

**109-1 迴歸分析\_期末報告**

**2018 年公投第 14 案：  
廣義線性迴歸模型探討**

**R08227112 林子堯**

國立台灣大學心理學研究所

## 壹、緒論

### 一、前言 ( 見附件 )

### 二、文獻回顧 ( 見附件 )

### 三、研究目的與假設

本研究在 2018 年公投第 14 案之同意率與同性戀態度有某種程度以上關聯的假設下，試以「廣義線性迴歸模式」(generalized linear regression model, GLM) 找出影響同意率的主要因素為何。此筆資料收集了全台 368 個「鄉、鎮、市、區」為單位的資料點，以該公投的同意率 (agree\_rate, Y) 作為反應變項，預測變項則有：投票率 (vot\_rate, X1)、位置 (location, X2，為類別變項，其包含了直轄市、縣市及離島這三種類別)、性別比 (gender\_ratio, X3)、年齡中位數 (age\_med, X4)、大學教育程度 (college\_ratio, X5)、結婚率 (married\_ratio, X6) 以及薪水中位數 (salary\_med, X7)。透過前述這些變項以及之間的交互作用項來建立一解釋性為主的模型。依照過往文獻的結果，本研究預期在直轄市比起非直轄市、或是教育程度越高地區，對於同意率會偏向較高；而性別比越高、結婚率越高的地區，則同意率越低。至於在文獻中尚未探討過的投票率，和過去研究沒有共同結論的年齡中位數和薪水中位數對於同意率，正向、負向或是無關連都有可能。

## 貳、資料來源與前處理

### 一、資料來源與前處理 ( 見附件 )

### 二、變數概況

表 1 中顯示反應變項 ( 同意率 ) 以及六個解釋變項以及權重 ( 有效票數 ) 的基本描述統計概況。其中位置為類別變項，標記出全台 368 個鄉鎮市區中，其中有 158 個地區是屬於直轄市，194 個屬於一般縣市以及剩下 16 個區域則屬於離島，此外各自變項的直方圖 ( 依照所在地區分色 ) 繪於圖 1 中，可發現其實許多變項在這三種位置類別下其實是有差異的，可能暗示著這些變項中有交互作用項的存在。圖 2 是將所有變項的資料繪製在[台灣地圖](#)上，其中各個鄉鎮市區的顏色代表著對此公同案的同意率程度，用滑鼠點擊該地區時 ( 如圖 2 所在示 )，則可以顯示其他變項的數值。值得注意的是，在直轄市的地區相較於非直轄市地區同意率相較起來較高，然而在所有地區同意率最高的地方卻是位於離島的「連江縣東引鄉」，比起其他同樣是離島地區 ( 普遍同意率偏低 ) 的來說可能是比較特別的存在。圖 3 為七個預測變項與反應變項

的散佈相關矩陣圖，其中除了投票率和性別比皆與同意率沒有顯著的相關，其他變項皆有顯著的相關，需要注意的是解釋變項之間也有高度相關，例如在大學教育比率分別與薪水中位數 ( $r(367) = .87, p < .001$ )、性別比 ( $r(367) = -.71, p < .001$ )，在後續分析上可能需要注意是否有共線性的問題。

## 參、分析方法

### 一、分析流程與軟體

本研究分析流程如圖 4 所示，並使用 R 軟體 (R Core Team, 2020)、tidyverse 套件 (Wickham et al., 2019) 與其他相關套件來進行資料整理、繪圖與統計分析。

### 二、二項式迴歸模型

透過先前的文獻回顧，本研究收集影響 14 案公投同意率的可能影響因子，並在前段文章中詳述各項獨變項的情形。由於投票行為（同意或不同意）服從於白努力分配，而本研究又以鄉政市區作為基本單位，將該區的有效票數  $N_i$  集結起來得到觀察到的同意比率  $\bar{Y}_i$  與其他解釋變項  $\mathbf{X}$  的數值，對此適合用二項式迴歸模型 (binomial regression model) 來模擬，背後受潛在的參數  $\pi_i$ （同意率）給控制，模型為：

$$N_i \bar{Y}_i \sim \text{Binomial}(N_i, \pi_i), \quad \pi_i = g(\eta_i) = g(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad i = 1, \dots, 368$$

其中  $g(\cdot)$  為鏈結函數， $\boldsymbol{\beta}$  為帶估計的參數， $\mathbf{X}$  也先考慮  $X_1$  至  $X_7$  的主要效果。對此，我嘗試使用四種鏈結函數，包含 identity, logit, probit 以及 cloglog function，來分別擬合此筆資料。結果如表 2，這四種鏈結函數下的模型實際上結果差異不大，若使用 AIC、BIC 與 deviance 做為模型表現評估的指標得知，identity link 在這三種數值中皆是最小的。主要原由在於觀察到的同意率  $\bar{Y}_i$  主要都集中在於 0.2~0.4 之間，根本沒有那些很接近 0 或是很接近 1 的數值，因此我們較不用擔心  $\mathbf{X}\hat{\boldsymbol{\beta}}$  會超出  $\pi$  的範圍。我們也可從圖 5 中得知，這些  $\hat{\eta}$  估計值也都坐落在於 logit, probit 以及 cloglog link 斜直線的區段，較沒使用到這些非線性函數轉換上的優勢，在此我便選擇使用 identity link 作為後續模型分析上使用的鏈結函數。

接續我去看淨殘差圖 (partial residuals plot)，來檢查殘差與解釋變項的關係，從圖 6 得知，每個變項確實有明顯的關係，但看不出來需要有高次項的解釋變項影響在其中。另一方面，從先前的變數概況章節中我們得知許多變數之間可能存在著某些關聯，因此接下來我將考慮所有變項之間的一階交互作用項而不考慮高次項效果，來搜

尋可能的最佳解釋模型。我使用 R 原生套件中的函式來執行逐步迴歸 (R Core Team, 2020)，並分別用 AIC、BIC 兩種準則以及前向、反向、雙向選擇 (forward, backward or both selection) 的方式，因此會找出六個候選模型 (有可能會找出相同模型)。然而逐步迴歸的缺點在於它並不會搜尋完所有的子模型，因此我又多使用 glmulti 套件 (Calcagno, 2020) 來搜尋前幾佳的 GLM，也分別尋找 AIC、BIC 兩種準則下的前三好模型，因此在此步驟我挑選了 12 個候選模型，並同時評估其 AIC、BIC、deviance 以及留一交叉驗證法 (leave-one-out cross-validation) 下的平均平方預測誤差 (mean squared prediction error) 之大小，並將結果放在表 3 中。

最終，我同時考量前述指標大小以及若模型存有交互作用時其相對應的主要效果也應在模型之中，因此我偏好 M2 作為我暫時的最佳模型。此結果大部分的估計值  $\hat{\beta}$  皆達顯著 ( $\alpha = .05$  下)，然而我們還是可以觀察到此模型還是有蠻大的過度離散 (overdispersion) 現象產生 ( $\hat{\phi}_p = 44.52, \hat{\phi}_p = 44.57$ )，因此我們可以再使用「準二項式迴歸模型」(quasi-binomial regression)，加入 dispersion 參數來降低 deviance 的大小。然而這麼做的結果就是  $\hat{\beta}$  的標準誤變大，原本找到的最佳模型 M2 的  $\hat{\beta}$  幾乎都變成不顯著了！與前者的比較結果可以見表 4，對此我面臨一個兩難的困境，一是如果我不使用 quasi-binomial regression 去擬合的話，M2 還是會有很大的過度離散現象存在，二是如果我選擇用 quasi likelihood 的做法，我便沒法使用 AIC、BIC 來做模型的選擇。但我後來想到我既然都是使用 identity function 作為鏈結函數的選擇，那何不直接使用一般線性模型 (general linear model) 來估計同意率 ( $\pi$ ) 呢？

### 三、加權最小平方估計之多元迴歸模型

一般線性模型是假設誤差的平均值為 0、變異數為定值  $\sigma^2$  的常態分佈，在此情況下我們可以使用最小平方方法或是最大概似估計法來估計參數。 $\bar{Y}_i | \mathbf{x}_i$  的條件期望為  $\pi_i$ ，而條件變異為  $\pi_i(1 - \pi_i)/N_i$ ，對此我假設同意率的 (條件) 分配為

$$N_i \bar{Y}_i \sim \text{Binomial}(N_i, \pi_i), \quad \bar{Y}_i | \mathbf{x}_i \sim N\left(\pi_i = \mathbf{x}_i^\top \boldsymbol{\beta}, \frac{\pi_i(1 - \pi_i)}{N_i}\right), \quad i = 1, \dots, 368$$

