

逢甲大學學生報告 ePaper

報告題名：影響台灣總人口數的因素

Factors that affect for the total population of Taiwan

作者：葉穎謙、周湘昀、魏君庭、許冠翎、宋子平、陳勁輝

系級：統計學系三甲

學號：D0065054、D0023916、D0065188、D0065071、D0063895、D0041453

開課老師：高秀蘭 教授

課程名稱：迴歸分析

開課系所：統計學系

開課學年：102 學年度 第一 學期

中文摘要

在此報告中，探討各項反應變數對解釋變數(台灣總人口數)具有較大的解釋能力，先對原始資料做 Excel 迴歸分析後，去除相關係數不符合範圍內者，查看常態圖是否為 45 度、殘差圖是否以 0 作為界線，上下均分。

對原始資料做迴歸分析，本組仍保留其分析結果，原先的七個變數利用 SAS 檢定，但初步結果發現解釋變數中的自然增加人數為出生人數減去死亡人數、總增加人數為自然增加人數加上社會增加人數，因此將自然增加人數和總增加人數從變數中剔除，再利用以下三種分析方法：第一為向前選取法(Forward Selection)，第二為向後消去法(backward method)，第三為逐步迴歸法(Stepwise)。發現「出生人口數」解釋變數不具有線性迴歸關係，進而減少一個變數，才可增加其餘變數之解釋能力。

將遺漏值加以刪減，漸進推出正確公式，交叉比較變數之結果，得出最具代表性之線性迴歸線。再以殘差分析檢定資料是否為常態，變異數是否為常數及迴歸模型是否為最佳的線性迴歸模型，如此才能加以將變數數據化，其影響程度一覽無遺。

最後所得到的結論為：在上述三種方法的篩選後，留下具可信的變數：死亡人數、社會增加人數、車禍死亡人數、出國人數，此四種變數為本組所選變數下影響總人口數較大之因素。

關鍵字：台灣、迴歸、殘差分析、選取法、總人口數

Abstract

In this report, explore each response variable on the independent variables (total population of Taiwan) has a greater explanatory ability, doing the regression analysis with the original data by Excel, then remove the correlation coefficient those who does not comply with range, check whether the regression curve was 45 degrees、whether residual plots with 0 as the boundary and evenly distributed.

Doing the regression analysis on the original data, we still retain its analysis and analyze the original seven variables with SAS. However, preliminary results showed that the natural increase in the number of explanatory variables is equal to the number of births minus deaths、the total increase in the number equal to the natural increase in the number plus the increase in the number of social. Therefore, the natural increase in the number and total increase in the number were removed from the variables, then use the following three analysis methods-the first method is "Forward Selection", the second method is "Backward method", the third method is "Stepwise".

We found that "the number of births." This variable does not have a linear regression relationship, so it is removed in order to increase explanatory ability of the remaining variable.

We delete the missing values in the data, then using the three test methods to select the best variable, it obtained to the most representative of the linear regression line. Using the residual analysis to test whether residual sum is zero and whether the model is the best linear regression model, so in order to obtain the best variable.

The final conclusion- after the filtering of the three methods, leaving credible variables "the number of deaths"、"the increase in the number of social"、"the number of car accident deaths"、"the number of people abroad".

The following of four variables are factors that we selected affect the total population of relatively large.

Keyword : Taiwan ; regression ; residual analysis ; selection methods ; total population

目錄

第壹章、緒論	
第一節：研究動機、目的	6
第二節：研究背景	7
第三節：研究流程圖	8
第四節：介紹變數	9
第貳章、基本統計資料分析	
第一節：基本敘述統計量	12
第二節：Correlation	16
第參章、原始模式檢定	
第一節：建立迴歸模型	17
第二節：參數檢定	18
第三節：模型適合度檢定	20
第四節：模型解釋能力	21
第肆章、模型的選取方法	
第一節：前進選擇法	22
第二節：向後消去法	24
第三節：逐步回歸法	25
第伍章、殘差分析	
第一節：檢定殘差加總為零	28
第二節：檢定常態性	29
第陸章、結論	31
第柒章、附錄	
第一節：原始檔案	32
第二節：SAS 程式碼	36
參考文獻與資料來源	38

圖目錄

圖 1	總人口成長趨勢圖	7
圖 2	台灣總人口數折線圖	13
圖 3	出生人數折線圖	13
圖 4	死亡人數折線圖	14
圖 5	社會增加人數折線圖	14
圖 6	車禍死亡人數折線圖	15
圖 7	出國人數折線圖	15
圖 8	各變數殘差分析圖	28
圖 9	常態機率圖之一	29
圖 10	常態機率圖之二	30



表目錄

表 1	未刪除相關解釋變數的基本統計量表	10
表 2	未刪除相關解釋變數的相關係數表	11
表 3	基本統計量表	12
表 4	變數相關性表	16
表 5	參數估計值表	17
表 6	變異數分析表	20
表 7	模型解釋能力分析表	21
表 8	前進選擇法總結	22
表 9	向後消去法總結	24
表 10	逐步迴歸法步驟 1	25
表 11	逐步迴歸法步驟 2	26
表 12	逐步迴歸法步驟 3	26
表 13	逐步迴歸法步驟 4	27
表 14	逐步迴歸法總結	27
表 15	殘差之 p-value 檢定	28
表 16	常態性檢定表	29



