

多分類反應變數的邏輯斯迴歸分析

梁文敏 教授

在迴歸分析中若反應變數為二元類別變數(binary variable)，例如手術的兩個結果(存活或死亡)，若以此為反應變數，則二元邏輯斯迴歸模型(binary logistic regression model)經常會用來分析；而若反應變數為超過二元的類別變數，例如研究者欲探討不同年齡層對睡眠品質重要性的看法，以三分法的李可特量尺(3-point Likert scale：1.不重要、2.中等重要、3.很重要)測量個案對睡眠品質重要性的看法，假設研究者針對研究族群依年齡分組，每組隨機抽樣 200 人，得到資料如下：

年齡組	睡眠品質重要性			合計
	1. 不重要	2. 中等重要	3. 很重要	
1. ≤ 50 歲	35	95	70	200
2. > 50 歲	15	55	130	200

二元邏輯斯迴歸模型(binary logistic regression model)

根據上述資料，假設研究者欲沿用傳統的二元邏輯斯迴歸模型，進行不同年齡層對睡眠品質重要性的分析，則需先將反應變數的三組結果合併為二組，例如：將「不重要」、「中等重要」合併為一組、另一組則為「很重要」，反應變數併組後，帶入二元邏輯斯迴歸分析，得到結果如下(SAS 程式碼如附錄 1)。

睡眠品質重要性	很重要/(不重要或中等重要)	P
OR (95% CI)		
年齡(以小於等於 50 歲為比較組)		
大於 50 歲	3.5 (2.3 ~5.2)	<0.0001

適合度指標：AIC=522.0、SC=521.5、-2LogL=518.0

OR：Odds Ratio 勝算比。

勝算比=ODD(大於 50 歲)/ODD(小於等於 50 歲)、ODD：勝算

上表可解釋為大於 50 歲組認為睡眠品質很重要的勝算是小於等於 50 歲組的 3.5 倍(P<0.0001)，亦即年紀愈大族群愈感受到睡眠品質的重要，且具統計上顯著的意義。

假設研究者不希望併組，則上述二元邏輯斯迴歸模型不再適用，此時可考慮用下列兩個模型：(1)廣義邏輯斯迴歸模型(generalized logistic regression

model)，亦稱為多元邏輯斯迴歸模型(multinomial logistic regression model)；(2) 等比勝算邏輯斯迴歸模型(proportional odds logistic regression model)，亦稱為序位邏輯斯迴歸模型(ordered logistic regression model)或累積邏輯斯迴歸模型(cumulative logistic regression model)。茲簡單介紹此兩個模型之定義，並探討其主要差別，同時提供如何選擇的一些建議。

廣義邏輯斯迴歸模型(generalized logistic regression model)

此模型首先指定某一組為參考組，接著其他組一一與此參考組做比較，其數學式如下：

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \alpha_j + \beta_j x, j = 2, \dots, J$$

若反應變數分三類，例如不重要、中等重要、很重要，則可得兩個數學式如下：

$$\log\left(\frac{\pi_{\text{中等重要}}}{\pi_{\text{不重要}}}\right) = \alpha_2 + \beta_2 x, \text{ 及 } \log\left(\frac{\pi_{\text{很重要}}}{\pi_{\text{不重要}}}\right) = \alpha_3 + \beta_3 x,$$

以上兩個數學式，可視為兩個二元邏輯斯迴歸模型。

今根據上述 400 人的資料，分析結果如下(SAS 程式碼如附錄 2)：

睡眠品質重要性	中等重要/不重要		很重要/不重要	
	OR (95%CI)	P	OR (95%CI)	P
年齡(以小於等於 50 歲為比較組)				
大於 50 歲	1.4 (0.7 ~ 2.7)	0.3930	4.3 (2.2 ~ 8.5)	<0.0001

適合度指標：AIC=750.1、SC=749.3、-2LogL=742.1

OR：Odds Ratio 勝算比。

勝算比=ODD(大於 50 歲)/ODD(小於等於 50 歲)、ODD：勝算

上表可解釋為大於 50 歲組認為睡眠品質中等重要比上不重要的勝算是小於等於 50 歲組的 1.4 倍(P=0.3930)，大於 50 歲組認為睡眠品質很重要比上不重要的勝算是小於等於 50 歲組的 4.3 倍(P<0.0001)，後者具有統計上顯著的意義。

等比勝算邏輯斯迴歸模型(proportional odds logistic regression model)

此模型原則上係假設反應變數為一個序位變數(即在同一條量尺上、但不等距)，在建立模型時，每次找一個切點將資料分為兩組，建立一個二元邏輯斯迴歸模型，其數學式如下：

$$\log\left(\frac{1-P(Y \leq j)}{P(Y \leq j)}\right) = \alpha_j + \beta x, j = 1, \dots, J-1$$

若反應變數分三類，例如不重要、中等重要、很重要，則可得兩個數學式如下：

$$\log\left(\frac{\pi_{\text{中等重要}+\text{很重要}}}{\pi_{\text{不重要}}}\right) = \alpha_1 + \beta x, \text{ 及 } \log\left(\frac{\pi_{\text{很重要}}}{\pi_{\text{不重要}+\text{中等重要}}}\right) = \alpha_2 + \beta x,$$

