時間序列分析 期末報告  
107508006 歐語四 陳葳芃 108405056 廣告四 楊岳錩

（之後會刪除）

<<報告內容>>

自行上網找時間序列資料、使用本學期講授的方法進行資料分析、預測。

(1) 保留最後s個(s為周期)觀測值(observations)做預測 (i.e.不要放下去建模)

(2) 資料分析應包含本學期講過的：  
ACF與PACF, order selection for SARIMA model, estimation, model diagnostic and model refinement  
(根據diagnostic的結果，SARIMA的order是否需要修正，...等), model selection (AIC, BIC, ...), and prediction.

報告大綱

1. 研究目的
2. 資料介紹(要詳細)   
    Time series、ACF、PACF
3. 原始時間序列觀察  
    Raw data ACF PACF 一階差分後的比較
4. 使用差分來de-trend進行order選取  
   每一個猜的模型都要以下步驟

參數估計 Estimation– 最高次係數是否異於0?  
模型診斷 Model diagnostic– 3方法  
模型修正 Model refinement?

1. 使用配飾迴歸模型來de-trend進行order選取  
   每一個猜的模型都要以下步驟

參數估計 Estimation– 最高次係數是否異於0?  
模型診斷 Model diagnostic– 3方法  
模型修正 Model refinement?

1. 模型選擇
2. 預測（1~12步預測）
3. 結論
4. 附錄（資料來源、R code不一定要）

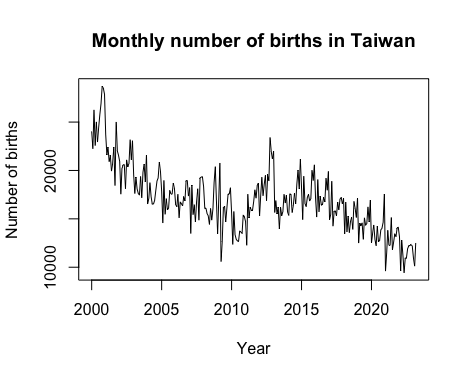
資料介紹：

本次報告研究的資料集為台灣月出生人口統計，來源自中華民國統計資訊網——總體統計資料庫，其中紀錄自2000年1月以來每月出生人數，資料目前更新至2023年3月。 樣本個數為279筆，變數個數為3個（可參考Table.1）   
變數介紹：

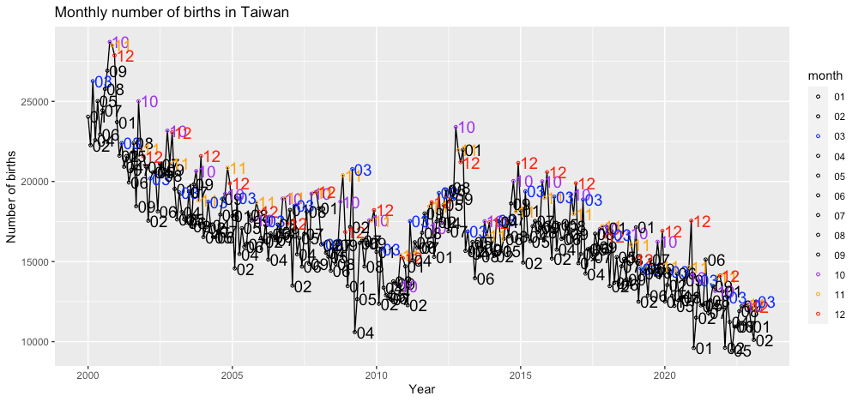
|  | time | amount | month |
| --- | --- | --- | --- |
| 類別 | 時間  time | 數值  numerical | 類別categorical |
| 說明 | 年月份 | 出生人數 | 月份 |

原始資料探索：

首先我們先將資料原始Time series給畫出來，可以看出整體生育人數是在往下降的，接著我們好奇哪些月份生育率會特別高？而哪些月份又特別低？又或者是哪些異於常理的值出現在我們的資料及當中。



