

# 香水百合價格預測

## 目錄

壹、研究動機與介紹-----	2
貳、資料來源與分析方法-----	2
參、資料分析-----	3
肆、模型建立-----	7
伍、殘差檢定-----	9
陸、其他模型之殘差分析-----	13
柒、模型預測-----	14
捌、結論-----	15
玖、參考資料-----	15

## 壹.研究動機與介紹

台灣地處亞熱帶地區，氣候優越而多樣，非常適合各種花卉的種植和育種。再者，可以通過海拔差異來進行花卉生產的季節性調整，從而使台灣花卉產業蓬勃發展。根據行政院農業委員會（農業合作社）的統計，台灣的花卉生產面積已從1983年的2996公頃增長到2015年的13864公頃，產值也從16億元增長到18億元。超過167億元。這表明台灣花卉業的產值在過去30年中增長了十倍以上。台灣的國內花卉產業主要集中在出口銷售上。由於其地理優勢，它靠近世界第三大花卉消費市場--日本，並且國內運輸和銷售技術得到了改善，因此國內花卉出口的產值逐年增加，從2005年的總出口額7167萬美元增長到2015年的19362萬美元。在國內市場，隨著國內經濟的發展和國民收入的增加，除了傳統的需求外，公眾也逐漸關注改善環境和家庭生活質量。因此，儘管花不是人們的必需品，但它們也是重要的國內農產品之一。

根據農業和食品管理局的2017年統計數據，批發市場約佔總產量的90%，這表明批發市場對花卉的重要性。批發市場的目的是為生產者和銷售者提供可靠的交易服務平台，並建立運輸和銷售訂單。通過批發市場建立拍賣系統和公開透明的交易流程，我們可以提高運輸和銷售效率，保護買賣雙方的權利，並在市場供求之間保持平衡。

由於花卉的批發市場通常是花店或是大型廠商，因此採購的數目都是非常龐大的，即使是些微的價差也有可能造成龐大的價差損失，就如同匯率價差一樣，因此我們這組希望藉由花卉的模型預測來降低價格的不確定性，以及提供最佳購入時機。

我們這組將預測香水百合的月平均花價，因為香水百合相較其他花種的成交量以及受歡迎程度都較高，且其花期也較其他花種長，因此市場的活絡程度相對較高。而我們採用月平均是因為我們研究的週期大約為22年，因此採用日資料的波動性太大，因此我們認為月平均的數據預測起來比較有參考依據。

