาณค่า Var(T) ตัวเดียวกัน)

คุณสมบัตินี้เรียกว่า ข้อคำถามทุกข้อมีความสามารถในการวัดภาวะสันนิษฐานเท่าเทียมกัน หรือที่เรียกว่า ความไวใกล้เคียง (Tau-equivalence)

รูปแบบของความเที่ยง



ความไวใกล้เคียงกัน



ความไวไม่ใกล้เคียงกัน

รูปแบบของความเที่ยง

- ข้อตกลงเบื้องต้นที่ว่าความไวใกล้เคียงกัน แทบเป็นไปไม่ได้ในความเป็นจริง
  - เช่น คำถามว่า “คุณเป็นคนสวย” กับ “คุณเป็นคนมีเสน่ห์” กับ “คุณเป็นคนหน้าตาดี” สามประโยคนี้สะท้อนถึงความสวยของคนแตกต่างกัน
  - ทำให้ความไวไม่เท่ากัน
- ดังนั้น ค่าอัลฟ่าโดยส่วนใหญ่จึงไม่ได้วัดความเที่ยง
- ค่าสัมประสิทธิ์โอเมก้า (Coefficient omega) สำหรับมาตรวัดมิติเดียว จะสะท้อนให้เห็นถึงความเที่ยงของคะแนนรวมได้ดีกว่า
  - ใช้หลักการของการวิเคราะห์องค์ประกอบ (Factor Analysis) โดยกำหนดให้มีมิติเดียว ใช้ R คำนวณได้

รูปแบบของความเที่ยง

- ถ้าคุณมั่นใจว่าข้อคำถามทั้งหมดวัดมิติเดียวกันจริง แต่ความไวของข้อคำถามแตกต่างกัน ค่าอัลฟ่าจะมีค่าน้อยกว่าค่าโอเมก้าเสมอ
- แสดงว่า ถ้าค่าอัลฟ่ามีค่าสูง มีความเที่ยงดี ก็แสดงว่ามาตรนี้น่าจะมีความเที่ยงดีด้วย

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

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

- ค่าสังเกตได้ (Observed score) จะเป็นการประมาณค่าแบบจุดที่ดีที่สุดของคะแนนที่แท้จริง (True score)

ค่าความผิดพลาดในการวัด

- แต่การประมาณค่าที่แท้จริงที่ดี ควรเป็นการประมาณค่าแบบช่วงผ่านช่วงเชื่อมั่น (Confidence interval)

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

ค่าความผิดพลาดในการวัด

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