其中同意率的變異本不會是定值，會受到有效票數以及本身的平均共同影響，因此該使用加權最小平方估計法 (weighted least square estimation) 來估計這些參數。最後我才發現如果是使用此方法的話結果與用 identity link 的 quasi-binomial regression 是相同的！只是差別在於前者還是有分配的假設，則後者沒有。

在估計完所有主要變數的模型後，檢查淨殘差圖發現應該不用考量變項之高次項加入模型中，同時也用 Box-Cox 轉換檢驗同意率經過轉換  $(\pi_i^{(\lambda)})$  是否會使 SSE 變小。圖 7 為在不同  $\lambda$  值下的 SSE 變化，其最小值發生在 1.0~1.5 之間，表示其實不用透過 Box-Cox 轉換就能達到較好的預測且用原始  $\pi_i$  方便後續的解釋，我便考量加入一階交互作用項來搜尋可能的最佳解釋模型。同樣是使用 R 來執行逐步迴歸，依據 AIC、BIC 指標和分別使用前向、反向、雙向選擇方式，找出六個候選模型外，另一方面，我也利用 leaps (Lumley, 2020) 套件來找其他候選模型，該套件的優點是能列出參數個數與不同模型選擇指標之大小，包括 adjust  $R^2$ 、Mallows's CP、AIC、BIC 等 (見圖 8)，方便我們選擇需要多少數量的解釋變項能得到最好的效果。我從中再挑選了 6 個候選模型，共 12 個候選模型最後再次去評估模型表現，包括：AIC、BIC、deviance 和 mean squared prediction error (LOOCV)，結果列於表 5。最終我選擇 LM10 為最佳模型，其殘差圖與常態分佈 Q-Q 圖於圖 9 與圖 10 所示，經過加權估計後的殘差沒有很明顯的變異數不同值，然而卻不完全符合常態分佈的假設，特別是再左右尾的位置有一些偏差，但在這邊我不認為有太大的影響。另外，在離群值 (outlier) 與影響點 (influence point) 判斷上 (見表 6)，雖然有些地區的資料點在 DFFIT 與 DFBETA 的指標上可能暗示出是影響值，但其影響的數值其實並不大，且在用 Cook's distance 指標上並沒有任何資料點超過經驗法則的 0.5 或是 1 (見圖 11)。除此之外，由與本研究是想要找出同意率大小的解釋性模型，而這 368 鄉政市區資料理論上都是我該考慮的範圍內，雖說有些資料點明顯與整體趨勢不同，但經果上述考量我並不打算刪除資料或視之為離群值。

## 肆、結果與討論

### 一、結果

本研究就探討全台 368 個「鄉、鎮、市、區」為單位下，針對公投第 14 案，找出影響同意率的人口學變相因素為何，最終得到的解釋性模型為：

$$\begin{aligned}
 (\text{Agree\_rate}) = \hat{\pi} &= \mathbf{X}\hat{\beta} \\
 &= 2.54 - 0.19(\text{Vote}) + 0.46(\text{Municipality}) + 1.06(\text{Island}) - 0.28(\text{Gender}) - 0.03(\text{Age}) \\
 &\quad - 3.44(\text{Married}) - 1.96(\text{College}) + 3.33 \times 10^{-4}(\text{Salary}) \\
 &\quad + 2.41 \times 10^{-3}(\text{Municipality: Age}) - 0.90(\text{Municipality: Married}) \\
 &\quad - 0.12(\text{Municipality: College}) + 1.00 \times 10^{-4}(\text{Municipality: Salary}) \\
 &\quad - 0.02(\text{Island: Age}) - 0.65(\text{Island: College}) + 0.76(\text{Gender: College}) \\
 &\quad + 0.06(\text{Age: Married}) - 7.04 \times 10^{-6}(\text{Age: Salary}) + 2.10(\text{Married: College})
 \end{aligned}$$

詳細的統計報表可見表 7，其中  $\hat{\pi}$  透過加權最小平方估計下多元迴歸模型估計獲得。

## 二、模型解釋

雖然說上式有點複雜，但我們可以依照不同位置 (location) 的類別來看上模型的結果，如果是在一般縣市 (Municipality = Island = 0)：

$$\begin{aligned}(\text{Agree\_rate}) = & 2.54 - 0.19(\text{Vote}) - 0.28(\text{Gender}) - 0.03(\text{Age}) - 3.44(\text{Married}) - 1.96(\text{College}) \\ & + 3.33 \times 10^{-4}(\text{Salary}) + 0.76(\text{Gender: College}) + 0.06(\text{Age: Married}) \\ & - 7.04 \times 10^{-6}(\text{Age: Salary}) + 2.10(\text{Married: College})\end{aligned}$$

如果是在直轄市 (Municipality = 1, Island = 0)：

$$\begin{aligned}(\text{Agree\_rate}) = & 3.00 - 0.19(\text{Vote}) - 0.28(\text{Gender}) - 0.029(\text{Age}) - 4.34(\text{Married}) - 2.08(\text{College}) \\ & + 4.33 \times 10^{-4}(\text{Salary}) + 0.76(\text{Gender: College}) + 0.06(\text{Age: Married}) \\ & - 7.04 \times 10^{-6}(\text{Age: Salary}) + 2.10(\text{Married: College})\end{aligned}$$

如果是在離島 (Municipality = 0, Island = 1)：

$$\begin{aligned}(\text{Agree\_rate}) = & 3.60 - 0.19(\text{Vote}) - 0.28(\text{Gender}) - 0.05(\text{Age}) - 3.44(\text{Married}) - 2.61(\text{College}) \\ & + 3.33 \times 10^{-4}(\text{Salary}) + 0.76(\text{Gender: College}) + 0.06(\text{Age: Married}) \\ & - 7.04 \times 10^{-6}(\text{Age: Salary}) + 2.10(\text{Married: College})\end{aligned}$$

整體來說，在主要效果的部分：

1. 投票率、性別比、年齡中位數、結婚率或是大學教育程度越高，同意率會越低；
2. 在直轄市或是離島的地區，同意率會大幅增高；
3. 相反的薪水中位數越高的地區對於此案公投的同意率也會越高。

但是考量地區與其他變項交互作用項的影響：

1. 直轄市比起一般縣市來說，年齡中位數所產生的負向效果變小，結婚率、大學教育程度的負向效果以及薪水所帶來的正向效果皆增強；
2. 離島比起一般縣市來說，則是使年齡中位數、大學教育程度越的負向效果變大。

除此之外，其他變相之間的交互作用項的影響：

1. 性別比與大學教育程度、年齡與結婚率、結婚率與大學教育程度這三者的交互作用項較效果越高，則同意率越高；
2. 年齡中位數與薪水中位數的交互作用項越高，則同意率越低。

實際上解釋變項的對於同意率的效果不能只單看主要效果而已，若在其他變項在固定的情況下，性別比、年齡中位數、結婚率、大學教育程度以及薪水中位數的效果方向可能會被反轉，此外，不同地區的效果程度也會有所變化。

