ความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไข (Conditional Standard Error of Measurement : CSEM)

รศ.ดร.สุทธิวรรณ พีรศักดิ์โสภณ*

บทนำ

ในการสอบวัดแต่ละครั้งคะแนนที่ได้จากการสอบจะแสดงถึงความรู้ความสามารถหรือพฤติกรรมของบุคคลนั้น ๆ ว่ามีอยู่เท่าใด ซึ่งคะแนนนี้จะบ่งบอกถึงคะแนนจริงได้มากน้อยเพียงใดย่อมขึ้นอยู่กับความคลาดเคลื่อนในการวัด (Error of measurement) ถ้าความคลาดเคลื่อนในการวัดมีมาก แสดงว่าคะแนนที่ได้จากการวัด (Observed score) จะแตกต่างจากคะแนนจริง (True score) มาก

ความคลาดเคลื่อนในการวัดมีผลมาจากองค์ประกอบหลาย ๆ องค์ประกอบ เช่น สภาพเงื่อนไขของการทดสอบ การดำเนินการสอบ เครื่องมือที่ใช้ทดสอบ ผู้เข้าสอบ โดยปกติความคลาดเคลื่อนหรือความไม่คงที่ของคะแนนนี้ แม้เราจะรู้ว่ามีอยู่จริง แต่เราก็ไม่สามารถบอกได้ว่าคะแนนที่สอบได้ของแต่ละคนมีความผิดพลาดเท่าใด แต่เราสามารถอาศัยวิธีการทางสถิติเพื่อประมาณความผิดพลาดโดยเฉลี่ยของคน ๆ นั้นว่าเป็นอย่างไรได้ ซึ่งเราเรียกว่า ความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurement)

---

* รองศาสตราจารย์ประจำภาควิชาวิจัยและพัฒนาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

คะแนนความคลาดเคลื่อนมาตรฐานของการวัดมีความสำคัญมากต่อการนำเสนอเกี่ยวกับค่าสถิติของแบบทดสอบ ดังนั้นในการรายงานผลเกี่ยวกับการวัดจึงควรนำเสนอค่าความคลาดเคลื่อนมาตรฐานในการวัดไว้ด้วยเพื่อจะช่วยทำให้การแปลความหมายของคะแนนมีความแม่นยำและชัดเจนขึ้น

โดยปกติเมื่อแปลความหมายของความคลาดเคลื่อนมาตรฐานในการวัด เรามักจะบอกว่าความคลาดเคลื่อนมาตรฐานในการวัดที่หาได้นี้มีค่าคงที่ตลอดทุกช่วงคะแนน นั่นคือไม่ว่าผู้เข้าสอบคนใดที่ได้คะแนนต่าง ๆ กันจะมีคะแนนความคลาดเคลื่อนในการวัดเท่ากัน ซึ่งแท้ที่จริงมิได้เป็นเช่นนั้น ดังที่ โมเลนคอป์ฟ (Mollenkopf. 1949 : 189) ได้ทำการศึกษามาเป็นเวลากว่า 50 ปีแล้ว พบว่า ความคลาดเคลื่อนมาตรฐานในการวัดจะมีค่าคงที่เมื่อการแจกแจงของคะแนนที่ได้จากการวัดเป็นการแจกแจงแบบสมมาตรหรือเป็นแบบ Mesokurtic

ดังนั้นในการหาความคลาดเคลื่อนมาตรฐานในการวัดจึงควรประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดของคะแนนที่ได้จากการทำแบบทดสอบถูก ณ ที่ระดับความสามารถเฉพาะ หรือที่เรียกว่า*เป็นความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไข (Conditional standard error of measurement)*

ความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไข

การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขของคะแนนดิบ ได้มีการศึกษากันมาแล้วกว่า 50 ปี ผู้ที่นำเสนอวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขในระยะแรก ก็คือ โมเลนคอป์ฟ (Mollenkopf. 1949) และ ธอร์นไดค์ (Thorndike. 1951) ต่อมา ลอร์ด (Lord. 1955, 1957) ก็ได้พัฒนาวิธีการประมาณค่าขึ้นมาโดยใช้โมเดลความคลาดเคลื่อนในการวัดแบบทวินาม (Binomial Error Model) ซึ่งเป็นวิธีที่รู้จักกันดี หลังจากนั้น เฟลต์ (Feldt. 1984) ก็ได้เสนอวิธีการประมาณค่าโดยใช้โมเดลความคลาดเคลื่อนในการวัดแบบทวินามประกอบ (Compound Binomial Error Model) โดยกำหนดไว้ล่วงหน้าว่าแบบทดสอบเป็นแบบคู่ขนานที่สุ่มมาจากข้อสอบที่แบ่งเป็นชั้นหรือเป็นกลุ่มตามเนื้อหาหรือตามความยาก จากนั้นในปี 1980 ลอร์ด ก็ได้ใช้วิธีการประมาณค่าโดยใช้ทฤษฎีการตอบข้อสอบ (IRT) และในปี 1998 เบรนนัน (Brennan) ก็นำเสนอวิธีการประมาณค่าโดยใช้ทฤษฎีสัมประสิทธิ์การสรุปอ้างอิง (G - Theory) ในแต่ละวิธีเหล่านี้ต่างก็มีข้อตกลงเบื้องต้นในการใช้ที่แตกต่างกันไป และต่อมา เฟลต์ และเบรนนัน (Feldt & Brennan. 1989) และ เฟลต์และควอลส์ (Feldt & Qualls. 1996) ก็ได้นำวิธีการทั้งหลายเหล่านั้นมาประยุกต์ใช้ในการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไข

สำหรับแบบทดสอบที่นำมาใช้เพื่อการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขนั้นส่วนใหญ่เป็นการดำเนินการสอบครั้งเดียวและลักษณะแบบทดสอบเป็นแบบคำถาม

โตต (Item-based) นั่นคือการตอบแต่ละข้อจะมีคำถามมาให้เพียงคำถามเดียว การดอบจะไม่เกี่ยวกับเงื่อนไขหรือสถานการณ์ใดที่กำหนดให้ ต่อมา ลี (Lee. 2000) ได้เสนอวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขของแบบทดสอบที่ประกอบด้วยข้อคำถามเป็นชุด แต่ละชุดเป็นสถานการณ์ บทความ โคลง กลอน กราฟ และนำไปใช้ดอบคำถามหลายข้อ โดยใช้วิธีการประมาณค่าตามวิธี G - Theory และวิธี IRT แบบสามพารามิเตอร์

ดังนั้นวิธีการประมาณค่าแบบ IRT และ G - Theory ใช้ได้กับแบบทดสอบที่ประกอบด้วยข้อคำถามโดด และ ข้อคำถามที่เป็นชุด ส่วนวิธีที่ใช้โมเดลความคลาดเคลื่อนทวินาม (Binomial Error Model) ของ ลอร์ด (Lord. 1955, 1957) และวิธีที่ใช้โมเดลความคลาดเคลื่อนทวินามประกอบ (Compound Binomial Error) ของ เฟลด์ เป็นกรณีเฉพาะของวิธี G - Theory นั่นคือทั้งวิธี IRT และ G - Theory เป็นวิธีที่ได้รวบรวมเอาความคิดพื้นฐานเกือบทั้งหมดของหลายวิธีเข้าไว้ด้วยกัน ยกเว้นวิธีของ ธอร์นไดค์ (Thorndike. 1951) และ วิธีของโมเลนคอป์ฟ (Mollenkopf. 1949)

สำหรับการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัด ณ แต่ละคะแนนดิบ สามารถประมาณค่าได้หลายวิธี พอสรุปได้ ดังนี้

วิธีที่ใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่ วิธีของc (1949) และวิธีของธอร์นไดค์ (1951)

