

量表發展工作坊 (3)

配適指標 & 修正指標

(Goodness-of-fit Index and Modification Index)

By
劉介宇

1

配適度檢定 (test of goodness-of-fit)

- 配適度檢定原理的虛無假設
 - 理論上 $H_0 : \Sigma = \Sigma(\theta)$
 - 實務上 $H_0 : S = \hat{\Sigma}$
 - S : 樣本共變數矩陣
 - Σ 與 $\Sigma(\theta)$: 未知母體共變數矩陣
 - $\hat{\Sigma}$: 根據母體建構參數所複製出(reproduced)的估計共變數矩陣
- CFA的基本檢定，是期望獲得「不要拒絕虛無假設」的檢定結果，也就是 H_0 被接受的結果，或獲得不顯著的檢定結果。
- 也就是: 你設定的model所跑出來的相關係數矩陣=原始資料的相關係數矩陣

2

Assessing Model Fit

- How well does the model fit the data?
- This can be answered by the Chi-square statistic but this test has many problems
 - It is sample size dependent, so with large sample sizes trivial differences will be significant
 - There are basic underlying assumptions are violated the probabilities are inaccurate

3

Example: Reproduced Correlation Matrix

Total Variance Explained									
Factor ^a	Initial Eigenvalues ^b			Extraction Sums of Squared Loadings ^f			Rotation Sums of Squared Loadings ^g		
	Total ^c	% of Variance ^d	Cumulative % ^e	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.249	52.076	52.076	5.851	48.759	48.759	2.950	24.583	24.583
2	1.229	10.246	62.322	.806	6.719	55.478	2.655	22.127	46.710
3	.719	5.992	68.313	.360	3.000	58.478	1.412	11.769	58.478
4	.613	5.109	73.423						
5	.561	4.676	78.099						
6	.503	4.192	82.291						
7	.471	3.927	86.218						
8	.389	3.240	89.458						
9	.368	3.066	92.524						
10	.328	2.735	95.259						
11	.317	2.645	97.904						
12	.252	2.096	100.000						

Extraction Method: Principal Axis Factoring.

4

Reproduced Correlations^a

	Item13 INSTRUC WELL PREPARED	Item14 INSTRUC SCHOLARLY GRASP	Item15 INSTRUC TOR CONFIDENCE	Item16 INSTRUC TOR FOCUS LECTURES	Item17 INSTRUC TOR OR USES CLEAR RELEVANT EXAMPLES	Item18 INSTRUC TOR SENSITIVE TO STUDENTS	Item19 INSTRUC TOR ALLOWS ME TO ASK QUESTIONS	Item20 INSTRUC TOR OR IS ACCESSIB LE TO STUDENTS OUTSIDE CLASS	Item21 INSTRUC TOR AWARE OF STUDENTS UNDERSTAN DING	Item22 I AM SATISFIED WITH STUDENT PERFORMAN CE EVALUATION	Item23 COMPARED TO OTHER INSTRUC TOR IS THIS INSTRUC TOR IS	Item24 COMPARED TO OTHER COURSES THIS COURSE WAS
Reproduced Correlations ^a	.819 ^b	.646	.622	.548	.594	.400	.289	.306	.475	.331	.563	.453
Item14 INSTRUC SCHOLARLY GRASP		.619 ^b	.601	.531	.582	.414	.308	.315	.475	.340	.551	.445
Item15 INSTRUCTOR CONFIDENCE			.622	.592 ^b	.525	.590	.461	.359	.349	.499	.376	.456
Item16 INSTRUCTOR FOCUS LECTURES				.548	.531	.525	.468 ^b	.529	.426	.338	.322	.452
Item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES					.594	.582	.590	.525	.623 ^b	.561	.462	.419
Item18 INSTRUCTOR SENSITIVE TO STUDENTS						.400	.414	.461	.426	.561	.679 ^b	.616
Item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS							.289	.308	.359	.338	.462	.619 ^b
Item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS								.306	.315	.348	.322	.419
Item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING									.475	.475	.499	.452
Item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION										.331	.340	.376
Item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS											.563	.551
Item24 COMPARED TO OTHER COURSES THIS COURSE WAS												.453
Residual ^a												
Item13 INSTRUC WELL PREPARED		.016	-.022	.019	-.017	.009	-.002	-.001	.000	.001	.001	.001
Item14 INSTRUC SCHOLARLY GRASP			.034	-.031	-.031	.019	.013	.000	-.026	-.007	.013	-.002
Item15 INSTRUCTOR CONFIDENCE				-.020	-.003	-.004	.000	.008	.010	-.007	.022	-.021
Item16 INSTRUCTOR FOCUS LECTURES					.019	-.031	-.030	.058	-.022	-.003	.000	.017
Item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES						-.017	-.031	-.003	.008	-.003	.031	-.009
Item18 INSTRUCTOR SENSITIVE TO STUDENTS							-.009	-.003	.001	-.009	-.007	.003
Item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS								.009	.019	-.004	-.022	-.007
Item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS									.009	.019	-.004	-.022
Item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING										.009	.019	-.004
Item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION											.009	.019
Item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS												.009
Item24 COMPARED TO OTHER COURSES THIS COURSE WAS												

