華醫學報　第 48 期
2018 年 6 月
頁數 60-71

以決策樹分析台灣地區近年之桿菌性痢疾

郭秀蘭
中華醫事科技大學食品營養系

摘要

桿菌性痢疾（Shigellosis）由腸道桿菌科（Enterobacteriaceae）志賀氏菌屬（*Shigella*）引起，其傳染方式與食物和飲水有關，根據疾病管制署官網資料顯示國內近年常出現之血清型別為 *S. sonnei* 及 *S. flexneri*。採集自傳染病統計資料查詢系統中確定病例數為依變數(共 2214 筆)，自變數為發病月份、縣市、性別、是否為境外移入,年齡層等六項進行決策樹 CRT (Decision tree，classification & regression trees)分析，結果發現自變數重要性依次為縣市(花蓮縣、台中市、南投縣)>發病月份(11 月、3 月)>是否為境外移入(是境外移入)，其餘未顯示在節點中之自變數則較不重要。此模式之風險值為 0.056，正確度為 94.4%，標準誤差為 0.005，此結果可提供桿菌性痢疾疫情管理工作之參考。

關鍵詞：桿菌性痢疾、決策樹、食物媒

華醫學報　第 48 期
2018 年 6 月
頁數 60-71

Application of decision tree for Shigellosis distribution analysis in Taiwan

Hsiu-Lan, Kuo
Department of Food and Nutrition

ABSTRACT

Shigellosis is caused by Enterobacteriaceae and Shigella, which are associated with food and drinking water. Recently years data from the official website of the CDC was displayed that serotype of *S. sonnei* and *S. flexneri* were the major disease-causing species. All of cases obtained from the TNIDSS is used by the number of dependent variables (2214 in total), the six independent variables were the month of onset, county, city, gender, overseas import cases and age were used for CRT (Decision tree，classification & regression trees) statistics analysis. It was found that the importance of the variables was counted by the counties (Hualien County, Taichung City, Nantou County)> Onset months (November, March)> indigenous/import cases (moved overseas), the remaining nodes that are not shown in the node are not important. In this model, the risk value is 0.056, the percent of correctness is 94.4%, and the value of standard error is 0.005. The obtained data could be application for control Shigellosis in Taiwan.

Keywords: Shigellosis ; decision tree; foodborne

前言

決策樹(Decision tree)是一種統計運算模式，在目前資料探勘領域中決策樹的運用亦能夠將資料適切的分類，分類的結果對於未來的預測和決策輔助非常有用，因此廣泛被應用於各個領域中，無論是國內外醫學領域皆然。以國內醫學研究而言，在疾病風險與預測方面例如類型風溼性關節炎(吳等，2016)、植牙(林和胡，2017)、阿茲海默症(侯等，2009)、冠狀動脈心臟(徐等，2007)、肝病(劉與陳，2016) 及其他醫院管理等(江等，2016；羅，2016；陳等，2013；侯等，2009；湯等，2008)；在食品與營養部分例如涂等(2006)早期進行食用油之市場區隔，此類顧客分析之運用最為普遍。羅(2016)結合資訊用以製作慢性病飲食推薦系統，龔等(2011)以此設計三高患者之飲食推薦系統，另外江等(2016)則用以進行廚房火災預測。

在我國疾病管制署法定傳染病分類中，和食物比較有關的包括第二類(腸道出血性大腸桿菌感染症、傷寒、霍亂、桿菌性痢疾、急性病毒性Ａ型肝炎)、第四類(李斯特菌症、肉毒桿菌中毒、庫賈氏病)與第五類中之沙門氏菌感染症等。近年來這些傳染病中以Ａ型肝炎和桿菌性痢疾居多，例如在 105 年法定傳染病確定病例人數統計年報(疾病管制署，2017)資料中以第二類之急性病毒性Ａ型肝炎人數最多(總計為 1,133 人)而桿菌性痢疾居次為 225 人，其餘則為零星案例，此顯示在感染型食物中毒案例中兩者之管理非常重要。然而因為我國對於餐飲從業人員有Ａ型肝炎檢查之規範，因此較為一般民眾認識。

桿菌性痢疾是由志賀氏菌(Shigella)引起的疾病，它有數種菌株，早期Gupta等人(2004)曾指出在已開發國家中以*Shigella sonnei* 居多，其次為*Shigella flexneri*。近年Dewanti-Hariyadi and Gitapratiwi (2014)指出在東南亞和中亞國家中志賀氏菌是這兩個地區國家中造成國民腹瀉的主要原因菌，而且也提到有75~95%的志賀氏菌具有抗藥性。在國內曾有幾家媒體在電子報中報導國內已經發現亞洲首例抗Azithromycin之菌株，據官方發佈疫情報導資料(廖等，2017)顯示2015及2016年國內發現之之志賀氏菌仍以*S. sonnei*最常見而*S. flexneri* 次之，另外在疾病管制署之監測報告中亦記載”對過去常用之抗生藥物streptomycin 、ampicillin 、sulfamethoxazole、TMP-SMX 及tetracycline 之抗藥性比率皆超過50%”，並且提及” 在測試的29 株菌株（6 株*S. sonnei*，23 株*S. flexneri*）中，有15株(51.7%)對azithromycin 有抗藥性，這些抗藥菌株皆屬於*S. flexneri* 血清型3a”，因此2017年疾病管制署曾發出抗Azithromycin之菌株之警示。因為桿菌性痢疾屬法定傳染病，因此在疾病管制署網站中對於該病菌之基本資料和防治均有詳載，一般民眾可上網查詢，在2017(廖等)報導資料

