

Confirmatory Factor Analysis

研究目的

想要研究一些有關老年者頭部外傷的記憶情形，藉由作答情形，來探討並將其歸納為幾個類型，希望達到輔助判讀症狀等一些參考。

使用變數

採用的變數為 Glasgow Outcome Scale-Extended(GOSE)評分表，皆採用受傷後 3 個月內(period 為 3 的 data)總共有 10 大問題，22 個小問題，分別編號，底下為此量表的編碼。

1	Gos1	是否能遵從一些簡單的指令或說任何話？	1. 否 (VS) 2. 是
2	Gos2	是否一些日常生活中的活動都需要家裡其他人的協助？ (答「否」要能照顧自己至少 24 小時)	1. 否 2. 是
3	Gos3	是否在家裡常常需要別人幫助或有人陪伴？(答「否」要能照顧自己至少 8 小時)	1. 否 (Upper SD) 2. 是 (Lower SD)
4	Gos4	受傷前在家是否需要協助？	1. 否 2. 是
5	Gos5	在無人協助的情形之下是否能獨自購物？(從計畫，算錢到與人溝通)	1. 否(Upper SD) 2. 是
6	Gos6	受傷前是否能獨自購物？	1. 否 2. 是
7	Gos7	在無人協助的情形之下是否能獨自(在當地)出遊？(能自己開車或使用大眾運輸工具，計程車)	1. 否(Upper SD) 2. 是
8	Gos8	受傷前是否能獨自出遊？	1. 否 2. 是
9	Gos9	目前工作的能力是否與受傷前相同？(能從事同級之工作或具備相同之謀職能力)	1. 否 2. 是
10	Gos10	受傷對工作造成的限制為？ a)降低工作量 b)目前不能工作/能在庇護下工作/從事無競爭性的工作	1. a (Upper MD) 2. b (Lower MD)
11	Gos11	受傷前有工作或正在找工作	1. 否 2. 是
12	Gos12	是否能重拾戶外的社交及休閒活動？(不一定要完全恢復以往活動，但必須不被現有能力的妨礙)	1. 否 2. 是
13	Gos13	戶外的社交及休閒活動受到影響的程度？ a)參與不如過去頻繁：約為過去的一半 → 1 b)較少參與：少於過去的一半 → 2 c)幾乎或無法參與：很少參與 → 3	1. a (Lower GR) 2. b (Upper MD) 3. c (Lower MD)
14	Gos14	受傷前是否就有從事戶外的社交及休閒活動？	1. 否 2. 是

15	Gos15	是否曾因心理上的問題而與家人或朋友間的關係變差？	1. 否 2. 是
16	Gos16	上述問題多久發生一次？ a)偶爾 - 少於一星期一次 b)常常 - 一星期至少一次，但可以忍受 c)經常 - 每天一次且無法忍受	1. a (Lower GR) 2. b (Upper MD) 3. c (Lower MD)
17	Gos17	受傷前是否與家人或朋友的關係有任何的問題？(若原本有問題但目前變嚴重，則選「是」)	1. 否 2. 是
18	Gos18	目前是否有其他因受傷而產生的問題影響到日常生活？ (例如：頭痛、疲倦、...)	1. 否(Upper GR) 2. 是(Lower GR)
19	Gos19	受傷前是否曾發生過類似的問題？	1. 否 2. 是
20	Gos20	是否在受傷後發生過癲癇的症狀？	1. 否 2. 是
21	Gos21	是否曾被告知會有發生癲癇的危險性？	1. 否 2. 是
22	Gos22	影響您目前狀況最重要的因子是？ a)頭傷 b)生病或是其他部位的傷 c)綜合前兩者	1. a 2. b 3. c

研究方法

因為採用 CFA 去分析，故要挑選 data 中，能夠有 latent 的 construct 去分析，但是也必須要有 continuous latent variable 去解釋每個題目的情形，而我覺得在頭部受傷後一些生活的表現，我認為會有一些能力的 continuous variable 去解釋他回答的情形，例如能力好的人，生活中表現會更好；反過來說，可能能力較差的人，需要更多人的幫助，才能夠生活等等。

首先發現 data 在 152 筆資料後，data 中出現 999 的數據，因而把它刪掉，以取其有用的 data 來分析。我採用的是 fix variance 的方式，即讓 factor variance=1. 一開始就將所有的題目都來做 factor analysis, 當然結果不是很理想。

因此就開始簡化 model, 從各個觀測值中先刪去 p-value 過小的那些，事實上經過這樣的改進後，如下表所示，會發現在 model 7 的時候，整個 model 的 p-value 就已經夠好了，也就是說 fit 的還不錯。

injury_1.sav

File name:	C:\統計方法\作業\injury_1.sav							
File size:	46492 bytes							
File date:	2009-五月 -10	下午 10:35:22						
			LL	BIC(LL)	Npar	L?/b>	df	p-value
Model1	Syntax (1)		-1256.7306	2884.2483	74	1159.7000	76	2.0e-193
Model2	Syntax (2)		-1220.1561	2881.2481	88	1086.5509	62	5.1e-187
Model3	Syntax (3)		-889.8196	2045.2028	53	543.6502	97	2.6e-63
Model4	Syntax (4)		-545.6753	1291.7759	40	131.7867	110	0.077
Model5	Syntax (5)		-609.6157	1374.5611	31	259.6676	119	1.8e-12
Model6	Syntax (6)		-559.1260	1253.5392	27	183.7800	123	0.00032
Model7	Syntax (7)		-497.6070	1110.4585	23	121.5646	127	0.62
Model8	Syntax (8)		-502.2726	1099.7473	19	130.8959	131	0.49

