



การประมาณจำนวนกลุ่มตัวอย่าง (Sample Size Estimation)



สันทัด พรประเสริฐมานิต



การประมาณค่ากลุ่มตัวอย่าง

- ปัจจัยที่ผลต่อการประมาณค่ากลุ่มตัวอย่าง
- วิธีการจำลองสถานการณ์
- การคำนวณกลุ่มตัวอย่างผ่านการตัดสินโมเดล
- การหาค่าพารามิเตอร์ที่เหมาะสม



ปัจจัยที่ผลต่อการประมาณค่ากลุ่มตัวอย่าง



- ความแม่นยำในการประมาณค่าพารามิเตอร์
 - ยิ่งโมเดลซับซ้อน ยิ่งมีพารามิเตอร์ให้ประมาณการจำนวนมาก ทำให้ต้องการกลุ่มตัวอย่างสูง
 - การกระจายไม่เป็น MVN ทำให้ต้องการกลุ่มตัวอย่างมากขึ้นในการใช้ robust method
 - การกระจายไม่เป็นตัวแปรต่อเนื่อง ต้องการกลุ่มตัวอย่างมากขึ้น เพื่อใช้เทคนิคเฉพาะ
- ขนาดของ SE ให้น้อยจนมีกำลังทางสถิติสูง หรือมีความกว้างของช่วงเชื่อมั่นต่ำ
 - ความเที่ยงขององค์ประกอบน้อย ทำให้ต้องการจำนวนกลุ่มตัวอย่างสูง
 - การมีน้ำหนักองค์ประกอบเด่นในมาตรทุกตัว ดีกว่าการมีน้ำหนักองค์ประกอบเด่นแค่บางตัว ดีกว่าการไม่มีน้ำหนักองค์ประกอบเด่นเลย (MacCallum et al., 1999)
 - จำนวนค่าสูญหายสูง ทำให้ SE จะมีค่าน้อยลง ต้องการกลุ่มตัวอย่างมาชดเชยมากขึ้น

ปัจจัยที่ผลต่อการประมาณค่ากลุ่มตัวอย่าง



- Jackson (2003) ได้ทดสอบจำนวนกลุ่มตัวอย่างต่อพารามิเตอร์ ($N:q$) ว่ามีผลช่วยให้เกิดความแม่นยำในการประมาณค่า Chi-square และ RMSEA แต่เขาแนะนำว่าให้สนใจจำนวนกลุ่มตัวอย่างเต็ม (N) มากกว่า เพราะมีผลมากกว่า
 - Kline (2023) ไปอ้างอิงงาน Jackson (2003) ว่าให้ใช้ $N:q = 20:1$ แต่ผมไม่เจอคำกล่าวนั้นในงาน
- Wolf et al. (2013) แสดงให้เห็นว่าการประมาณค่าจำนวนกลุ่มตัวอย่าง ไม่สามารถใช้คำแนะนำ (Rule of Thumbs) อย่างง่ายได้ และแนะนำให้ใช้การจำลองสถานการณ์ (Simulation Study) ในการหาจำนวนกลุ่มตัวอย่าง

ปัจจัยที่ผลต่อการประมาณค่ากลุ่มตัวอย่าง

- การคำนวณหากลุ่มตัวอย่าง สามารถพิจารณา 4 กรณี
 - โอกาสที่โมเดลสุ่มเข้า ต้องการให้โมเดลสุ่มเข้าเกือบทุกครั้งที่จะวิเคราะห์ข้อมูล
 - ความแม่นยำในการประมาณค่าพารามิเตอร์ ดูจากอคติในการประมาณค่า
 - ขนาดของ SE ที่ส่งผลให้เกิดกำลังทางสถิติ และความกว้างของช่วงเชื่อมั่น
 - โอกาสการปฏิเสธโมเดลที่ผิดพลาด และยอมรับโมเดลที่ถูกต้อง
- วิธีการทั้งหมดนี้สามารถคำนวณได้ผ่านการจำลองสถานการณ์ (Simulation Study)

การจำลองสถานการณ์

- การจำลองสถานการณ์ (Simulation Study) เป็นการกำหนดว่าโมเดลในประชากรเป็นอย่างไร และค่าพารามิเตอร์ในประชากรเป็นอย่างไร
- จากนั้น สร้างข้อมูลจากประชากรหลายชุด (เช่น 1,000 ชุด) แล้ววิเคราะห์ผลโดยโมเดลเป้าหมาย ซึ่งอาจจะเป็นโมเดลในประชากรหรือไม่ก็ได้
- หลังจากนั้น นำผลการวิเคราะห์จากข้อมูลหลายชุดมาสรุปผล
 - แต่ละชุด จะได้ค่าสถิติเหมือนกับการวิเคราะห์ทั่วไป เช่น χ^2 , RMSEA, $\hat{\theta}$, $SE(\hat{\theta})$
- นักวิเคราะห์สามารถปรับจำนวนกลุ่มตัวอย่าง หรือปรับโอกาสเกิดค่าสูญหาย เพื่อดูว่าจำนวนกลุ่มตัวอย่างระดับใดที่ได้ผลตามที่ต้องการ

ปัจจัยที่ผลต่อการประมาณค่ากลุ่มตัวอย่าง

- ค่าที่ใช้ตรวจสอบความแม่นยำในการประมาณค่าพารามิเตอร์ ในการจำลองสถานการณ์
 - อคติโดยสัมบูรณ์ (Absolute Bias)

$$\text{Absolute Bias} = \bar{\hat{\theta}} - \theta$$

- อคติสัมพัทธ์ (Relative Bias) ไม่ควรเกิน 5%

$$\text{Relative Bias} = (\bar{\hat{\theta}} - \theta) / \theta$$

- อคติสัมพัทธ์ใน SE (Relative Bias in SE) ไม่ควรเกิน 10%

$$\text{Relative Bias in SE} = \left(\overline{SE(\hat{\theta})} - SD(\hat{\theta}) \right) / SD(\hat{\theta})$$

ปัจจัยที่ผลต่อการประมาณค่ากลุ่มตัวอย่าง

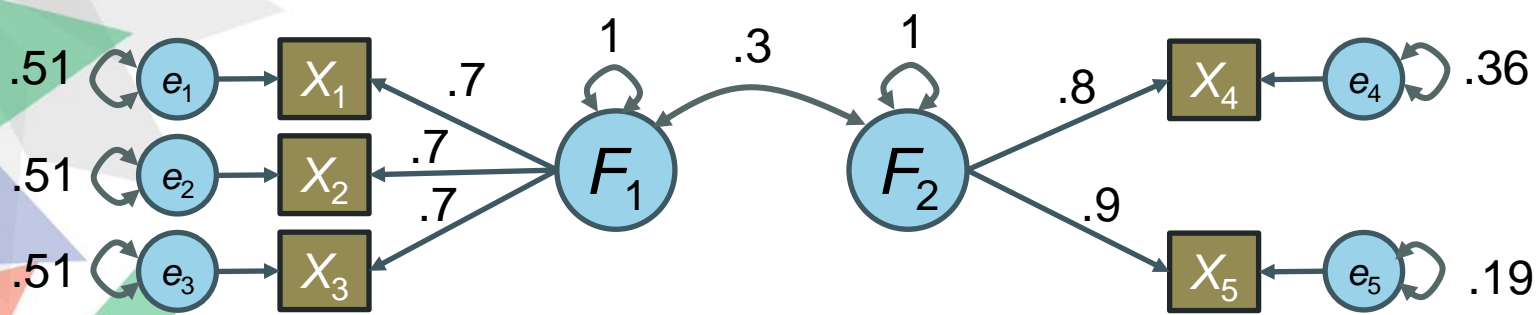
- กำลังในการทดสอบสถิติ จากผลการจำลองสถานการณ์ คือ

$$Power = K(p_{\hat{\theta}} < \alpha) / K$$

- ความกว้างของช่วงเชื่อมั่น จากผลการจำลองสถานการณ์ คือ

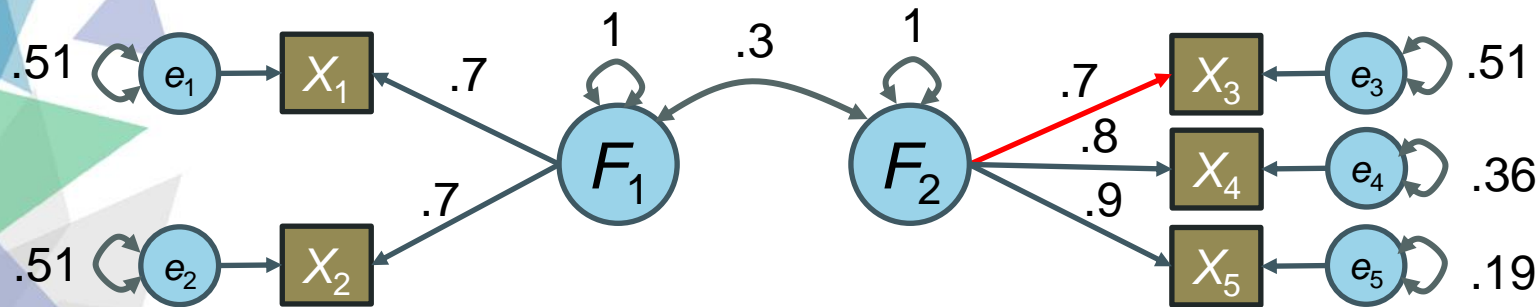
$$(1 - \alpha)\% \text{ CI Width} = \text{Upper Bound} - \text{Lower Bound}$$

$$= \left(\hat{\theta} + z_{1-\frac{\alpha}{2}} SE \right) - \left(\hat{\theta} - z_{1-\frac{\alpha}{2}} SE \right) = 2z_{1-\frac{\alpha}{2}} SE$$



จากโมเดลในประชากรข้างต้น ต้องการตรวจสอบว่าจำนวนกลุ่มตัวอย่างใดที่

1. การวิเคราะห์ห้อย่างน้อย **95%** ระบุเข้าสู่คำตอบ
2. ค่าความสัมพันธ์ระหว่างองค์ประกอบมีอคติไม่เกิน **5%**
3. ค่าความสัมพันธ์ระหว่างองค์ประกอบมีกำลังในการทดสอบทางสถิติเกิน **80%**
4. ค่าความสัมพันธ์ระหว่างองค์ประกอบมีความกว้างของช่วงเชื่อมั่นไม่เกิน **.5**
5. หากประชากรเป็นดังโมเดลด้านล่าง ให้มีโอกาสในการปฏิเสธโมเดลเกิน **80%** โดยใช้ดัชนีความเหมาะสม



```
> library(simsem)
```

```
> genmodel <- "  
f1 =~ 0.7*x1 + 0.7*x2 + 0.7*x3  
f2 =~ 0.8*x4 + 0.9*x5  
f2 =~ 0.3*f1  
f1 =~ 1*f1  
f2 =~ 1*f2  
x1 =~ 0.51*x1  
x2 =~ 0.51*x2  
x3 =~ 0.51*x3  
x4 =~ 0.36*x4  
x5 =~ 0.19*x5  
"
```

