

國立高雄大學資訊管理學系

統計學期末專題報告

市長連任與否統計分析

組別：第六組

組員：A1113306 曾梓庭

A1113316 莊依庭

A1113319 陳德諭

A1113322 陳思卉

A1113323 邱怡瑄

目錄

壹、動機與目的	1
貳、資料集敘述	1
參、統計分析與結果	2
一、敘述統計	1
二、推論統計	3
肆、討論與建議	12
伍、結論與對專題練習的建議	12

壹、動機與目的

一、動機

在當今社會，市民對於市長的滿意度是衡量政府績效的重要指標之一，不僅影響政府的公信力和聲譽，還直接關係到市長及其團隊的連任可能性。因此，政府單位若了解影響市民滿意度與連任支持度的關鍵因素，能幫助政府制定更有效的政策與曝光方針。

本研究旨在透過統計分析，深入了解市民對於高雄市市長滿意度的影響因素。具體而言，我們將根據問卷的調查結果，統整市民對市長政見兌現程度、政績能見度和上新聞頻率等方面的評價，並結合市民的性別、年齡、教育程度、行業和收入等特徵，進行全面的數據分析。

二、目的

1. 分析不同人口學特徵（性別、年齡、教育程度、行業和收入）對各項問卷指標的差異。
2. 評估政見兌現程度、政績能見度和上新聞頻率與市長滿意度乃至連任可能性有無顯著關係。
3. 提供實證依據與建議，提升市政管理水平，供市長及其團隊參考相關策略，提高市民滿意度或增強市民對市長連任的支持可能性。

貳、資料集敘述

項目	說明
領域	市政績效
資料集大小及欄位	1. 共有 180 筆 2. 共 29 欄： 性別、年齡、教育程度、行業別、平均月收入、政見兌現程度 1~5、政績能見度 1~5、上新聞頻率 1~5、滿意程度 1~5、連任可能性 1~4

參、統計分析與結果

一、敘述統計

1. 調查數據中男、女比例：

圖 1-1 用柱狀圖呈現在調查中男女受訪者的數量分布和差異，受訪者男女比差異小，若以性別來分群，說明該問卷調查具代表性和公正性。

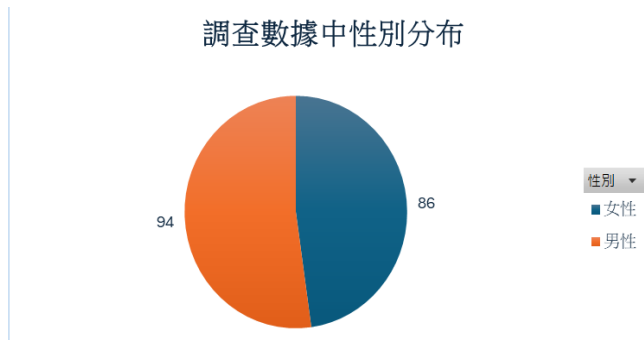


圖 1-1 男女比例

2. 調查數據中年齡層分布：

圖 1-2 用柱狀圖呈現在調查中不同年齡層的分布和比例，調查樣本中以 30~39 歲的年齡層佔最多數，其次是 20~29 歲以及 40~49 歲年齡區間，接著是 50 歲以上的受訪者，20 歲以下的年齡層最少。可分析不同年齡層對高雄市長的各項指標是否有顯著差異，進而傳遞出某年齡層對政治的關注和有感度。

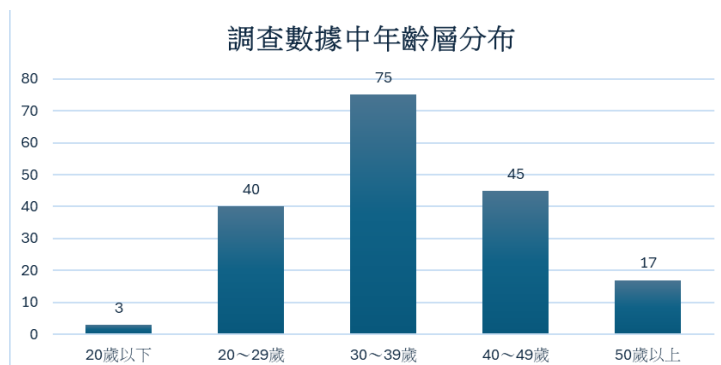


圖 1-2 年齡層分布

3. 調查數據中教育程度比例：

圖 1-3 以圓餅圖呈現在調查中受訪者的教育程度分布和比例，受訪者的教育程度以大專院校為最大宗，接著是碩士(含)以上的受訪者，高中(職)受訪者占少數，而國中以下的受訪者則為 0，受訪者的教育程度有明顯落差。

