

時序期末報告

路殺事件與汽機車數量的相關性

第七組 統計 109 陳冠頻 H24051304

目錄

一、研究動機	2
二、資料來源與介紹	2
三、資料分析	3
(一)路殺資料建模流程及結果	3
1. 模型判斷	3
2. 參數估計	8
3. 殘差分析	9
4. 模型統整	10
(二)汽機車數量資料建模流程及結果	11
1. 模型判斷	11
2. 參數估計	14
3. 殘差分析	15
4. 模型統整	17
(三)相關性分析	18
四、結論	19
五、參考資料	20
六、課堂心得與建議	21

一、研究動機

路殺(Roadkill)是指野生動物在道路上因為汽機車的撞擊或輾壓致死的現象。隨著都市的發展，有越來越多的地方建設並且開通越來越多的道路，這樣的發展可能因此破壞了這些野生動物原本的棲息地，讓牠們居無定所，使路殺事件發生的次數變的越來越頻繁，而這類型的事故並不只有涉及道路駕駛人的安全問題，很多原本就屬於保育類的動物更是因此而瀕臨絕種，因此現今很多國家都開始著手想要解決這樣的問題，像是美國就建造了動物天橋，而台灣近年來也開始制定一些公路動物友善的措施，期盼能減少路殺事件的發生。

而這次選擇這個主題是因為路殺所造成的生態議題，在近期有很多宣導與防範的新聞與活動，因此開始逐漸認識並了解這個議題，從很多的資料來看車輛的增加可能是造成路殺事故發生的一大因素，因此這次的主題想針對近年來汽機車數量的變化是否影響路殺事件的發生進行相關性分析，並預期會有一定程度的正相關。

二、資料來源與介紹

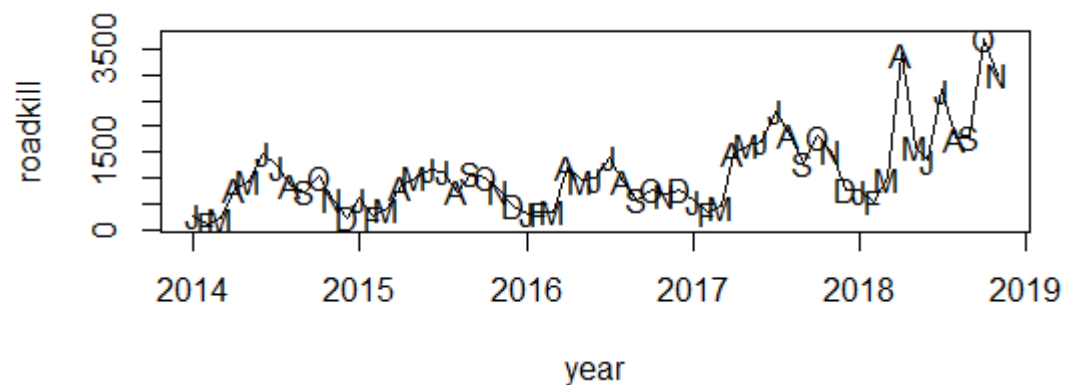
- 路殺(Roadkill)
資料來源：台灣動物路死觀察網
資料範圍：2014 年 1 月至 2018 年 11 月
資料數量：59 筆
時間間隔：以月為單位
- 汽機車數量
資料來源：交通部公路總局 統計查詢網
資料範圍：2014 年 1 月至 2018 年 11 月
資料數量：59 筆
時間間隔：以月為單位

