Fixed-effect and Random-effect variable selection in linear mixed models using R2 statistics

吳書恆

June, 2022

摘要

本研究旨在探討...。

目次

圖次

表次

第一章　緒論

這邊假設群集與另一個群集之間獨立，而每個群集之下數個樣本或觀測值可以具有某種程度的相關，由於這些樣本彼此之間具有相關性的假設只在該群集之下，因此這種資料具有巢狀特性（Nesting），換言之，樣本嵌套在群集裏頭，類似這樣的，（**link1995social**）

隨著...

第二章　文獻回顧

第一節　混合模型

首先先說明混合模型，定義特定觀測值表示型式（observation-specific）的混合模型為（Harville, 1977; Laird andWare, 1982)

$$Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i + e_{ij}, \quad i = 1,\ldots,m, \quad j = 1,\ldots,n_i \tag{1}$$

Y_{ij} 為第 i 個群集（cluster）之下第 j 個子樣本（subsample）的觀測值，$\boldsymbol{x}_{ij}$ 為第 i 個群集之下第 j 個子樣本且維度 $p\times 1$ 的固定效果設計矩陣（fixed-effects design matrix），$\boldsymbol{\beta}$ 為未知的常數向量且維度 $p\times 1$ 的固定效果參數（fixed-effect parameters），$\boldsymbol{z}_{ij}$ 為第 i 個群集之下第 j 個子樣本且維度為 $q\times 1$ 的隨機效果設計矩陣（random-effects design matrix），$\boldsymbol{b}_i$ 為第 i 個群集、維度 $q\times 1$ 且不可觀察到的隨機效果向量（vector of unobservable random effects），e_{ij} 為第 i 個群集之下第 j 個子樣本且不可觀察到的隨機誤差值。

另外比較常見的表達方式是將同個群體的觀測值用向量表示，稱為特定群體表示型式（subject-specific, matrix notation），如下，

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i \tag{2}$$

其中 $\boldsymbol{Y}_i = (Y_{i1},\ldots,Y_{in_i})'$、$\boldsymbol{X}_i = (\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{in_i})'$、$\boldsymbol{Z}_i = (\boldsymbol{z}_{i1},\ldots,\boldsymbol{z}_{in_i})'$ 以及 $\boldsymbol{e}_i = (e_{i1},\ldots,e_{in_i})'$，混合模型在統計上會假設不可觀察到的 $\boldsymbol{b}_i$ 服從 $N_q(\boldsymbol{0},\boldsymbol{G})$、$e_i$ 服從 $N_{n_i}(\boldsymbol{0},\boldsymbol{R}_i)$，$\boldsymbol{G}$ 為未知且維度為 $q\times q$ 的隨機效果共變異數矩陣，$\boldsymbol{R}_i$ 為第 i 個群集且維度為 $n_i\times n_i$ 的組內誤差共變異數矩陣，而且 $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_m$、$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m$、$\boldsymbol{b}_i$ 與 $\boldsymbol{e}_i$ 之間皆獨立，在這個假設成立之下，第 i 個群集 $\boldsymbol{Y}_i$ 的變異數可表示為 $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}'_i + \boldsymbol{R}_i$。式 (**??**) 和式 (**??**) 雖符號不同但表達同樣的意思，式 (**??**) 在定義共變數的拆解型式時會很好用，式 (**??**) 則是較好說明混合模型的統計模型假設。

定義時間相依共變數拆解（TDC）過後的混合模型：

$$Y_{ij} = \bar{\boldsymbol{x}}'_{i.}\boldsymbol{\beta}_B + \left(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i.}\right)'\boldsymbol{\beta}_W + u_i + \left(\boldsymbol{z}_{ij} - \bar{\boldsymbol{z}}_{i.}\right)'\boldsymbol{v}_i + e_{ij} \tag{3}$$

由於截距項和時間獨立共變數在上式的 $(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_{i.})'$ 和 $(\boldsymbol{z}_{ij}-\bar{\boldsymbol{z}}_{i.})'$ 中會全都等於 0，因此沒必要去估計這些變數的 β_W 和 v_i。重新定義拆解過後的混合模型，

$$y_{ij}=\bar{\boldsymbol{x}}_{i.}'\boldsymbol{\beta}_B+\boldsymbol{w}_{ij}'\boldsymbol{\beta}_W+u_i+\boldsymbol{s}_{ij}'\boldsymbol{v}_i+e_{ij} \tag{4}$$