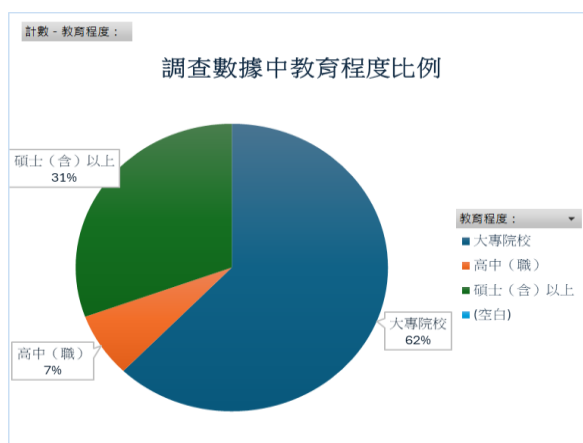


圖 1-3 教育程度比例

4. 調查數據中行業類別的分布：

圖 1-4 用圓餅圖呈現在調查中受訪者的教育程度分布和比例。受訪者的行業類別以其它的分類為最大宗，接著是製造業、資訊服務業和金融服務業，最後分別為零售貿易業和物流業，受訪者的行業分布並不平均，極度集中於製造業。

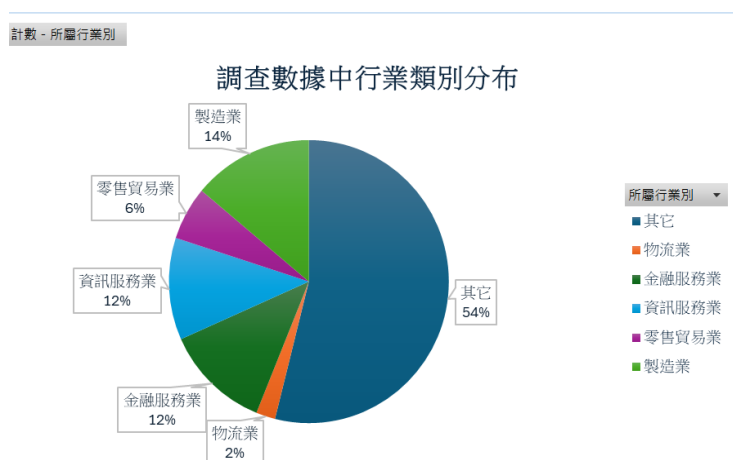


圖 1-4 行業類別分布

5. 調查數據中平均月收入的分布：

圖 1-5 用柱狀圖呈現在調查中受訪者平均月收入的分布和比例，受訪者的平均月收入大多位於 30,001~50,000 元的區間裡，而平均月收入為 20,000 元以下的受訪者人數則為 0，受訪者收入分布有明顯落差。

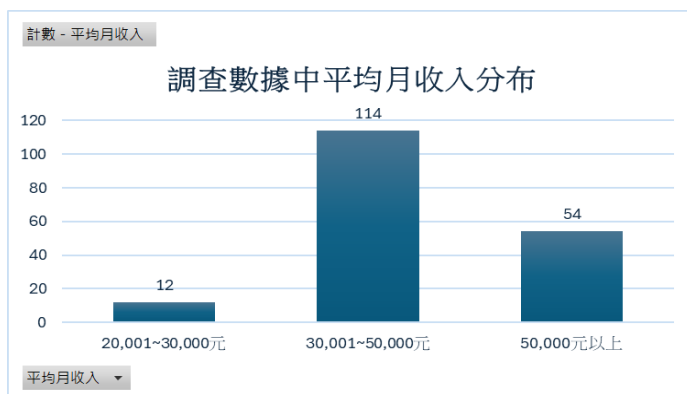


圖 1-5 平均月收入分布

二、推論統計

1. 假設檢定

因為製造業在政見兌現程度平均值最高，故想探討製造業對於政見兌現的在意程度，並和其他產業類別進行比較，此外，我們也想比較不同行業間對於市長上新聞頻率的在意程度有無顯著差異。

