

# 統計品管期末報告

## 台灣近十年失業率之探討

M122040017 吳俞憲

M122040018 張家輔

# 目錄

- ①研究動機與資料介紹
- ②統計分析與研究方法
- ③時間序列模型
- ④結論



## PART 01

# 研究動機與資料介紹

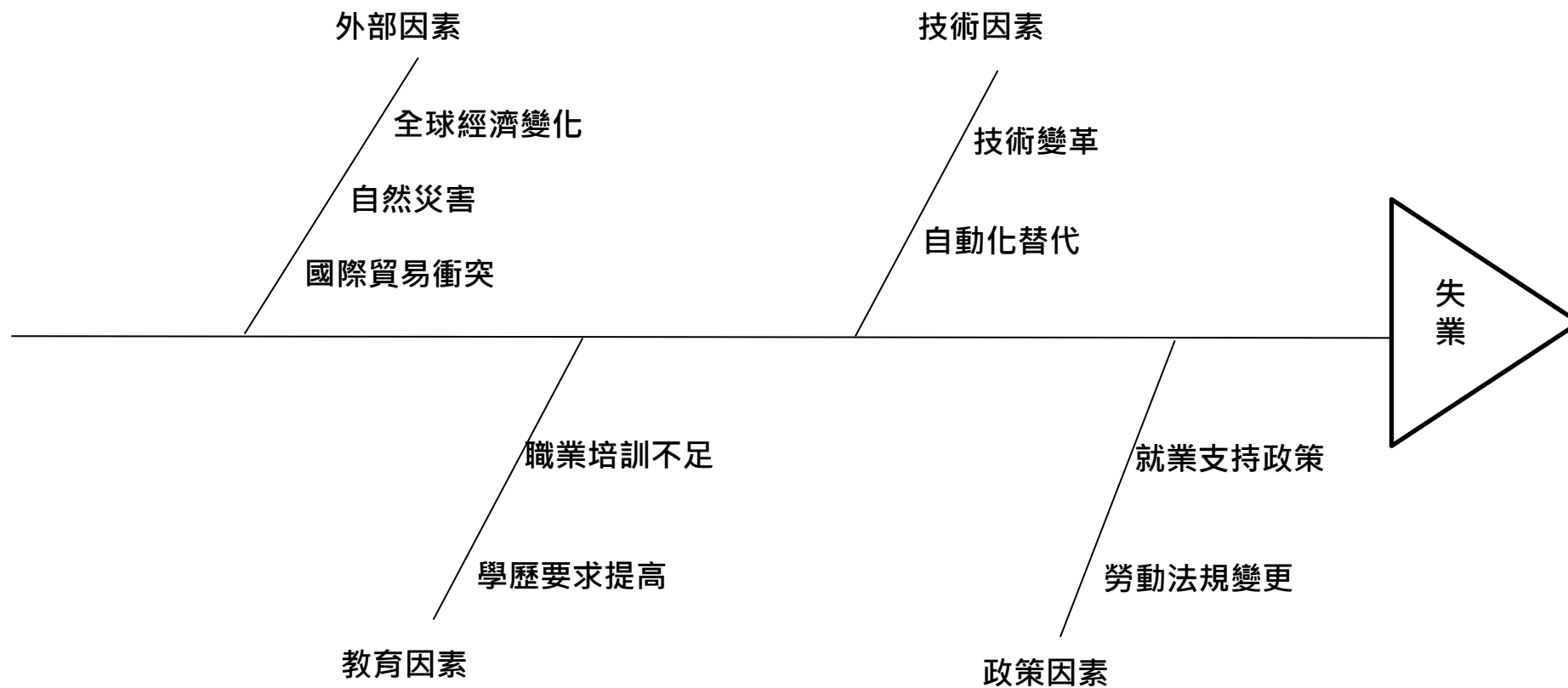


# 研究動機

- 隨著全球經濟的不確定性增加，失業率成為衡量經濟健康狀況的重要指標，影響社會穩定和人民生活水平。
- 後續會針對總失業率與青年失業率去做比較，探討青年失業率是否與總失業率的趨勢是否有所不同
- 觀察經濟指標，是否與失業率有關係

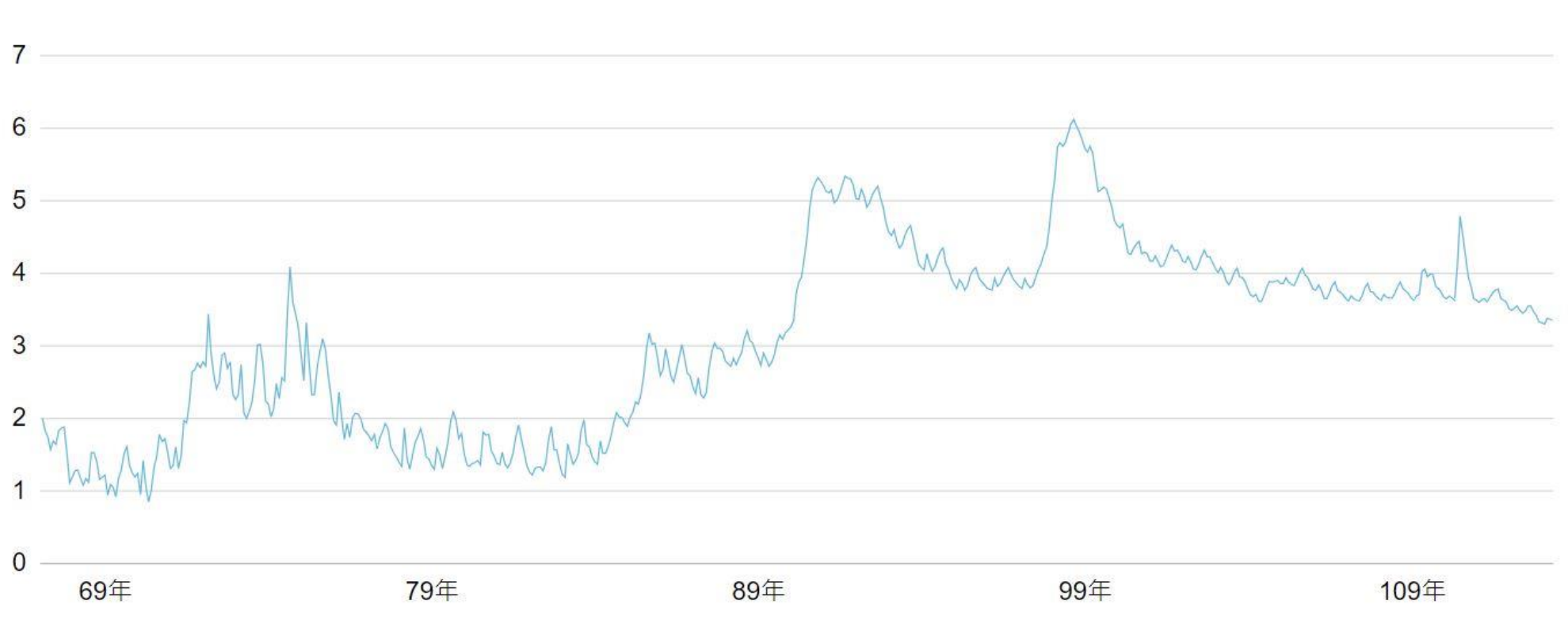


# 因果圖



# 資料介紹

- 資料來源：中華民國統計資訊網、勞動部統計查詢網
- 資料時間：民國100年至112年，共13年。分為月單位及年單位
- 分別做總失業率與青年失業率，其中青年失業率的範圍為15~29歲



