

การกำหนดขนาดตัวอย่างสำหรับการวิเคราะห์ตัวแบบสมการโครงสร้าง  
**Sample size determination for Structural Equation Modeling (SEM)**

รองศาสตราจารย์ ดร. มนต์รี พิริยะกุล\*

**บทคัดย่อ**

ขนาดตัวอย่างที่ใช้เพื่อทำการวิจัยโดยวิธีการวิเคราะห์ตัวแบบสมการโครงสร้างเป็นปัญหาที่นักวิจัยทุกคนสงสัยว่าควรกำหนดเท่าไรจึงถูกต้องและ/หรือเหมาะสม ควรเน้นที่จำนวนมากตามทฤษฎีการทดสอบหรือว่าควรน้อยลงตามเหตุผลอื่นๆ การศึกษาโดยวิธีจำลองแบบและโดยวิธีพัฒนาสูตรสถิติพบว่าขนาดตัวอย่างสำหรับ SEM มีได้หลายขนาดหลากหลายแตกต่างกันไปตามกฎเกณฑ์ทางสถิติที่คิดเผกรวมทั้งกฎอย่างง่าย แต่พอสรุปได้ว่าขนาดตัวอย่างที่พอเพียงควรมีประมาณ 200 แต่อาจเล็กกว่านี้หรือใหญ่กว่านี้ได้ขึ้นอยู่กับองค์ประกอบของตัวแบบและขนาดประชากร

**คำสำคัญ** ตัวแบบสมการโครงสร้าง การกำหนดขนาดตัวอย่าง กฎกำหนดขนาดตัวอย่างอย่างง่าย

**Abstract**

Sample size was the problem that researchers always in question of what size is correct or suitable for SEM analysis. Whether more subjects, according to classical test theory, is better or smaller is plausibly. Many investigation through simulation studies and some statistical formulas revealed that number of respondents can be variable subject to statistical criterions used including rule of thumb. But, as a rule of thumb, sample size of 200 seems sufficient. However, more than 200 respondents is possible subject to component of SEM model itself and population size constrained.

**Keywords:** Structural Equation Modeling (SEM), Determination of sample, Rule of thumb

---

\* ภาควิชาสถิติ มหาวิทยาลัยรามคำแหง E-mail: mpiriyakul@yahoo.com

บทความนี้อยู่ระหว่างรอเผยแพร่ ข้อเสนอแนะการอ้างอิงให้นักวิจัยอ้างอิงแหล่งเอกสารชื่อเว็บไซต์

## บทนำ

ในการวิจัยโดยใช้วิธีวิทยาสถิติการวิเคราะห์ตัวแบบสมการโครงสร้าง (Structural Equation Modeling, [SEM]) นักวิจัยจะพบปัญหาตลอดมาว่าควรกำหนดขนาดตัวอย่างเท่าไรจึงจะถือว่าเป็นขนาดที่เหมาะสม ไม่มากเกินไป (more is never too much) ตามทฤษฎีการทดสอบจนเป็นภาระหนัก หรือไม่น้อยเกินไป (more is ever too much) จนได้รับการโต้แย้งว่าน่าจะไม่มีพอที่จะมีผลให้ค่าประมาณมีความถูกต้อง (accuracy) และน่าเชื่อถือ (reliable)

เกณฑ์การกำหนดขนาดตัวอย่างสำหรับ SEM ในระยะเวลาที่ผ่านมามีอยู่หลายประการ เกณฑ์เหล่านี้นักวิจัยเองก็ไม่มั่นใจว่าถูกต้องหรือไม่? มีการศึกษาใดยืนยันความเหมาะสมหรือไม่? เป็นปัจจุบันหรือไม่? เกณฑ์ต่อไปนี้จะรวมเรียกว่ากฎอย่างง่าย (rule of thumb)

1. กำหนดให้  $n$  มีขนาดใหญ่มาก ๆ เข้าไว้ ยิ่งมากยิ่งดี
2. อัตราส่วนจำนวนข้อถามขั้นต่ำ โดยทั่วไปจะกำหนดให้ตัวแปรแฝงหนึ่งๆมีข้อถามอย่างน้อย 3 ข้อ (Anderson and Rubin, 1956) หรือ 2 ข้อถาม (Bollen, 1989) ซึ่งเมื่อย้อนคูณกลับกับจำนวนตัวแปรแฝง (latent variable, LV) จะได้ตัวอย่าง ( $n$ ) เท่ากับจำนวนตัวแปรแฝง \* 2 หรือจำนวนตัวแปรแฝง \* 3 ซึ่งยอมรับได้ยากเพราะ  $n$  จะมีขนาดเล็กเกินไปและขัดแย้งกับข้อ 1.
3. กำหนดให้  $n = 100, 200$  (Boomsma, 1985)
4. กำหนดให้  $\frac{n}{p} \geq 10$  โดยที่  $p$  คือจำนวนตัวชี้วัด (indicator, manifest variable) ดังนั้นขนาดตัวอย่างขั้นต่ำคือ  $n \geq 10p$  (Nunnally, 1967)
5. กำหนดให้  $\frac{n}{q} \geq 5, 10$  โดยที่  $q$  คือจำนวนพารามิเตอร์ที่ต้องประมาณ (คือจำนวนสัมประสิทธิ์เส้นทาง ( $\beta$ ) จำนวน loading ( $\lambda$ ) และจำนวนส่วนเหลือ ( $\delta$ ) รวมกัน) ดังนั้นขนาดตัวอย่างขั้นต่ำคือ  $n \geq 5q$  หรือ  $n \geq 10q$  (Bentler, 1989; Bollen, 1989)