(1) 假設：製造業別對於政見兌現程度大於(等於)平均值 5.7

H0：製造業對於市長的政見兌現在意程度 ≥ 5.7

H_a：製造業對於市長的政見兌現在意程度 < 5.7

平均數	5.648
標準差	1.2169
標準誤差	0.2434
T	-0.2137
P	0.4163

表 2-1 單母體檢定

因為 P 值(0.4163) \gg 假設的顯著水準 $\alpha = 0.05$ ，接受 H_a，因此我們的結論是：製造業對於市長的政見兌現在意程度不高

(2) 假設：製造業和資訊服務業對政見兌現在意程度無顯著差異

H0：製造業和資訊服務業對政見兌現在意程度無顯著差異

H_a：製造業和資訊服務業對政見兌現在意程度有顯著差異

F 檢定：兩個常態母體變異數的檢定

	製造業	資訊服務業
平均數	5.648	4.8857
變異數	1.4809	1.7383
觀察值個數	25	21
自由度	24	20
F 值	0.852	
P 值	0.6494	
臨界值	2.0825	

使用 F 檢定來判斷兩個常態母體變異數是否相同，分析數據為 F 值 $<$ 臨界值(2.0825)、P 值 > 0.05 ，因此沒有足夠的證據表明兩組數據的變異數不相等，假設兩者變異數相同。

在兩者變異數相同的情況下，我們使用 T 檢定判別製造業和資訊服務業對政見兌現在意程度無顯著差異。數據結果為 t 值 $>$ 臨界值(2.013)、P 值 < 0.05 ，拒絕原假設 H0，因此我們的結論為：製造業和資訊服務業對政見兌現在意程度有顯著差異。

表2-3 t 檢定：兩個母體平均數差的檢定，假設變異數不相等

	製造業	資訊服務業
平均數	5.648	4.8857
變異數	1.4809	1.7383
觀察值個數	25	21
假設的均數差	0	
自由度	41.26	
t 統計	2.024	
P 值(雙尾)	0.0119	
臨界值(雙尾)	2.013	

(3) 假設：製造業和資訊服務業對上新聞頻率在意程度無顯著差異

H0：製造業和資訊服務業對上新聞頻率在意程度無顯著差異

H_a：製造業和資訊服務業對上新聞頻率在意程度有顯著差異

表2-4 F 檢定：兩個常態母體變異數的檢定

	製造業	資訊服務業
平均數	4.984	4.9905
變異數	1.7464	1.1979
觀察值個數	25	21
自由度	24	20
F 值	1.4579	
P 值	0.1977	
臨界值	2.0825	

使用 F 檢定來判斷兩個常態母體變異數是否相同，數據結果為 F 值<臨界值(2.0825)、P 值>0.05，因此沒有足夠的證據表明兩組數據的變異數不相等，假設兩者變異數相同。

在兩者變異數相同的情況下，我們使用 T 檢定判別製造業和資訊服務業對上新聞頻率在意程度無顯著差異。數據結果為 t 值<臨界值(-0.015)、P 值>0.05，接受原假設 H0，因此我們的結論為：製造業和資訊服務業對上新聞頻率在意程度無顯著差異。

表2-5 t檢定：兩個母體平均數差的檢定

	製造業	資訊服務業
平均數	4.984	4.9905
變異數	1.7464	1.1979
觀察值個數	25	21
假設的均數差	0	
自由度	43.98	
t 統計	-0.018	
P值(雙尾)	0.247	
臨界值(雙尾)	-0.015	

2. ANOVA 變異數分析

a. ANOVA 單因子變異數分析

i. 連任可能性是否會因教育程度而有顯著的差異？

H0：不同教育程度的人，連任可能性**不會**因教育程度不同而有顯著差異

Ha：不同教育程度的人，連任可能性**會**因為教育程度不同而有顯著差異

表 2-6 單因子變異數分析(教育程度)

摘要				
組	個數	總和	平均	變異數
高中（職）	13	58	4.46154	0.6234
大專院校	112	535.25	4.77902	1.40737
碩士（含）以上	55	275.75	5.01364	1.8968

ANOVA						
變源	SS	自由度	MS	F	P-值	臨界值
組間	3.90988	2	1.95494	1.30023	0.27506	3.04701
組內	266.126	177	1.50354			
總和	270.036	179				