## 三、研究限制與改進方向（見附件）

## 參考文獻

- 王晴鋒 ( 2013 ) 。 「恐同症」 的根源——基于宗教, 现代性和文化的阐释。 *吉首大学学报：社会科学版* , 34(1) , 59-64 。
- Adamczyk, A., & Liao, Y. C. (2019). Examining public opinion about LGBTQ-related issues in the United States and across multiple nations. *Annual Review of Sociology*, 45, 401-423.
- Becker, A. B., & Scheufele, D. A. (2009). Moral politicking: Public attitudes toward gay marriage in an election context. *The International Journal of Press/Politics*, 14(2), 186-211.
- Bowman, R. (1979). Public attitudes toward homosexuality in New Zealand. *International Review of Modern Sociology*, 9, 229-238.
- Calcagno, V. (2020). *glmulti: Model selection and multimodel inference made easy* [R package version 1.0.8]. <https://CRAN.R-project.org/package=glmulti>
- Gelbal, S. & Duyan, V. (2006). Attitudes of university students toward lesbians and gay men in Turkey. *Sex Roles*, 55, 573-579.
- Glenn, N. D., & Weaver, C. N. (1979). Attitudes toward premarital, extramarital, and homosexual relations in the U.S. in the 1970s. *Journal of Sex Research*, 15(2), 108-118.
- Herek, G. M. (1984). Beyond “homophobia”: A social psychological perspective on attitudes toward lesbians and gay men. *Journal of Homosexuality*, 10(1-2), 1-21.
- Jäckle, S., & Wenzelburger, G. (2015). Religion, religiosity, and the attitudes toward homosexuality—A multilevel analysis of 79 countries. *Journal of Homosexuality*, 62(2), 207-241.
- Lumley, T. (2020). *leaps: Regression subset selection* [R package version 3.1]. <https://CRAN.R-project.org/package=leaps>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org>
- Seltzer, R. (1992). The social location of those holding antihomosexual attitudes. *Sex Roles*, 26, 391-398.
- West, K., & Cowell, N. M. (2015). Predictors of prejudice against lesbians and gay men in Jamaica. *The Journal of Sex Research*, 52(3), 296-305
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Kuhn, M. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

圖

圖 1 解釋變項與預測變項之直方圖（不同顏色表示三種地區：綠色為直轄市、藍色為離島、紅色為一般縣市的鄉政市區）。

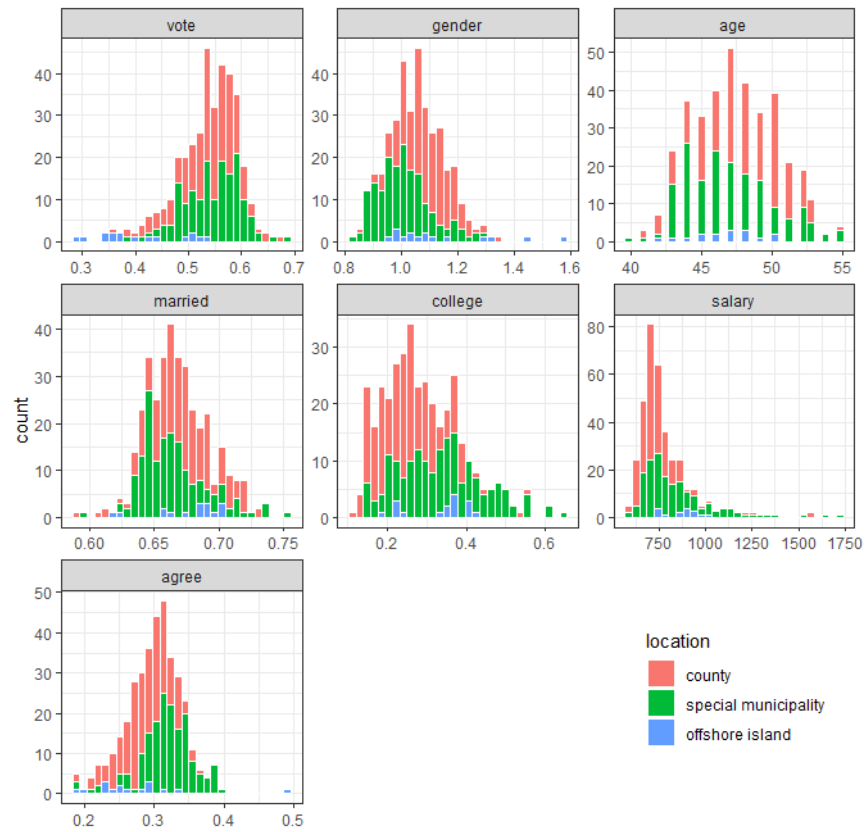
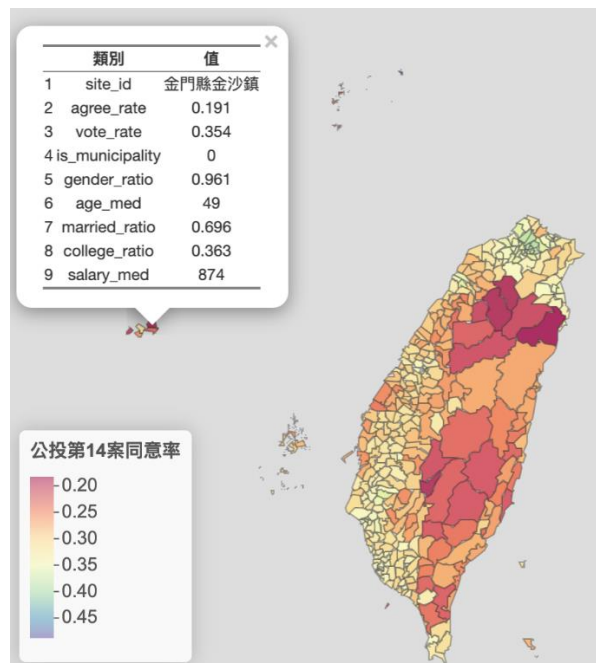


圖 2 全台 368 個鄉政市區同意率程度地圖。



動態網頁連結：[https://xup6y3ul6.github.io/Regression\\_Analysis/RA\\_final\\_report/referendum14\\_map.html](https://xup6y3ul6.github.io/Regression_Analysis/RA_final_report/referendum14_map.html)



