

Group2.Final Project

用不同貝氏分析方法

—— 預測2018高雄市長 選舉之得票率

流預所碩一 鄭凱元/張家甄/張芷榆

統計碩一 陳育婷/黃俊凱



Outline

- Topic & Motive
- Data source
- Method & Model Selection

Method 1 and Method 2

- Discussion
- Conclusion

Topic & Motive

2018正值縣市首長大選，尤其今年**高雄市市長**最後由哪位候選人當選是許多人關注的議題。

許多家民調公司都有對此議題做民意調查，而民調的樣本數都在**1068人**左右。



投票率7成



277萬人 → 投票數約230萬人

Goal



相較於民調公司所蒐集的一千多筆樣本，我們希望透過考慮不同家民調公司以及不同時間點的民調結果，結合上課所學的貝氏統計預測兩大黨候選人之得票率。



想要運用所學的gibbs sampling以及prior→posterior的更新方式，結合搜集的民調結果，估算得票率，並判斷預測的準確程度。

2018高雄市候選人



韓國瑜



陳其邁

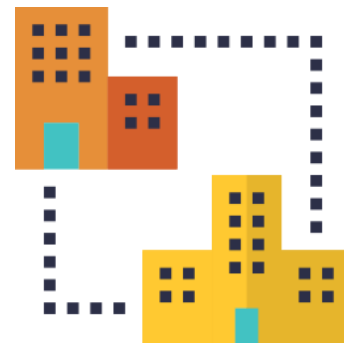
資料整理-民調(2018年)

公司	民調日期	樣本個數 (沒有就寫：No)	韓國瑜	陳其邁
聯合報系民意調查中心	11-08~11-10	1074	49	32
ET民調	10-30~11-08	1000	42.6	36.6
TVBS民意調查中心	11-04~11-06	1077	52	47
三立	11-03~11-04	1023	43.2	44.3
三立新聞	11-11~11-12	1063	40.5	43.6
大社會民調中心	7-25~7-26	1,317	17.5	60.8
台灣世代智庫	10月19日	1000	52.3	29.1
台灣世代智庫	9月19日	1,079	30.2	41.00
台灣競爭力論壇	10月5日	1075	30.8	33.5

Method 1-結合不同民調公司資訊

想法：

想要綜合**不同民調公司**的資訊做出預測。



方式：

採用上一次(2014年)高雄選舉之各家民調公司預測準確度當作權重大小標準(準確度越高，權重越高)；

藉此方式給予**不同公司**不同的權重，然後對每間公司的預測結果作加權平均。

資料整理- 2014年民調(n=5)

公司	2014年民調 (有就是1)	2014國民黨 (算權重)	和2014 真實之差	D取倒數	絕對值	權重wt
聯合報系民意調查中心	1	16	-14.89	-0.067159	0.067159	0.154859911
TVBS民意調查中心	1	19	-11.89	-0.084104	0.084104	0.193933059
三立	1	13.8	-17.09	-0.058514	0.058514	0.134924755
艾普羅民調	1	19.1	-11.79	-0.084818	0.084818	0.195577953
蘋果日報 (世新大學民調中心)	1	23.7	-7.19	-0.139082	0.139082	0.320704321
				sum	0.433677	

權重計算：將民調與真實得票率差取倒數後，再計算各間民調公司所有的權重。

$$\text{Ex. 聯合報 權重} = \frac{1}{(\text{此民調} - \text{真實得票率})} \div (0.434)_{\text{sum}}$$

計算權重 - 2014民調準度

公司	權重wt
聯合報系民意調查中心	0.154859911
TVBS民意調查中心	0.193933059
三立	0.134924755
艾普羅民調	0.195577953
蘋果日報（世新大學民調中心）	0.320704321

國民黨

公司	權重wt
聯合報系民意調查中心	0.304041367
TVBS民意調查中心	0.194378115
三立	0.142853101
艾普羅民調	0.229569041
蘋果日報（世新大學民調中心）	0.129158376

