

# 一般線性模型於醫學研究的應用

盧成皆<sup>1</sup> 鄒宗山<sup>2</sup>

有感於不少醫護研究者對一般線性模型認識不深，未能善加利用此方法於生物醫學研究上，本文作者於是希望在這裡以一個典型醫學研究的例子，來說明如何可使用一般線性模型進行如重覆測量變方分析等種種統計分析，及解決一些如遺漏數據等常見的應用問題。在解說的過程中，我們更把統計軟體 SPSS 內一般線性模型的操作過程加以說明，並闡釋如何判讀其報表，藉此協助讀者們了解，以期日後能實際應用於研究計劃中。（彰化醫學 2002;7:193-200）

**關鍵詞：**一般線性模型，遺漏數據，固定及隨機效果

## 前　　言

醫護研究者對常用的統計方法如 t-檢定、卡方檢定、複迴歸、甚至 Kaplan-Meier 方法或 Cox 迴歸等都不陌生，但根據我們的諮詢經驗，很多研究者對一般線性模型 (General Linear Model, GLM) 都不太認識，未能靈活運用於資料分析上。有鑑於此，我們希望在這裡有系統地介紹 GLM，並對怎樣使用 GLM 來解決一些常見的應用問題如重覆測量變方分析 (repeated measures anova)、遺漏數據 (missing data) 的處理等作較深入說明。

為方便解說，我們會使用一個假設的例子，如表一所示，共有 6 名病患，以隨機方式分為治療與對照兩組，並使用量表於入院時、出院當天、及兩個星期後的「追蹤」期收集每名病人的主觀疼痛程度 (0-50 分)，研究目的是探討治療效果。我們並會說明如何使用統計軟體 SPSS 來做適當的分析，所以若讀者們不知如何使用 SPSS 的話，可能需先學習使用，否則在閱讀本文時或許會有點困難。

### GLM 與 t-檢定及變方分析 (anova) 的關係

若我們要比較兩組於入院時的疼痛程度，並知道各統計前題假設 (statistical assumptions) 如常態分佈等都符合的話，可使用 t-檢定進行分析，結果是達統計差異 ( $18.3 \pm 3.1$  vs  $25.3 \pm 1.5$ ,  $P=0.024$ )。原來這個問題是可以用 GLM 來分析的，如果使用 SPSS V10 或 V11 的話，在進入了“GLM”後，選擇

“Univariate”，然後把「疼痛程度」的分數（即表一內的「入院」）放入“Dependent Variable”內、「組別」放入“Fixed Factor(s)”內，再按“OK”便可，其統計報表輸出列於表二。從表二中，我們可以清楚看到「組別」的 P-值為 0.024，與 t-檢定的結果一樣。

當然，以上只是一個例子，並不能視為正式的驗證，但本文的對象為沒有數理基礎的醫護研究者，故不打算在此做任何數理驗證，有興趣的讀者可自行翻閱數理統計課本，不過從應用的層面出發，我們或許可以介紹及說明 SPSS GLM 的使用方式。

“GLM-Univariate”內的第一個空白長方條為“Dependent Variable”，它的定義應該十分明確，不需再作解釋。至於第二個空白長方格則為“Fixed Factor(s)”，變方分析中的因子 (factor)，其意義與迴歸的自變項 (independent variable) 相類似，但自變項理論上可以是連續型 (continuous) 數據，也可以是類別型 (categorical) 數據，而因子則只可以是類別型數據。若某因子只有兩個分類的話（如性別—男或女），我們說性別有兩個等級 (levels)。至於所謂「固定」(fixed) 因子，是指某因子的等級已經完全包括了所有有可能的等級，最典型的例子是性別，因為除男與女外，已經沒有其它可能的等級了。但有時候在某研究內，對於某些因子，我們只對某幾個等級感興趣，如表一中的「組別」，治療組或許只代表了眾多治療方法的其中一種，可能還有其它的治療方法。雖然如此，若在這一個研究中我們只希望評估這一個治療方法的效用，也不會把研究結果推論至其它

<sup>1</sup> 澳洲雪梨大學國際衛生學院；<sup>2</sup> 國立中央大學統計研究所

索取抽印本請聯絡：盧成皆教授，Institute for International Health, University of Sydney, 144 Burren Street, Newtown, NSW, Australia 2042.

