

統計期末報告

【組員】

B06701129 毛譯萱 B06701137 林祐良 B06701130 江彥武

B06701158 楊舒涵 B06302344 陳奕心 B06701141 洪偉蘋 B06701101 薛祥呈

【研究主題】

台北市 Youbike 的營運狀況與各因子之間的影響

【研究動機】

大台北地區的 YouBike 使用率是相當高的，可知此交通工具在台北人以及觀光客的生活裡是扮演著不可或缺且相當便利的角色，然而，在眾多大眾交通中人們選擇交通方式的變因有許多種，因此我們想用數據來分析各種可能影響到 YouBike 使用率的原因，以此分析各種情況下人們使用 YouBike 的狀況。

【研究內容】

1. 透過分析 uibike 在 2015 年 4 月終止前 30 分鐘免費的優惠前後一年的每月使用次數，進行 one way anova 分析，探討是否因為減少了這項優惠，而造成使用次數下降。
2. 透過觀察每月降雨天數，以及月均溫這兩項變因，是否 YouBike 的每月使用次數會因為這兩項變因有所影響，我們是以月的角度來進行探討，所以是否造成影響，而兩項的因素又為何，將是我們進行探討的重點。
3. 用線性回歸分析 2017 年 4~8 月及 2018 年 10~2019 年 2 月的 YouBike 借用次數。選擇這兩個時段進行調查是因為 2017 年 4 月 obike 進入市場，2018 年 10 月 obike 大量裁撤，從 8000 台撤到 1800 台。我們想觀察 obike 的進駐與退出時 Youbike 的使用狀況，來分析究竟 obike 是不是影響 YouBike 營運的因子。
4. YouBike 公司在 2018 年 4 月 1 日起推出捷運轉乘優惠，在一個小時內轉乘可以折五元，我們想探討推出此優惠後是否有增加 YouBike 的使用次數，檢測此政策是否有帶來實質的效益。
5. 我們會用 ANOVA 分析 YouBike 借用次數多的站，與捷運站進出量大的站，是否存在相互影響的關係，又關係的密切程度為如何，看腳踏車數量是否供不應求，又是否需要增加腳踏車的數量。用線性回歸分析捷運搭乘次數和 YouBike 借用次數(尤其注重捷運與 YouBike 優惠方案運行之後)，目前以網站上的資料可能只能做到每月捷運搭乘次數與每月 YouBike 使用次數的比較，因為 YouBike 借用次數的資料不夠多(無各站的統計資料，無較久以前的資料)，我們往後會試圖聯繫 YouBike 公司詢問較細的資料。還有更進一步探討，YouBike 在 2018/4/1 日起推出捷運轉乘優惠，在一個小時內轉乘可以折五元，我們想探討推出此優惠後是否有增加 YouBike 與捷運的使用次數，檢測此政策是否有帶來

實質的效益。

6.利用時間序列分析預估未來的借用次數。用一階自我迴歸預測下一期的數據，再利用三指數平滑法，考慮了季節性因素，預測未來幾個月的 youbike 借用情形。

預期：

1. 預期台北市增加前三十分鐘付款會減少民眾的使用率
2. 預期下雨天或天氣過熱或過於寒冷會造成民眾使用率下降
3. 我們認為 obike 的進駐會使 YouBike 的使用人數下降，因 obike 的借用地點較為多元
4. 我們假設捷運大站 (ex:公館，市政府等等...) 的借用頻率會相對較高，還有我們覺得政府推出的捷運 YouBike 轉乘優惠也會同時使 YouBike 借用次數增加，提升民眾的使用率。

【研究結果】

1. 用 one way anova 分析台北市「前半小時免費」與 youbike 借用次數的關係

月份/時期	收費前	收費後
4	1905000	1662426
5	1810000	1554657
6	1830000	1534574
7	1970000	1584211
8	2130000	1345017
9	1910000	1483945
10	2250000	1611187
11	2150000	1601139
12	1990000	1490150
1	2270149	1100895
2	1880072	1024025
3	2065266	1242302

單因子變異數分析

摘要

組	個數	總和	平均	變異數
欄 1	12	24160487	2013374	2.5E+10
欄 2	12	17234528	1436211	4.44E+10

ANOVA

變源	SS	自由度	MS	F	P-值	臨界值
組間	2E+12	1	2E+12	57.66448	1.39E-07	4.30095
組內	7.63E+11	22	3.47E+10			
總和	2.76E+12	23				

