

政 治 大 學  
統 計 學 系 學 士 班

多 變 量 分 析  
期 中 報 告

Multivariate Analysis: Midterm Project

指導教授： 鄭宗記 教授

學 生： 陳葳芃 歐西四 107508006

學 生： 賴冠維 統計三 109304008

# 目錄

Part I. 資料集介紹 .....	2
Section 1-1 資料介紹 .....	2
Section 1-2 變數介紹 .....	2
Section 1-4 研究目的及分析方法 .....	3
Part II. 針對「地形」處理資料 .....	4
Section 2-1 資料前置處理 .....	4
Section 2-2 研究變數選取 .....	4
Section 2-3 Linear Discriminant Analysis .....	6
Section 2-4 Quadratic Discriminant Analysis .....	8
Section 2-5 Mixture Discriminant Analysis .....	10
Section 2-6 小結與發現 .....	11
Part III. 針對「車道寬度」處理資料 .....	12
Section 3-1 Linear Discriminant Analysis .....	13
Section 3-2 Mixture Discriminant Analysis .....	15
Section 3-3 Quadratic Discriminant Analysis .....	16
Part IV. 結論 .....	18

## Part I. 資料集介紹

### ➤ Section 1-1 資料介紹

本研究使用的資料為中華民國交通部公路總局每年發布的公路交通量調查統計表，資料調查年度為 111 年度調查期間涵蓋 111 年 5 月 12 日至 9 月 18 日，此分析我們將聚焦在資料中 18 個變數，共 1490 筆的樣本資料。

### ➤ Section 1-2 變數介紹

Table.1 十八筆變數的變數介紹

變數名稱	變數解釋	變數型態
X <sub>1</sub> 地形	包含 3 種不同的地形「丘嶺區」、「山嶺區」、「平原區」	類別
X <sub>2</sub> 經度	調查站的經度座標	數值
X <sub>3</sub> 緯度	調查站的緯度座標	數值
X <sub>4</sub> 里程 km	路段里程長度（單位：公里 km）	數值
X <sub>5</sub> 路面寬度	路面寬度（單位：公尺 m）	數值/類別 <sup>1</sup>
X <sub>6</sub> 路間寬度	路肩寬度（單位：公尺 m）	數值
X <sub>7</sub> 小型車數	小型車日行經車輛數	數值
X <sub>8</sub> 大客車數	大客車日行經車輛數	數值
X <sub>9</sub> 大貨車數	大貨車日行經車輛數	數值
X <sub>10</sub> 全聯結車數	全聯結車日行經車輛數	數值
X <sub>11</sub> 半聯結車數	半聯結車日行經車輛數	數值
X <sub>12</sub> 機車數	機車日行經車輛數	數值
X <sub>13</sub> 總流量 PCU	不同車種的車輛流量轉換為乘用車單位後的總流量	數值
X <sub>14</sub> 總計車公里	不同車種的車輛行經公里數的加總（單位：公里 km）	數值
X <sub>15</sub> 尖峰小時交通量 PCU	尖峰時刻不同車種的車輛流量轉換為乘用車單位後的總流量	數值
X <sub>16</sub> 快車道數量	路段所擁有的快車道數量	數值
X <sub>17</sub> 機慢車道數量	路段所擁有的機慢車道數量	數值
X <sub>18</sub> 合計車輛數	各類車種日行經車輛數	數值

<sup>1</sup> 在 Part I 後，我們嘗試在 Part II 對路面寬度進行分組，透過數值轉類別的方式，再進行一次類別分析，故此份研究報告中，X<sub>5</sub> 路面寬度會分別以數值、類別變數進行分析。

## ➤ Section 1-4 研究目的及分析方法

交通資料與施政者在做都市規劃及道路設計可謂環環相扣，而其中「地形」即是影響我們交通狀況的重要因素之一。此次研究我們想透過交通資料的觀察，理解不同地形條件下的交通行為，並根據觀察到的結果提供建議作為施政方針參考。

本研究採用的統計分析方法為區別分析 (Discriminant Analysis, DA)。首先，試圖從眾多變數當中，找出真正重要的變數；接著，根據 Wilks'lambda 及 minimum Mahalanobis Distance 的選取結果來建立模型，最後進行模型的正確率判斷，及視覺化模型的判斷準則。此次分析當中，將採用變數 2~11 項共 16 個數值型變數及 1 個類別型變數。

## Part II. 針對「地形」處理資料

### ➤ Section 2-1 資料前置處理

#### (1) 缺失值的處理

由於每筆樣本均為一個個調查站所蒐集到的觀測值，再缺失值處理上，我們不認為重新賦予值會是一個好的處理方式，因此我們分析的資料均移除含有 NA 的樣本。

#### (2) 異常值的處理

我們同時發現在其中一個觀測站，當初輸入的人員可能誤植導致經緯度值為一樣的，為了分析的合理性，我們一樣把這兩筆樣本予以移除。

#### (3) 資料的分割（訓練集、測試集）

在資料的建模與驗證部分，我們將樣本資料分割成 80% 的訓練集（模型建立）以及 20% 的測試集（驗證模型建立的準確性）。

### ➤ Section 2-2 研究變數選取

根據表（三）結果，Wilks'lambda test 中，在 95% 的信心水準下（ $\alpha=0.05$ ），所有針對預測變數（dependent variable）——『地形』的預測變數（independent variable）均具有統計顯著性，表示我們有足夠的證據顯示各預測變數組間具有差異。接著，我們計算各變數組間的 Mahalanobis Distance，並對其取最小值，再根據 *Minimum D<sup>2</sup>* 此欄位中依序選取較大的 mahalanobis distance 值來進行 stepwise 的模型建立。我們模型建立的比較順序為：全部變數（共 1 種組合）及『里程 km、路面寬度 m、快車道數量』這三個變數依序加入模型中（stepwise），總共 4 個模型來進行比較。

下方列出 Wilks'lambda test 的假設檢定虛無及對立假設

Hypothesis of Wilks'lambda test

H0: Group mean vectors are the same for all groups or they don't differ significantly.