ขนาดตัวอย่างตามกฎอย่างง่าย (rule of thumb) อาจมีผลกระทบหรือเกี่ยวพันไปถึงคำถามว่าขนาดตัวอย่างที่พอดีคืออะไร ขนาดตัวอย่างอาจสูงเกินไป (over sample size) หรือต่ำเกินไป (under sample size) ขนาดตัวอย่างที่ต่ำเกินไปอาจส่งผลกระทบต่อผลการวิเคราะห์ตัวแบบสมการโครงสร้างในเรื่องความถูกต้อง (accuracy) และความน่าเชื่อถือ (reliability) ของผลการศึกษาและของค่าสถิติเกี่ยวกับ model fit

### วิธีกำหนดขนาดตัวอย่าง

การวิจัยเพื่อหาขนาดตัวอย่างที่เหมาะสมมีมานานแล้ว โดยศึกษาด้วยการกำหนดสถิติตัดสินใจแตกต่างกันไป เช่น ตัดสินใจด้วย non-convergence และ improper solution หรือตัดสินใจด้วย non-convergence ratio, chi-square fit และ power of test หรือตัดสินใจด้วย non-convergence, improper solution, Biasness และ Power of test หรือตัดสินใจด้วย Power of test ในที่นี้จะแสดงการกำหนดขนาดตัวอย่าง 4 แนวทางดังกล่าว

#### 1. การกำหนดขนาดตัวอย่างให้สอดคล้องกับ **proper solution** และ **non-convergence**

Marsh et al. (1998) ศึกษาเปรียบเทียบวิธีกำหนดขนาดตัวอย่างด้วยการจำลองแบบรวมรวมตัวอย่างทั้งสิ้น 30,000 ตัวอย่าง โดยแปรค่าขนาดตัวอย่างจาก 50 ถึง 1,000 แปรค่าจำนวนตัวชี้วัดเฉลี่ยต่อ 1 ตัวแปรแฝง (indicator to construct,  $\frac{f}{p}$ ) จาก 2 ถึง 12 โดยที่  $p$  = จำนวนตัวชี้วัดทั้งหมด  $f$  = จำนวนตัวแปรแฝงทั้งหมด แล้วนับความถี่และหาค่าร้อยละของตัวอย่างที่มีปัญหา non-convergent และปัญหา improper solution

Non-convergence คือสถานการณ์ที่โปรแกรมไม่สามารถประมาณค่าตัวแบบและให้ได้คำตอบในรอบติดกันต่างกันอย่างน้อยกว่าทศนิยมหลักที่ 3 หรือไม่สามารถประมาณค่าตัวแบบได้ภายในจำนวนการวนซ้ำ (iteration) ที่กำหนดให้คือ 250 รอบ (Boomsma and Hoogland, 2001) หรือไม่สามารถประมาณค่าตัวแบบได้ภายในจำนวนการวนซ้ำ (iteration) ที่กำหนดให้คือไม่เกิน 3 เท่าของจำนวนพารามิเตอร์ (Joreskog and Sorbom, 1988)

Improper solution คือสถานการณ์ที่ตัวแบบที่ประมาณค่าได้แล้วมีค่า residual variance หรือ variance เป็นปริมาณลบ (เรียกว่า Heywood case) หรือ loading มีค่ามากกว่า 1 หรือเป็นปริมาณลบ หรือ residual variance มีค่าใหญ่มาก

หมายเหตุ เหตุการณ์นี้ถ้าพบในทางปฏิบัติอาจแก้ปัญหาได้ 3 วิธีคือตัดหน่วยวิเคราะห์ทิ้ง หรือตัดตัวชี้วัดทิ้งหรือปรับปรุงตัวแบบ

ผลการศึกษาของ Marsh et al. (1998) ปรากฏดังตาราง 1 ซึ่งแสดงสัดส่วนของ proper solution และสัดส่วนของ non-convergence จำแนกตามขนาดตัวอย่างและจำนวนตัวชี้วัดถ่วงเฉลี่ยต่อตัวแปรแฝงคือ  $\frac{\text{ยอดรวมจำนวนตัวชี้วัด (p)}}{\text{ยอดรวมจำนวนตัวแปรแฝง (f)}}$  โดยทดลองให้มีจำนวนจากน้อยไปหามากคือ 2, 3, 4, 6, 12 ตัว ขนาดตัวอย่างกำหนดให้มีจากขนาดเล็กไปหาขนาดใหญ่คือ 50, 100, 200, 400 และ 1,000

พบว่าขนาดตัวอย่าง  $n = 50$  เป็นขนาดที่เพียงพอถ้ามีตัวชี้วัด 4 ตัวขึ้นไป แต่ถ้ามีตัวชี้วัดมากกว่า 4 ตัวก็จะนับเป็นขนาดตัวอย่างที่ดียิ่งขึ้น สังเกตว่าสัดส่วนของ proper solution มีค่ามากขึ้นจนถึงร้อยละ 86.5 และมีสัดส่วนของ non-convergence เพียงร้อยละ 2.2 เมื่อมีตัวชี้วัด 4 ตัว และเมื่อกำหนดให้มีตัวชี้วัดมากกว่า 4 ตัวสัดส่วนของ proper solution นี้จะใกล้ร้อยละ 100 และสัดส่วนของ non-convergence จะเข้าใกล้ 0 และสำหรับตัวอย่างที่ใหญ่กว่า 50 คือ 100, 200, 400 และกำหนดให้มีตัวชี้วัดตั้งแต่ 4 ตัวขึ้นไปถือได้ว่าเป็นขนาดตัวอย่างที่เหมาะสม ซึ่งถ้าจะกำหนดให้มีตัวชี้วัดตั้งแต่ 3 ตัวขึ้นไปก็นับว่าเพียงพอ สังเกตได้ว่าเมื่อมีตัวชี้วัดมากขึ้นสัดส่วนของ proper solution จะสูงขึ้นขณะที่สัดส่วนของ non-convergence จะลดลง ขนาดตัวอย่างที่มากกว่าจะเหมาะสมกว่า แต่จะกำหนดให้มีขนาดตัวอย่างเท่ากับ 100 หรืออย่างมาก 200 ก็เพียงพอแล้ว และพบว่าจำนวนตัวชี้วัดเฉลี่ยต่อตัวแปรแฝง ( $\frac{p}{f}$ ) และขนาดตัวอย่างจะชดเชยกันได้คือถ้ามีจำนวนตัวชี้วัดเฉลี่ยต่อตัวแปรแฝง ( $\frac{p}{f}$ ) มากขึ้นก็สามารถลดขนาด  $n$  ลงได้ กลับกันถ้ามีจำนวนตัวชี้วัดเฉลี่ยต่อตัวแปรแฝง ( $\frac{p}{f}$ ) น้อยควรเพิ่ม  $n$  ขึ้น แต่ไม่แนะนำให้กำหนดให้มีจำนวน