拒絕 H0

前 30 分鐘收費與否有影響騎乘次數

結論：根據我們做的實驗結果來推斷，前 30 分鐘免費與否的確影響每個月的騎乘次數，有其實剛施行新政策的前幾個月能夠看到影響甚鉅，但過渡期之後，UBIKE 的騎乘次數又恢復穩定上漲。

而前 30 分鐘免費所減少的使用人次大部分來自於 30 分鐘內的短程使用者，因此對於 Youbike 來說，在度過過渡期之後，也不算壞事。

2.用 two factor anova 分析下雨天數、氣溫與 youbike 借用次數的關係

月均溫 20 以下為低 20-25 為中 25 以上為高

降雨頻率以降雨日數為標準 15 天以上為多

氣溫/降雨頻率	少	多
低	1866438	1100895
	1489878	1024025
	1626040	1242302
	1763642	1664401
	1725624	1993811
中	2221447	1853390
	1640296	2386920
	1123612	1483257
	2014758	1626040
	1900724	2380747
高	2128110	1550056
	1721659	1535254
	1685100	2014758

1870294 1617197
2074281 1719024

雙因子變異數分析：重複試驗

摘要	低	中	高	總和
少				
個數	5	5	5	15
總和	8471622	8900837	9479444	26851903
平均	1694324	1780167	1895889	1790127
變異數	2.05E+10	1.79E+11	4.03E+10	7.57E+10

多				
個數	5	5	5	15
總和	7025434	9730354	8436289	25192077
平均	1405087	1946071	1687258	1679472
變異數	1.7E+11	1.77E+11	3.88E+10	1.62E+11

總和				
個數	10	10	10	
總和	15497056	18631191	17915733	
平均	1549706	1863119	1791573	
變異數	1.08E+11	1.66E+11	4.72E+10	

ANOVA

變源	SS	自由度	MS	F	P-值	臨界值
樣本	9.18E+10	1	9.18E+10	0.881725	0.357085	4.259677
欄	5.39E+11	2	2.7E+11	2.589895	0.09584	3.402826
交互作用	2.95E+11	2	1.47E+11	1.415896	0.262264	3.402826
組內	2.5E+12	24	1.04E+11			
總和	3.43E+12	29				

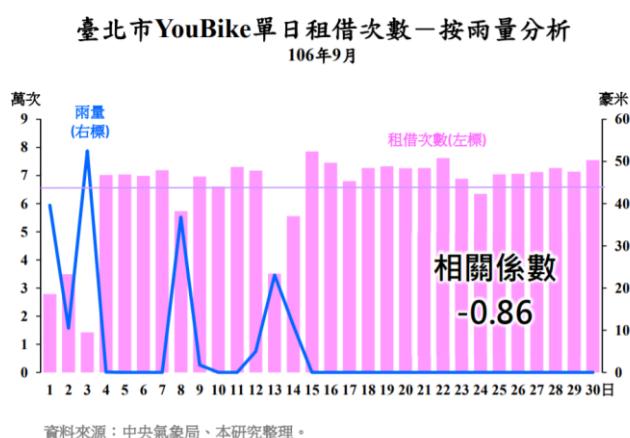
不拒絕 H0

代表以我們這次的分類，氣溫分類和降雨天數作為兩個因子的假設之下，這兩個因子對於騎乘次數是沒有影響的

但這似乎有一點點違背直覺，至少下雨這個因素似乎應該會影響騎乘次數。

我們認為有以下原因

1. 使用每月降雨次數作為因子規模太大，同時 15 天作為標準也有些模糊。除了少數幾個月，大部分月份降雨天數幾乎都為 10-20 天，而每次降雨的時數以及雨量多寡其實似乎是更精準的因子，但不見得會影響騎乘次數(下雨的時段以及地區)
2. 以月均溫來看其實每個月分的溫度都還算適合出外活動，扣除少數寒流幾天，其實對於如此大樣本的次數似乎也影響不大
3. 根據台北市統計局的研究結果顯示，如果按照每天以及降雨量來看的話，就能夠清楚發現，以一天的角度來說，下雨非常巨大的影響到每日騎乘次數，但如果以月(及降雨天數)的角度來看，效果就並不顯著。



3.使用各式檢定及 ANOVA 來分析 2017 年 4~8 月、2018 年 10~2019 年 2 月及其前後的 ubike 借用次數。

選擇這兩個時段進行調查是因為 2017 年 4 月 obike 進入市場，2018 年 10 月 obike 大量裁撤，從 8000 台撤到 1800 台。我們想觀察 obike 的進駐與退出時 ubike 的使用狀況，來分析究竟 obike 是不是影響 ubike 營運的因子。

ubike 租借次數 (車次)