以上兩個數學式，可視為兩個二元邏輯斯迴歸模型。

今根據上述 400 人的資料，分析結果如下(SAS 程式碼如附錄 3)：

睡眠品質重要性	(很重要或中等重要)/不重要		很重要/(中等重要或不重要)	
	OR1 (95%CI)	P	OR2 (95%CI)	P
年齡(以小於等於 50 歲為比較組)				
大於 50 歲	3.3 (2.2 ~ 4.9)	<0.0001	3.3 (2.2 ~ 4.9)	<0.0001

適合度指標：AIC=748.9、SC=748.3、-2LogL=742.9

OR：Odds Ratio 勝算比

勝算比=ODD(大於 50 歲)/ODD(小於等於 50 歲)、ODD：勝算

上表可解釋為大於 50 歲組認為睡眠品質重要的勝算是小於等於 50 歲組的 3.3 倍(P<0.0001)，亦即年紀愈大族群愈感受到睡眠品質的重要，且具統計上顯著的意義。

上述之「等比勝算邏輯斯迴歸模型」有一很強的假設，此假設為不同定義的勝算在不同組別之間呈相等的比例關係，例如：上表中「很重要或中等重要」的「勝算」(以 ODD1 表示)、與「很重要」的「勝算」(以 ODD2 表示)，在個別年齡組中，均存在一個相同的等比關係，亦即 $ODD1 = a * ODD2$ ，因此在計算「勝算比」時，即 $OR1 = ODD1(>50 \text{ 歲}) / ODD1(\leq 50 \text{ 歲})$ ；或 $OR2 = ODD2(>50 \text{ 歲}) / ODD2(\leq 50 \text{ 歲})$ ，兩者所得到的值是相同的。從上表之結果中可發現兩個不同定義的勝算比皆為 3.3，且 P 值皆為<0.0001，因此在結果的解釋上也變得簡單有力。

一般在運用「等比勝算邏輯斯迴歸模型」進行分析時，電腦分析報表中會提供檢定「等比勝算假設是否成立」的分析結果，以提供使用者判斷是否可使用此模型，檢定結果若 P 值大於 0.05，則表示等比勝算的假設成立，亦即等比勝算模型適合用來推論結果。上例中檢定「等比勝算假設」之結果如下表所示，P=0.3834 表示「等比勝算邏輯斯迴歸模型」適用於此筆資料。

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
0.7598	1	0.3834

比較與建議

從上列三個表格中，相信可以很容易了解到此三個模型所得到之勝算比，其定義各有所不同。依筆者意見，若您的反應變數是序位的，例如上例中以三分法的李可特量尺(1.不重要、2.中等重要、3.很重要) 測量睡眠品質的重要性，則上面三種模式均適用。但依筆者個人的選擇，若等比勝算假設成立，則會建議選用「等比勝算邏輯斯迴歸模型」，因為相較之下，它較「二元邏輯斯迴歸模型」忠於原始的資料、保有較大的檢定力，而且在解釋結果時，它較「廣義邏輯斯迴歸模型」容易解釋，尤其是當模型中還有其他共變數時。然而，當您的反應變數並非是序位資料，例如研究者欲探討的是不同年齡層的攝護腺癌患者其治療方式選擇之異同時，若治療方式分三類(1.開刀切除、2.放射療法、3.荷爾蒙療法)，則此類的分析，用「廣義邏輯斯迴歸模型」較為恰當。

附錄 1：

```
DATA NEW;
INPUT  AGE SLEEPIMP N @@;
IF SLEEPIMP = 1 OR SLEEPIMP = 2 THEN SLEEPIMP1=1;
IF SLEEPIMP = 3 THEN SLEEPIMP1= 3;
DATALINES;
1 1 35    1 2 95    1 3 70
2 1 15    2 2 55    2 3 130
;
PROC LOGISTIC DATA=NEW;
WEIGHT N;
CLASS  AGE(REF='1') /PARAM=REF; /*指定<=50歲組為參考組*/
MODEL SLEEPIMP1(REF='1')= AGE; /*指定以睡眠品質(不重要或中等重要)為參考組*/
RUN;
```

附錄 2：

```
DATA NEW;
```

```
...  
PROC LOGISTIC DATA=NEW;  
WEIGHT N;  
CLASS AGE(REF='1') /PARAM=REF; /*指定<=50歲組為參考組*/  
MODEL SLEEPIMP(REF='1')= AGE/ LINK=GLOGIT; /*指定以睡眠品質不重要為參考組*/  
RUN;
```

附錄 3：

```
DATA NEW;  
...  
PROC SORT DATA=NEW;  
BY DESCENDING SLEEPIMP; /*指定資料按照睡眠品質重要性由大排到小*/  
PROC LOGISTIC DATA=NEW ORDER=DATA;  
/*ORDER=DATA：指定計算「勝算」時，重要性大的放在分子*/  
WEIGHT N;  
CLASS AGE(REF='1') /PARAM=REF;  
MODEL SLEEPIMP= AGE/ LINK=CLOGIT;  
RUN;
```