表一、病患的主觀疼痛程度

病歷號	組別	入院	出院	追蹤
1	0	15	18	23
2	0	19	23	27
3	0	21	26	28
4	1	24	25	30
5	1	25	29	33
6	1	27	32	38

表二、使用 GLM 進行類 t-檢定來比較兩組平均的 SPSS 報表輸出

Dependent Variable: 入院

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	73.500	1	73.500	12.600	.024
Intercept	2860.167	1	2860.167	490.314	.000
組別	73.500	1	73.500	12.600	.024
Error	23.333	4	5.833		
Total	2957.000	6			

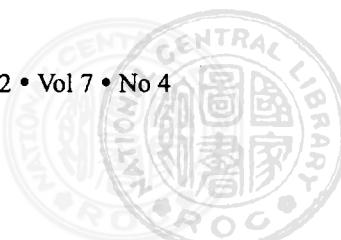
治療方法的話，那麼「組別」在這個研究中仍算是一個固定因子。再舉一個例子，在某個對老年人生活品質的研究中，「年齡層」共有三個等級，分別為“65-74”、“75-84”及“85+”，雖然嚴格來說，年齡層還包括 65 歲以下，但因為這是一個對老年人的研究，因此這三個等級已包括了所有有可能的等級，故年齡層在這個研究中仍算是一個固定因子。從年齡層這個例子，很多讀者可能都已經能夠想深一層，了解到當固定因子只有兩個等級時，GLM-Univariate 與 t-檢定無異；而當固定因子有三個或以上等級時，GLM-Univariate 便成為變方分析 (anova) 了。在此強調一點，使用 t-檢定或變方分析，無必要要求各組別標準差相同（即所謂的 equal variance assumption），但使用 GLM 時，雖然可以在 SPSS 內點選 “Options” 然後 “Homogeneity tests”，以檢定各組變異數是否相同，但 SPSS 在計算 P-值的過程中仍會自動假設各組的標準差是相同的，這一點不可不察。若報表中 “Levene's Test of Equality of Error Variances” 一欄（本論文並沒有顯示這一部份）的 P-值  $< 0.05$ ，代表推翻了相同變異數的假設，研究者便應考慮是否需要處理及如何處理，至於各種處理方式在很多課本內討論變方分析的章節都有論述，也非本文主要內容，故在此略過。至於 SPSS GLM-Univariate 中的其它空白長方格（如：“Random Factor(s)”、“Covariate(s)” … 等），我們會在下面繼續說明。

#### GLM 與線性複迴歸 (MLR) 的關係

現假設我們希望知道出院時的疼痛程度是否受

到入院時疼痛程度及接受那一種治療所影響，即利用入院分數及治療組別來估計出院分數。因為出院分數是一個「連續型」(continuous) 數據，可使用線性複迴歸來進行這個分析（嚴格來說，我們需要對有關之統計前題假設—如常態分佈、相同變異數等—先進行檢定，但本文的主旨不在討論複迴歸，故在此假設各統計前題均符合），其 SPSS 之報表輸出列於表三 (I)。其實這個分析也是可以用 GLM 來處理的，在進入了 SPSS 的 “GLM” 後，選擇 “Univariate”，並把「出院」的分數放入 “Dependent Variable”、「組別」放入 “Fixed Factor(s)”、「入院」放入 “Covariate(s)” 內，再按 “OK” 便可，其統計報表輸出列於表三 (II)。大家可注意到表三 (I) 及表三 (II) 「入院」的 P-值是一樣的，兩個表內「組別」的 P-值也完全一樣。我們用這個例子來說明（注意：是說明，不是證明）SPSS 內的 GLM 也可以處理 MLR，只要把所有「類別型」(categorical) 自變項放入 “Fixed Factor(s)”、「連續型」自變項放入 “Covariate(s)” 便可。但使用者需注意：在使用 GLM 來處理 MLR 時，所有的「交互作用」項目 (interaction terms) 都不能以點選 “Model” 內之 “Custom” 來增加，必須先以 “Transform” 及 “Compute” 來製造新的變項，再把這些新變項加入 “Covariate(s)” 內來檢視是否達統計意義。簡單地說，SPSS 內 GLM 與 MLR 處理交互作用的方法是一樣的。

雖然如此，使用 GLM 來跑 MLR 也有其優缺點。要討論它的優點，假設表一的六名病患共分三組（如表四），若使用 MLR，我們需設立「虛擬變項」



表三、以入院分數及治療組別來估計出院分數的 SPSS 報表輸出

## (I) 使用 MLR

	Unstandardized Coefficients B	Std. Error	t	Sig.
入院	1.500	.263	5.708	.011
組別	-4.167	2.111	-1.974	.143