## 貳.資料來源與分析方法

近年來，大多數農產品價格預測均採用時間序列分析作為主要預測方法，並利用以往的數據來構建預測模型。我們將在時間序列分析中使用ARIMA模型預測結果。

**研究題目:**百合的每月平均價格

**資料來源:**農產品批發市場交易站

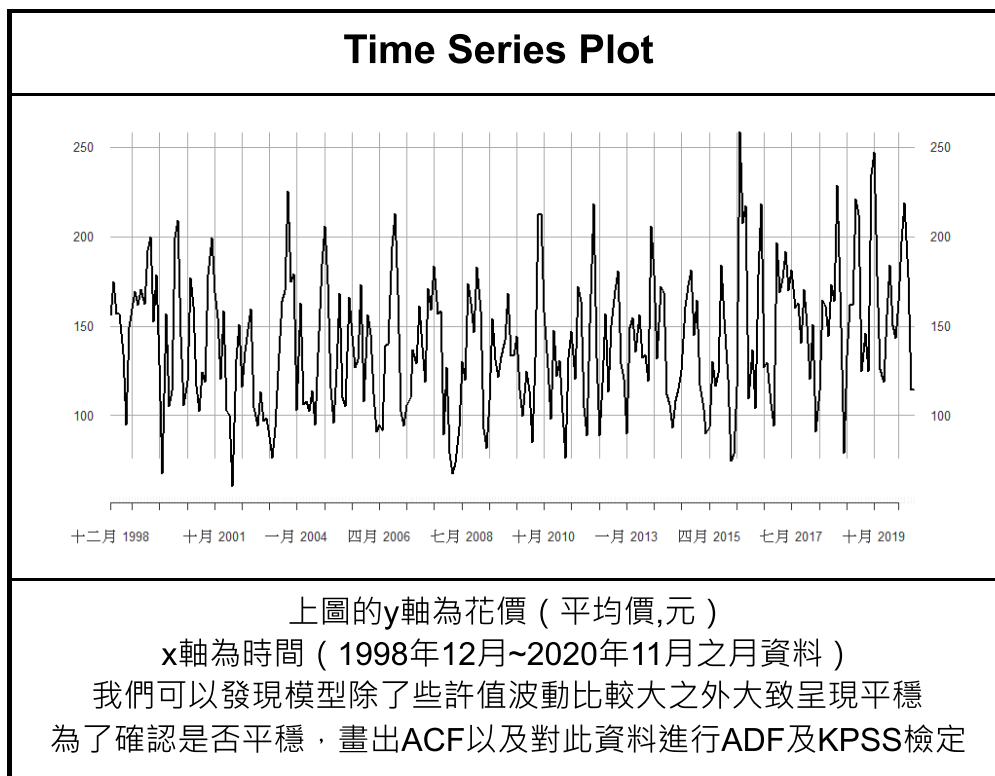
**資料期間:**1998年12月至2020年11月

**資料數目:**264筆

**時間間隔:**月

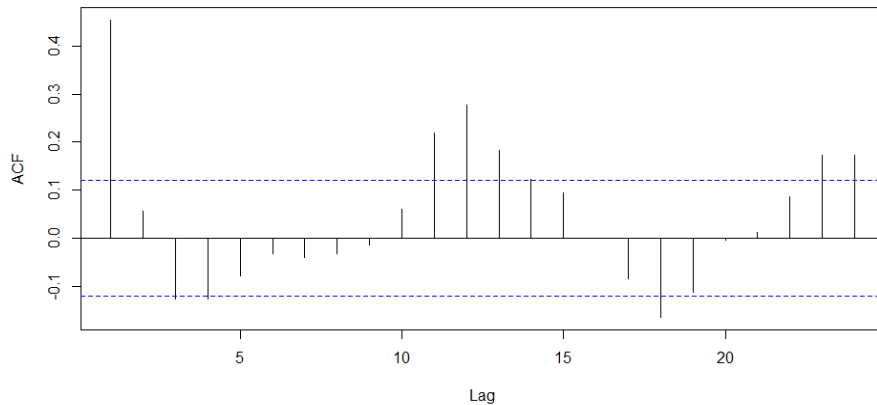


## 參.資料分析



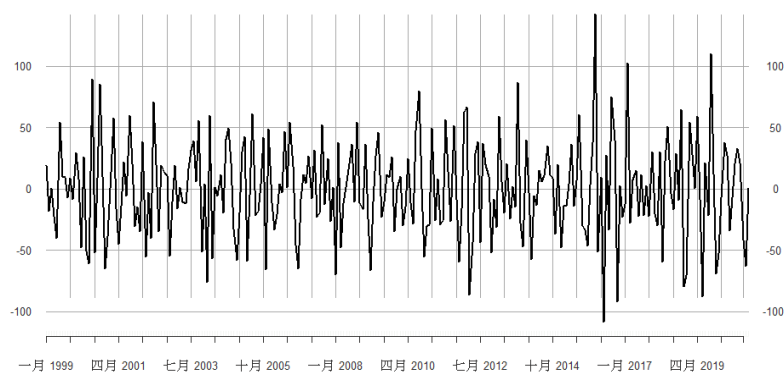
Dickey-Fuller test:		KPSS test:	
H0: $ \alpha  \geq 1$ , not stationary H1: $ \alpha  < 1$ , stationary		H0: $ \alpha  < 1$ , stationary H1: $ \alpha  \geq 1$ , not stationary	
Dickey-Fuller	- 6.9794	KPSS	0.54363
p-value( $\alpha=0.05$ )	0.01	p-value( $\alpha=0.05$ )	0.03184
Conclusion	stationary	Conclusion	not stationary
由Dickey-Fuller test 認為此資料為平穩, 不需要做差分 由KPSS test 認為此資料為不平穩, 需要做差分 接著我們來看ACF			