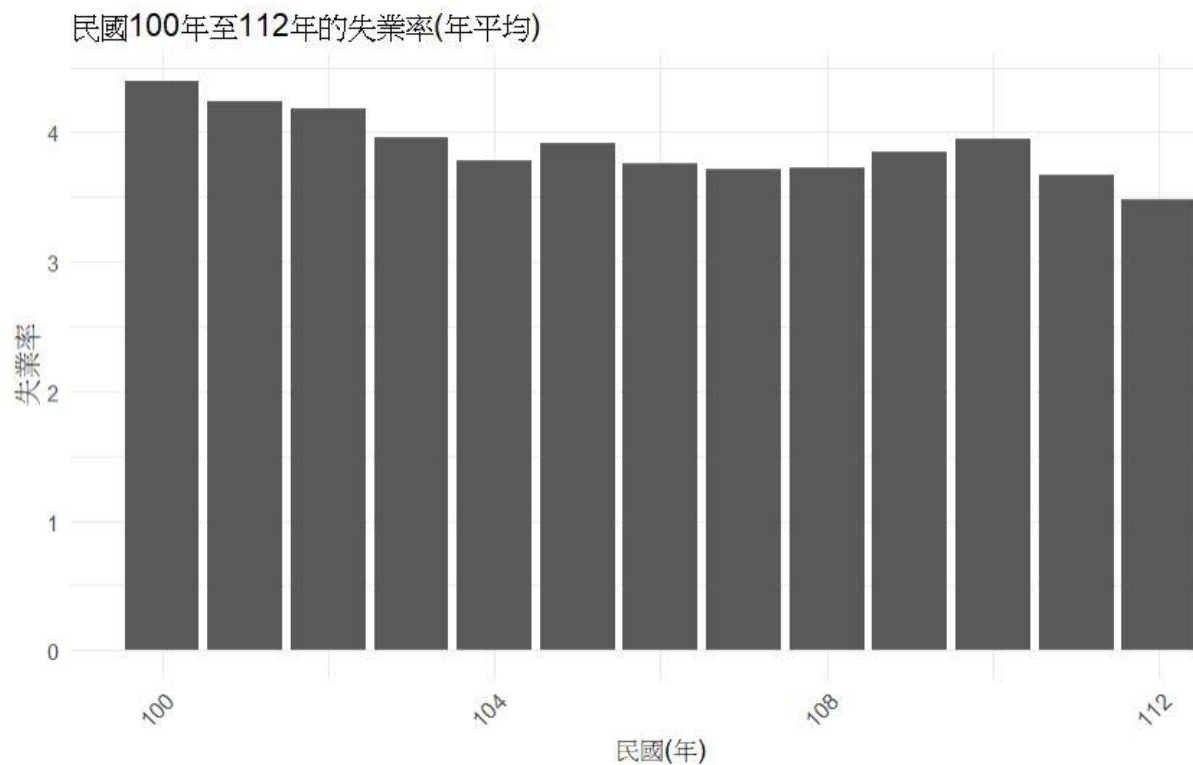


## 資料介紹

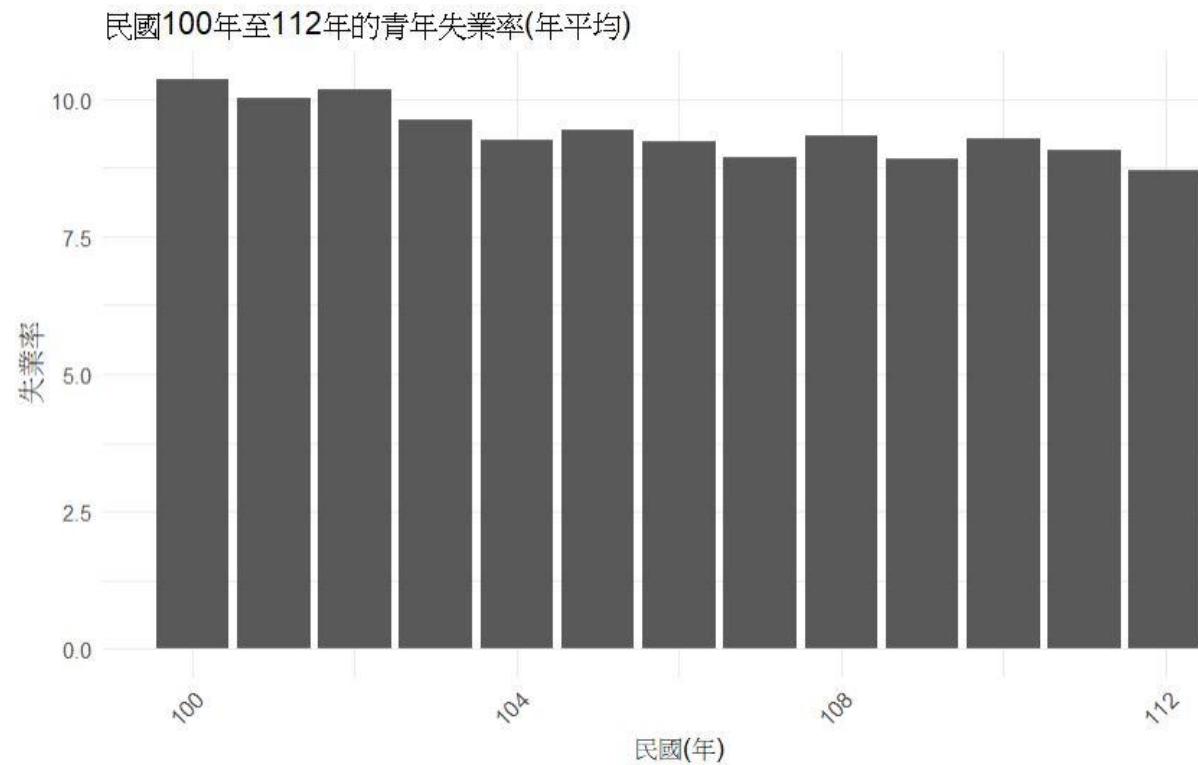
年分(民國)	總失業率(%)	青年失業率(%)
100	4.39	10.35
101	4.24	10.02
102	4.18	10.17
103	3.96	9.62
104	3.78	9.26
105	3.92	9.44
106	3.76	9.24
107	3.71	8.94
108	3.73	9.35
109	3.85	8.91
110	3.95	9.28
111	3.67	9.07
112	3.48	8.71

# 資料介紹

## 總失業率



## 青年失業率





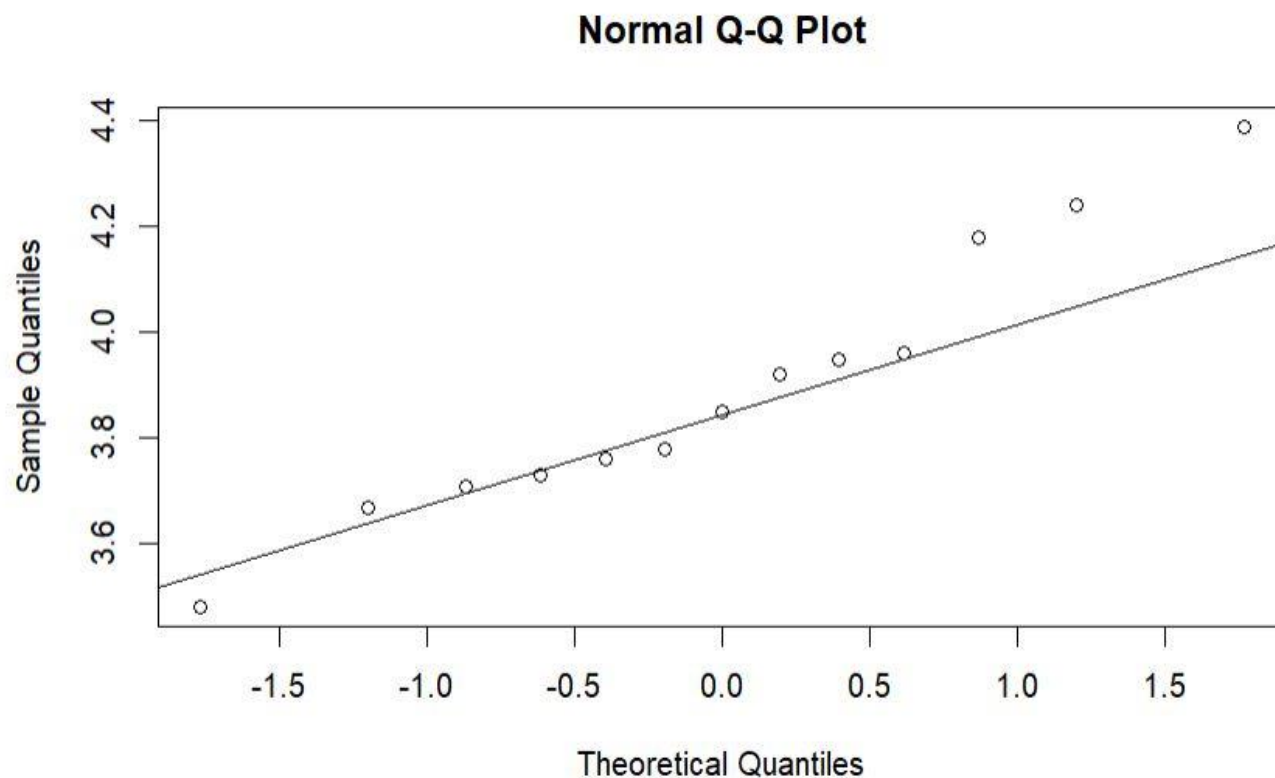


## PART 02

# 統計分析與研究方法

# 統計分析

資料是否屬於常態分佈(總失業率)