民進黨

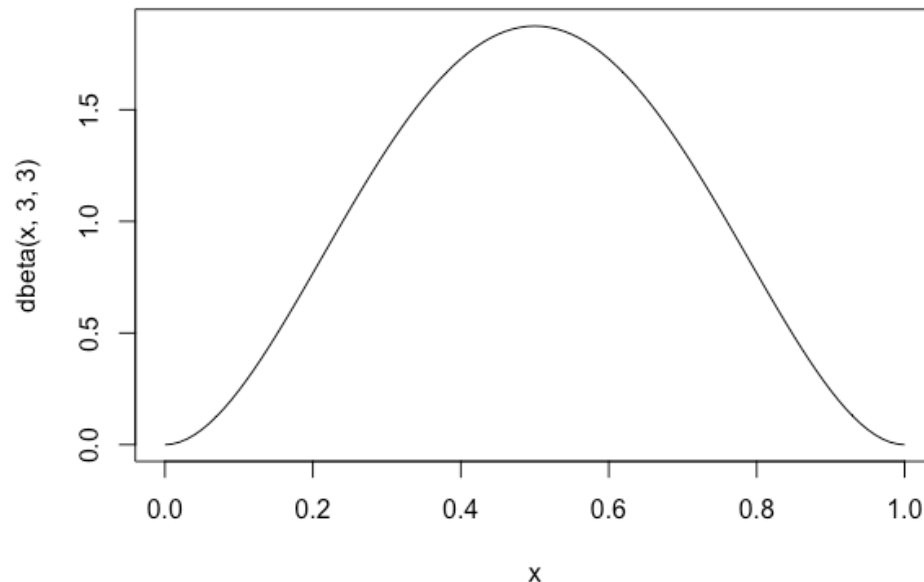
Statistical Method

Statistical Method: 用gibbs sampling的方式生得票率

Parameter Estimation:

$$Beta(\alpha, \beta) = Beta(3, 3)$$

$$E(\alpha, \beta) = \frac{\alpha}{\alpha + \beta} = 0.5$$



Model Selection



$$Y_i | p_i \sim \text{Binomial}(N_i, p_i)$$
$$p_i \sim \text{Beta}(3, 3)$$

- Y_i : 第*i*間民調中支持某地某候選人的的人數(已知)
 - N_i : 第*i*間民調總採訪人數(已知)
 - p_i : 某候選人的得票率
 - Goal: p_i
- i : 為第*i*間民調公司

OpenBUGS

將模型編碼成：

```
#model
model{
  for(i in 1:N){
    p[i] ~ dbeta(3, 3)  → prior
    Y[i] ~ dbin(p[i], n[i]) → likelihood
  }
}
```

OpenBUGS

```
#data
```

```
list(Y = c(526.26, 560.04, 430.515, 463.25, 450.996),  
N = 5, n=c(1074, 1077, 1063, 1090, 1274))
```

```
#initial values
```

```
list(p=0.5)
```