Extraction Method: Principal Axis Factoring

a. Residuals are computed between observed and reproduced correlations. There are 1 (1.0%) nonredundant residuals with absolute values greater than 0.05.

b. Reproduced communalities

5

Assessing Model Fit

- Fit indices
 - Read through text books and you'll find that there are tons of fit
 - Which do you choose?
 - Different researchers have different preferences and different cutoff criterion for each index
 - Examples:
 - CFI-Chi-square based calculation
 - RMSEA-residual-based calculation

6

Assessing Model Fit

- Assessing Model Fit Indices
 - Comparative Fit Index (CFI) – compares the proposed model to an independence model (where nothing is related)

$$CFI = 1 - \frac{\tau_{\text{est.model}}}{\tau_{\text{indep.model}}}$$

$$\text{where } \tau_{\text{indep.model}} = \chi^2_{\text{indep.model}} - df_{\text{indep.model}}$$

$$\text{and } \tau_{\text{est.model}} = \chi^2_{\text{est.model}} - df_{\text{est.model}}$$

7

Assessing Model Fit

- Root Mean Square Error of Approximation
 - Compares the estimated model to a saturated or perfect model

$$RMSEA = \sqrt{\frac{\widehat{F}_0}{df_{\text{model}}}}$$

$$\text{where } \widehat{F}_0 = \frac{\chi^2_{\text{model}} - df_{\text{model}}}{N} \text{ or } 0 \text{ whichever is smaller and positive}$$

8

$$GFI = 1 - \frac{\text{tr}[\Sigma^{-1}S - I]^2}{\text{tr}(\Sigma^{-1}S)}$$

and an adjusted version,

$$AGFI = 1 - \frac{p(p+1)}{2 \cdot df_0} (1 - GFI)$$

We should also mention that there exists a traditional measure of fit for any sort of model, the root mean square error, or

$$RMSE = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^i (s_{ij} - \sigma_{ij})^2}{p(p+1)/2}}$$

Note that the double summation operators in the numerator run through each of the unique elements in the covariance matrix. The RMSE gives you the average error across the elements of Σ as compared with S .

9

整體配適度指標

- **卡方值(χ^2)**
 - 希望未達顯著，容易受到樣本數(>200)與資料偏離常態的影響
- **均方根近似誤(RMSEA)和非趨中參數(NCP)**
 - RMSEA<0.05→良好配適；0.05~0.08→合理配適；0.08~0.10→普通配適；>0.10不良配適
 - NCP越小越好，檢驗其90%信賴區間是否涵蓋0
 - RMSEA不受樣本大小與模型複雜度的影響
- **適配度指標(GFI)和修正適配度指標(AGFI)**
 - 類似迴歸分析判定係數的觀念，>0.90→良好配適
- **均方根殘差(RMR)和標準化均方根殘差(SRMR)**
 - RMR或SRMR<0.08

比較配適度指標

- **非正規化適配指標(NNFI)**
 - 比較兩對立模式之間的配適度，門檻值為0.9
- **正規化適配指標(NFI)、增值適配指標(IFI)/比較適配指標(CFI)和相對非趨中性指標(RNI)**
 - NFI與IFI用來比較所提模式與獨立模式之間的卡方值差距
 - CFI與RNI則適用於一連串模式的比較
 - 門檻值為0.9
- **期望交叉驗證指標(ECVI)**
 - 可用來檢驗模式應用到同一母體的不同樣本是否仍有效
 - 越小越好，檢驗其值是否小於飽和模式與獨立模式下的ECVI值

11

精簡配適度指標

- **正規化卡方值(NC)**
 - 為卡方值/自由度，校正模型複雜的影響所造成的膨脹效應
 - 1~3為理想適配值；<1，表示模式容易受到機運影響而產生適配不良；>3或5，表示模式需要修正
- **精簡正規化適配指標(PNFI)和精簡適配度指標(PGFI)**
 - 0到1之間，其值越高表示模式具有精簡性
- **Akaike 訊息指標(AIC)和一致Akaike 訊息指標(CAIC)**
 - >0，越小越好，樣本數>200且資料符合多變量常態性假設
- **關鍵樣本數(CN)**
 - 基於統計檢定的考量，模式要獲得一個可被接受的適配程度，所需最低數量的樣本
 - 建議：CN>200