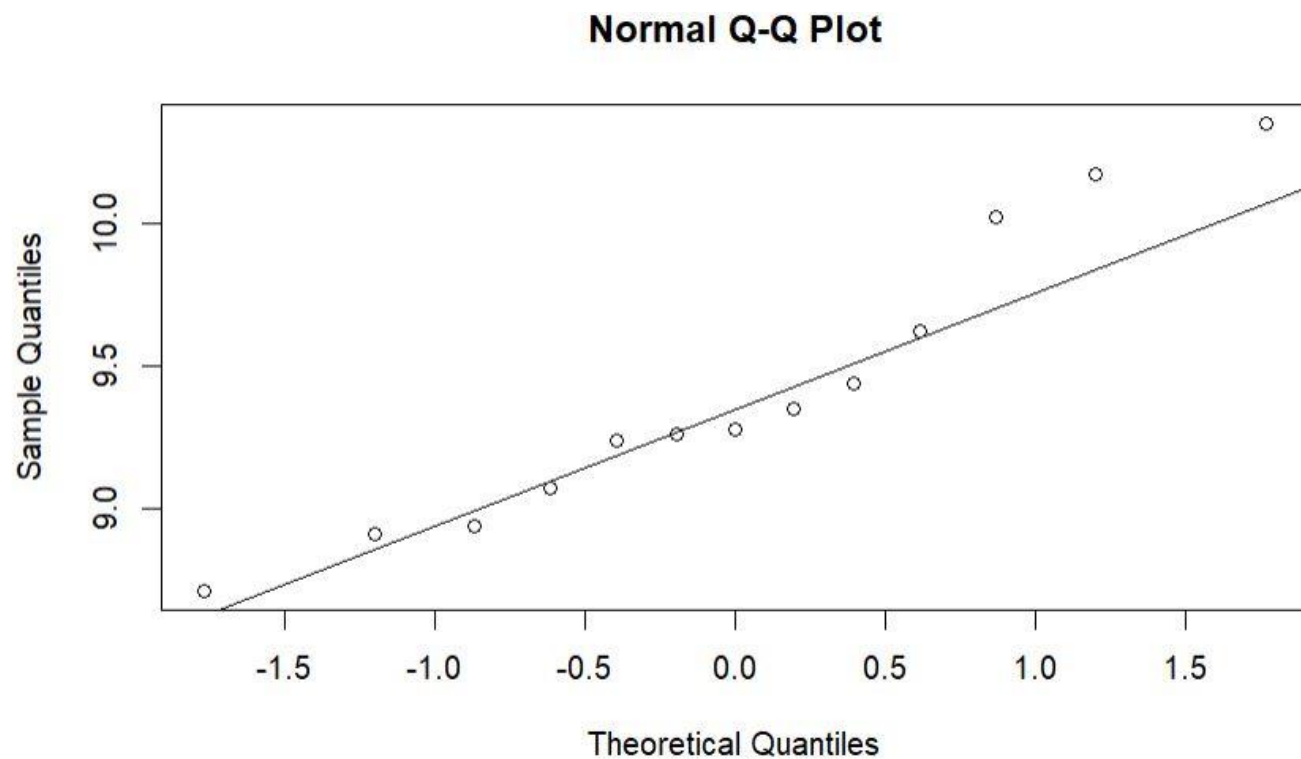
Shapiro-wilk normality test

data: data1\$rate

W = 0.95534, p-value = 0.6808

# 統計分析

資料是否屬於常態分佈(青年失業率)



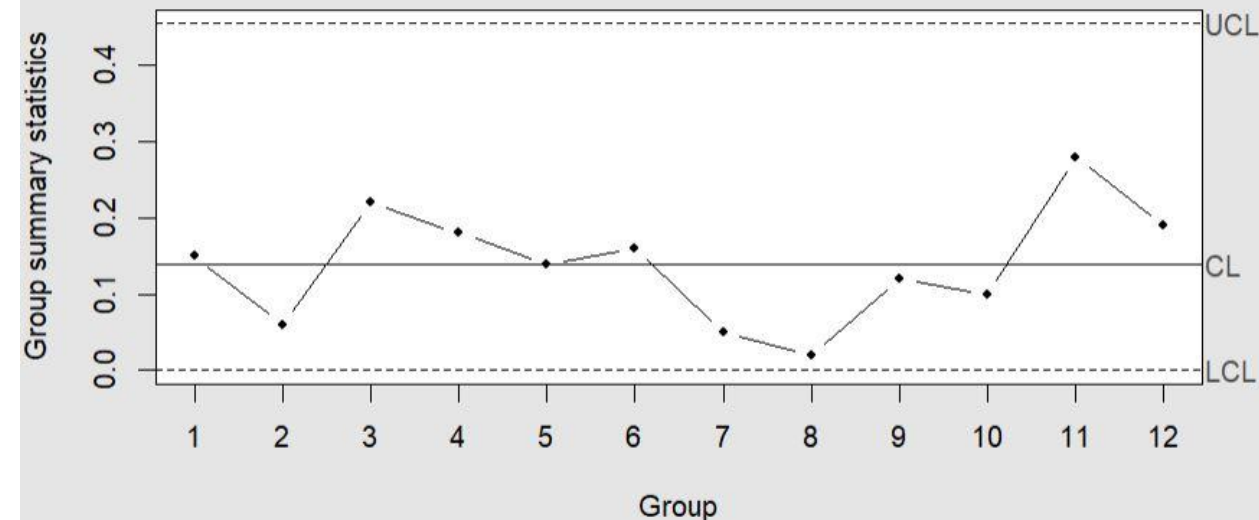
Shapiro-wilk normality test

```
data: data_year$rate  
W = 0.93125, p-value = 0.3539
```

