

資料科學實作 作業二

國立清華大學動機系

郭博洋

大綱

- ▶ 資料來源介紹
- ▶ 重要參考資料
- ▶ 我想知道各地區使用機車的概況
- ▶ 概況分析方法
- ▶ 分析1-平均每星期使用天數之差異
- ▶ 分析2-平均一天使用次數之差異
- ▶ 結論

資料來源介紹

- ▶ 政府資料開放平台-機車使用狀況調查資料
 - ▶ 網址：<http://data.gov.tw/node/6216>
- ▶ 資料提供機構：交通部
- ▶ 資料量：約11400筆
- ▶ 發布時間：105年05月11日
- ▶ 以「問卷調查」方式而蒐集的資料
 - ▶ 優點：蒐集資料數多、方便進行定量研究
 - ▶ 缺點：無法保證資料的可靠性與真實度、多為不連續性的選擇題資料

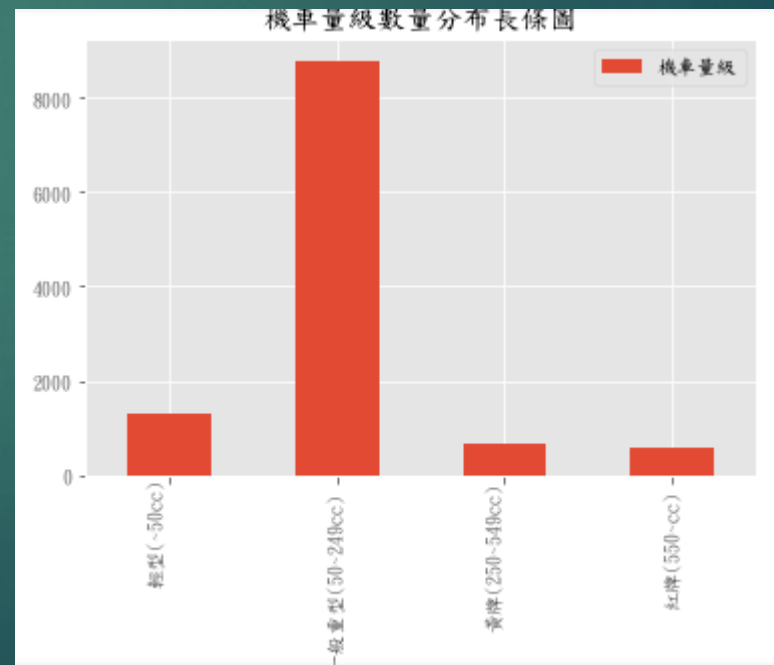
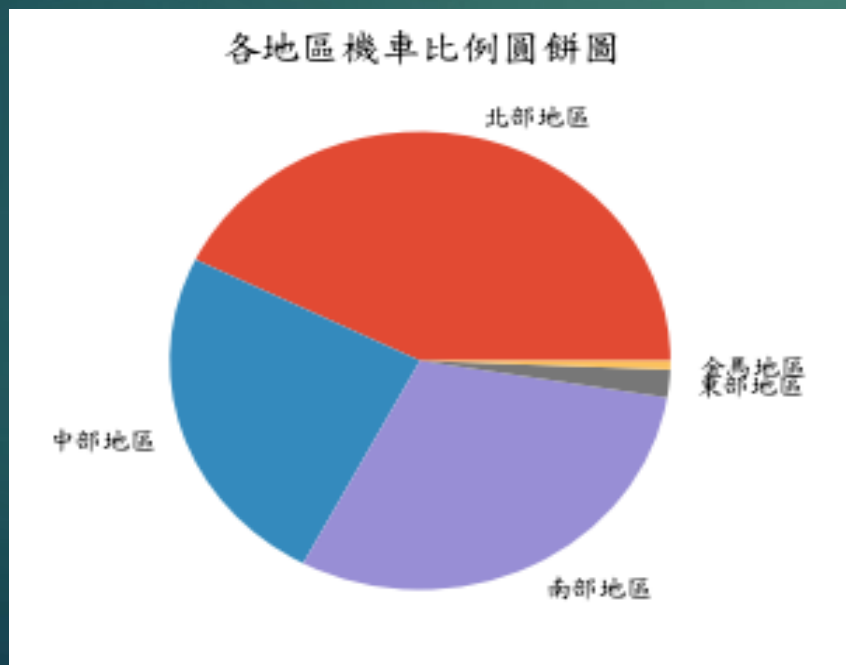
重要參考資料