其中，

$\boldsymbol{\beta}_B$ 維度 $(p\times 1)$ 的 fixed 時間獨立共變數、相依共變數組間效果係數，

$\boldsymbol{\beta}_W$ 維度 $(r\times 1)$ 的 fixed 時間相依共變數組內效果係數，

$u_i=$ 維度 (1×1) 的 random 截距項，

$v_i=$ 維度 $((q-1)\times 1)$ 的 random 斜率項，且 $(u_i,v_i)'\sim N_q(0,\boldsymbol{T})$，$T$ 可以是任何形式。

$w_{ij}=$ 維度 $(k\times 1)$ 的 fixed 時間相依共變數組內效果矩陣，

$s_{ij}=$ 維度 $((q-1)\times 1)$ 的 random 時間相依共變數組內效果矩陣，且 $q-1\le r$。

若 random 共變數矩陣放入所有時間相依共變數組內效果，拆解過後的混合模型可表示為：

$$\begin{aligned} Y_{ij} &= \bar{\boldsymbol{x}}_{i.}'\boldsymbol{\beta}_B+\boldsymbol{w}_{ij}'\boldsymbol{\beta}_W+u_i+\boldsymbol{w}_{ij}'\boldsymbol{v}_i+e_{ij} \\ &= \bar{\boldsymbol{x}}_{i.}'\boldsymbol{\beta}_B+\boldsymbol{w}_{ij}'(\boldsymbol{\beta}_W+\boldsymbol{v}_i)+u_i+e_{ij} \end{aligned} \tag{5}$$

第二節　解釋變異

解釋變異（explained variance）在迴歸模型中是一個很常見的指標，他描述模型的自變項（或解釋變數）解釋依變項（或反應變數）的變異程度，反過來說，未解釋變異可以提供模型的自變項解釋不足的程度，這個特性又叫做缺適性（lack-of-fit），在迴歸裏頭，描述模型解釋變異比例（explained proportion of variance）的指標為相關決定係數（coefficient of determination），表示為 R^2，在概念上，

$$\begin{aligned} R^2 &= \frac{\text{該模型所有自變數的解釋變異}}{\text{依變數的變異}} \\ &= 1 - \frac{\text{該模型所有自變數的未解釋變異}}{\text{依變數的變異}} \end{aligned}$$

在

$$100\% = R_C^2 + \frac{\sigma^2}{\mathrm{Var}\,(Y_{ij})}, \text{ where } R_C^2 = R_M^2 + R_u^2 + R_v^2 = (R_B^2 + R_W^2) + R_u^2 + R_v^2 \qquad (6)$$

第三節　AIC

$$AIC = 2\ln(f(y|\hat{\beta},\hat{\theta})) + 2k \qquad (7)$$

$$AICc = AIC + \frac{2(k+1)(k+2)}{nk2} \qquad (8)$$

where k is the number of parameters (including the intercept) in the model.

$$mAIC = 2\ln(f(y|\hat{\beta},\hat{\theta})) + 2\alpha_n(p+q) \qquad (9)$$

$\alpha_n = 1$ in the infinite sample form or $\alpha_n = n/(npq1)$ in the finite sample form (Sugiura, 1978).

$$cAIC = 2\ln(f(y|\hat{\beta},\hat{b},\hat{\theta}) + 2(\rho+1), \qquad (10)$$

$$\rho = trace\left(\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix}\right)$$

第三章　R^2 成分納入精簡指數

第一節　定義 $R^2(\alpha)$

通常模型變數越多解釋力也會越高，但當過多變數放入模型中，可能會導致共線性 (Collinearity) 的問題，或是資料樣本數不夠導致估計值無法收斂。不過，在進行迴歸分析之前，能夠發現這些問題是可以提早解決的，像是對變數做轉換來消除之間的相關性，又或是再補充樣本解決變數維度過高的問題，但模型變數的平衡與抉擇還是要歸咎回分析的目的與成本，倘落研究的目的著重在預測 (forecasting) 新樣本，那精簡就會是個不錯的選擇，已經有許多研究發現過多解釋變數會導致預測效果的邊際效應遞減，同時在解釋模型時也較容易；另一方面，當個體追蹤的時間點不夠多或中斷追蹤的時間點過多，隨機項係數的推估就有可能會估計不出來，因此追求較精簡模型是其有必要性的。