## ACF



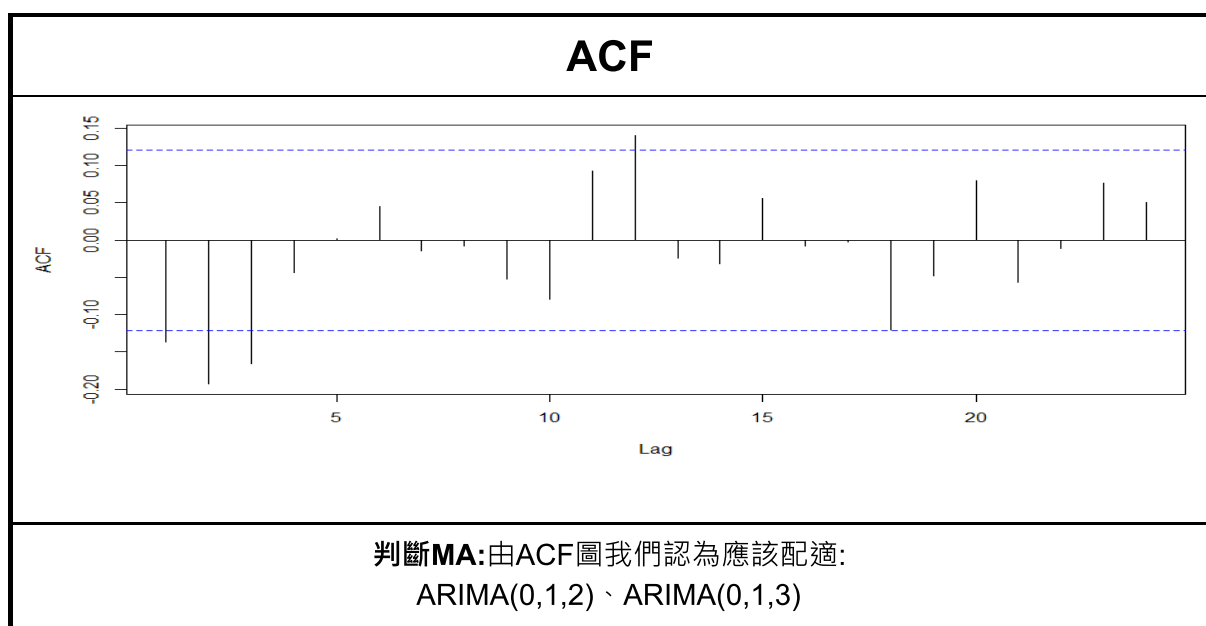
由ACF圖我們發現在第一步之外  
11、12、13及23、24步都有顯著的影响  
我們綜合上述檢定以及ACF圖  
我們認為要對資料進行一次差分的處理

## First Difference Time Series Plot

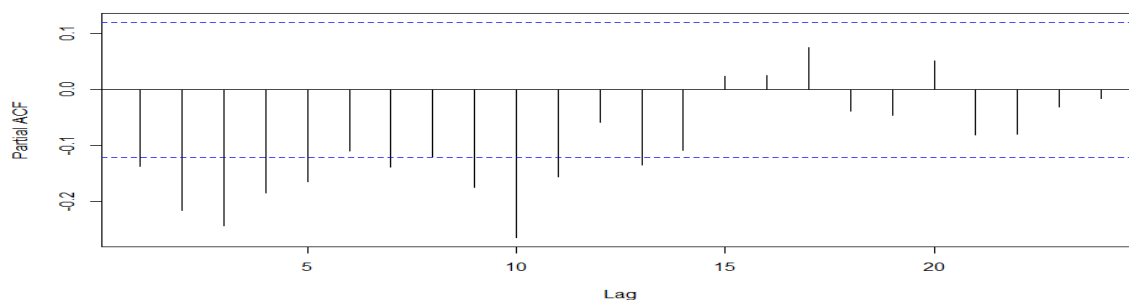


上圖的y軸為花價（平均價,元）  
x軸為時間（1998年12月~2020年11月之月資料）  
我們可以發現差分後資料的波動變小了

Dickey-Fuller test:		KPSS test:	
H0: $ \alpha  \geq 1$ , not stationary H1: $ \alpha  < 1$ , stationary		H0: $ \alpha  < 1$ , stationary H1: $ \alpha  \geq 1$ , not stationary	
Dickey-Fuller	- 10.034	KPSS	0.013155
p-value( $\alpha=0.05$ )	0.01	p-value( $\alpha=0.05$ )	0.1
Conclusion	stationary	Conclusion	stationary
<p>由Dickey-Fuller test 認為此資料為平穩，不需要做差分</p> <p>由KPSS test 認為此資料為平穩，不需要做差分</p> <p>因此由檢定結果，我們認為此資料已經為平穩資料，可以開始建立模型了</p> <p>接著我們畫出ACF、PACF、EACF來判斷要建立的模型</p>			



## PACF



**判斷AR:**由PACF圖我們認為應該配適:

ARIMA(4,1,0)、ARIMA(5,1,0)、ARIMA(7,1,0)、ARIMA(10,1,0)

## EACF

AR/MA														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	X	X	X	O	O	O	O	O	O	O	O	X	O	O
1	X	O	O	O	O	O	O	O	O	O	O	X	O	O
2	X	X	O	O	O	O	O	O	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O	O	O	O	O	O	O
4	X	X	O	X	O	O	O	O	O	O	O	O	O	O
5	X	X	O	X	O	O	O	O	O	O	O	O	O	O
6	X	X	X	X	O	O	O	O	O	O	O	O	O	O
7	X	X	X	O	O	X	O	O	O	O	O	O	O	O