ตัวชี้วัดเฉลี่ยต่อตัวแปรแฝง ( $\frac{P}{F}$ ) น้อยๆ ถ้าเป็นไปได้ควรเลือกให้สูงทั้ง 2 กรณีคือ กำหนดให้มีขนาดตัวอย่างมากและมีตัวชี้วัดมากจะเหมาะสมกว่า (Marsh et al., 1998)

นักวิจัยอาจมีคำถามว่าควรกำหนดจำนวนตัวชี้วัดเฉลี่ยต่อตัวแปรแฝงเท่าไร ซึ่งก็คือคำถามว่าควรมีคำถามรวมกันกี่ข้อ ให้นักวิจัยเลือกดูจากตาราง 1 ต่อไปนี้ ผู้เขียนให้ความคิดไว้ว่าการมีจำนวนข้อถามมากและไม่มีคำถามซ้ำในตัวแปรแฝงเดียวกันและไม่ซ้ำซ้อนกับคำถามในตัวแปรแฝงอื่น (นี่คือต้นทางของ CFA) นักวิจัยจะถามกี่ข้อก็ได้ จนกว่าจะสามารถใช้วัดผลตัวแปรแฝงได้ครบถ้วน เช่นมีข้อถาม 15 ข้อใน New Ecological Paradigm scale, NEP มีข้อถาม 20 ข้อใน short form Minnesota Satisfaction Questionnaire (MSQ) ซึ่งหากนักวิจัยเห็นว่าตัวชี้วัดมากเกินไปในแต่ละตัวแปรแฝงก็มีทางออกคือจัดตัวชี้วัดเป็นกลุ่มๆ เรียกว่า parcel แล้วใช้ค่าเฉลี่ยตัวชี้วัดหรือใช้ factor score ในแต่ละกลุ่มเป็นตัวแทนกลุ่มซึ่งเป็นการลดตัวชี้วัดไปในขณะเดียวกัน ซึ่งก็คือการออกแบบการวิจัยเป็น second order model หรือ third order model นั่นเอง ในกรณีทั้งสองนี้ก็มีวิธีจัดการกระบวนการวิเคราะห์ของตนเองอยู่ (มนตรี พิริยะกุล, 2556)

ตาราง 1 ร้อยละของ *proper solution* และ *non-convergence* จำแนกตามขนาดตัวอย่าง และจำนวนตัวชี้วัดเฉลี่ย

จำนวนตัวชี้วัดเฉลี่ย	ชนิดของผลการศึกษา	ขนาดตัวอย่าง				
		50	100	200	400	1,000
2	Proper solution	13.6	32.8	55.6	82.4	93.0
	Non-convergence	56.6	33.1	12.6	2.4	-
3	Proper solution	54.8	85.4	97.8	100.0	99.0
	Non-convergence	10.8	0.9	-	-	-
4	Proper solution	86.5	99.1	99.6	100.0	100.0
	Non-convergence	2.2	-	-	-	-
6	Proper solution	99.6	100.0	100.0	100.0	100.0
	Non-convergence	0.1	-	-	-	-
12	Proper solution	100.0	100.0	100.0	100.0	100.0
	Non-convergence	-	-	-	-	-

ที่มา. ดัดแปลงจาก Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. by Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). *Multivariate Behavioral Research*, p. 24

## 2. การกำหนดขนาดตัวอย่างให้สอดคล้องกับ convergence, power of test และระดับนัยสำคัญของ chi-square test

Koran (2017) ทำการศึกษาด้วยวิธีจำลองแบบเพื่อเปรียบเทียบวิธีกำหนดขนาดตัวอย่าง 7 วิธีโดยใช้ convergence, power of test และระดับนัยสำคัญของ chi-square test กำหนดให้มีสัดส่วนของ convergence เท่ากับร้อยละ 99 กำหนดระดับวิกฤติ ( $\alpha$ ) ของตัวสถิติ chi-square test ไว้ที่ระดับ 0.05 และกำหนดระดับ power of test ไว้ที่ร้อยละ 88 วิธีกำหนดขนาดตัวอย่างคือ

$$1. n = 8.2p$$

$$2. n = 10p$$

$$3. n = 2.7q$$

$$4. n = 10q$$

$$5. W1 = 6.3 * \left(\frac{p}{f}\right)^2 - 7.4 \left(\frac{p}{f}\right) + 665.9 \text{ จาก Westland (2010)}$$

$$6. W2 = 50 * \left(\frac{p}{f}\right)^2 - 405 \left(\frac{p}{f}\right) + 1,100 \text{ จาก Westland (2010)}$$

$$7. n = \alpha - \frac{\beta}{\frac{\beta}{\alpha - \gamma} + \left(\frac{p}{f} - \delta\right)^{\epsilon}} \text{ โดยกำหนดให้}$$