但是我想為了簡化 model，看能否每一個答題的選項，只受一個 latent continuous factor 的影響，因此我慢慢的從題意及數據中觀察，讓每個選項只受一個 latent continuous factor 影響。所以最後我選擇了 gos1, gos5, gos7, gos11, gos18, gos21, gos22. 並且我最後跑出來的結果如下所示：

Model8 - L?= 130.8959

Syntax (8) Model			
Estimation Warnings! See Iteration Detail			
Number of cases	150		
Number of parameters (Npar)	19		
Random Seed	266888		
Best Start Seed	2151421		
Chi-squared Statistics			
Degrees of freedom (df)	131	p-value	
L-squared (L?)	130.8959	0.49	
X-squared	134.5830	0.40	
Cressie-Read	125.5231	0.62	
BIC (based on L?)	-525.4973		
AIC (based on L?)	-131.1041		
AIC3 (based on L?)	-262.1041		
CAIC (based on L?)	-656.4973		
Dissimilarity Index	0.3795		
Log-likelihood Statistics			
Log-likelihood (LL)	-502.2726		
Log-prior	-11.5478		
Log-posterior	-513.8204		
BIC (based on LL)	1099.7473		
AIC (based on LL)	1042.5453		
AIC3 (based on LL)	1061.5453		

CAIC (based on LL)	1118.7473		
Classification Statistics	F1	F2	
Classification errors			
Reduction of errors (Lambda)			
Entropy R-squared			
Standard R-squared	0.7005	0.6992	
Files			
Infile	C:\統計方法\作業\injury_1.sav		
Options			
algorithm			
tolerance	1e-008		
emtolerance	0.01		
emiterations	250		
nritations	50		
algorithm	NR		
MstepNR	no		
startvalues			
seed	0		
sets	10		
tolerance	1e-005		
iterations	50		
bayes			
categorical	1		

variances	1		
latent	1		
poisson	1		
quadrature			
nodes	10		
missing	excludeall		
output			
parameters	effect		
standard errors	standard		
identification	no		
validation LL	no		
sample size BIC	150		
predictionstatistics	no		
Variable Detail			
Latent			
F1	Continuous	Case	10
F2	Continuous	Case	10
Dependent			
gos1	Nominal		2
1	1		
2	2		
gos5	Nominal		2
1	1		
2	2		
gos7	Nominal		2

	1	1		
	2	2		
gos11		Nominal	2	
	1	1		
	2	2		
gos18		Nominal	2	
	1	1		
	2	2		
gos21		Nominal	2	
	1	1		
	2	2		
gos22		Nominal	4	
	0	0		
	1	1		
	2	2		
	3	3		

Syntax

options

algorithm

```
tolerance=1e-008 emtolerance=0.01 emiterations=250
nriterations=50;
```

startvalues

```
seed=0 sets=10 tolerance=1e-005 iterations=50;
```

bayes

```
categorical=1 variances=1 latent=1 poisson=1;

montecarlo

seed=0 replicates=500 tolerance=1e-008;

quadrature nodes=10;

missing excludeall;

output

parameters=effect standarderrors probmeans=posterior profile
bivariateresiduals;

variables

dependent gos1 nominal,

    gos5 nominal, gos7 nominal,

    gos11 nominal, gos18 nominal,

    gos21 nominal, gos22 nominal;

latent

    F1 continuous,

    F2 continuous;

equations

(1) F1 ;

(1) F2 ;

F1 <-> F2;
```


gos1 <- 1 + F1;

gos5 <- 1 + F1;

gos7 <- 1 + F1;

gos11 <- 1 + F1;

gos18 <- 1 + F1;

gos21 <- 1 + F1;

gos22 <- 1 + F2;

Parameters

Regression Parameters									
term			coef	s.e.	z-value	p-value	Wald(0)	df	p-value
gos1(1)	<-	1	-3.9416	2.4476	-1.6104	0.11	2.5933	1	0.11
gos1(2)	<-	1	3.9416	2.4476	1.6104	0.11			
gos1(1)	<-	F1	-2.4451	1.7164	-1.4246	0.15	2.0295	1	0.15
gos1(2)	<-	F1	2.4451	1.7164	1.4246	0.15			
gos5(1)	<-	1	-1.1608	0.4556	-2.5477	0.011	6.4907	1	0.011
gos5(2)	<-	1	1.1608	0.4556	2.5477	0.011			
gos5(1)	<-	F1	-3.0937	0.9373	-3.3005	0.00097	10.8932	1	0.00097
gos5(2)	<-	F1	3.0937	0.9373	3.3005	0.00097			
gos7(1)	<-	1	-0.3110	0.5580	-0.5574	0.58	0.3107	1	0.58
gos7(2)	<-	1	0.3110	0.5580	0.5574	0.58			