由於混合模型的 R^2 成分容易選到較為複雜的模型，但 R^2 成分並沒辦法直接用一般迴歸調整後的 R^2 精簡的特性（分子分母除以相對應自由度），因此考慮對 R^2 成分做簡單地調整，這個調整是很直白的。令可調控的精簡指數為 α，且 α 介於 0 到 1 之間，則加入精簡指數 R^2 成分為

$$\begin{aligned} R^2(\alpha) &= (1-\alpha)\times R^2 + \alpha \times \frac{R^2}{\#(\text{parameter})}, \quad \alpha \in [0,1] \\ &= \left(1-\alpha+\frac{\alpha}{\#(\text{parameter})}\right)\times R^2 \end{aligned} \tag{11}$$

最不精簡的特性為考量總體參數解釋量，反映在 $(1-\alpha)\times R^2$，而最精簡模型的特性為考量平均一個參數解釋量，反映在 $\frac{R^2}{\#(\text{parameter})}\cdot\alpha$，其中 #(parameter) 表示該 R^2 成分對應模型的部分參數個數。想要達到多精簡是可以調整的，因此對這兩個解釋量取加權平均，權重由 α 決定，α 越接近 0 表示越不精簡，α 越接近 1 則越精簡。

第二節　套用至 $R_M^2(\alpha)$、$R_v^2(\alpha)$

$$R_M^2(\alpha) = \left(1-\alpha+\frac{\alpha}{p}\right) \times R_M^2 \tag{12}$$

$$R_v^2(\alpha) = \left(1-\alpha+\frac{\alpha}{\sum_{i=1}^{q} i}\right) \times R_v^2 \tag{13}$$

第三節　使用 $R^2(\alpha)$ 與選擇 α

假如要比較固定效果的解釋量，則計算要比較模型的 $R_M^2(\alpha)$，最後選擇最大 $R_M^2(\alpha)$ 的模型；假如比較隨機斜率效果的解釋量，則計算要比較模型的 $R_v^2(\alpha)$，最後選擇最大 $R_v^2(\alpha)$ 的模型。選混合模型時的變數時，通常不會同時比較固定效果與隨機效果，這兩者在估計時會彼此影響這特性已在第三節討論過，因此通常選模會考慮以下三種情形，一為已知 fixed 只選 random 部分；二為已知 random 只選 fixed 部分；三為 random 和 fixed 未知。

第四章　資料模擬與實際資料

第一節　資料模擬

為了方便表示模擬的各種模型，定義模型符號如下：

$M(fixed, random)$

下標前面表示隨機部分的模型類型，後面表示固定部分，模型類型符號包含：

$over$ = 超飽和模型，包含全部顯著與部分不顯著變數，

$full$ = 完整模型，包含全部顯著，

$reduce$ = 精簡模型，只包含部分不重要的顯著變數，

$-reduce$ = 過度精簡模型，只包含部分重要的顯著變數，

$other$ = 其他模型，包含部分顯著與部分不顯著變數。

舉例：$M(over, full)$ 表示模型的固定是飽和、隨機部分是超飽和。

模擬的主要流程

1. 產生具有相關性的資料：$m = 100$，$n_i = 10$，組間變數有 13 個，組內有 6 個

   $\beta_B = (-4, 5, -6, 8, -2, 0, 0, 0, 0, 0, 0, 0, 0)^T$,

   $\beta_W = (1, -1, 0, 0, 0, 0)^T$,

$$T=\begin{pmatrix}31 & 12 & 0 & 0 & 0 & 0 & 0\\ 12 & 28 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 25 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 7 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

顯著地的固定效果變數為 Vb1,Vb2,Vb3,Vb4,Vw1,Vw2，重要的顯著變數只有 Vb1,Vb2,Vb3；顯著地的隨機效果變數 Vb1,Vb2,Vb3,Vb4，重要的顯著變數只有 Vb1,Vb2。

2. 挑選出候選模型：決定候選模型就是在決定剔除變數的順序，在所有變數都放入的情況之下，透過向後選變數（backward selection）的概念，陸續剔除最不重要的變數，但是直至模型剩下最後一個變數才停止，並非完全等於向後選變數的方式。根據向後選變數的原理，會先從最不重要的變數開始剔除，也就是該參數的統計量最小的開始移除，最後模型剩下的變數通常是最重要，其參數的統計量也最大。例如，如果變數有 10 個，則候選模型就會有 10 個，其中第一個模型會包含所有 10 個變數，第二個模型會剔除 1 個最不重要的變數剩下 9 個，第三個模型剩下 8 個，以此類推，到第十個模型會剩下 1 個。
3. 算出候選模形的選模準則：根據前個步驟算出的候選模型，分別計算不同精簡程度的 R^2 成分。
4. 根據選模準則挑選最佳模型：不同精簡程度 R^2 成分挑選最佳模型的方式都是依據最大值（AIC 系列則皆是最小值），並記錄下該模型。
5. 重複執行一到四步驟，直到執行完 1000 次，觀察各種模型的比例。