# 總失業率

## 總失業率的管制圖

MR chart



Number of groups = 12

Center = 0.1391667

StdDev = 0.1233747

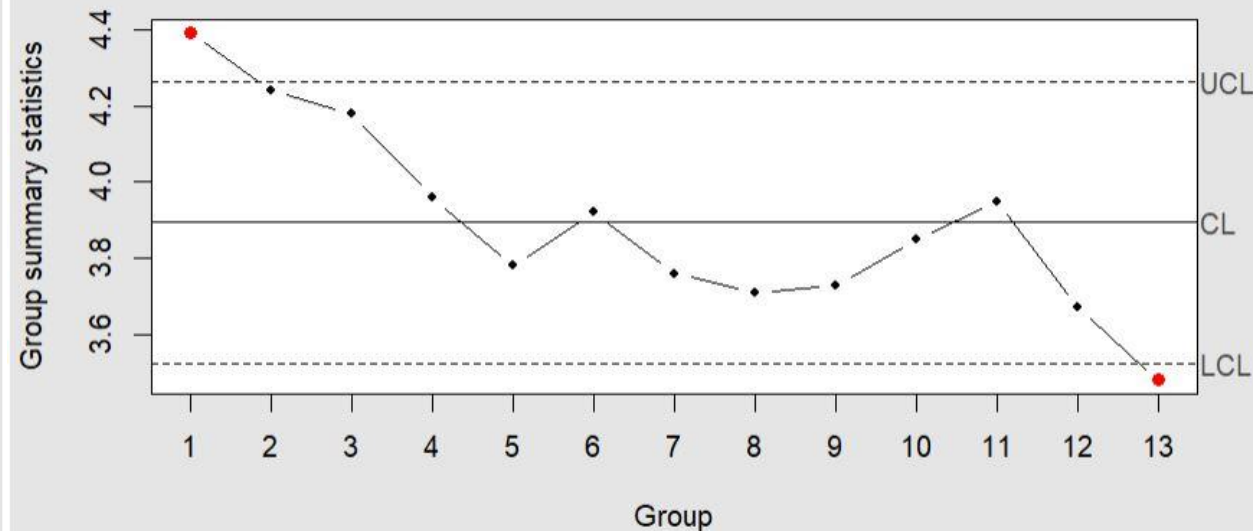
LCL = 0

UCL = 0.4546987

Number beyond limits = 0

Number violating runs = 0

I chart



Number of groups = 13

Center = 3.893846

StdDev = 0.1233747

LCL = 3.523722

UCL = 4.26397

Number beyond limits = 2

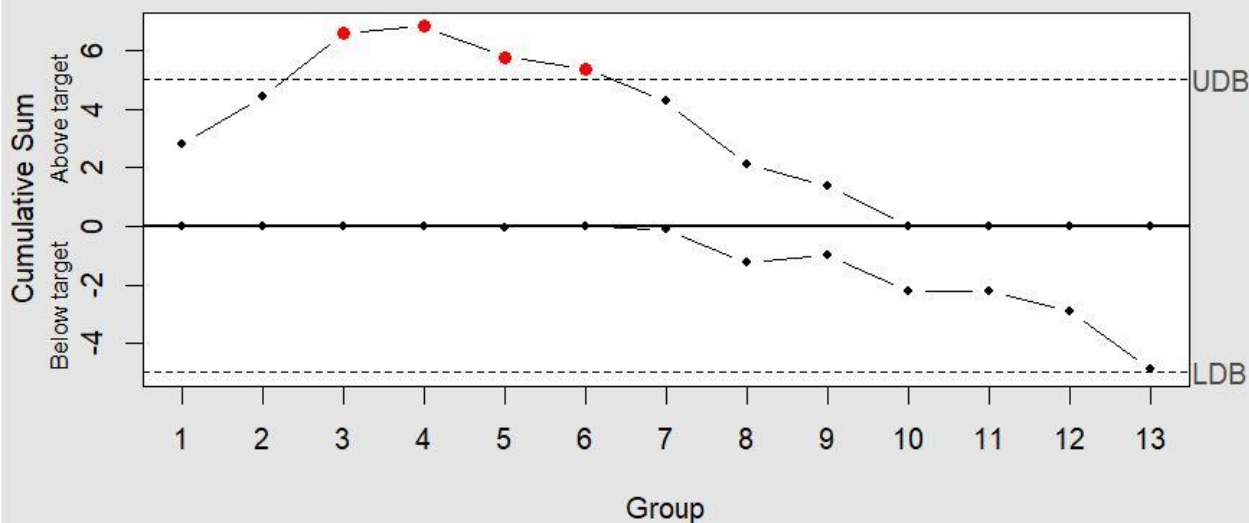
Number violating runs = 0

- 資料以年為單位( $n=1$ )，故使用I-MR chart來分析資料，可以看到在I chart裡面樣本100年跟112年都超出管制線外

# 總失業率

## 總失業率的管制圖

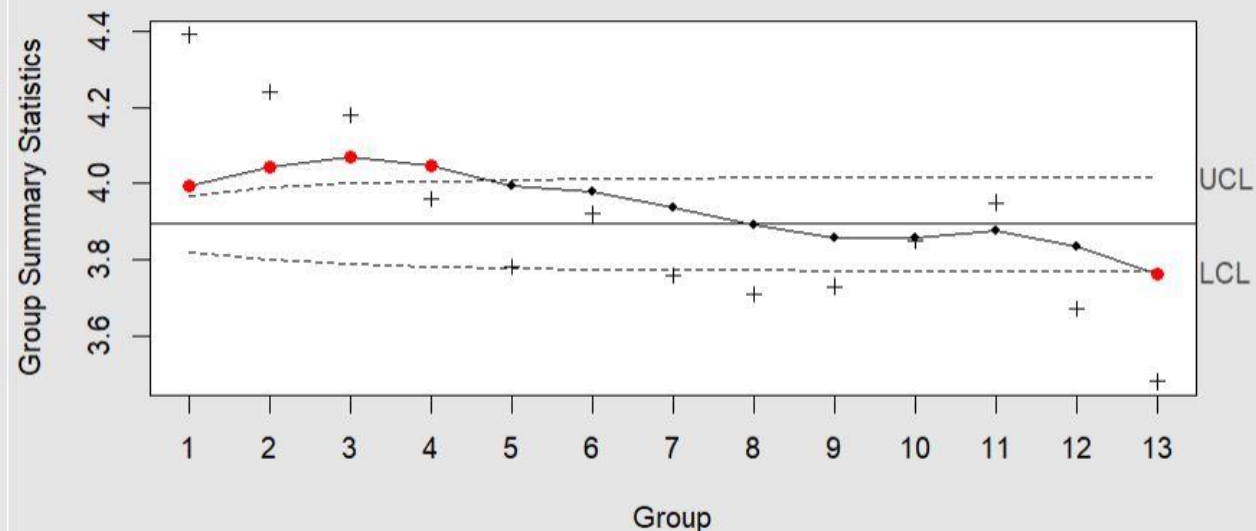
cusum Chart



Number of groups = 13  
Center = 9.412308  
StdDev = 0.2851655

Decision interval (std. err.) = 5  
Shift detection (std. err.) = 1  
No. of points beyond boundaries = 4

EWMA



Number of groups = 13  
Center = 3.893846  
StdDev = 0.1233747

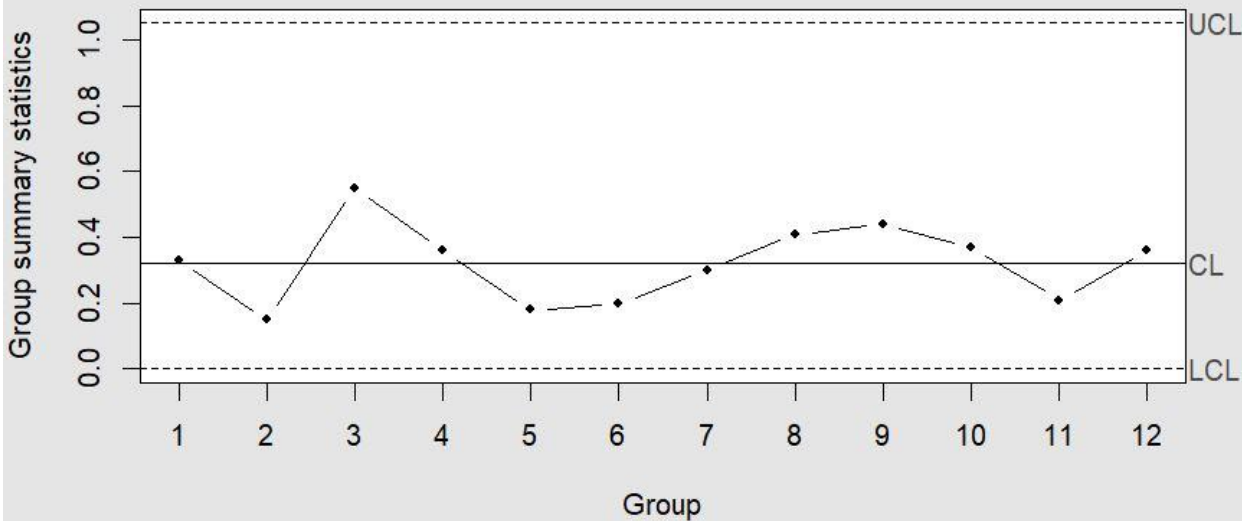
Smoothing parameter = 0.2  
Control limits at 3\*sigma  
No. of points beyond limits = 5

