

時間序列分析顯示疫情行為之空間與年齡差異性

— 台灣水痘疫情 2006-2012

Time Series Analysis Reveals the Variations in Areas and in Age Groups of Epidemic Behaviors -- Chickenpox in Taiwan 2006-2012

戴珮瑄
亞洲大學
生物與醫學資訊學系
peggy011520@gmail.com

李正宇
亞洲大學
生物與醫學資訊學系
leecheng2005@yahoo.com

摘要

時間序列分析不僅僅是一個預測工具，更重要的是能揭露疾病傳播的本質。針對台灣水痘的傳播在 2006-12 年間的疫情，本研究藉由各行政區的時間序列分析與時間模型的建立，比較分析在不同空間的行政區之時間自我相關行為的差異性，並推論歸納可能的原因，進而提供疾病管制參考方向。我們發現：1. 時間序列模型真實反應在不同年齡群的疫情行為的差異；2. 疫情的行為不但決定於地域，也決定於都市化程度，甚至於其它因素。

關鍵詞：時間序列分析、水痘、空間相關、年齡、都市化

Abstract

Time-series analysis is not only able to provide forecasts, but also, more essentially, to reveal the nature of disease spreading behaviors. In this study, as an example, Taiwan chickenpox epidemics during 2006-2012 were investigated and compared using time-series analysis among eight administrative regions of Taiwan. We found that: 1. time-series models factually reflected different behaviors between age groups; 2. the behavior of an epidemic was not only affected by spatial factors, but also by degree of urbanization, and even other non-spatial factors.

Keywords: ARMA, chickenpox, spatial relation,

age, urbanization.

1. 前言

在台灣，水痘是一個很常見的傳染病，水痘是帶狀疱疹病毒，是一種猝然發作的全身性疾病，一開始會先微燒、顫抖、腹痛，肌肉關節酸痛約 2~5 天。在水痘復原之後，會潛伏在寄主的神經節中，避免被免疫系統摧毀。水痘的傳染方式主要是經由皮膚直接接觸、飛沫或空氣傳染，或是接觸到帶狀泡疹的水疱也會造成傳染，若是經由水疱液和黏膜分泌物感染的物器也會造成間接傳染，痂皮則不具傳染性，二次感染機會也很小。

傳染病的傳播通常是和時間相關。所以利用時間序列的研究可發現現在和過去在時間上的關係，分析變數所隱藏的行為特性；或在不同的空間中，比較傳播行為在時間序列上的差異性，提供傳染病管控的重要資訊。本研究是利用時間自我迴歸 AR 模型(Autoregressive models)建模與分析[3]，在時間上觀察水痘的傳染對不同年齡群及不同空間的行為特性。為了探討不同行政區之間(可能地理上或都市行為上相鄰或相似)存在的差異性，我們使用同樣的建模條件與前置處理，對各行政區分別進行建模，最後並予以比較；基於時間序列的數學特性，我們對不同年齡群與不同空間之時間序列模型的異同點作出可能的推論，並進而提出對台灣地區水痘疾病管制的參考策略。

2. 方法

2.1 自我回歸移動平均模型

自我回歸移動平均模型 (ARMA, Autoregressive Moving Average models) 是一種通用的時間序列模型，由 Wold's theorem 可知：任何穩定的統計現象皆可以 ARMA 模型表示，是一極適合用以描述自然界的現象的模型。ARMA 包含了兩種不同的統計行為：1. 自我迴歸程序 (autoregressive process); 2. 移動平均程序 (moving-average process) [1][2][3]。

ARMA(p, q) 是一個自我迴歸 p 階及移動平均 q 階的時間序列模型，可表示如下式 [1][2][3]：

$$X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

其中 X_t 為時間序列變數， ε_t 為統計雜訊， ϕ_i 為自我回歸係數， θ_j 為移動平均係數。

傳染性疾病傳播的主要行為應屬自我迴歸的統計程序 [4]，本研究也使用 AR 模型進行分析。AR 模型的所有移動平均的效應均不存在 (θ_j 全為 0)，並可表示如下式 [1][2][3]：