這邊要注意，候選模型可能不包含以上指定的完整模型或精簡模型，因為何種變數會納入候選模型是交給向後選模的方式決定，不直接指定候選模形而

是透過選變數的方式的原因，是考量實務上變數量多的情形，因此模擬結果也會收到向後選變數的影響，這也比較貼近實務的情形。

以下將不同精簡模型的類型區分為三種，每一種都與要比較的完整模型的參數進行比較。首先先比較當隨機已知時，只選固定；再來是已知固定只選隨機部分；最後同時選固定和隨機。

表 1: Table to type1 for R_M.

	-R	R	F	O	M
$\alpha = 0$	NA	0.004	0.002	0.750	0.244
$\alpha = 0.1$	NA	0.089	0.053	0.556	0.301
$\alpha = 0.2$	NA	0.402	0.103	0.257	0.238
$\alpha = 0.3$	NA	0.711	0.071	0.101	0.117
$\alpha = 0.4$	NA	0.915	0.040	0.012	0.034
$\alpha = 0.5$	NA	0.990	0.008	NA	0.002
$\alpha = 0.6$	0.006	0.994	NA	NA	NA
$\alpha = 0.7$	0.034	0.966	NA	NA	NA
$\alpha = 0.8$	0.150	0.850	NA	NA	NA
$\alpha = 0.9$	0.495	0.505	NA	NA	NA
$\alpha = 1$	0.905	0.095	NA	NA	NA

表 2: Table to type1 for AIC.

	-R	R	F	O	M
AIC	NA	0.055	0.050	0.483	0.412
AICc	NA	0.059	0.053	0.463	0.424
mAIC	NA	0.891	0.034	0.002	0.073
cAIC	0.103	0.198	0.030	0.352	0.317

第二節　實際資料

表 3: Table to type2 for R_v.

	-R	R	F	O	M
$\alpha = 0$	NA	0.002	0.212	0.786	NA
$\alpha = 0.1$	NA	0.004	0.358	0.638	NA
$\alpha = 0.2$	NA	0.004	0.544	0.452	NA
$\alpha = 0.3$	NA	0.012	0.738	0.250	NA
$\alpha = 0.4$	NA	0.042	0.826	0.132	NA
$\alpha = 0.5$	NA	0.176	0.784	0.040	NA
$\alpha = 0.6$	NA	0.484	0.510	0.006	NA
$\alpha = 0.7$	NA	0.924	0.076	NA	NA
$\alpha = 0.8$	0.006	0.990	0.004	NA	NA
$\alpha = 0.9$	0.082	0.918	NA	NA	NA
$\alpha = 1$	0.794	0.206	NA	NA	NA

表 4: Table to type2 for AIC.

	-R	R	F	O	M
AIC	NA	NA	0.914	0.086	NA
AICc	NA	NA	0.932	0.068	NA
mAIC	0.100	0.900	NA	NA	NA
cAIC	NA	NA	0.680	0.320	NA

第五章　結論

過往有許多研究，...。

另外，這邊建議選變數的方式為向後或逐步向後（stepwise backward selection），比起向前（forward selection）和逐步向前（stepwise forward selection），向後比較能夠保留可能顯著的變數，但當樣本數遠小於變數個數時，向後的方式可能會估計不出來，此時向前和逐步向前的方式就會比較好，另一個解決這個情形的方式是納入懲罰項進去來解決資料維度較高的情形，也就是 Lasso 迴歸，但 Lasso…(待補)。倘落變數不多或是沒有時間成本的考量，也可考慮所有集合選變數的方式（Best subset selection）。

附錄一、R2 轉變

表 5: type1Rchange

	Mean(SD，Min，Max)		
	V3	V2	V1
1	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)
2	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.56 (0.05,0.39,0.70)
3	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.55 (0.05,0.38,0.70)
4	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.01 (0.01,0.00,0.04)
5	0.58 (0.05,0.42,0.73)	0.15 (0.03,0.07,0.25)	0.13 (0.02,0.07,0.23)
6	0.27 (0.03,0.17,0.38)	0.27 (0.03,0.18,0.38)	0.26 (0.03,0.18,0.38)