12

測量模式的配適指標

- 信度
 - 測量觀察變數的R平方值（收斂效度）
 - 組合信度(composite reliability, CR)
 - 一組測量觀察變項具有測量某個潛在變項的理想信度
 - $CR > 0.6$

$$CR = \frac{(\sum \lambda)^2}{(\sum \lambda)^2 + \sum (\theta)}$$

- 效度
 - 因素負荷量界於0.50~0.95之間
 - 測量誤差無負值且達到顯著
 - 參數間相關的絕對值不接近1
 - 平均變異數萃取量(average variance extracted, AVE)
 - 某各潛在變項相對於測量誤差，所能解釋到測量觀察變項變異數的量(百分比)
 - $AVE > 0.5$

$$AVE = \frac{(\sum \lambda^2)}{(\sum \lambda^2) + \sum (\theta)}$$

13

結構模式的配適指標

- 潛在變項彼此之間關係的路徑係數符號
 - 正負方向是否與研究預期方向相同
- 路徑係數的參數估計值是否達到統計上的顯著
 - 參數估計值的t值絕對值，至少須大於1.96
- 每一條結構方程式的R平方值是否達到顯著
 - R平方值越大越好(最好大於0.5)

14

各配適指標總整理

指標名稱與性質	範圍	判斷值	適用情形
整體適配度指標			
χ^2 test 理論模型與觀察模型的契合程度	-	P>.05	說明模型解釋力
RMSEA (Browne & Cudeck, 1993) 比較理論模式與飽和模式的差距	0-1	<.05	不受樣本數與模式複雜度影響
NCP (Bentler, 1988) 假設模型的卡方值距離中央卡方分配的離散程度	-	越接近0 越好	說明假設模型距離中央性卡方的程度
GFI (Bentler, 1983) 假設模型可以解釋觀察資料的比例	0-1	>.90	說明模型解釋力
AGFI (Bentler, 1983) 考慮模式複雜度後的GFI	0-1*	>.90	不受模式複雜度影響
RMR與SRMR 樣本共變數矩陣與適配矩陣的差距	>0	<0.08	修正模式時可參考
R平方值	0-1	越大越好	說明模型解釋力

15

指標名稱與性質	範圍	判斷值	適用情形
比較適配度指標			
NNFI (Bentler & Bonett, 1980) 考慮模式複雜度後的NFI	0-1*	>.90	不受模式複雜度影響
NFI (Bentler & Bonett, 1980) 比較假設模型與獨立模型的卡方差異	0-1	>.90	說明模型較虛無模型的改善程度
CFI (Bentler, 1988) 假設模型與獨立模型的非中央性差異	0-1	>.95	說明模型較虛無模型的改善程度特別適合小樣本
ECVI (Browne & Cudeck, 1989) 假設模型配適共變矩陣與期望共變矩陣的差異	-	越小越好	比較模式再同一母體下不同樣本是否適用
精簡適配度指標			
χ^2/df (Wheaton et al.) 考慮模式複雜度後的卡方值	-	2~5	不受模式複雜度影響
PNFI (James et al., 1982) / PGFI (Mulaik, 1989) 考慮模式的簡約性	0-1	>.50	說明模型的簡單程度
AIC/CAIC (Akaike, 1987) 經過簡約調整的模型契合度的波動性	-	越小越好	適用於效度複核 非巢套模式比較
CN (Hoelter, 1983) 產生不顯著卡方值的樣本規模	-	>200	反應樣本規模的適切性

16

指標名稱與性質	範圍	判斷值	適用情形
<i>測量模式的配適指標</i>			
組合信度CR	0-1	>0.6	檢測測量模式的信度
觀測變數的R平方值	0-1	越大越好	檢測測量模式的信度
誤差變異數	實數	無負值且達顯著	檢測測量模式的效度
參數間相關的絕對值	正的實數	不能太接近1	檢測測量模式的效度
因素負荷量	實數	0.5~0.95	檢測測量模式的效度
平均變異數萃取量AVE	0-1	>0.5	檢測測量模式的效度
<i>結構模式的配適指標</i>			
路徑係數符號	+或-	是否與研究預期	檢驗研究假設
參數估計值	實數	t值絕對值>1.96	檢驗研究假設
R平方值	0-1	>0.5	說明模式解釋力

17

模型評估的策略

- **策略一：模型設定的合理性**
 - SEM的模型與各項參數是否能被順利的辨識、收斂與估計
- **策略二：個別參數的檢視**
 - 檢查每一個參數的正負號、數值大小是否符合理論預期
 - 檢查測量誤差的大小，分析這些殘差項當中是否透露某些變項的測量品質不佳的訊息
- **策略三：適配度指標（goodness-of-fit index）的運用**
 - 利用統計顯著性考驗檢驗假設模型與實際觀察資料的適配情形
- **策略四：模型修飾的運用**
 - 利用模型修飾的功能尋找最佳的替代模型