- ▶ 機車使用地區
- ▶ 機車量級
- ▶ 每星期使用幾天
- ▶ 有使用的天數中，平均一天使用幾次
- ▶ 年齡
- ▶ 薪水
- ▶ 有使用的天數中，平均一天使用多少公里
- ▶ 擁有機車數
- ▶ 擁有汽車數

我想知道各地區使用機車的概況

► 2個分類+2種指標：

- 分類1—機車使用地區分布：北部、中部、南部、東部、金馬
- 分類2—使用機車量級：輕型(S)、一般重型(N)、黃牌(Y)、紅牌(R)
- 指標1—平均每星期使用天數
- 指標2—有使用的天數中，平均一天機車使用次數



概況分析方法

- ▶ 分析方法：樞紐分析
- ▶ 程式碼與結果：

備註：

每星期使用天數之數字含意

1:1天以下

2:1天

3:2天

4:3天

5:4天

6:5天

7:6天

8:7天

```
In [146]: df2 =df1
df2=df2.rename(columns={'F9_area':'地區','label':'機車量級'})

to_area={1:'北部地區',2:'中部地區',3:'南部地區',4:'東部地區',5:'金馬地區'}
df2['地區'] = df2['地區'].map(to_area)

# 在各縣市當中，不同CC數的騎士，平均每星期使用幾天
df2.pivot_table(values="B8A",index=['機車量級'],columns='地區',aggfunc="mean")

#df2[:20]
```

Out[146]:

	地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級						
N	5.736527	5.529859	5.878598	5.567901	5.000000	
R	2.294118	2.474164	2.504587	2.166667	4.000000	
S	4.923944	4.737194	5.167347	4.562500	5.222222	
Y	3.992754	4.012658	3.542636	4.095238	5.000000	

```
In [144]: # 在各縣市當中，不同CC數的騎士，一天平均使用幾次機車(來回算一次)
df2.pivot_table(values="B8B",index=['機車量級'],columns='地區',aggfunc="mean")
```

Out[144]:

	地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級						
N	2.020267	1.883822	2.146863	2.117284	2.285714	
R	1.375000	1.398176	1.330275	1.291667	0.000000	
S	1.808451	1.710468	1.861224	1.375000	2.222222	
Y	1.782609	1.726582	1.720930	2.000000	2.500000	

分析1-平均每星期使用天數之差異分析

- ▶ S、N作為一般使用之量級
 - ▶ 在各地區使用之頻率最高，除金馬地區
 - ▶ 使用天數為4~5天，可分析主要用途為通勤
 - ▶ 可視為台灣民眾用於通勤之主要使用量級
- ▶ Y、R作為非一般使用之量級
 - ▶ Y在各地區使用天數為2~3天
 - ▶ R在各地區使用天數為1~2天
 - ▶ 在台灣地區非主要通勤用途之機車量級

各地區使用各量級機車平均每星期使用天數統計表

地區 機車量級	中部地區	北部地區	南部地區	東部地區	金馬地區
N	5.736527	5.529859	5.878598	5.567901	5.000000
R	2.294118	2.474164	2.504587	2.166667	4.000000
S	4.923944	4.737194	5.167347	4.562500	5.222222
Y	3.992754	4.012658	3.542636	4.095238	5.000000

► 北部地區

- 主要通勤車N之使用頻率最少，可分析為其他交通選項的機會成本低：

分析原因1—其他大眾交通成本低且便利，如北部擁有全台最便捷之捷運系統，輔以其他大眾交通，選擇大眾交通的機會成本可以降低。

分析原因2—當地人對其他交通能負擔的預算高，如下表，北部的平均薪水最高。

各地區平均薪水統計表

薪水	
地區	
中部地區	29519.457337
北部地區	35252.974969
南部地區	29303.400174
東部地區	31547.085202
金馬地區	34428.571429

各地區使用各量級機車平均每星期使用天數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	5.736527	5.529859	5.878598	5.567901	5.000000
R	2.294118	2.474164	2.504587	2.166667	4.000000
S	4.923944	4.737194	5.167347	4.562500	5.222222
Y	3.992754	4.012658	3.542636	4.095238	5.000000

► 南部地區

- S、N、R之使用頻率為本島最高
- 可視為本島使用機車最頻繁之區域

► 金馬地區

- 只有70筆資料，數據誤差可能較大
- 各量級使用天數皆高

分析原因1—因較無其他交通選項作為通勤等用途，如右下表，Y量級只有金馬地區擁有平均小於2的機車數。

分析原因2—金馬地區使用機車之環境不會讓Y、R量級機車使用者造成不便，故可以較多機會使用。

各地區使用各量級機車平均每星期使用天數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	5.736527	5.529859	5.878598	5.567901	5.000000
R	2.294118	2.474164	2.504587	2.166667	4.000000
S	4.923944	4.737194	5.167347	4.562500	5.222222
Y	3.992754	4.012658	3.542636	4.095238	5.000000