進入市場前		進入市場後		減少前		減少後	
105 年 11 月	1640296	106 年 4 月	1900724	107 年 5 月	2653719	107 年 10 月	2386920
105 年 12 月	1866438	106 年 5 月	2074281	107 年 6 月	2383194	107 年 11 月	2380747
106 年 1 月	1626040	106 年 6 月	1664401	107 年 7 月	2521762	107 年 12 月	2219347
106 年 2 月	1489878	106 年 7 月	2014758	107 年 8 月	2390626	108 年 1 月	2334522
106 年 3 月	1719024	106 年 8 月	2128110	107 年 9 月	2182678	108 年 2 月	1928774

Obike 進入市場前後分析

F 檢定：兩個常態母體變異數的檢定

	變數 1	變數 2
平均數	1668335	1956455
變異數	1.91E+10	3.38E+10
觀察值個數	5	5
自由度	4	4
F	0.564089	
P(F<=f) 單尾	0.296387	
臨界值：單尾	0.156538	

透過 F 檢定，我們可以知道兩組資料的變異數不一樣

然後再透過 ANOVA 知道兩組資料的平均數也不一樣

ANOVA

變源	SS	自由度	MS	F	P-值	臨界值
組間	2.08E+11	1	2.08E+11	7.854181	0.023108	5.317655
組內	2.11E+11	8	2.64E+10			
總和	4.19E+11	9				

t 檢定：兩個母體平均數差的檢定，
假設變異數不相等

	變數 1	變數 2
平均數	1668335	1956455
變異數	1.91E+10	3.38E+10
觀察值個數	5	5
假設的均數差	0	
自由度	7	
t 統計	-2.80253	
P(T<=t) 單尾	0.013214	
臨界值：單尾	1.894579	
P(T<=t) 雙尾	0.026428	
臨界值：雙尾	2.364624	

最後我們使用變異數不相等的 t 檢定來檢視兩組資料的平均數差異，得到了以下的數字
T 統計量：-2.80253
 $|T| > t \text{ critical} = 1.894579$

結論：進入市場前的平均數減進入市場後的平均數小於 0，由以上分析可得知 Obike 進入市場後 youbike 的使用次數不僅沒有下降,甚至比以往增加,因此得知 Obike 並沒有對 youbike 造成使用率上的負面影響

Obike 大量撤出市場前後分析

一樣先透過 F 檢定，我們可以知道兩組資料的變異數不一樣

F 檢定：兩個常態母體變異數的檢定

	變數 1	變數 2
平均數	2426396	2250062
變異數	3.08E+10	3.68E+10
觀察值個數	5	5
自由度	4	4
F	0.83824	
P(F<=f) 單尾	0.434172	
臨界值：單尾	0.156538	

ANOVA

變源	SS	自由度	MS	F	P-值	臨界值
組間	7.77E+10	1	7.77E+10	2.299599	0.167879	5.317655
組內	2.7E+11	8	3.38E+10			
總和	3.48E+11	9				

t 檢定：兩個母體平均數差的檢定，假設變異數不相等

再透過 ANOVA 和 T 檢定，我們可以觀察到兩組資料的平均數在 $\alpha=0.05$ 的情況下是相等的

	變數 1	變數 2
平均數	2426396	2250062
變異數	3.08E+10	3.68E+10
觀察值個數	5	5
假設的均數差	0	
自由度	8	

t 統計 1.516443

P(T<=t) 單尾 0.08394

臨界值：單尾 1.859548

P(T<=t) 雙尾 0.167879

臨界值：雙尾 2.306004

結論：obike 大量撤出市場前後，對於 ubike 的租借次數並無明顯的影響

總結論：經過這些統計分析後，結果相當出乎我們意料之外，本來以為會是強大對手的 obike，在進入市場後會產生一定的競爭關係，造成 ubike 租借次數下降，沒想到在數字上 obike 不論在進入市場時或是大量退出時，對於 ubike 的影響力是非常小的。

4.YouBike 公司在 2018 年 4 月 1 日起推出捷運轉乘優惠，在一個小時內轉乘可以折五元，我們想探討推出此優惠後是否有增加 YouBike 的使用次數，檢測此政策是否有帶來實質的效益。