三、資料分析

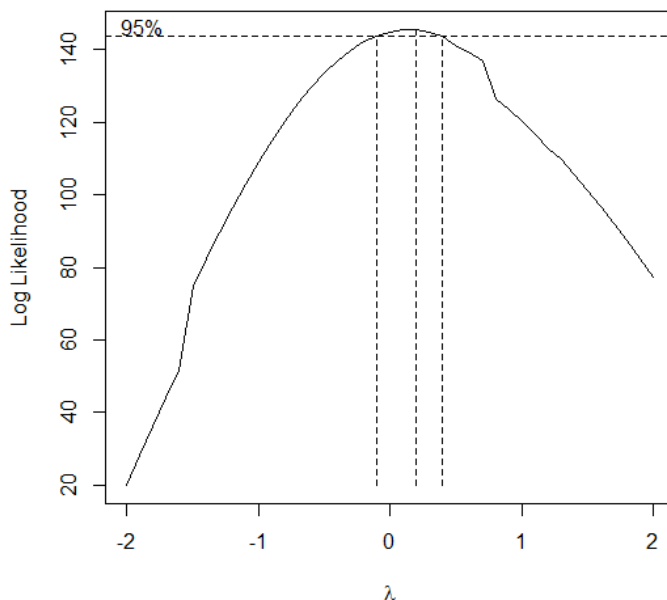
(一)路殺資料建模流程與結果

1. 模型判斷

此筆時間序列為全台灣從 2014 年 1 月至 2018 年 11 月路殺事件的月資料。



首先畫出此筆資料的時間序列圖，由圖可以看出此筆資料有一點規律的波動，所以猜測可能有季節性的趨勢，且在後半段有逐漸上升、變動幅度變大的趨勢。



接著對資料進行 BoxCox 轉換的分析，因為 $\lambda=0$ 在 95%信賴區間內，因此把整筆資料做對數轉換。

Dicky-Fuller test

檢定量	-4.7217
Lag order	3
P-value	<0.01

KPSS test

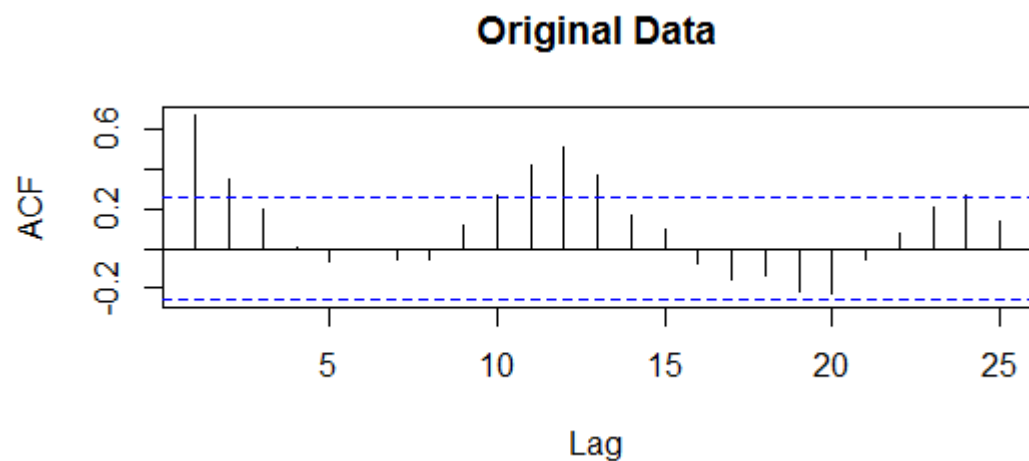
檢定量	1.1848
Lag order	1
P-value	<0.01

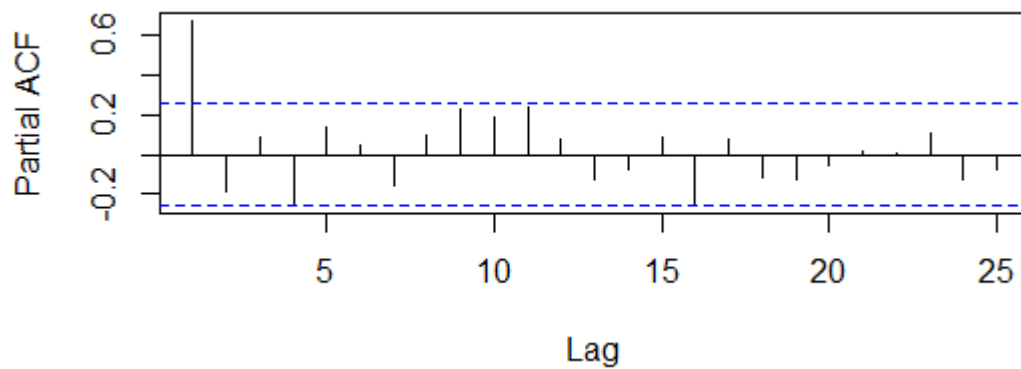
為了確定資料是否需要差分，因此分別做 Dicky-Fuller 以及 KPSS 檢定，而得到的結論分別是

(1)Dicky-Fuller 檢定的 $p\text{-value} < 0.05$ ，因此拒絕需要做差分的假設。

(2)KPSS 檢定中的 $p\text{-value} < 0.05$ ，因此拒絕不需要做差分的假設。由於兩個檢定出的結果不同，所以選擇分別做出差分與不差分的分析。

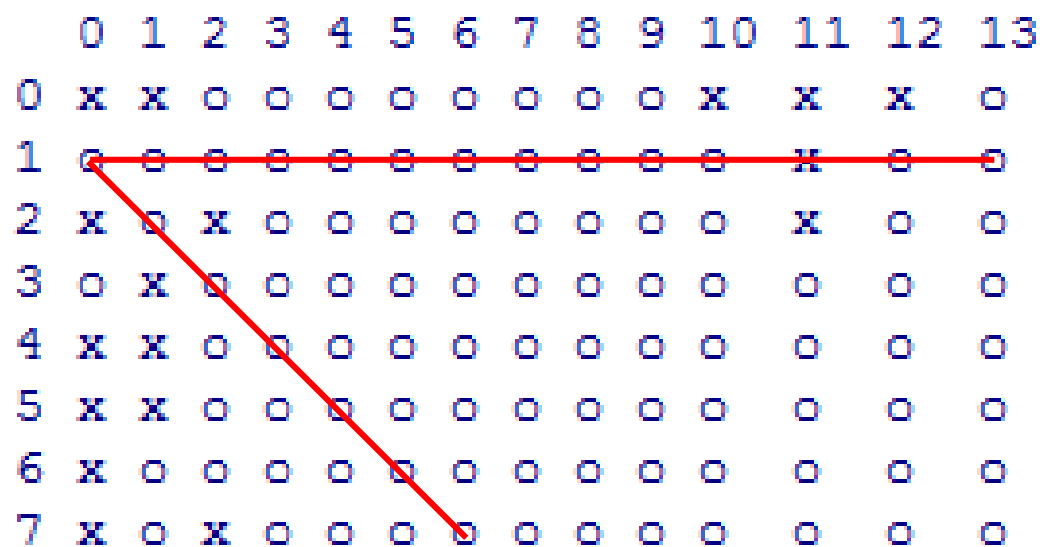
i. 未差分





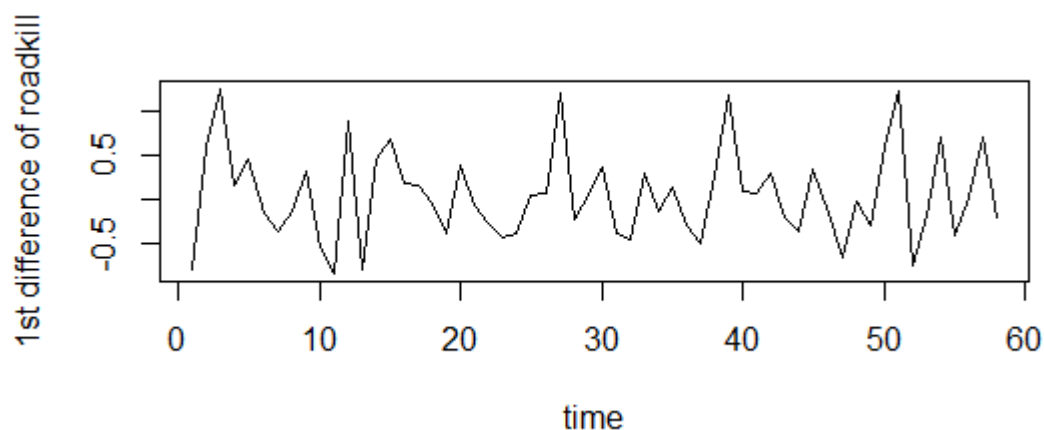
在 ACF 中每 12 步會有顯著的步數出現，而且非截斷，因此判斷模型可能為 $SARIMA(1,0,0) \times (1,0,0)_{12}$ ；而 PACF 的圖中可以明顯觀察到一步截斷的現象，因此判斷可能為 $AR(1)$ 。