第壹章、緒論

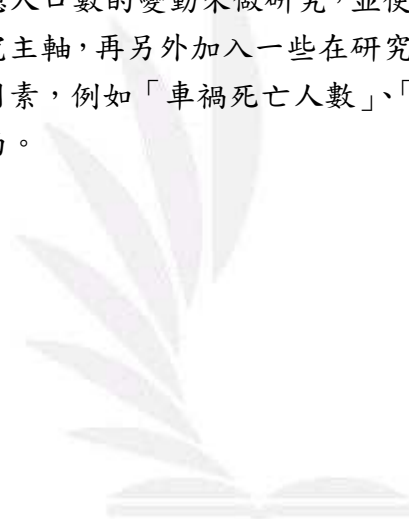
第一節：研究動機與目的

在生活中時常會閱讀或是看到一些有關台灣未來人口的問題，例如：少子化、高齡化社會、全世界出生率最低的國家等相關文章以及報導，因此想探討「台灣總人口數」是受到哪些重要的因素而改變，因為人口問題會是台灣未來重要的議題。

台灣人口的危機會慢慢衍生出更多的社會問題，以出生率全世界最低為例，因為出生人數少，導致勞動人口變少，加上人口高齡化，年輕人漸漸沒有足夠的能力扶養長者等一連串的問題，所以選擇了「出生人數」、「死亡人數」、「自然增加數」來當作解釋變數。

然而台灣人口不僅僅受限於國內因素的影響，近年來外籍配偶或是移民人口逐漸增加，相對的也有台灣人移民去其他國家，這樣的國際因素也會影響到台灣總人口數，因此選擇「社會增加數」當作其中的解釋變數。

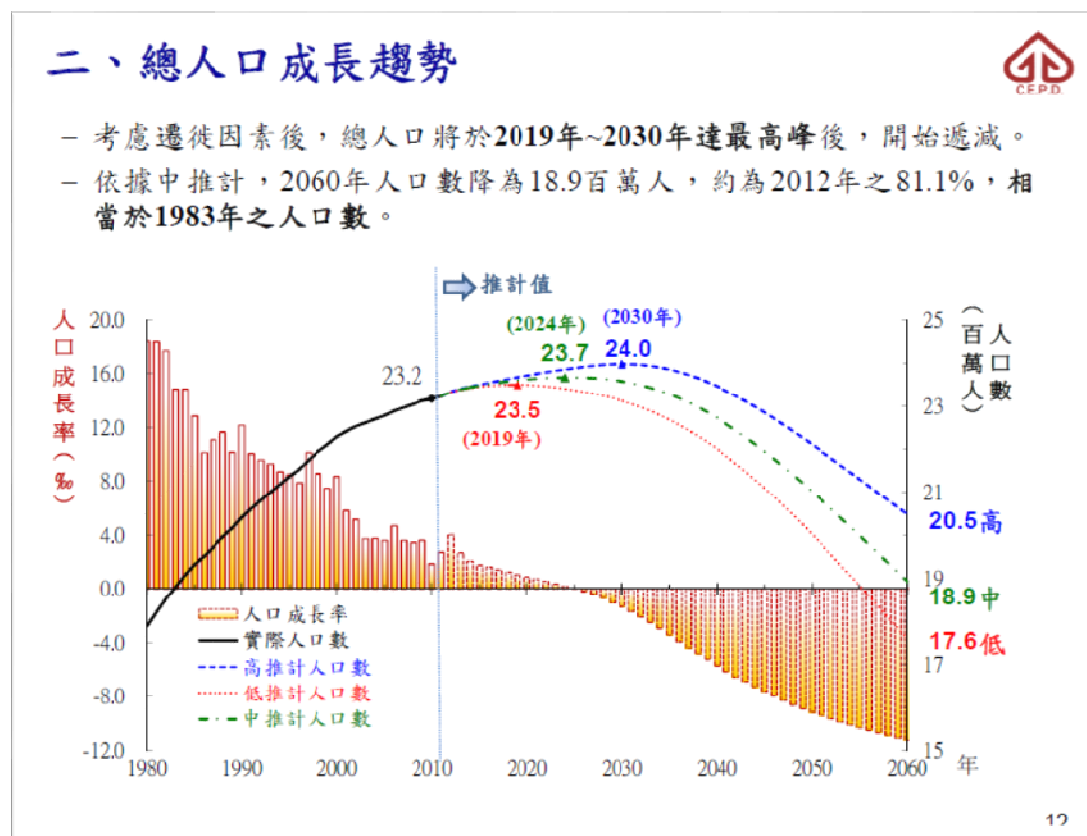
決定上述對於台灣總人口數的變動來做研究，並使用中華民國統計資訊網內提供的資料數據做為研究主軸，再另外加入一些在研究前無法得知對於台灣人口數的變化是否有影響的因素，例如「車禍死亡人數」、「出國人數」等，尋找台灣人口變化不為人知的一面。



第二節：研究背景

在瀏覽行政院經濟建設委員會網站，得到下圖圖 1，圖中以西元 1980 年至 2060 年為區間，由圖 1 中可以得之在 2060 年人口數將降至約 18.9 百萬人，約為 2012 年總人數的 81.1%，相當於 1923 年人口數。