**判斷ARMA:**由EACF圖我們不確定是紅色字體的ARIMA(1,1,1)，還是灰色底的ARIMA(2,1,2)比較好，因此我們決定兩個模型都配配看

肆.模型建立:

MODEL ARIMA(0,1,3)			
$Y_t = Y_{t-1} + e_t - 0.4549e_{t-1} - 0.3543e_{t-2} - 0.1575e_{t-3}$			
	ma1	ma2	ma3
coefficient	- 0.4549	- 0.3543	- 0.1575
S.E.	0.0617	0.0628	0.0643
AIC = 2592.5			

MODEL ARIMA(0,1,2)		
$Y_t = Y_{t-1} + e_t - 0.5318e_{t-1} - 0.4299e_{t-2}$		
	ma1	ma2
coefficient	- 0.5318	- 0.4299
S.E.	0.0491	0.0484
AIC = 2596.11		

MODEL ARIMA(5,1,0)					
$Y_t = 0.7021Y_{t-1} - 0.0594Y_{t-2} + 0.0157Y_{t-3} + 0.1064Y_{t-4} - 0.0667Y_{t-5} - 0.1685Y_{t-6} + e_t$					
	ar1	ar2	ar3	ar4	ar5
coefficient	- 0.2979	- 0.3573	- 0.3416	- 0.2352	- 0.1685
S.E.	0.0607	0.0621	0.0625	0.0621	0.0610
AIC = 2634.08					

<b>MODEL ARIMA(4,1,0)</b>				
<b><math>Y_t = 0.7338Y_{t-1} - 0.0423Y_{t-2} + 0.0191Y_{t-3} + 0.1010Y_{t-4} - 0.1884Y_{t-5} + e_t</math></b>				
	<b>ar1</b>	<b>ar2</b>	<b>ar3</b>	<b>ar4</b>
<b>coefficient</b>	- 0.2662	- 0.3085	- 0.2894	- 0.1884
<b>S.E.</b>	0.0605	0.0605	0.0604	0.0607
<b>AIC = 2639.6</b>				

<b>MODEL (7,1,0)</b>							
<b><math>Y_t = 0.6670Y_{t-1} - 0.0786Y_{t-2} + 0.0059Y_{t-3} + 0.0893Y_{t-4} + 0.0718Y_{t-5} + 0.0986Y_{t-6} + 0.0172Y_{t-7} - 0.1406Y_{t-8} + e_t</math></b>							
	<b>ar1</b>	<b>ar2</b>	<b>ar3</b>	<b>ar4</b>	<b>ar5</b>	<b>ar6</b>	<b>ar7</b>
<b>coefficient</b>	- 0.333	- 0.4116	- 0.4175	- 0.3282	- 0.2564	- 0.1578	- 0.1406
<b>S.E.</b>	0.061	0.0639	0.0670	0.0688	0.0670	0.0640	0.0611
<b>AIC = 2629.56</b>							

<b>MODEL ARIMA(10,1,0)</b>										
<b><math>Y_t = 0.5754 Y_{t-1} - 0.0916Y_{t-2} - 0.0436Y_{t-3} - 0.0533Y_{t-4} + 0.0267Y_{t-5} + 0.0727Y_{t-6} + 0.0226Y_{t-7} - 0.0769Y_{t-8} - 0.0325Y_{t-9} - 0.0129Y_{t-10} - 0.2713Y_{t-11} + e_t</math></b>										
	<b>ar1</b>	<b>ar2</b>	<b>ar3</b>	<b>ar4</b>	<b>ar5</b>	<b>ar6</b>	<b>ar7</b>	<b>ar8</b>	<b>ar9</b>	<b>ar10</b>
<b>coefficient</b>	- 0.4246	- 0.5162	- 0.5598	- 0.5065	- 0.4789	- 0.4062	- 0.3936	- 0.3167	- 0.2842	- 0.2713
<b>S.E.</b>	0.0592	0.0625	0.0675	0.0719	0.0743	0.0740	0.0740	0.0672	0.0624	0.0592
<b>AIC = 2602.58</b>										

