壹、前言

本研究旨在探討廣義估計方程式（Generalized Estimating Equation, GEE）在教育上的應用，而非理論上的創新。但為了讓讀者基於對GEE的瞭解而能正確地應用於教育研究上，首先簡介該方法的理論；接著，說明運用現有免費軟體R來操作分析時，所需注意的細節。最後，提出應用GEE於教育研究上的幾種可能性，以及如何解釋分析結果。受限於篇幅，若讀者想瞭解詳細的理論推導，可參考Diggle、Heagerty、Liang與Zeger（2002）的著作*Analysis of Longitudinal Data*一書。另外，在瞭解GEE的理論前，最好具備廣義線性方程的觀念，此部分可參酌McCullagh（1983）的*Generalized Linear Models*一書。為了讀懂本研究中應用軟體R的操作指令，可先參考陳景祥（2010）的著作《R軟體：應用統計方法》，以便對R有基本的認識。

貳、何謂廣義估計方程式

一般研究想探討影響某個依變項的解釋變數時，常用的分析方法為線性迴歸模式（linear regression），此方法的限制在於依變項為連續型，且需符合常態分布的假設。然而，在實際的研究中，研究者所感興趣的依變項常常是類別（categorical）或是計數（count）的資料型態。在此情形下，依變項不再符合常態分配的假設（normality）。為了能在線性模式的架構下分析資料，勢必要對依變項做適當的變換（transformation），而廣義線性模式（generalized linear model）即為此情形下所產生的模式（Nelder & Wedderburn, 1972）。假設針對每個個體所蒐集到的依變項為y_i，$i=1, 2, ..., n$；$E(y_i)=\mu_i$；解釋變項為$x_i=(x_{i1}, x_{i2}, ..., x_{ik})$；模式參數為$\beta=(\beta_1, \beta_2, ..., \beta_k)'$；則廣義線性模式可表示為：

$$g(\mu_i)=x_i\beta，i=1, 2, ..., n$$

其中，$g(\cdot)$為連結函數（link function），意即連結依變項和解釋變項的變換（transformation）函數；連結函數的用意在於使依變項在不同的資料型態下，與解釋變數的線性組合之間能有一合適的連結關係（McCullagh & Nelder, 1989），以對迴歸係數（coefficient）做出正確的估計。例如，針對依變項資料型態為連續時，常用的是單元（identity）連結函數：$g(\mu)=\mu$；或是針對依變項資料型態為二

元類別時，常用的是邏輯（logit）連結函數：$g(\mu) = \log[\mu / (1 - \mu)]$；以及針對依變項資料型態為個數時，常用的對數（log）連結函數：$g(\mu) = \log \mu$。

而在估計迴歸係數的部分，廣義線性模式可以藉由類概似函數（quasi-likelihood function）（McCullagh, 1983; Wedderburn, 1974）來進行參數的估計。類概似函數的定義為：

$$u(\beta_j) = \sum_{i=1}^{n} d_i v_i^{-1} (y_i - \mu_i)，j = 1, 2, ..., k$$

其中，$d_i = \partial\mu_i / \partial\beta_j$，$\mathrm{Var}(y_i) \equiv v_i = \phi \times v(\mu_i)$。

例如，依變項 y_i 在常態分布下：$\phi = \sigma^2$，$v(\mu_i) = 1$。

類概似函數與一般概似函數最大的不同點在於，前者無須假設依變項的機率分配，僅假設依變項的前兩階動差（first two moments），因此其估計值具有穩健的特性。

廣義線性模式廣泛地被使用於橫斷式（cross-sectional）資料的分析，但是當蒐集到的資料為縱貫式（longitudinal）資料時，我們又如何去分析這些資料呢？由於縱貫式資料是同一個體在不同時間點的觀察值，這些依變項觀察值為相依資料（related）。當未考量依變項間的相關性時，所估計出來的迴歸係數值是偏誤的（biased）或是無效的（inefficient）。在這種情形下會導致標準誤估計錯誤，形成不正確的推論結果。此時，研究者可能使用重複測量變異數分析（repeated measures ANOVA）或混合設計變異數分析（mixed design ANOVA）的方法來分析具相關性的資料。但是，當資料不符合常態分布的假設，或是資料間的相關矩陣不符合這兩種方法的假設時，也一樣會形成不正確的推論結果。而如何考慮依變項間的相關性以得到一個較為合理的估計值，便成了統計學家研究的方向，也是研究者應用統計方法時所需注意的。

在1986年，Zeger與Liang兩位學者提出以GEE處理上述問題。廣義線性模式和GEE的不同在於，前者是針對單一時間點的資料進行分析，而後者是針對同一對象之不同時間點所蒐集到的資料一起分析。在廣義線性模式的架構下，Zeger與Liang（1986）對不同時間點的觀察值做出相關性的假設，以此種方式來分析長期追蹤資料。此方法也已經被應用於教育研究中，例如：數學教育研究有Van Dooren、De Bock、Evers與Verschaffel（2009）將同一學生在四種數字結構的非比例問題的反應分成三種類別，發現整數倍的數字結構更容易誘發學生以比例解非比例問題。另外，Yang與Lin（2012）把前測作為共變數並用GEE分析比較實驗組和對照組的後測和延後測，結果顯示，閱讀取向比寫作取向的幾何證明教學更利