曾經說明”人與人之間的接觸傳染是桿菌性痢疾最主要的傳播模式,污染的食物與飲水則常引發大規模的流行”,因此除了人員的接觸和食物汙染之外,在山區,颱風淹水之後也都有部份風險存在,需要留意飲水衛生。,一般民眾對於桿菌性痢疾似乎較為陌生,因此,本研究擬透過官方資料分析提供相關資訊以供參考

材料與方法

一、資料採集

確定病例採集自衛生福利部疾病管制署之傳染病統計資料查詢系統(Taiwan National Infectious Disease Statistics System,TNIDSS),資料於107年8月上網站採集自2003年第1週~2017年第35週之發病病例共2214筆,先以Excel分類整理資料,再由統計軟體讀取後進行統計分析。

二、變數分類

設定確定病例數為依變數,自變數共有六項,分別為年份(2003~2017)、月份(1~12)、縣市(22個)、性別(男、女)是否為境外移入(本土、境外)以及不同之年齡層(<0,1-4,5-9,10-14,…….65-69,>70)等進行分析。

三、決策樹分析模式

1. 成長方法: 採用CRT (classification & regression trees)
2. 最大樹狀結構深度: 5
3. 父節點中最少觀察值個數: 400
4. 子節點中最少觀察值個數: 200

四、統計軟體

採用IBM SPSS statistics 第19 版進行統計分析。

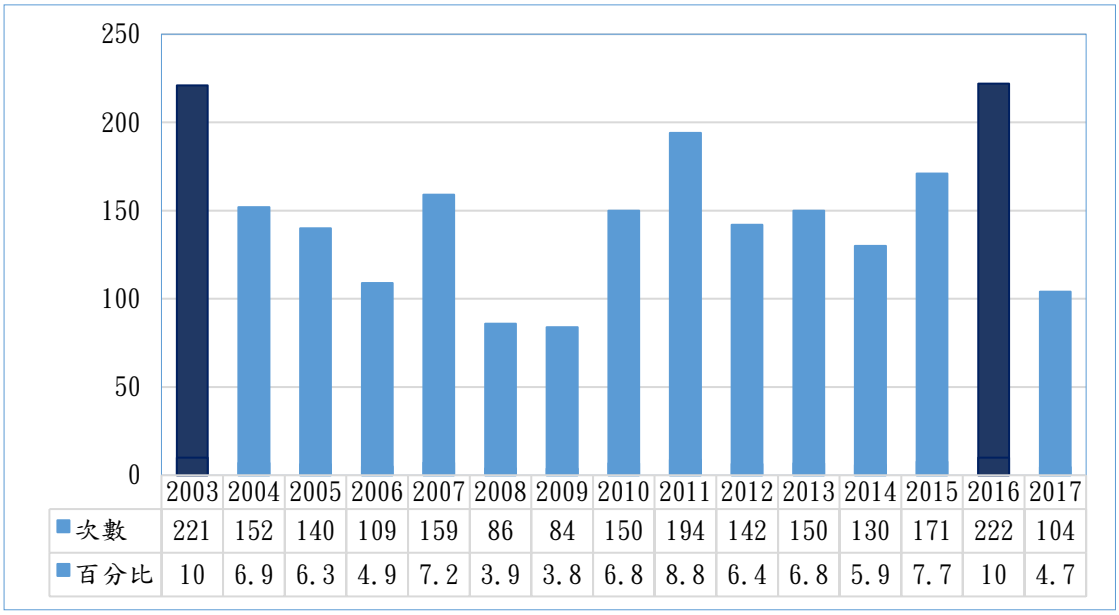
結果與討論

一、桿菌性痢疾各自變數之次數分配結果

比較查詢系統中2003~2017年和食物來源有關的感染性法定傳染病資料可以發現,除了A型肝炎(共1874人)和傷寒(共514人)之外,其他感染型案例均以零星發生居多,其中霍亂(*Vibrio cholerae*)最高紀錄為每年為10人、肉毒桿菌(*Clostridium botulinum*)曾經為11人、庫賈氏症(Creutzfeldt-Jakob disease, CJD)曾經在2009年偶發確診4人,而在國外有些案例發生的腸道出血性大腸桿菌(Enterohemorrhagic *E. coli*, EHEC)則未曾在國內發現,另外值得留意的是李斯特菌症(Listeriosis)在107年1月1日正式列為第四類傳染病,在這兩年也開始出現在官方統計資料中,2017年出現2個案例,2018年出現12個案例,且其中有