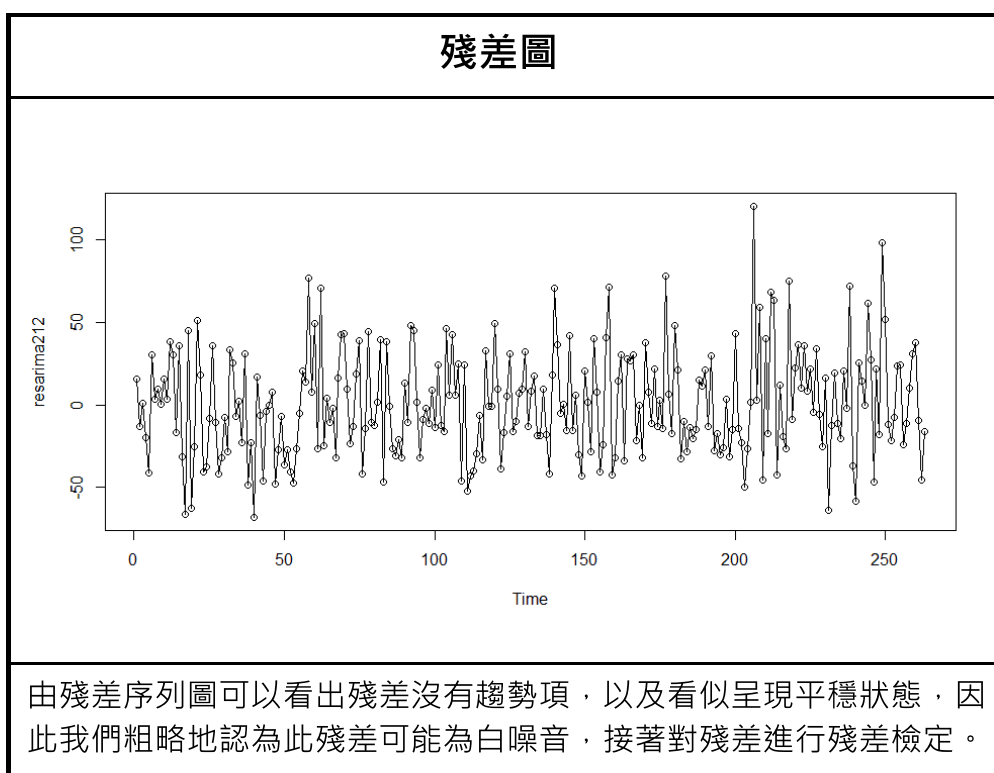


MODEL ARIMA(1,1,1)		
$Y_t = 1.4482Y_{t-1} + 0.4482Y_{t-2} + e_t - 1.0000e_{t-1}$		
	ar1	ma1
coefficient	0.4482	- 1.0000
S.E.	0.0555	0.0188
AIC = 2600.38		

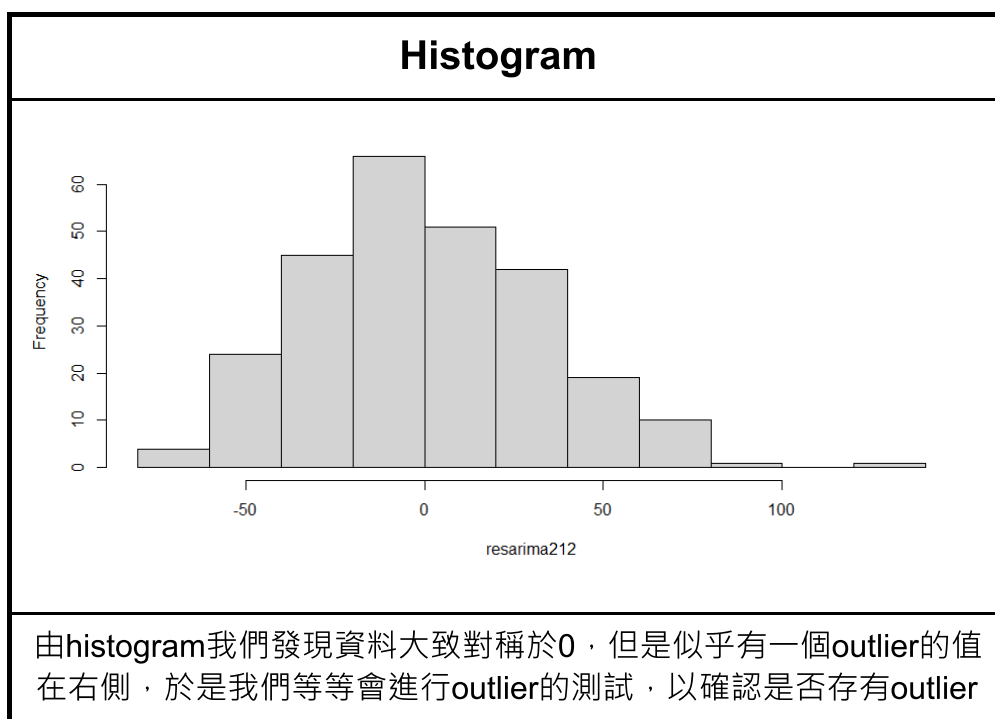
MODEL ARIMA(2,1,2)				
$Y_t = 2.2165Y_{t-1} - 1.7260Y_{t-2} - 0.5095Y_{t-3} + e_t + 1.17344e_{t-1} - 0.7527e_{t-2}$				
	ar1	ar2	ma1	ma2
coefficient	1.2165	- 0.5095	- 1.7344	0.7527
S.E.	0.0871	0.0564	0.0857	0.0842
AIC = 2582.42				