$$\alpha = \frac{(f - 9.25)^2}{4 * 0.18} + 990$$

$$\beta = 10^{[-1.06 + 36.717 * \frac{1}{f} + 6.491a - 27.202(\frac{a}{f})]}$$

$$\gamma = 398 + 375a + 2a * (f - 4.5)^2$$

$$\delta = 2.783 + 10.507\left(\frac{1}{a * f}\right)$$

โดยที่ a = จำนวน loading และ f = จำนวนตัวแปรแฝง

การทดลองดำเนินการโดยแปรค่า loading เป็น 2 ค่าคือ 0.4 กับ 0.8 เพื่อแสดงค่าน้อยกับค่ามาก ค่าเฉลี่ยจำนวนชี้วัด ( $\frac{p}{f}$ ) ต่อ 1 ปัจจัย (ตัวแปรแฝง) เท่ากับ 2, 3, 4, 5, 6, 7, 12 และจำนวนปัจจัยเท่ากับ 3, 6, 12, 16 ทำให้ได้จำนวนพารามิเตอร์ (q) เท่ากับ 21, 27, 33, 39, 45, 75 ตัว

ผลการจำลองแบบปรากฏดังตาราง 2 ต่อไปนี้ การเปรียบเทียบระหว่าง 7 วิธีตัดสินด้วย standard error of estimate (SEE) ผลการเปรียบเทียบพบว่าวิธีที่ 7 มี SEE ค่าที่ต่ำสุด วิธี

ทำให้ SEE ต่ำอันดับรองลงมาอีก 4 อันดับคือวิธีกำหนดให้  $n = 2.7q$  วิธี  $n = 8.2p$  และวิธี  $n = 10p$  ซึ่งเป็นวิธี rule of thumb ทั้ง 3 วิธี ส่วนวิธีของ Westland วิธีที่ 1 คือ

$W1 = 6.3 * \left(\frac{p}{f}\right)^2 - 7.4 \left(\frac{p}{f}\right) + 665.9$  ก็ให้ค่า SEE ไม่สูงมาก วิธีที่เหลืออีก 2 วิธีคือวิธีกำหนดให้  $n = 10q$  และ  $W2 = 50 * \left(\frac{p}{f}\right)^2 - 405 \left(\frac{p}{f}\right) + 1,100$  ให้ค่า SEE สูงมากเกินไป ขอให้สังเกตว่าวิธี  $n = 10q$  ให้ค่า SEE สูงมากแบบก้าวกระโดด

เฉพาะผลการศึกษาของ Koran สรุปกว้างๆ ได้ว่าขนาดตัวอย่างขั้นต่ำจะมากขึ้นถ้ามีจำนวนปัจจัย นักวิจัยสามารถเลือกใช้วิธีกำหนดขนาดตัวอย่างได้หลากหลายวิธีและหลายองค์ประกอบของตัวแบบจากตาราง 2

**ตาราง 2** ขนาดตัวอย่างจำแนกตามสูตรการกำหนดขนาดตัวอย่างและจำนวนปัจจัย ค่า loading จำนวนข้อถามเฉลี่ยต่อปัจจัย จำนวนข้อถามและจำนวนพารามิเตอร์ในตัวแบบ

f	a	p/f	p	q	C	2.7q	8.2p	10p	W2	W1	10q
					ที่ 1	ที่ 2	ที่ 3	ที่ 4	ที่ 5	ที่ 6	ที่ 7
3	0.4	3	9	21	754	58	74	90	491	200	210
3	0.4	4	12	27	563	74	98	120	458	100	270
3	0.4	5	15	33	388	91	123	150	437	100	330
3	0.4	6	18	39	293	107	147	180	429	200	390
3	0.4	7	21	45	260	123	172	210	434	400	450
3	0.4	12	36	75	250	205	294	360	647	2900	750
3	0.8	2	6	15	379	41	49	60	537	400	150
3	0.8	3	9	21	169	58	74	90	491	200	210
3	0.8	4	12	27	110	74	98	120	458	100	270
3	0.8	5	15	33	103	91	123	150	437	100	330
3	0.8	6	18	39	102	107	147	180	429	200	390
3	0.8	7	21	45	102	123	172	210	434	400	450
3	0.8	12	36	75	250	205	294	360	647	2900	750
6	0.4	3	18	51	999	140	147	180	491	200	510
6	0.4	4	24	63	953	173	196	240	458	100	630
6	0.4	5	30	75	565	205	245	300	437	100	750
6	0.4	6	36	87	255	238	294	360	429	200	870
6	0.4	7	42	99	250	271	343	420	434	400	990

ตาราง 2 (ต่อ)

f	a	p/f	p	q	C	2.7q	8.2p	10p	W2	W1	10q
					ที่ 1	ที่ 2	ที่ 3	ที่ 4	ที่ 5	ที่ 6	ที่ 7
6	0.4	12	72	159	1002	435	588	720	647	2900	1590
6	0.8	2	12	39	665	107	98	120	537	400	390
6	0.8	3	18	51	161	140	147	180	491	200	510
6	0.8	4	24	63	102	173	196	240	458	100	630
6	0.8	5	30	75	102	205	245	300	437	100	750
6	0.8	6	36	87	102	238	294	360	429	200	870
6	0.8	7	42	99	139	271	343	420	434	400	990
6	0.8	12	72	159	1004	435	588	720	647	2900	1590
12	0.4	3	36	138	986	377	294	360	491	200	1380
12	0.4	4	48	162	428	443	392	480	458	100	1620
12	0.4	5	60	186	294	508	490	600	437	100	1860
12	0.4	6	72	210	350	574	588	720	429	200	2100
12	0.4	7	84	234	977	639	686	840	434	400	2340
12	0.4	12	144	354	1001	967	1176	1440	647	2900	3540
12	0.8	2	24	114	540	312	196	240	537	400	1140
12	0.8	3	36	138	191	377	294	360	491	200	1380
12	0.8	4	48	162	189	443	392	480	458	100	1620
12	0.8	5	60	186	192	508	490	600	437	100	1860
12	0.8	6	72	210	592	574	588	720	429	200	2100
12	0.8	7	84	234	969	639	686	840	434	400	2340
12	0.8	12	144	354	1001	967	1176	1440	647	2900	3540
16	0.4	2	32	184	1053	503	262	320	537	400	1840
16	0.4	3	48	216	1006	590	392	480	491	200	2160
16	0.4	4	64	248	356	678	523	640	458	100	2480
16	0.4	5	80	280	355	765	654	800	437	100	2800
16	0.4	6	96	312	999	852	784	960	429	200	3120
16	0.4	7	112	344	1053	940	915	1120	434	400	3440
16	0.8	2	32	184	534	503	262	320	537	400	1840
16	0.8	3	48	216	310	590	392	480	491	200	2160
16	0.8	4	64	248	310	678	523	640	458	100	2480
16	0.8	5	80	280	348	765	654	800	437	100	2800
16	0.8	6	96	312	932	852	784	960	429	200	3120
16	0.8	7	112	344	1046	940	915	1120	434	400	3440
				SEE	200.5	308.7	314.5	327.1	333.3	877.3	1320.9

