

ดัชนีความสอดคล้องของผู้ประเมิน : KAPPA

ฉัตรศิริ ปิยะพิมลสิทธิ์*

ในการทดลองทางจิตวิทยาที่มีผู้สังเกตตั้งแต่ 2 คนขึ้นไปสังเกตเป้าหมายและจัดเข้ากลุ่มที่มีลักษณะเป็นนามบัญญัติ สมมติว่าให้การสังเกตของผู้สังเกตแต่ละคนเป็นอิสระจากกัน และผู้วิจัยปรารถนาจะวัดความสอดคล้องในการสังเกตว่ามีความแตกต่างกันอย่างมีนัยสำคัญจากความสอดคล้องที่คาดหวังหรือไม่

ถ้าผู้สังเกตแต่ละคนสังเกตตัวอย่าง n สิ่งแล้วตัดสินใจให้อยู่ในกลุ่มจำนวน C กลุ่ม ตารางความสอดคล้องจะเป็นตาราง 1 กลุ่มในแนวแถวต้องปรากฏเป็นลำดับเหมือนกับกลุ่มในแนวสดมภ์ และความสอดคล้องทั้งหมดจะปรากฏอยู่บนแนวทแยง ซึ่งผู้สังเกต 2 คนจัดเป้าหมายเข้ากลุ่มที่เหมือนกัน Goodman และ Kruskal อธิบายในกรณีนี้ว่ามีความเหมาะสมสำหรับการวัดความเชื่อมั่น สังเกตว่าไม่สามารถจะสลับแถวระหว่าง 2 แถวได้เช่นเดียวกับไม่สามารถสลับสดมภ์ระหว่าง 2 สดมภ์ในการสับเปลี่ยนอันดับ ดังนั้นลำดับของกลุ่มในแนวแถวและสดมภ์ต้องเป็นลักษณะเฉพาะสลับกันไม่ได้

ตาราง 1 สัญลักษณ์สำหรับความสอดคล้องของผู้สังเกต 2 คน

		Observer 2				
		1	2	...	C	
Observer 1	1	n_{11}	n_{12}	...	n_{1C}	n_{1+}
	2	n_{21}			:	n_{2+}
	:	:	n_{ij}		:	:
	C	n_{C1}	n_{CC}	n_{C+}
		n_{+1}	n_{+2}	...	n_{+C}	

$$\sum_{i=1}^C n_{i+} = n$$

$$= \sum_{i=1}^C n_{+i}$$

มีสัมประสิทธิ์ความเชื่อมั่นระหว่างผู้ประเมินหรือผู้สังเกต 2 คนที่สังเกตกลุ่มตัวอย่างและจัดเข้ากลุ่มเป็นนามบัญญัติ ดังนี้

* กศ.บ. (การวัดผลการศึกษา) <http://www.watpon.com/>

1. Asymmetric Kappa

ในการศึกษาต่าง ๆ ที่ใช้การสังเกต อาจจะมีผู้สังเกต (หรือผู้ประเมิน) 2 คน โดยที่ผู้ประเมินแต่ละคนเป็นอิสระจากกันในการประเมินข้อความ (หรือสิ่งที่สังเกต) แล้วจัดให้เป็นกลุ่มระดับนามบัญญัติ (nominal categories) สิ่งที่น่าสนใจก็คือความสอดคล้องของผู้สังเกตที่ประเมินสิ่งที่สังเกตได้สอดคล้องกันหรือไม่ ตัวอย่างเช่น มีนักจิตวิทยา 2 คนอาจจะประเมินกลุ่มของคนไข้และประเมินคนไข้แต่ละคนอย่างเป็นอิสระจากกันแล้วจัดอยู่ในกลุ่มที่วินิจฉัย

อันดับแรกเราจะพิจารณาตามาตรนามบัญญัติจำนวน i กลุ่ม กับผู้สังเกต 2 คนที่เป็นอิสระจากกัน ประเมินข้อสอบแต่ละข้อซึ่งในแต่ละข้อถูกจัดอยู่ในกลุ่มใดกลุ่มหนึ่ง สัญลักษณ์ p_{ij} แทนสัดส่วนของข้อสอบที่ผู้สังเกตคนที่ 1 จัดอยู่ในกลุ่มที่ i และผู้สังเกตคนที่ 2 จัดอยู่ในกลุ่มที่ j ซึ่ง i ไม่เท่ากับ j ดังนั้น p_{ij} กลายเป็นสัดส่วนของความไม่สอดคล้องของผู้สังเกต 2 คน เมื่อ p_{ii} เป็นสัดส่วนของความสอดคล้อง สัดส่วนทั้งหมดของความสอดคล้องของการสังเกตก็คือ $\sum p_{ii}$ เขียนเป็นสูตรประมาณค่าความสอดคล้องได้ว่า

$$K = \frac{\sum p_{ii} - \sum p_{i+} + p_{+i}}{1 - \sum p_{i+} + p_{+i}} \quad (1)$$

