บทที่ 9
มาตรการวัด

<u>วัตถุประสงค์</u> เพื่ออธิบายการสร้าง Rating Scale และ Index Construction

การวิจัยทางสังคมศาสตร์ ส่วนใหญ่จะเป็นการศึกษาพฤติกรรมของมนุษย์ในสังคมในเรื่องใดเรื่องหนึ่ง พฤติกรรมของมนุษย์นั้น อาจจะแบ่งได้เป็น

1. พฤติกรรมที่แท้จริง (Gross behavioral response)
2. ความรู้สึก หรือความคิดเห็น (Sentiment or opinion)
3. ทัศนคติ (Attitude)
4. พฤติกรรมสมมติ (Hypothetical Actions)

ในการศึกษาพฤติกรรมที่เป็นความคิดเห็นหรือทัศนคตินั้น ผู้วิจัยจำเป็นต้องใช้วิธีวัดความคิดเห็นหรือทัศนคติบางประการ ในเรื่องนี้ให้ถูกต้องและแน่นอน วิธีการที่จะวัดก็คือ การให้ประชากรที่ศึกษาจัดระดับตัวเองว่ามีความรู้สึกหรือทัศนคติในเรื่องนั้นอยู่ในขั้นไหน เช่น เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่มีความเห็น ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง และในแต่ละขั้นผู้วิจัยจะกำหนดคะแนนเอาไว้ วิธีการนี้เรียกว่า Rating Scale ซึ่งมีหลายชนิด ในที่นี้จะยกเอา Scale ที่เป็นที่นิยมใช้กันแพร่หลายมาอธิบาย 3 แบบ คือ

1. Semantic differential scale (Thurstone)
2. Summated scales (Likert)
3. Scalogram Analysis (Guttman)

1. Semantic differential scale

วิธีการวัดทัศนคติในเรื่องใดเรื่องหนึ่ง โดยจะมีคำสองคำที่แสดงถึงทัศนคติที่ตรงกันข้ามกัน เช่น ดี-เลว, เห็นด้วย-ไม่เห็นด้วย ผู้ตอบจะต้องทำเครื่องหมายใน

ระยะทางระหว่างคำสองคำนี้ ซึ่งจะแบ่งออกเป็น 7 ชั้น ดังนี้

ดี — — — — — — — เลว

(1) (2) (3) (4) (5) (6) (7)

ก่อนที่จะให้ผู้ตอบทำเครื่องหมาย จะมีข้อความที่เกี่ยวกับเรื่องที่จะวัดให้ผู้ตอบอ่านก่อน ตัวอย่างเช่น

"ท่านเห็นว่าระบบการให้คะแนนในปัจจุบันเป็นอย่างไร"

ผู้ตอบก็จะกาเครื่องหมาย (+) ลงไปในชั้นใดชั้นหนึ่งระหว่าง ดี — เลว ตัวเลขที่อยู่ในวงเล็บ มักจะไม่เขียนลงไปในแบบสอบถาม เป็นการให้คะแนนในตอนหลัง การวัดด้วยวิธีนี้ถือว่าเป็นการวัดในระดับ interval เพราะถือว่า ระยะทางระหว่างชั้นต่าง ๆ ในเจ็ดชั้นมีระยะทางเท่ากัน ซึ่งทำให้สามารถเปรียบเทียบการเปลี่ยนแปลงของทัศนคติได้ เช่น เมื่อทัศนคติเปลี่ยนจาก 3 เป็น 7 จะเท่ากับเปลี่ยนจาก 6 เป็น 10 เพราะเปลี่ยนไปในระยะที่เท่ากัน คือ 4 คะแนน การวัดด้วยวิธีนี้นิยมใช้กันมากแต่ยังมีข้อวิจารณ์ในเรื่องความเที่ยงตรง (Validity)