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$$

2.2 分析流程

傳統上，時間序列模型的建立分為三個階段：模型鑑定、參數估計、診斷檢定 [1] [2] [3]。

在這三個階段開始前，須先判斷序列資料的統計穩定性，此為 ARMA 建模的前提 [3]；若不穩定，則需視以下個別的情形進行建模前的前處理——1. 長期大趨勢：以一次到數次差分移除；2. 週期性或季節性變化：以重覆週期進行差分，移除週期性；3. 非常態分布 (non-Gaussian distribution) 的資料：進行轉換使之常態化 [3] (如：The Box-Cox Transformation [5])。

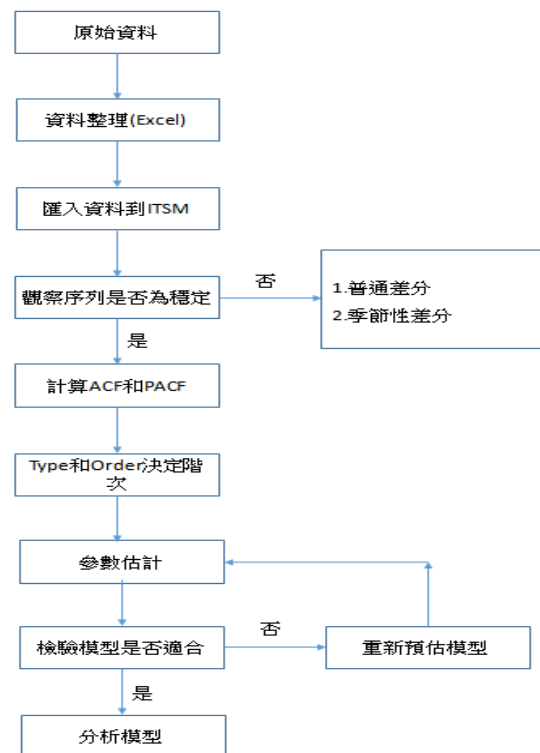
計算自我相關係數 (ACF, Autocorrelation Function) 和偏自我相關係數 (PACF, Partial Autocorrelation Function) 是 ARMA 分析的重要工具。ACF 可以先看出此時間序列的初步的時間自我相關特徵，PACF 可以決定模型的階次。當 ACF 的指數呈遞減 (exponentially decaying) 而 PACF 的 p 期後切斷 (cut-off)，此模型為 AR 模型；ACF 的 q 期後切斷而 PACF 的指數呈遞減，此模型為 MA 模型；若 ACF 和 PACF 的都呈現遞減，此模型為 ARMA 模型 [1][2][3]。

序列在相隔 k 的兩個時間點的自我相關係數 ACF (即 ρ_k) 的定義則如下式所示，

$$\rho_k = \frac{\text{cov}(X_t, X_{t+k})}{\sqrt{\text{var}(X_t)}\sqrt{\text{var}(X_{t+k})}}$$

其中 var 為變異數，cov 表示共變異。

建模的第一步為模型鑑定；透過 ACF 和 PACF 的行為，我們可決定模型的型態與階次。模型確定後，便可進行參數估計；一般而言，我們利用最小平方方法來估計最佳的參數值，而此過程通常為非線性的，是非常費時的；然而對於我們所使用的 AR 模型而言，只需使用 Yule-Walker 方程式，即可快速取得線性解 [3]。最後一步則為診斷檢定，有許多統計量值可讓我們估計模型的合適度，如：AICC, BIC [6] [7] 等等；如果模型過於複雜或合適度不足，必須回頭重新調整型態與階次後，再進行參數估計，直到滿足需求為止。本研究資料的分析與建模流程如圖一所示。



圖一 台灣地區水痘 2006~12 年疫情之時間序列 ARMA 建模過程

2.3 資料來源

本研究的資料來源是衛生福利部疾病管制局。根據 2006~12 年的水痘的疾病數，主要取地區的週資料，每年各 53 週。因本研究的需要，所以把行政區域劃分為台北區、台北新北區、北區、中區、南區、台南區、高屏區、高雄區；而東部和離島的病例數太少，在統計上