สร้างโมเดลในการจำลองข้อมูล ในที่นี้
จะบอกขนาดของพารามิเตอร์แต่ละตัวไว้ในโมเดล

```
> analysismodel <- "  
f1 =~ x1 + x2 + x3  
f2 =~ x4 + x5  
"
```

ใส่โมเดลที่ใช้ในการวิเคราะห์ข้อมูล จากข้อมูลที่สร้างขึ้นมา

คำสั่ง **lavaan** ที่ใช้วิเคราะห์
เป็น "cfa" หรือ "sem"

```
> simout200 <- sim(nRep=1000, model=analysismodel, generate=genmodel, n=200, lavaanfun="cfa", std.lv=TRUE)
```

จำนวนครั้งที่สร้าง
ข้อมูลแล้ววิเคราะห์

ขนาดกลุ่มตัวอย่าง

Argument อะไร
ที่จะใส่ใน cfa ก็สามารใส่ได้

แสดงจุดตัดของดัชนีความเหมาะสม ว่าถ้าต้องการ α เท่ากับเท่าไร

แล้วจุดตัดจะเท่ากับเท่าไร

```
> summary(simout200)
```

RESULT OBJECT

Model Type

[1] "lavaan"

==== Fit Indices Cutoffs =====

	Alpha				Mean	SD
Fit Indices	0.1	0.05	0.01	0.001		
chisq	8.155	9.617	12.928	17.864	4.125	2.875
aic	2609.360	2628.143	2658.588	2695.091	2553.754	45.198
bic	2645.641	2664.425	2694.870	2731.373	2590.035	45.198
rmsea	0.072	0.084	0.106	0.132	0.022	0.031
cfi	0.986	0.981	0.969	0.956	0.996	0.007
tli	0.964	0.953	0.921	0.890	0.999	0.025
srmr	0.036	0.042	0.051	0.058	0.023	0.010

ค่า SE เลื่อนจาก
1,000 ครั้ง $(SE(\hat{\theta}))$

ค่าพารามิเตอร์
ที่ใช้สร้างข้อมูล (θ)

==== Parameter Estimates and Standard Errors =====

	Estimate	Average Estimate	SD	Average SE	Power (Not equal 0)	Std Est	Std Est	SD	Std Ave SE	Average	Param	Average Bias	Coverage
f1~~x1	0.696	0.696	0.078	0.075	1.000	0.697	0.060	0.057	0.70	-0.004	0.933		
f1~~x2	0.693	0.693	0.076	0.075	1.000	0.696	0.058	0.057	0.70	-0.007	0.945		
f1~~x3	0.698	0.698	0.076	0.075	1.000	0.701	0.059	0.057	0.70	-0.002	0.935		
f2~~x4	0.806	0.806	0.151	0.151	0.980	0.809	0.143	0.140	0.80	0.006	0.961		
f2~~x5	0.920	0.920	0.286	0.412	0.980	0.922	0.297	0.421	0.90	0.020	0.962		
x1~~x1	0.506	0.506	0.082	0.079	0.999	0.510	0.083	0.079	0.51	-0.004	0.949		
x2~~x2	0.506	0.506	0.081	0.079	0.999	0.512	0.080	0.079	0.51	-0.004	0.953		
x3~~x3	0.499	0.499	0.082	0.079	1.000	0.506	0.082	0.079	0.51	-0.011	0.937		
x4~~x4	0.322	0.322	0.261	0.249	0.572	0.326	0.266	0.254	0.36	-0.038	0.970		
x5~~x5	0.070	0.070	1.903	3.930	0.268	0.061	2.176	4.476	0.19	-0.120	0.955		
f1~~f2	0.293	0.293	0.088	0.085	0.906	0.293	0.088	0.085	0.30	-0.007	0.952		

ค่าพารามิเตอร์ที่ประมาณค่าได้
เลื่อนจาก 1,000 ครั้ง $(\hat{\theta})$

ค่า SD ของพารามิเตอร์
ที่ประมาณค่าได้จาก 1,000 ครั้ง $(SD(\hat{\theta}))$

อคติในการประมาณ
ค่าพารามิเตอร์ (Absolute Bias)

กำลังในการทดสอบ

ว่าต่างจาก 0 อย่างมีนัยสำคัญ
กี่ % ใน 1,000 ครั้ง

ค่าพารามิเตอร์
มาตรฐานเฉลี่ย

SD ของค่าพารามิเตอร์
มาตรฐาน

SE เฉลี่ยของค่า
พารามิเตอร์มาตรฐาน

==== Parameter Estimates and Standard Errors =====

	Estimate	Average	Estimate	SD	Average	SE	Power (Not equal 0)	Std	Est	Std	Est	SD	Std	Ave	SE	Average	Param	Average	Bias	Coverage
f1~x1	0.696		0.078		0.075		1.000	0.697		0.060		0.057		0.70		-0.004		0.933		
f1~x2	0.693		0.076		0.075		1.000	0.696		0.058		0.057		0.70		-0.007		0.945		
f1~x3	0.698		0.076		0.075		1.000	0.701		0.059		0.057		0.70		-0.002		0.935		
f2~x4	0.806		0.151		0.151		0.980	0.809		0.143		0.140		0.80		0.006		0.961		
f2~x5	0.920		0.286		0.412		0.980	0.922		0.297		0.421		0.90		0.020		0.962		
x1~x1	0.506		0.082		0.079		0.999	0.510		0.083		0.079		0.51		-0.004		0.949		
x2~x2	0.506		0.081		0.079		0.999	0.512		0.080		0.079		0.51		-0.004		0.953		
x3~x3	0.499		0.082		0.079		1.000	0.506		0.082		0.079		0.51		-0.011		0.937		
x4~x4	0.322		0.261		0.249		0.572	0.326		0.266		0.254		0.36		-0.038		0.970		
x5~x5	0.070		1.903		3.930		0.268	0.061		2.176		4.476		0.19		-0.120		0.955		
f1~f2	0.293		0.088		0.085		0.906	0.293		0.088		0.085		0.30		-0.007		0.952		

ช่วงเชื่อมั่นระดับ .95
คลุมค่าพารามิเตอร์ที่ %

=====
Correlation between Fit Indices
=====

	chisq	aic	bic	rmsea	cfi	tli	srmr
chisq	1.000	0.012	0.012	0.949	-0.927	-0.991	0.813
aic	0.012	1.000	1.000	0.016	-0.025	-0.009	-0.006
bic	0.012	1.000	1.000	0.016	-0.025	-0.009	-0.006
rmsea	0.949	0.016	0.016	1.000	-0.938	-0.944	0.746
cfi	-0.927	-0.025	-0.025	-0.938	1.000	0.927	-0.702
tli	-0.991	-0.009	-0.009	-0.944	0.927	1.000	-0.817
srmr	0.813	-0.006	-0.006	0.746	-0.702	-0.817	1.000

ค่าสหสัมพันธ์ระหว่าง

ดัชนีความเหมาะสม

=====
Replications
=====

Number of replications = 1000

Number of converged replications = 744 จำนวนครั้งที่ Convergent ได้ 74.4%

Number of nonconverged replications:

1. Nonconvergent Results = 1
2. Nonconvergent results from multiple imputation = 0
3. At least one SE were negative or NA = 0
4. Nonpositive-definite latent or observed (residual) covariance matrix (e.g., Heywood case or linear dependency) = 255

ให้แสดงรายละเอียดเพิ่มเติม

ไม่ใช่ผลการวิเคราะห์จากผลลัพธ์ที่ไม่ convergent (จำนวนจากผล 744 ครั้ง แทนที่จะ
จำนวนจาก 1,000 ครั้ง) ให้ใช้ผลนี้ แทนที่จะใช้ผลจาก summary ในกรณีที่มี nonconvergent

```
> summaryParam(simout200, detail=TRUE, improper=FALSE)
```

	Estimate	Average Estimate	SD	Average SE	Power (Not equal 0)	Std Est	Std Est	SD	Std Ave SE	Average Param	Average Bias	Coverage
f1=~x1	0.6970982	0.07968484	0.07494684	1.0000000	0.6976029	0.06050454	0.05648238	0.70	-0.0029017706	0.9274194		
f1=~x2	0.6963704	0.07606392	0.07486430	1.0000000	0.6978047	0.05734523	0.05646388	0.70	-0.0036296147	0.9475806		
f1=~x3	0.6995982	0.07520998	0.07490020	1.0000000	0.7012055	0.05774656	0.05641270	0.70	-0.0004017632	0.9422043		
f2=~x4	0.8236899	0.08884775	0.11779228	1.0000000	0.8278387	0.07213091	0.10578581	0.80	0.0236898622	0.9825269		
f2=~x5	0.8707162	0.09025786	0.12190553	1.0000000	0.8728309	0.07237575	0.11005766	0.90	-0.0292837761	0.9758065		
x1~~x1	0.5059369	0.08186117	0.07910535	0.9986559	0.5096944	0.08381703	0.07832120	0.51	-0.0040630648	0.9489247		
x2~~x2	0.5052150	0.08016408	0.07887037	1.0000000	0.5097845	0.07987730	0.07834806	0.51	-0.0047850282	0.9543011		
x3~~x3	0.5005429	0.08144208	0.07903174	1.0000000	0.5049807	0.08032212	0.07866867	0.51	-0.0094571127	0.9408602		
x4~~x4	0.3044406	0.11724528	0.17250976	0.5188172	0.3094873	0.12058532	0.17603008	0.36	-0.0555594343	0.9959677		
x5~~x5	0.2300765	0.12318382	0.18986027	0.2930108	0.2329350	0.12544935	0.19273151	0.19	0.0400765414	0.9771505		
f1~~f2	0.3135581	0.08055679	0.08281016	0.9663978	0.3135581	0.08055679	0.08281016	0.30	0.0135580937	0.9516129		
	Rel Bias	Std Bias	Rel SE	SE Bias	Not Cover	Below Not Cover	Above Average	CI Width	SD	CI Width		
f1=~x1	-0.0041453866	-0.036415593	-0.05945920	0.03629032	0.036290323	0.2937862	0.01668842					
f1=~x2	-0.0051851639	-0.047717958	-0.01577120	0.02284946	0.029569892	0.2934626	0.01660879					
f1=~x3	-0.0005739474	-0.005341886	-0.00411892	0.02822581	0.029569892	0.2936034	0.01668790					
f2=~x4	0.0296123278	0.266634360	0.32577677	0.01612903	0.001344086	0.4617373	0.12724329					
f2=~x5	-0.0325375290	-0.324445719	0.35063618	0.00000000	0.024193548	0.4778609	0.12717901					
x1~~x1	-0.0079667937	-0.049633600	-0.03366462	0.01478495	0.036290323	0.3100873	0.03410162					
x2~~x2	-0.0093824082	-0.059690427	-0.01613833	0.01209677	0.033602151	0.3091662	0.03352816					
x3~~x3	-0.0185433583	-0.116120714	-0.02959574	0.01612903	0.043010753	0.3097987	0.03365714					
x4~~x4	-0.1543317618	-0.473873523	0.47135784	0.00000000	0.004032258	0.6762258	0.25914387					
x5~~x5	0.2109291655	0.325339325	0.54127605	0.02284946	0.000000000	0.7442386	0.25940077					
f1~~f2	0.0451936456	0.168304784	0.02797242	0.04166667	0.006720430	0.3246099	0.01729540					