Dependent Variable: 出院

## (II) 使用 GLM-Univariate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
入院	52.500	1	52.500	32.586	.011
組別	6.275	1	6.275	3.895	.143
Error	4.833	3			
Total	4019.000	6			
Corrected Total	117.500	5			

Dependent Variable: 出院

表四、把表一的數據由兩組改為三組

病歷號	組別	入院	出院	虛擬一	虛擬二
1	0	15	18	0	0
2	0	19	23	0	0
3	1	21	26	1	0
4	1	24	25	1	0
5	2	25	29	0	1
6	2	27	32	0	1

(dummy 或 indicator variables；如表四內之虛擬一及虛擬二) 來代替組別，虛擬一代表第“1”組與第“0”組之比較，而虛擬二則代表第“2”組與第“0”組之比較；SPSS 之有關報表輸出列於表五(I)；而 GLM 之報表輸出則列於表五 (II)。「入院」分數的 P-值在這兩部份的報表均為 0.278，沒有分別；但「組別」之結果看起來卻頗為不同：在 MLR 報表中，只有虛擬一及虛擬二之 P-值，亦即既沒有三組比較之 P-值，也沒有第“2”組與第“1”組比較之 P-值(需要重新設定虛擬變項，再重跑一次 MLR 才可得到此 P-值)。GLM 的報表則較為完整與詳盡，從「組別」中可以看到三組比較之 P-值 ( $P=0.754$ ,  $df=2$ )，從“Contrast Results”(在 GLM-Univariate 內點選“Contrasts”，再選擇“Simple”及按“Change”便可，使用者可選擇“First”或“Last”為“Reference Category”) 中亦可以看到三組兩兩比較之 P-值(類似變方分析後進行的多重比較)。亦即是說，當自變項中有類別型數據、而且該變項有三個或以上的層次 (levels；亦即分類) 時，MLR 的處理方式較複雜，報表亦沒有 GLM 的完整。而 SPSS 之 MLR 的優點則

在於它可以選擇“Stepwise”(包括“Forward”、“Backward”...) 等方法，但 GLM 則沒有選擇，而其內設之方式為“Enter”(即全部自變項均放入模型內)。當有很多自變項，或樣本數目不夠大時(樣本數目是否夠大較客觀的方式是以統計檢力來衡量)，沒有“Stepwise”的選擇時研究者便較難先以“Stepwise”方式先行作初步選擇，排除一些非常沒有顯著意義的自變項，因此在使用上可能較不方便。

## 使用 GLM 進行重覆測量變方分析 (repeated measures anova)

再回到表一的數據，若我們的目的是比較「入院」、「出院」及「追蹤」這三個時段的分數的話，可使用重覆測量變方分析。

## (一) 不考慮「組別」之比較

在進入“GLM-Repeated Measures”後，自訂一個“Within-Subject Factor”的名稱，表一中的“Within-Subject Factor”共有三個層次(levels)--「入院」、「出院」及「追蹤」--我們姑且稱這個

表五、以入院分數及治療組別（共三組）來估計出院分數的 SPSS 表輸出

## (I) 使用MLR

	Unstandardized Coefficients B	Std. Error	t	Sig.
入院	.793	.538	1.475	.278
虛擬一	.638	3.596	.177	.876
虛擬二	2.862	5.253	.545	.640

Dependent Variable: 出院

## (II) 使用GLM - Univariate

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
入院	9.121	1	9.121	2.177	.278
組別	2.729	2	1.365	.326	.754
Error	8.379	2	4.190		
Total	4019.000	6			
Corrected Total	117.500	5			

## Contrast Results

組 1 vs 組 0	Contrast Estimate	.638
	Std. Error	3.596
	Sig.	.876
組 2 vs 組 0	Contrast Estimate	2.862
	Std. Error	5.253
	Sig.	.640
組 2 vs 組 1	Contrast Estimate	2.224
	Std. Error	2.780
	Sig.	.508