วิธีของโมเลนคอป์ฟ

โมเลนคอป์ฟ (1949) ได้นิยามความคลาดเคลื่อนมาตรฐานในการวัด ดังนี้

$$\sigma_{meas_x} = \sigma_x \sqrt{1 - r_{xx}} \qquad \text{.....(1)}$$

เมื่อ r_{xx} เป็นความเชื่อมั่น และ σ_x เป็นค่าเบี่ยงเบนมาตรฐานของคะแนนแบบทดสอบที่ได้จากกลุ่มผู้สอบทั้งหมดซึ่งแปรเปลี่ยนไปตามคะแนนจริงถ้ากลุ่มผู้สอบเหล่านี้ได้ทำข้อสอบที่คู่ขนานจำนวนมาก ดังนั้นเมื่อผู้เข้าสอบทั้งหมด N คน ทำข้อสอบที่คู่ขนาน K แล้วค่าความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละคน สามารถนิยามได้ดังนี้

$$\sigma_{meas_x} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{K}(x_{ij} - t_i)^2}{KN}} \qquad \text{.....(2)}$$

เมื่อ x_{ij} แทน คะแนนที่สอบได้ของแต่ละคนที่ตอบแบบทดสอบ j

t_i แทน คะแนนจริงของแต่ละคน

N แทน จำนวนคนที่สอบ

K แทน จำนวนแบบทดสอบที่คู่ขนานกัน

ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่ได้จากสมการ (2) จะมีค่าใกล้เคียงกับค่าที่ได้จากสมการ (1) ก็ต่อเมื่อคะแนนจริงของแต่ละคนเป็นค่าเฉลี่ยของคะแนนสอบที่ได้จากการตอบข้อสอบคู่ขนานกันจำนวนมาก นั่นคือ

$$t_i = \frac{\sum_{j=1}^{K} x_{ij}}{K}$$

วิธีของธอร์นไดค์ (1951)

การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขของธอร์นไดค์นี้ ได้กำหนดให้ X เป็นคะแนนรวมที่ได้จากคะแนนแต่ละครึ่งฉบับ (X_1 และ X_2) ที่คู่ขนานกัน นั่นคือ $X = X_1 + X_2$ จากโมเดลทฤษฎีการทดสอบแบบมาตรฐานเดิมนั้น คะแนนของแบบทดสอบแต่ละครึ่งฉบับและคะแนนรวมทั้งหมดสามารถเขียนอยู่ในรูปส่วนประกอบของคะแนนจริงและคะแนนความคลาดเคลื่อนได้ดังนี้

$$X_1 = T_1 + E_1$$

$$X_2 = T_2 + E_2$$

$$X = X_1 + X_2$$

$$= (T_1 + T_2) + (E_1 + E_2)$$

$$= T + E$$

จากทฤษฎีการทดสอบแบบมาตรฐานเดิม ค่าสหสัมพันธ์ระหว่าง E_1 และ E_2 เป็นศูนย์ จึงทำให้

$$\sigma_E = (\sigma_{E_1}^2 + \sigma_{E_2}^2)^{1/2}$$

เพราะว่า $X_1 - X_2 = (T_1 - T_2) + (E_1 - E_2)$

สำหรับแบบทดสอบคู่ขนานถ้า $T_1 = T_2$ จะได้ว่าค่าเบี่ยงเบนมาตรฐานของความแตกต่างระหว่างคะแนนแต่ละครึ่งฉบับเท่ากับความคลาดเคลื่อนมาตรฐานของคะแนนแบบทดสอบทั้งฉบับ

$$\sigma_{X_1 - X_2} = \sigma_{E_1} + \sigma_{E_2}$$

$$= (\sigma_{E_1}^2 + \sigma_{E_2}^2)^{1/2}$$

$$= \sigma_E$$

วิธีนี้มีข้อจำกัดตรงจำนวนคนในกลุ่มตัวอย่างที่ได้คะแนนในแต่ละระดับคะแนน ถ้ามีผู้ได้คะแนน ณ ระดับคะแนนที่สนใจเป็นจำนวนน้อย ควรรวมจำนวนคนที่ได้คะแนนทั้งฉบับเป็นช่วงคะแนนดีกว่าที่จะใช้แต่ละคะแนน

