

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

林子堯

心理學研究所

2020/12/28

QR-code

Link: <https://bit.ly/3aHVNgv>



Outline

前言與研究目的

文獻回顧

資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

Outline

前言與研究目的

文獻回顧

資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

前言

2018年11月24日中華民國全國性公民投票結果

2018年公投開票結果

資料來源 | 中選會、製圖 | 郭怡君
最終數據以中選會公布資料為主
得票率是以有效票為計算基礎



編號	提出者	公投項目	贊成票	贊成率	反對票	反對率	結果
7	國民黨	反空污	7,955,753	79.04%	2,109,157	20.96%	通過
8		反深澳電廠	7,599,267	76.41%	2,346,316	23.59%	通過
9		反核食	7,791,856	77.74%	2,231,425	22.26%	通過
10	下一代幸福聯盟	婚姻定義	7,658,008	72.48%	2,907,429	27.52%	通過
11		適齡性平	7,083,379	67.44%	3,419,624	32.56%	通過
12		另立專法	6,401,748	61.12%	4,072,471	38.88%	通過
13	紀政	2020東京奧運正名	4,763,086	45.20%	5,774,556	54.80%	X
14	同運組織	同性婚姻	3,382,286	32.74%	6,949,697	67.26%	X
15		同志教育	3,507,665	34.01%	6,805,171	65.99%	X
16	黃士修	以核養綠	5,895,560	59.49%	4,014,215	40.51%	通過

性平公投 vs. 愛家公投

1124 九合一大選綁公投 兩好三壞這樣投

LGBTQ

投完大選
領公投票

同意 (O)	不同意 (X)	同意 (O)	不同意 (X)	同意 (O)	不同意 (X)	同意 (O)	不同意 (X)	同意 (O)	不同意 (X)
10		11		12		14		15	
你是否同意民法婚姻規定限定在一男一女的結合？		你是否同意在國民教育階段內（國中及國小），教育部及各級學校不應對學生實施性別平等教育法施行細則所定之同志教育？		你是否同意以民法婚姻規定以外之其他形式來保障同性別二人經長期共同生活的權益？		你是否同意，以民法婚姻保障同性別二人建立婚姻關係？		你是否同意，以「性別平等教育法」規定在國民教育各階段內實施性別平等教育，且內容應涵蓋情感教育、性教育、同志教育等課程？	

11月24日 領公投票投同意票

3好 2壞
10 11 12 14 15

守護最愛

愛家公投

同意	不同意

第10 11 12號蓋同意

- 10 婚姻應限定在一男一女
- 11 在國中小，不應對學生實施同志教育
- 12 婚姻以外其他形式保障同性生活權益



同運公投

同意	不同意

第14 15號蓋不同意

- 14 修改民法為同性婚姻
- 15 強迫國中小學習同志教育



探討對象

Outline

前言與研究目的

文獻回顧

資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

同性戀態度

負面 / 恐同可能因素

- 年齡較大、男性 > 女性、較少接觸到高等教育、鄉下或小城鎮
(Herk, 1984)
- 結過婚的 (Seltzer, 1992)
- 家中成員多 (Adamczyk & Pitt, 2009)
- 收入較低者 (West & Cowell, 2013)
- 大學教育以上的中年人比起老年人與年輕人較有正面態度
(Seltzer, 1992)
- 青少年與大學年齡的男性比中年人更可能持強烈的否定立場
(王晴峰, 2013)

並非所有研究都相同

- 年齡大者反而負面態度較小 (West & Cowell, 2013)
- 性別差異不顯著 (Bowman, 1979; Glenn & Weaver, 1979)
- 與收入高低無關 (Gelbal & Duyan 2006; Becker & Scheufele, 2009)

同性戀態度

其他可能因子

- 對 79 國的同性戀態度進行多元迴歸分析，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 (Jäckle & Wenzelburger, 2015)
- 透過後設分析，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 (Adamczyk & Liao, 2019)

總結來說，可能的影響因子：

- 年齡、性別、區域(都市/鄉下)、婚姻狀況、教育程度、經濟收入、種族、民主、宗教...等
- 以及變項間的交互作用
- 雖然有許多變項會影響同性戀態度，但受限於資料的取得 (如：台灣並無宗教信仰普查)，故期望後續研究得以加入。

Outline

前言與研究目的

文獻回顧

資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

資料來源

- 公投：
 - 政府資料開放平臺_第14案全國性公民投票計票結果
- 人口年齡組成：
 - 內政資料開放平臺_村里戶數、單一年齡人口（新增區域代碼）
- 結婚與教育：
 - 內政資料開放平臺_15歲以上現住人口按性別、年齡、婚姻狀況及教育程度分
- 薪水中位數：
 - 財政部財政資訊中心_105年度綜合所得稅申報核定統計專冊