不具意義，因而不列入分析中。此外，我們取2008年中的病例年齡分19歲以下和19歲以上的年週的資料分別進行建模分析。

2.4 分析平台—ITSM2000

我們使用的時間序列軟體為ITSM2000(An Interactive Time Series Modeling Package)，此軟體由Brockwell與Davis在2000年開發。功能主要是資料檢視、轉換、分析、參數估計、計算ACF、PACF，預測和建立ARMA的模型[3]。

3. 結果

本研究分析的地區以行政區域劃分，分為八區。因在2011年縣市合併，所以2006~10年的台中縣併入台中市，台南縣併入台南市，高雄縣併入高雄市，原本的台北縣改為為新北市。八區分別定義如下：

1. 台北區--台北市、新北市、基隆。
2. 新北台北--台北市、新北市。
3. 北區--桃園縣、新竹(縣市)、苗栗縣。
4. 中區--台中市、彰化縣。
5. 南區--台南市、嘉義(縣市)、雲林縣。
6. 台南--台南市。
7. 高屏--高雄市、屏東縣。
8. 高雄--高雄市。

原始資料經過一階差分後，移除平均值，並以ITSM2000計算ACF和PACF。再經由判斷ACF及PACF的行為，決定選用二階AR模型，即AR(2)；隨後分別估計出相對應的第一參數和第二參數，如下：

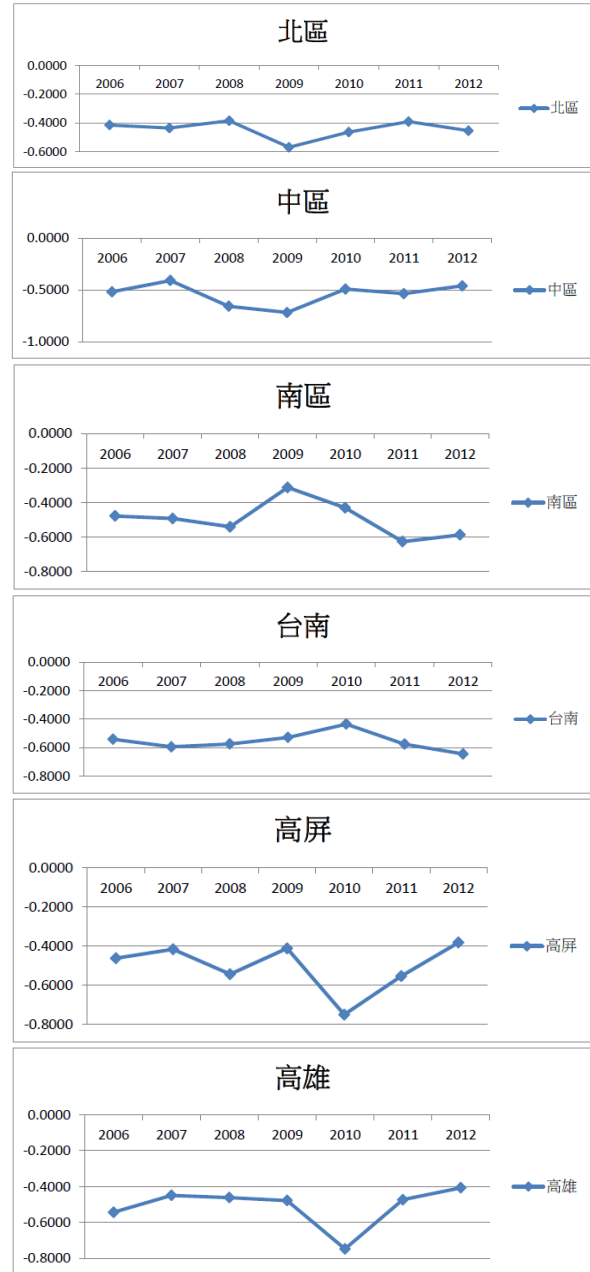
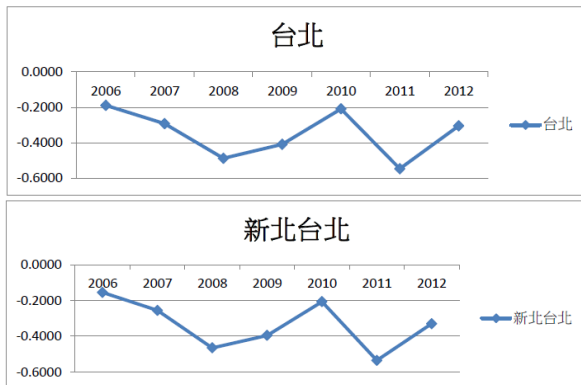
範例：台北區AR(2)模型為