- 兩張圖皆是比較前面的點有out of control的資訊，而EWMA在112年的數據也有問題

# 青年失業率

## 青年失業率的管制圖

MR Chart



Number of groups = 12

Center = 0.3216667

StdDev = 0.2851655

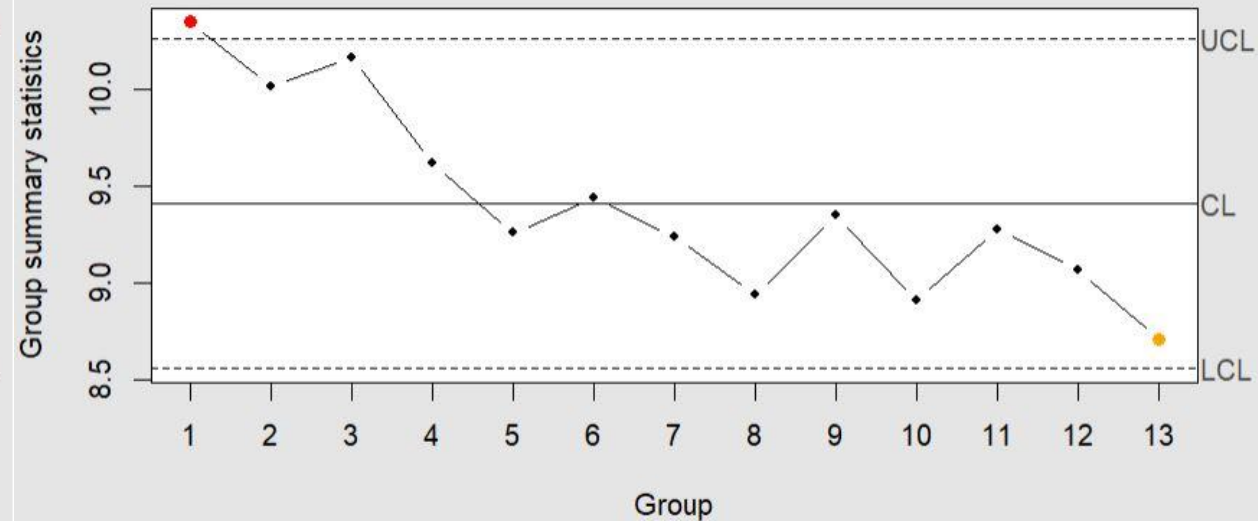
LCL = 0

UCL = 1.05098

Number beyond limits = 0

Number violating runs = 0

I Chart



Number of groups = 13

Center = 9.412308

StdDev = 0.2851655

LCL = 8.556811

UCL = 10.2678

Number beyond limits = 1

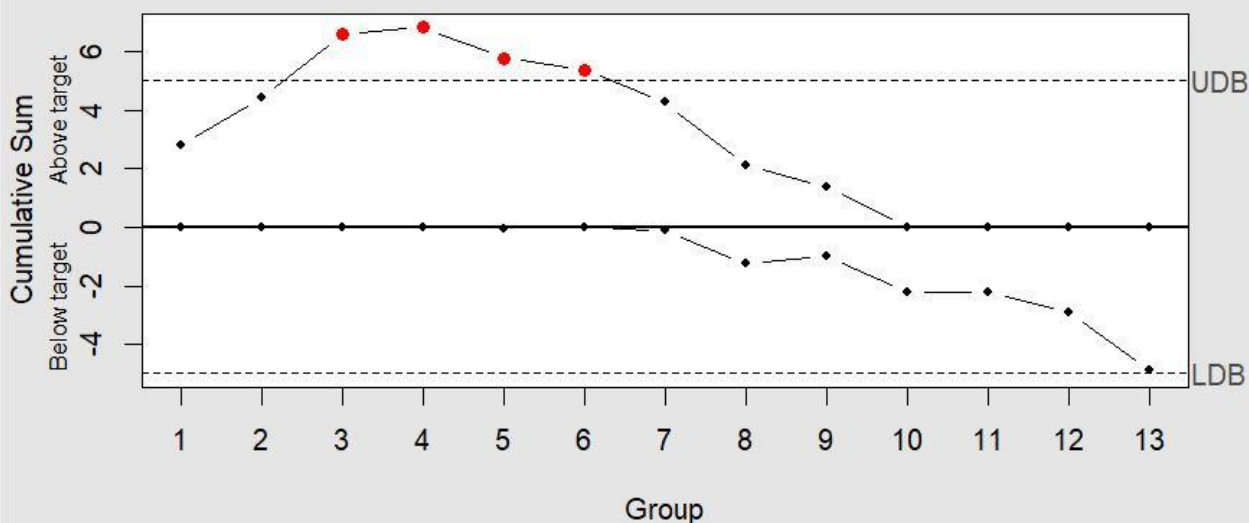
Number violating runs = 1

- 超出管制線的點與總失業率一樣，雖然112年的點有落在管制縣內，但還是有潛在的問題

# 青年失業率

## 青年失業率的管制圖

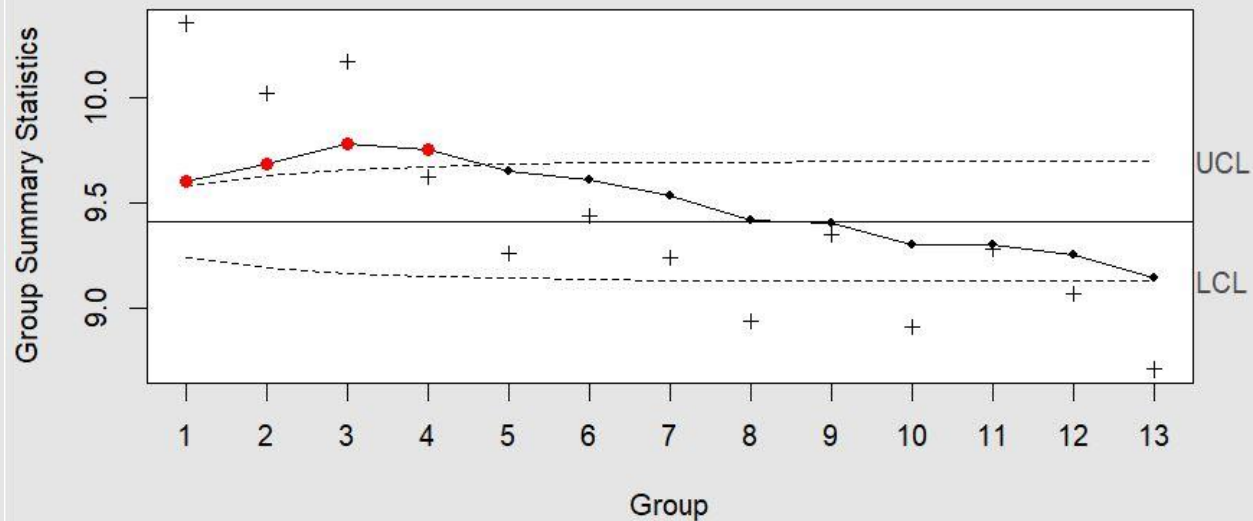
cusum Chart



Number of groups = 13  
Center = 9.412308  
StdDev = 0.2851655

Decision interval (std. err.) = 5  
Shift detection (std. err.) = 1  
No. of points beyond boundaries = 4

EWMA



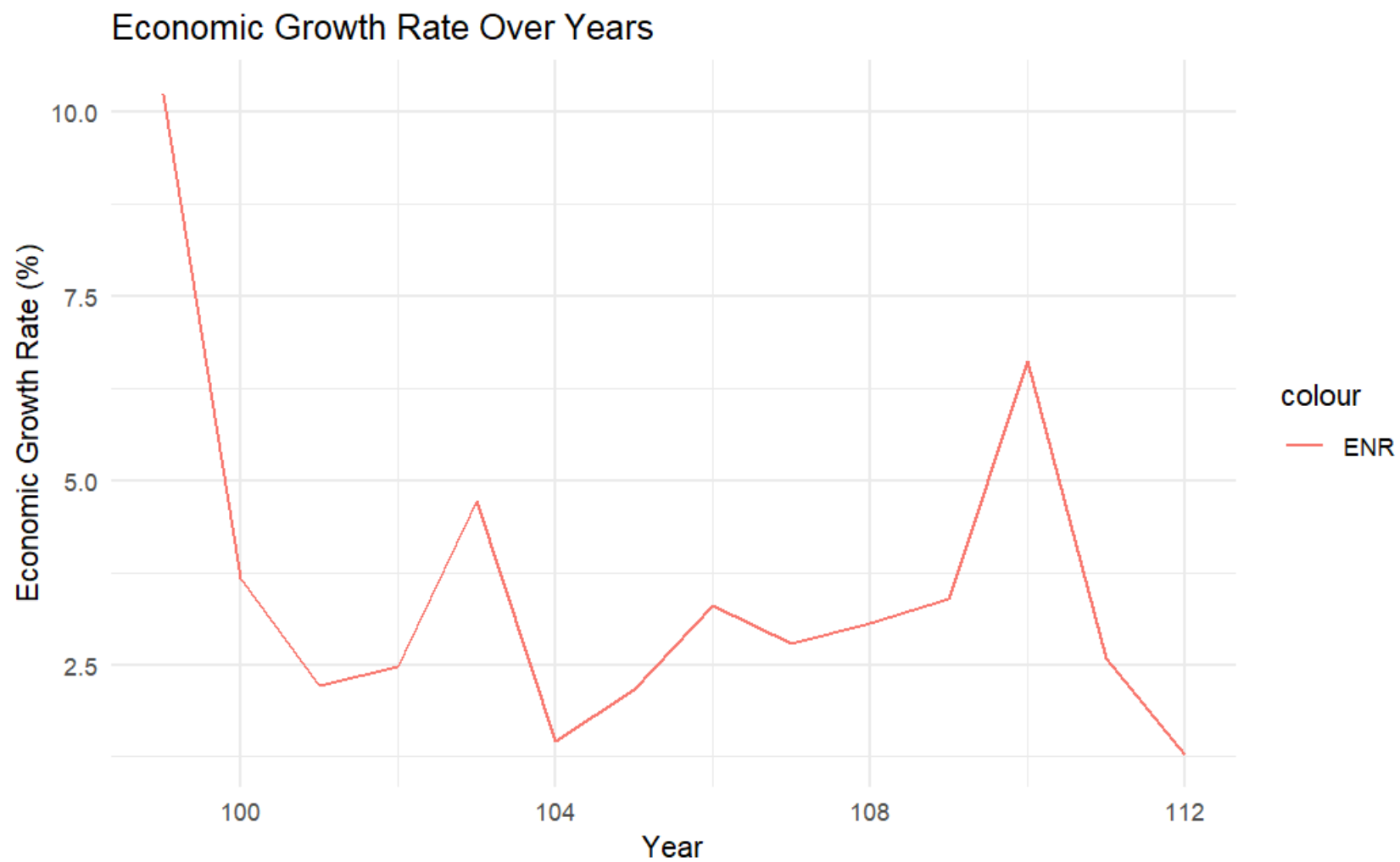
Number of groups = 13  
Center = 9.412308  
StdDev = 0.2851655

Smoothing parameter = 0.2  
Control limits at 3\*sigma  
No. of points beyond limits = 4

- 與總失業率的狀況一樣，這也說明不管是青年失業率或總失業率，兩個趨勢其實是相近的(年平均)