AR/MA

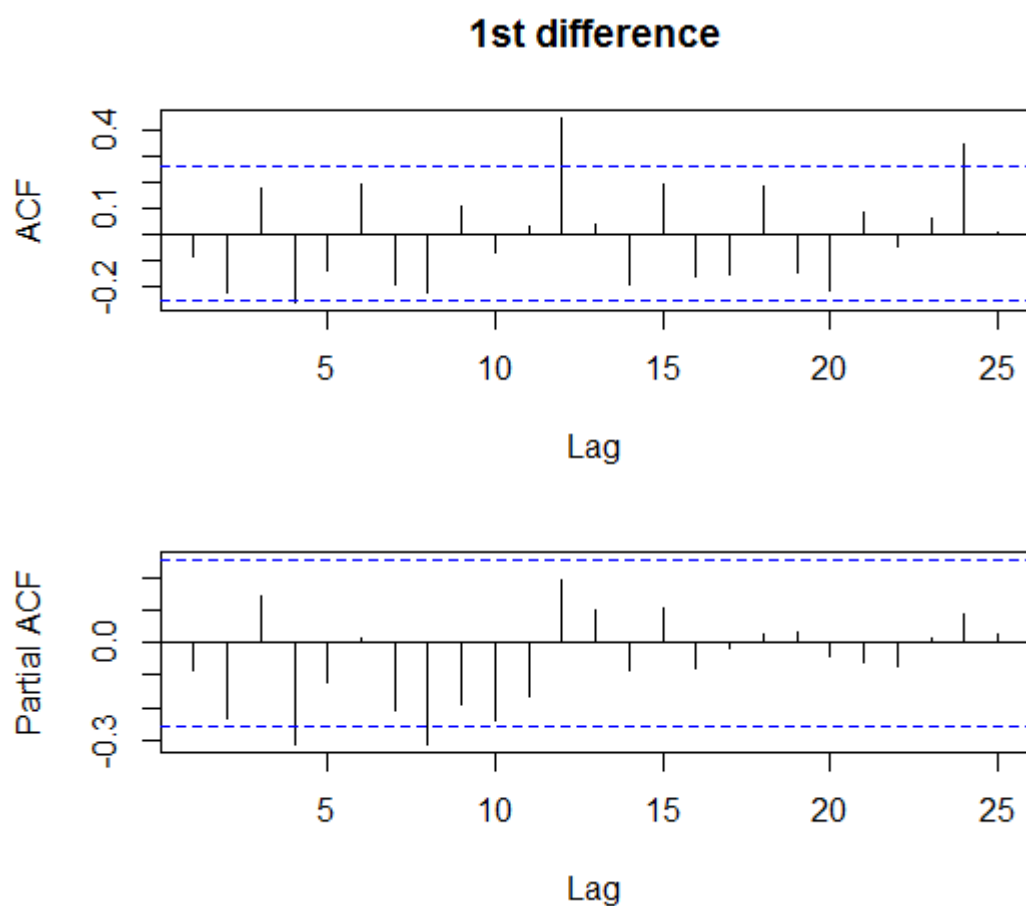


從 EACF 的圖中判斷可能的模型為 $AR(1)$ 。

ii. 一階差分



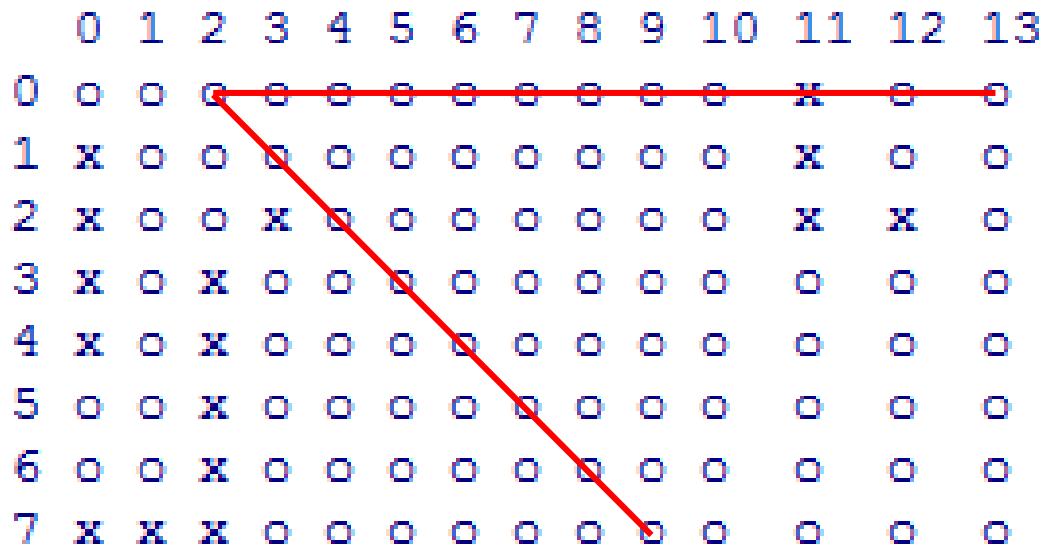
差分後的時間序列圖與原始數據相比變得比較平穩，不過看起來還是具有季節的趨勢。



進行差分後的 ACF 在第 12 以及 24 步時超出信賴界，因為步數過大不適合配置模型，不過從超出信賴界得這兩步中可以推知可能有季

節性的趨勢；而 PACF 在第 4 步以及第 8 步超出信賴界，但第 8 步的步數有點過大所以暫不考慮，因此選定 ARIMA(4, 1, 0) 為候選模型。

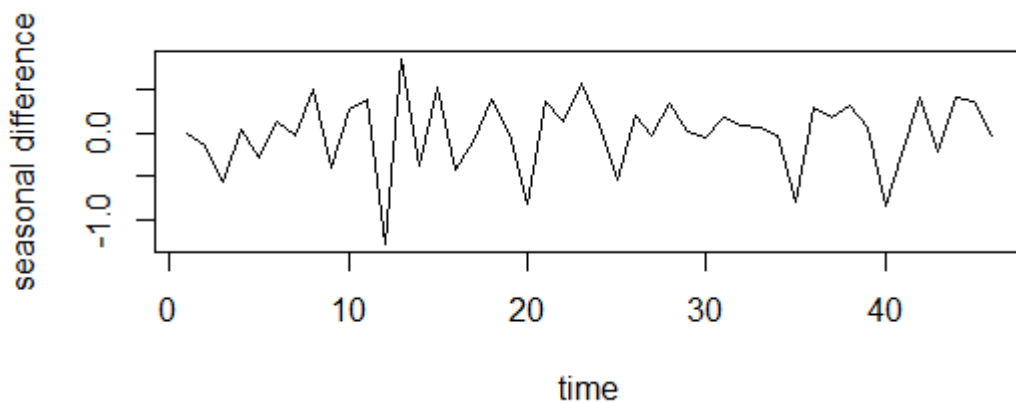
AR/MA



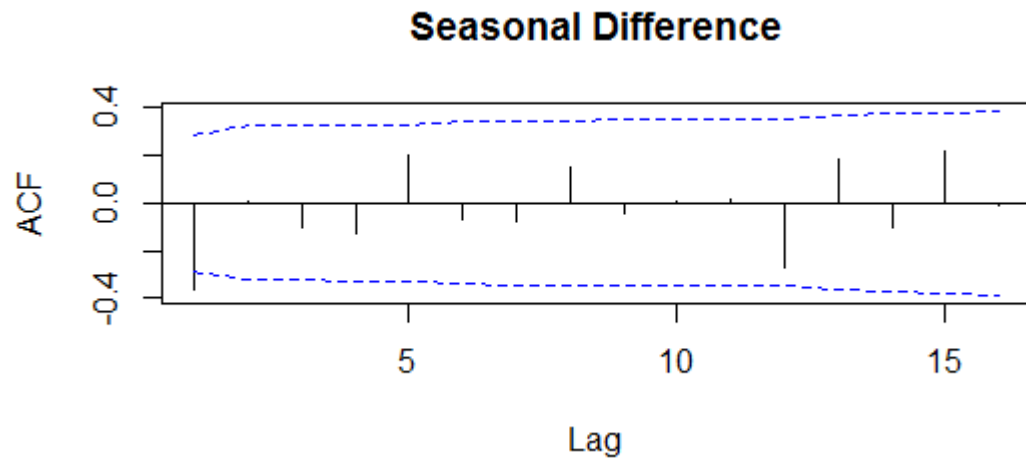
從 EACF 的圖中判斷可能的模型為 $ARIMA(0, 1, 2)$ 。

iii. 季節性差分

因為從未差分以及差分後的時間序列圖都可以看出此筆數據可能有季節趨勢，因此把此筆資料做季節性差分。



做完季節性差分後的时间序列圖看起來已經沒有特別的趨勢相了。



做完季節性差分後的 ACF 只有在第 1 步超出信賴界，因此可能的模型為 $\text{SARIMA}(0, 1, 1) \times (0, 1, 0)_{12}$ 。

從以上全部的分析當中所選出的候選模型如下：

候選模型	$\text{SARIMA}(1, 0, 0) \times (1, 0, 0)_{12}$
	$\text{SARIMA}(0, 1, 1) \times (0, 1, 0)_{12}$
	$\text{AR}(1)$
	$\text{ARIMA}(4, 1, 0)$
	$\text{ARIMA}(0, 1, 2)$