2. Summated Scale หรือบางทีเรียกว่า Likert Scale มีวิธีการคล้ายกับ Semantic differential ที่กำหนดให้มีข้อความที่เกี่ยวข้องกับเรื่องที่จะวัดจำนวนหนึ่งและประชากรจะถูกให้แสดงทัศนคติต่อข้อความนั้น โดยให้กาเครื่องหมายลงในข้อใดข้อหนึ่ง ดังต่อไปนี้

1. ไม่เห็นด้วยอย่างยิ่ง
2. ไม่เห็นด้วย
3. ไม่มีความเห็น
4. เห็นด้วย
5. เห็นด้วยอย่างยิ่ง

เมื่อผู้ตอบตอบหมดแล้ว ก็จะรวมคะแนนทั้งหมด เพื่อที่จะกำหนดว่าผู้ตอบแต่ละคนได้คะแนนเท่าไร หรือมีทัศนคติอย่างไรในเรื่องนี้ เมื่อเทียบกับผู้ตอบคนอื่น ๆ Likert Scale นี้ เป็นการวัดในระดับ ordinal ในการหาความเชื่อถือได้ (Reliability) ของ scale ประเภทนี้ ใช้วิธี Split - half

ส่วนการทดสอบความถูกต้องของ Scale ว่าได้คะแนนถูกต้องหรือไม่ หรือวัดในเรื่องที่ต้องการจะวัดหรือไม่มีอยู่ 2 วิธีคือ

1. Item Analysis
2. Internal Consistency

1. วิธีการทำ Item Analysis มีขั้นตอนดังนี้

ขั้นที่ 1 ให้คะแนนประจำข้อเฉพาะแบบสอบถามประเภททัศนคติ โดยคำนึงถึง คำถามแบบตอบรับ (Positive) และตอบปฏิเสธ (Nagative)

เกณฑ์การให้คะแนน คำถามแบบตอบปฏิเสธ (Nagative)

ตัวเลือกตอบ		ให้คะแนนเป็น
	เห็นด้วยอย่างยิ่ง	1
	เห็นด้วย	2
	ไม่แน่ใจ	3
	ไม่เห็นด้วย	4
	ไม่เห็นด้วยอย่างยิ่ง	5

เกณฑ์การให้คะแนนคำถามแบบตอบรับ (Positive)

ตัวเลือกตอบ	เห็นด้วยอย่างยิ่ง	ให้คะแนนเป็น	5
	เห็นด้วย	"	4
	ไม่แน่ใจ	"	3
	ไม่เห็นด้วย	"	2
	ไม่เห็นด้วยอย่างยิ่ง	"	1

สมมติว่าเรามีคำถามแบบตอบรับ (Positive) จำนวน 10 ข้อ และคำถามแบบตอบปฏิเสธ (Nagative) จำนวน 10 ข้อ รวมเป็น 20 ข้อ คละกันไปทั้งนี้เพื่อมิให้ผู้ตอบตอบไปในทิศทางเดียวกัน ซึ่งจะทำให้ข้อมูลผิดพลาดได้ คำถามแต่ละข้อถือว่าเป็นตัวแปรแต่ละตัว ในที่นี้มีคำถามอยู่ 20 ข้อ จำนวน 20 ตัวแปร โดยให้สัญญลักษณ์เป็น X_1 ——— X_{20} ตามลำดับ

<u>ขั้นที่ 2</u> รวมคะแนนของผู้ตอบแบบสอบถามแต่ละคน คะแนนที่รวมได้จะถือว่าเป็นตัวแปรใหม่ที่เกิดขึ้น โดยให้สัญลักษณ์เป็น X_{21}

<u>ขั้นที่ 3</u> หาค่าสหสัมพันธ์ (correlation) ระหว่างคำถามที่เป็นตัวแปรแต่ละตัว (X_1 — X_{20}) กับตัวแปรใหม่ที่เกิดจากการรวมคะแนนในข้อ 20 (X_{21}) ที่ละคู่