H1: At least one of the group mean vectors is different from the rest

**Table.2 Group Descriptive Statistics and Tests of Equality for the Estimation Sample in the Three-Group Discriminant Analysis**

Independent Variables	Dependent Variable			Test of Equality			Minimum	
	Group Means:			of Group Means*			<i>Mahalanobis D<sup>2</sup></i>	
	X <sub>1</sub> 地形							
	Group1 丘陵區 n=165	Group2 山嶺區 n=68	Group3 平原區 n=957	Wilks' Lambda	F value	Significance	Minimum D <sup>2</sup>	Between Groups
X <sub>2</sub> 經度	121.002	120.997	120.809	0.975	15.405	0	0.002	1 and 2
X <sub>3</sub> 緯度	23.781	23.718	24.033	0.982	10.986	0	0.016	1 and 2
X <sub>4</sub> 里程 km	9.975	18.753	5.677	0.747	200.488	0	0.614	1 and 2
X <sub>5</sub> 路面寬度 m	16.356	10.051	21.332	0.825	126.305	0	0.477	1 and 3
X <sub>6</sub> 路肩寬度	1.157	0.727	1.698	0.943	36.147	0	0.210	1 and 3
X <sub>7</sub> 小型車數	3988.305	1396.256	8803.132	0.915	54.903	0	0.374	1 and 3
X <sub>8</sub> 大客車數	60.190	25.156	111.967	0.973	16.158	0	0.097	1 and 3
X <sub>9</sub> 大貨車數	162.335	49.344	460.223	0.934	42.053	0	0.318	1 and 3
X <sub>10</sub> 全聯結車數	1.930	0.811	10.682	0.965	21.665	0	0.055	1 and 2
X <sub>11</sub> 半聯結車數	94.325	16.333	328.904	0.960	24.87	0	0.202	1 and 3
X <sub>12</sub> 機車數	1526.265	371.378	3792.004	0.953	29.522	0	0.206	1 and 3
X <sub>13</sub> 總流量 PCU	5924.035	2908.144	13174.177	0.903	63.505	0	0.264	1 and 2
X <sub>14</sub> 總計車公里	45702.100	31996.133	64274.994	0.977	13.848	0	0.088	1 and 3
X <sub>15</sub> 尖峰小時交通量 PCU	658.470	391.367	1354.568	0.905	62.038	0	0.155	1 and 2
X <sub>16</sub> 快車道數量	1.600	1.144	2.026	0.850	104.855	0	0.453	1 and 3
X <sub>17</sub> 機慢車道數量	0.365	0.067	0.413	0.974	15.854	0	0.009	1 and 3

\*Wilks'lambda (U statistic) and univariate F ratio with 1 and 1188 degrees of freedom

## ➤ Section 2-3 Linear Discriminant Analysis

### ☆ LDA 概述：

找出一條線性組合，讓這條線能夠區分不同類別的資料，並達到同群內的群內變異數小，不同群的群間變異數大的效果。

依據 Mahalanobis distance 選取結果 ( $>0.4$ )，我們將  $X_4$   $X_5$   $X_{15}$  依序加入模型中，並比較模型的預測準確率。

可以由 Table.3 發現模型當放入全部變數去建模時，後續分類可以獲得將近八成七的準確率，而放入只放入一個變數 ( $X_4$  里程 km) 建模同樣是有相當好的效果 (高達八成三)，而依序加入新的預測變數，對準確度越發無顯著性的增加，甚至加入  $X_{15}$  尖峰小時交通量 PCU 時，與只放兩個變數準確率無異，考量到模型簡約性，我們欲將進行只放里程  $X_4$  里程 km、 $X_5$  路面寬度 m 兩個預測變數與變數全放進行比較。

Table.3 LDA model

model	LD1	LD2	accuracy
$L_1: X_1 \sim .$	0.936	0.064	0.879
$L_2: X_1 \sim X_4$	1		0.835
$L_3: X_1 \sim X_4 + X_5$	0.998	0.002	0.842
$L_4: X_1 \sim X_4 + X_5 + X_{15}$	0.996	0.004	0.842

Table.4 Confusion Matrix of  $L_1$  (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	47	16	20	9	4	2
2	10	47	7	1	16	2
3	108	5	930	25	2	237

Table.5 Confusion Matrix of  $L_3$  (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	2	10	2	0	2	0
2	19	34	6	5	12	2
3	144	24	949	30	8	239