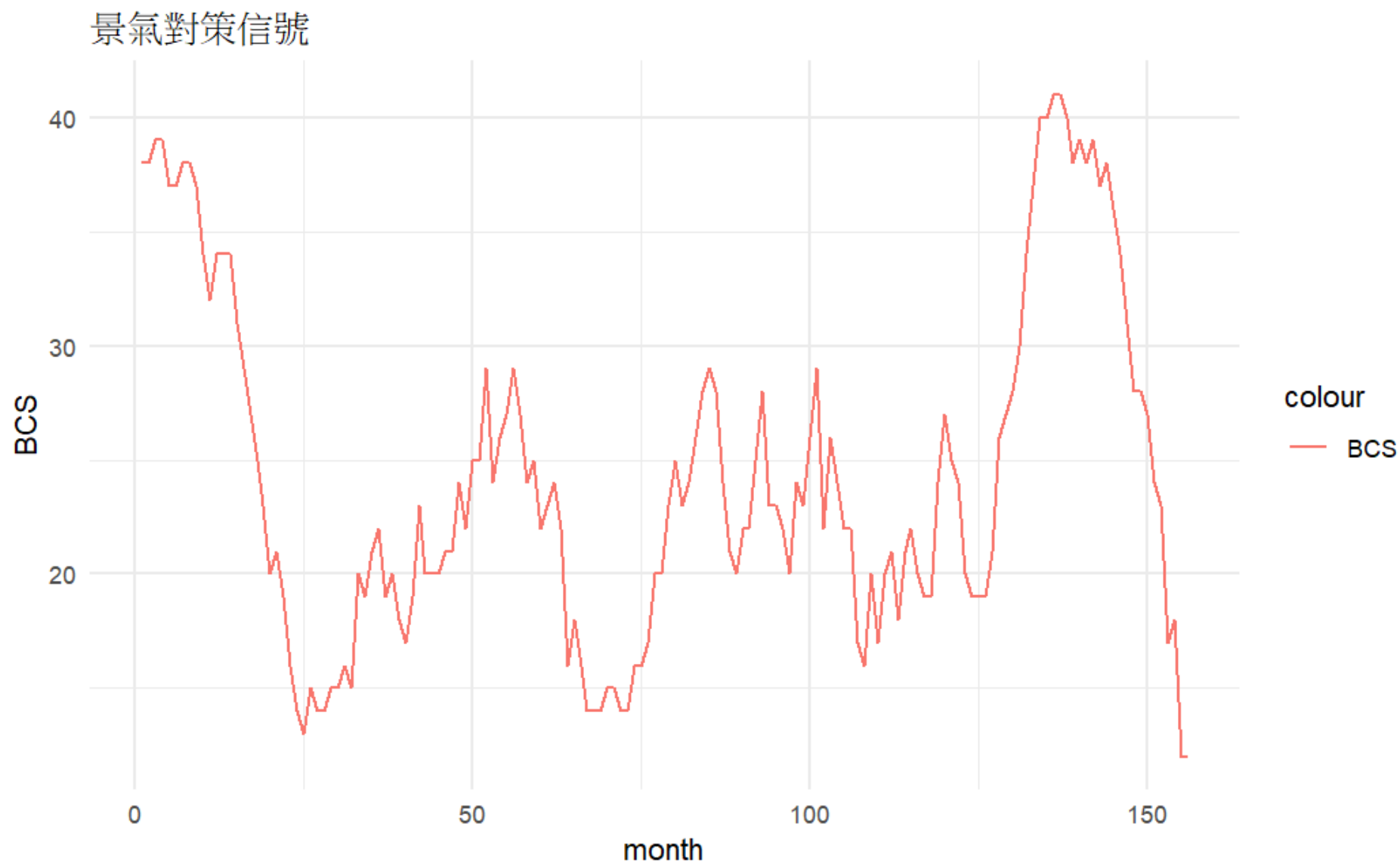


# 經濟成長率(%)





# 景氣對策信號



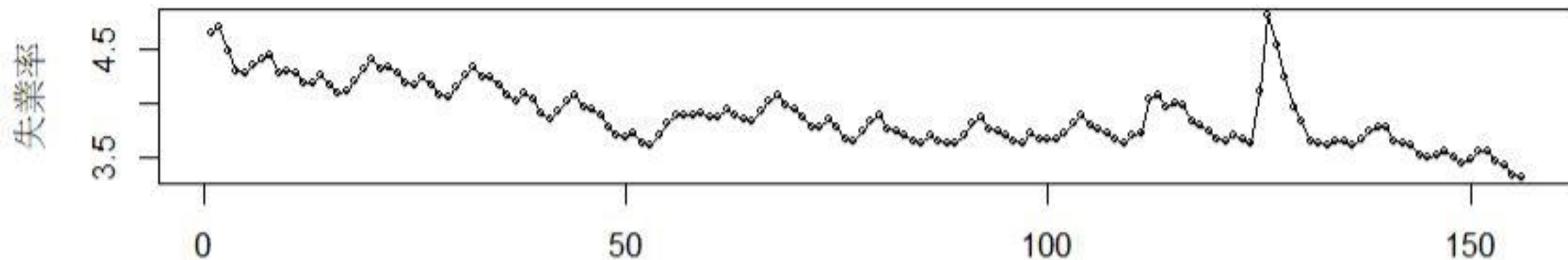


## PART 03

# 時 間 序 列 模 型

# 總失業率

## time plot



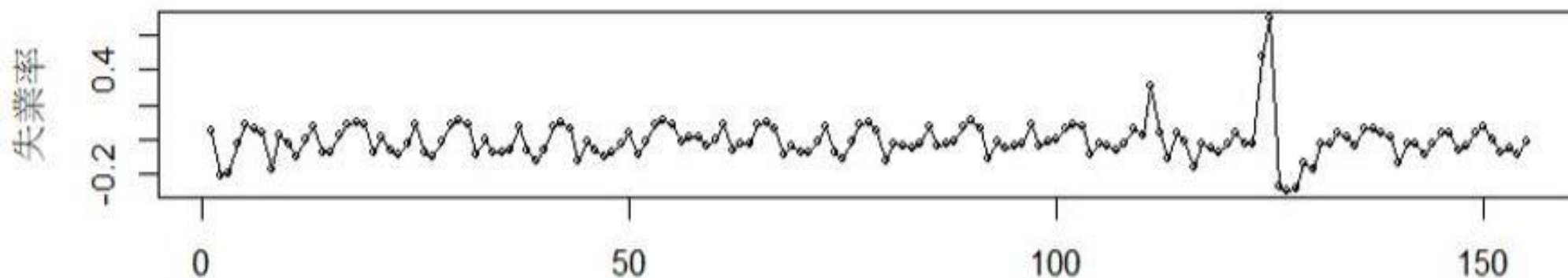
### Augmented Dickey-Fuller Test

```
data: data2$rate
Dickey-Fuller = -3.2342, Lag order = 5, p-value = 0.08475
alternative hypothesis: stationary
```

- 檢定結果的p value  $> 0.05$ ，代表序列是不平穩的

# 總失業率

time plot(經過一階差分)



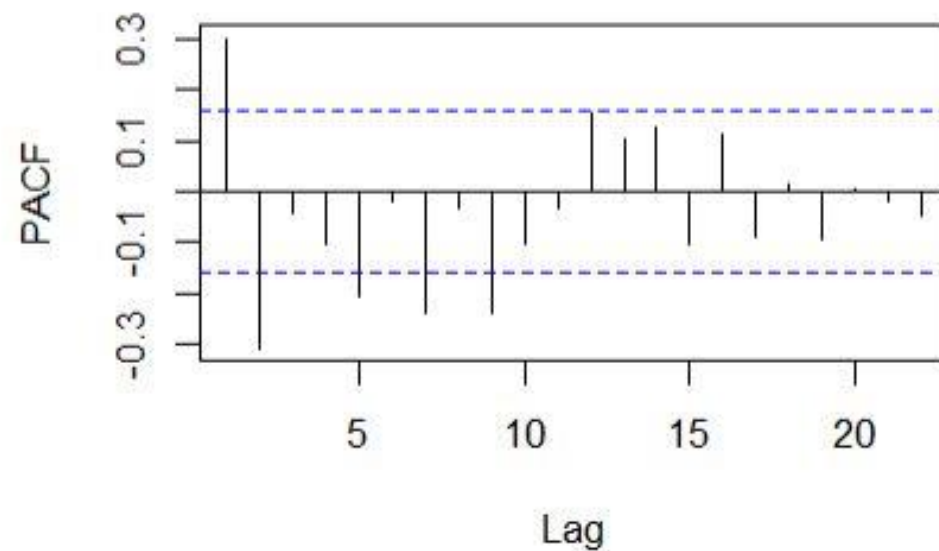
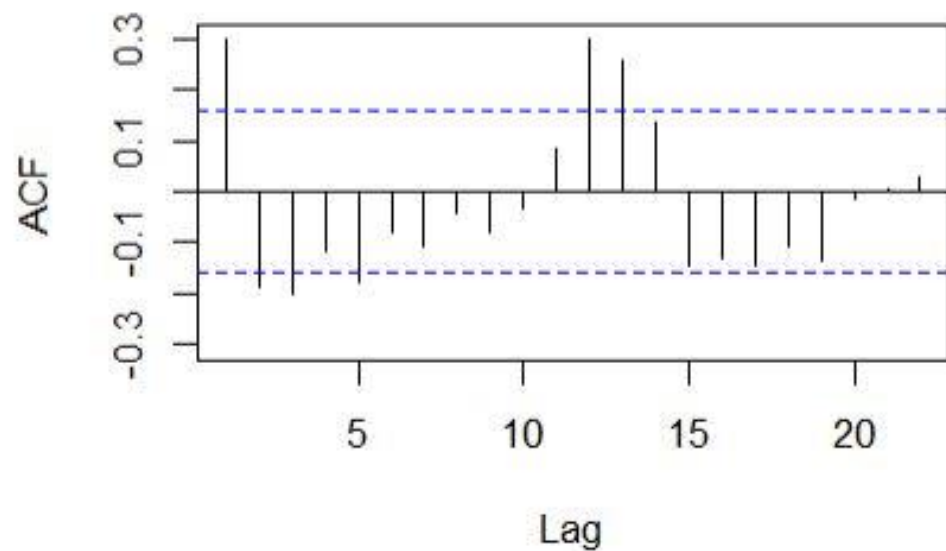
Augmented Dickey-Fuller Test

```
data: data2_diff  
Dickey-Fuller = -6.5451, Lag order = 5, p-value = 0.01  
alternative hypothesis: stationary
```