ค่าอคติสัมพัทธ์
ของการประมาณค่า
พารามิเตอร์
(Relative Bias)

ค่าอคติของค่าพารามิเตอร์
มาตรฐาน (Absolute Bias of
Standardized Parameter
Estimates)

ค่าอคติสัมพัทธ์ของ SE
ของการประมาณค่าพารามิเตอร์
(Relative Bias of SE)

```
> summaryParam(simout200, detail=TRUE, improper=FALSE)
```

	Estimate	Average Estimate	SD	Average SE	Power (Not equal 0)	Std Est	Std Est	SD	Std Ave SE	Average Param	Average Bias	Coverage
f1~x1	0.6970982	0.07968484	0.07494684	1.0000000	0.6976029	0.06050454	0.05648238	0.70	-0.0029017706	0.9274194		
f1~x2	0.6963704	0.07606392	0.07486430	1.0000000	0.6978047	0.05734523	0.05646388	0.70	-0.0036296147	0.9475806		
f1~x3	0.6995982	0.07520998	0.07490020	1.0000000	0.7012055	0.05774656	0.05641270	0.70	-0.0004017632	0.9422043		
f2~x4	0.8236899	0.08884775	0.11779228	1.0000000	0.8278387	0.07213091	0.10578581	0.80	0.0236898622	0.9825269		
f2~x5	0.8707162	0.09025786	0.12190553	1.0000000	0.8728309	0.07237575	0.11005766	0.90	-0.0292837761	0.9758065		
x1~x1	0.5059369	0.08186117	0.07910535	0.9986559	0.5096944	0.08381703	0.07832120	0.51	-0.0040630648	0.9489247		
x2~x2	0.5052150	0.08016408	0.07887037	1.0000000	0.5097845	0.07987730	0.07834806	0.51	-0.0047850282	0.9543011		
x3~x3	0.5005429	0.08144208	0.07903174	1.0000000	0.5049807	0.08032212	0.07866867	0.51	-0.0094571127	0.9408602		
x4~x4	0.3044406	0.11724528	0.17250976	0.5188172	0.3094873	0.12058532	0.17603008	0.36	-0.0555594343	0.9959677		
x5~x5	0.2300765	0.12318382	0.18986027	0.2930108	0.2329350	0.12544935	0.19273151	0.19	0.0400765414	0.9771505		
f1~f2	0.3135581	0.08055679	0.08281016	0.9663978	0.3135581	0.08055679	0.08281016	0.30	0.0135580937	0.9516129		
	Rel Bias	Std Bias	Rel SE Bias	Not Cover	Below Not Cover	Above Average	CI Width	SD	CI Width			
f1~x1	-0.0041453866	-0.036415593	-0.05945920	0.03629032	0.036290323	0.2937862	0.01668842					
f1~x2	-0.0051851639	-0.047717958	-0.01577120	0.02284946	0.029569892	0.2934626	0.01660879					
f1~x3	-0.0005739474	-0.005341886	-0.00411892	0.02822581	0.029569892	0.2936034	0.01668790					
f2~x4	0.0296123278	0.266634360	0.32577677	0.01612903	0.001344086	0.4617373	0.12724329					
f2~x5	-0.0325375290	-0.324445719	0.35063618	0.00000000	0.024193548	0.4778609	0.12717901					
x1~x1	-0.0079667937	-0.049633600	-0.03366462	0.01478495	0.036290323	0.3100873	0.03410162					
x2~x2	-0.0093824082	-0.059690427	-0.01613833	0.01209677	0.033602151	0.3091662	0.03352816					
x3~x3	-0.0185433583	-0.116120714	-0.02959574	0.01612903	0.043010753	0.3097987	0.03365714					
x4~x4	-0.1543317618	-0.473873523	0.47135784	0.00000000	0.004032258	0.6762258	0.25914387					
x5~x5	0.2109291655	0.325339325	0.54127605	0.02284946	0.000000000	0.7442386	0.25940077					
f1~f2	0.0451936456	0.168304784	0.02797242	0.04166667	0.006720430	0.3246099	0.01729540					

สัดส่วนที่พารามิเตอร์
อยู่ต่ำกว่าช่วงเชื่อมั่น

สัดส่วนที่พารามิเตอร์
อยู่สูงกว่าช่วงเชื่อมั่น

ค่าความกว้าง
ของช่วงเชื่อมั่น
ระดับ .95 เหลือ

SD ของความกว้าง
ของช่วงเชื่อมั่น
ระดับ .95

ปรับจำนวนกลุ่มตัวอย่าง เพื่อคว้าจำนวนเท่าไร ถึงจะทำให้ลู่อื่นเกิน 95%

==== Replications =====

Number of replications = 100

Number of converged replications = 88

Number of nonconverged replications:

1. Nonconvergent Results = 0
2. Nonconvergent results from multiple imputation = 0
3. At least one SE were negative or NA = 0
4. Nonpositive-definite latent or observed (residual) covariance matrix (e.g., Heywood case or linear dependency) = 12

$n = 400, nRep = 100$

==== Replications =====

Number of replications = 100

Number of converged replications = 86

Number of nonconverged replications:

1. Nonconvergent Results = 0
2. Nonconvergent results from multiple imputation = 0
3. At least one SE were negative or NA = 0
4. Nonpositive-definite latent or observed (residual) covariance matrix (e.g., Heywood case or linear dependency) = 14

$n = 600, nRep = 100$

==== Replications =====

Number of replications = 100

Number of converged replications = 93

Number of nonconverged replications:

1. Nonconvergent Results = 0
2. Nonconvergent results from multiple imputation = 0
3. At least one SE were negative or NA = 0
4. Nonpositive-definite latent or observed (residual) covariance matrix (e.g., Heywood case or linear dependency) = 7

$n = 800, nRep = 100$

==== Replications =====

Number of replications = 100

Number of converged replications = 98

Number of nonconverged replications:

1. Nonconvergent Results = 0
2. Nonconvergent results from multiple imputation = 0
3. At least one SE were negative or NA = 0
4. Nonpositive-definite latent or observed (residual) covariance matrix (e.g., Heywood case or linear dependency) = 2

$n = 1000, nRep = 100$

ใช้ $nRep$ น้อยลง เพื่อประหยัดเวลา

ถ้ากลัวว่าเก็บข้อมูลมาแล้วเจอผลไม่

Convergent ต้องใช้กลุ่มตัวอย่าง

1000 คน นักวิจัยอาจปรับโมเดลโดย

หลีกเลี่ยงองค์ประกอบที่มี **2** ตัวบ่งชี้

หรือเพิ่มจำนวนองค์ประกอบ เพื่อให้

ต้องการจำนวนกลุ่มตัวอย่างน้อยกว่านี้


```

> genalternativemodel <- "
+ f1 =~ 0.7*x1 + 0.7*x2
+ f2 =~ 0.7*x3 + 0.8*x4 + 0.9*x5
+ f2 ~~ 0.3*f1
+ f1 ~~ 1*f1
+ f2 ~~ 1*f2
+ x1 ~~ 0.51*x1
+ x2 ~~ 0.51*x2
+ x3 ~~ 0.51*x3
+ x4 ~~ 0.36*x4
+ x5 ~~ 0.19*x5
+ "

```

สร้างโมเดลในการสร้างข้อมูลใหม่ ตอนนี้ X3 ไปอยู่กับ F2

แล้วจะลองวิเคราะห์ด้วยโมเดลที่ X3 อยู่กับ F1

แล้วคิดว่าดัชนีความเหมาะสมสามารถปฏิเสธโมเดลได้หรือไม่

```

>
> simoutalt200 <- sim(nRep=1000, model=analysismodel, generate=genalternativemodel,
+ n=200, lavaanfun="cfa", std.lv=TRUE)

```

เปลี่ยนโมเดลในการสร้างข้อมูล

```
> summary(simoutalt200)
```

RESULT OBJECT

Model Type

[1] "lavaan"

==== Fit Indices Cutoffs =====

Alpha

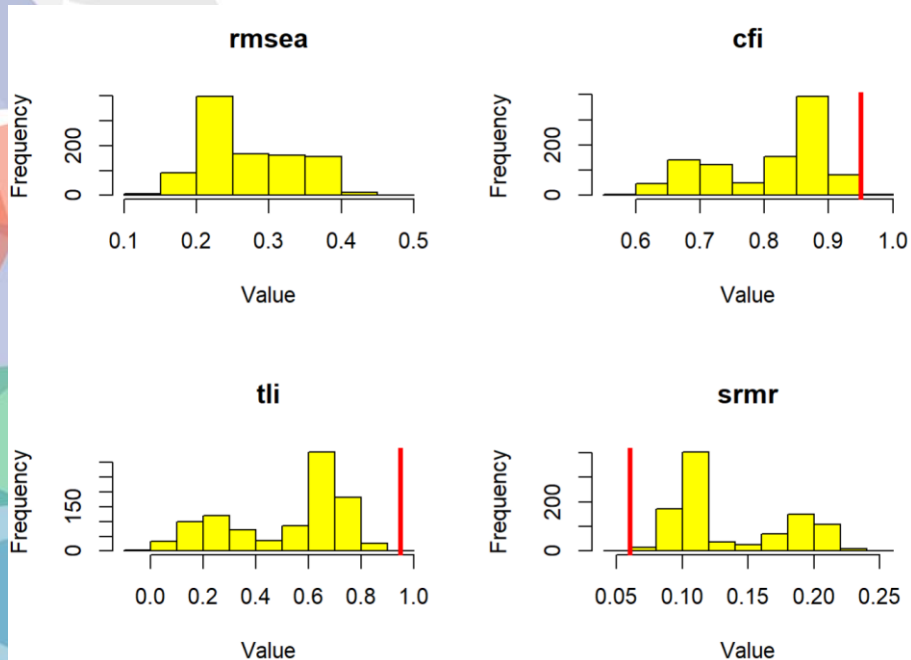
Fit Indices	0.1	0.05	0.01	0.001	Mean	SD
chisq	110.898	117.779	134.117	157.316	65.562	29.360
aic	2651.875	2676.527	2715.190	2753.214	2585.229	53.318
bic	2688.156	2712.808	2751.471	2789.496	2621.510	53.318
rmsea	0.366	0.377	0.403	0.438	0.270	0.064
cfi	0.668	0.650	0.621	0.584	0.807	0.090
tli	0.170	0.125	0.052	-0.040	0.517	0.224
srmr	0.203	0.210	0.220	0.241	0.136	0.044