18

模式修飾 (model modification)

- 意義：模式需要重新建立的一種界定搜尋 (specification searches) 過程，以增進模式精簡性、整體配適程度與解釋性，降低測量誤差與預測殘差的目標。
- 模型修飾使SEM分析失去了驗證性的特性，而帶有探索性的意味
- 模式修正方向
 - 測量模式
 - 放寬或限制潛在變項對觀察變項的連結參數
 - 允許或限制測量誤差之間具有相關
 - 結構模式
 - 放寬或限制外生潛在變數對內生潛在變數的連結參數
 - 允許或限制外生潛在變數之間具有相關存在
 - 允許或限制內生潛在變數的殘差之間具有相關存在

19

We can also look at lack of fit for any individual fixed parameter. Of course, any free parameter estimated from the sample covariance matrix $\mathbf{S}$ does not contribute to lack of fit. It is the fixed parameters, generally the 0's in $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and $\mathbf{\Theta}$ that are being tested in H_0 and it is these elements that cause a model to not fit. Given that we are picking free parameters in such a way that the derivative of Chi Square with respect to those parameters is 0, or assuming all of our free parameters are in the vector α' , we have solved for the free parameters when

$$\frac{\partial \hat{\chi}^2}{\partial \alpha'} = \mathbf{0}'$$

because when the derivatives are zero, Chi Square is minimized. But this suggests a way to judge the fixed parameters. For any fixed parameter, say π , in general

$$\frac{\partial \hat{\chi}^2}{\partial \pi} \neq 0.$$

20

because when the derivatives are zero, Chi Square is minimized. But this suggests a way to judge the fixed parameters. For any fixed parameter, say π , in general

$$\frac{\partial \hat{\chi}^2}{\partial \pi} \neq 0.$$

These first derivatives provide a clue as to which parameter can be changed from fixed to free for the maximal benefit to $\hat{\chi}^2$. All that remains is that we scale the first derivative with the second derivative and we have what is called a *modification index*, or MI:

$$MI = \frac{\frac{n}{2} \left(\frac{\partial \hat{\chi}^2}{\partial \pi} \right)^2}{\frac{\partial (\hat{\chi}^2)^2}{\partial \pi \partial \pi}}.$$

21

模式修飾技術與依據

- LISREL報表上的診斷指標
 - 殘差統計數 (residual statistics)
 - 優先修正標準化殘差值之絕對值最大者
 - 標準化殘差絕對值大於1.96，表示該殘差值達到顯著
 - 修正指標 (modification indices, MI)與期望參數改變量(expected parameter change, EPC)
 - 表示重新估計參數，所能降低整個模式卡方值的數量與期望參數改變量
 - 優先修正MI與EPC之值較大的參數
 - 修正指標>3.84時，表示該參數已大到值得被修正
- 模式修正最好有理論文獻支持

22

模式修飾的建議

模型修飾的可能問題	操作建議
界定搜尋不一定帶來一個正確合理的模型	界定搜尋不宜使用在模型驗證的初期階段（因為模型尚未穩定）
越多界定錯誤，界定搜尋成功的機會越少	審慎小心的進行模型界定的工作
無法拒絕一個錯誤模型（第二類型錯誤）機率的提高	檢驗研究的統計檢定力是否足夠（例如樣本數的影響）；避免過度依賴卡方統計量來進行契合度研判
測量模型的界定搜尋較不易成功結構模型的界定搜尋較容易進行	同時兼顧測量模型與結構模型的修正可能性
以資料為基礎的無限制搜尋比以理論為基礎的有限度搜尋較不易成功	避免過度依賴計量指標來進行修飾的決策，應適當的納入理論的考量
先處理測量模型不見得會有助益而可能造成修正的混淆	針對模型當中最重要、最有意義的參數進行討論，而非取決於參數的類型
小型樣本的模型修正易造成統計結果的較大波動	若要進行模型修飾，盡可能的擴大樣本規模，樣本規模小於100時不宜進行界定搜尋
模型修飾造成模型契合度指標的不理性波動	如果樣本足夠大，建議將樣本隨機切割為二，以進行交叉檢驗
修正後模型往往無法獲得理想的效度交叉檢驗結果（除非樣本數夠大）	如有可能，可進行雙重交叉檢驗（double cross-validation）

23

建議

- 先檢定測量模式是否成立，再檢定結構模式中的潛在關係是否存在
- 建議寫入研究報告的適配度指標
 - GFI/AGFI、RMSEA、ECVI、SRMR、CFI
 - Hoyle & Panter (1995, pp.165-169)
 - 卡方值、正規化卡方值、GFI/AGFI、NNFI、IFI、CFI、RNI

Thank you for your attention!!