ถ้ามีความสอดคล้องอย่างสมบูรณ์ระหว่างผู้สังเกต 2 คนแล้ว $\sum p_{ii} = 1$ และ $K = 1$ แต่ $K = 0$ เมื่อความสอดคล้องของผู้สังเกตเท่ากับความสอดคล้องที่คาดหวัง และ K มีค่าเป็นลบถ้าความสอดคล้องของผู้สังเกตมีค่าน้อยกว่าความสอดคล้องที่คาดหวัง

2. Symmetric Kappa

ในสูตรก่อนหน้านี้เป็นการวัดความสอดคล้องที่เรียกว่า K ในสมการ (1) สำหรับผู้สังเกต 2 คนที่เท่าเทียมกัน นั่นคือ K มีผลเป็น symmetric measure ในบางสถานการณ์อาจจะต้องการผู้สังเกตเพียงคนเดียวที่สนใจโดยเฉพาะและพิจารณาผู้สังเกตนี้ให้เป็นมาตรฐาน (standard) หรือเกณฑ์ ซึ่งจะนำไปเปรียบเทียบกับผลการประเมินของผู้สังเกตอีกคนหนึ่ง

วิธีการของการวัดความสอดคล้องกับ “มาตรฐาน” จะถูกพัฒนาโดย Light (1971), Williams, และ Wackerly, McClave and Rao บุคคลเหล่านี้ได้นำ asymmetric Kappa มาขยายผลต่อจากสมการ (1) กลายเป็นสมการใหม่

$$K_{2/1} = \sum p_{i+} K_i = \sum \left(\frac{p_{ii} - p_{i+} + p_{+i}}{1 - p_{+i}} \right) \quad (2)$$

$K_{2/1}$ ถูกเรียกว่า asymmetric Kappa, $K_{2/1}$ เป็นการเสนอการวัดความสอดคล้องระหว่างผู้สังเกต 2 คนเมื่อคนที่ 1 ถูกพิจารณาให้เป็น “มาตรฐาน” (standard) อาจจะได้ว่า $K_{2/1}$ เป็นการวัดความสอดคล้องของผู้สังเกตคนที่ 2 กับผู้สังเกตคนที่ 1 ส่วนการวัด K เป็นความสอดคล้องที่เป็นอิสระจากกันระหว่างผู้สังเกต 2 คนที่เท่าเทียมกัน

ตาราง 2 สัดส่วนของกลุ่มตัวอย่างที่ถูกประเมินโดยผู้เชี่ยวชาญและนักศึกษา

ผู้เชี่ยวชาญ	นักศึกษา			Totals
	อาการทางจิต	อาการทางประสาท	อาการทางกาย	
อาการทางจิต	.40	0	.05	.45
อาการทางประสาท	.10	.20	.02	.32
อาการทางกาย	.14	.01	.08	.23
Total	.64	.21	.15	1.00

ในบางคุณสมบัติของ $K_{2/1}$ ในสมการ (2) มีทำนองเดียวกับ K ในสมการ (1) นั่นคือ ถ้ากลุ่มการตอบสนองหรือข้อสอบโดยผู้สังเกตคนที่ 2 สอดคล้องอย่างสมบูรณ์กับผู้สังเกตคนที่ 1 แล้ว $K_{2/1} = 1$ ถ้าความสอดคล้องที่ถูกสังเกตมากกว่าหรือเท่ากับความสอดคล้องที่คาดหวังแล้ว $K_{2/1} \geq 0$ นอกนั้น $K_{2/1}$ จะมีค่าติดลบ และ $K_{2/1}$ ยังเป็นค่าเฉลี่ยถ่วงน้ำหนักของ K_i ในสมการ (3) หรือ (4) $K_{2/1}$ จะมีค่าระหว่างค่าต่ำสุดและค่าสูงสุดของ K_i สำหรับข้อมูลชุดเดียวกัน