經由上 Table.4 Table.5 Confusion Martix 的結果，我們可以看見分類最大宗是在 Group1 與 Group3 的辨別、其次為 Group1 與 Group2，而 Group2 與 Group3 在兩種模型、測試集資料驗證下，都符合我們對 Maha-distance 的最小值計算結果理解，Group2 與 Group3 通常相距較遠，故分錯的機會相較其他兩種可能還小很多。

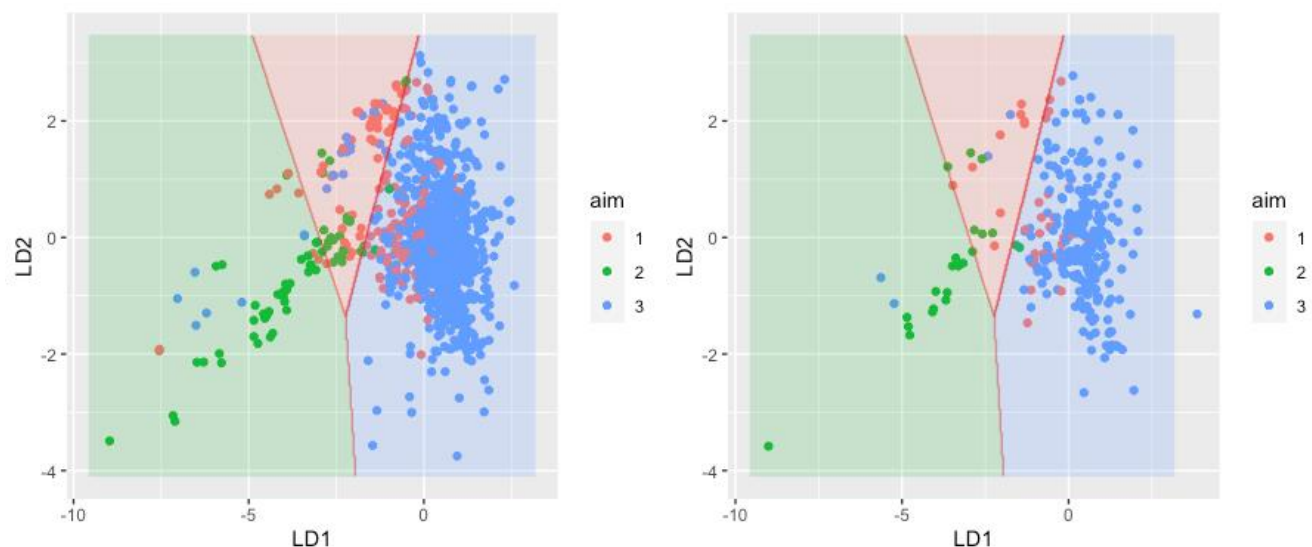


Figure.1  $L_1$  LDA boundaries plot (Training and Test)

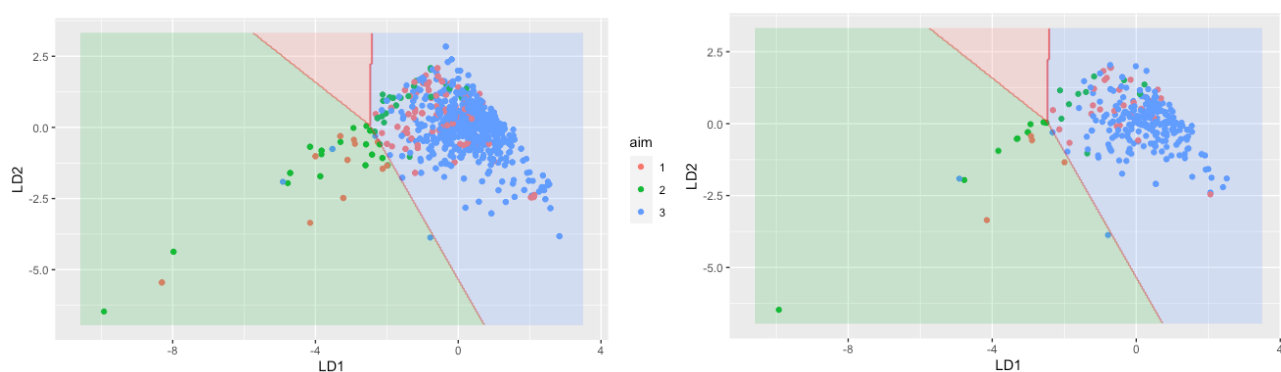


Figure.2  $L_3$  LDA boundaries plot (Training and Test)

觀察 Figure.1 可以很好的視覺化判斷三群的分類情形，初步可以看出 Group1 與 Group3 有高度的重疊（距離較小）



## ➤ Section 2-4 Quadratic Discriminant Analysis

### ✧ QDA 概述：

與 LDA 相似，但 LDA 是依一直線去區分資料來分類，QDA 則用二次曲線將資料進行分類。要注意的是使用 QDA，資料假設為服從多元常態，但根據我們檢定這筆資料的結果，此筆資料並無服從多元常態，但為了練習方便，我們還是假設這筆資料服從多元常態進行分析。

依據 Maha distance 選取結果 ( $>0.4$ )，我們將  $X_4$   $X_5$   $X_{15}$  依序加入模型中，並比較模型的預測準確率。

常理來說 QDA 能使用二次曲線將資料進行分類效果有可能會提升，但推測本次的資料集所有變數均不符合常態假設，違反了模型的使用的假設，也因此模型放入全部的情形下，預測結果既不如預期，也有明顯下降。