2. 參數估計

以 $\text{SARIMA}(1, 0, 0) \times (1, 0, 0)_{12}$ 為例：

$$(1 - \psi B^{12})(1 - \phi B)(Y_t - \mu) = e_t$$

係數	ψ	ϕ	intercept
估計量	0.7291	0.6598	6.8220
標準差	0.0945	0.1049	0.3438

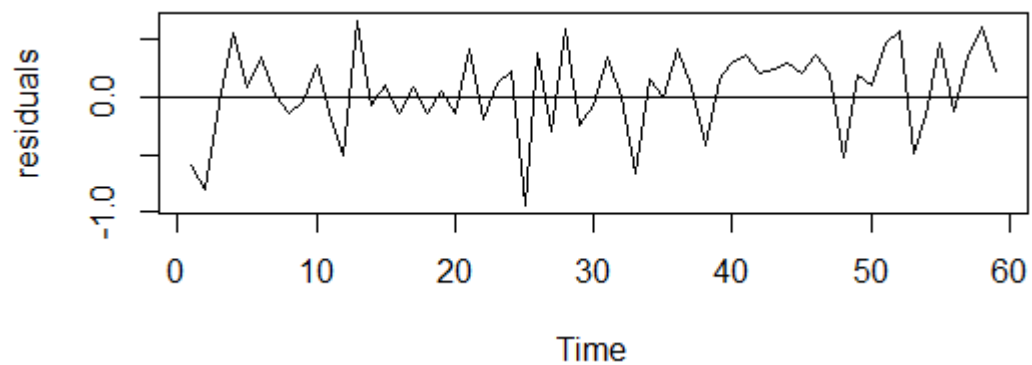
AIC=61.8

此模型的係數顯著不為零。

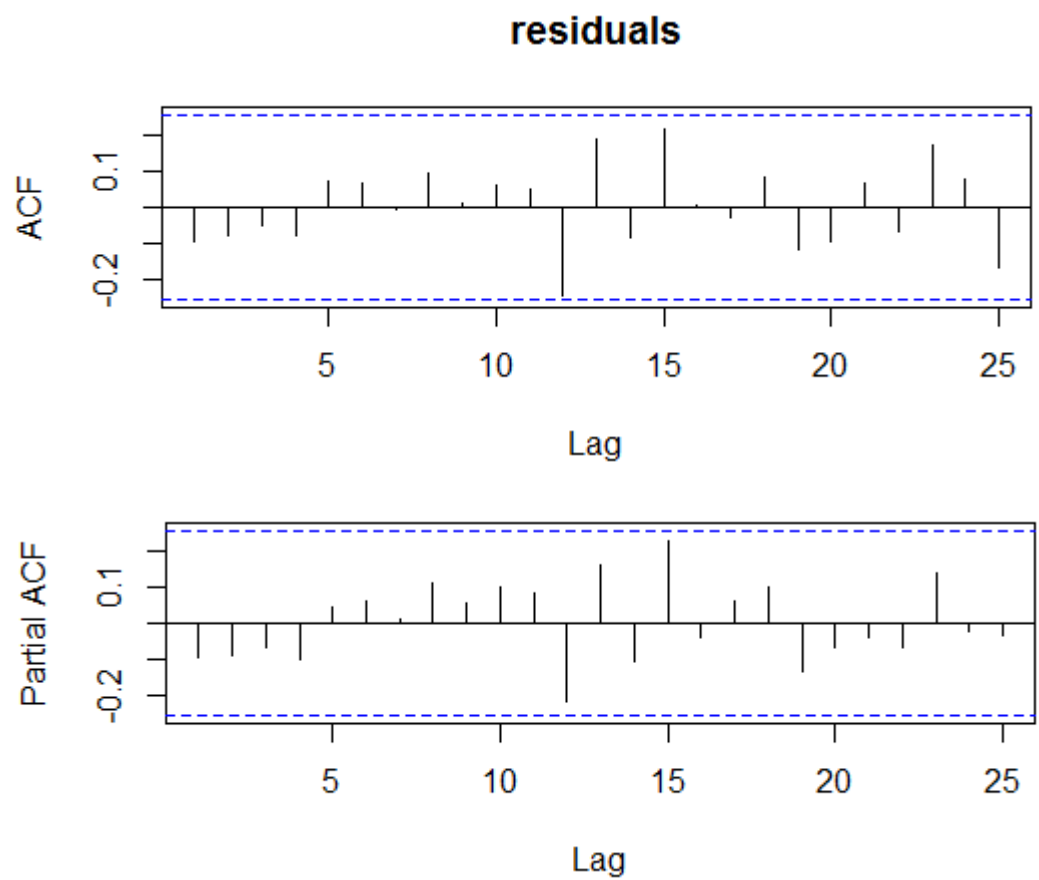
3. 殘差分析

以 $\text{SARIMA}(1, 0, 0) \times (1, 0, 0)_{12}$ 為例：

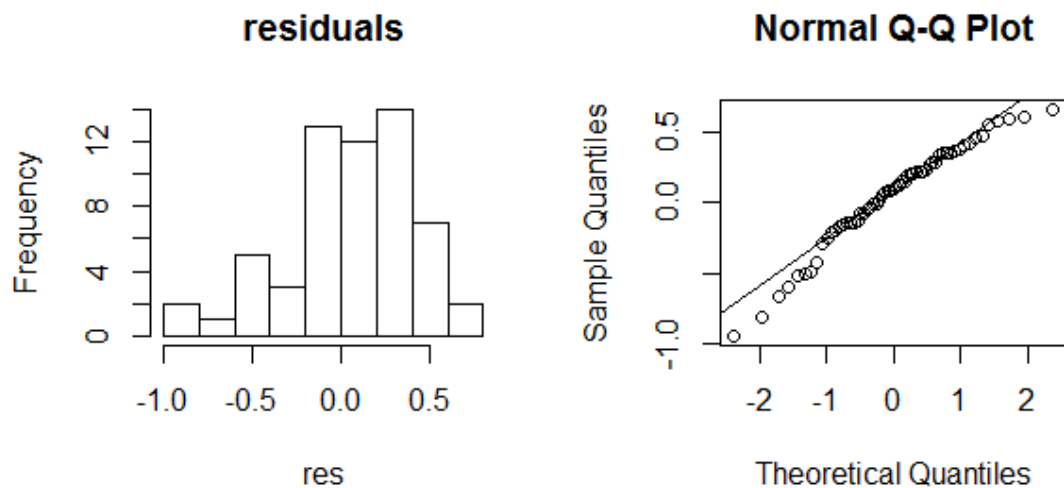
- 進行模型診斷



首先觀察原始的模型殘差圖，從圖中看不出有任何趨勢相。



殘差的 ACF 與 PACF 也都沒有超出信賴界，代表殘差之間沒有自相關。



從直方圖中可以看出殘差的值大概對稱於 0，而在 qqplot 中兩端的點有偏離直線的現象，因此需要進一步做常態檢定。

t-test

檢定量	1.0876
自由度	58
p-value	0.2813

在 t 檢定中的 $p\text{-value} > 0.05$ ，因此不拒絕平均為 0 的假設。

Shapiro-Wilk test

檢定量	0.96281
p-value	0.06857

在 Shapiro-Wilk 的檢定當中， $p\text{-value} > 0.05$ ，所以不拒絕殘差為常態分配的假設。

4. 模型統整

把所有可能模型的模型診斷整理如下：

	$\text{SARIMA}(1, 0, 0) \times (1, 0, 0)_{12}$	$\text{SARIMA}(0, 1, 1) \times (0, 1, 0)_{12}$
t-test	接受 H_0	接受 H_0
Shapiro-Wilk test	接受 H_0	拒絕 H_0
ACF(超出信賴界)	都不超過	都不超過
Ljung Box test	都接受 H_0	都接受 H_0