วิธีที่ใช้โมเดลความคลาดเคลื่อนแบบทวินาม (Binomial Error Model : BEM) เช่น วิธีของลอร์ด (1955) คีตส์ (1957) และวิธีของเฟลต์ (1984)

วิธีของลอร์ด

ลอร์ด (1955) ได้เสนอวิธีการประมาณค่าความคลาดเคลื่อนมาตรฐาน ณ แต่ละระดับคะแนนภายใต้โมเดลแบบทวินามซึ่งกล่าวว่าข้อคำถามจำนวน k ข้อ เป็นข้อที่สุ่มมาจากประชากรข้อสอบทั้งหมดที่คู่ขนานกัน และ X_i เป็นคะแนนที่ผู้สอบคนที่ i ตอบข้อคำถามได้ถูกเมื่อมีการให้คะแนนแบบ 0 - 1 สูตรที่ใช้ในการประมาณค่าความแปรปรวนคลาดเคลื่อนของผู้เข้าสอบแต่ละคนสามารถประมาณค่าได้ดังนี้

$$\tilde{\sigma}^2_{e_i} = \frac{X_i(k - X_i)}{k-1} \qquad \text{.....(3)}$$

ความแปรปรวนคลาดเคลื่อนเฉลี่ยของผู้เข้าสอบ N คน คือ

$$\text{ค่าเฉลี่ย } \tilde{\sigma}^2_e = \frac{[\bar{X}\ (k - \bar{X}) - S^2_X]}{k-1} \qquad \text{.....(4)}$$

เมื่อ $\bar{X}$ และ S^2_X เป็นค่าเฉลี่ยและความแปรปรวนของคะแนนที่สอบได้ เมื่อนำค่าเฉลี่ยความแปรปรวนคลาดเคลื่อนนี้ไปแทนค่าในสูตรความเชื่อมั่น จะได้ว่า

$$\text{reliabilty} = 1 - \frac{\sigma^2_e}{\sigma^2_X} \qquad \text{.....(5)}$$

ค่าที่ได้จากสูตรนี้จะเท่ากับค่าที่ได้จากสูตรของคูเดอร์ - ริชาร์ดสัน 21

ให้ $p_1, p_2, \ldots, p_n$ เป็นสัดส่วนของการตอบถูกใน k ข้อ ที่สุ่มมาจากประชากรข้อสอบทั้งหมด ดังนั้น

$$\sum_{h}^{k} p_h = k\bar{p} = \bar{X}_A$$

ความแปรปรวนที่ถูกประมาณค่าเฉลี่ยของแบบทดสอบที่เป็นฟอร์มคู่ขนานที่ได้จากการทำข้อสอบซ้ำ k ข้อ จากประชากรทั้งหมดคือ

$$\bar{\sigma}_X^2 = \bar{\sigma}_{kp}^2 = k^2\bar{\sigma}_p^2 = k^2\left(\frac{S_p^2}{k-1}\right) \qquad \text{.....(6)}$$