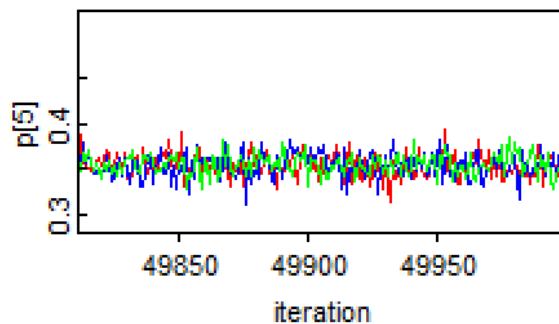
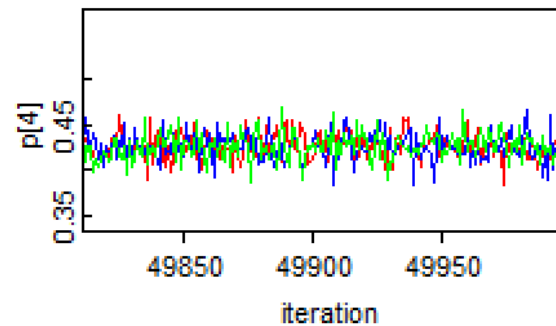
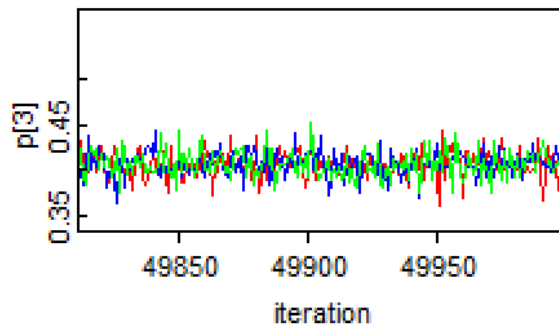
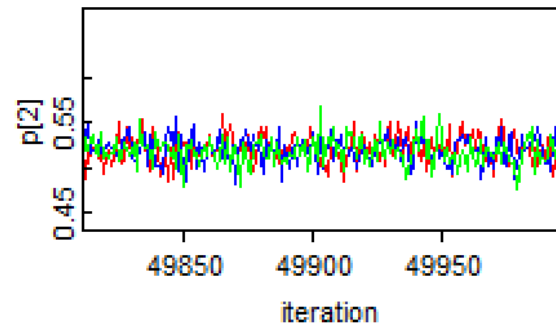
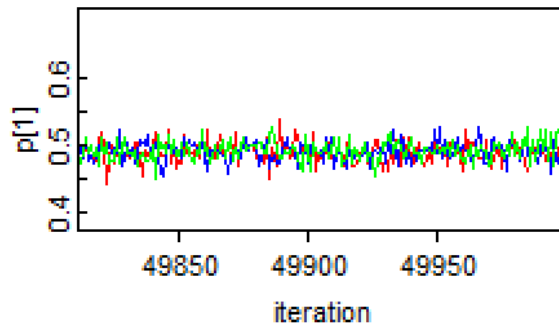
```
list(p=0.3)
```

```
list(p=0.7)
```

OpenBUGS

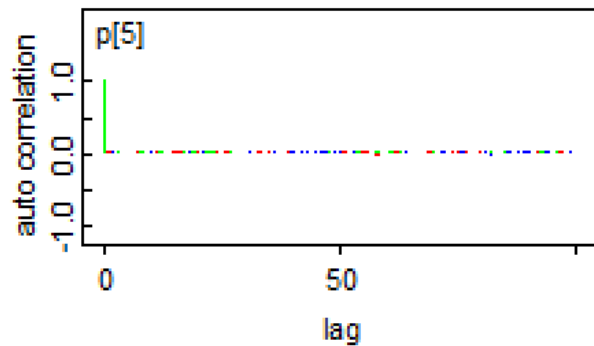
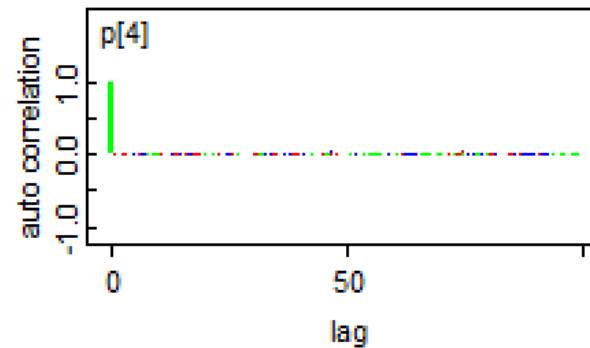
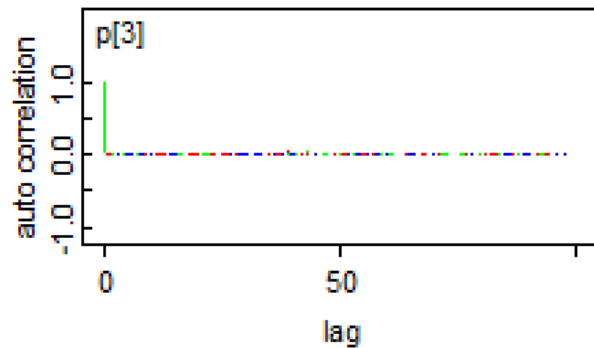
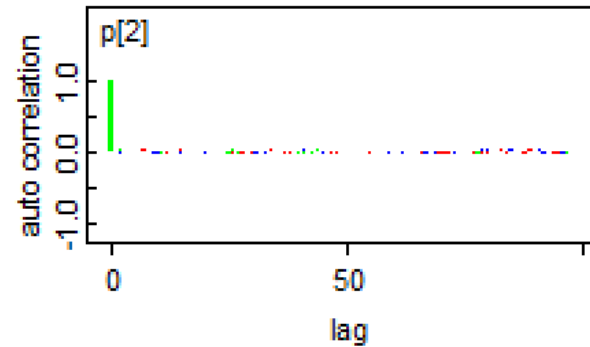
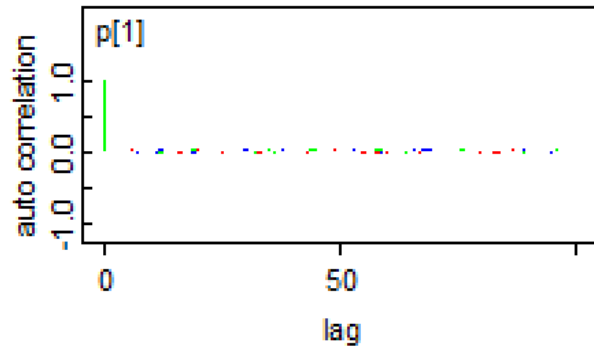
- number of chains: 3
- type number of samples : 50,000
- the first number of initial samples: 1,000
- thin: 10 #不想要和前幾個有關聯所以就取10