資料數量：全台鄉、鎮、市、區共 368 個

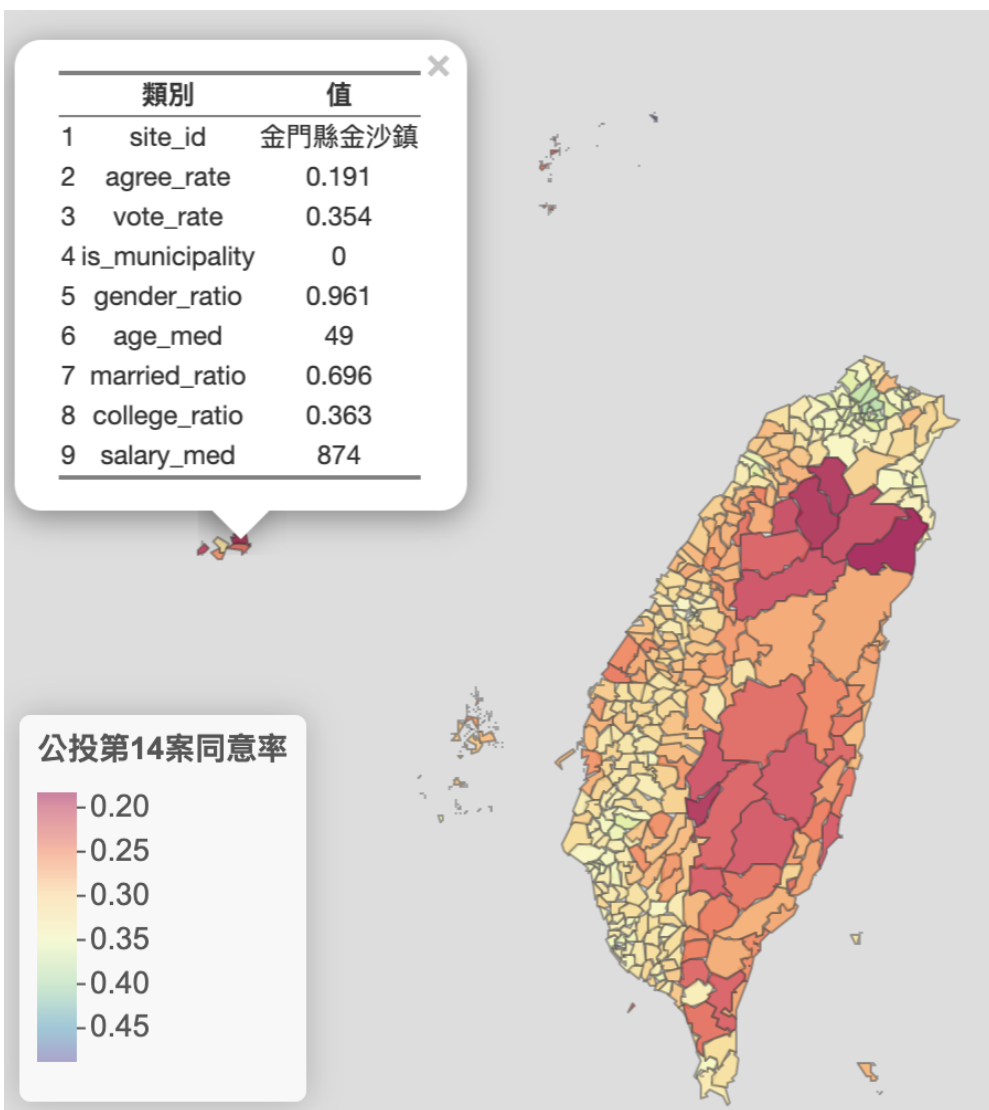
反應變項：公投第 14 案同意率

預測變項：投票率、地區、性別比、年齡、婚姻、教育程度、薪水收入

變數介紹

- \bar{y} : Agree_rate 第14案公投同意率
- X1: Vote_rate 投票率
- X2: Location 地理位置 (類別變項：直轄市、一般縣市、離島)
 - X21:Municipality = $\begin{cases} 1, \text{直轄市} \\ 0, \text{非直轄市} \end{cases}$
 - X22:Island = $\begin{cases} 1, \text{離島} \\ 0, \text{非離島} \end{cases}$
- X3: Gender_ratio 性別比
- X4: Age_med 年齡中位數
- X5: Married_ratio 結過婚比例
- X6: College_ratio 教育程度 (大專生畢業比例)
- X7: Salary_med 薪水中位數 (千元)

資料地圖 (連結)



公投第 14 案同意率

$$\text{同意率} = \frac{\text{同意票數}}{\text{有效票數}}$$

Stat.	Value
Min.	0.184
Q1	0.277
Med.	0.305
Mean	0.301
Q3	0.326
Max.	0.487
Std.	0.040

投票率

$$\text{投票率} = \frac{\text{投票數}}{\text{可投票人數}}$$

Stat.	Value
Min.	0.291
Q1	0.505
Med.	0.542
Mean	0.534
Q3	0.577
Max.	0.689
Std.	0.060

地理位置

一般縣市：194 個

直轄市：158 個

離島：16 個

性別比

$$\text{性別比} = \frac{\text{男性人數}}{\text{女性人數}}$$

Stat.	Value
Min.	0.833
Q1	0.986
Med.	1.047
Mean	1.053
Q3	1.118
Max.	1.582
Std.	0.102

年齡中位數

Stat.	Value
Min.	40.000
Q1	45.000
Med.	47.000
Mean	47.394
Q3	50.000
Max.	55.000
Std.	2.978

結婚率

結婚率

$$= \frac{\text{有偶} + \text{喪偶} + \text{離婚}}{\text{總人數}}$$

Stat.	Value
Min.	0.590
Q1	0.652
Med.	0.666
Mean	0.668
Q3	0.683
Max.	0.753
Std.	0.025

大學教育程度

大專畢業率

$$= \frac{\text{大學} + \text{大專} + \text{碩士} + \text{博士}}{\text{總人數}}$$

Stat.	Value
Min.	0.117
Q1	0.214
Med.	0.275
Mean	0.290
Q3	0.356
Max.	0.653
Std.	0.101