圖 3 解釋變項與預測變項之散佈相關矩陣圖：左下為相關係數，右上圖中綠線為 LOESS 迴歸線。

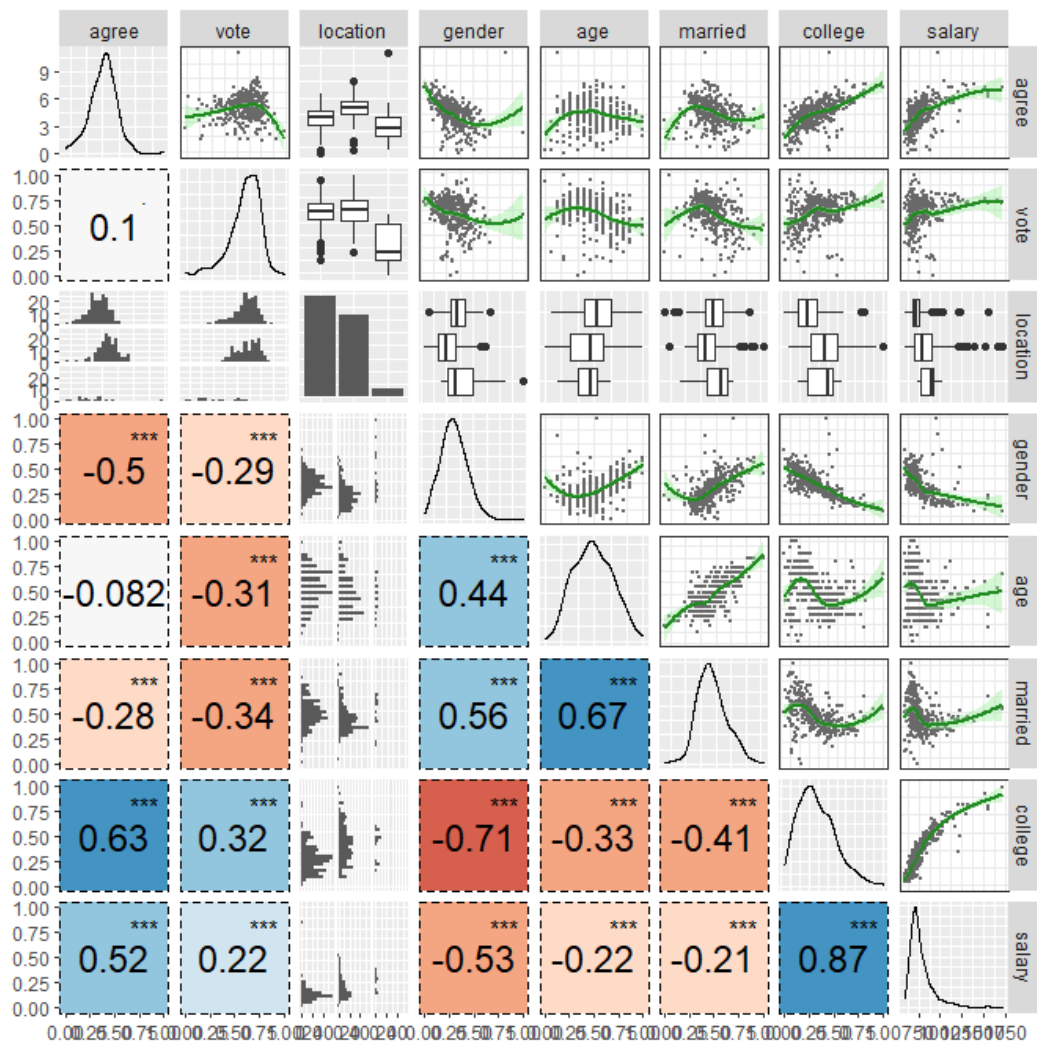


圖 4 本研究模型建立之流程。

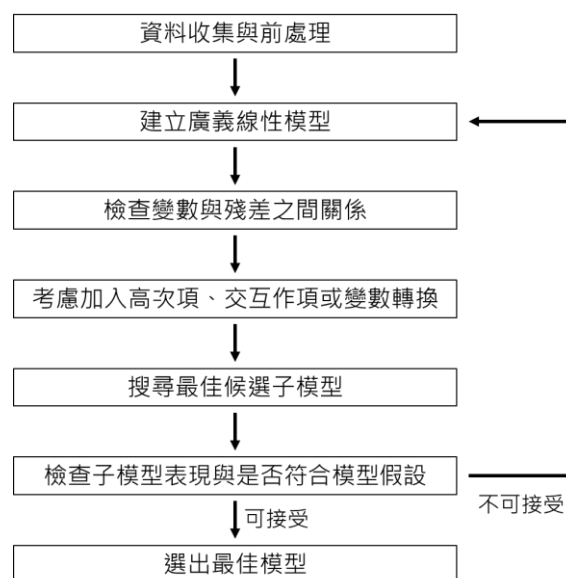


圖 5 二項式迴歸四種鏈結函數  $g(\cdot)$  下之表現：( 左半部 ) 預測值  $\hat{\pi} = g(\hat{\eta}) = g(X\hat{\beta})$  與真實值  $\pi$  散佈圖；( 右半部 ) 則是未經轉換前  $\hat{\eta}$  與真實值  $\pi$  散佈圖，其中紅線表示鏈結函數  $g(\cdot)$  樣貌。

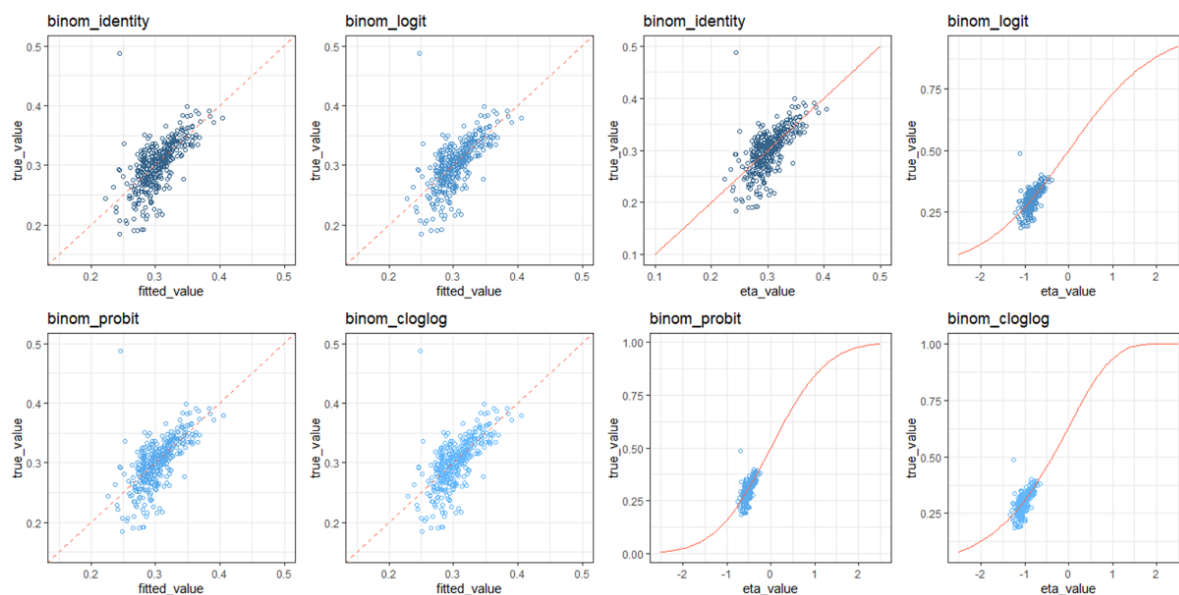


圖 6 淨殘差圖。每一張圖代表的是拿掉該某一變項下的模型殘差與該變項之散佈圖。

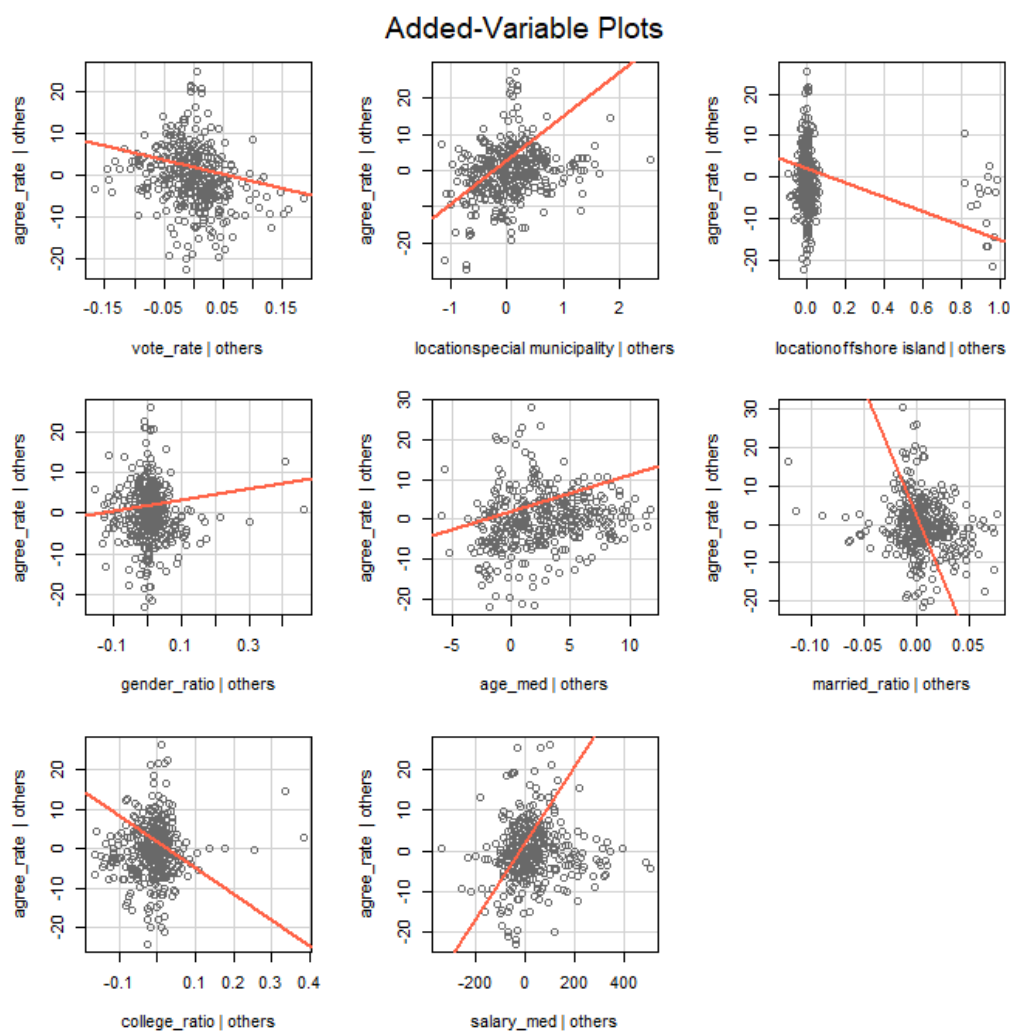


圖 7 Box-Cox 轉換圖。將預測變項經過不同  $\lambda$  值轉換 ( $\pi_i^{(\lambda)} = \{\frac{\pi_i^{\lambda-1}}{\lambda(GM(\pi))^{\lambda-1}}\}^{\lambda \neq 0} \{GM(y)\ln(\pi_i)\}^{\lambda=0}$ ，其中  $GM(\pi) = \sqrt[n]{\pi_1\pi_2\cdots\pi_n}$ )，找尋何者值能使 SSE 最小。