5位為70歲以上之高齡者。

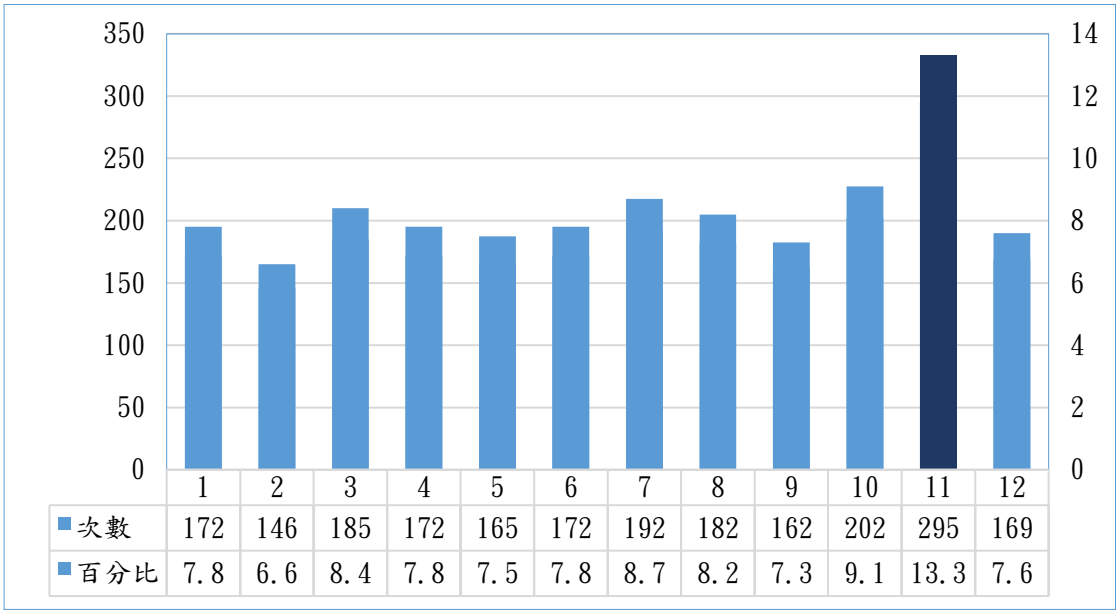
國內桿菌性痢疾確定病例分析後之次數分配結果如圖一~圖六所示。從圖一可發現歷年來桿菌性痢疾確定病例大多為100例以上，而在2003年和2016年曾經超過200例，均佔總案例之10%，另經查詢2017全年較新資料為162例。在桿菌性痢疾發生月份中可以發現以11月(圖二)最常發生，總計有295案例，佔總案例之



	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
次數	221	152	140	109	159	86	84	150	194	142	150	130	171	222	104
百分比	10	6.9	6.3	4.9	7.2	3.9	3.8	6.8	8.8	6.4	6.8	5.9	7.7	10	4.7