薪水中位數

Stat.	Value
Min	587.0
Q1	691.0
Median	735.0
Mean	784.9
Q3	823.2
Max	1722.0
Std	162.8

Scatter-correlation matrix

Outline

前言與研究目的

文獻回顧

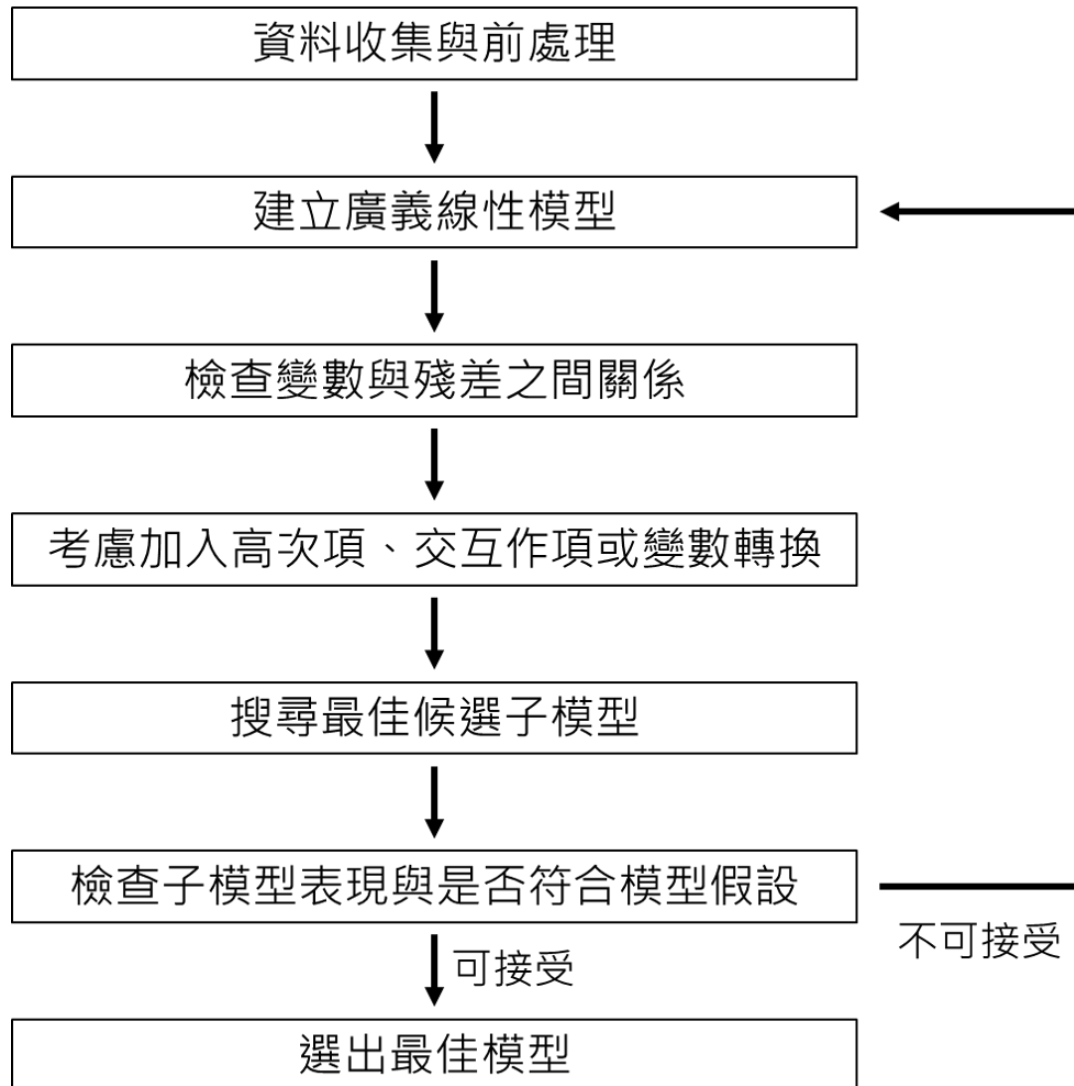
資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

分析流程



Binomial regression model

Let

- \bar{y}_i be the observed agree_rate in the i th 鄉鎮市區
- N_i be the number of the valid vote in the i th 鄉鎮市區
- π_i be the probability of agree.

Our preliminary model is

$$N_i \bar{y}_i \sim \text{Binomial}(N_i, \pi_i)$$
$$\pi_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$$

where $i = 1, \dots, 368$, $\mathbf{x}_i = (x_{i1}, \dots, x_{i7})$ and $g(\cdot)$ is a link function.