ที่มา. ดัดแปลงมาจาก Preliminary Proactive Sample Size Determination for Confirmatory Factor Analysis Models. By Koran, J., p. 28-30, Measurement and Evaluation in Counseling and Development (Jan 2017), pp.1-30.



### 3. การกำหนดขนาดตัวอย่างให้สอดคล้องกับ biasness และ power of test

Wolf et al. (2013) ทำการศึกษาวิธีกำหนดขนาดตัวอย่างด้วยการจำลองแบบโดยใช้ biasness และ power of test เป็นเกณฑ์ตัดสินใจ

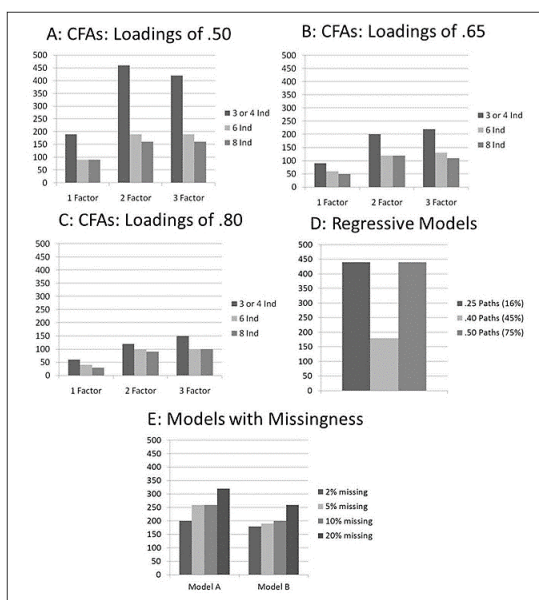
$$\text{Bias}(\hat{\beta}_i) = \sum_{j=1}^{n_r} \frac{\hat{\beta}_{ij} - \beta_i}{n_r}$$

คือค่าเฉลี่ยจากจำนวนการวนซ้ำ  $n_r$  รอบของอัตรา

ความเอนเอียงของค่าประมาณสัมประสิทธิ์เส้นทาง

Power of test  $W = \text{Prob}\left(\frac{|\hat{\beta}|}{\sigma_{\hat{\beta}}} > t_{.05}\right)$  คือ โอกาสที่จะปฏิเสธสมมุติฐานหลักที่ไม่เป็นจริง (Kock and Hadaya, 2018) Wolf et al. (2013) กำหนดให้มี power of test ไม่ต่ำกว่าร้อยละ 80 ณ ระดับ  $\alpha = 0.05$

การทดลองกำหนดให้ loading แปรค่าไป 3 ค่าคือ 0.50, 0.65, 0.80 จำนวนตัวชี้วัดมี 3 ระดับคือ 4, 6, 8 ตัว จำนวนปัจจัยเท่ากับ 1, 2, 3 และอัตราการสูญหายของข้อมูล (missing data) เท่ากับร้อยละ 2, 5, 10, 20 ผลการศึกษาปรากฏดังภาพ



ภาพ 1 ขนาดตัวอย่างจำแนกตามขนาดของ loading จำนวนตัวชี้วัดเฉลี่ยและชนิดตัวแบบที่มา. จาก Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety, by Wolf, et al., p. 934, Educ Psychol Meas, December 2013; 76(6)

ผลการศึกษาของ Wolf et al. (2017) พบว่าเมื่อกำหนดให้มีจำนวนตัวชี้วัดมากขึ้น จะใช้ขนาดตัวอย่างน้อยลงดังภาพ A, B, C ยกเว้นกรณีมีข้อมูลสูญหายมากกว่าจะต้องมีขนาดตัวอย่างที่ใหญ่กว่า (ภาพ E model A สำหรับกรณี CFA ส่วนภาพ E model B สำหรับกรณี SEM) มีข้อสรุปได้ว่าขนาดตัวอย่างเท่ากับ 100, 150, 200 มากพอสำหรับกรณีที่มีตัวชี้วัด 3-8 ตัวต่อตัวแปรแฝง 1 ตัวและ loading สูงกว่า 0.50 (เกณฑ์คือ loading ต้องสูงไม่น้อยกว่า 0.707 อาจต่ำกว่านี้ได้คือไม่ต่ำกว่า 0.50 แต่ AVE ต้องไม่ต่ำกว่า 0.50) และเป็นกรอบการวิจัยขนาดเล็กมีตัวแปรแฝง 2, 3 ตัว