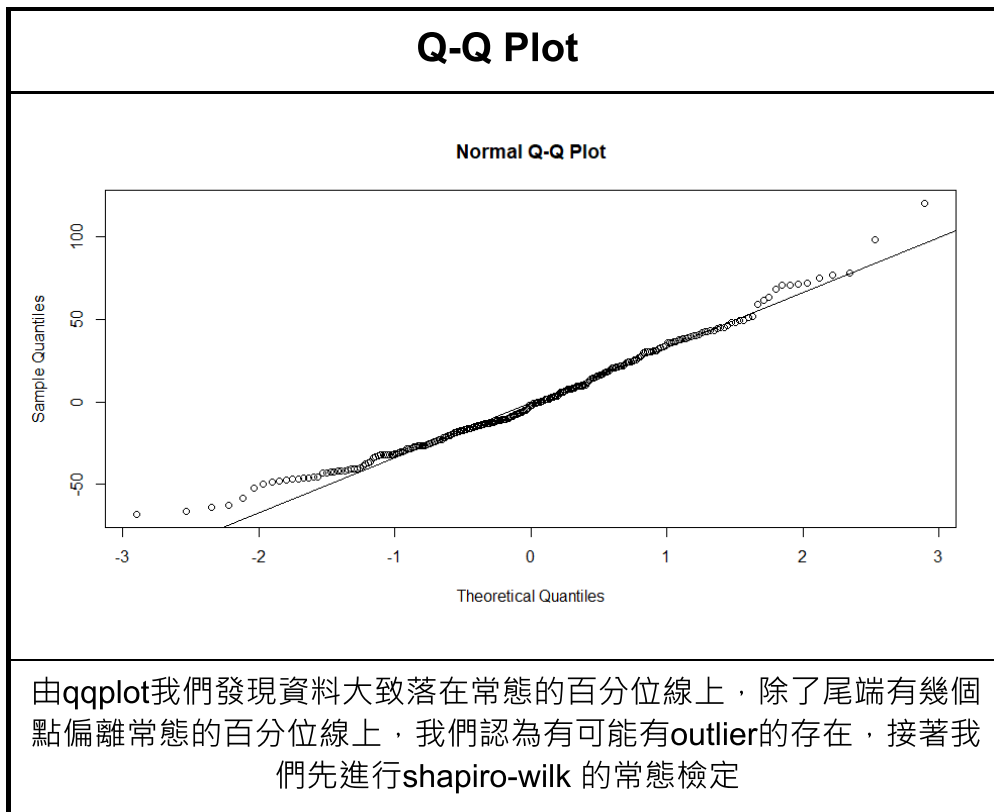
## 伍.殘差檢定:

此部份我們以配適最好的模型**ARIMA(2,1,2)**為例，其餘會以表格呈現



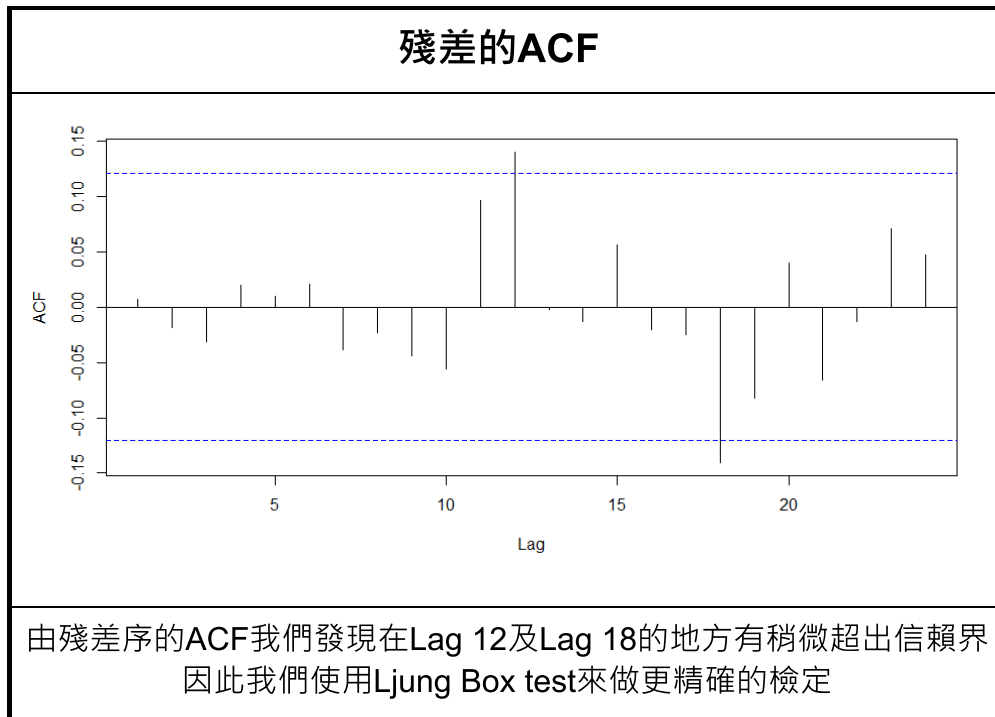
T - test		
H0 : mean is equal to 0 H1: mean is not equal to 0		
t	df	p-value( $\alpha=0.05$ )
0.2875	262	0.774
Conclusion = True mean is equal to 0		
95% C.I. = [-3.338751,4.480397]		
由殘差檢定我們不能拒絕虛無假設，因此我們認定此殘差為white noise的資料		