- identity link
- logit link
- probit link
- cloglog link

Binomial regression model

考慮所有變項的都加入下，identity link 表現比較好

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

True value \bar{y} vs. Fitted value $\hat{\pi}$

True value \bar{y} vs. $\hat{\eta} = X\hat{\beta}$

Standardized residual plot

Partial residual plot

Model selection

考慮原始兩兩變項間之交互作用項加入於原始模型中，但不考慮高次項

R 原生套件中的函式 `step()` 來執行逐步迴歸 (R Core Team, 2020)

- AIC、BIC 兩種準則
- 前向、反向、雙向選擇

找出六個候選模型 (有可能會找出相同模型)，但其缺點在於它並不會搜尋完所有的子模型

`glmulti` 套件的函式 `glmulti()` 來搜尋前幾佳的 GLM (Calcagno, 2020)

- AIC、BIC 兩種準則

再找出六個候選模型

Model selection

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

選擇基準：

- AIC、BIC、MSPE (LOOCV) 較小者
- 若兩變項之交互作用存在時其相對應的主要效果也應在模型之中

偏好 **M2** 作為暫時的最佳模型

The temporal best model: M2

Call:

```
glm(formula = .f_M2, family = binomial(link = "identity"), data = LM_data,  
     weights = valid_vote)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-23.806	-3.996	-0.042	3.325	23.244

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.441e+00	2.236e-01	15.393	< 2e-16	***
college	-4.410e+00	2.790e-01	-15.807	< 2e-16	***
locationmunicipality	5.588e-01	1.902e-02	29.382	< 2e-16	***
locationisland	7.961e-01	8.535e-02	9.328	< 2e-16	***
vote	-1.074e+00	1.947e-01	-5.517	3.45e-08	***
married	-7.610e+00	4.198e-01	-18.128	< 2e-16	***
salary	1.636e-03	1.125e-04	14.538	< 2e-16	***
age	-1.684e-02	3.657e-03	-4.606	4.10e-06	***
gender	6.604e-02	2.144e-01	0.308	0.75813	
locationmunicipality:married	-1.125e+00	3.309e-02	-33.990	< 2e-16	***
locationisland:married	2.173e-01	1.649e-01	1.317	0.18770	
college:locationmunicipality	-1.171e-01	1.300e-02	-9.002	< 2e-16	***
college:locationisland	-4.439e-01	1.339e-01	-3.314	0.00092	***
locationmunicipality:salary	1.021e-04	5.129e-06	19.896	< 2e-16	***
locationisland:salary	-7.850e-05	1.253e-04	-0.626	0.53111	
college:married	5.197e+00	3.724e-01	13.953	< 2e-16	***

Overdispersion model

Call:

```
glm(formula = .f_M2, family = quasibinomial(link = "identity"),  
     data = LM_data, weights = valid_vote)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-23.806	-3.996	-0.042	3.325	23.244

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.441e+00	1.492e+00	2.307	0.02167	*
college	-4.410e+00	1.862e+00	-2.369	0.01840	*
locationmunicipality	5.588e-01	1.269e-01	4.404	1.43e-05	***
locationisland	7.961e-01	5.695e-01	1.398	0.16306	
vote	-1.074e+00	1.299e+00	-0.827	0.40891	
married	-7.610e+00	2.801e+00	-2.717	0.00693	**
salary	1.636e-03	7.508e-04	2.179	0.03004	*
age	-1.684e-02	2.440e-02	-0.690	0.49048	
gender	6.604e-02	1.431e+00	0.046	0.96322	
locationmunicipality:married	-1.125e+00	2.208e-01	-5.094	5.86e-07	***
locationisland:married	2.173e-01	1.100e+00	0.197	0.84360	
college:locationmunicipality	-1.171e-01	8.676e-02	-1.349	0.17819	
college:locationisland	-4.439e-01	8.937e-01	-0.497	0.61974	
locationmunicipality:salary	1.021e-04	3.423e-05	2.982	0.00308	**
locationisland:salary	-7.850e-05	8.363e-04	-0.094	0.92527	
college:married	5.197e+00	2.485e+00	2.091	0.03726	*

兩難

1. 如果我不用 quasi-binomial regression 去擬合的話，M2 還是會有很大的過度離散現象存在
2. 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
(但似乎可以用 QAIC)

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 (π) 呢？

Outline

前言與研究目的

文獻回顧

資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

考慮直接做多元迴歸

$$\bar{y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$
$$\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$$

但這會有點怪怪的是

- 每個的資料點 (鄉鎮地區) 的權重都相同，沒有考量到投票人數
- 例如：同樣是同意率為 0.366
 - 臺北市文山區 (有效票數為 47668 票) vs.
 - 臺南市官田區 (有效票數為 3074 票)
- 變異數不符合同質性假設

Weighted least square estimation

$$N_i \bar{y}_i \sim \text{Binomial}(N_i, \pi_i)$$

- $E[\bar{y}_i | X] = \pi_i = \mathbf{x}_i^\top \boldsymbol{\beta}$
- $\text{Var}[\bar{y}_i | X] = \frac{\pi_i(1-\pi_i)}{N_i}$

$$\bar{y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

$$\varepsilon_i \stackrel{i.i.d.}{\sim} \text{Normal}\left(0, \frac{\pi_i(1-\pi_i)}{N_i}\right)$$

- 變異數不同質
- 加權最小平方估計，其中權重為 $\frac{N_i}{\pi_i(1-\pi_i)}$

結果發現其實與用 quasi-binomial regression (with identity link) 的參數估計是相同的！

Weighted least square estimation

```
.formula <- agree_rate ~ 1 + vote_rate + location + gender_ratio + age_med +  
gauss_ident_w <- glm(formula = .formula, weights = valid_vote,  
                      family = gaussian(link = "identity"), data = Data)  
  
.weight <- Data$valid_vote / (predict(gauss_ident_w)*(1-predict(gauss_ident_w))  
gauss_ident_w2 <- glm(formula = .formula, weights = .weight,  
                      family = gaussian(link = "identity"), data = Data)  
summary(gauss_ident_w2)
```

