第二章　文獻回顧

一般文獻中所提到的統計方法，如：t 檢定、變異數分析、卡方檢定甚至是無母數統計方法等，均要求所蒐集的整組資料彼此間需相互獨立，即所謂的獨立性資料。但是在很多研究方面，資料間獨立性並不是一個合理的假設，例如同一個人每個月觀察一次並持續一年做重複觀測實驗，這是一個時序性追蹤資料的例子，因為同一個人隨著時間被重複觀測。上面的例子中，觀測值可以被分類成群集(Cluster)，像是同一個人的重複觀測值為一群集。對於同一群集裡的觀測值，通常假設彼此間存在相關性；至於不同的群集，則視為獨立。因此針對具有相關性資料的研究分析裡，若是忽略了其相關性結構，則可能導致錯誤的推論。

由於本研究為時序性追蹤資料，同一個人所組成之群集的觀測值具有相關性，所以必須使用考慮觀測值間相關性之模型來進行配適。以下將介紹廣義估計方程式及 Alternating Logistic Regressions 模型。此外在觀測值之間有較為複雜相關性結構時，可以使用的多層結構分析方法以增加模型估計的有效性。

第一節　廣義估計方程式

1986 年，Liang 及 Zeger 提出廣義估計方程式(Generalized Estimating Equations，簡稱 GEE)的估計方法後，克服資料需獨立性的限制，無論觀察值為連續型或是類別型，廣義估計方程式皆可以用來進行參數的估計。廣義估計方程式為一種類概似(quasi-likelihood)的估計，對於相關性資料，不需要假定觀測值的聯合分配，只需要給定操作相關矩陣(Working Correlation Matrix)以調整在同一群集內觀測值間相關性的結構即可。

由於廣義估計方程式為一邊際效應模型，因此第一步是將觀測值的邊際期望值與解釋變數的一個線性組合透過一個連結函數(Link Function)來結合

$$\mathbf{y}_i = (y_{i1},\ldots,y_{it_i})', \qquad \mu_{ij} = E(y_{ij}), \qquad g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta}$$

其中 y_{ij} 為第 i 個對像在第 j 個時間的觀測值，$\mathbf{x}_{ij} = (x_{ij1},\ldots,x_{ijp})'$ 為相對應的解釋變數所建構而成的一個 $p\times 1$ 向量，$\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)'$ 為未知參數的 $p\times 1$ 向量，而 $g(\cdot)$ 則是連結函數。

廣義估計方程式的第二步是敘述 y_{ij} 之變異數與期望值間的函數關係

$$\mathrm{Var}(y_{ij}) = V(\mu_{ij})\phi$$

其中 $V(\cdot)$ 為變異數函數，而 ϕ 是一個未知的尺度參數(Scale Parameter)。舉例來說，對於一個常態分配的觀測值，

$$g(\mu_{ij}) = \mu_{ij}, \qquad V(\mu_{ij}) = 1, \qquad \mathrm{Var}(y_{ij}) = \phi$$

如果是一個二項分配的觀測值，

$$g(\mu_{ij}) = \log\frac{\mu_{ij}}{1-\mu_{ij}}, \qquad V(\mu_{ij}) = \mu_{ij}(1-\mu_{ij}), \qquad \phi = 1$$

若觀測值為波松計數，通常則使用

$$g(\mu_{ij}) = \log(\mu_{ij}), \qquad V(\mu_{ij}) = \mu_{ij}, \qquad \phi = 1$$

第三個步驟是針對每個對象 $\mathbf{y}_i = (y_{i1},\ldots,y_{it_i})'$ 選擇一 $t_i \times t_i$ 操作相關矩陣 $\mathbf{R}_i(\boldsymbol{\alpha})$ 的形式，$\mathbf{R}_i(\boldsymbol{\alpha})$ 的第 (j, j') 元素是 y_{ij} 及 $y_{ij'}$ 之間已知或被假設的相關性。操作相關矩陣是由一個未知參數向量 $\boldsymbol{\alpha}$ 所構成，因此 $\boldsymbol{\alpha}$ 必須從資料中來估計，一旦 $\boldsymbol{\alpha}$ 給定之後，$\mathbf{R}_i(\boldsymbol{\alpha})$ 便完全是已知的。即使每個對象的相關性矩陣都可以不同，我們通常利用一個操作相關矩陣 $\mathbf{R} = \mathbf{R}(\boldsymbol{\alpha})$ 表示每個對象之間重複觀測值的一個