$$X_t = -0.1895X_{t-1} - 0.1891X_{t-2} + Z_t$$

其中-0.1895為第一參數，-0.1891為第二參數

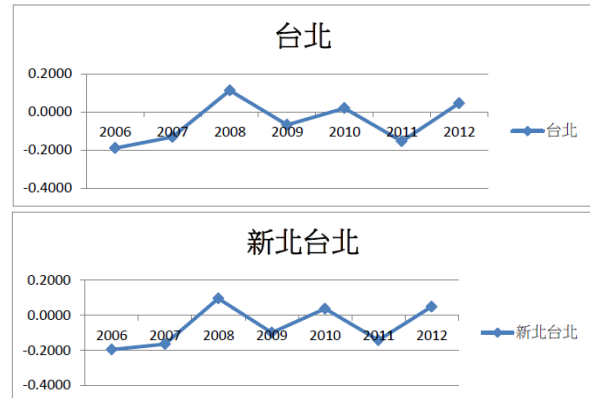
本研究主要探討在不同地區或年齡層的差異性，以此兩參數進行比較並作出統計推論。

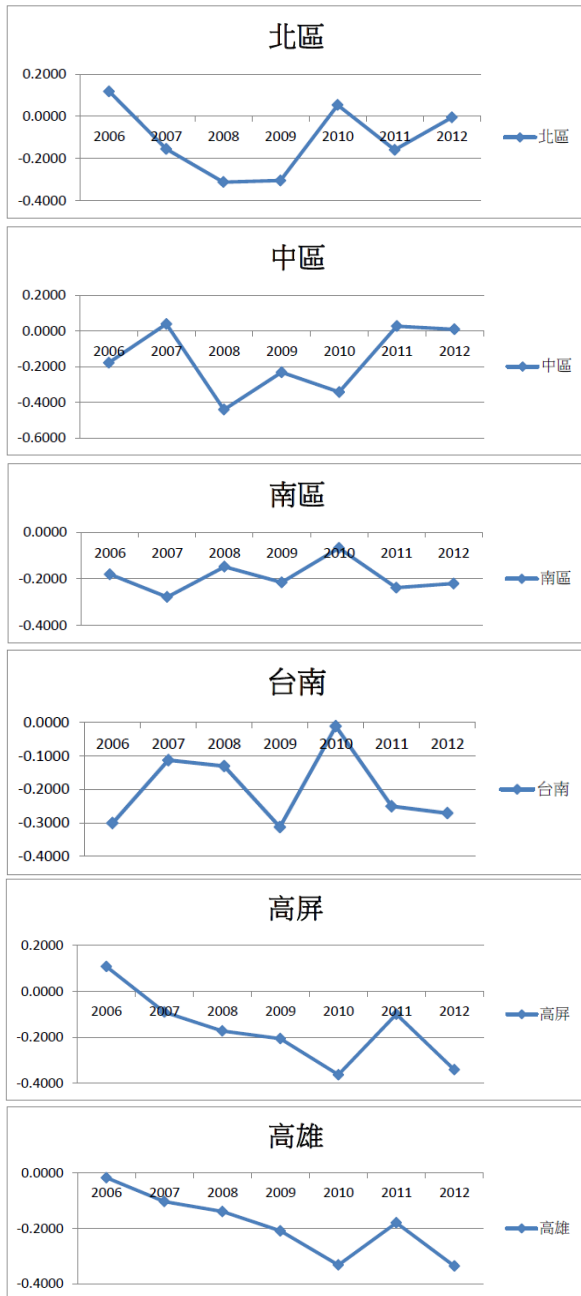
3.1 以地區分(第一參數)