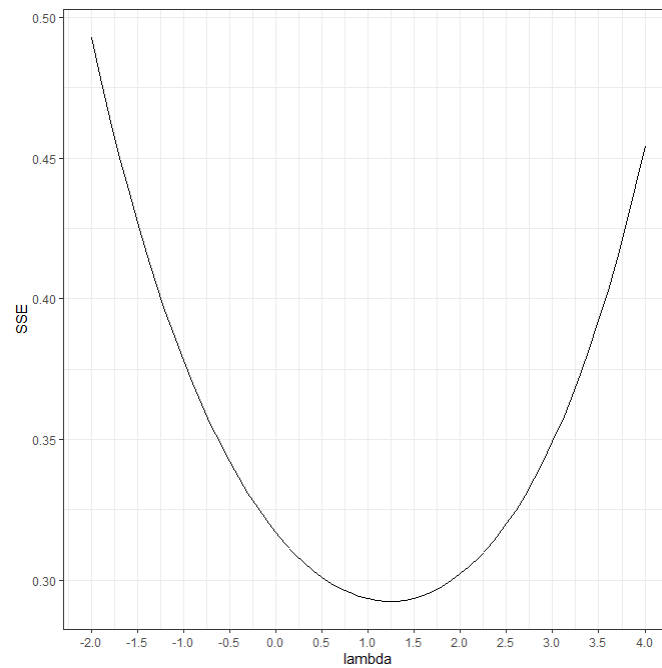


圖 8 透過 R 中 leaps 套件 (Lumley, 2020) 來找尋不同參數個數下子模型與不同模型選擇指標 (包括  $R^2$ 、adjust  $R^2$ 、Mallows's CP、AIC、BIC 與 PRESS) 之大小比較。