#### 4. การกำหนดขนาดตัวอย่างให้สอดคล้องกับ power of test

Kock and Hayada (2018) ทำการศึกษาโดยการพัฒนาสูตรกำหนดขนาดตัวอย่างให้สอดคล้องกับ power of test และปรับสูตรให้สามารถใช้ได้ง่าย สามารถคำนวณได้ด้วยเครื่องคำนวณหรือด้วยโปรแกรมเอกเซล ซึ่งผู้เขียนจะแสดงตารางให้เห็นต่อไป

Power of test นิยามได้ดังนี้คือ

จาก  $\beta$  และ  $\sigma_\beta$  ซึ่งเป็นค่าสัมประสิทธิ์เส้นทางและ SEE พบว่า  $t = \frac{\beta}{\sigma_\beta}$  จะมีค่าสูง

ถ้า  $\beta$  มีค่าสูงและ/หรือ  $\sigma_\beta$  มีค่าต่ำและมีผลให้  $\Pr\left(\frac{\beta}{\sigma_\beta} > t_{.05}\right)$  มีค่าสูงไปด้วย

เราเรียก  $W = \Pr\left(\frac{|\beta|}{\sigma_\beta} > t_{.05}\right) = \Pr\left(\frac{|\beta|}{\sigma_\beta} - t_{.05} > 0\right)$  เมื่อ  $\beta \neq 0$  ว่า power of test

กำหนดให้  $W = 0.80$  ดังนั้น  $\Phi\left(\frac{|\beta|}{\sigma_\beta} - t_{.05}\right) > 0.80$  หรือ  $\left(\frac{|\beta|}{\sigma_\beta} - t_{.05}\right) > Z_{0.80}$

ถ้าตัวอย่างมีขนาดใหญ่สามารถปรับเป็น  $\left(\frac{|\beta|}{\sigma_\beta} - z_{.95}\right) > Z_{0.80}$  ดังนั้น

$$\frac{|\beta|}{\sigma_\beta} > Z_{0.80} + Z_{.95} \quad (1)$$

แต่ SEE ของ standardized regression coefficient มีค่าประมาณ  $\frac{1}{\sqrt{n}}$  (Petraitis, Dunham, and Niewiarowski, 1996) (นี่คือที่มาของชื่อว่า inverse square root )

ดังนั้นสมการ (1) สามารถเสนอได้เป็น  $\frac{|\beta|}{1} > Z_{0.80} + Z_{.95}$  โดยที่  $Z_{.80} = 0.842$ ,

$Z_{.95} = 1.645$

$$\text{ดังนั้น } n = \frac{2.487^2}{|\beta_{\min}|^2} \quad (2)$$

สมการ (2) เป็นสมการสำหรับกำหนดขนาดตัวอย่างที่มีผลให้ power of test มีค่าไม่ต่ำกว่าร้อยละ 80 มีระดับความคลาดเคลื่อน  $\alpha$  ไม่เกินร้อยละ 5 เรียกวิธีนี้ว่าวิธี inverse square root

และจากสมการ (1) เราสามารถประมาณค่า SEE ของ standardized regression coefficient ได้เป็น  $s = \frac{1}{\sqrt{n}} e^{-\left(\frac{e^{|\beta|}}{\sqrt{n}}\right)}$  (Kock and Hadaya, 2018)

$$\text{ดังนั้นสมการ (1) จึงเปลี่ยนรูปเป็น } |\beta|_{\min} \sqrt{n} e^{\left(\frac{e^{|\beta|}}{\sqrt{n}}\right)} > 2.487 \quad (3)$$

จัดรูปสมการ (3) ได้เป็น

$$\ln\sqrt{n} + \frac{e^{|\beta|_{\min}}}{\sqrt{n}} > \ln\left(\frac{2.487}{|\beta|_{\min}}\right) \quad (4)$$

เราเรียกสมการ (4) ว่า Gamma-Exponential method การกำหนดขนาดตัวอย่างตามวิธีนี้สามารถคำนวณได้ด้วยโปรแกรมเอกเซลโดยกำหนดค่า  $|\beta|_{\min}$  และเงื่อนไข  $\ln\left(\frac{2.487}{|\beta|_{\min}}\right)$  แล้วแปรค่าของ  $n$  ไปเรื่อยๆจนกระทั่งพบค่า  $n$  ที่ตรงกับเงื่อนไข

ตารางกำหนดขนาดตัวอย่างทั้ง 2 สูตรปรากฏดังนี้

ตาราง 3 ขนาดตัวอย่างตามสูตร inverse square root method

Standardized path coefficient	n
0.05	2474
0.10	619
0.15	275
0.20	155
0.25	99
0.30	69
0.35	50
0.40	39
0.45	31
0.50	25
0.55	20
0.60	17
0.65	15
0.70	13
0.75	11
0.80	10

จากตาราง 3 จะสังเกตเห็นว่าขนาดตัวอย่างจะเล็กลงตามค่าสัมประสิทธิ์เส้นทางที่สูงขึ้น ค่าสัมประสิทธิ์เส้นทางสามารถมีได้มากเท่าจำนวนเส้นทางในภาพเส้นทางซึ่งทราบได้จากการทบทวนวรรณกรรม