附表為實施前一年與實施後一年的租借次數統計：

(由於資料無法細分成每日借用次數或其他小單位，而且考量到腳踏車借用可能會有季節性波動，因此無法僅分析實施前後短暫時間內造成的變化，這點敬請見諒)

	實施前借用次數		實施後借用次數
Apr-17	1900724	Apr-18	2221447
May-17	2074281	May-18	2653719
Jun-17	1664401	Jun-18	2383194
Jul-17	2014758	Jul-18	2521762
Aug-17	2128110	Aug-18	2390626
Sep-17	1993801	Sep-18	2182678
Oct-17	1725624	Oct-18	2386920
Nov-17	1853390	Nov-18	2380747
Dec-17	1763642	Dec-18	2219347
Jan-18	1596746	Jan-19	2334522
Feb-18	1123612	Feb-19	1928774
Mar-18	2191938	Mar-19	2229157

建立假設: $H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 < 0$

先用 F-test 檢驗變異數是否相等:

F 檢定：兩個常態母體變異數的檢定		
	變數 1	變數 2
平均數	1835918.9	2319407.75
變異數	8.561E+10	33585839865
觀察值個數	12	12
自由度	11	11
F	2.5490963	
P(F<=f) 單尾	0.0679597	
臨界值：單尾	2.8179305	

p-value=0.0679597，兩母體變異數相等

⇒ t 檢定：兩個母體平均數差的檢定，假設變異數相等

t 檢定：兩個母體平均數差的檢定，假設變異數相等		
	變數 1	變數 2
平均數	1835919	2319408
變異數	8.56E+10	3.36E+10
觀察值個數	12	12
Pooled 變異數	5.96E+10	
假設的均數差	0	
自由度	22	
t 統計	-4.8511	
P(T<=t) 單尾	3.78E-05	
臨界值：單尾	1.717144	
P(T<=t) 雙尾	7.56E-05	
臨界值：雙尾	2.073873	

單尾 p-value= 0.0000378017778236009

我們有足夠的理由拒絕 H_0 ，說明 $\mu_1 - \mu_2 < 0$ ，實施一個小時內轉乘折五元的優惠可以增加 YouBike 借用次數，所以說明此策略獲得台北市民親賴，可以有效增加租借次數。

接下來要探討 YouBike 租借次數與捷運站進出量是否有關係。

5.使用回歸分析 YouBike 借用次數多的站，與捷運站進出量大的站，是否存在相互影響的關係，又關係的密切程度為如何，看腳踏車數量是否供不應求，又是否需要增加腳踏車的數量。**如果能從迴歸模型中獲得適當的回歸直線，在以後建置新捷運站時，能觀察附近居民數與人流流動次數，推算該捷運站附近應該放置多少輛 YouBike。**

2019 年 4 月份 YouBike 借用次數與捷運站進出量

借用次數排名		YouBike借用次數	捷運站進出量
1	捷運公館站	37915	857140
2	捷運芝山站	29861	703841
4	捷運市政府站	26511	1798935
7	捷運科技大樓站	23540	417860
8	捷運大安站	23289	736769
9	捷運台北101/世貿站	22145	979592
10	捷運國父紀念館	20503	696856
11	捷運東門站	19863	576128
12	捷運圓山站	18797	915175
13	捷運松山站	17248	707591
14	捷運永春站	17150	575487
15	捷運劍潭站	17121	1016903
16	捷運松江南京站	16507	1062949
18	捷運象山站	15669	363221
19	捷運信義安和站	15334	541622
20	捷運中山國中站	15304	416222
21	捷運行天宮站	15163	833893
22	捷運大安森林公園站	15087	271736
23	捷運北投站	15067	576128
26	捷運六張犁站	14585	329326
27	捷運台電大樓站	14579	537030
29	捷運忠孝復興站	14431	1400305

22 筆資料

• 從 YouBike 官方網站取得借用次數前 30 高的 YouBike 站，與台北捷運官方網站所提供的月搭乘次數做比對，其中我們只列出捷運站附近的 YouBike 站，因為這樣可以有比較準確的比較。