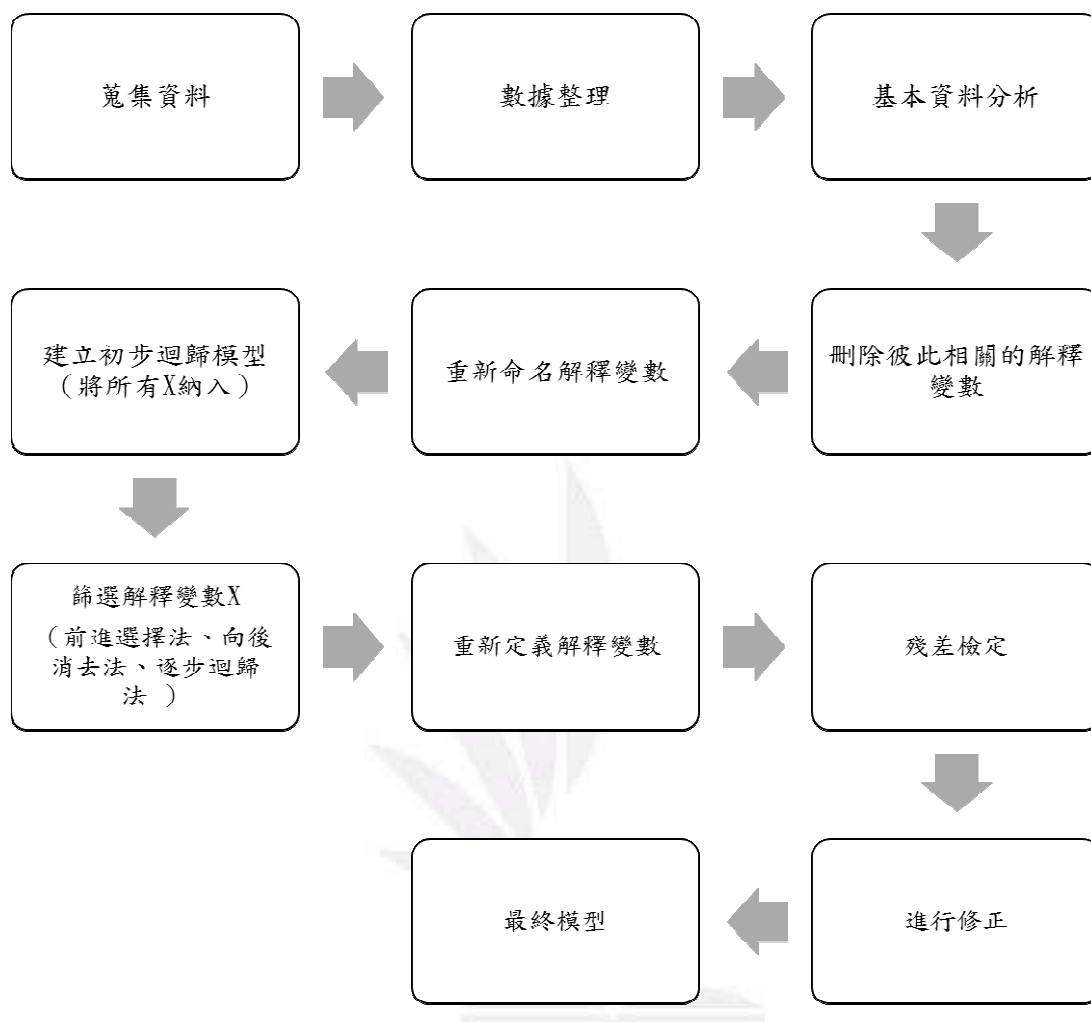
圖 1. 總人口成長趨勢圖



(資料來源：行政院經濟建設委員會「中華民國 2012 年至 2060 年人口推計」簡報。<http://www.cepd.gov.tw/ml.aspx?sNo=0000455>)

上圖 1 中，約在西元 2025 年左右，人口成長率趨近於 0，之後則由正成長轉為負成長，人口不增反減，是因為何種解釋變數有相當的影響程度，導致人口負成長的現象發生，我們將近幾年來的對人口有重要關係的變數一一做分析，從研究的過程中尋找真正的答案

第三節：流程圖



第四節：介紹變數

一、反應變數(Y)

■ 台灣總人口數

醫療進步、衛生條件上升的時代，人口大爆炸是不可避免的，成長的幅度是否也如預期的指數一樣成長或是減少，其中的影響條件甚多，因此作為反應變數，由以下解釋變數去分析。

二、解釋變數(X)

- X_1 ：出生人口數：最直接聯想到的變數，出生人口多，相對總人口也會隨之上升，幾乎是呈現正相關，並觀察近幾年來少子化後，其影響程度的變化。
- X_2 ：死亡人口數：與總人口數呈現負相關，老年化後的台灣，死亡人口數銳減，在另一方面，慢性病及意外事故層出不窮，有時又有各式流感或突變病毒的感染，死亡人數也有時會有大幅的漲跌。
- X_3 ：自然增加人數：是為出生人口數及死亡人數之差，但解釋變數之間有相關性，因此從解釋變數中剔除。
- X_4 ：社會增加人數：包括外籍人士遷戶，是移入人口及移出人口之差，交通改善，到異國定居有許多不同原因：近期因女性結婚率節節下降，許多男士選擇外籍新娘。而外勞遷台工作也是趨勢，其他還有異國婚姻或在台求學外籍生等。
- X_5 ：總增加人數：是為自然增加人數及社會增加數之和，但解釋變數之間有相關性，因此從解釋變數中剔除。
- X_6 ：車禍死亡人數：加入其中一種與死亡有關的變數，查看車禍死亡之人數對總人口數之影響，是否國內對交通意外上仍有疏失，若其關係越大，更應當做好防護措施。
- X_7 ：出國人口數：此變數為「出境」人口數，包含各地區並不分國籍，若總人口與出國人口數具有正相關，即人口越多，出國越多人，此關係到經濟發展、個人所得的上升等，若為負相關則原因可能需更進一步討論。此人口數是對社會增加人口數多做一個解釋的空間。