13.3%；10月份也有202案例，其他月份均在100例以上。

圖一　台灣地區桿菌性痢疾確定病例發病年份次數分配圖表

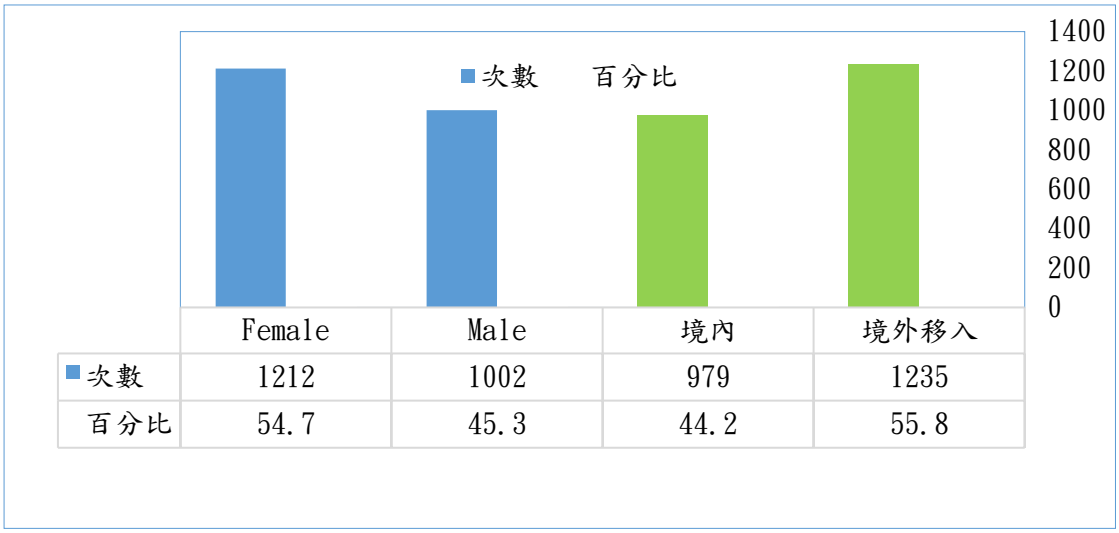


	1	2	3	4	5	6	7	8	9	10	11	12
次數	172	146	185	172	165	172	192	182	162	202	295	169
百分比	7.8	6.6	8.4	7.8	7.5	7.8	8.7	8.2	7.3	9.1	13.3	7.6

圖二　台灣地區桿菌性痢疾確定病例發病月份次數分配圖表

性別分析結果顯示女性稍為高於男性(圖三)，分別佔54.7%和45.3%；而圖三

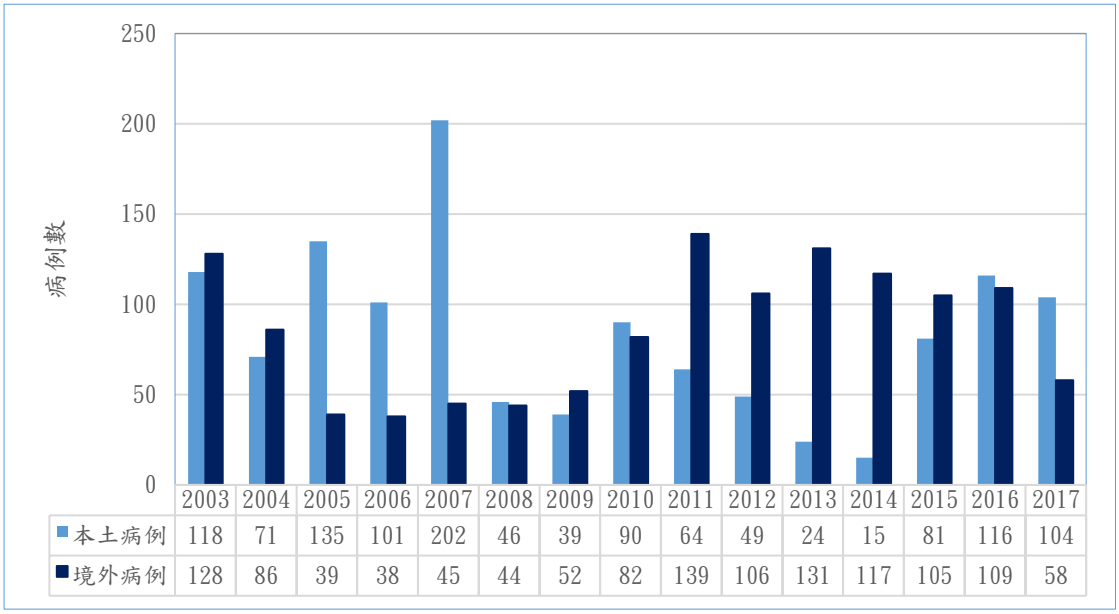
同時顯示在境內和境外移入判定方面則以境外移入居多佔55.8%。進一步交叉分析各年份資料顯示，歷年間境外移入病例最多者為2011年之139例，在此之前境外移入與本土病例互有消長，但2011~2015年境外移入案例均高於本土病例(圖四)。在縣市別方面顯示於圖五，歷年來以新北市最多佔21%，其次為台北市、桃園市、花蓮縣、台中市和南投縣等均超過100個案例，這幾個縣市即佔總案例之73.6%。進一步分析查證各鄉鎮模式評估我們發現在花蓮縣玉里鎮曾經發生過之案例數最多，為113案例，佔總數之5.3%，可能也因此影響了縣市別的結果。將確診人數歸納為3個規模(案例數>200以上，介於100~200和<100)分析各年齡層之差別可以發現高於200個案例以上的年齡層：集中在20~44歲之間，其中25~29歲之青年人發生率最高，案例數達431位；介於100~200個案例的年齡層：為1~9歲