- 數據經過一階差分之後，檢定的p value < 0.05，代表序列是平穩的

# 總失業率

## Autocorrelation



- 從ACF與PACF還是能看到序列是有相關性的



# 總失業率

## ARMA模型

ARIMA(2,0,2) with non-zero mean

Coefficients:

	ar1	ar2	ma1	ma2	mean
	0.8993	-0.2844	-0.6017	-0.2619	-0.0063
s.e.	0.1444	0.1381	0.1427	0.1385	0.0028

$\sigma^2 = 0.00887$ : log likelihood = 148.36

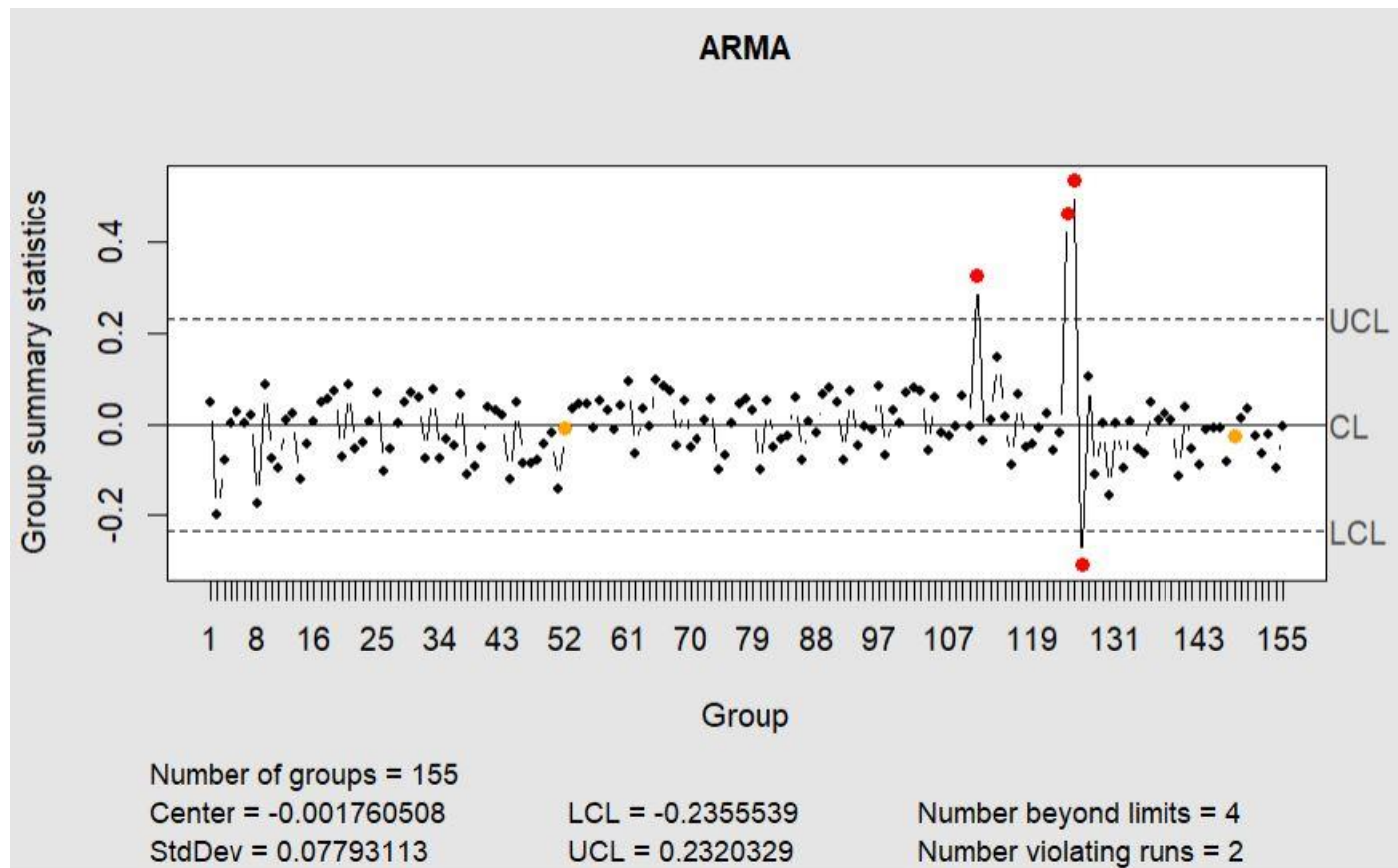
AIC=-284.73    AICc=-284.16    BIC=-266.47

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.001760508	0.09265147	0.06148228	NaN	Inf	0.7635702	-0.005132592

