

統計品管期末書面報告

台灣近十年失業率之探討

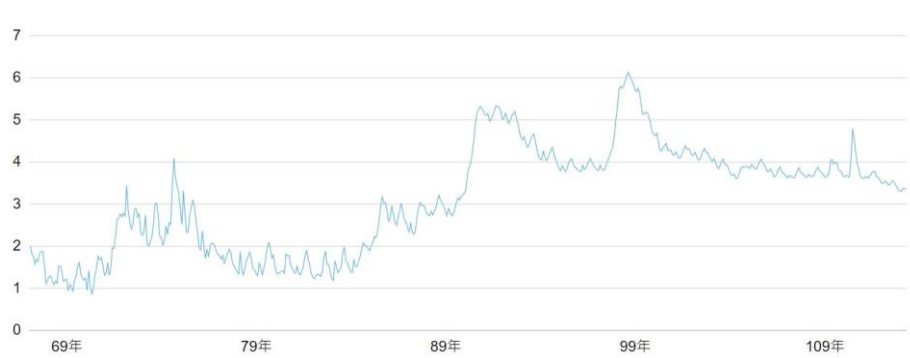
一、研究動機

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。在台灣，過去十年來的失業率變化反映出多方面的經濟動態，本報告著重在比較總失業率及青年失業率的趨勢，並探討影響失業率的重要事件。

此外，本報告將觀察其他經濟指標與失業率之間的關係，以期揭示更全面的經濟健康狀況評估方法。這不僅有助於理解失業率的變動規律，還可以為政策制定者提供有價值的參考，促進更有效的經濟政策制定。

二、資料介紹與初步分析

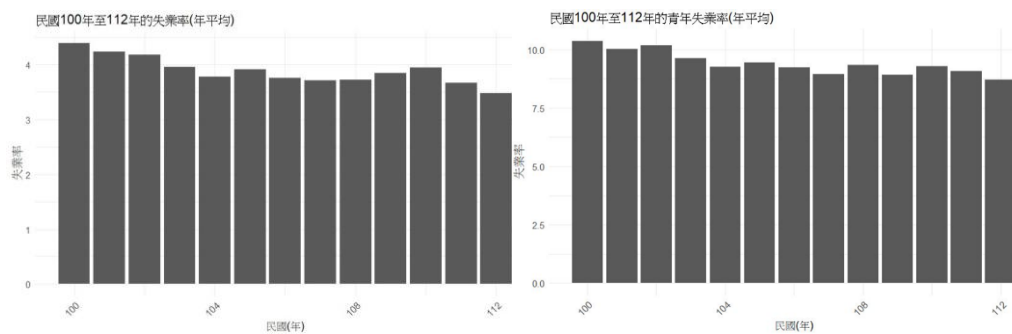
1. 資料名稱:總失業率與青年失業率，其中青年失業率的範圍為 15~29 歲。
2. 資料來源:中華民國統計資訊網、勞動部統計查詢網。
3. 資料時間:民國 100 年至 112 年，共 13 年。
4. 資料單位:月單位及年單位。
5. 資料初步分析:
 - I. 民國 90 年以前的平均失業率遠低於民國 90 年以後，欲將管制圖應用於失業率，因而選擇近幾年的失業率資料（民國 100 年至 112 年失業率資料）。