จะเห็นว่าค่าเฉลี่ยของดัชนีความเหมาะสมค่อนข้างไม่ดี



```
> ruleThumb <- c(RMSEA=0.05, CFI=0.95, TLI=0.95, SRMR=0.06)
> plotPowerFit(simoutalt200, cutoff=ruleThumb, alpha=0.05,
+             usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))
```

ใช้จุดตัดที่ได้รับความนิยมสูง



ภาพการกระจาย (Sampling Distribution)
ของดัชนีความเหมาะสมของการวิเคราะห์ทั้ง 1000 ครั้ง
พบว่ามียาค่าดัชนีความเหมาะสมเกินจุดตัดเกือบทั้งหมด

```
> getPowerFit(simoutalt200, cutoff=ruleThumb, alpha=0.05,
+             usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))
```

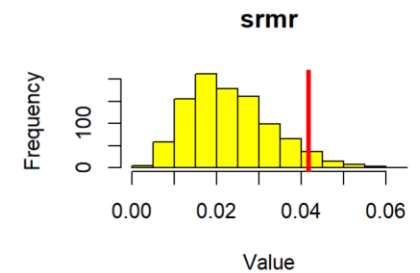
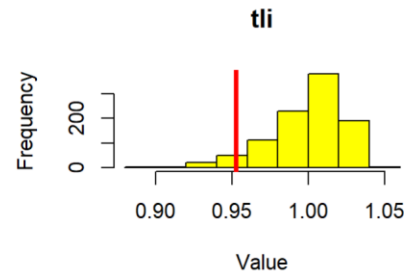
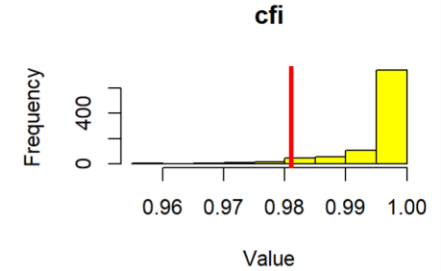
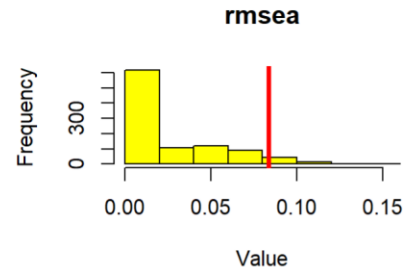
```
      cfi      tli      rmsea      srmr
0.9969849 1.0000000 1.0000000 0.9989950
```

สามารถปฏิเสธโมเดลได้ดี

ลองดูว่า ถ้านำ **Cutoff** จากการกำหนด $\alpha = .05$ จากข้อมูลที่สร้างมาจากโมเดลที่ถูกต้อง

```
> plotCutoff(simout200, alpha=0.05,  
+           usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))  
> getCutoff(simout200, alpha=0.05,  
+           usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))  
          rmsea      cfi      tli      srmr  
95% 0.08379104 0.9810692 0.9526731 0.04163366
```

คำนวณค่าจุดตัดที่ $\alpha = .05$



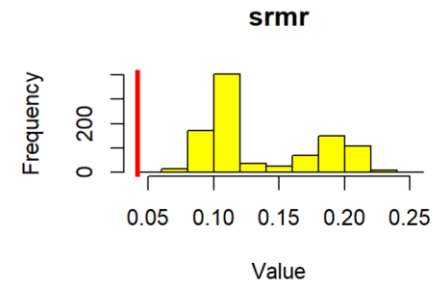
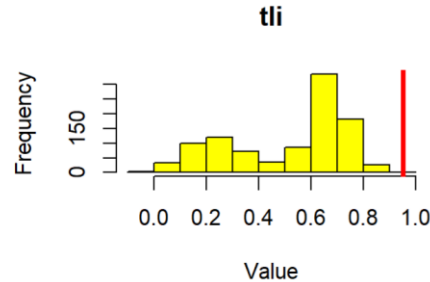
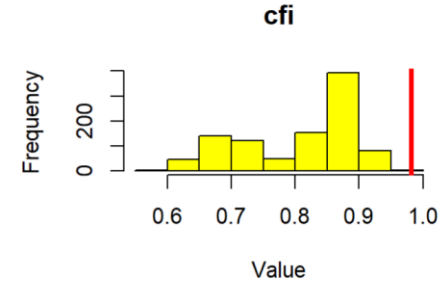
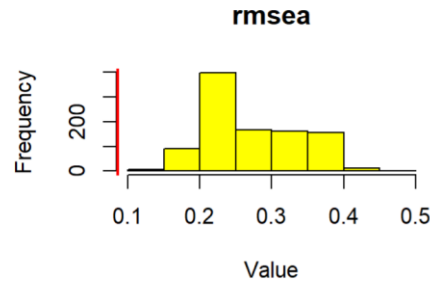
```

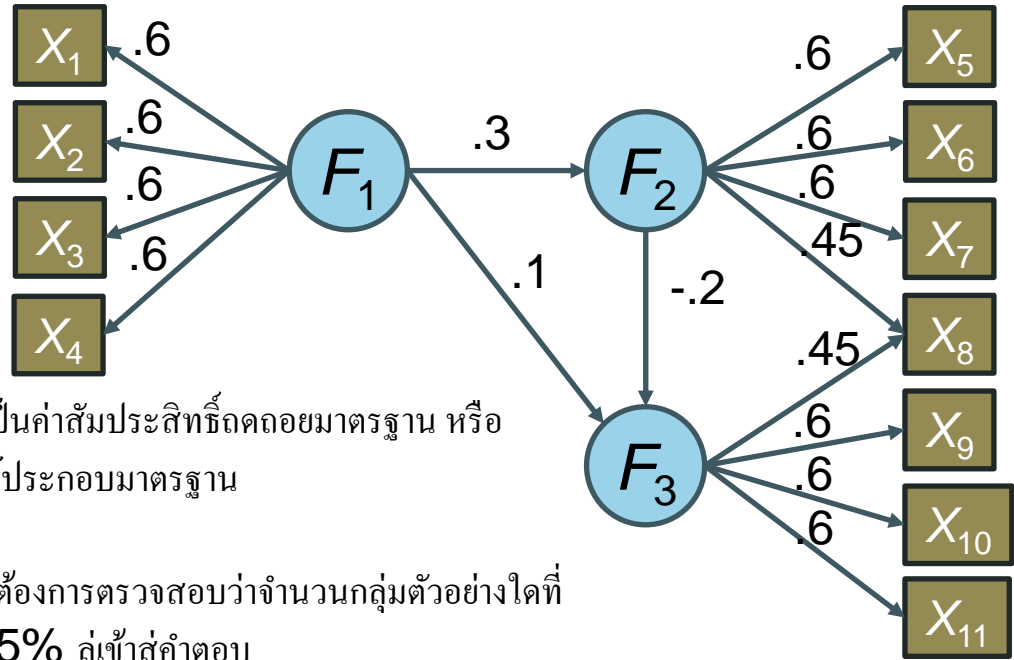
> dfiCutoff <- getCutoff(simout200, alpha=0.05,
+                       usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))
> getPowerFit(simoutalt200, cutoff=dfiCutoff, alpha=0.05,
+            usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))
  cfi  tli rmsea srmr
  1    1    1    1
> plotPowerFit(simoutalt200, cutoff=dfiCutoff, alpha=0.05,
+             usedFit=c("RMSEA", "CFI", "TLI", "SRMR"))

```

ใช้จุดตัดจากที่หามาข้างต้น

กำลังในการปฏิเสธ โมเดลที่ไม่เหมาะสมดีมาก





ค่าทั้งหมดเป็นค่าสัมประสิทธิ์ถดถอยมาตรฐาน หรือ
น้ำหนักองค์ประกอบมาตรฐาน

จากโมเดลในประชากรข้างต้น ต้องการตรวจสอบว่าจำนวนกลุ่มตัวอย่างใดที่

1. การวิเคราะห์ห้อย่างน้อย **95%** ลู่เข้าสู่คำตอบ
2. ค่าสัมประสิทธิ์ถดถอยระหว่างองค์ประกอบมือคติไม่เกิน **5%**
3. ค่าสัมประสิทธิ์ถดถอยทั้งหมด มีกำลังในการทดสอบทางสถิติเกิน **80%**
4. ค่าอิทธิพลทางอ้อมมีกำลังในการทดสอบทางสถิติเกิน **80%**
5. ค่าสัมประสิทธิ์ถดถอยทั้งหมดมีช่วงเชื่อมั่นไม่เกิน **.10**
6. ค่าอิทธิพลทางอ้อมมีช่วงเชื่อมั่นไม่เกิน **.10**

```
> library(simstandard)
```

```
> set.seed(123321)
```

```
> genmodel <- "
```

```
+ f1 =~ 0.6*x1 + 0.6*x2 + 0.6*x3 + 0.6*x4
```

```
+ f2 =~ 0.6*x5 + 0.6*x6 + 0.6*x7 + 0.45*x8
```

```
+ f3 =~ 0.45*x8 + 0.6*x9 + 0.6*x10 + 0.6*x11
```

```
+ f2 ~ 0.3*f1
```

```
+ f3 ~ 0.1*f1 + -0.2*f2
```

```
+ "
```

```
> popdat <- sim_standardized(genmodel, n = 100000, observed=TRUE,
```

```
+ latent=FALSE, errors=FALSE)
```

```
> head(popdat)
```

```
# A tibble: 6 x 11
```

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1.98	-0.447	0.645	-0.374	0.552	0.532	-0.338	0.127	0.765	-0.0842	0.407
2	-2.55	-1.27	0.120	0.563	-1.43	-0.870	-0.595	0.269	-0.322	-0.616	-0.905
3	-0.732	0.0824	0.186	0.319	-0.117	-0.576	-0.0226	-0.773	1.32	0.121	-0.684
4	-1.23	-1.98	-0.211	-0.0946	1.15	0.899	2.11	1.29	0.848	0.748	2.34
5	0.433	0.183	0.0466	-0.342	1.55	0.216	0.146	-0.229	1.80	-0.181	1.01
6	0.642	-0.167	-1.76	-0.236	0.139	-0.884	-0.289	-0.764	0.330	-1.24	-1.59

ใช้ package **simstandard** มาช่วยสร้างข้อมูลประชากรจำนวน 1,000,000 คน

แล้วสุ่มคนมาจากประชากรดังกล่าว

สาเหตุที่ใช้ **simstandard** สามารถคำนวณความแปรปรวน
ค่าคงเหลือของทั้งระดับองค์ประกอบและระดับตัวบ่งชี้ได้อัตโนมัติ
โดยไม่ต้องกำหนดในโมเดล

ไม่แสดงคะแนนองค์ประกอบและคะแนน
ค่าคงเหลือ

โมเดลจำลองนี้ จะถือว่า มีประชากร 1,000,000 คน แล้วจะสุ่มข้อมูลที่ละ 200 คนออกมา 1,000 ครั้ง

```

> medmodel <- "
+ f1 =~ x1 + x2 + x3 + x4
+ f2 =~ x5 + x6 + x7 + x8
+ f3 =~ x8 + x9 + x10 + x11
+ f2 ~ a*f1
+ f3 ~ c*f1 + b*f2
+ ind := a*b
+ "