表 1. 未刪除相關解釋變數的基本統計量表

$x3 =$	$x1 - x2$
$x5 =$	$x1 - x2 + x4$

參數估計值							
變數	標籤	DF	參數估計值	標準誤差	t 值	Pr > t	變異數膨脹
Intercept	Intercept	1	23021148	304891	75.51	<.0001	0
x1	出生人數	B	-18.11292	9.61132	-1.88	0.0643	1.34272
x2	死亡人數	B	49.43918	15.17341	3.26	0.0018	1.11054
x3	自然增加	0	0
x4	社會增加	B	23.58748	10.16712	2.32	0.0238	1.35710
x5	總增加	0	0
x6	車禍死亡人數	1	-3159.61692	523.95606	-6.03	<.0001	1.31095
x7	出國人數	1	0.35178	0.15977	2.20	0.0315	1.52074



表 2. 未刪除相關解釋變數的相關係數表

Pearson 相關係數, N = 66								
Prob > r (位於 H0 底下): Rho=0								
	y	x1	x2	x3	x4	x5	x6	x7
y	1.00000	-0.46939	0.42110	-0.60737	0.41743	-0.25491	-0.70862	0.56892
總人口數		<.0001	0.0004	<.0001	0.0005	0.0389	<.0001	<.0001
x1	-0.46939	1.00000	-0.03664	0.87110	-0.39069	0.52543	0.32130	-0.40654
出生人數	<.0001		0.7703	<.0001	0.0012	<.0001	0.0085	0.0007
x2	0.42110	-0.03664	1.00000	-0.52269	0.00480	-0.49175	-0.28897	0.17601
死亡人數	0.0004	0.7703		<.0001	0.9695	<.0001	0.0186	0.1575
x3	-0.60737	0.87110	-0.52269	1.00000	-0.33566	0.68990	0.41611	-0.43331
自然增加	<.0001	<.0001	<.0001		0.0059	<.0001	0.0005	0.0003
x4	0.41743	-0.39069	0.00480	-0.33566	1.00000	0.45033	-0.16548	0.45393
社會增加	0.0005	0.0012	0.9695	0.0059		0.0001	0.1842	0.0001
x5	-0.25491	0.52543	-0.49175	0.68990	0.45033	1.00000	0.26724	-0.06188
總增加	0.0389	<.0001	<.0001	<.0001	0.0001		0.0301	0.6216
x6	-0.70862	0.32130	-0.28897	0.41611	-0.16548	0.26724	1.00000	-0.38693
車禍死亡人數	<.0001	0.0085	0.0186	0.0005	0.1842	0.0301		0.0013
x7	0.56892	-0.40654	0.17601	-0.43331	0.45393	-0.06188	-0.38693	1.00000
出國人數	<.0001	0.0007	0.1575	0.0003	0.0001	0.6216	0.0013	

- 解釋變數之間如果相關性過高則會有線性重合的問題，因此剔除彼此之間有較高相關性的解釋變數後，將剩餘的解釋變數重新命名為：

X_1 ：出生人數 X_2 ：死亡人數 X_3 ：社會增加

X_4 ：車禍死亡人數 X_5 ：出國人數

- 之後經由 SAS 的向後消去法、前進選擇法、逐步選擇法篩選解釋變數，判定將 X_1 (出生人數) 剔除。

第貳章、基本統計資料分析

第一節：基本敘述統計量

一、基本敘述統計量

Y：臺灣總人口數

X_1 ：出生人數 X_2 ：死亡人數 X_3 ：社會增加

X_4 ：車禍死亡人數 X_5 ：出國人數

表 3. 基本統計量表

簡單統計值							
變數	N	平均值	標準差	總和	最小值	最大值	標籤
y	132	22890972	258109	3021608310	22414047	23315822	總人口數
x1	132	17339	2421	2288783	10597	23398	出生人數
x2	132	11750	1173	1551021	9373	16007	死亡人數
x3	132	1307	2020	172492	-3656	6936	社會增加
x4	132	207.12121	39.37509	27340	136.00000	333.00000	車禍死亡人數
x5	132	702373	136024	92713227	112983	1041365	出國人數