ตาราง 4 ขนาดตัวอย่างตามสูตร *Gamma-Exponential method*

beta	เงื่อนไข	beta	เงื่อนไข	beta	เงื่อนไข	beta	เงื่อนไข	beta	เงื่อนไข
0.1	3.213662	0.15	2.808197	0.2	2.520515	0.25	2.297372	0.3	2.11505
n	LH	n	LH	n	LH	n	LH	n	LH
1	1.105171	1	1.161834	1	1.221403	1	1.284025	1	1.349859
2	1.128047	2	1.168114	2	1.210236	2	1.254517	2	1.301068
3	1.187377	3	1.220091	3	1.254483	3	1.290639	3	1.328647
4	1.245733	4	1.274064	4	1.303849	4	1.335164	4	1.368077
5	1.298966	5	1.324307	5	1.350947	5	1.378953	5	1.408394
41	2.029385	41	2.038234	41	2.047537	41	2.057317	41	2.067599
42	2.039366	42	2.04811	42	2.057301	42	2.066964	42	2.077123
43	2.049137	43	2.057778	43	2.066862	43	2.076412	43	2.086452
44	2.058706	44	2.067248	44	2.076228	44	2.085669	44	2.095594
45	2.068084	45	2.076527	45	2.085407	45	2.094742	45	2.104556
46	2.077269	46	2.085624	46	2.094407	46	2.103644	46	2.113347
47	2.086284	47	2.094545	47	2.103234	47	2.112368	47	2.121971
47	2.086284	47	2.094545	47	2.103234	47	2.112368		
72	2.268579	72	2.275257	72	2.282277	72	2.289657		
74	2.280506	74	2.287093	74	2.294017	74	2.301297		
124	2.509388	124	2.514477	124	2.519826				
125	2.513006	125	2.518074	125	2.523402				
126	2.516597	126	2.521645						
236	2.803856	236	2.807545						
237	2.805819	237	2.809499						
238	2.807773								
559	3.209818								
564	3.214063								

ตาราง 4 คำนวณได้จากสมการ (4) โดยผู้เขียนแสดงไว้สำหรับค่าสัมประสิทธิ์เส้นทางเพียง 5 ค่า ค่าเหล่านี้เป็นเพียงตัวอย่างที่ผู้เขียนกำหนดจากการคาดหมายว่าค่าของ

สัมประสิทธิ์เส้นตรงที่ต่ำที่สุดไม่ควรสูงกว่าค่าเหล่านี้ หากนักวิจัยพบว่าสัมประสิทธิ์เส้นตรงแตกต่างไปจากค่าตามตารางนี้ก็ขอให้ใช้ค่านั้นคำนวณหาขนาดตัวอย่าง

ตาราง 4 ผู้เขียนย่อตารางเอกเซลเหลือเพียงให้พอดี 1 หน้า คำว่าเงื่อนไขหมายถึงด้านขวาของสมการ (4) คือ  $\ln\left(\frac{2.487}{|\beta|_{\min}}\right)$  และ LH (left hand side) หมายถึงด้านซ้ายมือของสมการ (4) คือ  $\ln\sqrt{n} + \frac{e^{|\beta|_{\min}}}{\sqrt{n}}$  จากนั้นให้แปรค่า  $n$  ไปจนกระทั่งพบว่าผลการคำนวณคือขนาดตัวอย่างที่ปรากฏใน LH มีค่ามากกว่าค่าของเงื่อนไข

สูตรทั้ง 2 ข้างต้นให้คำตอบต่างกัน สูตรตาม inverse square root คำนวณได้ง่ายกว่า ขอให้สังเกตด้วยว่าขนาดตัวอย่างจากทั้ง 2 สูตรจะลดลงเมื่อค่าสัมประสิทธิ์เส้นตรง (จากผลการทบทวนวรรณกรรม) มีค่าสูง เหตุที่ต้องใช้  $|\beta|_{\min}$  เพราะต้องการให้ได้ขนาดตัวอย่างใหญ่ครอบคลุมขนาดตัวอย่างที่สอดคล้องกับ  $\beta$  ที่มีขนาดใหญ่กว่า

ผลจากการทบทวนวรรณกรรมอาจพบว่าค่าสัมประสิทธิ์เส้นตรงของเส้นทางหนึ่ง ๆ มีหลายค่าแตกต่างกันไปตามผลการศึกษาเชิงประจักษ์ซึ่งขึ้นอยู่กับบริบทของการวิจัยนั้นๆ ให้บันทึกค่าเหล่านี้เอาไว้ ภาพกรอบการวิจัยมีหลายเส้นทางเราจึงอาจมีจำนวนสัมประสิทธิ์เส้นตรงเป็นจำนวนมากหลายเท่าของจำนวนเส้นทางในภาพเส้นทาง ให้เลือกเอาค่าที่ต่ำที่สุดจากทั้งหมดมาใช้คำนวณขนาดตัวอย่างตามสูตร  $n = \frac{2.487^2}{|\beta|_{\min}^2}$  หรือ  $\ln\sqrt{n} + \frac{e^{|\beta|_{\min}}}{\sqrt{n}} > \ln\left(\frac{2.487}{|\beta|_{\min}}\right)$  จึงจะมีผลให้ได้ขนาดตัวอย่างที่ใหญ่ที่สุด ทั้งนี้เพื่อป้องกันมิให้เกิดปัญหาจากการใช้ตัวอย่างน้อยเกินไป ขอให้สังเกตจากตาราง 3 ว่าถ้าสัมประสิทธิ์เส้นตรงมีค่าประมาณ 0.35 และที่สูงกว่านี้ขนาดตัวอย่างจะเล็กมากจนอาจไม่เหมาะสมในด้านความถูกต้องและความน่าเชื่อถือของผลการวิจัย

## สรุป

สูตรสำหรับกำหนดขนาดตัวอย่างเพื่อการวิจัยด้วยตัวแบบสมการ โครงสร้างเป็นสิ่งที่นักวิจัยทุกคนถือว่าเป็นเรื่องยุ่งยากและมักเป็นปัญหาต่อเนื่องตั้งแต่เริ่มออกแบบการวิจัยสืบเนื่องไปจนถึงการประเมินผลการวิจัยโดยผู้เชี่ยวชาญ เพราะจะได้รับความถามเสมอว่าขนาดเท่านี้พอเพียงให้ผลการวิเคราะห์ตัวแบบถูกต้อง (accurate) แม่นยำ (precise) และมี