Shapiro-Wilk Test	
<b>H0 : the distribution is normal</b> <b>H1 : the distribution is not normal</b>	
D	p-value( $\alpha=0.05$ )
0.98192	0.002029
<b>Conclusion:</b> the distribution is normal	
由shapiro-wilk test我們拒絕分配為常態的虛無假設，因為shapiro-wilk 是檢定力很強的檢定，因此我們決定再試試ks test，如ks test也拒絕，則確定此殘差不服從常態分配。	

One-Sample Kolmogorov-Smirnov Test	
<b>H0 : the distribution is normal</b> <b>H1 : the distribution is not normal</b>	
D	p-value( $\alpha=0.05$ )
0.068662	0.1674
<b>Conclusion:</b> the distribution is normal	
由One-Sample Kolmogorov-Smirnov檢定我們不能拒絕分配為常態的虛無假設，因此雖然shapiro-wilk檢定沒有過，我們仍勉強接受此模型為常態分布	



Ljung-Box test		
H0: not difference from 0 H1: difference from 0		
	Lag 12	Lag 18
X-squared	10.59	17.459
df	12	18
p-value( $\alpha=0.05$ )	0.5643	0.4918
conclusion	not difference from 0	not difference from 0
由結果來看，我們不能拒絕兩者的虛無假設，因此全部的值都會落在信賴界裡面，所以我們認為此模型的配適結果是良好的，但接下來我們仍會測試是否有outliers，如果有的話，對outliers處理過後的模型是否配適的更完善。		

Outlier Test		
	Detcet AO	Detcet IO
outcome	"No AO detected"	"No IO detected"
檢定顯示沒有outliers，因此以上模型即為最終模型		

## 陸.其他模型之殘差分析:

--灰色底為我們採用的最佳模型**ARIMA(2,1,2)**

模型	殘差分析				AIC
	t-test	shapiro-wilk	ks normal	ACF&Ljung box	
	p-value ( $\alpha=0.05$ )				
ARIMA(2,1,2)	0.774 不拒絕	0.0020 拒絕	0.1674 不拒絕	Lag 12 不拒絕 Lag 18 不拒絕	2582.42
ARIMA(0,1,3)	0.285 不拒絕	0.0042 拒絕	0.4293 不拒絕	Lag 11 不拒絕 Lag 12 不拒絕 Lag 18 不拒絕	2592.50
ARIMA(0,1,2)	0.836 不拒絕	0.01352 拒絕	0.7959 不拒絕	Lag 03 拒絕 Lag 11 拒絕 Lag 12 拒絕 Lag 18 拒絕	2596.11
ARIMA(5,1,0)	0.981 不拒絕	0.07214 不拒絕	0.6828 不拒絕	Lag 06 拒絕 Lag 07 拒絕 Lag 11 拒絕 Lag 12 拒絕 Lag 18 拒絕 Lag 19 拒絕	2634.08
ARIMA(4,1,0)	0.971 不拒絕	0.1116 不拒絕	0.8817 不拒絕	Lag 05 拒絕 Lag 11 拒絕 Lag 12 拒絕 Lag 18 拒絕	2639.60
ARIMA(7,1,0)	0.994 不拒絕	0.7149 不拒絕	0.7149 不拒絕	Lag 08 拒絕 Lag 11 拒絕 Lag 12 拒絕 Lag 18 拒絕	2629.56
ARIMA(10,1,0)	0.973 不拒絕	0.02027 拒絕	0.4506 不拒絕	Lag 18 不拒絕	2602.58
ARIMA(1,1,1)	- 0.996 不拒絕	0.0050 拒絕	0.5997 不拒絕	Lag 03 拒絕 Lag 11 拒絕 Lag 12 拒絕 Lag 18 拒絕	2600.38