Dependent Variable: 出院

“Within-Subject Factor”為「時間」，再在“Number of Levels”旁邊輸入“3”，然後按“Add”；之後按“Define”，並把「入院」、「出院」及「追蹤」按入右面的三個“\_\_?”；若我們只想檢定「時間」是否具統計意義，這時便可按“OK”。但若「時間」有統計意義，我們想更進一步檢視究竟是「入院」與「出院」有統計差異、抑或是「入院」與「追蹤」有統計差異、...等等的話，便需進行「對比」(contrast)分析。在“Define”內選點“Contrasts”，並把對比的方法改為“Simple”（其實選擇其它對比方法如“Polynomial”、“Difference”等都沒有錯，但如何選擇使用那一種方法已經超越「入門」的水平，故不在此詳述），再選擇“First”或“Last”為“Reference Category”便可（註：在這個例子中，“First”即「入院」與其它層次比較；而“Last”則為「追蹤」與其它層次比較），報表列於表六。

使用重覆測量變方分析時，有一個統計前題假設，稱為“Sphericity”，它的數理定義不適宜在此詳

述，而用語言簡單來說，是“入院與出院的差距”及“出院與追蹤的差距”兩者之間並無關係。表六的第一部份“Mauchly's Test of Sphericity”的P-值為0.081，未達統計意義，未能推翻“Sphericity”這個假設，因此在判讀表六第二部份“Tests of Within-Subjects Effects”（即比較「入院」、「出院」、「追蹤」三個時段的疼痛程度是否有顯著差異），我們選擇第一行“Sphericity Assumed”的結果，「時間」的P-值  $P < 0.001$ ，三個時段的平均疼痛分數並非完全一樣。在這裡附帶一提，若“Test of Sphericity”的P-值達統計意義，則表六的第二部份可使用“Greenhouse-Geisser”或“Huynh-Feldt”的結果。

又因為三個時段的平均疼痛分數並非完全一樣，所以我們再判讀表六第三部份有關“Contrasts”的輸出。Level1、2、3分別代表「入院」、「出院」及「追蹤」，故三個時段兩兩比較時均有顯著差異。此外，也請讀者們注意，我們共跑了兩次SPSS--分別以“First”及“Last”為“Reference Category”才能



表六、使用 GLM 進行重覆測量變方分析的 SPSS 報表輸出

## Mauchly's Test of Sphericity

	Mauchly's W	Approx. Chi-Square	df	Sig.
時間	.285	5.026	2	.081

## Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
時間	Sphericity Assumed	2187.111	2	1093.556	173.887	.000
	Greenhouse-Geisser	2187.111	1.166	1875.835	173.887	.000
	Huynh-Feldt	2187.111	1.303	1678.567	173.887	.000
	Lower-bound	2187.111	1.000	2187.111	173.887	.000

## Tests of Within-Subjects Contrasts

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
時間	Level 2 vs. Level 1	2730.667	1	2730.667	176.552	.000
	Level 3 vs. Level 1	3750.000	1	3750.000	187.500	.000
	Level 2 vs. Level 3	80.667	1	80.667	35.588	.002

## 包括組別之分析

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
時間	Sphericity Assumed	192.444	2	96.222	75.304	.000
	時間 * 組別	1.333	2	.667	.522	.612
	組別	73.500	1	73.500	7.531	.052

產生三個層次的兩兩互相比較。若研究中“Within-Subject Factor”有第四個層次（如追蹤二），則可在“Define”內將「入院」、「出院」、「追蹤一」及「追蹤二」按入右面的四個“\_\_?”時把各層次的先後次序互調，舉個例說，若輸入次序為「出院」、「入院」、「追蹤一」及「追蹤二」，那麼「出院」便成為“Level 1”。利用以上互調方式及多跑幾次SPSS，便可包含所有兩兩比較的組合。

## (二) 考慮「組別」之比較

處理方式與不考慮組別時頗為類似，只是除了上述的步驟外，還要把「組別」放入“Between-Subjects Factor(s)”空格內。從應用的角度來看，SPSS 內“GLM-Repeated Measures”之“Between-Subjects Factor(s)”與“GLM-Univariate”之“Fixed Factor(s)”差不多完全一樣，讀者可參看本文前述“GLM 與 t-檢定及變方分析的關係”中有關“Fixed Factor(s)”的說明。

加入「組別」後之分析結果列於表六的最後一部份，因為“Test of Sphericity”的結果未達統計意義，因此我們只保留“Sphericity Assumed”的結果，其餘如“Greenhouse-Geisser”、“Huynh-Feldt”等部份均