近似平均相關性。原則上，我們應該選擇一個與實際資料差不多的操作相關係數矩陣結構，不過即使給定一個錯誤的結構，廣義估計方程式方法所估計出來的迴歸係數及迴歸係數的變異數仍然具有一致性(Consistency)，只是錯誤之結構會降低估計的有效性(Efficiency)。除此之外，儘管選擇錯誤的操作相關係數矩陣結構，但是只要觀察對象夠多，損失的估計有效性通常都可以被忽略。

廣義估計方程式的第四個步驟就是估計參數向量$\boldsymbol{\beta}$以及它的共變異數矩陣。對於第 i 個對象，令$\mathbf{A}_i$為$t_i \times t_i$的對角矩陣，其中$V(\mu_{ij})$為第 j 個對角線元素。令$\mathbf{R}_i(\boldsymbol{\alpha})$為對應於第 i 個人的$t_i \times t_i$操作相關矩陣，因此$\mathbf{y}_i = (y_{i1}, \ldots, y_{it_i})'$的操作共變異數矩陣則為

$$\mathbf{V}_i(\boldsymbol{\alpha}) = Cov(\mathbf{y}_i) = \phi \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$$

$\boldsymbol{\beta}$的廣義估計方程式估計值則是下列方程式的解

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \left[\mathbf{V}_i(\hat{\boldsymbol{\alpha}})\right]^{-1} \left(\mathbf{y}_i - \boldsymbol{\mu}_i\right) = \mathbf{0}_p \tag{1}$$

其中$\hat{\boldsymbol{\alpha}}$為$\boldsymbol{\alpha}$的一致性估計值，而$\mathbf{0}_p$是一個$p \times 1$的 0 向量。

在(1)式中，存在著未知參數$\boldsymbol{\beta}$、$\boldsymbol{\alpha}$以及ϕ，Liang 及 Zeger (1986)提出建議以$\boldsymbol{\alpha}$和ϕ的一致性估計值$\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$及$\hat{\phi}(\boldsymbol{\beta})$取代之，這兩個估計值則是利用標準化皮爾森殘差(Standardized Pearson Residuals)來估計

$$\hat{\phi}^{-1} = \sum_{i=1}^{n} \sum_{t=1}^{n_i} r_{it}^2 / \left(\sum n_i - p\right), \quad \hat{\boldsymbol{\alpha}} = \sum_{i=1}^{n} r_{iu} r_{iv} / \left(\sum n_i - p\right)$$

其中

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{[\mathrm{Var}(\hat{\mu}_{ij})]}}$$

因此廣義估計方程式的估計程序如下：

1. 給定 $\mathbf{R}_i(\boldsymbol{\alpha})$ 和 ϕ 一個現有的估計值，並利用重複加權最小平方法計算及更新 $\boldsymbol{\beta}$ 的估計值。
2. 利用 $\boldsymbol{\beta}$ 的估計值計算標準化皮爾森殘差
3. 利用標準化皮爾森殘差估計 $\boldsymbol{\alpha}$ 和 ϕ
4. 重複 1 到 3 步驟直到結果收斂為止

接下來關於 $\hat{\boldsymbol{\beta}}$ 的變異數-共變異數矩陣估計則是利用費雪資訊矩陣(Fisher Information Matrix)的反矩陣

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \mathbf{M}_0^{-1} \tag{2}$$

其中

$$\mathbf{M}_0 = \sum_{i=1}^{n} \left(\frac{\partial \hat{\boldsymbol{\mu}}_i}{\partial \boldsymbol{\beta}} \right)' \mathbf{V}_i^{-1} \left(\frac{\partial \hat{\boldsymbol{\mu}}_i}{\partial \boldsymbol{\beta}} \right)$$

且

$$\mathbf{V}_i = \mathbf{V}_i(\hat{\boldsymbol{\alpha}})$$

(2)式中，若是模型不正確，則 $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$ 不會有一致性的估計值，但通常在時序性追蹤資料分析中，我們不太可能確定所選擇的操作相關矩陣是真正的相關性結構。

Liang 和 Zeger (1986)建議 $\hat{\boldsymbol{\beta}}$ 的變異數-共變異數矩陣可利用下面的式子估計

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \mathbf{M}_0^{-1}\mathbf{M}_1\mathbf{M}_0^{-1} \tag{3}$$

其中

$$\mathbf{M}_1 = \sum_{i=1}^{n} \left(\frac{\partial \hat{\boldsymbol{\mu}}_i}{\partial \boldsymbol{\beta}} \right)' \mathbf{V}_i^{-1} \left(\frac{\partial \hat{\boldsymbol{\mu}}_i}{\partial \boldsymbol{\beta}} \right) (\mathbf{y_i} - \hat{\boldsymbol{\mu}}_i)(\mathbf{y_i} - \hat{\boldsymbol{\mu}}_i)' \mathbf{V}_i^{-1} \left(\frac{\partial \hat{\boldsymbol{\mu}}_i}{\partial \boldsymbol{\beta}} \right)$$