<u>ขั้นที่ 4</u> พิจารณาค่าสหสัมพันธ์ (Correlation) ที่คำนวณได้ถ้าค่าที่คำนวณได้มีค่าติดลบหรือค่าที่คำนวณได้ไม่ใกล้เคียงกัน แสดงว่า เครื่องมือนั้นไม่สามารถวัดในทิศทางเดียวกันได้ หรืออาจจะเป็นเพราะการให้คะแนนประจำข้อที่เป็นคำถาม

แบบตอบรับ และตอบปฏิเสธผิดพลาดไป ผู้วิจัยจึงควรได้มีการตรวจสอบก่อนพิจารณาคัดคำถาม (Item) ที่มีค่าไม่ใกล้เคียงกัน หรือมีค่าติดลบออกไป

2. วิธีการทำ Internal Consistency มีขั้นตอนดังนี้

ขั้นที่ 1 รวมคะแนนทั้งหมดของผู้ตอบแต่ละคน

ขั้นที่ 2 นำคะแนนของผู้ที่ได้คะแนนสูงสุดร้อยละ 10 และผู้ที่ได้คะแนนต่ำกว่าร้อยละ 10 มาหาผลรวมของคะแนนในแต่ละข้อหรือแต่ละ Item

ขั้นที่ 3 หาค่าความแตกต่างระหว่างกลุ่มคะแนนสูงและกลุ่มคะแนนต่ำของแต่ละ Item

ขั้นที่ 4 หาค่าเฉลี่ยของผลต่างที่ได้จากขั้นที่ 3 โดยหารด้วย จำนวน Item

ขั้นที่ 5 เรียงลำดับค่าเฉลี่ยหากค่าเฉลี่ยที่ได้ใกล้เคียงกันก็แสดงว่าข้อคำถามเหล่านั้นใช้ได้

3. Scalogram Analysis หรือ Guttman Scale

วิธีการวัดทัศนคติอีกวิธีหนึ่ง แต่ในปัจจุบันได้มีผู้นำเอาวิธีนี้มาใช้วัด Response Consistency ของคนในสังคม นอกจากนี้ยังใช้เป็นเครื่องมือที่ใช้วัดความแตกต่างในโครงสร้างสังคม (Structural differentiation) ซึ่งถือว่าเป็นเครื่องชี้ถึงการพัฒนาสังคม วิธีการวัดแบบนี้คล้ายกับสองวิธีแรก คือ ต้องมีข้อความชุดหนึ่งประมาณ 5 - 20 ข้อ ซึ่งจะต้องเป็นข้อความที่จะวัดเรื่องใดเรื่องหนึ่ง (Unidimension) สำหรับคำตอบนั้นอาจจะเป็น Dichotomus Choice หรือ Multiple Choice ก็ได้ แต่ถ้าเป็น

เป็น Multiple Choice จะต้องทำให้เป็น Dichotomus Choice เสียก่อน เมื่อได้คำตอบแล้วก็จะเรียงลำดับความสำคัญของข้อความในชุดนั้น โดยถือเอาเปอร์เซนต์ที่ได้จากการตอบในแต่ละข้อเป็นหลัก ถ้าผู้ตอบคนใดตอบได้ตรงตามลำดับก็ถือว่าถูกต้อง แต่ถ้าตอบผิดไปจากลำดับก็จะถือว่าเป็นการคลาดเคลื่อน (error) การที่จะยอมรับว่าเป็น Scale ที่ใช้ได้นั้น มีหลักเกณฑ์อยู่ 2 อย่างคือ

1. Coefficient of Reproducibility
2. Coefficient of Scalability

การคำนวณทั้ง 2 วิธีนี้ ถือว่า errors ที่ผู้ตอบไม่ตรงตามที่ได้เรียงลำดับไว้เป็นหลัก โดยมีหลักเกณฑ์ว่าจะยอมรับ Scale ก็ต่อเมื่อ