Table.6 LDA model

model	accuracy
Q1: $X_1 \sim .$	0.724
Q2: $X_1 \sim X_4$	0.812
Q3: $X_1 \sim X_4 + X_5$	0.832
Q4: $X_1 \sim X_4 + X_5 + X_{15}$	0.842

Table.7 Confusion Matrix of Q<sub>1</sub> (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	142	1	270	32	0	78
2	9	67	10	2	22	1
3	14	0	677	1	0	162

Table.8 Confusion Matrix of Q<sub>3</sub> (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	2	10	2	3	4	7
2	19	34	6	4	12	1
3	144	24	949	28	6	233

根據 Table.7 Table.8 Confusion Martix 的結果，與 LDA 相比，可以知道在放入全部預測變數時 QDA 模型能夠以相當優異的表現辨別出 Group2；反觀若只考量到模型的簡約採用模型 3，Group2 的

正確率又有下降。同樣，使用測試集資料驗證下， $L_1$  下則可以很好的進行區別分析出 Group2， $Q_3$  在各項預測表現都有出錯。

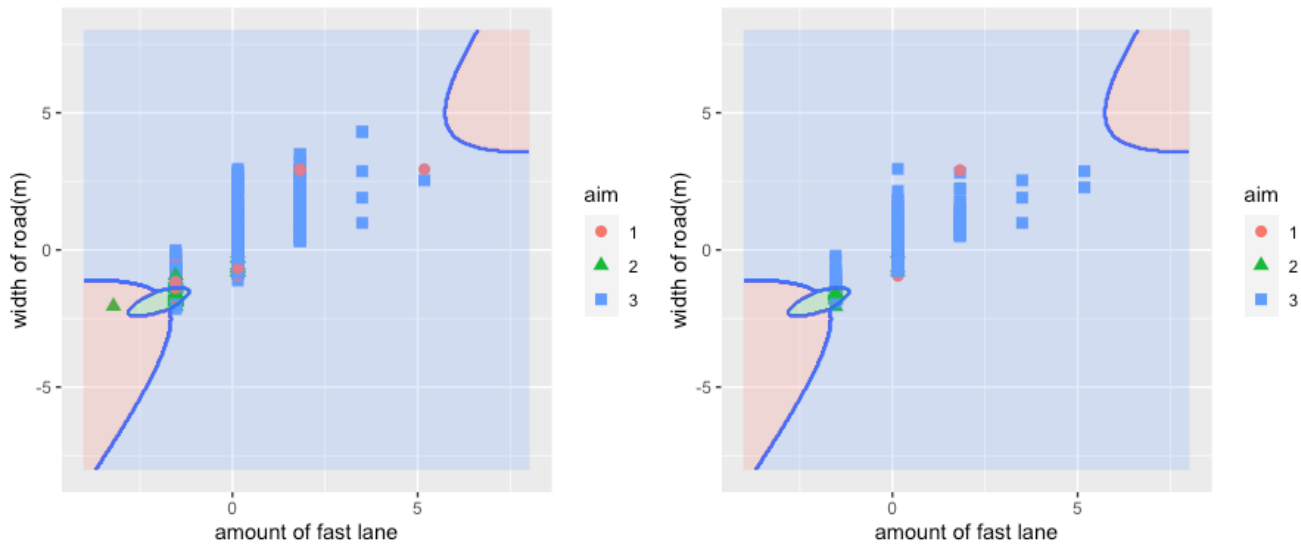


Figure.3  $Q_3$  LDA boundaries plot (Training and Test)

## ➤ Section 2-5 Mixture Discriminant Analysis

### ✧ MDA 概述：

為 LDA 的過展，但 LDA 是每個類別數據來自單一常分佈，MDA 假設混合的常態分配，其通常需要通過迭代完成參數的估計。MDA 會受到起始選擇參數的影響，因此最好進行多次 MDA 運算以選取較佳結果。

Table.9 LDA model

model	accuracy
$M_1: X_1 \sim .$	0.882
$M_2: X_1 \sim X_4$	0.822
$M_3: X_1 \sim X_4 + X_5$	0.849
$M_4: X_1 \sim X_4 + X_5 + X_{15}$	0.845

Table.10 Confusion Matrix of  $L_1$  (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	55	10	32	9	5	4
2	11	55	6	2	17	2
3	99	3	919	24	0	235

Table.11 Confusion Matrix of  $L_3$  (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	5	5	7	1	1	1
2	15	42	11	5	12	3
3	145	21	939	29	9	237

根據 Table.10 Table.11，MDA 做出的結果感覺優於 QDA，但同樣在 Group1 的分類上相當不精準，值得注意的是在對於 Group3 來說分類中似乎更有效的分類，也因此準確率有上升。