建立假設 H0: YouBike 借用次數與捷運站進出量沒有關係

H1: YouBike 借用次數與捷運站進出量有關係

計算檢定統計量

迴歸統計

R 的倍
數 0.327787

R 平方 0.107444

調整的
R 平方 0.060467

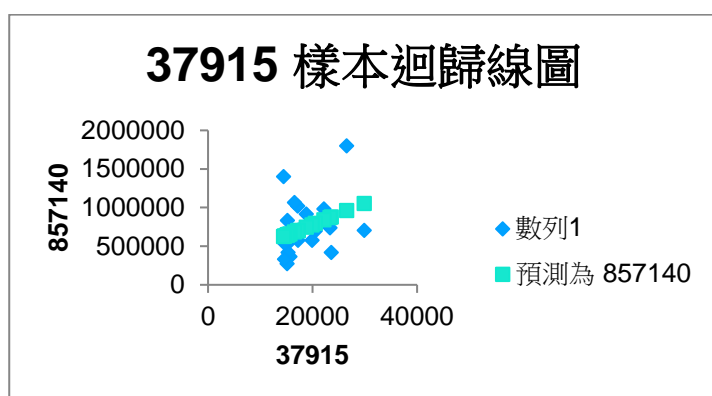
標準誤 358701.1

觀察值
個數 21

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	2.94E+11	2.94E+11	2.287181	0.146902
殘差	19	2.44E+12	1.29E+11		
總和	20	2.74E+12			

	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	222333.2	348600.7	0.637788	0.531224	-507296	951962.9	-507296	951962.9
37915	27.82324	18.39744	1.512343	0.146902	-10.683	66.32952	-10.683	66.32952



◆標準誤相當大，說明此回歸模型 X 與 Y 之間關係薄弱

經 excel 計算得出 p-value=0.146902，因此在 $\alpha=0.05$ 下，我們沒有足夠的理由拒絕 H_0 ，所以我們認為前 30 高的 YouBike 租借站與捷運站進出量沒有強烈的關係。

推測的造成此現象的原因:

- 1) 其中 4 捷運市政府站、15 捷運劍潭站、29 捷運忠孝復興站月進出量過大，在回歸分析中屬於離群值，造成殘差受到劇烈影響，又這三站附近都有北市著名的觀光商圈，許多觀光客出捷運站後便會步行至商圈，只有低比例的乘客會選擇轉乘 YouBike 至他處，可能是造成資料和回歸線較為偏差的原因。

(然而扣除此三站與同樣是步行距離內有商圈的其他兩站後，R square 有增加至 0.22214298，但仍不足說明借用次數與捷運站進出量有關係)

- 2) 影響租借 YouBike 的原因還有很多，可能是捷運站附近有沒有學校，住戶的生活習慣，轉乘公車的方便性等，需分開探討各面向影響的原因。
- 3) 降雨造成 YouBike 租借次數的波動，每天的租借次數波動大，無法一概而論，而雨量造成租借次數的影響，會在之後的部分提到。

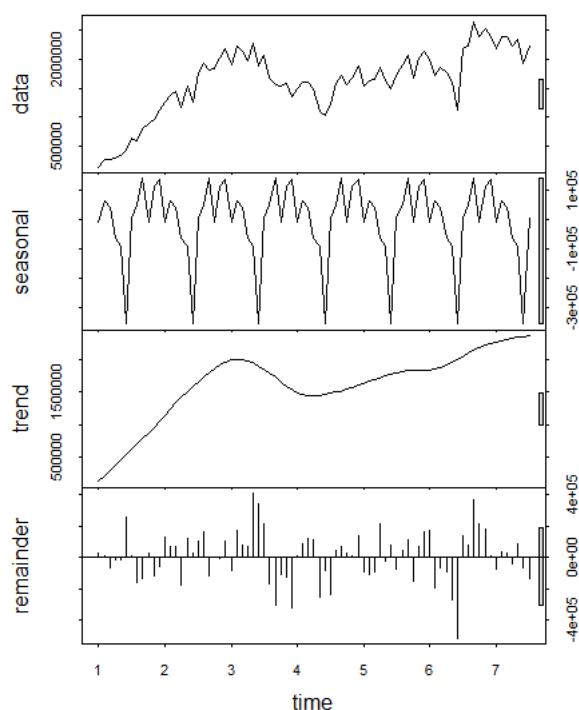
6. 嘗試以時間分析與預測未來數據:

(1) 我們嘗試利用一階自我回歸預測下一期的結果，得出:

預測下一期腳踏車借用次數 = $1.14299 \times \text{本期腳踏車借用次數} - 0.66804$

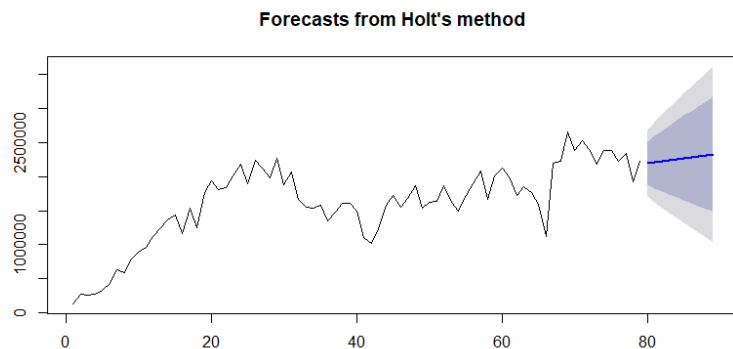
將 2019 年 4 月的 2229157 帶入，我們預估 2019 年 5 月的結果為 2547901。但因數據具有的 seasonal effect 和 residuals，因此此方法的預測效果有限。

(2) 觀察統計數據，我們發現 2 月、6 月、9 月的腳踏車借用次數偏低，尤其是 2 月，但 1 月、3 月卻無此現象，因此藉由課堂所學的時間序列分析，我們試著將 youbike 的使用次數分成 trend + seasonal effect(以月分區分，而非季節) + residuals，如下圖：

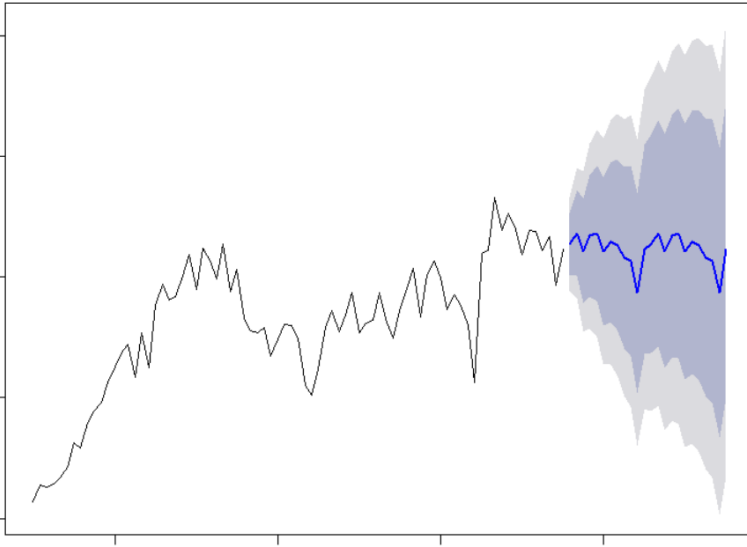


(3) 單純的指數平滑法因為忽略季節性及趨勢的因素，因此也不適合用來預測未來的結果，因此我們決定利用 Holt-Winters Forecast Method 三指數平滑法，同時考量季節性及趨勢，來預測未來結果。

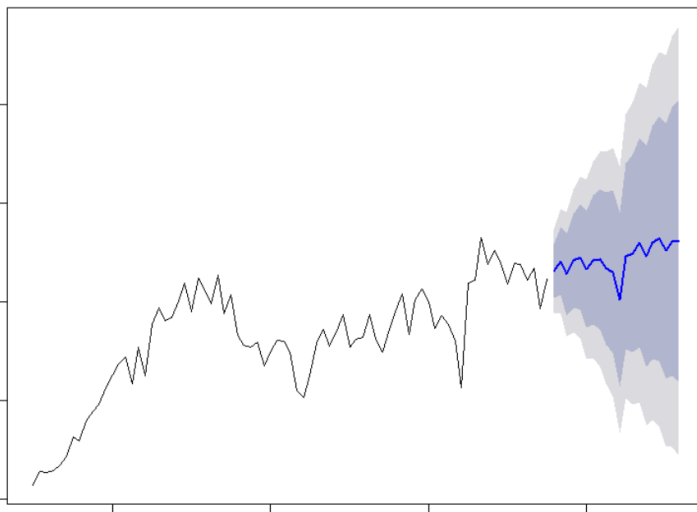
利用(2)的分析結果，我們先考量 trend 對未來數據的影響，如下圖：



再考量 seasonal effect，採用加法模型預測法，預測未來幾個月的 youbike 借用情形，如下圖：



將兩者合併，我們可以得到更準確的估計，如下圖：



結論: 有了這樣的預估區間，youbike 營運者可以根據此結果做出一些決策，例如借用次數多的月份，需要多準備一些 youbike、分配多一點人力補充 youbike，而借用次數較少的月份就可以收回一些 youbike 做保養，也同時才能增加 youbike 的流通性，避免太多使用的 youbike 卡在站上，占掉 youbike 的歸還車位。