Coefficient of Reproducibility มีค่ามากกว่า .90 และ

Coefficient of Scalability มีค่ามากกว่า .60

Guttman ได้เสนอ Scale ที่สมบูรณ์ของเขาดังตารางต่อไปนี้

<u>ตารางที่ 1</u> The Perfect Guttman Scale

(เรียงลำดับจากมากไปหาน้อย)

	item						
Patterns	A	B	C	D	E	F	scores
1	1	1	1	1	1	1	6
2	1	1	1	1	1	0	5
3	1	1	1	1	0	0	4
4	1	1	1	0	0	0	3
5	1	1	0	0	0	0	2
6	1	0	0	0	0	0	1
7	0	0	0	0	0	0	0

Perfect Scale ของ Guttman นี้เป็นไปได้เฉพาะในทางทฤษฎีเท่านั้น แต่ในทางปฏิบัติเป็นไปได้ยากมากย่อมจะเกิด errors ได้เสมอ Guttman เองก็ยอมรับความจริงในข้อนี้ จึงได้ยอมให้มี errors ได้ร้อยละ 10 ซึ่งหมายความว่า การที่จะยอมรับเป็น scale ที่ใช้ได้จะต้องมีค่าของ Coefficient of Reproducibility .90 ขึ้นไป ซึ่งคำนวณค่า Coefficient of Reproducibility (C.R.)นี้ได้มาจากสูตร

$$\text{Coefficient of Reproducibility (C.R.)} = \frac{1 - \text{number of errors}}{\text{total responses}}$$

แต่อย่างไรก็ตามต่อมาในระยะหลัง Herbert Manzel มีความเห็นว่าค่าของ C.R. ที่กำหนดให้ตั้งแต่ .90 ขึ้นไปนั้น มีโอกาสเป็นไปได้ยาก ทั้งนี้ขึ้นอยู่กับข้อมูล ถ้าข้อมูลนั้นเป็นคำถามสุดโต่ง (Extreme) ในข้างใดข้างหนึ่ง ถ้าเป็นคำถามที่เห็นชัดและง่ายแก่การตอบ โอกาสที่จะได้ค่า C.R. = .90 ก็เป็นไปได้ง่าย ซึ่งเป็นการวัดที่ไม่ถูกต้องอย่างแท้จริง (Validity) เขาจึงเสนอสูตร Coefficient of Scalability ขึ้นมาอีกสูตรหนึ่ง ซึ่งหาได้จากสูตรที่ว่า

$$\text{Coefficient of Scalability (C.S.)} = \frac{1 - \text{number errors}}{\text{maximum errors}}$$

Scale ที่สร้างขึ้นนั้นย่อมจะยอมรับได้ก็ต่อเมื่อค่าของ C.S. ตั้งแต่ .60 ขึ้นไป

Domian Sampling Model

วิธีการทดสอบหา reliability ของการวัด concept ใด concept หนึ่งที่ประกอบไปด้วยหลาย subconcept เป็นวิธีหนึ่งของ Index Construction หรือในการที่ผู้วิจัยได้นำแนวความคิดและทฤษฎีต่าง ๆ มาเป็นเครื่องมือในการทดสอบการวิจัยแล้ว โดยใช้ item ต่าง ๆ วัดตัวแปรแต่ละตัวนั้น item นั้น ๆ สามารถวัดในสิ่งเดียวกันได้จริงหรือไม่ โดยถ้าวิเคราะห์มาแล้วว่า item ต่าง ๆ ที่ใช้นั้นสามารถวัดในสิ่งเดียวกันจริง ก็ถือว่า item นั้นอยู่ใน domain เดียวกัน

ในการหาค่า item ต่าง ๆ อยู่ใน domain เดียวกันนี้ Domain Sampling เป็นวิธีการหนึ่งที่ใช้ในการวิเคราะห์ โดยใช้หาค่าสหสัมพันธ์ของ item ย่อยและมีวิธีการวิเคราะห์ในการที่จะใช้วิธีการนี้ จะต้องทำ Index Matrix ออกมาก่อนซึ่งเป็นค่าสหสัมพันธ์ของ item แต่ละตัว