เมื่อ S_p^2 เป็นความแปรปรวนของ $p_1, p_2,...,p_k$ ของข้อคำถามในฟอร์ม A นำสมการ (6) ลบออกจากสมการ (4) จะได้

$$\begin{aligned}
\bar{\sigma}_e^2 &= \frac{\bar{X}(k-\bar{X})-S_X^2}{k-1} - k^2\left(\frac{S_p^2}{k-1}\right) \\
&= \frac{\bar{X}(k-\bar{X})-S_X^2}{k-1} - \left(\frac{k^2}{k-1}\right)\left(\frac{\sum p_h^2}{k} - \bar{p}^2\right) \\
&= \frac{\bar{X}(k-\bar{X})-S_X^2}{k-1} - \left(\frac{k^2}{k-1}\right)\left(\frac{\sum p_h^2}{k} - \frac{\bar{X}^2}{k^2}\right) \\
&= \frac{k\bar{X}-\bar{X}^2-S_X^2-k\sum p_h^2+\bar{X}^2}{k-1} \\
&= \frac{k\sum p_h - k\sum p_h^2 - S_X^2}{k-1} \\
&= \frac{k\sum p_h(1-p_h)-S_X^2}{k-1} \qquad \text{.....(7)}
\end{aligned}$$

ให้ $q_h = 1 - p_h$ แทนค่าความแปรปรวนคลาดเคลื่อนในสมการ (7) ลงในสมการ (5) จะได้สูตรการหาความเชื่อมั่นดังนี้

$$r_{XX} = 1 - \frac{k\sum p_h q_h - S_X^2}{(k-1)S_X^2}$$

$$= \frac{k}{k-1}\left(\frac{S_X^2 - \sum p_h q_h}{S_X^2}\right) \qquad \text{.....(8)}$$

สมการ (8) ก็คือสูตรการหาความเชื่อมั่นของคูเดอร์ - ริชาร์ดสัน นั่นเอง นั่นแสดงว่าทฤษฎีความคลาดเคลื่อนแบบทวินามได้ถูกปรับมาเพื่อให้ได้มีการยอมรับสูตร KR20 มากกว่า KR21

วิธีของคีตส์

จากการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขวิธีของลอร์ดในสมการ (3) นั้น นับเป็นวิธีการที่มีจุดอ่อนมาก เพราะไม่ได้คำนึงถึงเรื่องของการจับคู่ฟอร์มของแบบทดสอบ ความยากของแบบทดสอบหรือคุณลักษณะอื่นใดของแบบทดสอบเลย จึงทำให้การประมาณค่าที่ได้เกินความเป็นจริง ถ้าพิจารณาให้ดีแล้วค่ากำลังสองของสมการนี้ก็คือสัมประสิทธิ์ความเชื่อมั่นที่เป็นสูตรของคูเดอร์ - ริชาร์ดสัน 21 นั่นเอง ดังนั้นคีตส์ (1957) จึงเสนอวิธีการประมาณค่าความ คลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขขึ้นมาโดยนำค่าคงที่ไปคูณทางด้านขวาของสมการที่ลอร์ดเสนอเรียกว่าเป็นวิธีปรับแก้ของคีตส์ บางครั้งเขียนเป็นสัญลักษณ์ว่า ${}_{KL}S_E^2$ สูตรที่ใช้คือ

$$S_E = \left[\left(\frac{X(k-X)}{k-1}\right)\left(\frac{1-r_{xx'}}{1-r_{21}}\right)\right]^{1/2} \qquad \text{.....(9)}$$

เมื่อ r_{21} เป็นค่าความเชื่อมั่นของแบบทดสอบที่หาโดยสูตรของคูเดอร์ - ริชาร์ดสัน 21 และ $r_{xx'}$ เป็นค่าประมาณความเชื่อมั่นของแบบทดสอบซึ่งคีตส์เสนอว่าควรเป็นสัมประสิทธิ์ความเชื่อมั่นที่ได้จากฟอร์มคู่ขนาน แต่ในทางปฏิบัติอาจใช้สัมประสิทธิ์ความเชื่อมั่นแบบแบ่งครึ่งหรือสัมประสิทธิ์แอลฟ่าของครอนบัคก็ได้ แด่เฟลด์ - ควอลส์ (1998) ได้ใช้สัมประสิทธิ์ความเชื่อมั่นสูตรของคูเดอร์ - ริชาร์ดสัน 20