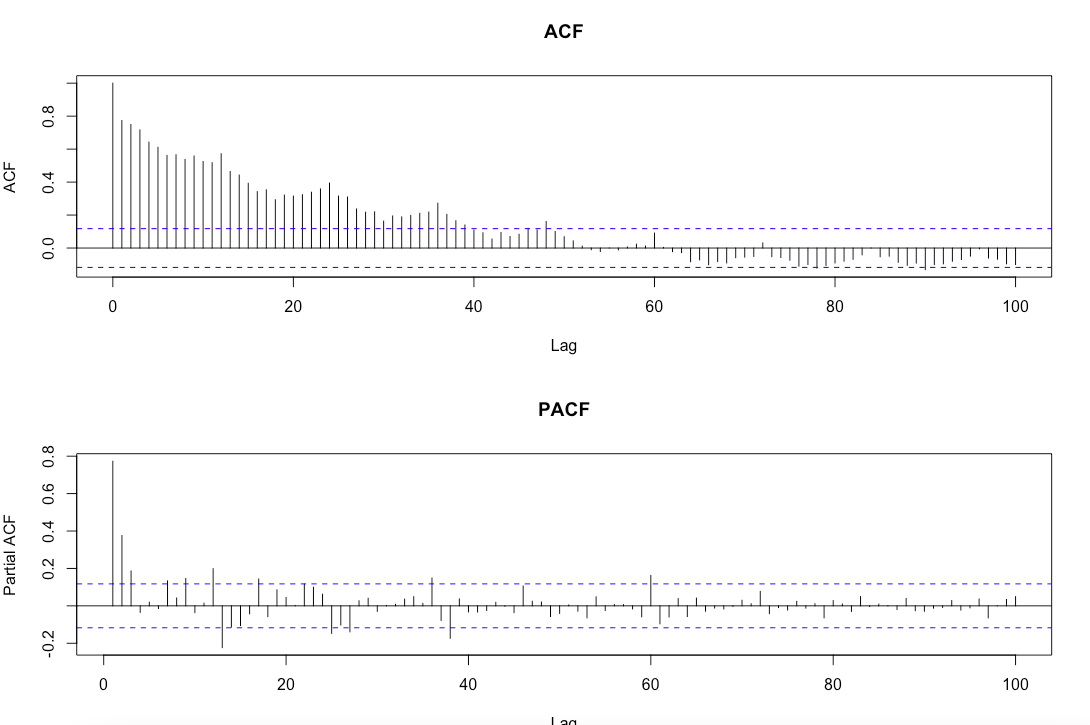
因此，我們使用ggplot將每個點標上月份，此時可以發現每年的相對極大值幾乎都是發生在10、12月，故將生育率高峰的此2個月特別用顏色標注起來。除此之外我們也可以發現2個異常之處：  
(1) 2009年3月為一個分水嶺，4月開始急遽下降且2010年整年生育率呈現一個低迷的狀況  
(2) 2012年 生育率又出現了拉升的情況，有點違反剛開始觀測的長期觀測趨勢



關於第1點細究原因，我們可以想到有可能是早些年發生的全球金融海嘯風暴，而根據戴庭玉於2011在陽明交通大學的研究論文《美國2007年至2009年引發環球金融風暴的制度因素》第七頁也指出貨幣流通速度與2008年第三季來到最高，象徵市場蕭條來臨。另一方面，王英傑於2015年中央大學完成的研究論文《景氣循環對婚育行為之影響》第八頁提及：『如過人們認為景氣蕭條持續，那對將來預期財富是下降的，......，造成生育率下降。』  
  
故根據2008年第三季產期懷胎10月左右的先備知識來看，剛好會對應上2009年3月，也就是說以上兩篇論文都可以佐證該出生人口下降分水嶺的很好原因。

關於第2點，2012年生育率恢復但卻更甚經濟蕭條前的數字，根據時任行政院政務委員薛承泰的採訪表示，政府於該段時期前的虎年就開始開始鼓勵結婚生育，並鼓吹年輕夫婦民國99年「愛情久久」年訂定婚約的風潮、生個民國百年寶寶或再下個年度生個龍寶寶，應是生育率竄高原因之一。

ACF PACF(raw)  
  
探索資料的時間序列圖結束後，可以藉由原始資料ACF PACF判斷出我們此筆資料，介於ACF值下降的不算緩慢，但同時也沒有明顯的快速下降，關於去除趨勢的一項抉擇——要不要差分，則是在經驗法則判斷的邊緣。因此我們也決定兩種方式都進行，最後再從中選取比較好的模型進行預測。



Augmented Dickey-Fuller Test(raw)

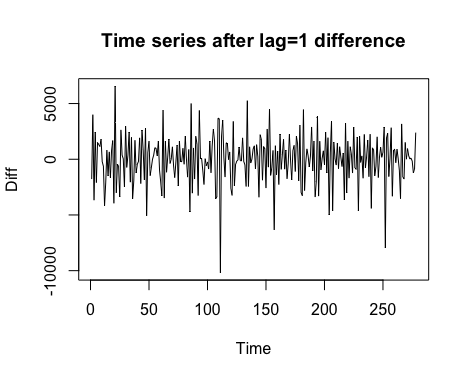
data: data

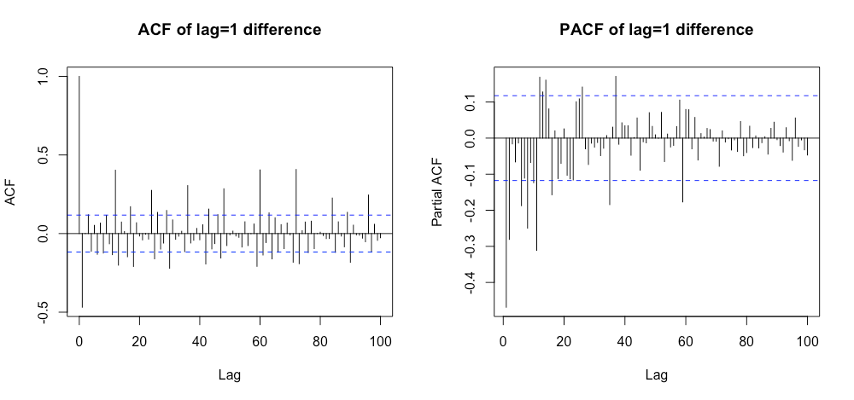
Dickey-Fuller = -3.6269, Lag order = 6, p-value = 0.03096

alternative hypothesis: stationary

【差分】

tsplot ACF PACF(diff=1)





Augmented Dickey-Fuller Test (diff1)

data: data\_diff1

Dickey-Fuller = -9.1596, Lag order = 6, p-value = 0.01

alternative hypothesis: stationary

Warning message:

In adf.test(data\_diff1) : p-value smaller than printed p-value