```

ลองเอาข้อมูลประชากร มาวิเคราะห์ข้อมูล จะได้ค่าพารามิเตอร์ออกมา (คอลัมน์ **est**)
 (ในที่นี้ ไม่ต้องสนใจ **se**, **z**, **pvalue**, และ **CI** เพราะเราวิเคราะห์ประชากร)

```

> paramout <- sem(medmodel, popdat)
> parameterEstimates(paramout)

```

	lhs	op	rhs	label	est	se	z	pvalue	ci.lower	ci.upper
1	f1	==	x1		1.000	0.000	NA	NA	1.000	1.000
2	f1	==	x2		0.999	0.003	391.958	0	0.994	1.004
3	f1	==	x3		0.999	0.003	391.998	0	0.994	1.004
4	f1	==	x4		0.997	0.003	391.617	0	0.992	1.002
5	f2	==	x5		1.000	0.000	NA	NA	1.000	1.000
6	f2	==	x6		0.998	0.003	364.366	0	0.993	1.004
7	f2	==	x7		0.998	0.003	364.268	0	0.992	1.003
8	f2	==	x8		0.749	0.002	306.827	0	0.744	0.754
9	f3	==	x8		1.000	0.000	NA	NA	1.000	1.000
10	f3	==	x9		1.339	0.004	301.814	0	1.330	1.347
11	f3	==	x10		1.338	0.004	301.765	0	1.329	1.346
12	f3	==	x11		1.338	0.004	301.731	0	1.330	1.347
13	f2	~	f1	a	0.297	0.002	184.757	0	0.294	0.301
14	f3	~	f1	c	0.074	0.001	61.262	0	0.072	0.076
15	f3	~	f2	b	-0.149	0.001	-103.026	0	-0.152	-0.147
16	x1	~~	x1		0.640	0.001	528.035	0	0.637	0.642
17	x2	~~	x2		0.639	0.001	528.240	0	0.637	0.641
18	x3	~~	x3		0.639	0.001	528.110	0	0.636	0.641
19	x4	~~	x4		0.641	0.001	529.351	0	0.638	0.643
20	x5	~~	x5		0.638	0.001	497.959	0	0.636	0.641
21	x6	~~	x6		0.639	0.001	499.016	0	0.637	0.642
22	x7	~~	x7		0.641	0.001	499.701	0	0.638	0.643
23	x8	~~	x8		0.662	0.001	520.316	0	0.660	0.665
24	x9	~~	x9		0.639	0.001	489.846	0	0.637	0.642
25	x10	~~	x10		0.640	0.001	490.378	0	0.637	0.642
26	x11	~~	x11		0.641	0.001	490.739	0	0.639	0.644
27	f1	~~	f1		0.361	0.001	266.777	0	0.359	0.364
28	f2	~~	f2		0.329	0.001	249.728	0	0.327	0.332
29	f3	~~	f3		0.193	0.001	188.274	0	0.191	0.195
30	ind	:=	a*b	ind	-0.044	0.000	-91.998	0	-0.045	-0.043

กำหนด **rawData** ว่าข้อมูลนี้เป็นข้อมูลประชากร

```
> simout200 <- sim(nRep=1000, model=medmodel, n=200, rawData=popdat, lavaanfun="sem")
```

```
> pv <- simout200@paramValue
```

 หากกำหนด **rawData** ใน **simsem** ยังมี **bug** อยู่เลยใส่ค่าพารามิเตอร์ลงในโมเดล

```
> pv[,names(coef(paramout, type="all"))] <- coef(paramout, type="all")
```

```
> simout200@paramValue <- pv
```

```
> summary(simout200)
```

RESULT OBJECT

Model Type

[1] "lavaan"

=====
Fit Indices Cutoffs =====

	Alpha					
Fit Indices	0.1	0.05	0.01	0.001	Mean	SD
chisq	53.748	58.204	65.174	74.523	41.044	9.326
aic	6025.808	6048.412	6095.132	6141.070	5941.470	67.450
bic	6111.564	6134.169	6180.888	6226.826	6027.226	67.450
rmsea	0.041	0.048	0.056	0.066	0.015	0.017
cfi	0.959	0.945	0.929	0.904	0.987	0.019
tli	0.943	0.925	0.902	0.868	0.997	0.041
srmr	0.054	0.056	0.061	0.068	0.046	0.006

===== Parameter Estimates and Standard Errors =====

	Estimate	Average Estimate	SD	Average SE	Power (Not equal 0)	Std Est	Std Est SD	Std Ave SE	Average Param	Average Bias	Coverage
f1~x2	1.018	0.188	0.185	1.000	0.602	0.067	0.065	0.999	0.019	0.947	
f1~x3	1.010	0.188	0.184	1.000	0.597	0.067	0.065	0.999	0.011	0.946	
f1~x4	1.019	0.190	0.186	1.000	0.602	0.067	0.065	0.997	0.022	0.956	
f2~x6	1.014	0.212	0.199	1.000	0.598	0.074	0.069	0.998	0.016	0.942	
f2~x7	1.012	0.199	0.199	1.000	0.598	0.071	0.069	0.998	0.014	0.953	
f2~x8	0.765	0.185	0.179	0.998	0.453	0.086	0.081	0.749	0.016	0.943	
f3~x9	1.371	0.360	0.340	0.997	0.596	0.073	0.071	1.339	0.032	0.926	
f3~x10	1.382	0.393	0.344	0.997	0.599	0.074	0.071	1.338	0.045	0.922	
f3~x11	1.391	0.375	0.346	0.997	0.601	0.073	0.071	1.338	0.052	0.922	
a	0.308	0.123	0.116	0.796	0.303	0.104	0.097	0.297	0.011	0.950	
c	0.079	0.094	0.089	0.111	0.102	0.117	0.110	0.074	0.005	0.946	
b	-0.161	0.119	0.109	0.215	-0.201	0.124	0.118	-0.149	-0.011	0.929	
x1~x1	0.632	0.087	0.085	1.000	0.635	0.076	0.077	0.640	-0.007	0.941	
x2~x2	0.628	0.085	0.084	1.000	0.633	0.080	0.077	0.639	-0.011	0.946	
x3~x3	0.634	0.088	0.085	1.000	0.639	0.080	0.077	0.639	-0.005	0.935	
x4~x4	0.632	0.091	0.085	1.000	0.633	0.081	0.077	0.641	-0.009	0.934	
x5~x5	0.626	0.088	0.090	1.000	0.632	0.082	0.083	0.638	-0.012	0.949	
x6~x6	0.632	0.094	0.090	0.999	0.637	0.088	0.083	0.639	-0.007	0.938	
x7~x7	0.632	0.092	0.090	1.000	0.637	0.086	0.083	0.641	-0.008	0.945	
x8~x8	0.644	0.091	0.089	1.000	0.651	0.086	0.081	0.662	-0.018	0.935	
x9~x9	0.632	0.094	0.091	1.000	0.639	0.086	0.084	0.639	-0.007	0.945	
x10~x10	0.627	0.092	0.091	1.000	0.636	0.088	0.085	0.640	-0.012	0.937	
x11~x11	0.630	0.095	0.092	0.999	0.633	0.087	0.085	0.641	-0.011	0.944	
f1~f1	0.367	0.093	0.095	1.000	1.000	0.000	0.000	0.361	0.005	0.947	
f2~f2	0.329	0.090	0.091	1.000	0.897	0.064	0.058	0.329	0.000	0.933	
f3~f3	0.198	0.076	0.071	0.922	0.937	0.052	0.049	0.193	0.005	0.915	
ind	-0.049	0.042	0.038	0.047	-0.062	0.047	0.044	-0.044	-0.005	0.896	

หากกลุ่มตัวอย่างมี 200 คน กำลังในการทดสอบผลทางอ้อมมีเพียง 4.7% ต้องเพิ่มจำนวนกลุ่มตัวอย่าง

==== Replications =====

Number of replications = 1000

Number of converged replications = 1000

Number of nonconverged replications:

1. Nonconvergent Results = 0
2. Nonconvergent results from multiple imputation = 0
3. At least one SE were negative or NA = 0
4. Nonpositive-definite latent or observed (residual) covariance matrix (e.g., Heywood case or linear dependency) = 0

เมื่อวิเคราะห์ด้วย 200 คน พบว่าผลการสุ่มทุกครั้ง **convergent**

n = 700, nRep = 100

==== Parameter Estimates and Standard Errors =====

	Estimate	Average Estimate	SD Estimate	Average SE	Power (Not equal 0)	Std Est	Std Est SD	Ave SE	Average Param	Average Bias	Coverage
f1=~x2	0.996	0.092	0.096	1.00	0.600	0.033	0.035	0.999	-0.003	0.93	
f1=~x3	1.000	0.083	0.096	1.00	0.602	0.036	0.035	0.999	0.001	0.96	
f1=~x4	0.993	0.095	0.096	1.00	0.598	0.037	0.035	0.997	-0.004	0.94	
f2=~x6	0.998	0.098	0.105	1.00	0.594	0.038	0.038	0.998	-0.001	0.96	
f2=~x7	1.017	0.098	0.107	1.00	0.603	0.035	0.038	0.998	0.019	0.96	
f2=~x8	0.754	0.102	0.095	1.00	0.446	0.047	0.043	0.749	0.005	0.92	
f3=~x9	1.348	0.182	0.172	1.00	0.603	0.040	0.038	1.339	0.009	0.93	
f3=~x10	1.333	0.182	0.170	1.00	0.596	0.038	0.038	1.338	-0.004	0.94	
f3=~x11	1.333	0.183	0.170	1.00	0.596	0.039	0.038	1.338	-0.006	0.95	
a	0.286	0.065	0.060	1.00	0.289	0.057	0.053	0.297	-0.011	0.94	
c	0.074	0.044	0.046	0.34	0.099	0.057	0.059	0.074	0.000	0.94	
b	-0.155	0.052	0.057	0.88	-0.201	0.055	0.063	-0.149	-0.006	0.97	
x1~~x1	0.629	0.049	0.045	1.00	0.632	0.042	0.042	0.640	-0.011	0.91	
x2~~x2	0.636	0.045	0.045	1.00	0.639	0.039	0.042	0.639	-0.004	0.95	
x3~~x3	0.636	0.042	0.046	1.00	0.636	0.043	0.042	0.639	-0.002	0.95	
x4~~x4	0.641	0.052	0.046	1.00	0.641	0.044	0.042	0.641	0.001	0.91	
x5~~x5	0.645	0.046	0.049	1.00	0.645	0.041	0.045	0.638	0.007	0.97	
x6~~x6	0.640	0.048	0.048	1.00	0.646	0.044	0.045	0.639	0.001	0.96	
x7~~x7	0.632	0.045	0.049	1.00	0.635	0.042	0.045	0.641	-0.008	0.97	
x8~~x8	0.667	0.052	0.049	1.00	0.664	0.049	0.044	0.662	0.005	0.91	
x9~~x9	0.636	0.053	0.050	1.00	0.635	0.048	0.046	0.639	-0.003	0.97	
x10~~x10	0.644	0.046	0.050	1.00	0.643	0.046	0.046	0.640	0.005	0.95	
x11~~x11	0.645	0.050	0.049	1.00	0.644	0.046	0.046	0.641	0.004	0.95	
f1~~f1	0.366	0.046	0.051	1.00	1.000	0.000	0.000	0.361	0.005	0.96	
f2~~f2	0.325	0.046	0.050	1.00	0.913	0.033	0.030	0.329	-0.004	0.98	
f3~~f3	0.199	0.044	0.039	1.00	0.956	0.022	0.024	0.193	0.005	0.93	
ind	-0.044	0.018	0.018	0.79	-0.058	0.021	0.022	-0.044	0.000	0.94	