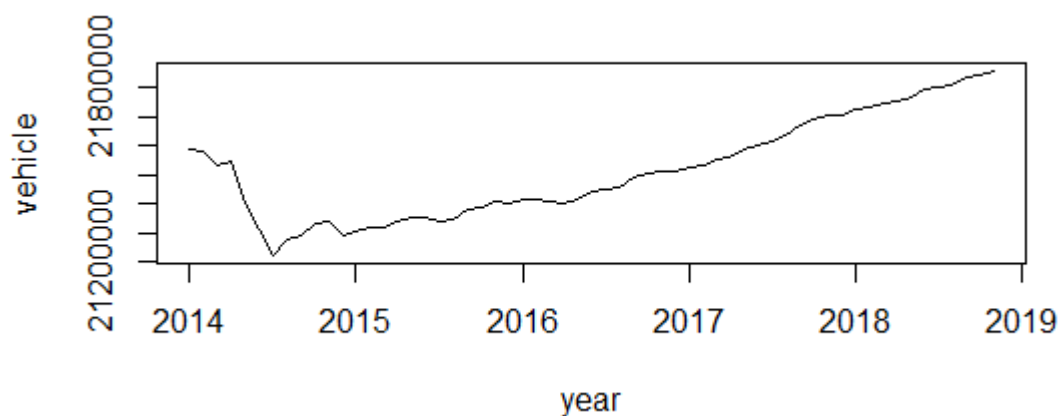
	AR(1)	ARIMA(4, 1, 0)	ARIMA(0, 1, 2)
t-test	接受 H_0	接受 H_0	接受 H_0
Shapiro-Wilk test	接受 H_0	接受 H_0	接受 H_0
ACF(超出信賴界)	Lag=12, 24	Lag=12, 24	Lag=12, 24
Ljung Box test	兩步都拒絕 H_0	兩步都接受 H_0	兩步都拒絕 H_0

由上表可得最佳模型應該是 $SARIMA(1, 0, 0) \times (1, 0, 0)_{12}$ 。

(一) 汽機車數量資料建模流程與結果

1. 模型判斷

此筆時間序列為全台灣 2014 年 1 月到 2018 年 11 月汽機車數量的月資料。



首先觀察此筆時間序列，由圖可以明顯察覺在前期有短暫的下降，到 2014 年中就一直持續上升，而且資料的數值過大，在配模型時的 e_t 也會因此變得很大而難以配適模型，因此選擇把資料進行對數轉換。

Dicky-Fuller test

檢定量	-9.2688
Lag order	3
P-value	<0.01

KPSS test

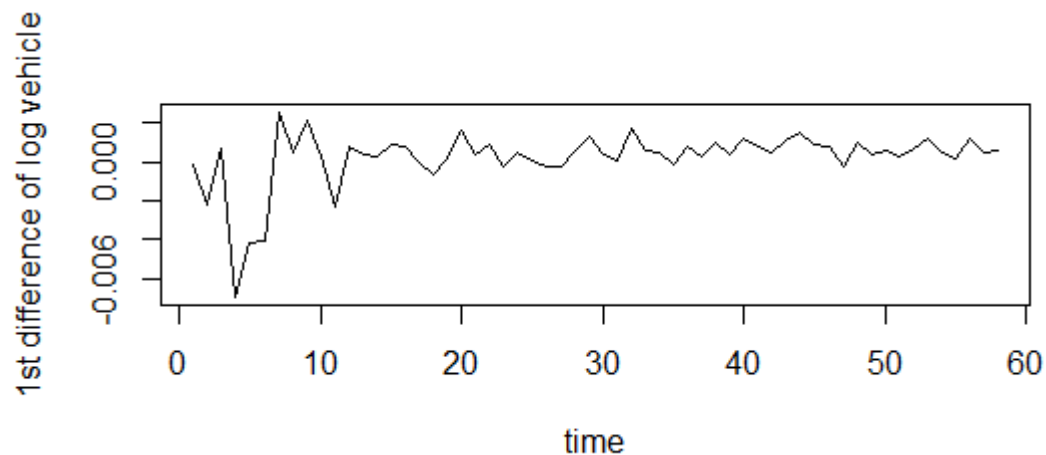
檢定量	2.3867
Lag order	1
P-value	<0.01

為了確定資料是否需要差分，因此分別做 Dicky-Fuller 以及 KPSS 檢定，而得到的結論分別是

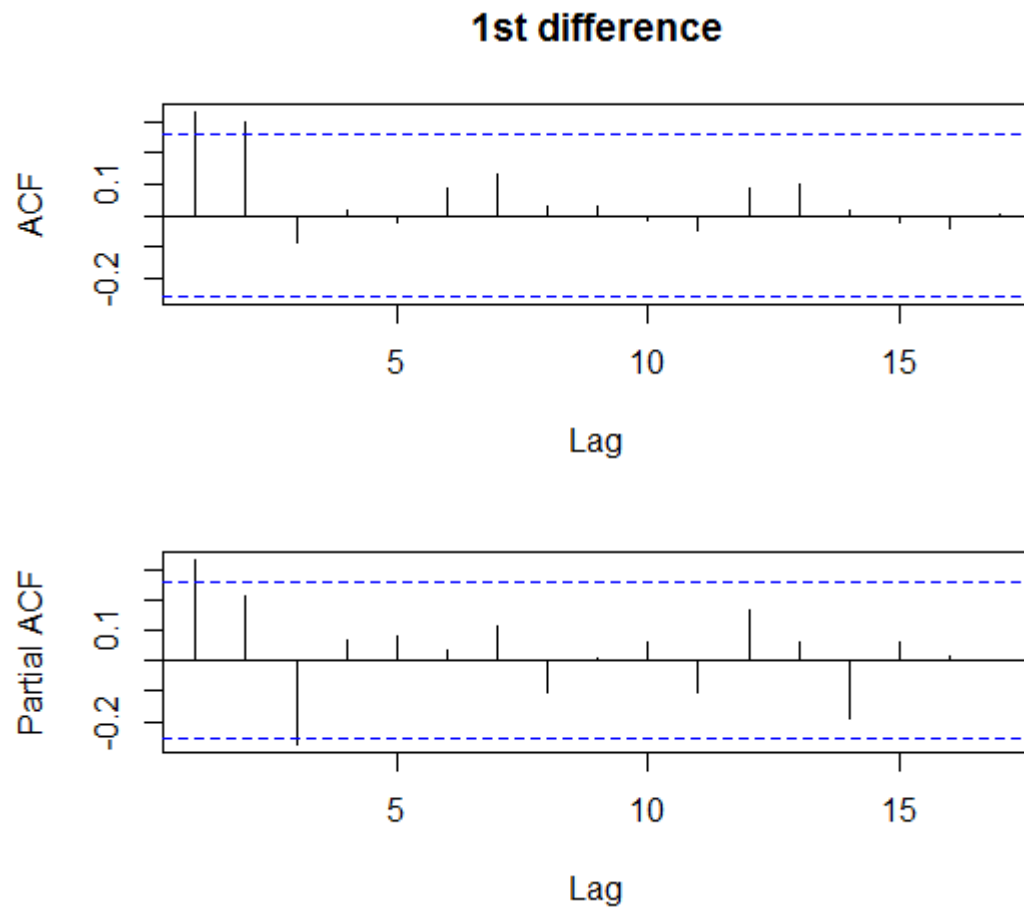
(1)Dicky-Fuller 檢定的 $p\text{-value} < 0.05$ ，因此拒絕需要做差分的假設。