圖中可見，民國 99 年及民國 110 年有出現失業率高峰的現象。

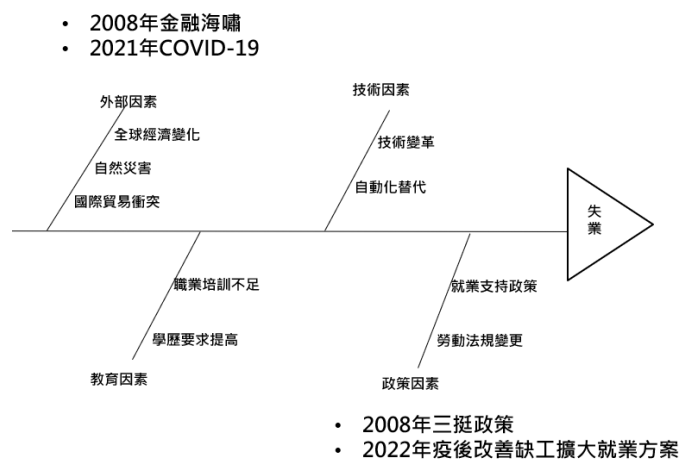
II. 選擇民國 100 年至 112 年總失業率及青年失業率資料。

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

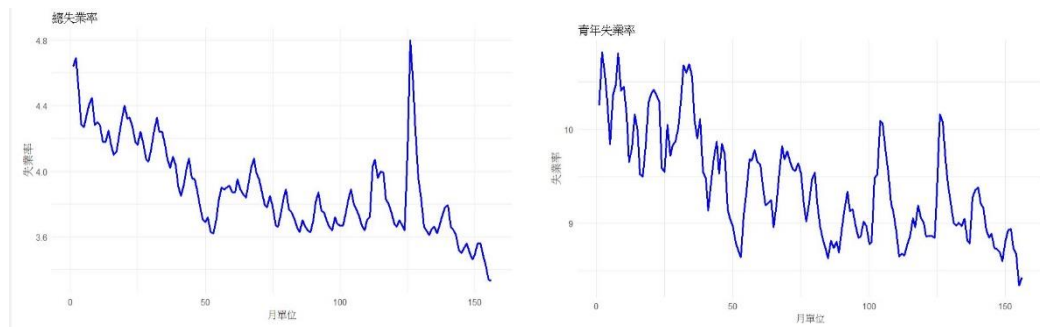


圖中可見，前幾年有失業率較高的情況。

III. 失業率因果圖



IV. 總失業率與青年失業率比較

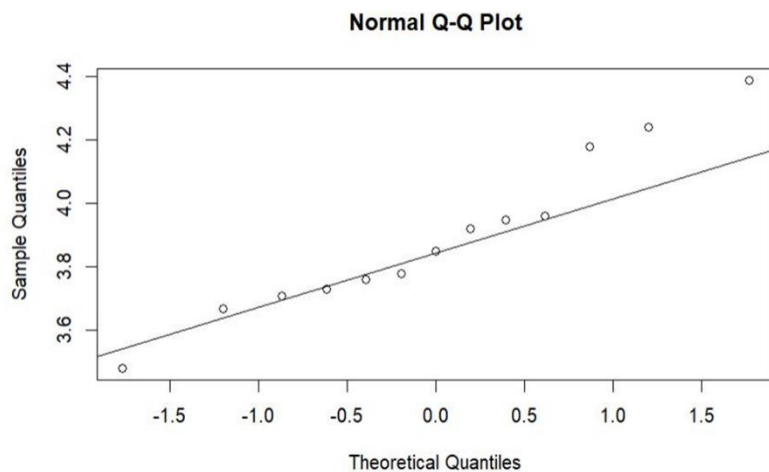


可以看到兩個趨勢的後半段來看是比較不一樣的，後續會試著去找尋原因

三、 統計分析與研究方法

1. 資料是否屬於常態分佈

I. 總失業率

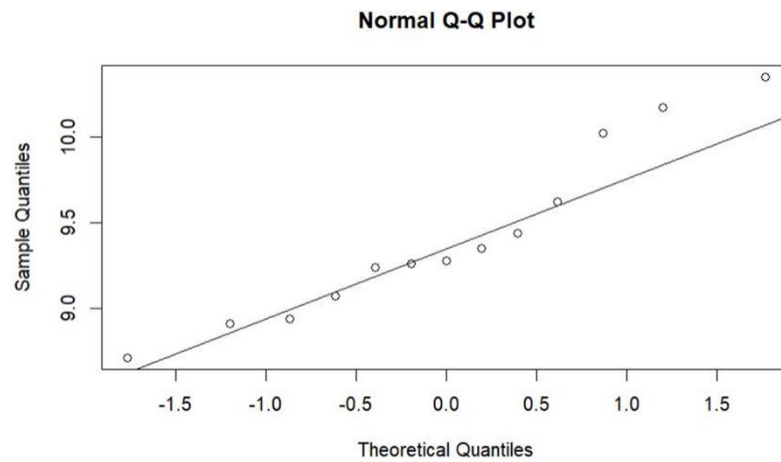


Shapiro-wilk normality test

```
data: data1$rate  
W = 0.95534, p-value = 0.6808
```

檢定結果: $p\text{-value} > 0.05$ ，表示資料服從常態假設。

II. 青年失業率



Shapiro-wilk normality test

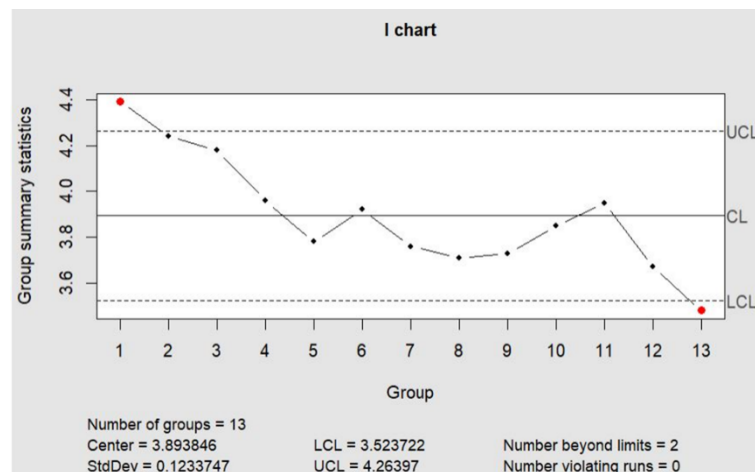
```
data: data_year$rate
W = 0.93125, p-value = 0.3539
```