由表 2-6 可知，F 值<臨界值，P 值>0.05，因此可推論：不同教育程度的人，連任的可能性**不會**因為教育程度不同而有顯著的差異。

ii. 連任可能性是否會因所屬行業別而有顯著的差異？

H0：所屬不同行業別的人，連任的可能性**不會**因為行業別不同而有顯著的差異。

Ha：所屬不同行業別的人，連任的可能性**會**因為行業別不同而有顯著的差異。

表 2-7 單因子變異數分析(所屬行業別)

摘要					
組	個數	總和	平均	變異數	
製造業	25	125.25	5.01	1.4452	
物流業	4	16	4	2.6667	
零售貿易業	11	56.25	5.1136	1.267	
金融服務業	22	108	4.9091	1.914	
資訊服務業	21	101.75	4.8452	1.403	
其它	97	461.75	4.7603	1.4725	

ANOVA						
變源	SS	自由度	MS	F	P-值	臨界值
組間	5.0633	5	1.0127	0.665	0.6505	2.2661
組內	264.97	174	1.5228			
總和	270.04	179				

由表 2-7 可知，F 值 < 臨界值，P 值 > 0.05，因此可推論：所屬不同行業別的人，連任的可能性不會因為行業別不同而有顯著的差異。

b. ANOVA 雙因子變異數分析

- i. 不同性別與不同的年齡是否會影響連任可能性，而因此有顯著的差異？

H0：不同性別但相同年齡的人，連任可能性不會有顯著差異。

Ha：不同性別但相同年齡的人，連任可能性會有顯著差異。

H0：不同年齡但相同性別的人，連任可能性不會有顯著差異。

Ha：不同年齡但相同性別的人，連任可能性會有顯著差異。

表 2-8 雙因子變異數分析：無重複試驗(連任可能性)

摘要	個數	總和	平均	變異數		
男	5	25.02089947	5.004179894	0.027412802		
女	5	22.50526316	4.501052632	0.476723376		
20歲以下	2	10.125	5.0625	0.0078125		
20~29歲	2	10.0933584	5.046679198	0.006864161		
30~39歲	2	9.491666666	4.745833333	0.0003125		
40~49歲	2	9.435185185	4.717592593	0.437285665		
50歲以上	2	8.380952381	4.190476191	1.197562357		

ANOVA						
變源	SS	自由度	MS	F	P-值	臨界值
列	0.632842605	1	0.632842605	2.489069733	0.189776247	7.708647422
欄	0.999550134	4	0.249887533	0.982847063	0.506487834	6.388232909
錯誤	1.016994578	4	0.254248644			
總和	2.649387317	9				

由表 2-8 可知，在年齡中，因 $P > 0.05$ ，所以我們推論：不同年齡但相同性別的人，連任可能性不會有顯著差異。而在性別中，因 $P > 0.05$ ，所以我們推論：不同性別但相同年齡的人，連任可能性不會有顯著差異。

3. 迴歸分析

(1) 簡單迴歸

- 第一部分，先建構三組模型，分別設定自變項為上新聞頻率、政績能見度、政見兌現程度，而依變項為滿意程度，但三組模型的 R 平方皆小於 0.5，模型擬合有問題，可信度不高，所以嘗試再將應變數改成「連任可能性」，開啟第二部分，再度建構三組新模型。
- 其中，當設定自變項為上新聞頻率或政績能見度，而依變項為連任可能性時，R 平方分別為 0.6983、0.5541，說明這兩個模型解釋變異的能力較強，適合進行迴歸分析。

觀察兩模型各自的統計表，可以歸納以下共同點：

■ $\alpha=0.05$ ，且 ANOVA 的顯著值和 T 檢定的 p 值皆小於 0.05，拒絕 H_0 ，也就是說，自變項的變化顯著影響依變項，這可推論以自變項「上新聞頻率」或是「政績能見度」去預測依變項「連任可能性」的結果是合適的。

■ 根據係數，得出兩組線性關係公式

1. 連任可能性 = $0.694001 + 0.84436 * \text{上新聞頻率}$

- a. 將上新聞頻率的實際值代入公式，得出「連任可能性的預測值」，再與「連任可能性的實際值」對比，計算得出均方根誤差 $RMSE=0.667365$ ，因為該值介於 0~1 之間，一般來說，可以稱此模型預測績效是好的（誤差小）。

2. 連任可能性 = $1.08022 + 0.84659 * \text{政績能見度}$

- a. 將政績能見度的實際值代入公式，得出「連任可能性的預測值」，再與「連任可能性的實際值」對比，計算得出均方根誤差 $RMSE=6.48496$ ，該值較大，也就代表此模型預測績效相對較不精準。