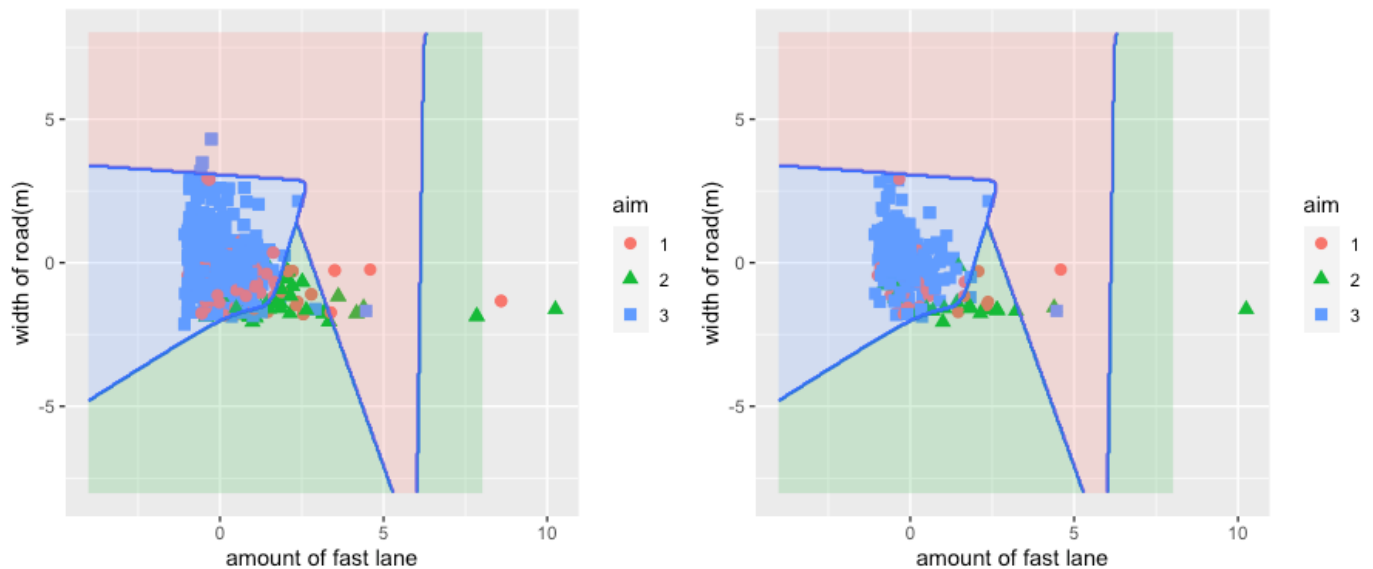


Figure.4 Q<sub>3</sub> LDA boundaries plot (Training and Test)

可藉由 Figure.3 發現 MDA 在 Group1 分類效果更為差強人意，無論是在訓練集或測試集中結果都相當不高，我們也可藉由此 Figure 發現只單看模型準確度的問題，將在後續 Section 詳細提及。

## ➤ Section 2-6 小結與發現

綜合以上多種分析方法，可以發現放入越多變數使預測結果會更為精準，關鍵會在於 Group2 分類情況。從其他圖形檢驗 Group1 與 Group3 經常會有許多重疊的部分，使模型預測失準，在特定區域就甚至只判斷為 Group3，使得 Group1 的預測效果在不同模型下都表現不佳。

若只驗證模型分類的準確率，乍看之下以為模型只要一個預測變數就能得到很好的預測效果 (>80%)，但在檢驗視覺化結果時，就可以發現模型可能嚴重分錯，將所有 testing data 中大部分類別變數都判別為 Group3 平原區，回頭檢驗這筆類別資料，Group 其實本賴就已經佔比資料的將近 8 成，故模型就算全部分錯也會有表現很好的錯覺，多種模型下結果均顯示模型時常在 Group1 時出錯分類為 Group3。

在下一個部分，我們也嘗試探索此筆資料將對路面寬度進行分組，透過數值轉類別的方式，再進行一次類別分析，而路面寬度也將由預測變數轉變為目標變數。

### Part III. 針對「車道寬度」處理資料

**TABLE 5 Group Descriptive Statistics and Tests of Equality for the Estimation Sample in the Two-Group**

Independent Variables	Dependent Variable			Test of Equality			Minimum	
	Group Means:			of Group Means*			Mahalanobis D <sup>2</sup>	
	area							
	Group1 width=1 n=306	Group2 width=2 n=664	Group3 width=3 n=520	Wilks' Lambda	F value	Significance	Minimum D <sup>2</sup>	Between Groups
經度	121.039	120.798	120.779	.955	28.284	.000	.001	1 and 2
緯度	23.842	23.918	24.115	.982	11.019	.000	.106	2 and 3
里程 km	11.672	6.537	5.012	.861	95.780	.000	.227	2 and 3
路肩寬度	0.836	1.565	1.972	.868	90.445	.000	.200	2 and 3
小型車數	2730.235	7412.722	11119.914	.847	107.443	.000	.093	2 and 3
大客車數	36.992	84.844	162.654	.907	60.759	.000	.263	2 and 3
大貨車數	98.437	394.670	577.246	.889	74.299	.000	.045	2 and 3
全聯結車數	1.806	10.646	12.117	.962	23.268	.000	.058	2 and 3
半聯結車數	69.895	338.316	337.282	.958	26.229	.000	.083	2 and 3
機車數	1202.146	2735.721	5554.668	.888	75.056	.000	.239	2 and 3
總流量 PCU	4147.709	11008.276	17080.807	.803	146.030	.000	.268	2 and 3
總計車公里	32995.53	58728.46	79847.29	.920	51.724	.000	.122	2 and 3
尖峰小時交通量 PCU	477.000	1139.441	1777.119	.781	166.434	.000	.314	2 and 3
快車道數量	1.081	1.949	2.358	.401	888.915	.000	.500	2 and 3
機慢車道數量	0.275	0.222	0.683	.828	123.598	.000	.036	1 and 2