gos7(1)	<-	F1	-4.1982	1.1488	-3.6545	0.00026	13.3554	1	0.00026
gos7(2)	<-	F1	4.1982	1.1488	3.6545	0.00026			
gos11(1)	<-	1	0.1203	0.0936	1.2850	0.20	1.6512	1	0.20
gos11(2)	<-	1	-0.1203	0.0936	-1.2850	0.20			
gos11(1)	<-	F1	-0.3735	0.1208	-3.0907	0.0020	9.5522	1	0.0020
gos11(2)	<-	F1	0.3735	0.1208	3.0907	0.0020			
gos18(1)	<-	1	-0.1687	0.0894	-1.8866	0.059	3.5591	1	0.059
gos18(2)	<-	1	0.1687	0.0894	1.8866	0.059			
gos18(1)	<-	F1	0.1873	0.1098	1.7062	0.088	2.9110	1	0.088
gos18(2)	<-	F1	-0.1873	0.1098	-1.7062	0.088			
gos21(1)	<-	1	2.1399	0.5475	3.9082	9.3e-5	15.2739	1	9.3e-5
gos21(2)	<-	1	-2.1399	0.5475	-3.9082	9.3e-5			
gos21(1)	<-	F1	-0.5168	0.4664	-1.1082	0.27	1.2281	1	0.27
gos21(2)	<-	F1	0.5168	0.4664	1.1082	0.27			
gos22(0)	<-	1	1.4349	0.2294	6.2545	4.0e-10	52.8837	3	1.9e-11
gos22(1)	<-	1	1.4287	0.2261	6.3189	2.6e-10			
gos22(2)	<-	1	-0.9863	0.3601	-2.7388	0.0062			
gos22(3)	<-	1	-1.8773	0.5354	-3.5064	0.00045			
gos22(0)	<-	F2	0.0260	0.6656	0.0391	0.97	0.0015	3	1.00
gos22(1)	<-	F2	0.0056	0.1442	0.0390	0.97			
gos22(2)	<-	F2	-0.0056	0.1443	-0.0389	0.97			
gos22(3)	<-	F2	-0.0260	0.6663	-0.0391	0.97			

Variances									
term			coef	s.e.	z-value	p-value	Wald(0)	df	p-value
F1		(chol)	1.0000	.	.	.			
F2		(chol)	1.0000	.	.	.			
Covariances / Associations									
term			coef	s.e.	z-value	p-value	Wald(0)	df	p-value
F1	<->	F2 (chol)	22.7987	583.0205	0.0391	0.97	0.0015	1	0.97
Variances / Covariances continuous latent									
term			coef	s.e.	z-value	p-value	Wald(0)	df	p-value
F1			1.0000	.	.	.			
F2			520.7792	26584.1787	0.0196	0.98			
F1	<->	F2	22.7987	583.0205	0.0391	0.97			

Profile

	Overall
gos1	
1	0.0655
2	0.9345
gos5	
1	0.3855
2	0.6145
gos7	
1	0.4925

2	0.5075
gos11	
1	0.5533
2	0.4467
gos18	
1	0.4191
2	0.5809
gos21	
1	0.9778
2	0.0222
gos22	
0	0.4713
1	0.4599
2	0.0443
3	0.0244

ProbMeans-Posterior

	F1	F2
Overall	0.2356	5.3712
Dependent		
gos1		
1	-1.7933	-40.8846
2	0.3349	7.6354
gos5		
1	-0.9207	-20.9883
2	0.6853	15.6221

gos7		
1	-0.8199	-18.6938
2	0.8293	18.9077
gos11		
1	-0.0711	-1.6200
2	0.5591	12.7454
gos18		
1	0.4280	9.7634
2	0.0844	1.9202
gos21		
1	0.2120	4.8334
2	1.0970	24.9984
gos22		
0	0.4647	10.6041
1	0.0551	1.2469
2	-0.1894	-4.3379
3	-0.6695	-15.3016

而且我發現經過這樣的簡化，並沒有讓 p-value 減少太多，但是還是可以 fit 不錯。綜合前面的部份，我將這兩個 continuous factor 分別命名為，F1:能力，F2:受傷程度。

最後我的 model 為 $\bar{X} = \Lambda \bar{f} + \bar{\delta}$ ，其中

$$E[\bar{X}] = \Lambda E[\bar{f}] \bar{f} = \begin{bmatrix} 1 \\ f_1 \\ f_2 \end{bmatrix}, \hat{\Phi} = \begin{bmatrix} 1 & 22.7987 \\ 22.7987 & 1 \end{bmatrix}$$

舉幾個例子說明例如 $E[\text{logit}(\text{gos1}(1))] = -3.9416 - 2.4451F1$,

$E[\text{logit}(\text{gos22}(0))] = 1.4349 + 0.026F1$ 等等。