表 3-1 上新聞頻率與連任可能性的關係分析表

迴歸統計		ANOVA					
R 的倍數	0.83567	自由度	SS	MS	F	顯著值	
R 平方	0.698345	迴歸 R	1	185.5465	185.5465	409.7622	6.18E-48
調整的 R	0.69664	殘差 E	177	80.14829	0.452815		
標準誤	0.672915	總和 T	178	265.6948			
觀察值個數	179						

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.694001	0.210871	3.291108	0.001205	0.277855	1.110146	0.277855	1.110146
2.6	0.84436	0.041712	20.24258	6.18E-48	0.762043	0.926677	0.762043	0.926677

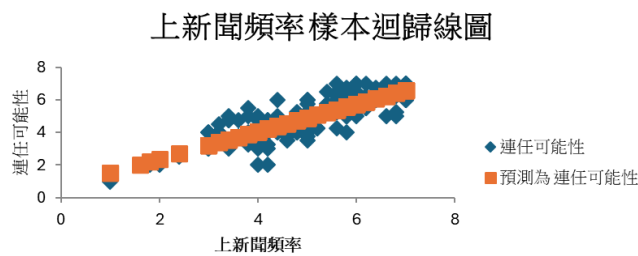


圖 3-1 上新聞頻率樣本迴歸線圖

表 3-2 政績能見度與連任可能性的關係分析表

迴歸統計		ANOVA					
R 的倍數	0.744348		自由度	SS	MS	F	顯著值
R 平方	0.554053	迴歸	1	149.6144	149.6144	221.151	4.88E-33
調整的 R	0.551548	殘差	178	120.4217	0.676526		
標準誤	0.822512	總和	179	270.0361			
觀察值個數	180						

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	1.080221	0.259352	4.165076	4.85E-05	0.56842	1.592021	0.56842	1.592021
X 變數 1	0.846587	0.056928	14.87115	4.88E-33	0.734246	0.958928	0.734246	0.958928

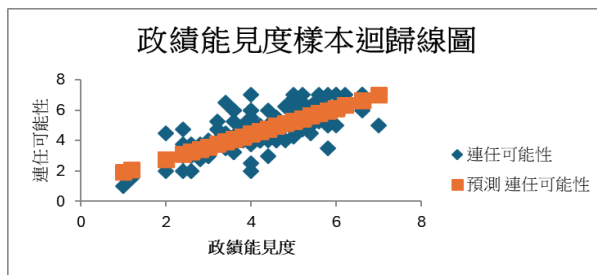


圖 3-2 政績能見度樣本迴歸線圖

(2) 複迴歸

- 在複迴歸分析中，我們以政見兌現程度、政績能見度、上新聞頻率及滿意程度為自變數，而連任可能性為依變數做迴歸分析，以此來探討四種自變數(政見兌現程度、政績能見度、上新聞頻率及滿意程度)對連任可能性是否有顯著影響。
- 根據分析結果(表 3-3)來看，R 平方為 0.757193，而從這個值中可得知該迴歸式的全部自變數對依變數的解釋能力是偏高的。
- 接著我們用 F 檢定來檢查在迴歸模型中的係數是否全部為 0，當係數不全部為 0 時，迴歸模型才具有預測力，否則，我們需要考慮使用非線性迴歸模型來預測變數。
 - 首先我們先設 $\alpha = 0.05$ ，接著將虛無假說設為係數全部為 0，而對立假說設為係數不全部為 0，ANOVA 表格的計算結果顯示顯著值為 1.09E-52，遠小於 0.05
 - 因此，我們可以拒絕虛無假說，接受該迴歸模型迴歸係數不全部為 0 的聲稱，換句話說，在迴歸式中至少存在一個係數不為 0。

- 最後進行 T 檢定檢查每個迴歸係數是否顯著，在「政績能見度均值」這列中的 p 值為 0.187，大於 0.05，可看出該自變數不顯著，只有「政見兌現程度均值」、「上新聞頻率均值」以及「滿意程度均值」用來預測「連任可能性」的結果為顯著的。