(2)KPSS 檢定中的 $p\text{-value} < 0.05$ ，因此拒絕不需要做差分的假設。兩個檢定出的結果不同，但是從原始資料的時間序列圖來看，發現有明顯的上升趨勢，因此還是選擇做差分再配模型。

i. 一階差分

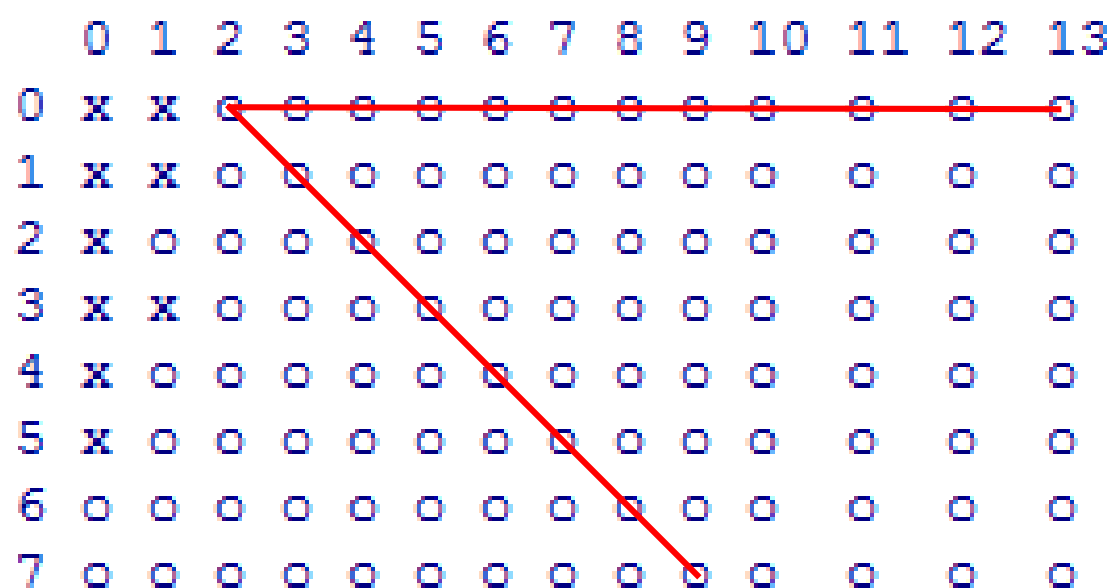


進行一階差分後的时间序列圖在前面有一段變動幅度比較大，後面看起來就沒有其他趨勢相。



把資料進行差分後觀察 ACF 以及 PACF 的圖，ACF 為兩步截斷，因此可能為 $ARIMA(0, 1, 2)$ ；PACF 只有在第 1 步明顯超出信賴界，因此可能的模型為 $ARIMA(1, 1, 0)$ 。

AR/MA



從 EACF 的圖中判斷可能的模型為 ARIMA(0, 1, 2)。

從以上全部的分析當中所選出的候選模型如下：

候選模型	ARIMA(0, 1, 2)
	ARIMA(1, 1, 0)

2. 參數估計

以 ARIMA(0, 1, 2)為例：

$$(1 - B)Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

係數	θ_1	θ_2
估計量	0.3664	0.4155
標準差	0.1291	0.1163

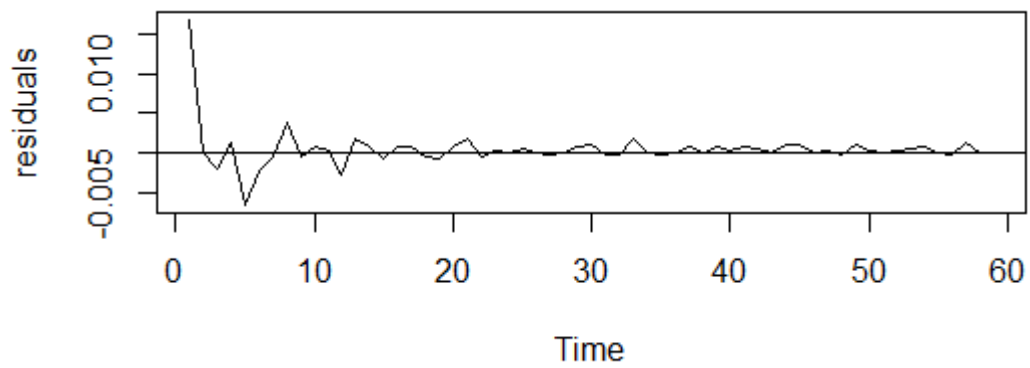
AIC= -599.18

此模型的係數皆顯著不為 0。

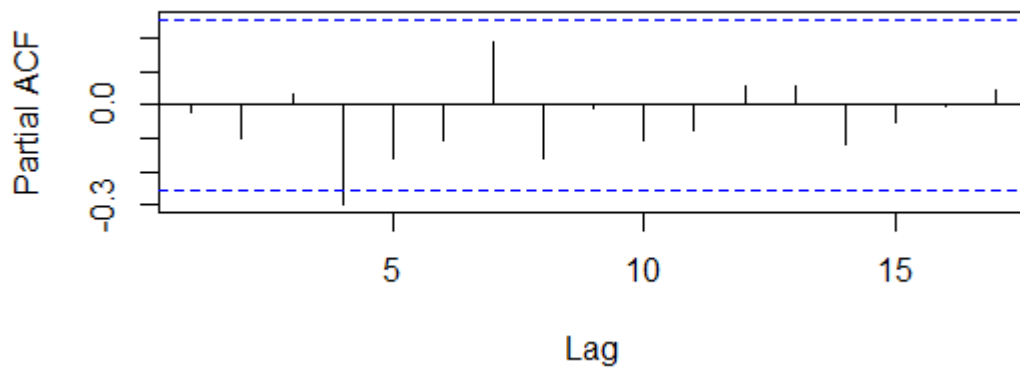
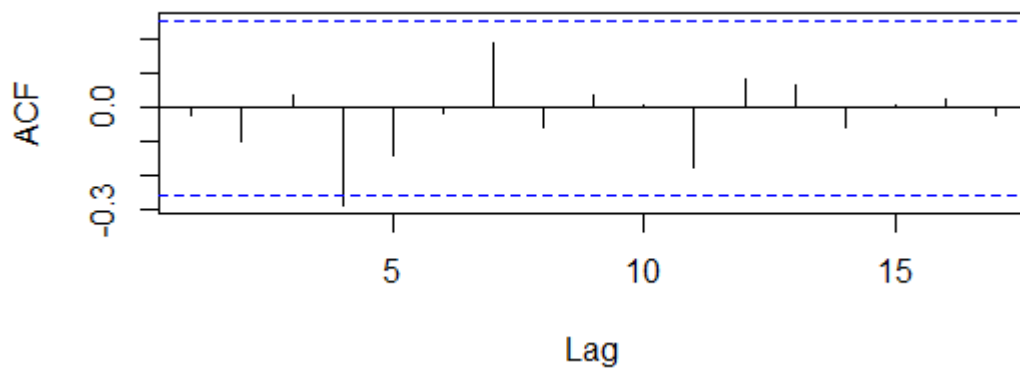
3. 殘差分析