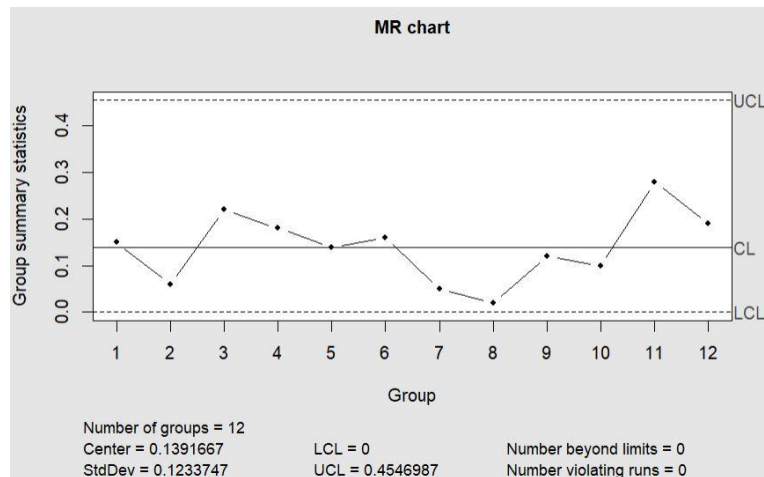
檢定結果: $p\text{-value} > 0.05$ ，表示資料服從常態假設。

2. 總失業率管制圖

I. I-MR chart

資料以年為單位($n=1$)，故使用 I-MR chart 來分析資料。



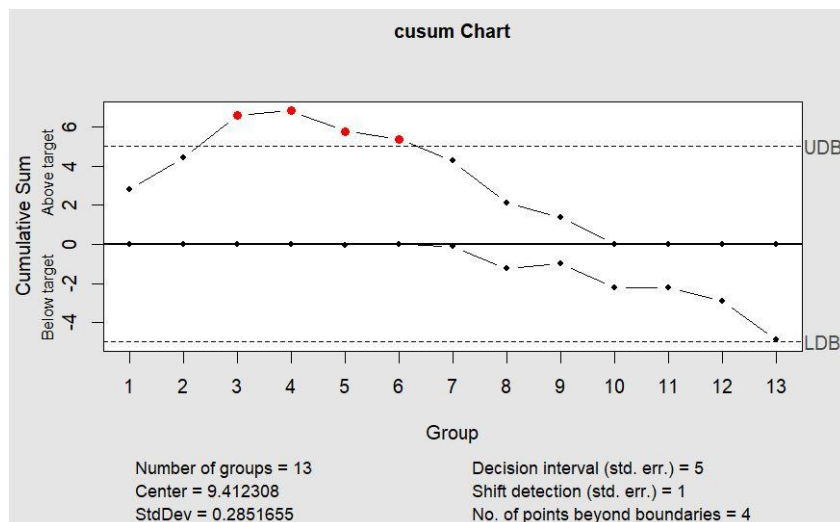


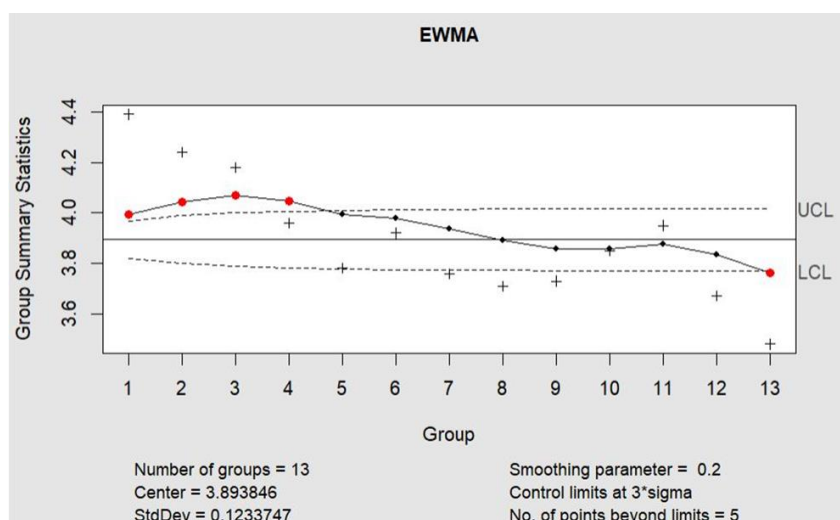
由 I chart 可見，樣本 1（民國 100 年）及樣本 13（民國 112 年）落在管制線之外。