即使 $\mathbf{R}_i(\alpha)$ 不是 $\mathbf{y}_i$ 真正的相關矩陣，$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})$ 是一個穩健(robust)的估計值。

估計完迴歸係數矩陣之後就可以進行參數值的檢定，其假設如下

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$$

其中 $\mathbf{C}$ 為一個限制 $\boldsymbol{\beta}$ 為線性獨立之 $c \times p$ 矩陣，而 $\mathbf{d}$ 為 $p \times 1$ 的常數向量，由於 $\hat{\boldsymbol{\beta}}$ 為漸近常態，所以其華德統計量(Wald Statistic)為

$$Q_C = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})'[\mathbf{C}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})\mathbf{C}'](\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})$$

若是虛無假設為真，則 Q_C 有一近似 χ_c^2 之分配。

第二節　Alternating Logistic Regressions

當同一個群集之觀測值具有相關性時，若觀測值為一個二項的反應值，也可以使用 Alternating Logistic Regressions (Carey, Zeger and Diggle, 1993)。此方法與廣義估計方程式非常類似，二者的差異在於廣義估計方程式著眼於觀測值間的相關係數，而 ALR 則是利用勝算比(Odds Ratio)的形式來建立模型。對於第 i 個對象，其第 j 及第 k 個觀測值的勝算比定義如下

$$\mathrm{OR}_{ijk} = \frac{P(Y_{ij}=1, Y_{ik}=1)P(Y_{ij}=0, Y_{ik}=0)}{P(Y_{ij}=1, Y_{ik}=0)P(Y_{ij}=0, Y_{ik}=1)}, \quad i = 1, \ldots, n \quad 1 \le j < k \le n_i$$

在廣義估計方程式中，與相關係數關聯之參數 $\boldsymbol{\alpha}$ 是利用迴歸係數的估計值 $\hat{\boldsymbol{\beta}}$ 來估計，而迴歸係數估計值則再利用最新估計得到之相關係數參數再做更新。如此的

演算法會不斷繼續直到α及β收斂為止。

Alternating Logistic Regressions 的演算法與廣義估計方程式相似，不同的地方是α參數所要解釋的是勝算比而不是相關係數。除此之外，對於一個相同的資料，觀測值 j 及 k 的勝算比，利用 ALR 模型所估計的值若是大於 1，則與廣義估計方程式中兩觀測值具有正相關之關係是相同的。反之，若是估計的勝算比小於 1，代表兩觀測值是負相關。對於許多醫療上的研究，利用如 ALR 之勝算比的計算方法比起相關係數較為常見且容易解釋。然而 ALR 模型只能用在反應值為二項的時候，GEE 則沒有此限制。

第三節　多層結構分析

在時序性追蹤資料分析中，觀測值間的相關性必須被考慮。雖然上面兩節的模型皆可以用來配適重複觀測資料，但是當觀測值彼此間的相關矩陣較為複雜時，如果沒有選擇一個正確的操作相關矩陣，則模型的估計可能會變得沒有效率。Chao (2006)提出了一種可以考慮觀測值間相關性之複雜結構的方法，即多層結構(Multilayer) 相關性和多層區塊(Multiblock)相關性。此方法在指定操作相關矩陣是較具有彈性的，不同於傳統上在使用廣義估計方程式時，只能選擇如 unstructured、AR1、exchangeable 及 independent 等的相關矩陣，這些選擇可能無法適用於較為複雜結構的相關係數矩陣。多層結構的分析方法在群集內觀測次數多，但是群集個數少的時候特別有用，並且可以讓廣義估計方程式配適一任意群集數目及多階層之複雜資料，而且計算過程非常有效率。

多階層資料包含了一個較複雜的結構，當在配適模型時，不同的因子及其不同之水準會將相關性矩陣分割為三角或四角形的區集，一個區集表示是利用同一個參數估計，而同一組區集代表一個同質相關性結構的層。下面為一些多階層資料結構的例子，由於這些資料的階層結構較複雜，對於同一個群集內的相關性結構隨著不同的階層而不同，所以這些例子是在傳統的相關性結構都不太適合使用

的情況。

例一：不平衡且巢狀資料

假若研究中除了精神病患者外，每位病患都有一個健康的兄弟姊妹當作控制組。這項實驗對每位病患取得一個參加試驗前及試驗後的測量值，每人共測量四次，圖 2-1 為一群集包含兩個來自相同家庭的對象，由於病患(試驗組)比控制組多了一次試驗，因此控制組為一個 2 階層結構，試驗組則為 3 階層結構。