ความน่าเชื่อถือ (reliable) หรือไม่ บ่อยครั้งที่เป็งานถูกหักท้วงจากผู้ทรงคุณวุฒิเพราะไม่มั่นใจในความเหมาะสมของขนาดตัวอย่างที่ใช้ ด้วยเหตุดังกล่าวนักวิจัยจำนวนมากจึงใช้วิธีกำหนดขนาดตัวอย่างให้มีขนาดใหญ่มาก ๆ โดยสุ่มตัวอย่างเพื่อเอาไว้โดยเชื่อว่าน่าจะเหมาะสมในเรื่องความถูกต้องและความเชื่อถือได้ของผลการวิจัย ที่จริงแล้วการใช้ขนาดตัวอย่างที่ใหญ่มากไม่จำเป็นต้องมีผลให้งานวิจัยมีความถูกต้องและความเชื่อถือได้กว่าการใช้ตัวอย่างที่มีขนาดเล็กกว่า ผลการศึกษาในปัจจุบันหรือใกล้ปัจจุบันพบว่าขนาดตัวอย่างที่ใช้เพื่อวิเคราะห์ตัวแบบสมการ โครงสร้างขึ้นอยู่กับองค์ประกอบหลายอย่างคือ ค่าของน้ำหนักปัจจัย (loading) จำนวนปัจจัย (latent variable, construct:  $f$ ) จำนวนตัวชี้วัดเฉลี่ยต่อ 1 ปัจจัย ( $\frac{p}{f}$ ) ขนาดของค่าสัมประสิทธิ์เส้นทาง (จากวรรณกรรม) อัตราข้อมูลสูญหาย และจำนวนพารามิเตอร์ ด้วยเหตุนี้ขนาดตัวอย่างสำหรับการวิจัยด้วย SEM จึงแปรไปได้ นักวิจัยสามารถเลือกใช้สูตรต่าง ๆ ได้ตามความเหมาะสม เช่น ขนาดประชากรที่อาจมีจำนวนสมาชิกไม่มาก หรือองค์ประกอบของตัวแบบที่อาจมีตัวแปรแฝงมากมีตัวชี้วัดมากหรือกลับกัน

ผลการศึกษาโดยรวมพบว่าขนาดตัวอย่างขึ้นอยู่กับจำนวนปัจจัย (คือตัวแปรแฝง) ถ้ามีปัจจัยมากคือเป็นตัวแบบที่มีขนาดใหญ่ขนาดตัวอย่างก็ใหญ่ และขึ้นอยู่กับจำนวนตัวชี้วัดเฉลี่ยต่อ 1 ตัวแปรแฝง (indicator to construct,  $\frac{f}{p}$ ) และค่า loading ถ้าจำนวนตัวชี้วัดเฉลี่ยต่อ 1 ตัวแปรแฝง ( $\frac{f}{p}$ ) มากและ loading มีค่าสูงจะใช้ตัวอย่างขนาดเล็กซึ่งกรณีนี้  $n = 50$  ก็เพียงพอ ถ้าหากไม่คิดขีดว่าประชากรมีขนาดเล็กการกำหนดให้  $n = 200$  ถือได้ว่าเป็นขนาดที่เพียงพอเพราะไม่มีปัญหา non-convergence และ improper solution (Marsh et al., 1998; Boomsma and Hoogland, 2001)

นักวิจัยสามารถกำหนดขนาดตัวอย่างโดยเลือกจากตาราง 1-4 เฉพาะตาราง 4 ผู้เขียนคำนวณไว้เพียงบางระดับของค่าสัมประสิทธิ์เส้นทางที่เห็นว่าเป็นค่าต่ำ นักวิจัยสามารถคำนวณได้เองจากสมการ (4)

### เอกสารอ้างอิง

มนตรี พิริยะกุล (2556). การวิเคราะห์ตัวแบบสมการ โครงสร้างชนิด Second Order Model, วารสารการจัดการ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏรำปาง, ปีที่ 6 ฉบับที่ 1, หน้า 97- 111.

Anderson, T. W.; Rubin, H. (1956). Statistical Inference in Factor Analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry, 111--150, University of California Press, Berkeley, Calif.

Bentler, P. (1989). EQS structural program manual, Los Angelis, CA. BMPD statistical software.

Bollen, K. A. (1989): Structural Equations with Latent Variables, John Wiley & Sons, New York.

Boomsma, A. (1985). Non-convergence, improper solutions, and starting values in LISREL maximum likelihood estimation, in: Psychometrika, 50(2), pp. 229–242.

Boomsma, A., and Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), Structural equation modeling: Present and future (pp. 139–168). Chicago: Scientific Software

Joreskog, K.G. and Sorbom, D. (1988). LISREL7-A guide to program and applications (2<sup>nd</sup> ed.), Chicago, Illinois, SPSS.

Kock, N. (2014). Stable P value calculation methods in PLS-SEM. Laredo, TX: ScriptWarp Systems.

Koch, N. and Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods Info Systems J (2018) 28, pp. 227–261.

Koran, J. (2017). Preliminary Proactive Sample Size Determination for Confirmatory Factor Analysis Models. *Measurement and Evaluation in Counseling And Development* (Jan 2017), pp. 1-30.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. *Multivariate Behavioral Research*, 33, pp. 181-220.

Muthén, L. K. & Muthén, B. O. (2002) How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power, *Structural Equation Modeling: A Multidisciplinary Journal*, 9:4, pp. 599-620.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Petraitis, P. S., Dunham, A. E. and Niewiarowski, P. H. (1996). Inferring multiple causality: the limitations of path analysis, *Functional ESSAY REVIEW Ecology* 1996 10, pp. 421-431.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling, *Electronic Commerce Research and Applications*, pp.476-487.

Wolf, E. J., Harrington, K. M. Clark, S. L. and Miller, M. W. (2013), Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety, *Educ Psychol Meas*, December ; 76(6): 913–934.