各地區使用各量級機車平均擁有機車數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	2.081990	1.869164	2.164945	1.907407	1.767857
R	2.345588	2.319149	2.642202	2.666667	3.000000
S	2.050704	1.797327	1.957143	1.625000	1.555556
Y	2.202899	2.268354	2.534884	2.095238	1.500000

分析2-平均一天使用次數之差異分析

- ▶ N作為一般通勤主要使用之量級
 - ▶ 各地區平均一天使用次數在2上下，可分析為通勤加上非通勤之使用，而達到2次以上。
 - ▶ 北部低區使用次數較最低，與先前北部每星期使用次數最低為相同因素

各地區使用各量級機車平均一天使用次數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	2.020267	1.883822	2.146863	2.117284	2.285714
R	1.375000	1.398176	1.330275	1.291667	0.000000
S	1.808451	1.710468	1.861224	1.375000	2.222222
Y	1.782609	1.726582	1.720930	2.000000	2.500000

- ▶ R作為非一般使用之量級
 - ▶ 各地區平均一天使用次數靠近接近1
 - ▶ 已知為非通勤、運人為主之用途
 - ▶ 使用者擁有機車和汽車數高於其他量級
 - ▶ 因交通工具替代選項多，故分析使用R量級的使用者，多為無特定用途取向的使用，如出遊。
- ▶ 金馬地區
 - ▶ 只有70筆資料，數據誤差可能較大
 - ▶ 各量級使用天數皆高，原因如前言所示

各地區使用各量級機車平均一天使用次數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	2.020267	1.883822	2.146863	2.117284	2.285714
R	1.375000	1.398176	1.330275	1.291667	0.000000
S	1.808451	1.710468	1.861224	1.375000	2.222222
Y	1.782609	1.726582	1.720930	2.000000	2.500000

各地區使用各量級機車平均擁有機車數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	2.081990	1.869164	2.164945	1.907407	1.767857
R	2.345588	2.319149	2.642202	2.666667	3.000000
S	2.050704	1.797327	1.957143	1.625000	1.555556
Y	2.202899	2.268354	2.534884	2.095238	1.500000

各地區使用各量級機車平均擁有汽車數統計表

地區	中部地區	北部地區	南部地區	東部地區	金馬地區
機車量級					
N	0.909258	0.696254	0.785978	0.882716	0.678571
R	1.235294	1.221884	1.302752	1.416667	1.000000
S	0.954930	0.677060	0.702041	0.750000	0.777778
Y	1.246377	1.007595	1.015504	0.952381	0.500000

結論

- ▶ 雖然一開始只用2個分類與2種指標，但是從數據看出的現象，佐以薪水、擁有機車數、擁有汽車數等資料，更可清楚解釋現象的形成原因。
- ▶ 現在政府欲淘汰的二行程機車多以S量級為主，雖然S量級的數量已經不多，與Y和R量級數量差不多，但可看出仍有多數使用者以S量級機車作為主要通勤使用，這是淘汰過程的阻因之一。
- ▶ 選擇Y、R量級的使用者，與選擇其他交通工具的機會成本有關。
- ▶ 而機會成本來自於環境、環境的大眾交通工具便捷性、自身薪水、自身擁有的交通工具選項。

- 若不考慮環境因素，將使用者自身的資源：擁有機車數、擁有汽車數、機車量級、薪水混合分析其相關性，可以發現雖為低度相關，但是還是擁有較為突出的相關性，且考慮問卷調查資料之不連續性會降低相關性，此相關性仍是有參考價值的。
- 其中最為特別的是擁有機車數與汽車數之間的相關性最高，我猜想與使用者周遭環境是否有足夠的停車土地有關，這部分可以做更深入之調查。

各項資料之相關係數表

	平均每日機車使用次數	年齡	擁有機車數	擁有汽車數	機車量級	薪水
平均每日機車使用次數	1.000000	-0.032005	0.013330	-0.020626	-0.109481	-0.035829
年齡	-0.032005	1.000000	0.061974	0.046542	-0.034412	-0.061140
擁有機車數	0.013330	0.061974	1.000000	0.325580	0.084420	0.046548
擁有汽車數	-0.020626	0.046542	0.325580	1.000000	0.146683	0.145300
機車量級	-0.109481	-0.034412	0.084420	0.146683	1.000000	0.207761
薪水	-0.035829	-0.061140	0.046548	0.145300	0.207761	1.000000