	Female	Male	境內	境外移入
次數	1212	1002	979	1235
百分比	54.7	45.3	44.2	55.8

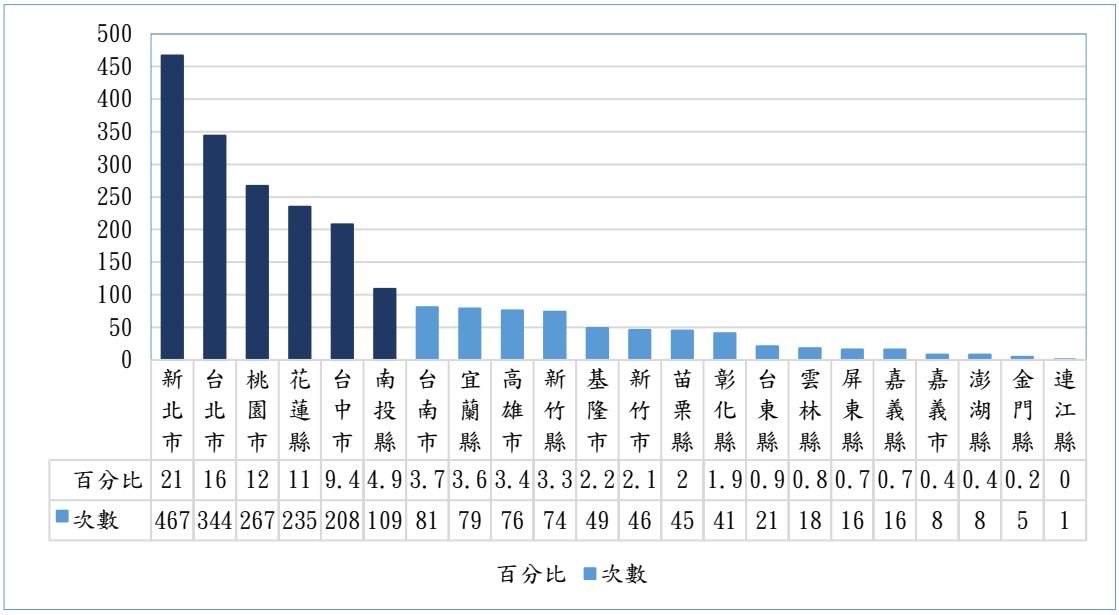
和>70歲以上的國民；其餘則均低於100例以下。

圖三　台灣地區桿菌性痢疾確定病例發病性別及是否為境外移入次數分配



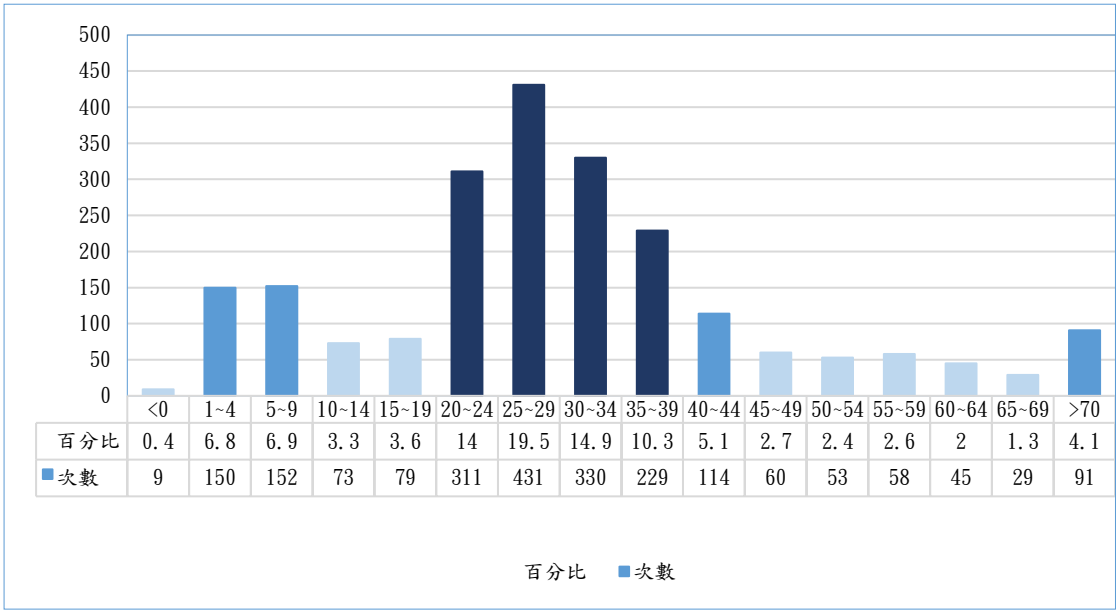
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
本土病例	118	71	135	101	202	46	39	90	64	49	24	15	81	116	104
境外病例	128	86	39	38	45	44	52	82	139	106	131	117	105	109	58

圖四台灣地區桿菌性痢疾確定病例境內/境外別病例數



	新北市	台北市	桃園市	花蓮縣	台中市	南投縣	台南市	宜蘭縣	高雄市	新竹縣	基隆市	新竹市	苗栗縣	彰化縣	台東縣	雲林縣	屏東縣	嘉義縣	嘉義市	澎湖縣	金門縣	連江縣
百分比	21	16	12	11	9.4	4.9	3.7	3.6	3.4	3.3	2.2	2.1	2	1.9	0.9	0.8	0.7	0.7	0.4	0.4	0.2	0
次數	467	344	267	235	208	109	81	79	76	74	49	46	45	41	21	18	16	16	8	8	5	1