### ➤ Section 3-1 Linear Discriminant Analysis

表 12 中呈現了不同組別的各變量平均數、F 值、Lambda 值等不同數據。在這邊我們嘗試利用不同的選取方式進行 LDA 的變數挑選，挑選標準為從擁有最大 F 值的變數最先挑選，並且根據 F 值大小依序放入模型，直到模型內有四個自變數。由上面的挑選標準，將會對以下五種模型進行分析：

L<sub>1</sub>:  $X_6 \sim .$

L<sub>2</sub>:  $X_6 \sim X_{16}$

L<sub>3</sub>:  $X_6 \sim X_{15} + X_{16}$

L<sub>4</sub>:  $X_6 \sim X_{13} + X_{15} + X_{16}$

L<sub>5</sub>:  $X_6 \sim X_{13} + X_{15} + X_{16} + X_{17}$

Table.13 LDA model

Model	LD1	LD2	Accuracy(Test)
L <sub>1</sub> : $X_6 \sim .$	0.936	0.064	0.876
L <sub>2</sub> : $X_6 \sim X_{16}$	1.578		0.738
L <sub>3</sub> : $X_6 \sim X_{15} + X_{16}$	0.989	0.011	0.725
L <sub>4</sub> : $X_6 \sim X_{13} + X_{15} + X_{16}$	0.989	0.011	0.742
L <sub>5</sub> : $X_6 \sim X_{13} + X_{15} + X_{16} + X_{17}$	0.942	0.058	0.812

根據表 13 可以觀察到雖然當所有變數都放入模型時區分的精確度達 87.6%，但是 L<sub>5</sub> 僅用了四個自變數就可以使精確度達到 81.2%。因此我們根據模型精確度和簡約度之挑選標準挑選出 L<sub>1</sub> 和 L<sub>5</sub> 並對其進行比較分析。

Table.14 Prior probabilities of groups

1	2	3
0.207	0.441	0.352

Table.15 Confusion Matrix of L<sub>1</sub> (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	226	18	1	56	5	0
2	21	477	94	3	127	23
3	0	31	324	0	6	78

Table.16 Confusion Matrix of L<sub>5</sub> (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	226	27	2	56	8	0
2	21	404	73	3	104	19
3	0	95	344	0	26	82

表 15、表 16 中可以觀察到 L<sub>1</sub> 和 L<sub>5</sub> 在區分 Group1、Group2 和 Group1、Group3 時均有較高的精確度，但在區分 Group2、Group3 時較不精確，其中 L<sub>5</sub> 在這兩組之間的區分能力較弱。但根據表 14 可看到 Group2 和 Group3 佔了整體資料 79.3%，故即使 L<sub>5</sub> 在 Group2 和 Group3 的區分錯誤數量明顯大於 L<sub>1</sub>，但是其整體區分錯誤率在測試集中分別為 L<sub>5</sub> :19.5%和 L<sub>1</sub>:12.4%，僅相差了 7%。

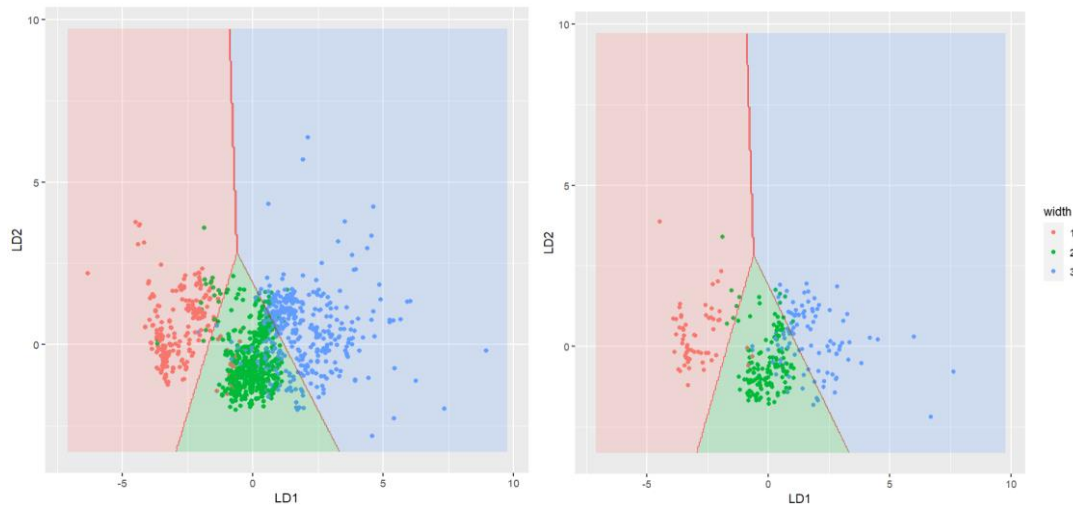
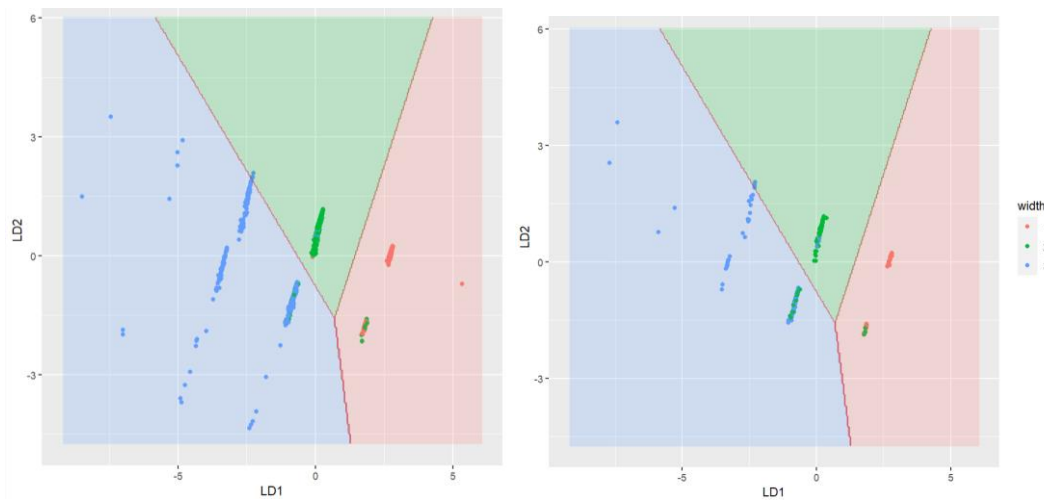
Figure.5 L<sub>1</sub> LDA boundaries plot (Training and Test)Figure.6 L<sub>5</sub> LDA boundaries plot (Training and Test)