由 I chart 可見，樣本 1 落在管制線之外且樣本 2、3 接近上管制線，經由查詢資訊及推測認為是由於西元 2008 年金融風暴後導致失業率異常飆升，而樣本 1、2、3 仍處於金融風暴後失業率校正回歸的時間區段，因而失業率較高。

由 I chart 可見，樣本 13 落在管制線之外，經由查詢資訊及推測認為是由於民國 110 年 COVID-19 疫情嚴峻導致三級警戒，造成失業率增高，而後疫情趨緩解封後，大量缺工的情況發生，導致樣本 13 失業率低的情形。

II. Cusum chart and EWMA

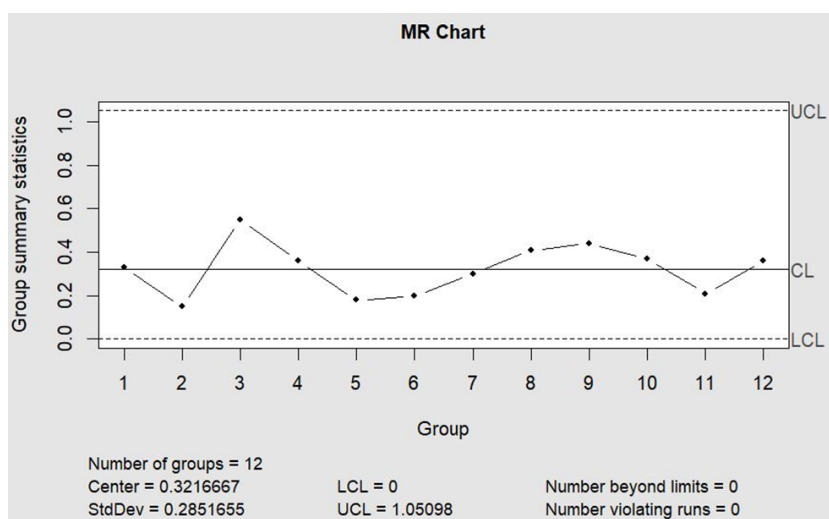
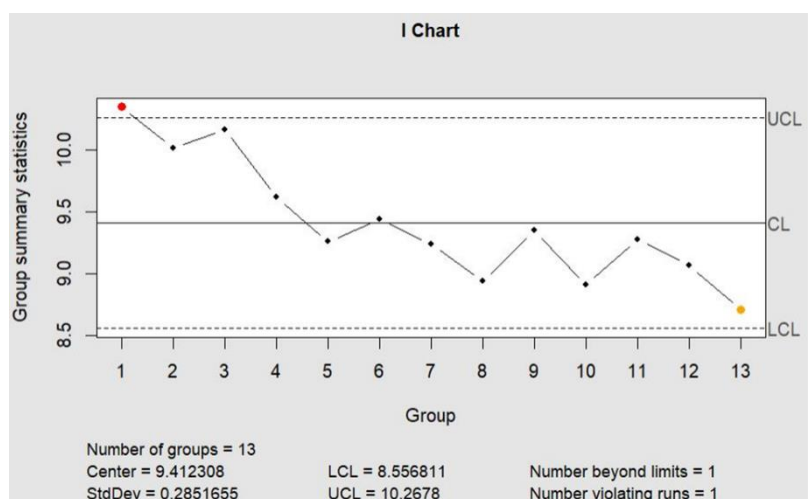