圖五 台灣地區桿菌性痢疾確定病例發病縣市別次數分配圖表



	<0	1~4	5~9	10~14	15~19	20~24	25~29	30~34	35~39	40~44	45~49	50~54	55~59	60~64	65~69	>70
百分比	0.4	6.8	6.9	3.3	3.6	14	19.5	14.9	10.3	5.1	2.7	2.4	2.6	2	1.3	4.1
次數	9	150	152	73	79	311	431	330	229	114	60	53	58	45	29	91

圖六　台灣地區桿菌性痢疾確定病例發病年齡層次數分配圖表

二、桿菌性痢疾以決策樹CRT分析結果

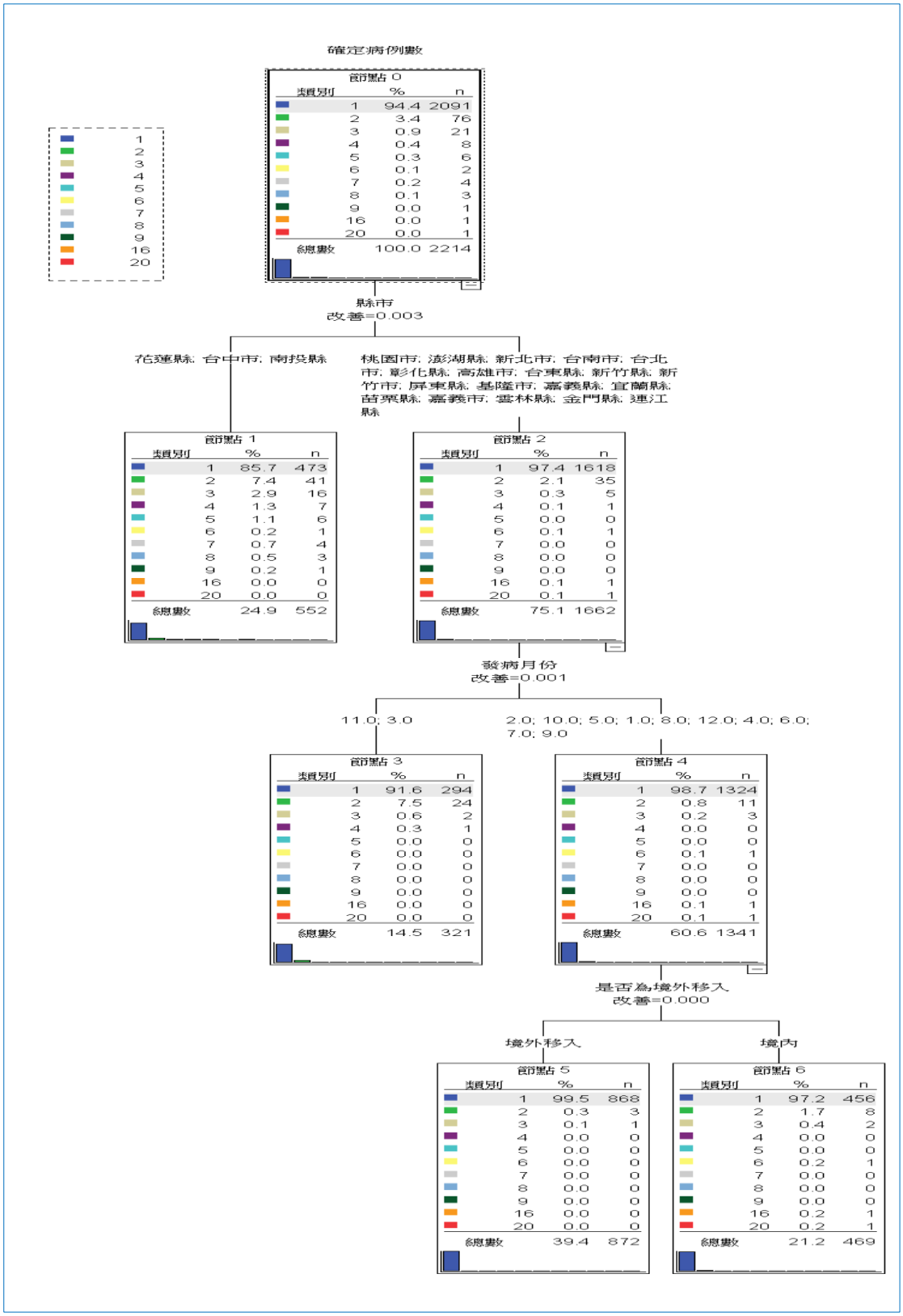
以決策樹CRT模式進行桿菌性痢疾資料分析，得到其節點數量為7，終端節點數量為4，而其深度為3，其風險與各個自變數重要性如表一所示，決策樹節點模式圖如圖七所示，運用傳染病統計資料查詢系統繪製之分佈圖如圖八所示。從表一得知分析結果之風險值僅0.056，表示在評估模式中有94.4%的資料分類是適切的，且其標準誤差低於0.005。經過CRT分類結果亦發現，在自變數中最重要的項目為”縣市”，其次為”發病月份”和”是否為境外移入”等幾個自變項，特別是在”縣市”變項中其正規化重要性為100%，顯示在幾個發病人數較多的縣市投注較多的管理工作是非常有必要的。透過CRT模式分析可以了解在縣市工作部分以花蓮縣、台中市以及南投縣最為重要，此與圖五之次數分配表略有不同，此可能因這幾個縣市集中於少數鄉鎮區的關係，例如花蓮縣玉里鎮即曾累積113個案；因而出現在第一個節點。若以次數分配表觀之，台北市新北市和桃園縣之案例數分別位居一到三名，因此在疾病管制署之互動式圖表中(圖八)這三個縣市亦為警示區域。另外，如先前所述在發病月份中以10月和11月居多，因此在此之前加強宣導教育工作應該可行。根據疾管局資料顯示境外移入國家以東南亞國家最多，建議亦可以加強聘顧外勞之企業團體宣導工作。在決策樹模式中會依自變數