Call:

```
glm(formula = .formula, family = gaussian(link = "identity"),  
     data = Data, weights = .weight)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-23.2350	-4.4457	-0.3962	3.7651	25.9829

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.003e-01	5.680e-02	10.569	< 2e-16	***
vote_rate	-1.589e-01	3.328e-02	-4.776	2.61e-06	***
locationmunicipality	1.663e-02	2.700e-03	6.161	1.94e-09	***
locationisland	-6.259e-02	1.223e-02	-5.116	5.09e-07	***
gender_ratio	-2.182e-02	3.698e-02	-0.590	0.555	

Quasi-binomial regression (with identity link)

```
quasibinom_ident <- glm(formula = .formula, weight = valid_vote,  
                        family = quasibinomial(link = "identity"), data = Da  
summary(quasibinom_ident)
```

Call:

```
glm(formula = .formula, family = quasibinomial(link = "identity"),  
    data = Data, weights = valid_vote)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-23.6032	-4.4651	-0.3957	3.7544	25.7171

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.004e-01	5.680e-02	10.570	< 2e-16	***
vote_rate	-1.589e-01	3.327e-02	-4.777	2.60e-06	***
locationmunicipality	1.663e-02	2.700e-03	6.160	1.95e-09	***
locationisland	-6.259e-02	1.224e-02	-5.114	5.13e-07	***
gender_ratio	-2.184e-02	3.698e-02	-0.590	0.555	
age_med	2.036e-03	5.074e-04	4.013	7.29e-05	***
married_ratio	-5.471e-01	9.383e-02	-5.831	1.23e-08	***
college_ratio	6.482e-02	4.237e-02	1.530	0.127	
salary_med	7.214e-05	1.406e-05	5.133	4.69e-07	***

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

AIC (likelihood) 怪怪的？

使用 weighted least square estimation 在 general linear regression 下的 AIC 為 -1571.9

但先前用 binomial regression (with identity link) 下的 AIC 大多為 18XXX 以上

原因在於

- 前者的 likelihood 計算上是用 normal distribution

$$l(\beta, \sigma) = \sum \ln \left((2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(\frac{\bar{y}_i - \hat{\pi}_i(\beta)}{2\sigma^2}\right) \right)$$

- 而後者則是用 binomial distribution

$$l(\beta) = \sum \ln \left(\binom{N_i}{N_i \bar{y}_i} [\hat{\pi}_i(\beta)]^{N_i \bar{y}_i} [1 - \hat{\pi}_i(\beta)]^{N_i (1 - \bar{y}_i)} \right)$$

Partial residual plot

Box-Cox transformation

Model selection

考慮原始兩兩變項間之交互作用項加入於原始模型中，但不考慮高次項

R 原生套件中的函式 `step()` 來執行逐步迴歸 (R Core Team, 2020)