兩張圖皆是前面的點有 out of control 的資訊，說明前幾年的失業率明顯高於平均，也反映前面金融風暴後失業率校正回歸的推測。

3. 青年失業率管制圖

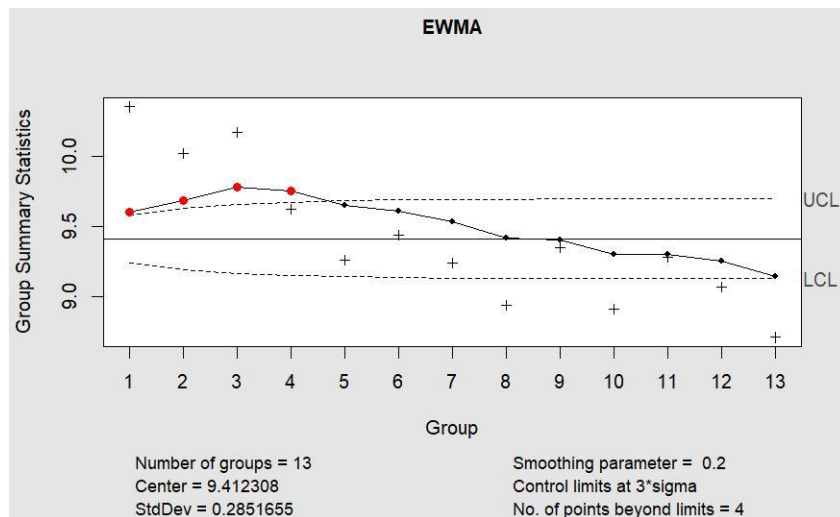
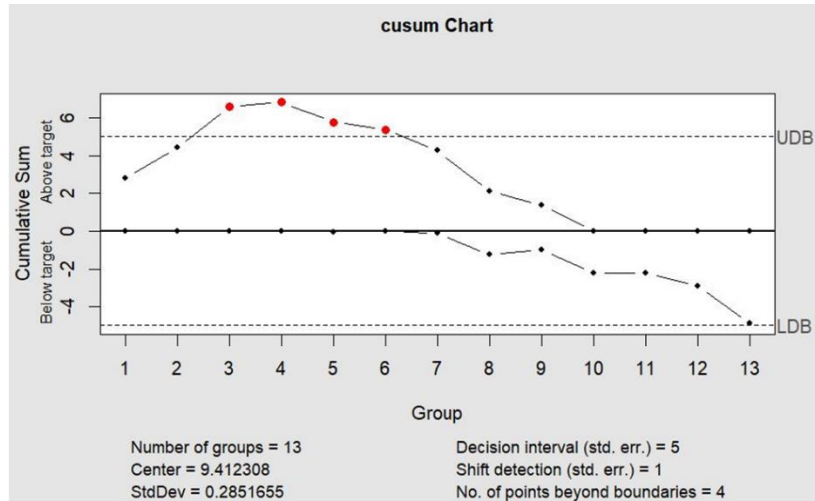
I. I-MR chart



由 Ichart 可見，樣本 1（民國 100 年）落在管制線之外，且樣本 13（民國 112 年）接近下管制線。

Out of control 的原因與總失業率相同。

II. Cusum chart and EWMA



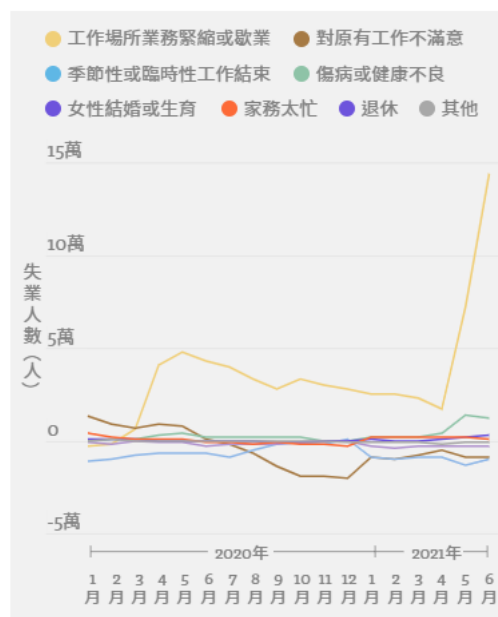
與總失業率的狀況一樣，這也說明不管是青年失業率或總失業率，兩個趨勢其實是相近的(年平均)