決策樹模式能做明確的分類，雖然也有一些缺點，但其節點圖型易於判別(黃等，2006)，本分析結果會有助於輕重緩急之判定，例如在本研究中台北市、新北市和桃園縣並沒有出現在第一個節點，但就分析結果而言[區域管理]被認為是最重要的因子。第三個節點出現的是月份，亦說明做好[時間管理]有其意義。分

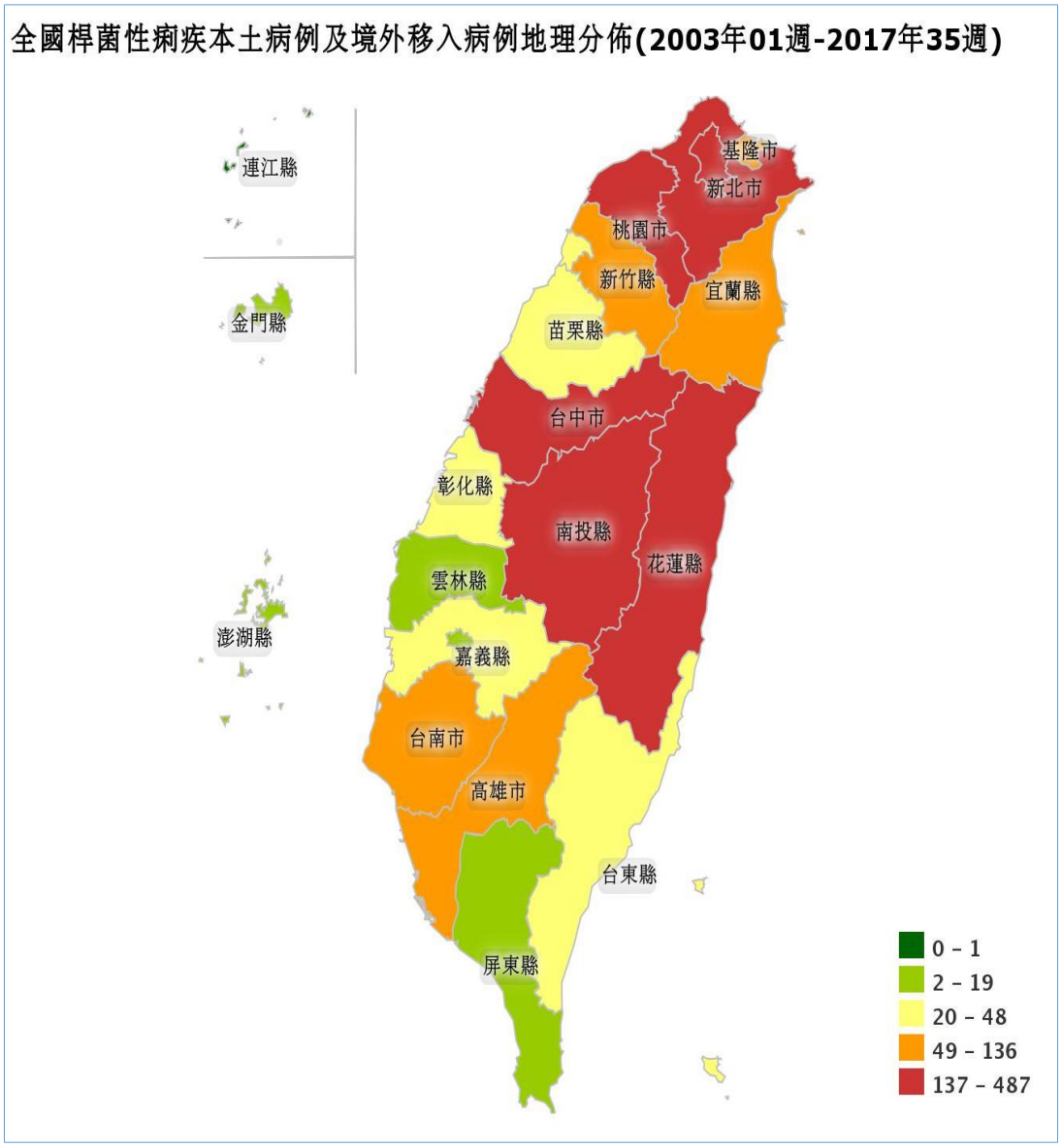
類或預測是大數據分析中非常重要的功能，本研究僅初步探討單一模式的分析結果，期望未來能進一步作不同模式比較分析，以便獲得更適切的結果。