- AIC、BIC 兩種準則
- 前向、反向、雙向選擇

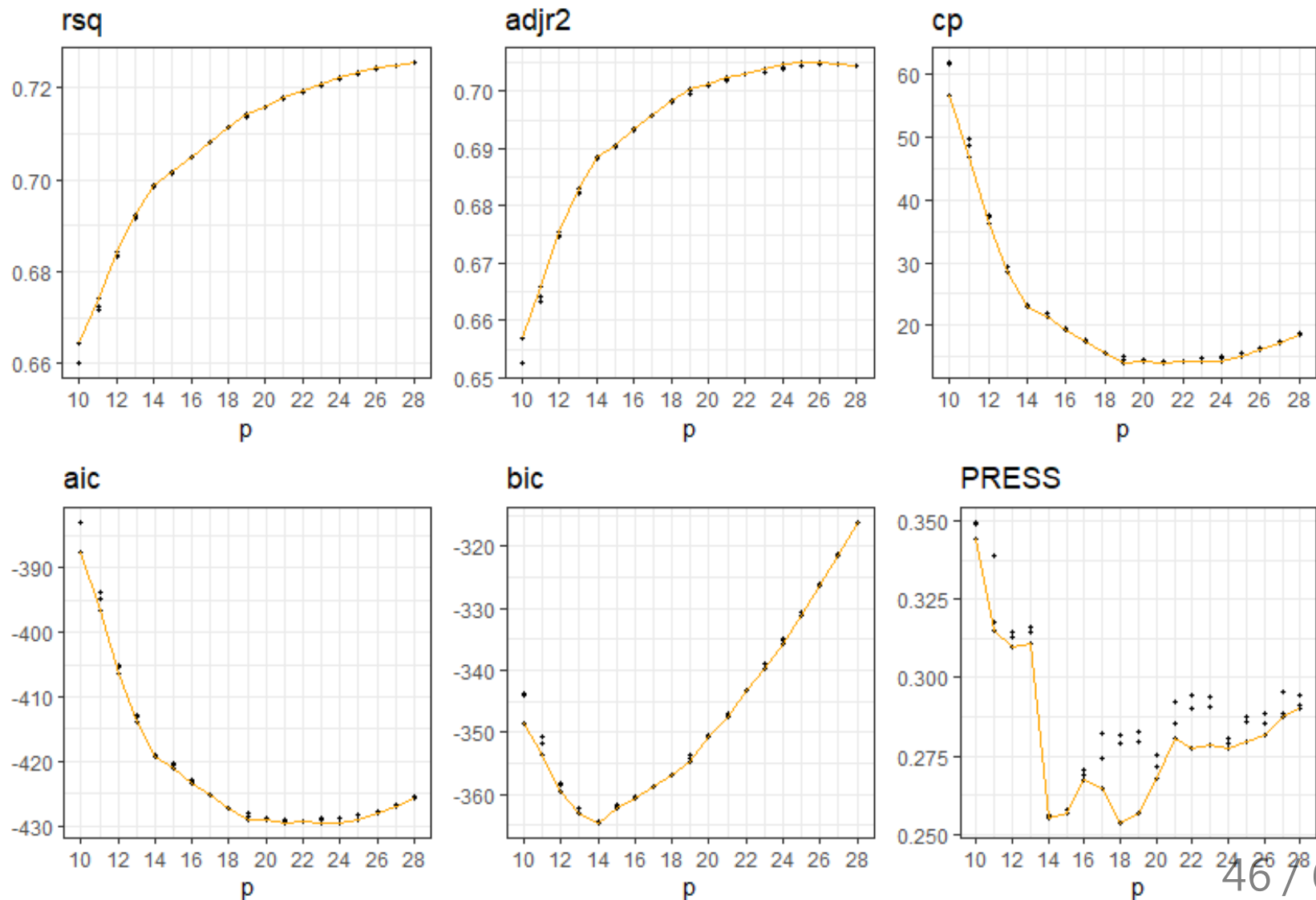
找出六個候選模型

leaps 套件的函式 `regsubsets()` 來搜尋模型候選 (Lumley, 2020)

- adjust R2、Mallows' s CP、AIC、BIC、PRESS

再挑選了6個候選模型

Model selection



Model selection

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

- AIC、BIC、MSPE (LOOCV) 較小者
- 若兩變項之交互作用存在時其相對應的主要效果也應在模型之中

偏好 **LM10** 作為暫時的最佳模型

Final model: LM10

Call:

```
glm(formula = .f_LM10, family = gaussian(link = "identity"),  
     data = LM_data, weights = weighted)
```

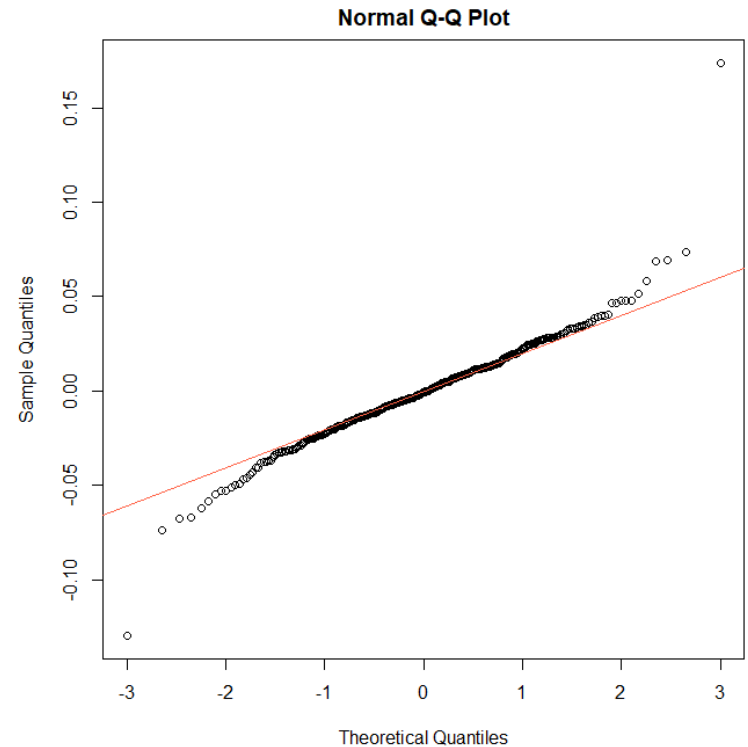
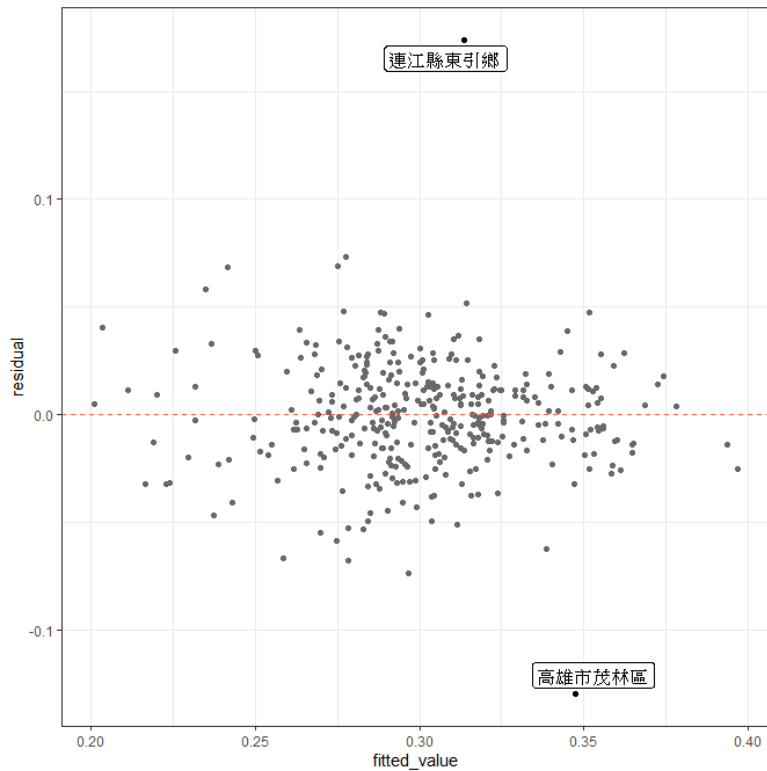
Deviance Residuals:

Min	1Q	Median	3Q	Max
-21.6078	-3.8120	-0.1633	3.6229	24.2597

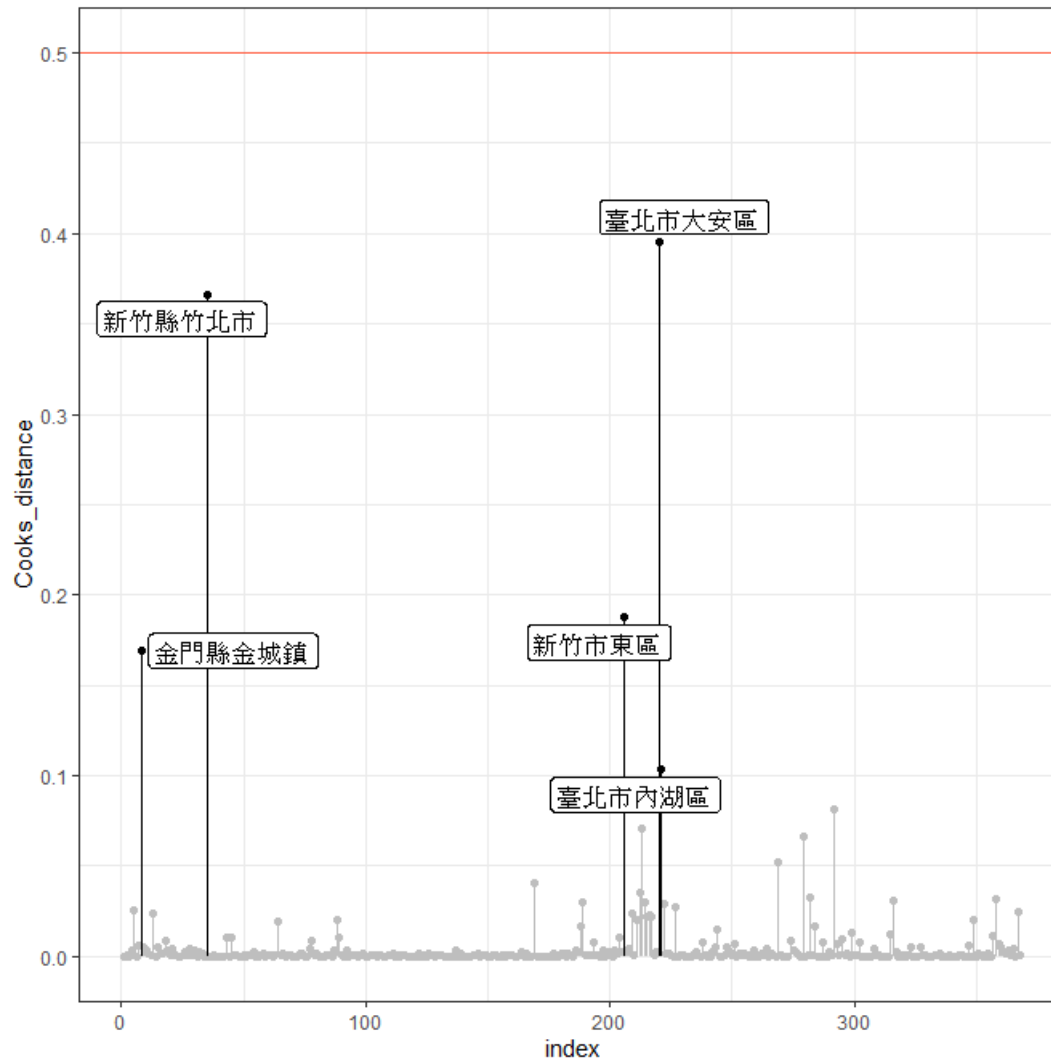
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.539e+00	8.351e-01	3.040	0.002542	**
vote	-1.888e-01	3.262e-02	-5.788	1.59e-08	***
municipality	4.585e-01	1.089e-01	4.210	3.25e-05	***
offshoreIsland	1.059e+00	2.834e-01	3.735	0.000219	***
gender	-2.796e-01	7.309e-02	-3.825	0.000155	***
age	-3.193e-02	1.606e-02	-1.989	0.047525	*
married	-3.446e+00	1.249e+00	-2.759	0.006097	**
college	-1.955e+00	5.302e-01	-3.687	0.000263	***
salary	3.332e-04	1.193e-04	2.793	0.005502	**
municipality:age	2.408e-03	1.251e-03	1.925	0.055061	.
municipality:married	-9.022e-01	1.672e-01	-5.396	1.26e-07	***
municipality:college	-1.233e-01	6.467e-02	-1.907	0.057351	.
municipality:salary	1.001e-04	3.159e-05	3.169	0.001663	**
offshoreIsland:age	-1.969e-02	5.634e-03	-3.494	0.000536	***
offshoreIsland:college	-6.488e-01	1.724e-01	-3.764	0.000196	***
gender:college	7.611e-01	2.101e-01	3.622	0.000335	***

Residual and Q-Q plot



Cook's distance



Influence points

Show entries

Search:

	DFFITS	Cooks_D	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
1	-0.014	0	0	0	0	0	0	0	-0.001	-0.001	0	0
2	0.093	0	0.003	0	-0.001	-0.013	0.001	0	-0.005	0.006	0	0
3	0.02	0	-0.001	0	0	-0.003	0	0	0.001	0.001	0	0
4	0.255	0.003	-0.017	0	0.004	0.047	0.001	0	0.027	0.006	0	0
5	-0.694	0.025	-0.057	-0.001	0.005	0.146	0.003	0.001	0.084	0.014	0	0
6	-0.014	0	0.002	0	0	-0.001	0	0	-0.003	0	0	0
7	-0.345	0.006	-0.015	0	-0.003	0.048	0.002	0	0.018	0.017	0	0
8	1.797	0.169	-0.037	-0.001	0	-0.255	0.005	0.001	0.048	0.009	0	0
9	-0.316	0.005	0.029	0.001	-0.005	0.01	-0.001	-0.001	-0.045	0.005	0	0
10	-0.238	0.003	-0.011	0.001	0	-0.044	0	0	0.014	-0.002	0	0

Showing 1 to 10 of 368 entries

Previous

1

2

3

4

5

...

37

Next

51 / 60

Outline

前言與研究目的

文獻回顧

資料來源與前處理

二項式迴歸模型

加權最小平方估計之多元迴歸模型

結果與討論

最終模型

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，找出影響同意率的可能因素為何，最終得到的解釋性模型為：

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

最終模型

依照不同位置的類別：

- 一般縣市 ($Municipality = Island = 0$)

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

- 直轄市 ($Municipality = 1, Island = 0$)

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

- 離島 ($Municipality = 0, Island = 1$)

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

模型解釋 (1/4)

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

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，以及薪水中位數越高，同意率會越高

模型解釋 (2/4)

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

但考量他變項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. 年齡中位數與薪水交互作用項越高，則同意率越低。

模型解釋 (3/4)

- 一般縣市 ($Municipality = Island = 0$)

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

- 直轄市 ($Municipality = 1, Island = 0$)

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

除此之外，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；

模型解釋 (4/4)

- 一般縣市 ($Municipality = Island = 0$)

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

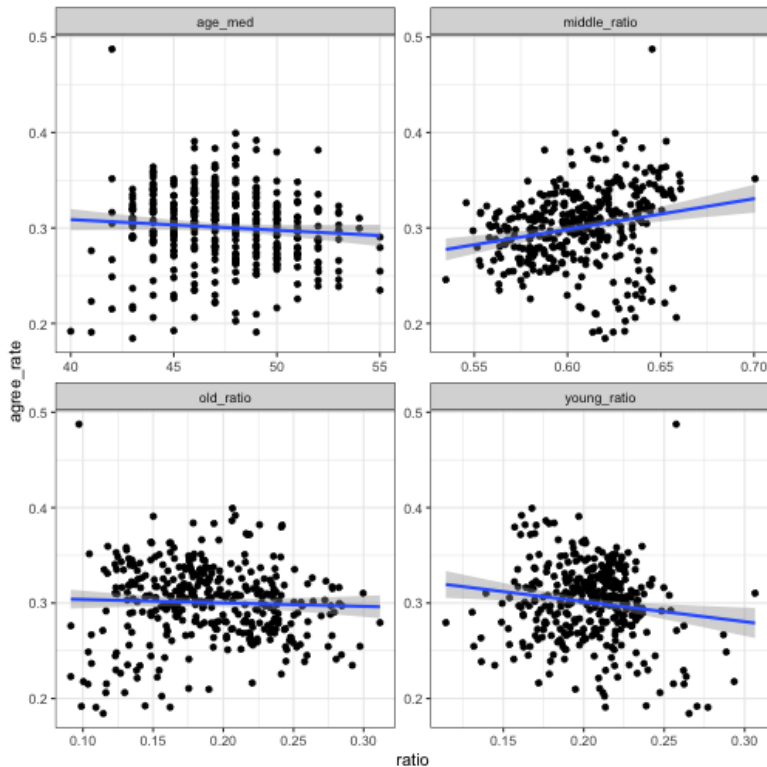
- 離島 ($Municipality = 0, Island = 1$)

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

除此之外，考量地區與其他變項交互作用項的影響：

1. 離島比起一般縣市來說，則是使年齡、大學教育程度越的負向效果變大。

年齡議題



依照年齡拆成三組與同意率做散佈圖

- 年輕人 (18~30) 負相關 ($r = -0.14$)
- 中年人 (31-65) 正相關 ($r = 0.21$)
- 老年人 (65-) 無相關 ($r = -0.04$)

- 過往文獻中年齡對於同性戀的態度也尚未有定論，且會與其他因素交雜在一起
- 我們都以為年紀較長者不容易接受婚姻平權，因此比較容易投不同意票。依據這次的模型結果，年齡中位數越高的地區同意率較低 (當然還要考慮交互作用項，因此關係沒那麼簡單)
- 但如果我們沒有考量其他變項的因素，或者是換不同種的年齡變項的計算方式，結果有可能會大不相同

研究限制與改進方向

- 公投14案：同性婚姻納入民法 \neq 同性戀態度
- 鄉鎮市區為每筆資料單位，但真的有代表性嗎
- 無法得之確切的投票人行為和態度
- 缺少一些可能的潛在變項 (宗教、家庭人數等)
- 離島 (界外值) 特別性質
- 假若我們有每個投票者的個人資料與行為，或許以後可以考慮使用 hierarchical linear model