4. 影響失業率之事件



疫後增加失業人數，按失業原因分類（限於非初次尋職者）

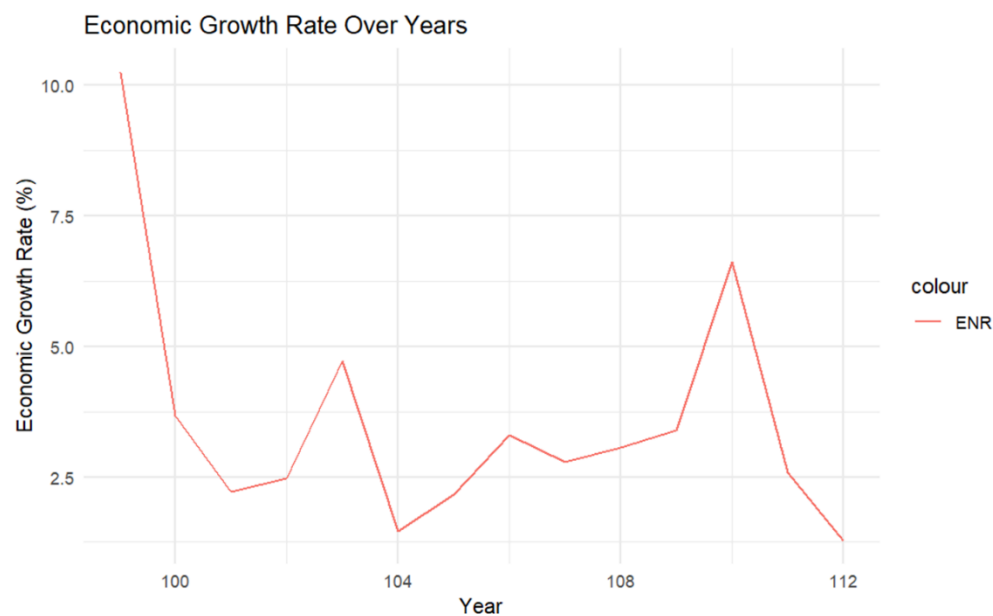
與2019年同月份相比



資料來源：主計總處人力資源調查統計月報

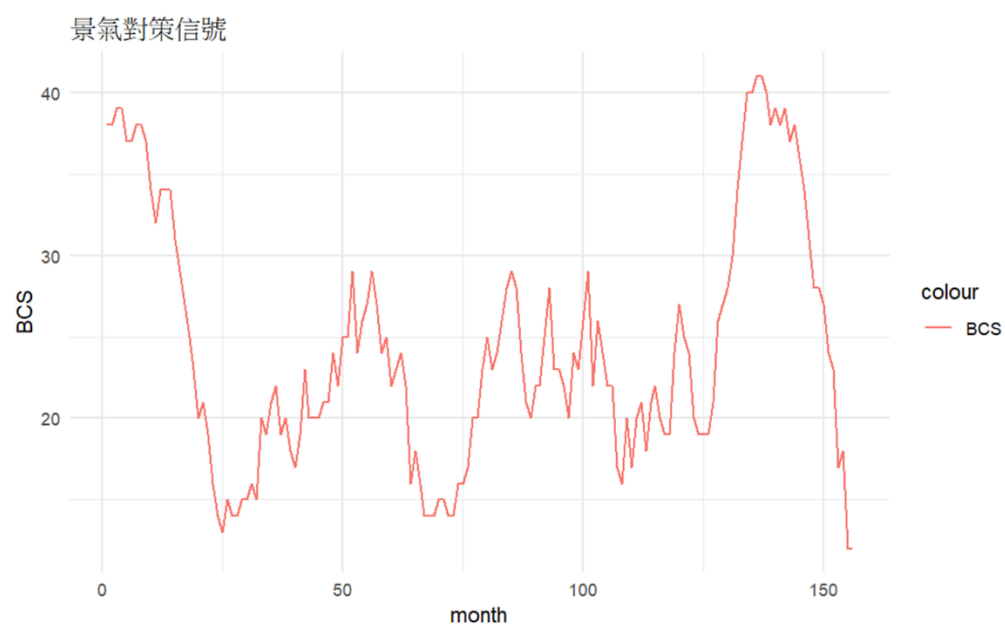
5. 比較其他經濟指標

I. 經濟成長率(%)



經濟成長率下降劇烈的區段，失業率較高，如民國 99 至 100 年及民國 110 年至 111 年。

II. 景氣對策信號

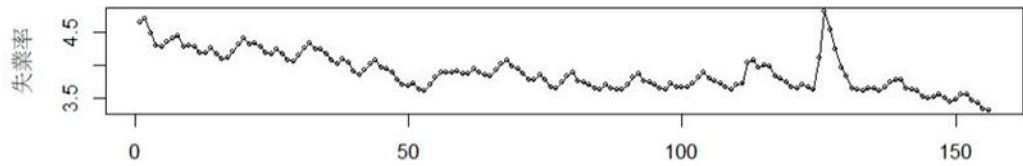


景氣對策信號下降劇烈的區段，失業率較高，如民國 99 至 101 年及民國 110 年至 111 年。

四、 時間序列模型

1. 總失業率

I. Time plot



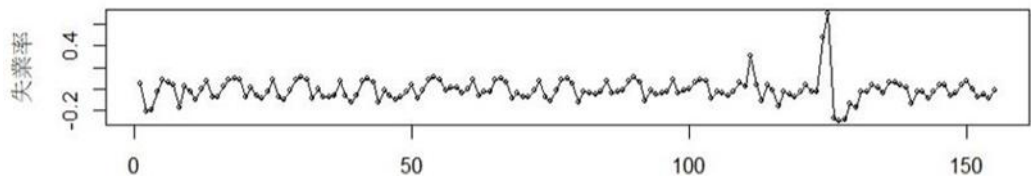
可以發現除了後面波動較大的部分以外，是可以看出有一個週期性的感覺，而且是慢慢下降的。