圖 4 圖 5 中可以觀察到不管是在  $L_1$  或  $L_5$  的散佈圖中，Group2 和 Group3 均有大量重疊的地方，因此這可能為用直線進行區分的 LDA 沒辦法精確區分這部分之原因。

### ➤ Section 3-2 Mixture Discriminant Analysis

相較於 LDA，我們為了繪圖方便故在 MDA 選擇較多僅包含兩個自變數的模型，以下會對這五種模型進行分析比較：

$M_1: X_6 \sim .$

$M_2: X_6 \sim X_{15} + X_{16}$

$M_3: X_6 \sim X_{13} + X_{16}$

$M_4: X_6 \sim X_{13} + X_{15}$

$M_5: X_6 \sim X_{13} + X_{15} + X_{16} + X_{17}$

Table.17 MDA model

Model	Accuracy(Test)
$M_1: X_6 \sim .$	0.829
$M_2: X_6 \sim X_{15} + X_{16}$	0.762
$M_3: X_6 \sim X_{13} + X_{16}$	0.748
$M_4: X_6 \sim X_{13} + X_{15}$	0.560
$M_5: X_6 \sim X_{13} + X_{15} + X_{16} + X_{17}$	0.802

根據表 17 可以觀察到  $M_1$  將所有變數都放入模型使精確度達 82.9%， $M_5$  用了四個自變數使精確度達到 80.2%，而  $M_2$  因更是只用了兩個自變數就使精確度達 76.2%，並且其精確度與  $M_1$  和  $M_5$  沒有太大的差異，故以下將會對  $M_2$  進行分析。

Table.18 Confusion Matrix of  $M_2$ (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	220	19	2	55	8	0
2	27	464	76	4	120	29
3	0	43	341	0	10	72

從表 18 我們可以發現 MDA 與 LDA 面臨著相同的問題；即在區別 Group2 和 Group3 的能力相對於其他組合較弱。接下來會對挑選出之  $L_5$  和  $M_2$  在 Group2、Group3 間的區分錯誤率進行分析。



Table.19 Error rate between Group2 and Group3 in LDA and MDA

Model	Training set	Test set
L <sub>5</sub>	0.178	0.195
M <sub>2</sub>	0.129	0.169

表 19 中可以看到 M<sub>4</sub> 僅用了兩個變數其區分 Group2 和 Group3 的能力就較用了四個變數的 L<sub>5</sub> 強，且整體的區別能力並沒有特別大的差異，因此若要用較少的變數區分類別的話可以考慮選用 MDA。

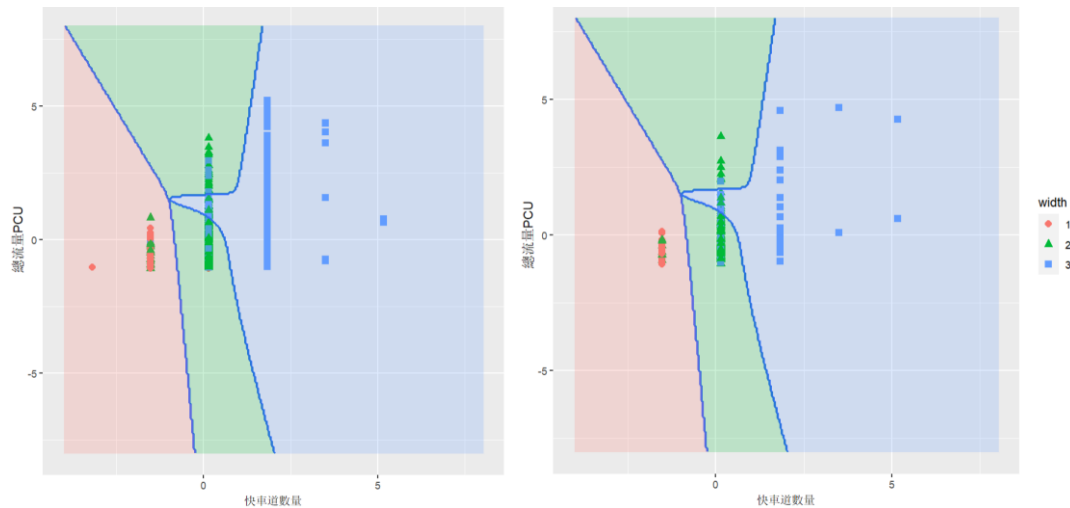


Figure.7 M<sub>4</sub> MDA boundaries plot(Training and Test)