先行刪掉。這部份的報表中有一個很重要的項目，就是驗証「時間\*組別」的交互作用(interaction)，當其 P-值小於顯著水平(significance level：一般為 0.05)時，代表有證據顯示交互作用的存在；反過來說，若「時間\*組別」的 P-值大於顯著水平，便是交互作用未達統計意義。什麼是交互作用呢？以本文所討論的例子來解釋，若有交互作用，即表示兩組從「入院」到「出院」再到「追蹤」疼痛程度的改變有顯著差異，但若果要了解如何改變、差異在那個時段等問題，則需以“Contrasts”來進一步分析。不過表六內「時間\*組別」的 P-值為 0.612，交互作用未達統計意義，故兩組疼痛程度的改變沒有顯著差異(讀者請注意，是「改變」沒有差異，並非疼痛程度沒有差異)，亦即仍沒有足夠證據證明兩組中任何一組的效果較另一組為佳。

## 使用 GLM 處理具「遺漏值」(missing data)之重覆測量變方分析

很多研究者都有以下經驗：在追蹤病人的過程中會因種種原因，沒法收集到每名病人的所有資料，而產生了「遺漏數值」。舉例說，假設三號病人在「追蹤」時及五號病人在「出院時」的疼痛程度無法取得，



表七、病患主觀疼痛程度：使用“GLM – Univariate”所需的輸入方式

病歷號	組別	時段*	疼痛程度
1	0	1	15
1	0	2	18
1	0	3	23
2	0	1	19
2	0	2	23
2	0	3	27
3	0	1	21
3	0	2	26
4	1	1	24
4	1	2	25
4	1	3	30
5	1	1	25
5	1	3	33
6	1	1	27
6	1	2	32
6	1	3	38

\* 時段： 1 = 入院、2 = 出院、3 = 追踪

即表一中的“28”及“29”為遺漏數值，是空白的。有遺漏數值的數據若使用“GLM-Repeated Measures”來處理，SPSS（或任何常用的統計軟體）便會在分析前先把三號病人及五號病人完全排除，即這兩名病人的其它四個數值（三號病人的入院及出院；五號病人的入院及追蹤）也不會使用，這樣的處理方式最大的缺點是會使統計檢力（statistical power）下降，至於下降多少則視乎遺漏數值的分佈情況。遺漏數值的處理方法有好幾種，其中一種是以“GLM-Univariate”來跑，但使用此方法前，必須先把數據的輸入方式修改，新的輸入方式列於表七。

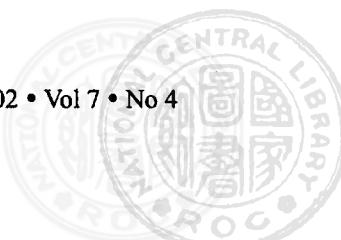
### （一）不考慮「組別」之比較

在“GLM-Univariate”內，點選「疼痛程度」為“Dependent Variable”，因為不比較組別，所以“Fixed Factor(s)”只有「時段」（時段是“Fixed Factor”因為在這研究中，我們僅希望比較入、出院及追蹤三個時段），此外，需把「病歷號」放入“Random Factor(s)”內。在這裡我們第一次提到GLM內之“Random Factor”，顧名思義，放入“Random Factor(s)”內之因子都應具有統計學上「隨機」的意義，在這個例子中，「病歷號」其實具備了「這一個樣本」的含意，需知表一內之六名病人僅為從母體（population）內抽出之一個樣本（理論上需以隨機方式抽出，但實際上幾乎沒有臨床研究能做到），並非「所有」病人，而且研究結果也希望引伸到其他同一疾病的病人，「病歷號」因此並非“Fixed factor”而是“Random Factor”。

在點選“Fixed factor”及“Random Factor”後，便要進入“Model”內、再點選“Custom”，然後把「時段」及「病歷號」分別放入右面“Model”的空格內，再按“Continue”及“OK”，便會有如表八（I）之報表。讀者需注意表八內並無“Mauchly's Test of Sphericity”，因為使用這個方式必須先假定“Sphericity”，故軟體並無再行檢定，也沒有提供“Greenhouse-Geisser”或“Huynh-Feldt”等方法來校正。若研究者不願先行假定“Sphericity”，便不能使用這個處理方式。因「時段」達統計意義，本來是需要使用“Contrast”來辨別差異在那裡，但因其進行方式與前述“GLM-Repeated Measures”時十分類似，故在此不覆述。此外，「病歷號」的P-值僅代表病患與病患間是否不一樣，一般來說病人各異，故其P-值理應達統計意義，若此P-值並未達統計意義，便代表病人依變項的同質性非常非常高，雖然不會影響如何判讀報表內「時段」部份的結果，但研究者應考慮樣本是否具代表性的問題。