表 3-3 四個自變數與連任可能性的關係分析表

迴歸統計							
R 的倍數	0.870168	ANOVA					
R 平方	0.757193		自由度	SS	MS	F	顯著值
調整的 R 平方	0.751643	迴歸	4	204.4693	51.11733	136.4339	1.09E-52
標準誤	0.612101	殘差	175	65.56677	0.374667		
觀察值個數	180	總和	179	270.0361			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.08677	0.22156	0.39164	0.6958	-0.3505	0.52405	-0.3505	0.52405
政見兌現程度均值(X1)	0.1092	0.04884	2.23575	0.02663	0.0128	0.20559	0.0128	0.20559
政績能見度均值(X2)	0.10076	0.07606	1.32473	0.18699	-0.04935	0.25087	-0.04935	0.25087
上新聞頻率均值(X3)	0.56409	0.06656	8.47509	9.4E-15	0.43273	0.69545	0.43273	0.69545
滿意程度均值(X4)	0.23636	0.04939	4.78594	3.6E-06	0.13889	0.33383	0.13889	0.33383

- 為何政績能見度使用簡單迴歸分析的結果是顯著的，但使用複迴歸卻變不顯著？
- 可能的原因為變數間存在交互作用，變數間會互相影響，在做簡單迴歸時通常不會考慮交互作用，但在做複迴歸時會考慮，也因此複迴歸中此項變數變為不顯著。
 - 不顯著的變數必須從函數中被移除，若未移除，則在未來應用該函數時會導致預測精準度降低。
- 將不顯著變數移除後重新再做一次複迴歸分析，這次我們以政見兌現程度、上新聞頻率及滿意程度為自變數，而連任可能性為依變數來探討兩者間是否有顯著關係，結果如表 3-4 所示。

表 3-4 刪除不顯著變數後重新做複迴歸

迴歸統計							
R 的倍數	0.868768	ANOVA					
R 平方	0.754758		自由度	SS	MS	F	顯著值
調整的 R 平方	0.750577	迴歸	3	203.8118	67.93728	180.5525	1.78E-53
標準誤	0.613412	殘差	176	66.22429	0.376274		
觀察值個數	180	總和	179	270.0361			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	0.12493	0.220152	0.567472	0.571117	-0.30955	0.559407	-0.30955	0.559407
政見兌現程度平均(X1)	0.128422	0.046734	2.747945	0.006622	0.036191	0.220653	0.036191	0.220653
上新聞頻率平均(X2)	0.61556	0.054156	11.36649	8.39E-23	0.508682	0.722438	0.508682	0.722438
滿意程度平均(X3)	0.250317	0.048352	5.176931	6.11E-07	0.154892	0.345742	0.154892	0.345742

- 透過結果計算出該迴歸模型的公式為：

連任可能性=0.1249+0.1284*政見兌現程度+0.6156*上新聞頻率+0.2503*滿意程度

- 小結：統整迴歸分析結果，除了政績能見度外的其他自變數(政見兌現程度、上新聞頻率及滿意程度)都與連任可能性有關。

肆、討論與建議

根據上述的統計結果，我們從假設檢定中得知製造業對政見兌現的在意程度不高，且製造業和資訊服務業對政見兌現在意程度有顯著差異，製造業的政見兌現平均是所有行業別中最高，因此推論各行業對高雄市長政見兌現之在意程度普遍不高，並且能夠從製造業和資訊服務業對上新聞頻率在意程度無顯著差異之結果推論，不同行業別與市長上新聞頻率之在意程度間無顯著相關性。

從迴歸分析之結果得知，政績能見度 p 值大於 0.05，得知此變數不顯著後，將其刪除再做一次複迴歸的 R 平方值為 0.75，從這值可看出自變數對依變數的解釋能力偏高，也因此我們能推斷除了政績能見度外的其他因素(政見兌現、上新聞頻率及滿意程度)都與連任可能性有關，我們建議市長能夠增加上新聞的曝光度、定期參考追蹤市民的滿意程度變化，以得知政策與市政加強方向有無符合市民期望，進而提升連任可能性，最後是市長之政見內容可以跨越各產業，讓所有行業之市民更能主動關注其政見內容，就能提高每個產業對政見兌現在意程度，並影響滿意程度，這也有助於提升市長連任之可能性。

伍、結論與對專題練習的建議

在學完兩個學期的統計學理論後，我們終於有機會能夠將所學實際應用並做出一份專案。在這過程中，我們發現有許多知識盲點都是在實際操作過後才能被真正理解，同時，透過專案我們也不再將思考僅僅侷限在數據層面，更多的是去判斷該如何利用統計分析找出我們感興趣的問題解答，並給出具體管理層面的建議。