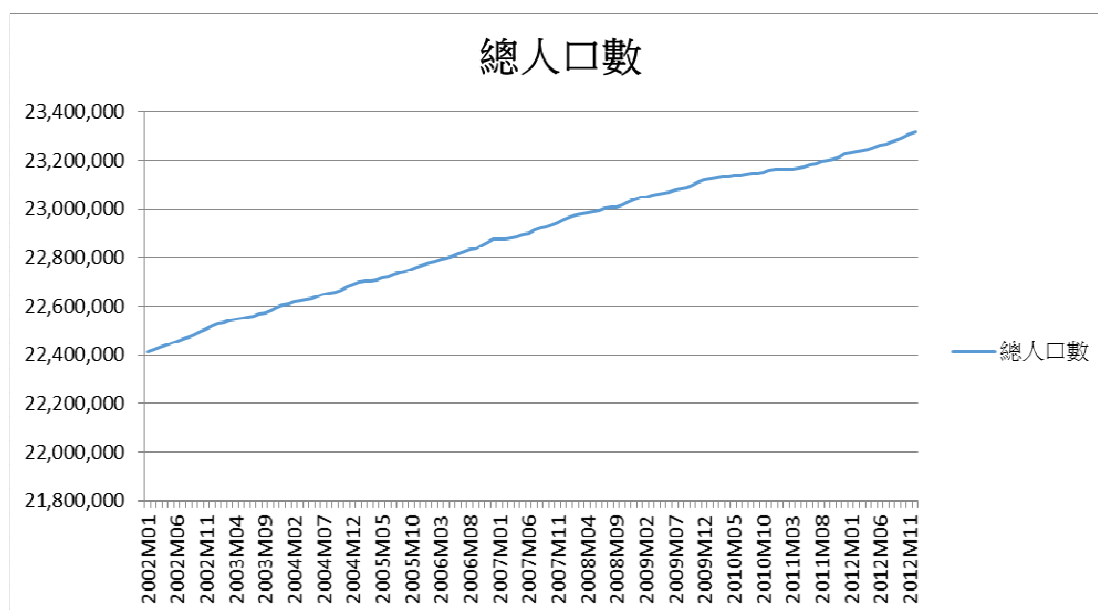


圖 2. (Y: 臺灣總人口數)

由上圖 2 可知，台灣在 2002 年後的總人口數逐年很穩定的增加，但是從 2009 年後曲線趨向平緩，即表示增加幅度逐漸趨緩，台灣總人口數於十年間增加約 900 萬人。

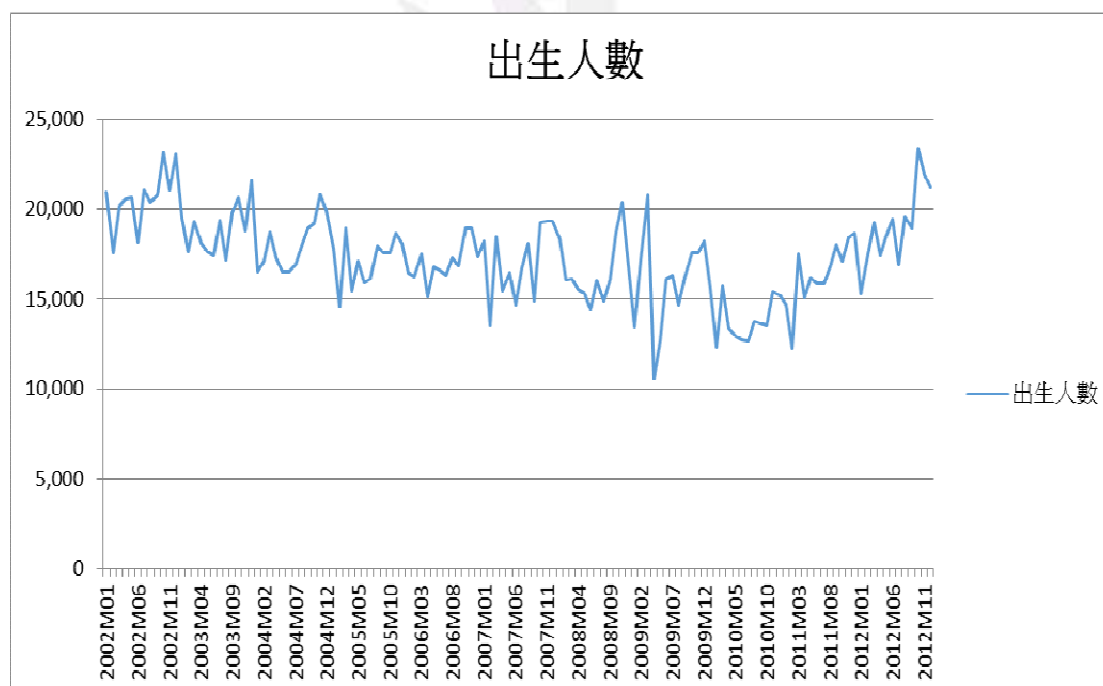


圖 3. (X_1 : 出生人數)

由上圖 3 可知，出生人數從 2002 年開始至 2010 年都是呈現下降的趨勢，但會受到中國習俗—生肖的影響而會有高低起伏，2009 年是最明顯的低點，可能是因為經歷全球金融風暴的影響導致出生人數銳減，而從 2010 年後的出生人數又有逐漸增加的現象。