Augmented Dickey-Fuller Test

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

我們也透過 ADF 檢定，得到 p value 為 0.085，代表序列不是平穩，處理方式就是對序列做一階差分。

II. Time plot (經過一階差分)



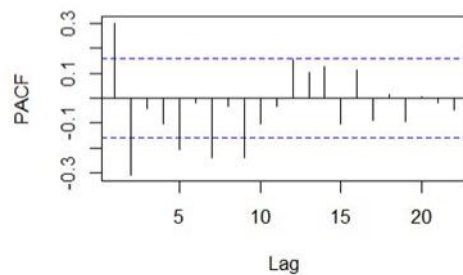
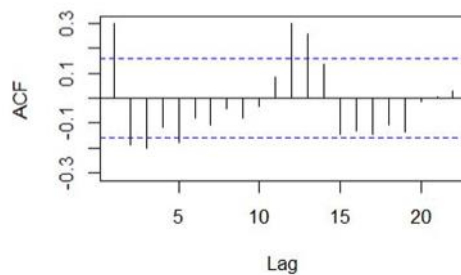
可以發現除了後面波動比較大的部分外，已看不出有什麼趨勢。

Augmented Dickey-Fuller Test

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

透過 ADF 檢定，得到 p-value 為 0.01，代表序列經過一階差分之後已經變平穩了。

III. 序列自相關性(經過一階差分)



從兩張圖來看，序列是有自相關的，所以後續會利用 ARMA 模型來處理自相關的問題。

IV. 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

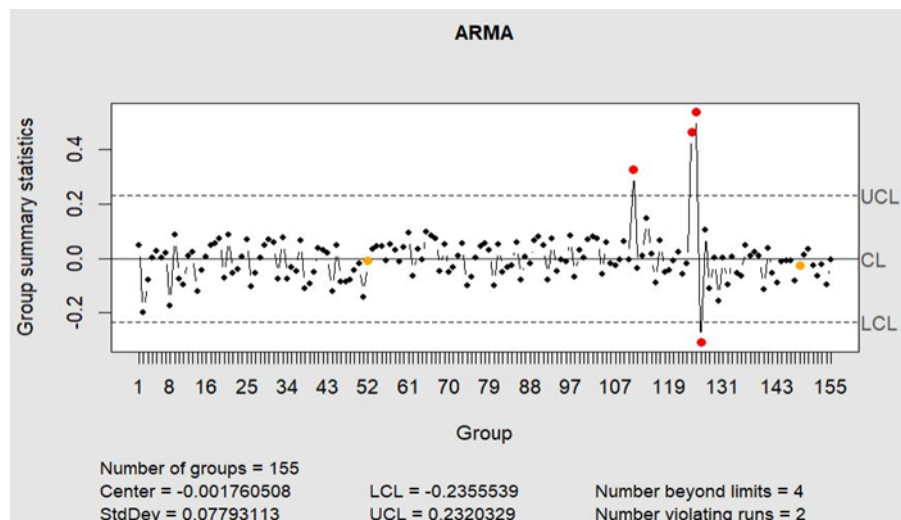
sigma^2 = 0.00887: log likelihood = 148.36

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

模型的數據是有經過一階差分，利用套件得到的模型，係數為 AR(2)、MR(2)。

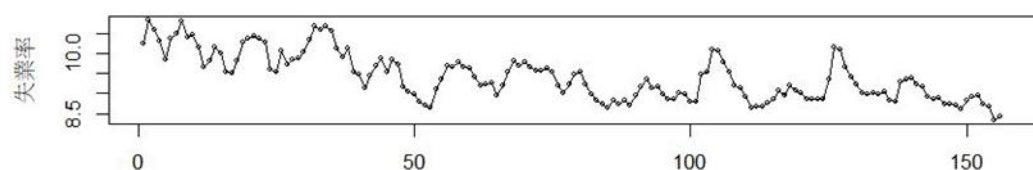


這是利用殘差所畫出來的管制圖，可以發現有 4 個超出管制線的點。第一個點是在 2020 年 3~4 月，是疫情剛始便嚴重的時期，那時候政策有：全國酒店、舞廳停止營業、觀光景點限制人流、社交安全距離。