以 $ARIMA(0, 1, 2)$ 為例：

- 進行模型診斷



此殘差圖只有在前面變動比較大，不過看不出有特別的趨勢。

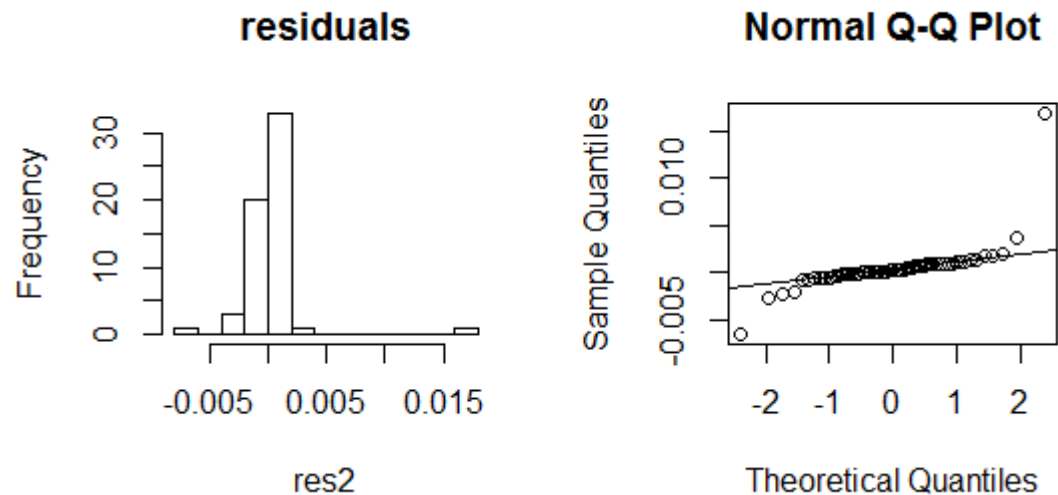


殘差的 ACF 與 PACF 在第 4 步的地方都超出信賴界因此對第 4 步做 Ljung-Box 檢定。

Ljung-Box test

檢定量	6.0796
自由度	4
p-value	0.1933

檢定出的結果為 $p\text{-value} > 0.05$ ，代表此殘差沒有顯著的自相關。



從直方圖中可以看出殘差的值大概對稱於 0，而在 qqplot 中兩端的點有偏離直線的現象，因此需要進一步做常態檢定。

t-test

檢定量	1.2225
自由度	58
p-value	0.2265

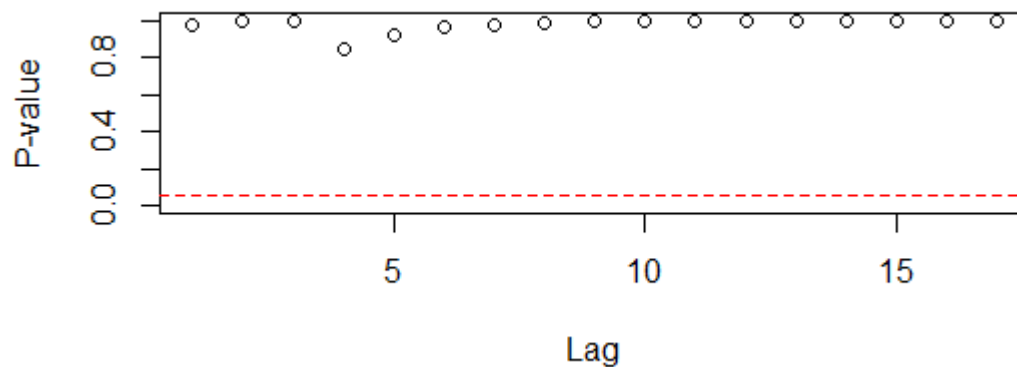
在 t 檢定中的 $p\text{-value} > 0.05$ ，因此不拒絕平均為 0 的假設。

Shapiro-Wilk test

檢定量	0.51701
p-value	1.21×10^{-12}

在 Shapiro-Wilk 的檢定當中， $p\text{-value} < 0.05$ ，所以拒絕殘差為常態分配的假設。

因為在 Shapiro-Wilk 的檢定當中拒絕殘差為常態分配的假設，所以殘差之間可能有異質相關，因此再做 McLeod.Li test。



上圖為檢定的結果，因為所有的 p-value 都落在紅色的虛線之上，代表沒有顯著的異質相關。

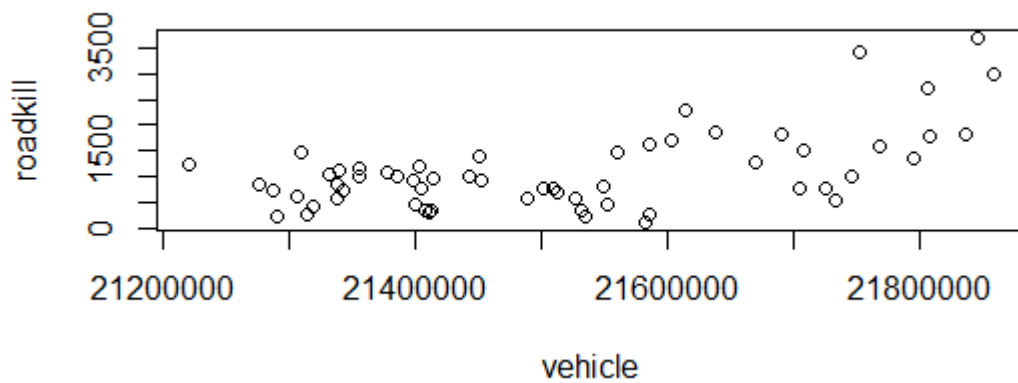
4. 模型統整

把所有可能模型的模型診斷整理如下：

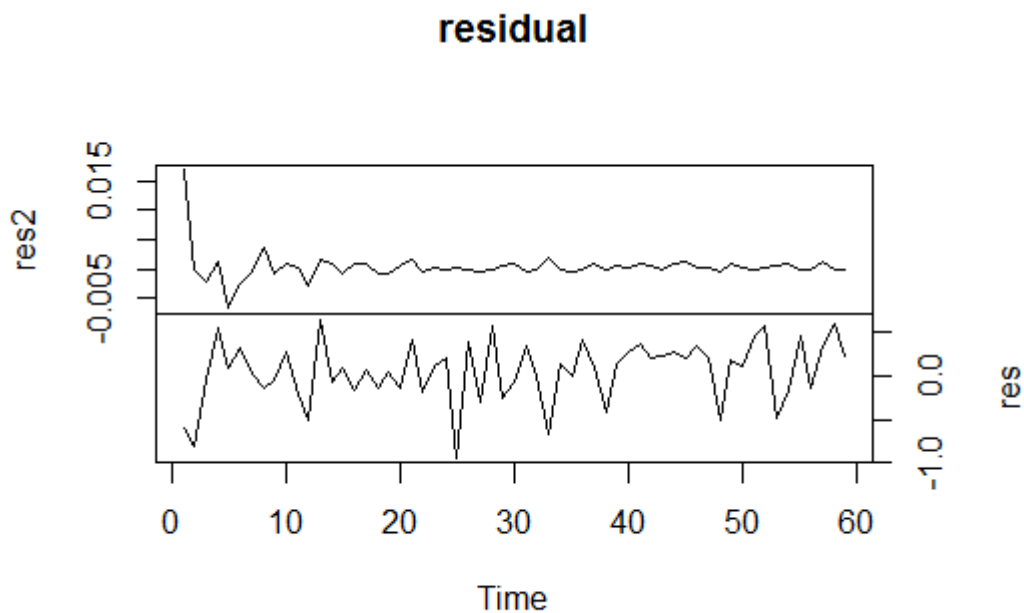
	ARIMA(0, 1, 2)	ARIMA(1, 1, 0)
t test	接受	接受
Shapiro-Wilk test	拒絕	拒絕
McLeod.Li test	接受	接受
ACF (超出信賴界)	Lag=4	Lag=4
Ljung Box test	接受	接受