根據圖 6 可以觀察到即使 Group2 和 Group3 有重疊的部分。相較於 LDA 僅能用直線進行區分，MDA 可以利用曲線將這兩種類別較精確的區分出來

### ➤ Section 3-3 Quadratic Discriminant Analysis

基於與挑選 MDA 相同的原因，我們將會對以下模型進行分析：

$$Q_1: X_6 \sim .$$

$$Q_2: X_6 \sim X_{15} + X_{16}$$

$$Q_3: X_6 \sim X_{13} + X_{16}$$

$$Q_4: X_6 \sim X_{13} + X_{15}$$

$$Q_5: X_6 \sim X_{13} + X_{15} + X_{16} + X_{17}$$

Table.20 QDA model

Model	Accuracy(Test)
Q1: $X_6 \sim .$	0.832
Q2: $X_6 \sim X_{15} + X_{16}$	0.724
Q3: $X_6 \sim X_{13} + X_{16}$	0.721
Q4: $X_6 \sim X_{13} + X_{15}$	0.477
Q5: $X_6 \sim X_{13} + X_{15} + X_{16} + X_{17}$	0.742

根據表 20 可以觀察到 Q1 將所有變數都放入模型使精確度達 83.2%，Q3 用了兩個自變數便使精確度與包含四個自變數的 Q5 沒有太大的差異，故將會 Q2 進行分析。

Table.21 Confusion Matrix of Q2 (Training and Test)

	Training			Test		
	1	2	3	1	2	3
1	226	25	2	56	8	0
2	19	472	261	3	124	65
3	2	29	156	0	6	36

由表 21 可以發現 QDA 在此資料中區分 Group2、Group3 的能力在三種方法中為最弱的，且其利用較少變數區分類別表現也最差。

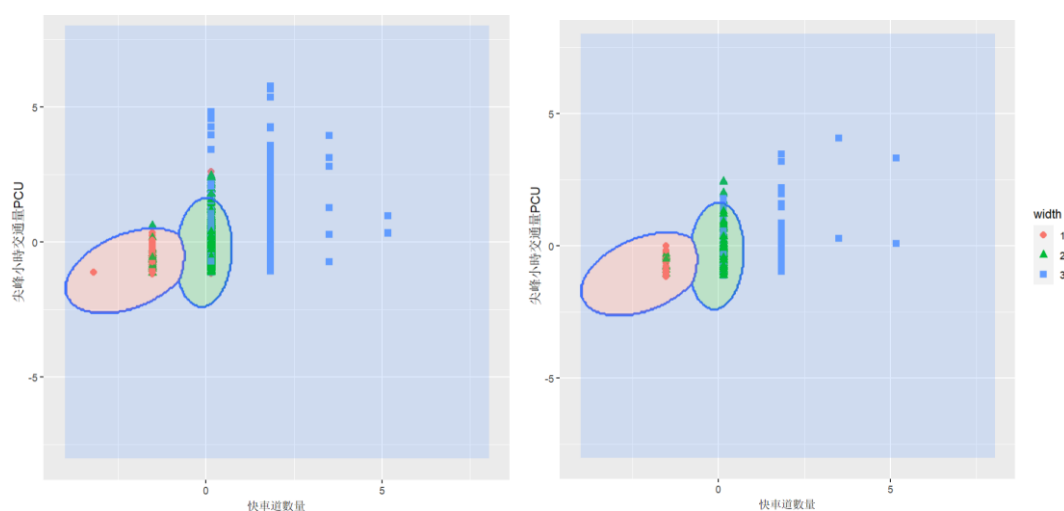


Figure.8 Q2 QDA boundaries plot(Training and Testing)

圖 7 中可以看到在此筆資料中 QDA 的篩選方式更像是利用不同的直條進行區分，而不像 MDA 在同一個直條中也能區分出 Group2 和 Group3。

## Part IV. 結論

根據 Part I 結果顯示丘陵區與平原區會較不易分清楚。從另一個角度來看，推測丘陵區與平原區連結性相對較高，而顯示了山嶺區獨特較易分出，建議高公局在做都市交通規劃時可能需要為了山嶺區做順應地勢的適度調整，可專門為了山嶺區投入經費及資源來進行建設。

從上面三種區分方法中我們可以觀察到不管是哪一種方法，其區分能力與 X15、X16 均有極大的相關，即快車道數量以及尖峰小時交通量為區分道路寬度之主要依據。另外我們也可以發現若要在這筆資料中對車道寬度進行區別分析的話，利用 LDA 或是 MDA 其區分 Group2、Group3 的效果會較 QDA 顯著。