$$K_i = \frac{p_{ii} / p_{i+} - p_{+i}}{1 - p_{+i}} \quad (3)$$

$$= \frac{p_{ii} - p_{i+}p_{+i}}{p_{i+} - p_{i+}p_{+i}} \quad (4)$$

ต่อไปนี้จะแสดงตัวอย่างเชิงจำนวนให้เห็นการคำนวณค่า $K_{2/1}$ ซึ่งจะให้ผู้เชี่ยวชาญที่ถูกพิจารณาเป็น “มาตรฐาน” และนักศึกษาที่เรียนจิตวิทยา มาดำเนินการสังเกตกลุ่มตัวอย่างอย่างเป็นอิสระจากกันและจัดเข้ากลุ่มตัวอย่างเข้ากลุ่มลักษณะอาการของคนไข้ 3 กลุ่มคือ อาการทางจิต, อาการทางประสาท หรืออาการทางกาย ผลในการจัดเข้ากลุ่มแสดงในตาราง 2 ดังนั้นในตัวอย่างนี้มี 5% ของกลุ่มตัวอย่างที่ถูกวินิจฉัยให้อยู่ในกลุ่มอาการทางจิตโดยผู้เชี่ยวชาญหรือกลุ่มอาการทางกายโดยนักศึกษา ดังนั้น 45% ของกลุ่มตัวอย่างจะถูกวินิจฉัยว่าเป็นกลุ่มอาการทางจิตโดยผู้เชี่ยวชาญเปรียบเทียบกับ 64% ที่ถูกประเมินโดยนักศึกษา เมื่อข้อมูลในตาราง 2 ใช้กับสมการ (2) เราจะได้

$$K_{2/1} = \frac{.40 - (.45)(.64)}{1 - .64} + \frac{.20 - (.32)(.21)}{1 - .21} + \frac{.08 - (.23)(.15)}{1 - .15} = .53$$

ค่าจะบ่งชี้ว่าการสังเกตของนักศึกษาสอดคล้องอย่างดีกับการสังเกตโดยผู้เชี่ยวชาญ ถ้าไม่ต้องการพิจารณาการสังเกตของผู้เชี่ยวชาญที่เป็นมาตรฐานเปรียบเทียบกับ การสังเกตของนักศึกษา แต่ทั้งสองคนนี้มีความเท่าเทียมกันแล้ว ก็ควรจะใช้ symmetric K จากสมการ (1) และจากผลในตาราง 2 จะได้ค่า $K = .48$ และ K_1 ในสมการ (3) หรือ (4) ถูกคำนวณโดยใช้ข้อมูลในตาราง 1 จะพบว่า $K_1 = .69$, $K_2 = .53$ และ $K_3 = .23$ สำหรับความสอดคล้องของกลุ่มอาการทางจิต, อาการทางประสาท และอาการทางกาย ดังนั้นระดับสูงสุดของความสอดคล้อง มีก็คือกลุ่มตัวอย่างที่ผู้เชี่ยวชาญจะจัดให้อยู่ในกลุ่มอาการทางจิต ตามสมการ (2) $K_{2/1}$ สามารถจะคำนวณได้จากค่าเฉลี่ยถ่วงน้ำหนัก K_1 นั่นคือ $K_{2/1} = 45(.69) + 32(.53) + 23(.23) = .53$



บรรณานุกรม

- Brenan, Robert L. and Prediger, Dale J. "Coefficient Kappa : Some Uses, Misuses, and Alternatives," *Educational And Psychological Measurement*. 41(1981). P.687-699.
- Collin, Glyn M. "Kappa, Measures of Marginal Symmetry and Intraclass Correlation," *Educational and Psychological Measurement*. 45(1985), p.55-62.
- Kvalseth ,Tarald O. "A Coefficient of Agreement for Nominal Scales : An Asymmetric Version of Kappa," *Educational and Psychological Measurement*. 51(1991). p.95-101.
- Light, Richard J. "Measures of Response Agreement for Qualitative Data : Some Generalizations and Alternatives," *Psychological Bulletin*. 5(76) : 1971. p.364-377.
- Rae, Gordon. "The Equivalence of Multiple Rater Kappa Statistics and Intraclass Correlation Coefficients," *Educational and Psychological Measurement*. 48(1988). p.367-374.