由上表可知 ARIMA(0, 1, 2)的結果與 ARIMA(1, 1, 0)一樣，不過 ARIMA(0, 1, 2)的 AIC= -599.18 小於 ARIMA(1, 1, 0)的 AIC= -593.98，所以此筆資料最佳模型選擇的是 ARIMA(0, 1, 2)。

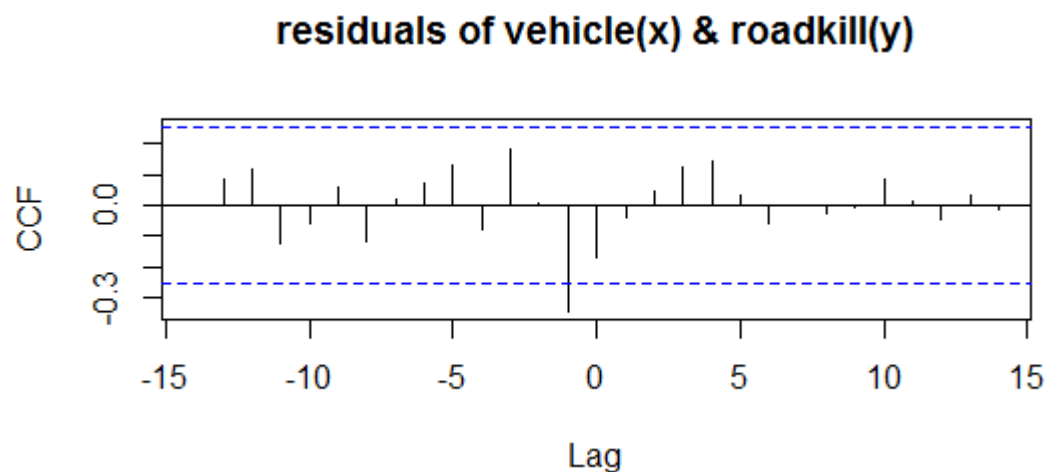
(三)相關性分析



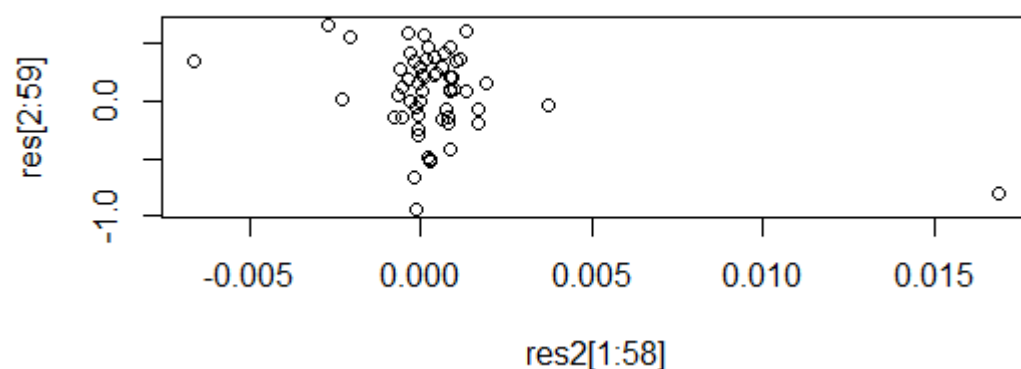
接著要進行兩筆數據的相關性分析，首先畫出兩筆數據的散佈圖，由圖以及計算出的相關係數= 0.5762344 可以看出這兩筆數據可能有一定程度的正相關，不過還是要由兩筆數據殘差間的關聯才能判斷是否真的有相關性。



把兩筆數據的殘差圖做比較，從上圖看不出兩筆資料的殘差有相關性，且計算出的相關係數= -0.1712432，只有一點點負相關而且跟原始數據的結果不同，所以接著畫 CCF 的圖觀察有沒有延遲步數的相關。



觀察畫出的 CCF 圖，只有在-1 步的地方超出信賴界，代表汽機車的數量可能會影響後一期路殺事件發生的次數，因此把資料平移再計算相關係數=-0.3554915，結果比未平移前的負相關性還大。



再把平移後的殘差畫出散布圖，由圖發現其實兩筆資料看起來並沒有相關性。

四、結論

在分析兩筆時間序列的資料之前，原本以為汽機車的數量要是 一直變多的話，路殺事件發生的可能性也會有直接的影響，兩者之間有很大的正相關才對，但是完整的分析之後，雖然以兩筆資料的原始數據來看有一定程度的正相關，但是分析這兩筆資料模型的殘差，卻變成負相關不過相關性不大，代表在分析的過程中把所有時間趨勢拿掉之後，兩者的關聯性反而跟一開始預期的結果不同，因此這樣的結果說明了汽機車數量的多寡可能不是影響路殺事件的主要因素。

五、參考資料

台灣路死觀察網：

<https://roadkill.tw/>

交通部公路總局 統計查詢網：

<https://stat.thb.gov.tw/hb01/webMain.aspx?sys=100&funid=11100>

兇手就是你爸的車！ 國內首齣環境生態劇 化身石虎控訴路殺危害野生動物：

https://www.newsmarket.com.tw/blog/101186/?fbclid=IwAR26qFs0u2HXnvFji7faRpL38wfVUtaGXP2ocv7nKCa814XNwCLE6FCMn_0

暢行無阻動物廊道 各國比美：

<http://www.tanews.org.tw/info/12415>

六、課堂心得與建議

之前大一大二的課程幾乎都是理論方面的，很少會應用在實際的例子當中，所以過去會不知道自己學的東西要怎麼用，也覺得那時候的課程很抽象跟難懂，到了大三才開始有比較深入的學一些如何應用統計在真實數據的課程，而在時間序列這堂課算是學了很多實證資料的分析，除了讓我比較了解如何運用統計方法作分析並且用統計學觀念去解釋分析出來的結果之外，也讓我明白之前的課程都是這些應用背後的基礎，對於統計理論有比較具體的理解，而不像過去可能只會計算跟證明。

而建議的部分，我覺得以後可以換一間教室上課，因為這堂課是以寫板書為主要的上課方式，但是教室的座位配置會讓兩側的座位視線太斜而看不清楚，當修課人數很多的時候就會有人坐在兩邊的位置，影響同學上課的品質，因此或許可以考慮跟其他主要是用投影片上課的課程協調交換教室。