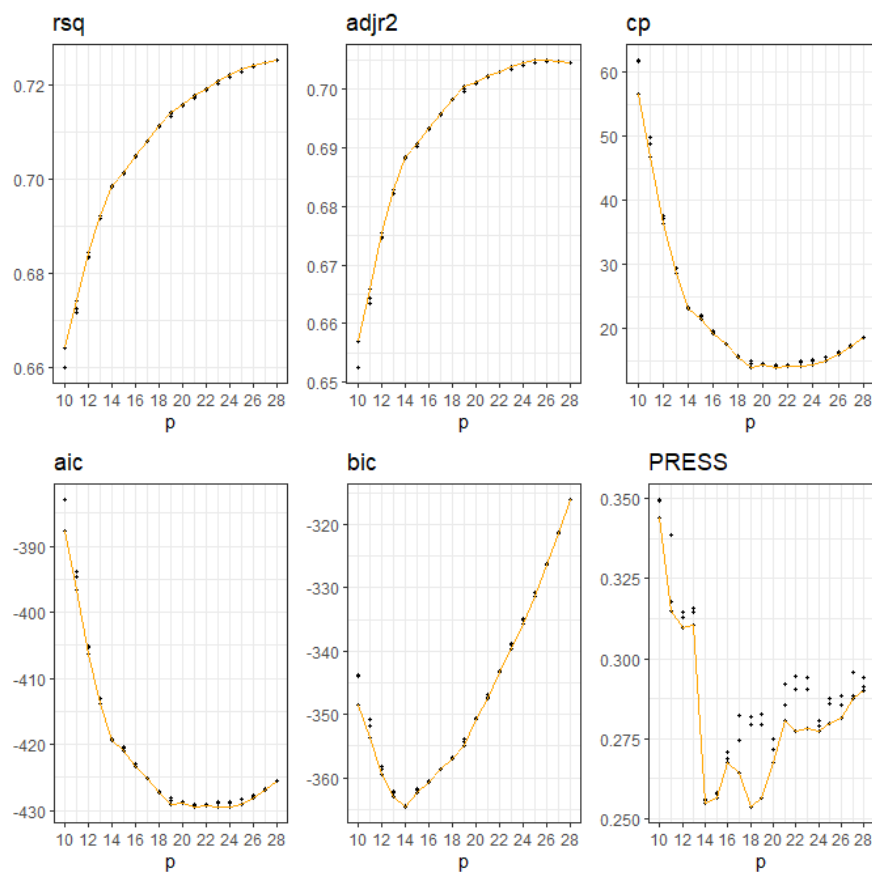


圖 9 本研究最終選擇的最佳模型 LM10 ( 見表 5、7 ) 之殘差圖。

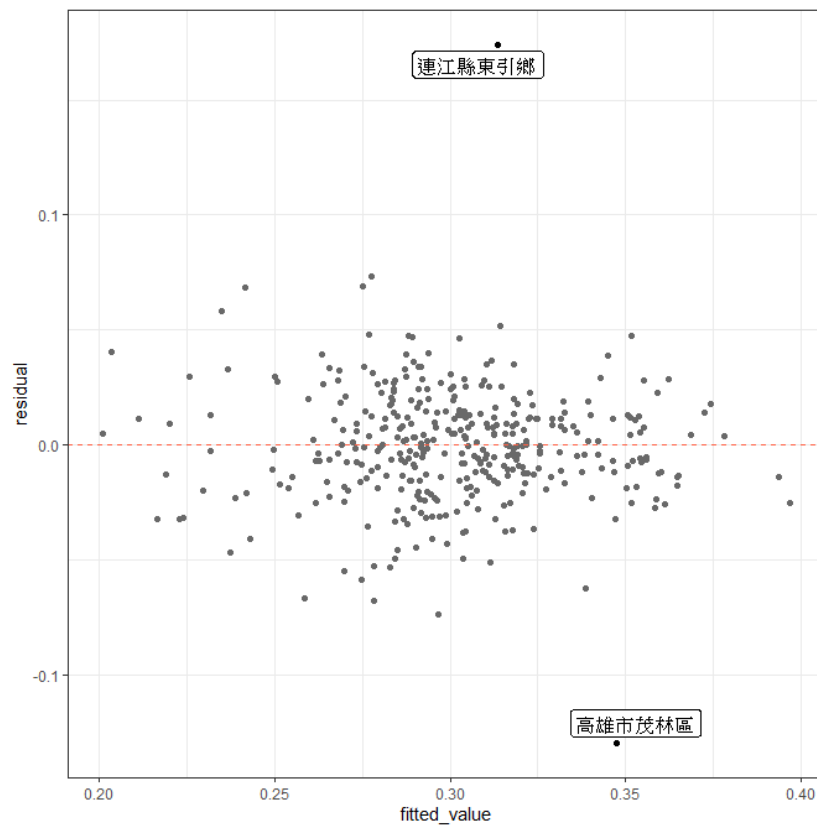


圖 10 本研究最終選擇的最佳模型 LM10 ( 見表 5、7 ) 之常態分配 Q-Q 圖。

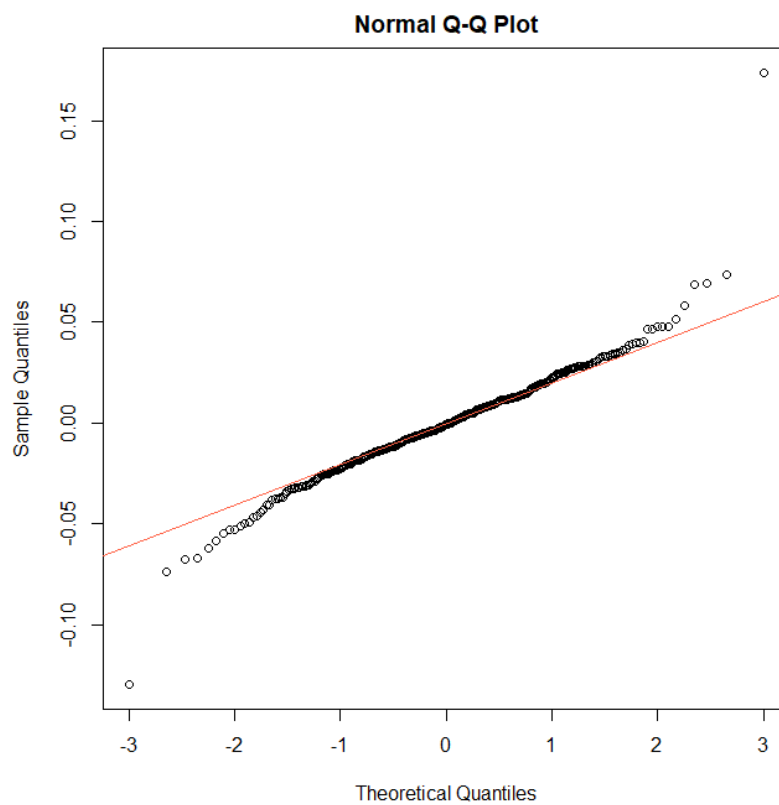
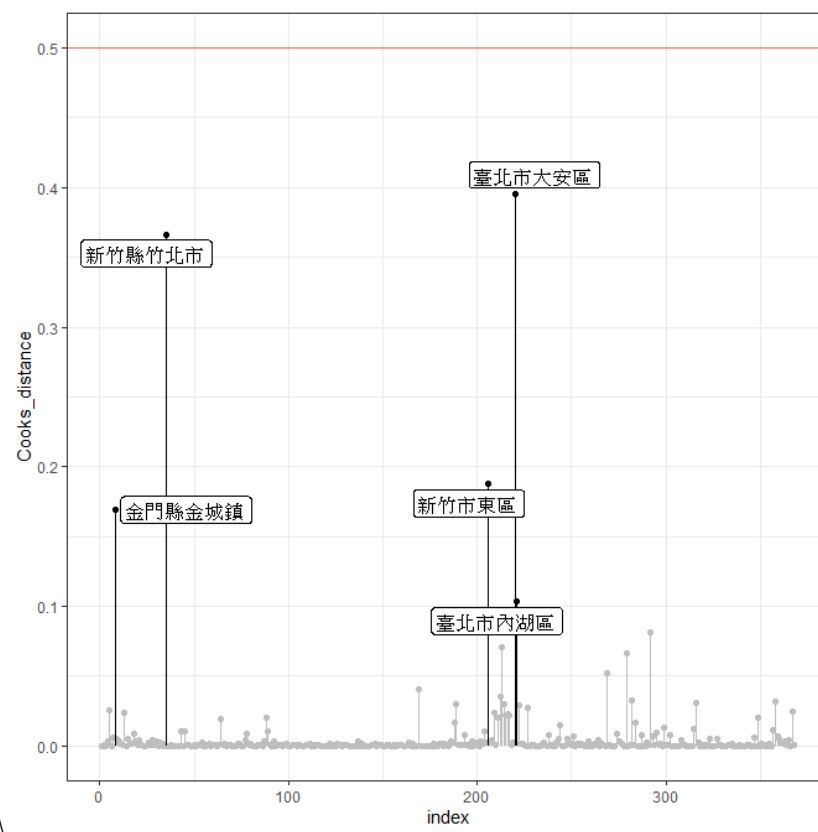


圖 11 本研究最終選擇的最佳模型 LM10 ( 見表 5、7 ) Cook's distance 值，其中紅線為經驗標準 0.5。



# 表

表 1 解釋變項、預測變項與權重之基本描述統計。

	反應 變項	預測變項						權重
	同意率	投票率	性別比	年齡 中位數	結婚率	教育 程度	薪水中位 數 (千元)	有效 票數
最小值	0.184	0.291	0.833	40.000	0.590	0.117	587.0	161
Q1	0.277	0.505	0.986	45.000	0.652	0.214	691.0	5904
中位數	0.305	0.542	1.047	47.000	0.666	0.275	735.0	13260
平均值	0.301	0.534	1.053	47.394	0.668	0.290	784.9	28076
Q3	0.326	0.577	1.118	50.000	0.683	0.356	823.2	34170
最大值	0.487	0.689	1.582	55.000	0.753	0.653	1722.0	247263
標準差	0.040	0.060	0.102	2.978	0.025	0.101	162.8	37197

表 2 二項式迴歸四種鏈結函數  $g(\cdot)$  下參數估計結果，其中 dispersion parameter 為預設值 1。

Parameter	Identity link		Logit link		Probit link		Cloglog link	
	estimate	s.e.	estimate	s.e.	estimate	s.e.	estimate	s.e.
(Intercept)	6.00e-01***	7.91e-03	6.00e-01	3.68e-02	3.48e-01	2.23e-02	1.91e-01	3.06e-02
Vote_rate	-1.59e-01***	4.63e-03	-7.44e-01	2.16e-02	-4.51e-01	1.31e-02	-6.20e-01	1.80e-02
Location (municipality)	1.66e-02***	3.76e-04	7.71e-02	1.75e-03	4.68e-02	1.06e-03	6.40e-02	1.46e-03
Location (island)	-6.26e-02***	1.70e-03	-3.02e-01	8.56e-03	-1.82e-01	5.11e-03	-2.54e-01	7.31e-03
Gender_ratio	-2.18e-02***	5.15e-03	-9.78e-02	2.47e-02	-5.99e-02	1.49e-02	-7.89e-02	2.07e-02
Age_med	2.04e-03***	7.06e-05	8.57e-03	3.22e-04	5.32e-03	1.96e-04	6.79e-03	2.66e-04
Married_ratio	-5.47e-01***	1.31e-02	-2.54e+00	6.13e-02	-1.54e+00	3.71e-02	-2.11e+00	5.11e-02
College_ratio	6.48e-02***	5.90e-03	3.13e-01	2.75e-02	1.89e-01	1.67e-02	2.67e-01	2.29e-02
Salary_med	7.21e-05***	1.96e-06	3.15e-04	8.85e-06	1.94e-04	5.41e-06	2.53e-04	7.26e-06
Deviance (df)	18585 (359)		18693 (359)		18667 (359)		18767 (359)	
AIC	22205		22313		22288		22387	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

表 3 考慮主要效果以及一階交互作用項作為解釋變數在二項式迴歸 (identity link) 下，使用 R 軟體原生套件逐步迴歸與 glmulti 套件 (Calcagno, 2020) 方法，尋找出 12 個候選模型之模型表現標準比較。

候選模型	AIC	BIC	Deviance	MSPE (LOOCV)
M1	18637.38	18774.16	14964.77	0.000927
M2	18643.16	18768.22	14976.55	0.000848
M3	18637.38	18774.16	14964.77	0.000927
M4	18643.16	18768.22	14976.55	0.000848
M5	18637.38	18774.16	14964.77	0.000927
M6	18643.16	18768.22	14976.55	0.000848
M7	18637.38	18774.16	14964.77	0.000927
M8	18638.56	18771.44	14967.95	0.000913
M9	18639.28	18779.97	14964.66	0.000933
M10	18643.16	18768.22	14976.55	0.000848
M11	18638.56	18771.44	14967.95	0.000913
M12	18642.72	18771.69	14974.11	0.000861