後面三個點為疫情三級解戒的那段時期(2021 年 5 月中開始實施)，所以導致其失業率波動較大，以上可以知道疫情對總失業率是有很大影響的。

2. 青年失業率

I. Time plot



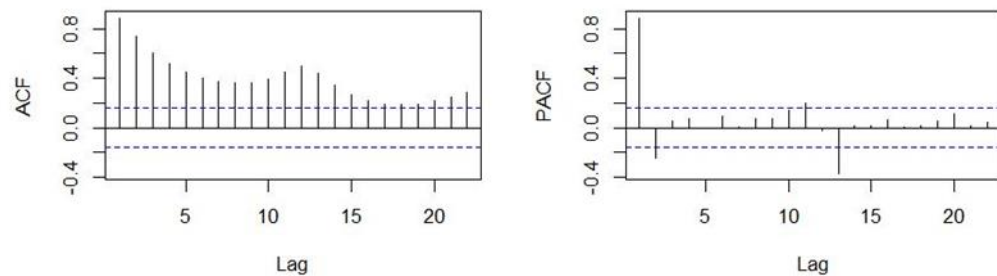
從圖來看是沒有發現有任何週期，但也是有慢慢下降的趨勢的。

Augmented Dickey-Fuller Test

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

利用 ADF 檢測來驗證，得到的 p-value 也是小於 0.05 的，代表序列是穩定的，所以現階段我們就沒有額外再做任何處理。

II. 序列自相關性



從兩張圖都可以看出序列是有相關性的，所以後續會用 ARMA 模型來處理。

III. ARMA 模型

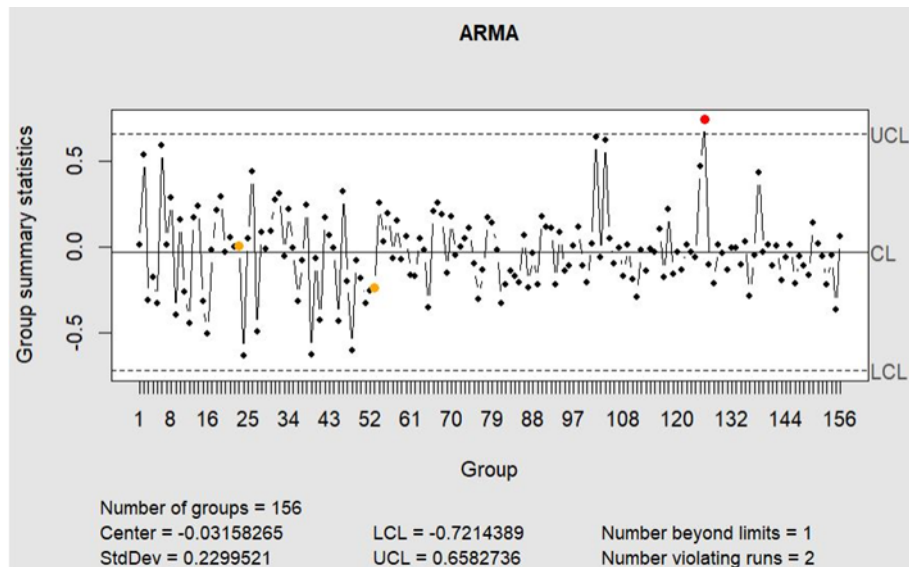
```
Call:
arima(x = data_month$rate, order = c(3, 1, 1))

Coefficients:
      ar1      ar2      ar3      ma1
    0.9800 -0.2063 -0.0849 -0.9090
s.e.  0.0922  0.1146  0.0860  0.0466

sigma^2 estimated as 0.05578:  log likelihood = 3.39,  aic = 3.22

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -0.03158265 0.2354164 0.1737514 -0.3847458 1.83371 0.8947332 -0.02715328
```

雖然前面 ADF 檢定出來序列是平穩的，但這邊用套件得到的模型是有經過一階差分的，且係數為 AR(3)與 MA(1)。



這是利用模型殘差畫出的管制圖，青年失業率只有一個點超出管制圖，時間就是在三級警戒剛開始的時候。這邊可以知道相比於總失業率，疫情的影響是比總失業率還要小的。

四. 結論

1. 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，這是我們從折線圖管製圖所得到的結果。
2. 利用時間序列模型之後，可以發現總失業率與青年失業率在疫情期間都會有預測比較不穩定的情形，但是總失業率相較於青年失業率受的影響可能是比較大的。所以我們以此得出青年失業率相較於總失業率比較不會受到外部因素的影響。
3. 後續可以利用比較多年的資料來探討兩者之間的差異，因為這幾年比較有影響的事件只有疫情，所以很難了解外部因素對兩個種類的差別。我們也需要比較多資料跟年份來確定是否青年失業率是比較不受外部因素影響的。

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://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>