เมื่อวิเคราะห์ด้วย 700 คน กำลังในการทดสอบผลทางอ้อมถึง 80%

อย่างไรก็ตาม ผลทางตรงยังมีกำลังเพียงแค่ 34% ซึ่งขึ้นอยู่กับว่าผู้วิจัยสนใจให้ถึงระดับนัยสำคัญหรือไม่

n = 2000, nRep = 100

===== Parameter Estimates and Standard Errors =====

	Estimate	Average Estimate	SD	Average SE	Power (Not equal 0)	Std Est	Std Est	SD	Std Ave SE	Average Param	Average Bias	Coverage
f1~x2	0.995	0.066	0.057	1.00	0.598	0.023	0.021	0.999	-0.004	0.91		
f1~x3	1.006	0.059	0.057	1.00	0.604	0.020	0.021	0.999	0.007	0.96		
f1~x4	0.996	0.062	0.057	1.00	0.598	0.022	0.021	0.997	-0.001	0.93		
f2~x6	1.008	0.066	0.062	1.00	0.599	0.023	0.022	0.998	0.009	0.96		
f2~x7	1.003	0.061	0.062	1.00	0.598	0.020	0.022	0.998	0.005	0.97		
f2~x8	0.755	0.063	0.056	1.00	0.450	0.029	0.025	0.749	0.006	0.91		
f3~x9	1.338	0.109	0.100	1.00	0.600	0.023	0.023	1.339	0.000	0.96		
f3~x10	1.332	0.094	0.100	1.00	0.598	0.020	0.023	1.338	-0.005	0.96		
f3~x11	1.334	0.101	0.100	1.00	0.596	0.024	0.023	1.338	-0.004	0.95		
a	0.293	0.036	0.036	1.00	0.296	0.033	0.031	0.297	-0.004	0.94		
c	0.074	0.026	0.027	0.79	0.099	0.034	0.035	0.074	0.000	0.95		
b	-0.152	0.031	0.033	1.00	-0.201	0.036	0.038	-0.149	-0.003	0.97		
x1~~x1	0.635	0.028	0.027	1.00	0.637	0.027	0.025	0.640	-0.005	0.93		
x2~~x2	0.639	0.028	0.027	1.00	0.642	0.027	0.025	0.639	0.000	0.94		
x3~~x3	0.634	0.024	0.027	1.00	0.635	0.024	0.025	0.639	-0.005	0.95		
x4~~x4	0.640	0.031	0.027	1.00	0.642	0.026	0.025	0.641	-0.001	0.90		
x5~~x5	0.640	0.029	0.029	1.00	0.643	0.026	0.027	0.638	0.002	0.93		
x6~~x6	0.641	0.031	0.029	1.00	0.640	0.028	0.027	0.639	0.001	0.93		
x7~~x7	0.639	0.025	0.029	1.00	0.642	0.024	0.027	0.641	-0.001	0.98		
x8~~x8	0.662	0.027	0.029	1.00	0.664	0.028	0.026	0.662	0.000	0.97		
x9~~x9	0.638	0.027	0.029	1.00	0.640	0.027	0.027	0.639	-0.001	0.97		
x10~~x10	0.639	0.028	0.029	1.00	0.642	0.025	0.027	0.640	0.000	0.96		
x11~~x11	0.647	0.031	0.029	1.00	0.644	0.029	0.027	0.641	0.006	0.90		
f1~~f1	0.363	0.034	0.030	1.00	1.000	0.000	0.000	0.361	0.001	0.92		
f2~~f2	0.325	0.030	0.029	1.00	0.912	0.019	0.018	0.329	-0.005	0.95		
f3~~f3	0.194	0.023	0.023	1.00	0.959	0.014	0.014	0.193	0.001	0.96		
ind	-0.045	0.011	0.011	1.00	-0.060	0.013	0.013	-0.044	0.000	0.94		

เมื่อวิเคราะห์ด้วย 2000 คน กำลังในการทดสอบผลทางตรงถึง 80%

```
> summaryParam(simout700, detail=TRUE)[c("a", "b", "c", "ind"),]
  Estimate Average Estimate SD Average SE Power (Not equal 0) Std Est Std Est SD Std Ave SE Average Param Average Bias Coverage
a      0.28630860  0.06506798 0.06014715          1.00  0.28878473 0.05673444 0.05291524   0.29742913 -1.112053e-02   0.94
b     -0.15506425  0.05208745 0.05659281          0.88 -0.20108966 0.05487674 0.06319006  -0.14935626 -5.707993e-03   0.97
c      0.07389426  0.04353265 0.04604037          0.34  0.09870264 0.05658764 0.05895055   0.07402254 -1.282716e-04   0.94
ind    -0.04435121  0.01804250 0.01832624          0.79 -0.05845069 0.02111173 0.02180950  -0.04442290  7.169682e-05   0.94
  Rel Bias      Std Bias Rel SE Bias Not Cover Below Not Cover Above Average CI width SD CI width
a  -0.037388824 -0.170906274 -0.07562586          0.01          0.05          0.23577251  0.02313637
b   0.038217303 -0.109584805  0.08649596          0.02          0.01          0.22183973  0.03448819
c  -0.001732873 -0.002946562  0.05760558          0.03          0.03          0.18047494  0.02218212
ind -0.001613961  0.003973774  0.01572625          0.05          0.01          0.07183755  0.01821628
```

```
> summaryParam(simout2000, detail=TRUE)[c("a", "b", "c", "ind"),]
  Estimate Average Estimate SD Average SE Power (Not equal 0) Std Est Std Est SD Std Ave SE Average Param Average Bias Coverage
a      0.29335526  0.03563233 0.03581442          1.00  0.29558740 0.03254902 0.03131948   0.29742913 -0.0040738651   0.94
b     -0.15246469  0.03116742 0.03306067          1.00 -0.20130692 0.03574080 0.03755027  -0.14935626 -0.0031084255   0.97
c      0.07437196  0.02637358 0.02716711          0.79  0.09908869 0.03386580 0.03509038   0.07402254  0.0003494196   0.95
ind    -0.04477426  0.01073843 0.01090186          1.00 -0.05966245 0.01309374 0.01306370  -0.04442290 -0.0003513582   0.94
  Rel Bias      Std Bias Rel SE Bias Not Cover Below Not Cover Above Average CI width SD CI width
a  -0.013696927 -0.11433060  0.005110308          0.02          0.04          0.14038994  0.009337089
b   0.020812154 -0.09973316  0.060744417          0.02          0.01          0.12959544  0.012003326
c   0.004720448  0.01324885  0.030088159          0.02          0.03          0.10649313  0.008039894
ind  0.007909394 -0.03271969  0.015218930          0.05          0.01          0.04273451  0.006330240
```

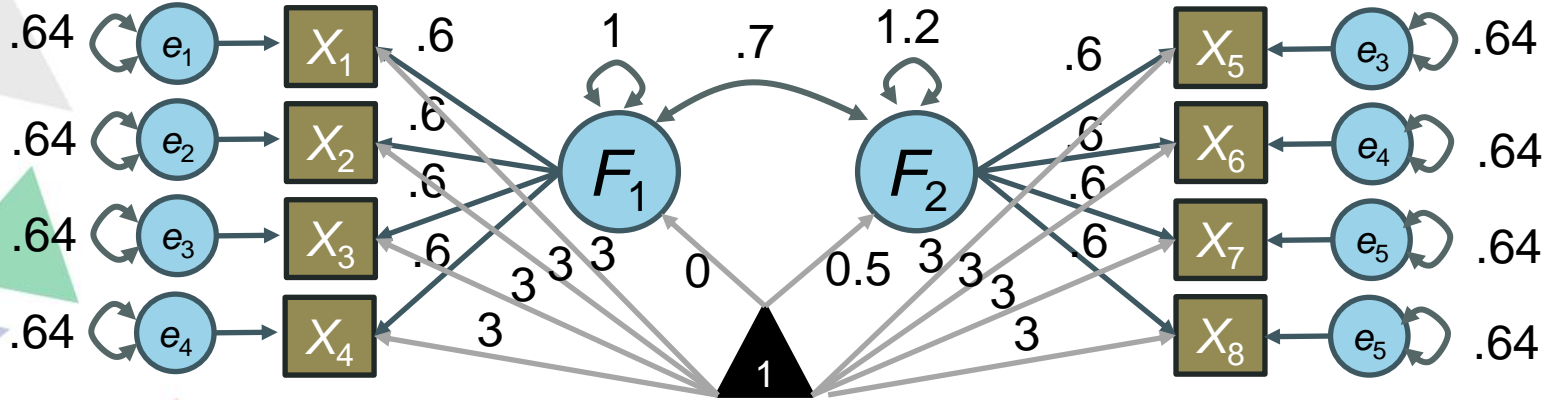
หากต้องการผลทางอ้อมมีช่วงเชื่อมั่นไม่เกิน .1 กลุ่มตัวอย่างระดับ 700 คนถือว่าเพียงพอ
 แต่หากต้องการให้ทั้ง a, b, c มีช่วงเชื่อมั่นไม่เกิน .1 ต้องเพิ่มจำนวนกลุ่มตัวอย่างมากกว่า 2000 คน

n = 4000, nRep = 100

```
> summaryParam(simout4000, detail=TRUE)[c("a", "b", "c", "ind"),]
```

	Estimate	Average Estimate	Estimate SD	Average SE	Power (Not equal 0)	Std Est	Std Est SD	Std Ave SE	Average Param	Average Bias	Coverage
a	0.29764832	0.025558119	0.025449961		1.00	0.2985934	0.023189090	0.022130585	0.29742913	0.0002191890	0.95
b	-0.15066063	0.020186005	0.023173011		1.00	-0.2003111	0.022764838	0.026548073	-0.14935626	-0.0013043665	0.99
c	0.07355970	0.019290268	0.019211291		0.98	0.0982611	0.025177475	0.024851673	0.07402254	-0.0004628387	0.96
ind	-0.04487003	0.007252868	0.007723727		1.00	-0.0599054	0.008951304	0.009261508	-0.04442290	-0.0004471260	0.97
	Rel Bias	Std Bias	Rel SE Bias	Not Cover Below	Not Cover Above	Average CI Width	SD CI Width				
a	0.0007369454	0.008576101	-0.004231851	0.02	0.03	0.09976201	0.004800854				
b	0.0087332561	-0.064617366	0.147974074	0.01	0.00	0.09083653	0.005666608				
c	-0.0062526726	-0.023993378	-0.004094135	0.01	0.03	0.07530688	0.004227356				
ind	0.0100652127	-0.061648157	0.064920286	0.03	0.00	0.03027645	0.003135580				