表 4 M2 模型 ( 見表 3 ) 下使用二項式迴歸與準二項式迴歸 (quasi-binomial regression) 之參數估計結果。

Parameter	Binomial (identity link)			Quasi-binomial (identity link)		
	estimate	s.e.	signif.	estimate	s.e.	sign.
(Intercept)	3.44e+00	2.24e-01	***	3.44e+00	1.49e+00	*
college	-4.41e+00	2.79e-01	***	-4.41e+00	1.86e+00	*
municipality	5.59e-01	1.90e-02	***	5.59e-01	1.27e-01	***
island	7.96e-01	8.54e-02	***	7.96e-01	5.70e-01	
vote	-1.07e+00	1.95e-01	***	-1.07e+00	1.30e+00	
married	-7.61e+00	4.20e-01	***	-7.61e+00	2.80e+00	**
salary	1.64e-03	1.13e-04	***	1.64e-03	7.51e-04	*
age	-1.68e-02	3.66e-03	***	-1.68e-02	2.44e-02	
gender	6.60e-02	2.14e-01		6.60e-02	1.43e+00	
unipality:married	-1.12e+00	3.31e-02	***	-1.12e+00	2.21e-01	***
island:married	2.17e-01	1.65e-01		2.17e-01	1.10e+00	
college:municipality	-1.17e-01	1.30e-02	***	-1.17e-01	8.68e-02	
college: island	-4.44e-01	1.34e-01	***	-4.44e-01	8.94e-01	
municipality:salary	1.02e-04	5.13e-06	***	1.02e-04	3.42e-05	**
island:salary	-7.85e-05	1.25e-04		-7.85e-05	8.36e-04	
college:married	5.20e+00	3.72e-01	***	5.20e+00	2.49e+00	*
college:salary	-8.56e-05	1.88e-05	***	-8.56e-05	1.25e-04	
vote_rate:age	-2.83e-02	2.32e-03	***	-2.83e-02	1.55e-02	.
municipality:age	1.84e-03	2.09e-04	***	1.84e-03	1.40e-03	
island:age	-1.95e-02	2.43e-03	***	-1.95e-02	1.62e-02	
married:age	8.02e-02	5.36e-03	***	8.02e-02	3.57e-02	*
salary:age	-6.83e-06	6.27e-07	***	-6.83e-06	4.18e-06	
municipality:gender	6.90e-02	1.18e-02	***	6.90e-02	7.89e-02	
island:gender	1.00e-01	3.33e-02	**	1.00e-01	2.22e-01	
married:salary	-1.26e-03	1.52e-04	***	-1.26e-03	1.01e-03	
college:gender	6.36e-01	5.48e-02	***	6.36e-01	3.65e-01	.
vote_rate:married	4.71e+00	3.37e-01	***	4.71e+00	2.25e+00	*
vote_rate:gender	-6.30e-01	1.25e-01	***	-6.30e-01	8.31e-01	
vote_rate:salary	-7.75e-04	8.00e-05	***	-7.75e-04	5.34e-04	
college:vote_rate	1.05e+00	1.44e-01	***	1.05e+00	9.58e-01	
age:gender	-1.26e-02	1.92e-03	***	-1.26e-02	1.28e-02	
married:gender	8.77e-01	2.89e-01	**	8.77e-01	1.93e+00	
Deviance (df)	14977 (336)			14977 (336)		
AIC	18643			NA		

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

表 5 考慮主要效果以及一階交互作用項作為解釋變數在一般多元迴歸下，使用 R 軟體原生套件逐步迴歸與 leaps 套件 (Lumley, 2020) 方法，尋找出 12 個候選模型之模型表現標準比較。

候選模型	AIC (Normal)	BIC (Normal)	AIC (Binomial)	BIC (Binomial)	Deviance	MSPE (LOOCV)
LM1	-1611.72	-1533.55	19427.75	19431.66	15814.23	0.000817
LM2	-1588.44	-1545.45	21308.6	21312.51	17691.33	0.000921
LM3	-1617.41	-1519.71	18775.92	18779.83	15154.08	0.000782
LM4	-1608.16	-1553.45	20106.78	20110.69	16496.92	0.000695
LM5	-1617.41	-1519.71	18775.92	18779.83	15154.08	0.000782
LM6	-1608.16	-1553.45	20106.78	20110.69	16496.92	0.000695
LM7	-1608.16	-1553.45	20106.78	20110.69	16496.92	0.000695
LM8	-1609.53	-1550.91	19959.64	19963.54	16346.68	0.000700
LM9	-1612.33	-1549.8	19750.67	19754.58	16134.85	0.000730
LM10	-1617.75	-1539.59	19176.71	19180.61	15557.17	0.000728
LM11	-1618.52	-1536.45	19057.21	19061.11	15440.52	0.000794
LM12	-1618.29	-1536.22	19070.18	19074.09	15450.06	0.000763

表 6 本研究最終選擇的最佳模型 LM10 ( 見表 5 ) 之離群值 (outlier) 與影響點 (influence point) 。

地點	DFFITS	Cook's distance	DFBETA									
			b1	b2	b3	b4	b5	b6	b7	b8	...	b19
臺北市 大安區	-2.76	0.395	-0.411	0.008	-0.049	-0.077	-0.036	0.008	0.477	0.199		-0.509
新竹縣 竹北市	-2.649	0.366	0.177	0.002	-0.074	-0.112	-0.027	-0.006	-0.089	0.491		-0.925
新竹市 東區	1.894	0.188	0.051	0.003	-0.008	-0.05	0.002	-0.002	-0.085	0.149		-0.262
金門縣 金城鎮	1.797	0.169	-0.037	-0.001	0	-0.255	0.005	0.001	0.048	0.009		0.017
臺北市 內湖區	1.426	0.104	-0.044	0.004	-0.036	-0.032	0.025	0.001	-0.093	-0.032		0.143
臺中市 西屯區	-1.253	0.081	0.073	-0.008	0.01	0.063	-0.006	-0.001	0.009	-0.218		0.314
臺北市 中正區	-1.164	0.07	-0.11	-0.005	0.001	0.011	0.017	0.002	0.179	0.116		-0.082
新北市 新店區	-1.143	0.067	0.176	0.006	0.01	-0.005	0.019	-0.003	-0.338	-0.075		0.224
新北市 新莊區	1.002	0.052	0.363	0.005	0.033	-0.015	-0.01	-0.007	-0.542	-0.049		0.049
臺東縣 臺東市	-0.885	0.041	-0.07	-0.002	0.053	0.066	0.016	0	0.111	0.044		0.027
臺北市 中山區	0.821	0.035	-0.244	0	0.012	0.008	0.008	0.004	0.397	0.195		-0.238
新北市 三重區	0.788	0.033	0.068	-0.001	0.037	0.001	-0.003	-0.001	-0.115	-0.068		0.091
桃園市 中壢區	-0.78	0.032	0.028	0.009	0.016	-0.017	0.006	-0.001	-0.054	0.064		-0.076
花蓮縣 花蓮市	-0.765	0.03	0.087	0.003	-0.013	-0.034	-0.004	-0.002	-0.116	-0.04		-0.04
臺北市 士林區	0.755	0.03	-0.048	0.003	-0.006	-0.018	-0.026	0.001	0.103	0.046		-0.212
臺中市 南屯區	-0.762	0.03	0.055	-0.009	0.007	0.038	-0.008	0	-0.003	-0.143		0.183
...	...	...	...	...	...	...	...	...	...	...	...	...



表 7 本研究最終選擇的最佳模型 LM10 (一般多元迴歸・見表 5) 之參數估計結果。

	estimate	s.e.	t value	Pr(> t )
(Intercept)	2.539e+00	8.351e-01	3.040	0.003
vote	-1.888e-01	3.262e-02	-5.788	<0.001
municipality	4.585e-01	1.089e-01	4.210	<0.001
offshoreIsland	1.059e+00	2.834e-01	3.735	<0.001
gender	-2.796e-01	7.309e-02	-3.825	<0.001
age	-3.193e-02	1.606e-02	-1.989	0.048
married	-3.446e+00	1.249e+00	-2.759	0.006
college	-1.955e+00	5.303e-01	-3.687	<0.001
salary	3.332e-04	1.193e-04	2.793	0.006
municipality:age	2.408e-03	1.251e-03	1.925	0.055
municipality:married	-9.022e-01	1.672e-01	-5.396	<0.001
municipality:college	-1.233e-01	6.467e-02	-1.907	0.057
municipality:salary	1.001e-04	3.159e-05	3.169	0.002
island:age	-1.969e-02	5.634e-03	-3.494	0.001
island:college	-6.488e-01	1.724e-01	-3.764	<0.001
gender:college	7.611e-01	2.101e-01	3.622	<0.001
age:married	6.007e-02	2.420e-02	2.482	0.014
age:salary	-7.039e-06	2.734e-06	-2.575	0.010
married:college	2.100e+00	8.106e-01	2.591	0.010
Dispersion parameter for gaussian family			44.576	
Deviance (df)		15557 (349)		
AIC		-1617		