วิธีที่ใช้โมเดลทฤษฎีการตอบข้อสอบ

ทฤษฎีการดอบข้อสอบเป็นทฤษฎีที่แสดงความสัมพันธ์ระหว่างการตอบข้อสอบแต่ละข้อกับคุณลักษณะที่ข้อสอบนั้นจะวัด ความสัมพันธ์นี้แทนได้ด้วยฟังก์ชันการตอบข้อสอบ $P_j(\theta)$ หรือเส้นแนวทาง (trace line) และสามารถใช้ประมาณค่าโอกาสที่แต่ละบุคคลหนึ่งที่มีความสามารถ θ จะตอบข้อสอบข้อใดได้ถูกต้อง ในการประมาณค่าความแปรปรวนคลาดเคลื่อนในการวัดของแต่ละบุคคลมีสองขั้นตอน คือ

ขั้นแรก ประมาณค่าความสามารถ θ ของแต่ละบุคคลซึ่งอาจใช้โมเดลที่มีหนึ่งพารามิเตอร์ สองพารามิเตอร์ หรือสามพารามิเตอร์ ก็ได้

ขั้นที่สอง หาค่าฟังก์ชันต่อไปนี้ ซึ่งเป็นการประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดของแต่ละบุคคล

$$S^2_{E(i)} = \left\{ \sum_j^k [P_j(\theta_i)][1 - P_j(\theta_i)] \right\}^{1/2} \qquad \text{.....(10)}$$

เมื่อ $P_j(\theta_i)$ เป็นค่าฟังก์ชันของข้อคำถาม j ณ ระดับความสามารถ θ ในแต่ละบุคคล i

วิธีนี้คล้ายกับวิธีแบบทวินามประกอบที่ให้การประมาณค่าในผู้สอบแต่ละคน ในการใช้วิธีนี้ผู้เข้าสอบจะถูกจัดกลุ่มตามระดับคะแนนที่ได้ และคำนวณ $S^2_{E(i)}$

บทสรุป

การประมาณค่าความคลาดเคลื่อนมาตรฐานในการวัดอย่างมีเงื่อนไขหรือการประมาณค่าความคลาดเคลื่อน ณ แต่ละระดับคะแนนเฉพาะ เป็นสิ่งสำคัญที่ควรได้มีการรายงานไว้ควบคู่กับค่าสถิติของแบบทดสอบ เพราะจะทำให้เราสามารถประมาณค่าความสามารถจริงของผู้สอบแต่ละคนได้ด้วยความแม่นยำและชัดเจนยิ่งขึ้น วิธีการประมาณค่ามีหลายวิธีแต่ละวิธีก็มีข้อจำกัดต่างกัน ดังนั้นการที่จะเลือกใช้วิธีใดควรศึกษาแนวคิดของแต่ละวิธีให้ชัดเจน

เอกสารอ้างอิง

Feldt, Leonard S. (1984). "Some Relationships between the Binomial Error Model and Classical Test Theory," ***Educational and Psychological Measurement.*** 44 : 883 - 591.

Feldt, Leonard S.& Qualls, Audrey L. (1996, Summer). " Estimation of Measurement Error Variance at Specific Score Levels," ***Journal of Educational Measurement.*** 33(2) : 141 - 156.

Keats, J.A. (1957). "Estimation of Error Variances of Test Scores,' ***Psychometrika.*** 22 : 29 - 41.

Lord, Federic M. (1955). "Estimating Test Reliability'," ***Educational and Psychological Measurement.*** 15 : 325 - 336.

--------. (1957, Winter). "Do tests of the same length have the same standard errors of measurement?," ***Educational and Psychological Measurement.*** 17 : 510 - 521.

Lord, Federic M. & Novick, Melvin R. (1968). ***Statistical Theories of Mental Test Scores.*** CA : Addison - Wesley

Mollenkop, William G. (1949, September). "Variation of the Standard Error of Measurement," ***Psychometrika.*** 14(3) : 189 - 215.