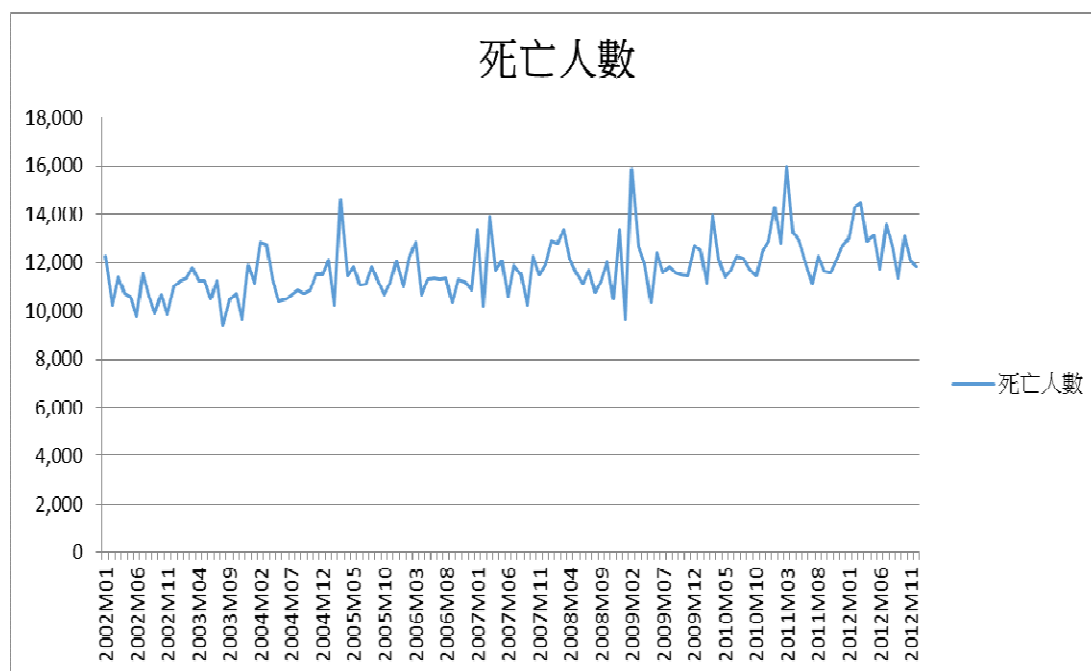


圖 4. (X_2 : 死亡人數)

由上圖 4 可看出歷年來的死亡人數變動幅度不大，約在一萬人至一萬四千人之間，在 2009 年後有稍微上升，可能是因為全球金融風暴和國內經濟不景氣等環境因素影響導致死亡人數遽增。

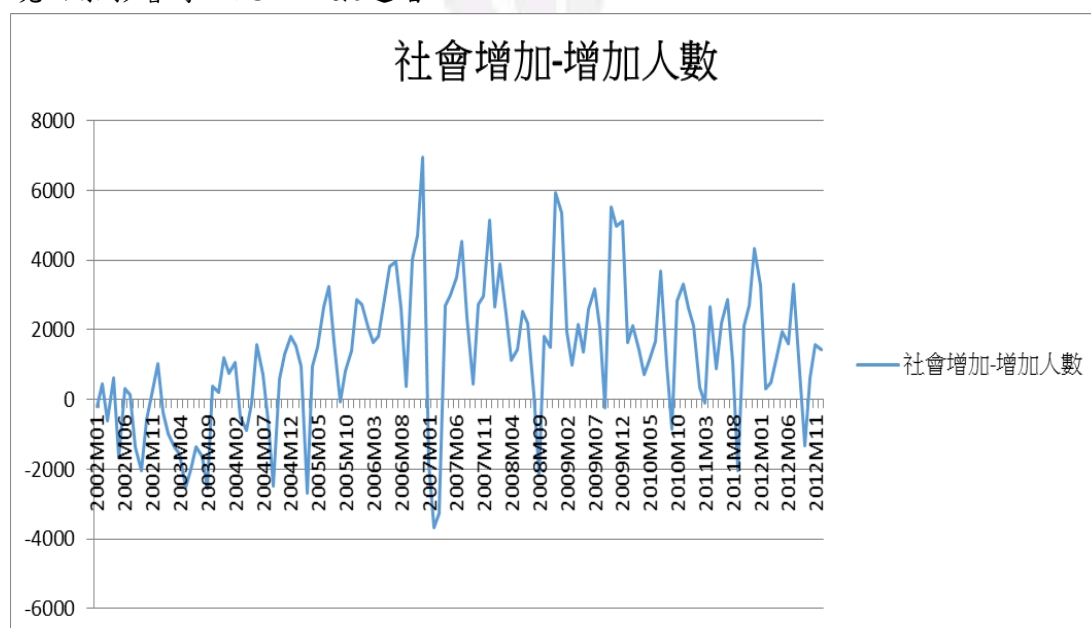


圖 5. (X_3 : 社會增加人數)

由上圖 5 可得知社會增加人數在 2002 年後呈現緩緩上升的狀態，但在 2007 年後開始變成微幅的下降，同年每月變動幅度很大。

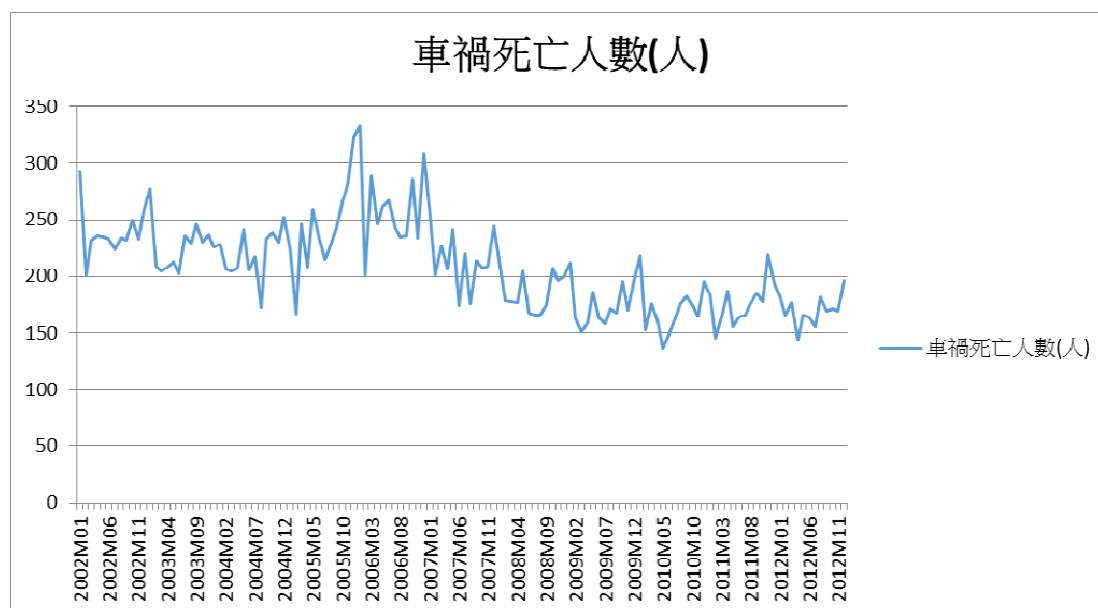


圖 6. (X_4 :車禍死亡人數)