表 6: type2Rchange

	Mean(SD，Min，Max)		
	V3	V2	V1
1	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)	0.96 (0.01,0.93,0.98)
2	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.56 (0.05,0.39,0.70)
3	0.11 (0.05,0.00,0.29)	0.53 (0.05,0.31,0.70)	0.55 (0.05,0.38,0.70)
4	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.01 (0.01,0.00,0.04)
5	0.58 (0.05,0.42,0.73)	0.15 (0.03,0.07,0.25)	0.13 (0.02,0.07,0.23)
6	0.27 (0.03,0.17,0.38)	0.27 (0.03,0.18,0.38)	0.26 (0.03,0.18,0.38)

表 7: type3Rchange

	Mean(SD，Min，Max)			
	V3	V2	V1	V4
1	0.81 (0.03,0.73,0.88)	0.92 (0.01,0.89,0.95)	0.81 (0.03,0.72,0.88)	0.92 (0.01,0.88,0.95)
2	0.11 (0.05,0.01,0.31)	0.11 (0.05,0.01,0.31)	0.53 (0.05,0.34,0.68)	0.53 (0.05,0.35,0.67)
3	0.11 (0.05,0.01,0.31)	0.11 (0.05,0.01,0.31)	0.53 (0.05,0.34,0.68)	0.53 (0.05,0.35,0.67)
4	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)	0.00 (0.00,0.00,0.00)
5	0.58 (0.05,0.41,0.70)	0.58 (0.05,0.43,0.71)	0.15 (0.03,0.08,0.25)	0.15 (0.02,0.09,0.24)
6	0.13 (0.02,0.07,0.20)	0.23 (0.03,0.15,0.34)	0.13 (0.02,0.07,0.20)	0.24 (0.03,0.15,0.33)

附錄二、精簡模型與完整模型模擬參數比較

表 8: type1beta

	Mean(SE，SD，CP(95%))		
	V3	V2	V1
1	0.01 (1.14,1.20,0.94)	-0.02 (0.56,0.59,0.93)	-0.02 (0.53,0.55,0.94)
2	NA (NA,NA,NA)	-0.00 (0.56,0.58,0.94)	-0.01 (0.53,0.54,0.94)
3	NA (NA,NA,NA)	-0.00 (0.56,0.59,0.93)	-0.01 (0.53,0.56,0.94)
4	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.01 (0.53,0.57,0.93)
5	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.00 (0.54,0.56,0.93)
6	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.00 (0.51,0.52,0.94)

表 9: type2T

	Mean(SD，CP(95%))		
	V3	V2	V1
1	0.57 (5.18,0.97)	0.32 (4.57,0.96)	0.35 (4.42,0.95)
2	0.52 (4.13,0.94)	0.27 (3.66,0.96)	0.31 (3.62,0.96)
3	0.64 (4.41,0.97)	0.84 (4.07,0.98)	0.75 (4.06,0.96)
4	NA (NA,NA)	-0.09 (3.73,0.97)	-0.13 (3.67,0.95)
5	NA (NA,NA)	NA (NA,NA)	0.06 (0.90,0.97)
6	NA (NA,NA)	NA (NA,NA)	0.02 (0.63,0.95)

表 10: type3betaT

	Mean(SE，SD，CP(95%))			
	V3	V2	V1	
1	0.03 (1.17,1.16,0.95)	0.04 (1.15,1.15,0.95)	0.04 (0.60,0.61,0.94)	0.05 (0.57,0.60,0.93)
2	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.02 (0.60,0.62,0.94)	-0.01 (0.57,0.61,0.93)
3	NA (NA,NA,NA)	NA (NA,NA,NA)	-0.00 (0.61,0.63,0.94)	0.01 (0.58,0.61,0.94)
4	104.72 (NA,19.82,0.00)	104.77 (NA,19.32,0.00)	4.43 (NA,5.79,0.90)	4.41 (NA,5.38,0.90)
5	0.44 (NA,7.24,0.49)	0.51 (NA,6.88,0.46)	0.66 (NA,4.10,0.94)	0.68 (NA,3.82,0.94)
6	1.32 (NA,4.98,0.93)	1.25 (NA,4.50,0.93)	1.33 (NA,4.94,0.94)	1.25 (NA,4.49,0.94)
7	NA (NA,NA,NA)	0.63 (NA,3.93,0.95)	NA (NA,NA,NA)	0.63 (NA,3.93,0.95)