圖 2-1、不平衡之多階層資料

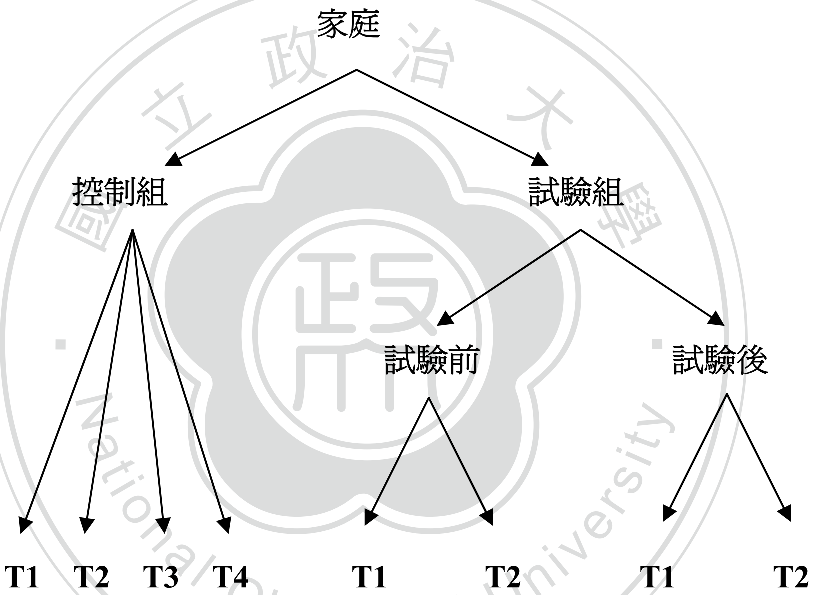

圖 2-2 為配適此資料的一個可能之相關性結構，其中對角線為同一觀測值的相關係數(d=1)，若只看上三角矩陣且不包含對角線，每一個數字表示一個相關性結構的層，同樣的層使用同一種相關性結構，如：層 1 有三個區集，層 2 跟層 3 都只有一個區集。在此例中，層 1 表示控制組四次觀測(T1、T2、T3、T4)的相關性，以及試驗組在試驗前觀測的相關性跟試驗後觀測的相關性，因此層 1 表示一個未跨越階層的相關性；同樣層 2 跟層 3 各代表跨越 1 個及 2 個階層的相關性，所以在本例中層 1、層 2 及層 3 可各使用一種傳統的相關性結構估計。

圖 2-2、多階層相關矩陣之結構(不平衡巢狀資料)

	1	2	3	4	5	6	7	8
1	d	1	1	1	3	3	3	3
2	1	d	1	1	3	3	3	3
3	1	1	d	1	3	3	3	3
4	1	1	1	d	3	3	3	3
5	3	3	3	3	d	1	2	2
6	3	3	3	3	1	d	2	2
7	3	3	3	3	2	2	d	1
8	3	3	3	3	2	2	1	d

*d 表示同觀測值的相關性，因此相關係數為 1

例二：平衡且巢狀資料

如果有一資料與例一相似，只是每個群集包含兩個來自相同家庭且有精神病的患者，每個患者有兩次試驗且每次試驗有兩次測量，此資料則變成一平衡且巢狀的 2^3 因子實驗。在此為一個典型的三層相關性矩陣，其結構如圖 2-3 所示。

圖 2-3、多階層相關矩陣之結構(平衡巢狀資料)

	1	2	3	4	5	6	7	8
1	d	1	2	2	3	3	3	3
2	1	d	2	2	3	3	3	3
3	2	2	d	1	3	3	3	3
4	2	2	1	d	3	3	3	3
5	3	3	3	3	d	1	2	2
6	3	3	3	3	1	d	2	2
7	3	3	3	3	2	2	d	1
8	3	3	3	3	2	2	1	d

一般來說，相關性矩陣的選擇高度依賴於資料的結構及分析資料之目的。傳統的相關性結構在有較大群集數或是較不注重相關性結構時是非常有用的。當層與層之間的相關性很小時，利用一個簡單相關性結構配適模型是非常實際且實用的。但是當觀測值間出現複雜的相關性結構時，多層結構相關性都可以配適而且非常彈性，無論多少因子及其不同的水準，多層結構相關性在區集的維度及層的數目上沒有限制，因此在廣義估計方程式模型的計算上是相當有效率的，比使用 unstructured 結構減少許多參數的估計，也能減少自由度的使用量。