由上圖 6 可以看出車禍死亡人數再 2005 到 2007 年間是最高峰，約在 200 人至 300 人之間變動，而在 2007 年後車禍死亡人數明顯減少至約 150 人至 200 人之間變動。

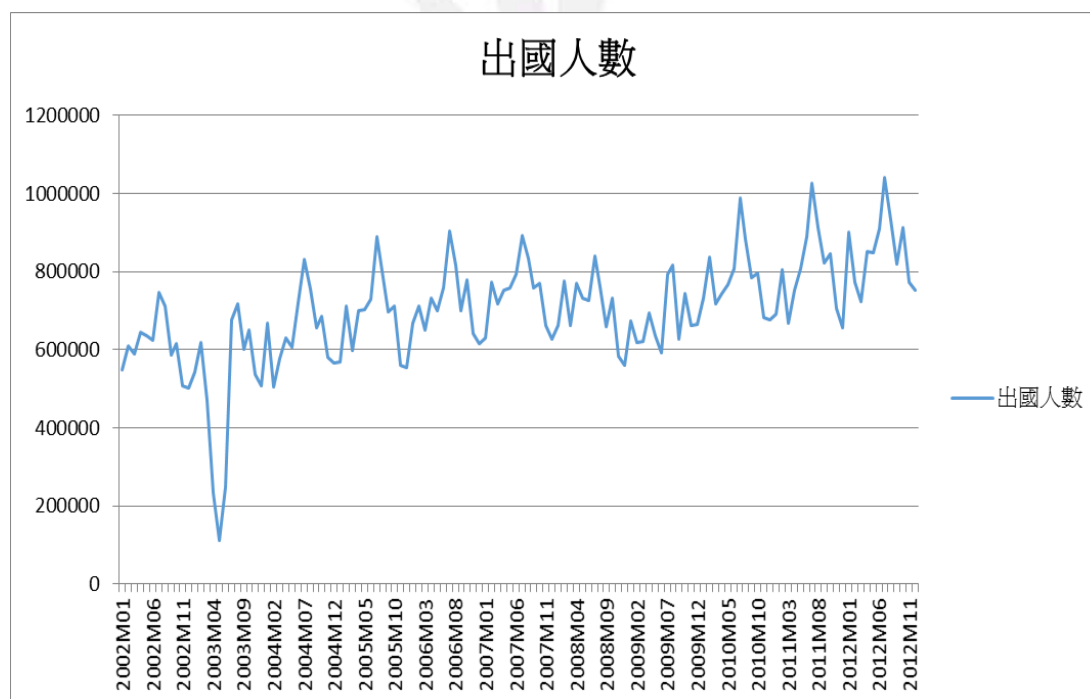


圖 7. (X_5 :出國人數)

由上圖 7 可以看出 2003 年的出國人數為最低點，可能是因為台灣爆發 SARS 疫情和美伊戰爭的影響，除此之外皆呈現穩定增加的趨勢，約在 60 萬至 100 萬之間變動。

第二節、相關(Correlation)

表 4. 變數相關性表

Pearson 相關係數, N = 132 Prob > r (位於 H0 底下): Rho=0						
	y	x1	x2	x3	x4	x5
y	1.00000	-0.38417	0.51027	0.40732	-0.64144	0.58708
總人口數		<.0001	<.0001	<.0001	<.0001	<.0001
x1	-0.38417	1.00000	-0.03025	-0.23187	0.40351	-0.32279
出生人數	<.0001		0.7306	0.0075	<.0001	0.0002
x2	0.51027	-0.03025	1.00000	0.13582	-0.31767	0.15779
死亡人數	<.0001	0.7306		0.1205	0.0002	0.0708
x3	0.40732	-0.23187	0.13582	1.00000	-0.03805	0.31942
社會增加	<.0001	0.0075	0.1205		0.6649	0.0002
x4	-0.64144	0.40351	-0.31767	-0.03805	1.00000	-0.39601
車禍死亡人數	<.0001	<.0001	0.0002	0.6649		<.0001
x5	0.58708	-0.32279	0.15779	0.31942	-0.39601	1.00000
出國人數	<.0001	0.0002	0.0708	0.0002	<.0001	

判斷方式：

(1) 相關係數 (r) 介於 1 到 -1 之間，r 值為正代表為正相關，r 值為負代表為負相關，相關係數大小可將相關程度分為高度、中度、低度。

(2) 相關係數判定：

(a) $0 \leq |r| < 0.3$ → 低度相關

(b) $0.3 \leq |r| \leq 0.7$ → 中度相關

(c) $0.7 < |r| \leq 1$ → 高度相關

第參章、原始模式檢定

第一節：建立迴歸模型

影響總人口數(Y)的因素有出生人數(X_1)、死亡人數(X_2)、社會增加人數(X_3)、車禍死亡人數(X_4)、出國人數(X_5)

表 5. 參數估計值表

參數估計值							
變數	標籤	DF	參數估計值	標準誤差	t 值	Pr > t	變異數膨脹
Intercept	Intercept	1	22349615	208233	107.33	<.0001	0
x1	出生人數	1	-7.14328	5.95694	-1.20	0.2327	1.31529
x2	死亡人數	1	66.98473	11.53291	5.81	<.0001	1.15749
x3	社會增加	1	31.33853	6.79922	4.61	<.0001	1.19337
x4	車禍死亡人數	1	-2600.49445	392.56268	-6.62	<.0001	1.51123
x5	出國人數	1	0.53504	0.10757	4.97	<.0001	1.35428