วิธีการหา

1. จะต้องมีการกำหนด statement ก่อนว่ามีเท่าไร
2. จากนั้นทำ Correlation matrix ตารางแสดงค่าความสัมพันธ์ (Correlation ระหว่างแต่ละ Statement
3. จากนั้นหาค่า Column Sum แต่ละ Statement
4. หาค่า Column Average แต่ละ Statement
5. หาค่า $\bar{r}_{ij}$ โดยเอาค่า Correlation ทั้งหมดบวกกันแล้วหารด้วยจำนวน Correlation
6. เปรียบเทียบค่า Column Average กับค่า $\bar{r}_{ij}$
7. พิจารณาตัด statement หา statement ที่มีค่าพิจารณาค่า $\bar{r}_{ij}$ มาก

8. จากนั้นเอา Statement ที่คัดแล้วมาพิสูจน์หาค่า r_{kk} ซึ่งจะต้องได้ค่าไม่ต่ำกว่า .75 จึงจะยอมรับ statement เหล่านั้น

นั่นคือ Domain Sampling Model เป็นการทดสอบว่า Subconcept เหล่านั้นจะอยู่ใน Domain (concept ใหญ่) เดียวกันหรือเปล่า โดยการทดสอบกับค่า r_{ij} และพิสูจน์หา Reliability ของ Subconcept เหล่านั้นว่าจะใช้วัด concept ใหญ่ได้จริงหรือเปล่า โดยใช้ค่า r_{kk} เข้ามาเป็นตัววัด

<u>ตัวอย่าง</u> วิธีการหา เราต้องกำหนดก่อนว่าตัวแปรของเราประกอบด้วย item อะไรบ้าง เราจะต้องนำเอาทฤษฎีมาเป็นแนวทางในการใช้

(1) เอา item ทั้งหมดมาหาค่าความสัมพันธ์ สมมติมี 4X สมมติได้ค่า

	X_1	X_2	X_3	X_4
X_1	1	.52	.30	.48
X_2	.55	.1	.34	.56
X_3	.52	.50	.1	.60
X_4	.54	.50	.46	.1
Column Sum	1.61	1.52	1.04	1.64
Column Average	.53	.51	.34	.54

(2) เมื่อได้ค่าเฉลี่ยแล้วดูค่าของความแตกต่างกับของค่าเฉลี่ยแต่ละ Column ถ้าพบว่า X_3 น้อยกว่าค่าอื่นมีความสัมพันธ์ค่าเฉลี่ยแตกต่างไปจาก X_1, X_2, X_4 สรุปได้ว่า ตัวแปรที่อยู่ใน Domain เดียวกัน จากการ Domian Sampling ได้ ก็คือ X ประกอบด้วย X_1, X_2, X_4

เราอาจมีการทดสอบด้วยวิธีการหาค่า r_{ij} คือเอาค่าของ column และ row ทั้งหมดมารวมกัน แล้วหารด้วยจำนวน x จากสูตร

$$r_{ij} = \frac{\text{ผลบวกของสหสัมพันธ์ทั้งหมด}}{\text{จำนวนทั้งหมด}}$$

ซึ่งได้แก่ .55 + .52 + .54 + .52 + .50 + .50 + .30 + .34 + .46 + .48 + .56 + .60 แล้วหารด้วย 12 ได้ค่า .48 เอาไปเทียบกัน

x_1, x_2, x_3, x_4 ดูก็พบว่า x_3 ได้ .34 มีค่าห่างไปจาก .48 จึงไม่อยู่ใน Domain เดียวกัน

สรุปว่า ความเชื่อถือได้ (Reliability) มากน้อยเพียงใดนั้นขึ้นอยู่กับการเอาค่า

$$r_{kk} = \frac{k\,\bar{r}_{ij}}{1 + (k-1)\,\bar{r}_{ij}}$$

k = จำนวน item ทั้งหมด

ถ้าได้ค่า 0.75 ขึ้นไปถือว่า มีความเชื่อถือได้ ในลักษณะของการวิเคราะห์แบบนี้สามารถเอา item เข้าไปอีกได้ถ้า ค่า r_{kk} ต่ำกว่า 0.75 และลดค่า item ได้ ค่า r_{kk} สูงมาก