表一 台灣地區桿菌性痢疾確定病例以CRT分析結果之風險與自變數重要性

風　險	自變數的重要性		
估計	自變數	重要性	正規化重要性
0.056	縣市	.003	100.0%
標準誤差	發病月份	.001	34.0%
0.005	是否為境外移入	.000	3.1%
成長方法:CRT	性別	5.428E-5	1.7%
依變數: 確定病例數	年齡層	1.491E-5	0.5%
正確率94.4%			



圖七 台灣地區2003~2017年(第35週)桿菌性痢疾確定病例決策樹節點模式圖



圖八　台灣地區2003~2017年桿菌性痢疾確定病例地理分佈圖(疾病管制署)

參考文獻

江秋蓉、李春松、陳佩葶、張淑卿、盧育聘（2016）。利用根本原因分析（RCA）預防廚房火災。臺灣膳食營養學雜誌，8(1)，53-66。

吳建廷、程秀蘭、胡雅涵、童建學、彭子安（2016）。比較三種資料探勘演算法預測類型風溼性關節炎預後之研究。北市醫學雜誌，13(3)，98-110。

林承俊、胡雅涵（2017）。運用資料探勘技術建構植牙手術成敗之預測模型。醫學與健康期刊，6(2)，57-69。

侯藹玲、許晏賓、江志民（2009）。資料採礦技術應用於微陣列資料分析以篩選

阿茲海默症候選基因之研究。Journal of Data Analysis，4(3)，179-213。

涂嘉峪、曾光華、何雍慶（2006）。應用資料採礦技術分析生物科技新產品之市場區隔：以健康食用油為例。科技管理學刊，11(4)，63-98。

徐敏耀、劉夷生、馬作錡、張木信、張丁權、賴昭宏、鍾國屏（2007）。冠狀動脈心臟病危險因子之老年人心導管檢查預測模型研究。台灣老年醫學暨老年學雜誌，3(1)，25-33。

陳月蓉、張光昭、邱俊誠（2013）。台灣地區保健食品消費力探討。Journal of Data Analysis，8(5)，97-112。

湯宗泰、簡守維、賴仲亮、陳永福、高琳詠(2008)。利用決策樹方法分析探討糖尿病人血糖控制之成效。資訊科技國際研討會論文集。No. 290。

黃美玲、許佑新、黃智仁、陳正誼、陳俊誠(2006)。應用決策樹於青光眼患者之鑑別。中華民國品質學會第 42 屆年會會議論文。1-11。

廖盈淑、廖春杏、梁綉雲、王佑文、曹其森、邱乾順(2017)。2015 年臺灣桿菌性痢疾流行病學與抗藥性分析。疫情報導。33(4)，61-70。

劉振隆、陳建良（2016）。以決策樹分析與模糊邏輯技術建立肝病罹病風險評估系統。Electronic Commerce Studies。14(4)，475-498。

衛生福利部疾病管制署(2017)。2016 年台灣志賀氏桿菌抗藥性監測報告-掛網版。3-4。

衛生福利部疾病管制署 (2017)。 http://www.cdc.gov.tw/professional/knowdisease.aspx。

羅育文(2016)。實作飲食本體結合決策樹之慢性病飲食推薦系統。碩士論文。朝陽科技大學資訊管理系。

龔旭陽、郭庭歡、蔡京玶(2011)。基於語意感測網路之智慧型適性化健康飲食推薦系統 設計與實作—以三高患者為例。資訊科學應用期刊。7(2)，21-39。

Dewanti-Hariyadi, R and Gitapratiwi, D. (2014). Prevalence of foodborne diseases in South East and Central Asia. ***Encyclopedia of Food Safety. 1***, 287-294.

Gupta A., Polyak, C. S., Bishop, R. D., Sobel, J. and & Mintz, E. D. (2004). Laboratory-confirmed shigellosis in the United States, 1989–2002: epidemiologic trends and patterns. ***Clin. Infect. Dis. 38***, 1372–1377.