### （二）考慮「組別」之比較

處理方式與不考慮組別時一樣，但需把「組別」也放入“Fixed Factor(s)”內，而且在“Model”中，需分別把「時段」、「組別」、「時段\*組別」之交互作用及「組別\*病歷號」放入右面“Model”的空格內。其報表列於表八（II）。「組別\*病歷號」的P-值的判讀方式與不考慮「組別」之比較時之「病歷號」頗為類似，在此也不重覆。



表八、使用“GLM – Univariate”進行重覆測量變方分析的 SPSS 報表輸出

## (I) 不考慮「組別」之比較

## Tests of Between-Subjects Effects

Dependent Variable: 疼痛

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
時段	Hypothesis	181.369	2	90.684	79.449	.000
	Error	9.131	8	1.141		
病歷號	Hypothesis	321.302	5	64.260	56.299	.000
	Error	9.131	8	1.141		

## (II) 考慮「組別」之比較

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
時段	Hypothesis	173.692	2	86.846	63.805	.000
	Error	8.167	6	1.361		
組別	Hypothesis	160.000	1	160.000	5.888	.072
	Error	109.134	4.01	27.174		
時段*組別	Hypothesis	.965	2	.482	.354	.715
	Error	8.167	6	1.361		
組別*病歷號	Hypothesis	113.000	4	28.250	20.755	.001
	Error	8.167	6	1.361		

## 結論

一般線性模型之應用頗為廣泛，但綜觀各醫學期刊，很多醫護研究者似乎仍未對它有足夠之認識，我們希望透過此論文大家可以有基本的認識。但我們必須強調，若追蹤的過程包含了多個時段，又或者有些「共變項」在各時段會有差異（在此補充說明：假設男、女對疼痛的忍受程度不同，而兩組的男女比例又不一樣的話，那麼「性別」便可算是一個「共變項」；但「性別」不會因時段不同而有所改變；若「血壓」高低也會影響疼痛程度，而「血壓」則有可能會因時段不同而有所改變，它便是一個依時間而改變之共變項），GLM 便不太適合，研究者應考慮使用“Generalized Estimating Equation”（簡稱 GEE），這個頗新的統計方法我們日後有機會時再討論。

## 參考文獻

1. Dobson AJ: An introduction to generalized linear models. Chapman and Hall, New York, USA, 1990.
2. Fahrmeir L & Tutz G: Multivariate statistical modelling based on generalized linear models (2nd ed.) Springer, New York, USA, 2001.
3. Horton RL: The general linear model: data analysis in the social and behavioral sciences. McGraw-Hill, London, England, 1978.
4. Lindsey JK: Applying generalized linear models. Springer, New York, USA, 1997.
5. McCulloch CE & Searle SR: Generalized, linear, and mixed models. John Wiley & Sons, New York, USA, 2001.
6. McCullagh P & Nelder JA: Generalized linear models (2nd ed.) Chapman and Hall, New York, USA, 1989.
7. McNeil K., Newman I & Kelly FJ: Testing research hypotheses with the general linear model. Southern Illinois University Press, Carbondale, USA, 1996.



## On the Use of General Linear Models in Medical Research

*Sing Kai Lo<sup>1</sup>, Daniel T.S. Tsou<sup>2</sup>*

General Linear Model (GLM) is a statistical method that can be applied to analyze data generated from different fields of medical research. However, accordingly to our consulting experience, bio-medical researchers do not possess adequate understanding of the technique. Therefore, we present a hypothetical data set similar to those generated by intervention studies to illustrate the use of GLM for various analyses such as repeated measures of analysis of variance. We also examine the usefulness of GLM when there are missing observations in the data set, a problem widely encountered in follow-up studies. Throughout the article, we demonstrate the procedures required to run GLM using the commonly available statistical software package, SPSS, and explain how the outputs should be interpreted. Readers will be equipped with the basic skills to analyze their own research data using GLM after reading this manuscript. (*Changhua J Med* 2002;7:193-200)

**Key words:** general linear model, missing data, fixed and random effect

---

<sup>1</sup>Institute for International Health, University of Sydney, Australia; <sup>2</sup>Graduate Institute of Statistics, National Central University, Taiwan

Reprint requests and correspondence to: Dr. Sing Kai Lo, Institute for International Health, University of Sydney, 144 Burren Street, Newtown, NSW, Australia 2042