Trace



{ all chains wiggle
overlap well }

Auto correlation



Summary Stats

	mean	聯合報	val2.5pc	median	val97.5pc	start	sample
p[1]	0.49		0.4599	0.4903	0.5191	1001	14700
p[2]	0.52	TVBS	0.4898	0.5199	0.5505	1001	14700
p[3]	0.4056	三立	0.3761	0.4056	0.4352	1001	14700
p[4]	0.4254	艾普羅	0.3964	0.4254	0.454	1001	14700
p[5]	0.3548	蘋果日報	0.3289	0.3549	0.3811	1001	14700

Summary Stats-民進黨

	mean		val2.5pc	median	val97.5pc	start	sample
p[1]	0.3209	聯合報	0.2935	0.3208	0.3482	1001	14700
p[2]	0.47	TVBS	0.4401	0.4701	0.4993	1001	14700
p[3]	0.4365	三立	0.4072	0.4365	0.4663	1001	14700
p[4]	0.3898	艾普羅	0.3613	0.39	0.4185	1001	14700
p[5]	0.3268	蘋果日報	0.3014	0.3266	0.3532	1001	14700

整理結果（國民黨）

公司	樣本個數	韓國瑜人數	openbugs模擬p	p*wt	標準化
聯合報系民意調查中心	1074	526.26	0.49	0.075881356	52.27%
TVBS民意調查中心	1077	560.04	0.5201	0.100864584	
三立	1063	430.515	0.4056	0.054725481	
艾普羅民調	1090	463.25	0.4256	0.083237977	
蘋果日報（世新大學民調中心）	1274	450.996	0.3547	0.113753823	
				0.428463221	

將模擬生成之後驗分佈p的mean乘以權重wt，再進行標準化。（以2014年藍綠兩黨總得票率為基底）

得票率為52.27%

轉換百分比總和 (校正未表態)



✓ 民調普遍狀況：

有一定比例的民眾未表態，通常高於實際選舉總得票率。

✓ 民調未表態平均比率：21.3125%

✓ 修正方法：

利用上次高雄市長選舉(2014年)國民黨、民進黨總得票率為基底，等比例放大。

ex. 韓國瑜
得票率



98.98%
總得票率



韓國瑜得票率估計

韓國瑜得票率估計 + 陳其邁得票率估計

整理結果（民進黨）

公司	樣本個數	陳其邁	openbugs模擬p	p*wt	標準化
聯合報系民意調查中心	1074	343.68	0.3209	0.097566875	46.71%
TVBS民意調查中心	1077	506.19	0.4699	0.091338276	
三立	1063	463.468	0.4364	0.062341093	
艾普羅民調	1090	424.01	0.3898	0.089486012	
蘋果日報（世新大學民調中心	1274	415.324	0.3267	0.042196041	
				0.382928298	