ต้องเพิ่มให้กลุ่มตัวอย่างถึง 4000 คน ถึงทำให้ทั้ง a, b, c มีช่วงเชื่อมั่นไม่เกิน .1



$$\text{ทดสอบขนาดอิทธิพล} = \frac{\alpha_2 - \alpha_1}{\sqrt{\phi_{11}}}$$

- จากโมเดลในประชากรข้างต้น ต้องการตรวจสอบว่าจำนวนกลุ่มตัวอย่างใดที่
1. การวิเคราะห์ห้อย่างน้อย **95%** ทุ่มเข้าสู่คำตอบ
 2. ค่าขนาดอิทธิพลมีอคติไม่เกิน **5%**
 3. ค่าความแตกต่างระหว่างค่าเฉลี่ยมีกำลังในการทดสอบเกิน **80%**
 4. ค่าความแตกต่างระหว่างค่าเฉลี่ยมีช่วงเชื่อมั่นไม่เกิน **.50**

```

> library(simsem)
>
> genmodel <- "
+ f1 =~ 0.6*x1 + 0.6*x2 + 0.6*x3 + 0.6*x4
+ f2 =~ 0.6*x5 + 0.6*x6 + 0.6*x7 + 0.6*x8
+ f2 =~ 0.7*f1
+ f1 =~ 1*f1
+ f2 =~ 1.2*f2
+ x1 =~ 0.64*x1
+ x2 =~ 0.64*x2
+ x3 =~ 0.64*x3
+ x4 =~ 0.64*x4
+ x5 =~ 0.64*x5
+ x6 =~ 0.64*x6
+ x7 =~ 0.64*x7
+ x8 =~ 0.64*x8
+ f1 ~ 0*1
+ f2 ~ 0.5*1
+ x1 ~ 3*1
+ x2 ~ 3*1
+ x3 ~ 3*1
+ x4 ~ 3*1
+ x5 ~ 3*1
+ x6 ~ 3*1
+ x7 ~ 3*1
+ x8 ~ 3*1
+ "

> analysismodel <- "
+ f1 =~ x1 + L2*x2 + L3*x3 + L4*x4
+ f2 =~ x5 + L2*x6 + L3*x7 + L4*x8
+ f1 =~ NA*f2
+ f1 =~ NA*f1 + V1*f1
+ f2 =~ NA*f2 + V2*f2
+ x1 ~ 0*1
+ x2 ~ I2*1
+ x3 ~ I3*1
+ x4 ~ I4*1
+ x5 ~ 0*1
+ x6 ~ I2*1
+ x7 ~ I3*1
+ x8 ~ I4*1
+ f1 ~ M1*1
+ f2 ~ M2*1
+ eff := (M2 - M1)/sqrt(V1)
+ "
>
> simout200 <- sim(nRep=1000, model=analysismodel, generate=genmodel, n=200, lavaanfun="cfa")

```

นิยามพารามิเตอร์ขนาดอิทธิพลเพิ่มเติม

โมเดลสำหรับสร้างข้อมูลตามรูปภาพ


```
> summary(simout200)
```

```
RESULT OBJECT
```

```
Model Type
```

```
[1] "lavaan"
```

การกระจายของดัชนีความเหมาะสม

```
==== Fit Indices Cutoffs =====
```

Alpha

Fit Indices	0.1	0.05	0.01	0.001	Mean	SD
chisq	34.752	38.258	43.543	47.087	25.798	6.852
aic	4366.097	4389.898	4418.672	4471.883	4296.062	56.647
bic	4428.765	4452.566	4481.340	4534.551	4358.730	56.647
rmsea	0.044	0.051	0.061	0.066	0.016	0.019
cfi	0.969	0.959	0.940	0.926	0.990	0.014
tli	0.966	0.954	0.933	0.917	0.998	0.025
srmr	0.055	0.059	0.066	0.072	0.045	0.008

===== Parameter Estimates and Standard Errors =====

	Estimate	Average	Estimate	SD	Average	SE	Power (Not equal 0)	Std Est	Std Est	SD	Std	Ave SE
L2 <- (f1=~x2)	1.006		0.119		0.111		1.000	0.598		0.051		0.049
L3 <- (f1=~x3)	1.005		0.118		0.110		1.000	0.598		0.050		0.049
L4 <- (f1=~x4)	1.004		0.117		0.110		1.000	0.598		0.051		0.049
L2 <- (f2=~x6)	1.006		0.119		0.111		1.000	0.630		0.049		0.048
L3 <- (f2=~x7)	1.005		0.118		0.110		1.000	0.633		0.048		0.048
L4 <- (f2=~x8)	1.004		0.117		0.110		1.000	0.632		0.050		0.048
f1~f2	0.251		0.058		0.054		1.000	0.637		0.077		0.074
V1	0.361		0.075		0.071		1.000	1.000		0.000		0.000
V2	0.431		0.087		0.082		1.000	1.000		0.000		0.000
I2 <- (x2~1)	-0.019		0.379		0.353		0.069	-0.015		0.380		0.355
I3 <- (x3~1)	-0.016		0.375		0.352		0.056	-0.013		0.379		0.354
I4 <- (x4~1)	-0.011		0.373		0.352		0.053	-0.006		0.375		0.354
I2 <- (x6~1)	-0.019		0.379		0.353		0.069	-0.015		0.368		0.343
I3 <- (x7~1)	-0.016		0.375		0.352		0.056	-0.011		0.366		0.343
I4 <- (x8~1)	-0.011		0.373		0.352		0.053	-0.006		0.363		0.343
M1	2.998		0.064		0.062		1.000	5.078		0.573		0.532
M2	3.300		0.067		0.065		1.000	5.107		0.543		0.498
x1~x1	0.633		0.079		0.078		1.000	0.638		0.061		0.058
x2~x2	0.638		0.079		0.078		1.000	0.640		0.061		0.058
x3~x3	0.636		0.077		0.078		1.000	0.640		0.060		0.058
x4~x4	0.637		0.082		0.078		1.000	0.640		0.060		0.058
x5~x5	0.635		0.081		0.079		1.000	0.597		0.064		0.060
x6~x6	0.641		0.080		0.080		1.000	0.600		0.061		0.060
x7~x7	0.633		0.078		0.079		1.000	0.598		0.060		0.060
x8~x8	0.634		0.084		0.079		1.000	0.598		0.063		0.060
eff	0.507		0.103		0.100		1.000	0.029		0.489		0.476

ค่าพารามิเตอร์ที่ประมาณการได้เฉลี่ย ต่างจากค่าในประชากร (0.5) เล็กน้อย
กำลังในการทดสอบ 100%

```
> summaryParam(simout200, detail=TRUE)["eff",]
      Estimate Average Estimate SD Average SE Power (Not equal 0)      Std Est Std Est SD Std Ave SE
eff      0.5069574    0.1027855 0.09979624                1 0.02911386  0.4886727  0.4757374
> mean(2*1.96*simout200@se[, "eff"])
[1] 0.3912012
```

กลุ่มตัวอย่างเท่ากับ **200** มีความกว้างช่วงเชื่อมั่นเท่ากับ **.39**

```
> summaryParam(simout120, detail=TRUE)["eff",]
      Estimate Average Estimate SD Average SE Power (Not equal 0)      Std Est Std Est SD Std Ave SE
eff      0.4925369    0.1188993 0.1277008                0.99 0.07304969  0.6509034  0.6311804
> mean(2*1.96*simout120@se[, "eff"])
[1] 0.5005872
```

ลดกลุ่มตัวอย่างเหลือ **120** มีความกว้างช่วงเชื่อมั่นเท่ากับ **.50**

โปรแกรมยังมี **bug** ที่ไม่แสดงความกว้างของช่วงเชื่อมั่น เมื่อมีการ **labels** ค่าพารามิเตอร์บางตัว
เลยต้องคำนวณจาก **$2 \times 1.96 \times SE$**

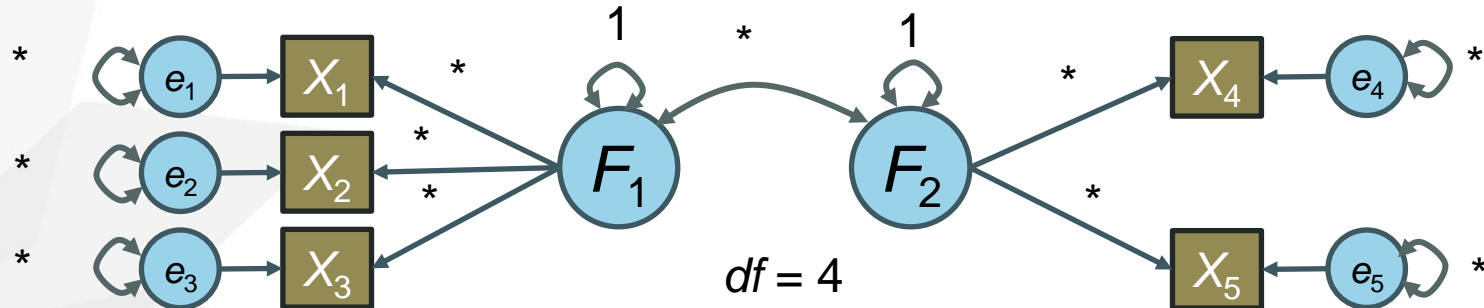
การคำนวณกลุ่มตัวอย่างผ่านการตัดสินใจโมเดล

- MacCallum, Browne, & Suguwara (1996) เสนอว่านักวิจัยสามารถทดสอบว่า RMSEA ของโมเดลตนเอง มากกว่าค่าที่กำหนดหรือไม่ เช่น

$$H_0: \varepsilon \leq .05$$

$$H_1: \varepsilon > .05$$

- โดย ε คือค่า RMSEA ในประชากร การทดสอบสามารถทำได้โดยสร้าง Sampling Distribution ของ RMSEA แล้วดูว่าค่า critical value อยู่ที่จุดใด เช่น จากโมเดล