圖二 2006~12年，八個地區模型的第一參數

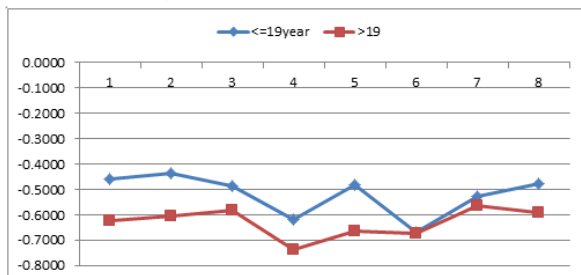
3.2 以地區分(第二參數)





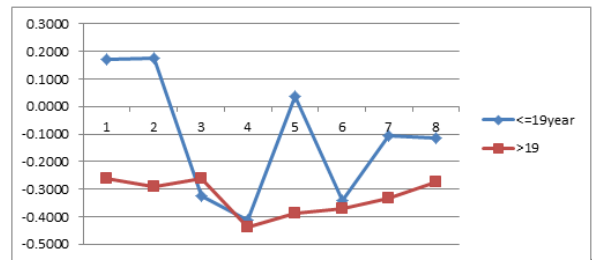
圖三 2006~12年，八個地區模型的第二參數

3.3 以年齡群分(第一參數)



圖四 2008年兩個年齡群(19歲以前和19歲以後)在不同區域(1:台北區,2:台北新北市,3:北區,4:中區,5:南區,6:台南,7:高屏,8:高雄)的模型第一參數

3.4 以年齡群分(第二參數)



圖五 2008年兩個年齡群(19歲以前和19歲以後)在不同區域(1:台北區,2:台北新北市,3:北區,4:中區,5:南區,6:台南,7:高屏,8:高雄)的模型第二參數

4. 討論

水痘的潛伏期大約是兩週的時間，且具有高度傳染性，水痘的互相傳染率可高達85%，但痊癒後即免疫。以下分別討論前節結果中的彼此異同及可能的推論。第一參數關連到前一週，第二參數是關連到兩週前。以下分述2008年不同年齡群的模型比較，與2006~12年各分區的模型間的交叉比較。

4.1 不同年齡群的模型

以圖四和圖五來比較，討論如下。

一般而言，19歲以前無論是在第一參數(時間遲延一週)或第二參數(時間遲延二週)都較19歲以後較大，即19歲以前的病例數改變量在兩週內的關連性較19歲以後為大，此原因應為19歲以前病例多數以就學為主，使群聚感染更容易，且部分成年者在少年期已免疫，二次感染的機會很低。

19歲以前的第二參數在台北區、台北新北及南區都有很大的正值，我們推論在這三區的第二週關連是很大的，應以適當的管制措施(建議感染者至少停學五天)以降低或斷絕此疾病的時間關連性，減少擴散的可能。

北區、中區及台南(即桃、竹、苗、中、彰、及台南市)19歲以前與以後，在模型上幾乎沒有差別；這可能是與人口結構的類似有關。

此外，南區和台南在19歲前後的第二參數相差甚大，主要的差異應是南區尚包含雲林和嘉義這兩個縣市，而台南市為都會區，所以導致較大的變化。

4.2 不同地區的模型

在所有地區，第一參數全為負值，而第二參數僅有少部分地區在某幾年是負值；亦即前

一週疫情變化對於本週疫情的變化是負向的；兩週前疫情變化對於本週的影響也是負向的；由此特性看來，此疫情並不易擴大。

在所有地區，在 2006~08 年間，第一參數幾乎全為遞減（愈來愈負），即前一週疫情變化對於本週疫情變化的影響是愈來愈大，而且是負向的，也就是 2006~08 年間，每年的疫情消弭的速度是愈來愈快；但在 2008 或 2009 年起，多數地區的此一參數為大幅正向反彈，疫情在前後一週的關連性大增，疫情漸漸失去穩定控制。2008~09 年當中有金融風暴和 H1N1 疫情發生，導致願意看診的人數不多，或是施打疫苗者的意願下降，才會導致此現象。