將模擬生成後驗分佈p的mean乘以權重wt為各民調公司的p，再進行標準化。（以2014年藍綠兩黨總得票率為基底）

得票率為46.71%

Summary



	韓國瑜	陳其邁
用Gibbs sampling 估計	52.27%	46.71%
2018 真實得票率	53.87%	44.80%

Method 2-結合不同時間點的民調資訊

想法：

當新的資料進來時，想要以先前時間點的民調資料為先驗知識，不斷更新後驗資訊。

方式：Bayesian Updating

按照時間先後將民調資料排序。



Statistical Model

Prior : $p \sim \text{Beta}(\alpha + y_i, \beta + N_i - y_i)$

Likelihood : $Y_{i+1} | p \sim \text{Binomial}(N_{i+1}, p)$

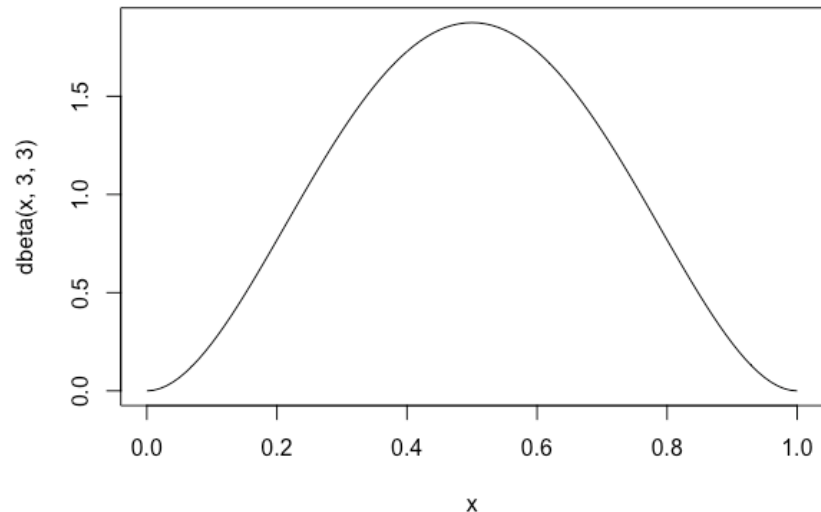
*Posterior: $p | Y_{i+1} \sim \text{Beta}(\alpha + y_i + y_{i+1},$
 $\beta + N_i - y_i + N_{i+1} - y_{i+1})$*

代
入

i : 第 i 筆資料(第 i 次迭代)

Model Notes

- ***Program*** : write loops in Python
(can actually derive exact posterior distribution)
- ***Initial Prior*** : $p \sim \text{Beta}(3,3)$



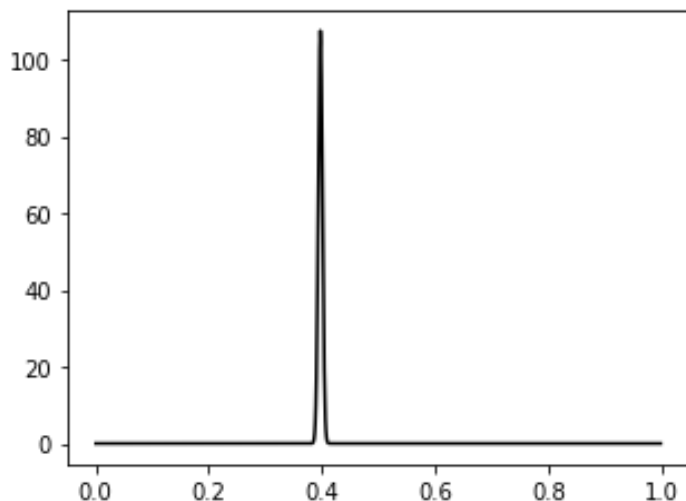
Results (I) - 全部16筆資料(7/25-11/12)