## 附件

### 壹一、前言

在 2018 年 11 月 24 日，中華民國全台灣各地舉行了全國性的公民選舉投票。在公投編號的第 10 到 12 案中與第 14 到 15 案分別由「下一代幸福聯盟」與「同運組織」提出，在內容上對於同志婚姻平權、立法保障和同志教育的議題上卻有完全分歧的觀點。在社會上偏向性別平權的一方，倡導要投出「兩好三壞」（14、15 提案同意，10、11、12 案投不同意）。相反的，若是偏向傳統觀念或是愛家聯盟觀點的人民，則是說要投成「三好兩壞」（10、11、12 提案同意，14、15 案投不同意）。雖然說在者兩組提案並非完全相反或互斥，只是在社會、網路社群上的意見常常是完全支持一方而排斥另一組。

對於這樣的現象，本研究以公投的第 14 案作為主要探討對象，該提案的內文為：「**您是否同意，以民法婚姻章保障同性別兩人建立婚姻關係？**」，說明的是要不要採用民法來規範同志婚姻制度。因此對此案表示同意表示著可以用民法來保障。相反的，若是表示不同意，也僅僅說明的指示不採用民法來保障同志婚姻，但不排除用其他的法律形式來保障。然而，大眾對此案的傾向上同意者通常會對同性戀議題表現出正向的態度，而對此案表示不同意者通常也會對性戀議題表現出負面的態度。因此本研究有預設可以透過第 14 案之同意率結果來與同性戀正負向態度連結，但需謹記著，本公投的結果和同性戀態度實際上是兩回事，之中可能有某種程度的相關而已。

### 壹二、文獻回顧

關於同性戀與不同社會人口因子的關係在過往便有相當多研究探討。Herek (1984) 探討恐同症的文化因素時即列舉了許多對同性戀持負面態度的人較可能擁有的性質，包含年齡較大、較沒受到完好教育、住在鄉下或小城鎮（特別是那些持負面態度為其常態的區域），且男性通常比女性更容易持負面態度；Seltzer (1992) 也得到與上述相同的結果，除此之外，他的研究還顯示結過婚的人也較易持有負面態度。然而並非所有研究都得到相同的結果，Bowman (1979) 及 Glenn & Weaver (1979) 都沒有找到顯著的性別差異；West & Cowell (2013) 的研究則顯示年齡較大並不能預測較強的負面態度，甚至年齡較大者反而負面態度較小，此外他們也發現高收入者持較正面的態度，但也有其他研究顯示並沒有關係（see also Gelbal & Duyan, 2006；Becker & Scheufele, 2009）。另一方面，Seltzer (1992) 發現教育程度與年齡有交互作用的影響：

受到大學教育以上的中年人比起老年、青年人都持有較正面的態度，王晴峰（2013）也表示處於青少年與大學年齡組的男性比中年人更可能持強烈的否定立場。

還有其他研究將焦點放在區域、宗教、種族，如 Jäckle & Wenzelburger (2015) 其中包含：宗教、年紀、家中小孩人數、有無結婚、收入、教育狀態.....等，都有顯著影響。其中特別是宗教信仰最能影響同性戀態度的取向。Adamczyk & Liao (2019) 也於近期進行過後設分析，將主要影響各國同性戀態度的原因歸類至經濟、民主、宗教三個主要因子。或許這些變項對同性戀態度亦有相當程度的影響，但我們並無此方面的資料且台灣並無宗教信仰的普查，故能期望後續研究得以加入。

## 貳一、資料來源與前處理

為了研究全台灣 2018 年公投第 14 案之同意率之可能影響因子，試以建立廣義線性迴歸模型來找出之間的關聯。在本研究中，將以全台 368 個「鄉、鎮、市、區」（中華民國的直轄市、縣、市所管轄的下級行政區）作為研究資料單位，其中反應變項（應變項）為該地區的同意率，預測變項（獨變項）分別有：投票率、是否位於六都、性別比、年齡中位數、教育程度、結婚率以及薪水中位數，共七項。其中，各鄉鎮市區之同意率與投票率從政府資料開放平臺「第 14 案全國性公民投票計票結果」而來；內政資料開放平臺「村里戶數、單一年齡人口」和「15 歲以上現住人口按性別、年齡、婚姻狀況及教育程度分」分別提供了 2018 年 10 月的人口年齡結構分佈和婚姻狀況及教育程度；最後，財政部財政資訊中心提供了「105 年度綜合所得稅申報核定統計專冊」，我們可以從中獲取各鄉鎮市區的薪水中位數，然而可惜的是目前財政部最新的資訊只提供到了民國 105 年（西元 2016 年），因此只能先以此筆資料作為薪水中位數的處理。在下方將一一介紹每個變項的定義。

公投第 14 案之同意率為該地區在該案的同意票數除以總有效票數；投票率為該地區實際對該案進行投票人數在符合資格選民中所佔的比例；地理位置為類別變項，共有三個類別，因此可以再拆分成兩個虛擬變項：一、是否為直轄市 (Imunicipality)：若該地區位於六個直轄市地區（中華民國之第一行政區，包括：台北市、新北市、桃園市、台中市、台南市、高雄市）之內，則標記為 1，其他地區則標記為 0；二、是否為離島(island)：若該地區位於澎湖縣、金門縣或連江縣內的化，則標記為 1，其他地區則標記為 0。性別比為該地區的符合資格選民（18 歲以上）之男性人口總數除以女性人口總數；年齡中位數如同字面上所意，然而我們選定的範圍是該地區符合資格選民（18 歲以上）的年齡中位數；結婚率則是定義為曾經有結果婚的比率，因此在計算上則是將該地區目前是有配偶、喪偶以及離婚的人數總和除以該地的 15 歲以上之人口總

數，由於原始資料的將每五歲作為一個級距統計，因此無法切出符合資格選民的年齡（18歲以上）人數；教育程度則是定義在具有大專生以上教育程度的比例，因此在計算上則是將該地區的大學、大專、碩士和博士以上的人數總和除以該地的15歲以上之人口總數；最後的薪水中位數也如同字面上所意，但此變項則是直接採財政部的以統計好的資料，因此我們並未對數據進行任何計算處理。

### 肆三、研究限制與改進方向

首先，本研究是針對公投第14案的同意率，透過廣義迴歸模型的建立找出相關重要的影響變項。然而在本案中的議題是針對「以民法的方式保障同性別兩人的婚姻建立關係」，並非如過往研究探討的是「同性戀態度」以及相關的因子。即便這兩個議題是有高度的相關，但本質上還是兩個完全不同的議題，因此在結論的連結上必須謹慎小心，兩者是不能互相推論的。舉例來說，對於同性戀態度正向者不一定會在此公投投下同意票，他的考量可能會認為修該民法的成本過高，而對此次公投中的第14案，「以民法婚姻規定以外之其他形式來保障同性別兩人...」表示贊同。

第二，迴歸模型是以鄉鎮市區作為每筆資料的單位，來看預測變項與反應變項之間的關聯。雖然說我們一開始預設可以以一個地區來預設該區的投票者整體的傾向，但這真的具有代表性嗎？以及以鄉鎮市區所帶表的意義性又會是什麼？畢竟，我們最終獲得的是整體群族資料，並非是投票人確切的行為和態度。因此在此研究結果，或許可以說在某變數上，整體的選民是偏向於某個面向（同意或不同意），但絕對不能推論到個體實際的想法和投票行為。

第三，在資料來源上有諸多的限制與困難。由於在民主社會中是採以不記名投票，我們無法追溯每個投票人的投票行為的，也因此造就上述第二點我們採以地區整體的表現作為模型資料的基本單位。另一方面，在預測變項的選擇上很有可能忽略掉重要的選擇變項，如宗教信仰、家庭人數、政黨傾向...等，特別是宗教信仰上在先前文獻中特別提及它大大的影響對同性戀的態度。然而在台灣目前沒有對宗教信仰進行普查，因此在目前無法得知他們對於公投同意率的影響，實為可惜。

最後，在本次研究最終選出的模型實際上並沒有完全符合一般多元迴歸模型的假設，一些潛在的離群值與影響點也沒有完善的探討與處理等，可能有待繼續去改進。另一方面，使用一般多元迴歸模型來估計二項是分配中的參數，在一般二項式迴歸上較不常見的做法，有可能是此筆資料的特性才比其他種透過鏈結函數轉換表現要來的好，但整體的方法學上的使用還有帶進一步地去確認。期望在後續的研究中，能納入潛在的重要變項（如宗教信仰、家庭人數...等）做探討，並驗證是否跟先前文獻的結

果是否有出入。有可能在這些變項加入後，會使得本研究最終選出的迴歸模型有所改變。