# 總失業率

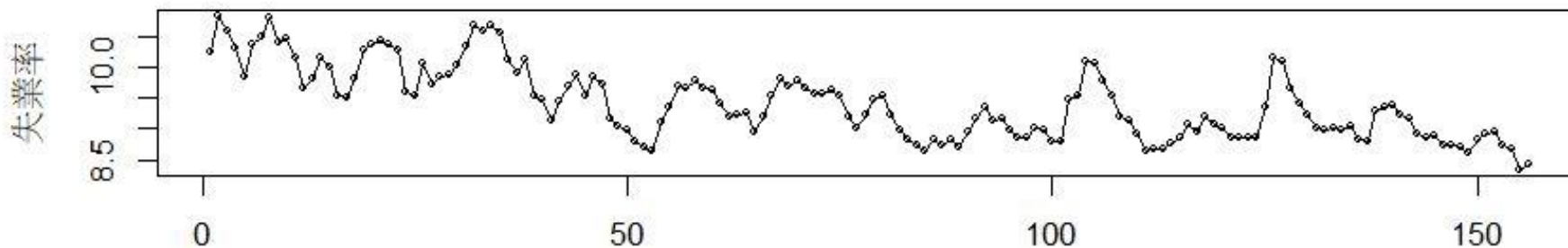
## 殘差管制圖



- 超出管制線的點代表實際的點與預測的點相差較大，那這幾個點都是在疫情爆發的時期，代表疫情對其影響是很重大的。

# 青年失業率

## time plot



### Augmented Dickey-Fuller Test

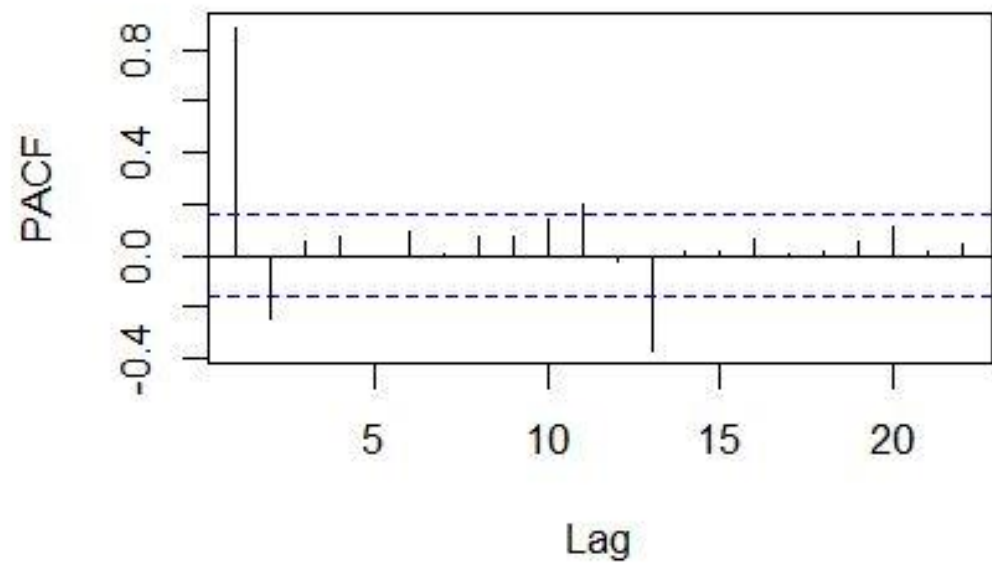
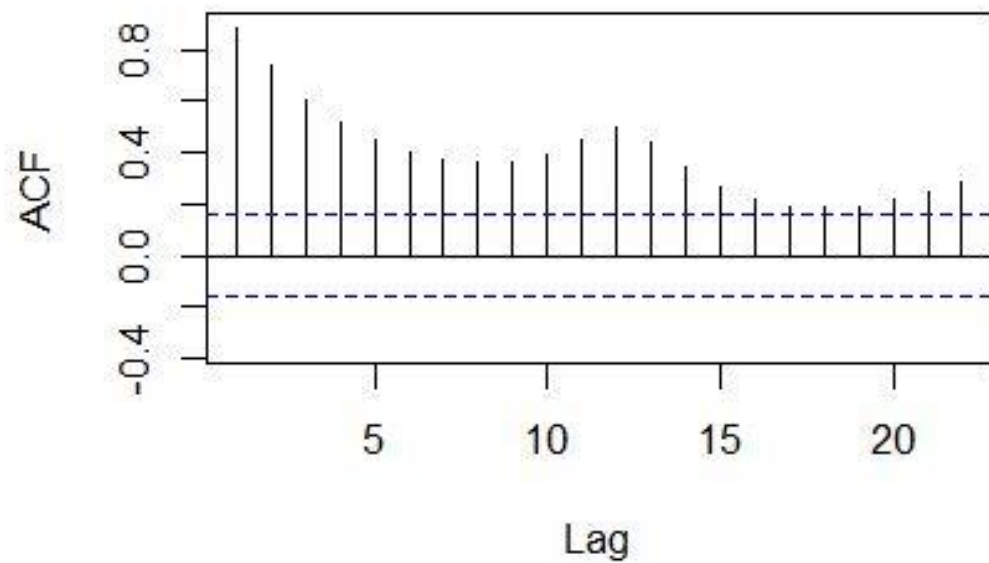
```
data: data_month$rate
Dickey-Fuller = -4.101, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

- 從檢定可以得到序列是平穩的，所以就不用對其做額外的處理



# 青年失業率

## Autocorrelation



- 從圖中可以看到序列是有自相關的



# 青年失業率

## ARMA模型

```
Call:
arima(x = data_month$rate, order = c(3, 0, 1))
```

Coefficients:

	ar1	ar2	ar3	ma1	intercept
	0.1735	0.8807	-0.2914	0.9904	9.4032
s.e.	0.0775	0.0331	0.0796	0.0151	0.1515

sigma^2 estimated as 0.05413: log likelihood = 4.41, aic = 3.19

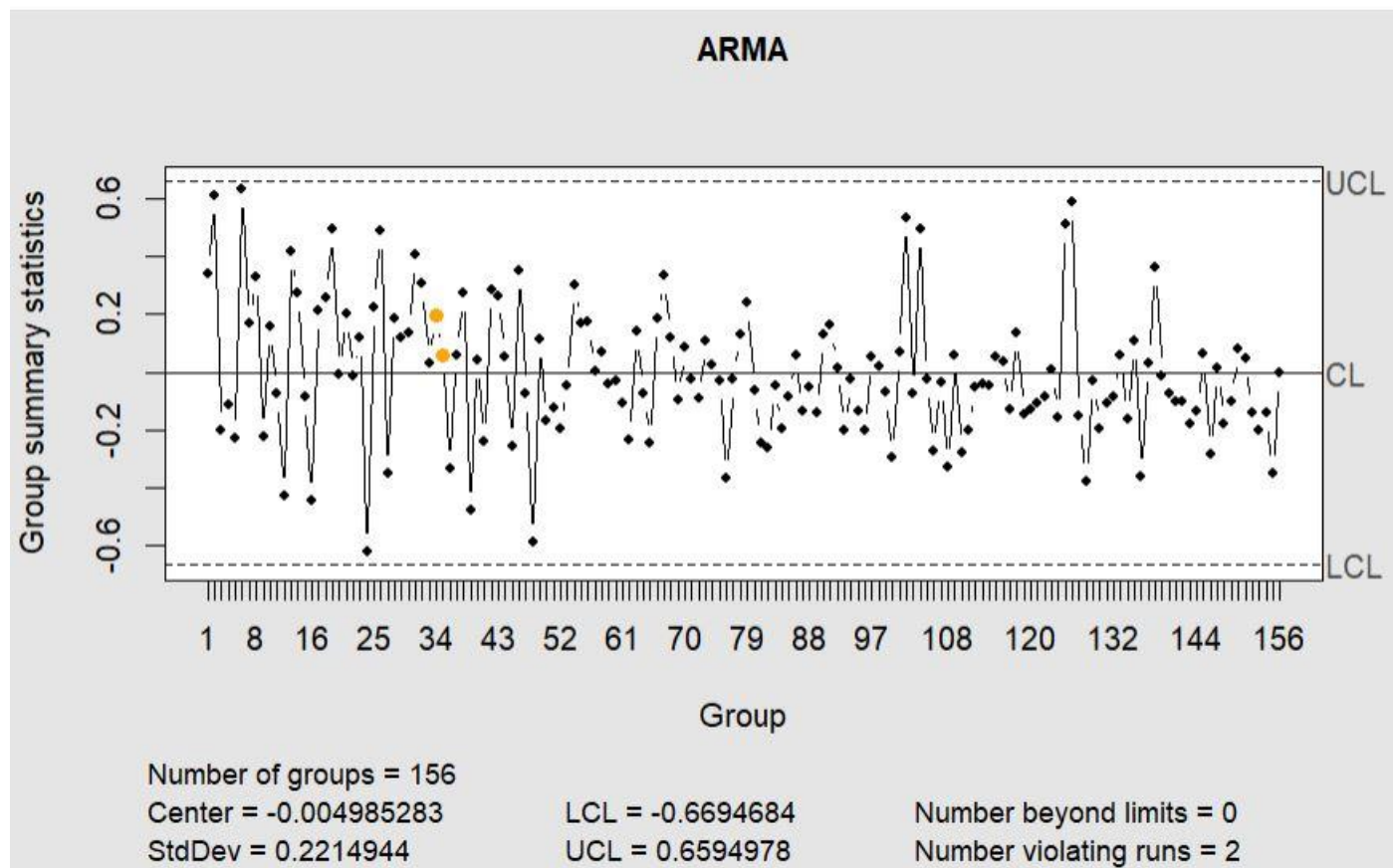
Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.004985283	0.2326594	0.1798736	-0.1137355	1.892037	0.9262594	0.03428081



# 青年失業率

## 殘差管制圖



- 從圖來看所有點都落在管制線內，我們也可以知道相比於總失業率來看，疫情對其的影響是比較低的。



# PART 04

## 結 論



## 結論

- 以年為單位來看的話，雖然青年失業率相較總失業率高出不少，但總失業率與青年失業率的總體趨勢其實是差不多的。
- 利用時間序列模型之後，可以發現總失業率在疫情嚴重期間會有預測比較不穩定的狀況，那相較於青年失業率比較沒有影響
- 後續可以利用比較多年的資料來探討兩者的差異，能更了解兩者的趨勢



# References

<https://www.stat.gov.tw/>

<https://statfy.mol.gov.tw/index02.aspx>

[https://medium.com/@cindy050244\\_52136/%E6%99%82%E9%96%93%E5%BA%8F%E5%88%97%E6%8E%A2%E7%B4%A2-%E4%BA%8C-arima%E5%AE%B6%E6%97%8F%E7%B0%A1%E4%BB%8B-8d533f0b18d6](https://medium.com/@cindy050244_52136/%E6%99%82%E9%96%93%E5%BA%8F%E5%88%97%E6%8E%A2%E7%B4%A2-%E4%BA%8C-arima%E5%AE%B6%E6%97%8F%E7%B0%A1%E4%BB%8B-8d533f0b18d6)

<https://zh.wikipedia.org/zh-tw/2021%E5%B9%B4%E4%B8%AD%E8%8F%AF%E6%B0%91%E5%9C%8B%E5%85%A8%E5%9C%8B%E7%96%AB%E6%83%85%E7%AC%AC%E4%B8%89%E7%B4%9A%E8%AD%A6%E6%88%92>

The background features several organic, flowing shapes. In the top left, a dark brown shape with a fine, repeating geometric pattern is partially visible. In the top right, a dark green shape with a similar pattern is present. The bottom left and bottom right corners contain light beige or tan shapes, also with the same pattern. Thin, wavy lines in a light brown or tan color meander across the white background, some passing through the patterned shapes.

**Thanks for listening**