การคำนวณกลุ่มตัวอย่างผ่านการตัดสินโมเดล

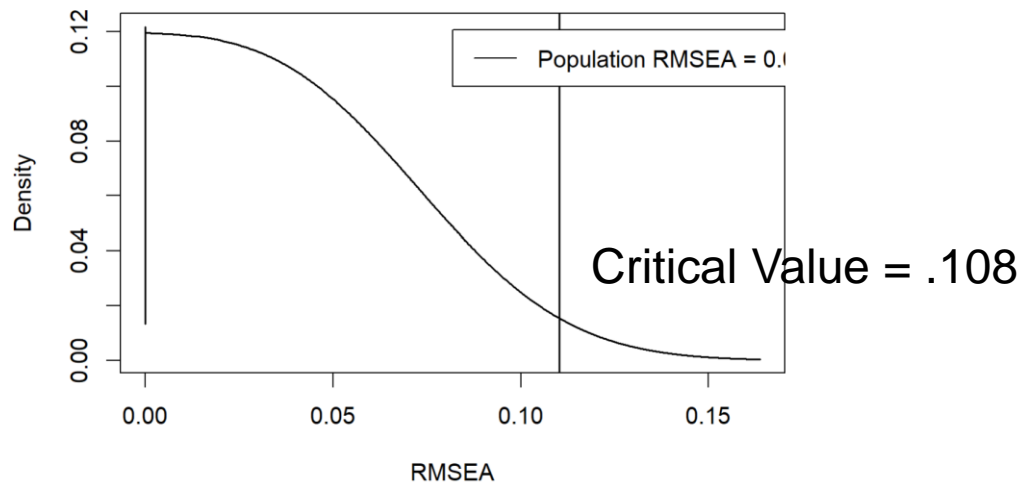
- ตัวอย่าง Sampling Distribution of RMSEA เมื่อ $\varepsilon = .05$ จำนวนกลุ่มตัวอย่างเท่ากับ 200 และ $df = 4$ คำนวณผ่าน semTools

```
> plotRMSEAdist(0.05, n = 200, df = 4, ptile = .95, rmseaScale = TRUE)
```

Warning message:

```
In plotRMSEAdist(0.05, n = 200, df = 4, ptile = 0.95, rmseaScale = TRUE) :
```

```
The density at RMSEA = 0 cannot be trusted because the plots are truncated.
```



การคำนวณกลุ่มตัวอย่างผ่านการตัดสินใจโมเดล

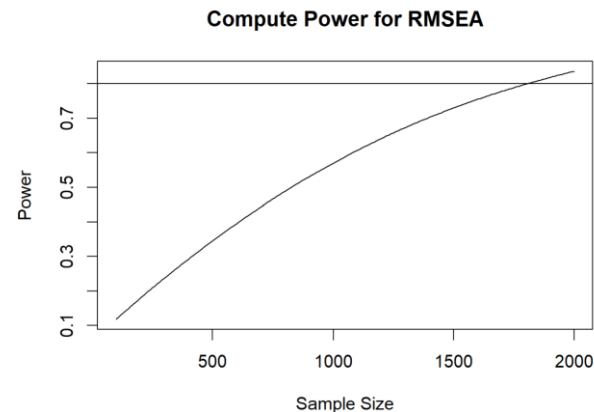
- หากทราบโมเดลที่นักวิจัยต้องการปฏิเสธ เช่น เกณฑ์ RMSEA ที่ยอมรับได้คือ .05 และต้องการปฏิเสธหากโมเดลที่เก็บข้อมูลมา มีระดับ $\varepsilon = .08$
- ดังนั้น นักวิจัยสามารถสร้าง Sampling Distribution เมื่อ $\varepsilon = .08$ แล้วสามารถหาค่ากำลังในการทดสอบ ด้วยการดูสัดส่วนที่มีค่าเกิน Critical Value
- เช่น ถ้า $H_0: \varepsilon \leq .05$ โดย $\alpha = .05, N = 200, df = 4$ แล้ว กำลังในการทดสอบทางสถิติเมื่อประชากรมี $\varepsilon = .08$ จะเท่ากับ 18%

```
> findRMSEApower(rmseao = 0.05, rmseaA = 0.08, df = 4, n = 200, alpha = 0.05, group = 1)
[1] 0.1788338
```

การคำนวณกลุ่มตัวอย่างผ่านการตัดสินใจโมเดล

- ลองเปลี่ยนแปลงจำนวนกลุ่มตัวอย่างจาก 100 ไปถึง 2000 แล้วดูว่ากลุ่มตัวอย่างมีผลกระทบต่อกำลังในการทดสอบทางสถิติอย่างไร

```
> plotRMSEpower(rmse0 = 0.05, rmsea = 0.08, df = 4,  
+              nlow = 100, nhigh = 2000, steps = 10)  
> abline(h=0.8)
```



- จำนวนกลุ่มตัวอย่างที่ทำให้กำลังเท่ากับ 80% คือ 1803

```
> findRMSEAsamplesize(rmse0 = .05, rmsea = .08, df = 4, power = 0.80)  
[1] 1803
```

การคำนวณกลุ่มตัวอย่างผ่านการตัดสินใจโมเดล



- Kelley & Lai (2011) ใช้ความกว้างของช่วงเชื่อมั่นของ RMSEA ในการกำหนดจำนวนกลุ่มตัวอย่าง

```
> library(MBESS)
```

```
> ci.rmsea(rmse=.05, df=4, N=200, conf.level=.95)
```

Note: The lower confidence limit is negative and thus set to 0 based on RMSEA's definition.

```
$Lower.Conf.Limit
```

```
[1] 0
```

```
$RMSEA
```

```
[1] 0.05
```

```
$Upper.Conf.Limit
```

```
[1] 0.1388551
```

```
> ci.rmsea(rmse=.05, df=4, N=1803, conf.level=.95)
```

```
$Lower.Conf.Limit
```

```
[1] 0.02684087
```

```
$RMSEA
```

```
[1] 0.05
```

```
$Upper.Conf.Limit
```

```
[1] 0.07512014
```

ถ้ากลุ่มตัวอย่าง = 200 ช่วงเชื่อมั่นจะอยู่ระหว่าง 0 ถึง .139

ความกว้าง = .139

ถ้ากลุ่มตัวอย่าง = 1803 ช่วงเชื่อมั่นจะอยู่ระหว่าง .027 ถึง .075

ความกว้าง = .048

การคำนวณกลุ่มตัวอย่างผ่านการตัดสินใจโมเดล

- จำนวนกลุ่มตัวอย่างที่ทำให้ความกว้างของช่วงเชื่อมั่นระดับ .95 เท่ากับ .04 ถ้า RMSEA เท่ากับ .05 คือ 2554

```
> ss.aipe.rmsea(RMSEA=.05, df=4, width=0.04, conf.level=.95)
```

```
Necessary sample size so that the expected width of the 95% confidence interval  
is no greater than 0.04, given a population RMSEA of 0.05, is:
```

```
[1] 2554
```

การคำนวณกลุ่มตัวอย่างผ่านการตัดสินใจโมเดล

- การทดสอบว่า RMSEA น้อยกว่าระดับ .05 หรือระดับใดอย่างมีนัยสำคัญ ไม่ค่อยได้ใช้ในการทดสอบความเหมาะสมของโมเดลในเชิงปฏิบัติ ดังนั้นการคำนวณกลุ่มตัวอย่างของ MacCallum et al. จึงไม่ค่อยได้ใช้
- แนวคิดเรื่องความกว้างของ RMSEA จึงดูเหมือนจะเหมาะสมมากกว่า เพราะนักวิจัยต้องการประมาณค่า RMSEA ในประชากรอย่างแม่นยำ
 - จากตัวอย่างเหมือนต้องใช้กลุ่มตัวอย่างเยอะ แต่หาก df เยอะ ยังต้องการจำนวนกลุ่มตัวอย่างน้อยลง
 - ใช้ช่วงเชื่อมั่นระดับ .9 (เหมือนทดสอบทางเดียว กับทั้งด้านขวา และด้านซ้าย) จะทำให้จำนวนกลุ่มตัวอย่างน้อยลง

```
> ss.aipe.rmsea(RMSEA=.05, df=40, width=0.04, conf.level=.90)
```

```
Necessary sample size so that the expected width of the 90% confidence interval  
is no greater than 0.04, given a population RMSEA of 0.05, is:
```

```
[1] 296
```

เพิ่ม df เป็น 40 แล้ว จำนวนกลุ่มตัวอย่างเหลือ 296 ที่ทำให้ความกว้างของ CI of RMSEA = .04

การหาค่าพารามิเตอร์ที่เหมาะสม



- การวิเคราะห์ปริมาณ บ่งบอกถึงความสัมพันธ์ระหว่างองค์ประกอบ
- งานวิจัยในอดีต งานวิจัยใกล้เคียง
- ใช้เกณฑ์แนะนำขนาดของอิทธิพล
 - Cohen (1988) แนะนำค่าอิทธิพลของสหสัมพันธ์ระดับ .1 = น้อย, .3 = ปานกลาง, .5 = สูง
 - Gignac & Szodorai (2016) ได้ทำการวิเคราะห์ปริมาณจากค่าสหสัมพันธ์ 708 ค่า ได้ 25th, 50th, และ 75th percentile เท่ากับ .11, .19, .29 จึงแนะนำขนาดอิทธิพลของสหสัมพันธ์ระดับ .1 = น้อย, .2 = ปานกลาง, .3 = สูง
 - ประเด็นคือ คิดว่าขนาดอิทธิพลระดับเท่าไรที่อยากให้เห็นผลถึงนัยสำคัญ ถ้าต่ำกว่านี้ไม่เห็นผลถึงระดับนัยสำคัญ ก็ไม่เป็นไร

การหาค่าพารามิเตอร์ที่เหมาะสม

- ใส่น้ำหนักองค์ประกอบมาตรฐาน โดยการคำนวณจากความเที่ยงองค์ประกอบ
 - จากสูตรความเที่ยง กำหนดให้ความแปรปรวนองค์ประกอบเท่ากับ 1 น้ำหนักองค์ประกอบของทุกตัวบ่งชี้เท่ากัน และเป็นน้ำหนักองค์ประกอบมาตรฐาน

$$\omega_k = \frac{\phi_{kk} \left(\sum_{j=1}^J \lambda_{jk} \right)^2}{\phi_{kk} \left(\sum_{j=1}^J \lambda_{jk} \right)^2 + \sum_{j=1}^J \theta_{jj}}$$
$$\omega = \frac{1(J\lambda)^2}{1(J\lambda)^2 + J(1 - \lambda^2)} = \frac{J\lambda^2}{J\lambda^2 + 1 - \lambda^2}$$

$$\omega J\lambda^2 + \omega - \omega\lambda^2 = J\lambda^2$$

$$J\lambda^2 - \omega J\lambda^2 + \omega\lambda^2 = \omega$$

$$\lambda^2(J - \omega J + \omega) = \omega$$

$$\lambda = \sqrt{\frac{\omega}{J - \omega J + \omega}}$$

การหาค่าพารามิเตอร์ที่เหมาะสม

- เช่น มาตรการมีความเที่ยง .75 และใช้ 3 Parcels

$$\lambda = \sqrt{\frac{.75}{3 - 3(.75) + .75}} = .707$$

- เช่น มาตรการมีความเที่ยง .8 และเป็นมาตร 5 ข้อคำถาม

$$\lambda = \sqrt{\frac{.8}{5 - 5(.8) + .8}} = .667$$