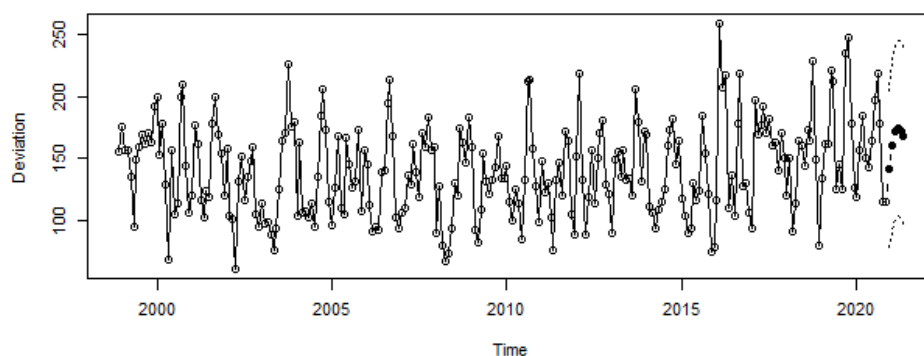
由上表可以看到ARIMA(2,1,2)除了shapiro test 拒絕之外，其餘的檢定都顯示出相較於其他模型此模型是一個相對配適成功的模型，且ARIMA(2,1,2)之AIC也是裡面最小的，因此我們便可以確定我們的最終模型為 **ARIMA(2,1,2)**。

## 柒.模型預測:

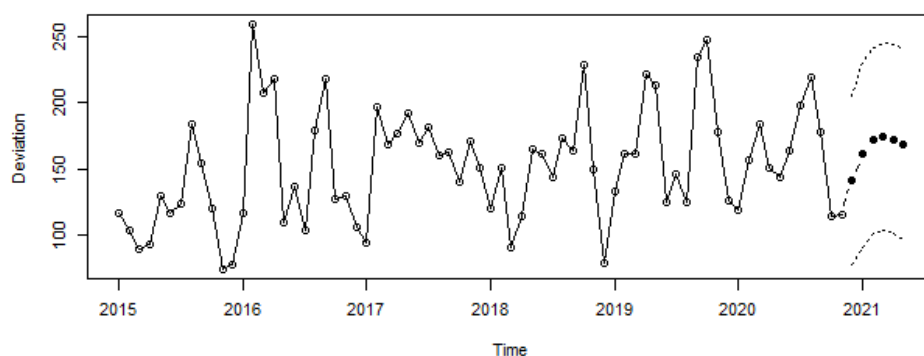
最終模型 $ARIMA(2,1,2)$ :

$$Y_t = 2.2165Y_{t-1} - 1.7260Y_{t-2} - 0.5095Y_{t-3} + e_t + 1.17344e_{t-1} - 0.7527e_{t-2}$$

預測2020年12月至2021年5月之每月花價平均值



下圖我們將時間軸縮短以方便觀看預測值



預測值的波動看起來合理，但信賴區間稍嫌大了一些，不過上下界仍落在合理範圍內

模型預測值

	Dec	Jan	Feb	Mar	Apr	May
上界	204.3800	230.9083	241.6385	244.7483	243.5379	240.1686
估計值	141.3793	160.9705	171.4441	174.2034	172.2238	168.4099
下界	78.37851	91.03279	101.24980	103.65848	100.90962	96.65116

註:經確認12月的月均價為**159.1**。

## 捌.結論:

本次的報告，針對百合的月均價進行模型分析，最後選定ARIMA(2,1,2)進行6步預測(預測六個月的月均價)，在比較預測值與截至年底的月均價，可以知道預測值雖與真實狀況有所差距，但真實值仍落在95%的信賴區間中，因此初步判斷此模型仍能有效透過歷史資料抓到未來趨勢進行預測，而由於現在仍無法獲得未來幾個月的月均價，因此利用2020年12月前的資料對未來半年所做出的後幾步預測是否準確則尚未無法下絕對的定論。

## 玖.參考資料:

1. 林彥甫.(2017). 影響花卉拍賣價格之因素：實證研究．碩士論文．淡江大學 資訊管理系.
2. 林銘昌.(2001). 台灣芒果批發價格預測模型之比較研究．博士論文．國立臺灣大學農業 經濟學研究所．
3. 農產品批發市場交易行情站  
<https://amis.afa.gov.tw/flower/FlowerProdDayTransInfo.aspx>