韓國瑜

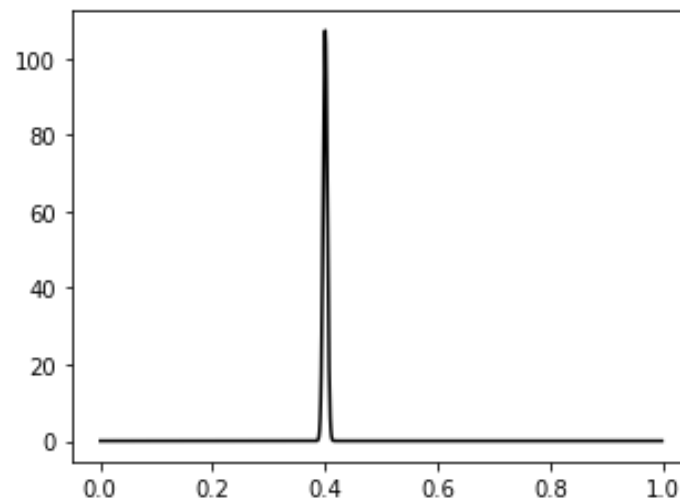
Final posterior dist.	mean	mode
Beta(6918, 10476)	0.3977	0.3977

Final posterior dist.	mean	mode
Beta(6969,10425)	0.4007	0.4006

陳其邁



韓國瑜-Beta(6918, 10476)



陳其邁-Beta(6969,10425)

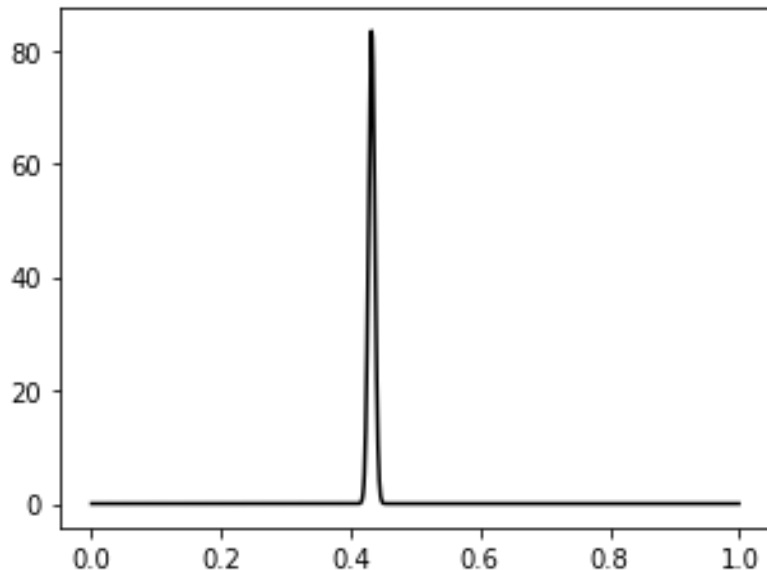
Results (II) - 10筆資料(10/30-11/12)

韓國瑜

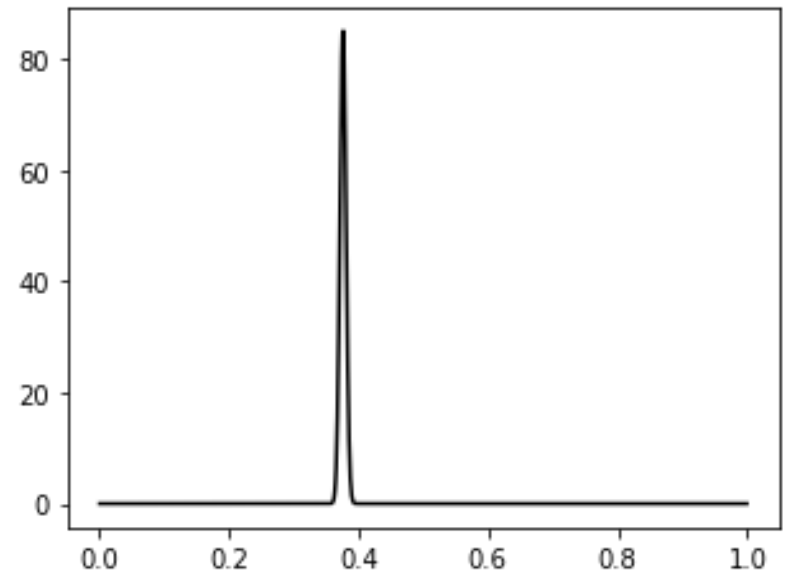
Final posterior dist.	mean	mode
Beta(4655, 6101)	0.4328	0.4328