在各區分析結果的觀察中，比較台北區和台北新北、高屏區和高雄區，和台北區和高屏區，我們發現空間上相近的區域所呈現出來的 ARMA 模型特性較相似；由此可推估各地區在空間上疫情的相關性。

我們進一步作了各地模型相似性的分析。表一的數值為各地區模型的第一參數在 2006~12 的相關性係數，很明顯的，疫情的時間行為在空間上是有區域性，或區域相關的；此外，地區的都市化程度也影響很大，如：台北區雖然在地理上與北區（桃竹苗）近，但時間上的傳播行為卻與中區（中彰投）較接近；另，高雄與台南雖然近，但時間上的傳播行為卻背道而馳，此應與台南高度都市化的區域比例較低有關。

下表中所列的數值為相關係數(correlation coefficients)，1.0 為完全一致，-1.0 為完全相反；格中的淡色表高相似度，深色表低相似度。

水痘 AR 模型的相關性	台北	新北台北	北區	中區	南區	台南	高屏	高雄
台北	1.00	0.99	-0.18	0.53	0.34	0.34	-0.08	-0.50
新北台北	--	1.00	-0.14	0.51	0.37	0.36	-0.07	-0.50
北區	--	--	1.00	0.38	-0.83	-0.28	-0.23	0.09
中區	--	--	--	1.00	-0.46	-0.20	-0.02	-0.10
南區	--	--	--	--	1.00	0.60	-0.01	-0.37
台南	--	--	--	--	--	1.00	-0.78	-0.94
高屏	--	--	--	--	--	--	1.00	0.86
高雄	--	--	--	--	--	--	--	1.00

表一 2006 年~2012 年，八個地區模型相似度分析

5. 結論

在本研究中，我們以時間序列來分析

2006-2012 水痘時間相關的行為在不同的年齡群及在不同空間的差異性，從結果與討論中，我們看到時間序列模型真實反應了不同族群的疾病傳播行為，這樣的傳播行為差異不僅出現在年齡差異上，也會出現在生活習慣、民族、氣候、生活水平等差異上，提供未來使用時間序列研究者多面向的考量。

此外，經由不同空間中疾病傳播行為的比較，我們更進一步發現：疾病傳播的空間分析，不可僅考慮地理上的距離，其它如：都市化程度、交通便捷性、人口稠密度等都有很深的影響。

經由細心的觀察與分析，時間序列分析提供給我們的不再只是時間上或數量上的預測，而是對於疾病管制更有意義的疾病傳播行為的本質；此一資訊，對如何擬定管制政策、控制疫情、推行公衛教育、衛生改善是更為關鍵的與有效的。

6. 參考文獻

- [1] 江麗玲”時間序列自我迴歸模型應用於 1998 年-2006 年臺南市與高雄市登革熱案例之研究”，碩士論文，亞洲大學，2009。
- [2] 施惠蘭”時間序列分析法-自我迴歸模型應用於台灣 1999~2008 之腸病毒疫情”，碩士論文，亞洲大學，2009。
- [3] P. J. Brockwell and R. A. Davis., Introduction to Time Series and Forecasting, Second Edition, 2nd ed., Springerl, 2001.
- [4] Mathias Hofmann. Statistical Models for Infectious Disease Surveillance Counts, Dissertation, Ludwig Maximilians Universitaet, Muenchen, 16. November 2006.
- [5] Box, George E. P.; Cox, D. R. (1964). "An analysis of transformations". *Journal of the Royal Statistical Society, Series B* 26 (2): 211–252.
- [6] H. Akaike. Time series analysis and control through parametric models. In D. F. Findley, editor, Applied Time Series Analysis, pages 1-23, New York, 1978. Academic Press.
- [7] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6 (2):461-64, 1978.