Final posterior dist.	mean	mode
Beta(4047,6709)	0.3763	0.3762

陳其邁



韓國瑜-Beta(4655, 6101)



陳其邁-Beta(4047,6709)

Summary



	韓國瑜	陳其邁
Bayesian Updating 方法估計	52.78%	45.89%
2018 真實得票率	53.87%	44.80%

轉換百分比總和 (校正未表態)



✓ 民調普遍狀況：

有一定比例的民眾未表態，通常高於實際選舉總得票率。

✓ 民調未表態平均比率：21.3125%

✓ 修正方法：

利用上次高雄市長選舉(2014年)國民黨、民進黨總得票率為基底，等比例放大。

ex. 韓國瑜
得票率



98.98%
總得票率



韓國瑜得票率估計
韓國瑜得票率估計 + 陳其邁得票率估計

Final Summary

	韓國瑜	陳其邁	韓差距	陳差距	總差距
真實值	53.87%	44.80%			
Method 1 估計	52.27%	46.71%	-1.60%	1.91%	3.51%
Method 2 result 2估計(Nov.)	52.78%	45.89%	-1.09%	1.09%	2.18%
Method 2 result 1估計(All)	49.15%	49.52%	-4.72%	4.72%	9.44%
距離選舉最近之單筆民調估計 (11/11-11/12)	59.85%	39.13%	5.98%	-5.67%	11.65%



is Method 2 result 2估計(Nov)

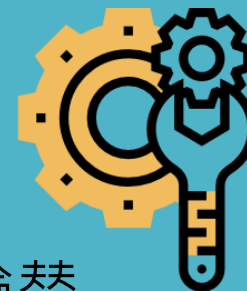
Discussion-Method 1



權重問題 - 政黨輪替？光看2014年夠嗎？

60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%

可以再改善權重的做法：



1. 希望之後能加入2018年這次政黨輪替
做之後民調公司權重的調整!!

2. 可以納入考量 2018.2014.2010年等
今年和之前的準確度 再調配權重。

結果日投入廿三十大學已

重

31

Discussion-Method 2



1.並沒有把民調公司的異質性區分開來-上下筆民調所屬之調查公司的立場可能也會影響迭代結果。

公司	民調日期	樣本個數 (沒有就寫：No)	韓國瑜	陳其邁
聯合報系民意調查中心	11-08~11-10	1074	49	32
ET民調	10-30~11-08	1000	42.6	36.6
TVBS民意調查中心	11-04~11-06	1077	52	47
三立	11-03~11-04	1023	43.2	44.3
三立新聞	11-11~11-12	1063	40.5	43.6
大社會民調中心	7-25~7-26	1,317	17.5	60.8
台灣世代智庫	10月19日	1000	52.3	29.1
台灣世代智庫	9月19日	1,079	30.2	41.00
台灣競爭力論壇	10月5日	1075	30.8	33.5

2.兩筆資料間的時間差距大小也可以試著融入模型。

Conclusion

1. 同一問題可從不同角度切入，衍生出相對應的分析方法。
2. 民調預測結果比時間最接近選舉的單筆民調預測結果還要準確，總誤差大小分別為**3.51%**及**2.18%**。當選人之預測也與真實結果相同。
3. 就本次用於分析的資料和方法而言，發現雖然結合過去資訊有助於預測未來；但是離選舉過於久遠的民調反而可能混淆資訊，使結果不能反映真實狀況。因此，可考慮採納距選舉固定一段時間的資料參考即可。

References

網站：

無情真實的未來預測：<https://tsjh301.blogspot.com/>

維基百科

論文：

張順全、莊文忠(2008)。探索選民的投票行為變化—應用機率分配模型的預測方法。選舉研究，15(2)，91-117。

Using state polls to forecast presidential election outcomes in the American states (Thomas M. Holbrook, Jay A. DeSart)

Thank You for
Your Attention!
&
Questions