第二節：參數檢定

為了檢定所有的解釋變數是否與反應變數(台灣總人口數)存在線性關係。

- (1) 我們想要判斷出生人數(X_1)與台灣總人口數(Y)是否存在線性關係，首先

假定其他變數固定之下的情況下，設立假設 $H_0: \beta_1=0$

$$H_1: \beta_1 \neq 0$$

因為 $p\text{-value}=0.2327 > \alpha=0.05$ ，所以不拒絕 H_0 的假設，表示我們有充分的

資料表示 $\beta_1=0$ 。即出生人數(X_1)與總人口數(Y)不存在線性關係。

- (2) 我們想要判斷死亡人數(X_2)與台灣總人口數(Y)是否存在線性關係，首先

先假定其他變數固定之下的情況下，設立假設 $H_0: \beta_2=0$

$$H_1: \beta_2 \neq 0$$

因為 $p\text{-value}<0.0001 < \alpha=0.05$ ，所以拒絕 H_0 的假設，表示我們有充分的

資料表示 $\beta_2 \neq 0$ 。即死亡人數(X_2)與總人口數(Y)存在線性關係。

- (3) 我們想要判斷社會增加人數(X_3)與台灣總人口數(Y)是否存在線性關係，

首先假定其他變數固定之下的情況下，設立假設 $H_0: \beta_3=0$

$$H_1: \beta_3 \neq 0$$

因為 $p\text{-value}<0.0001 < \alpha=0.05$ ，所以拒絕 H_0 的假設，表示我們有充分的

資料表示 $\beta_3 \neq 0$ 。即社會增加人數(X_3)與總人口數(Y)存在線性關係。

(4) 我們想要判斷車禍死亡人數(X_4)與台灣總人口數(Y)是否存在線性關係, 首

先假定其他變數固定之下的情況下, 設立假設 $H_0: \beta_4=0$

$$H_1: \beta_4 \neq 0$$

因為 $p\text{-value} < 0.0001 < \alpha = 0.05$, 所以拒絕 H_0 的假設, 表示我們有充分的

資料表示 $\beta_4 \neq 0$ 。即車禍死亡人數(X_4)與總人口數(Y)存在線性關係。

(5) 我們想要判斷出國人數(X_5)與台灣總人口數(Y)是否存在線性關係, 首先先

假定其他變數固定之下的情況下, 設立假設 $H_0: \beta_5=0$

$$H_1: \beta_5 \neq 0$$

因為 $p\text{-value} < 0.0001 < \alpha = 0.05$, 所以拒絕 H_0 的假設, 表示我們有充分的

資料表示 $\beta_5 \neq 0$ 。即出國人數(X_5)與總人口數(Y)存在線性關係。

本節結論：

將每個變數做假設檢定, 發現出生人口數(X_1)之 $p\text{-value} = 0.2327 > \alpha = 0.05$, 所以不拒絕 H_0 , 及 $\beta_1 = 0$, 表示出生人口數(X_1)與總人口數(Y)沒有存在線性關係。

第三節：模型適合度檢定

以下為利用剩下的五個變數利用 SAS 做適合度檢定，變異數分析表如下表 6：

表 6. 變異數分析表

變異數分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	5	6.117654E12	1.223531E12	59.08	<.0001
誤差	126	2.609597E12	20711083658		
已校正的總計	131	8.727251E12			

統計假設如下：

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1: \beta_i \text{ 不完全為 } 0, i=1, 2, 3, 4, 5$$

因為 $P\text{-value} < 0.0001 < \alpha = 0.05$ ，所以拒絕 H_0 的假設，表示有充分資料顯示

β_i 不完全為 0，即解釋變數(出生人數、死亡人數、社會增加人數、車禍死亡人數、出國人數)與反應變數(台灣總人口數)有線性迴歸相關。

第四節：模型解釋能力

表 7. 模型解釋能力分析表

根 MSE	143913	R 平方	0.7010
應變平均值	22890972	調整 R 平方	0.6891
變異係數	0.62869		

在模型下 $R^2=0.7010$ ，表示樣本迴歸線已經解釋了總變異的 70%，而經過校正的 R^2 只有比原本 R^2 低 0.0119，所以代表 X_i ， $i=2、3、4、5$ 對 Y 有一定程度的解釋能力，接著要做選取變數的分析來完成最終模型



第肆章、模型的選取方法

第一節：前進選擇法(*Forward Selection*)顯著水準=0.05

在 SAS 的 REG 程序中選取解釋變數，將解釋變數依序輸入模型的一種變數選取程序，以參數估計值的 F 統計量最顯著者（即 p-value 最小）優先進入，直到所有變數加進模式內，其 F 統計量皆不顯著時，則選取解釋變數的程序停止。向前選取法一旦變數加入程序中，則此變數將不會從模式中剔除，保留在模式之中。

令 X_1 ：出生人數 X_2 ：死亡人數 X_3 ：社會增加

X_4 ：車禍死亡人數 X_5 ：出國人數

表 8. 前進選擇法總結

前進選擇 的摘要								
步驟	輸入的變數	標籤	變數 Vars In	部分 R 平方	模型 R 平方	C(p)	F 值	Pr > F
1	x4	車禍死亡人數	1	0.4114	0.4114	120.006	90.88	<.0001
2	x3	社會增加	2	0.1468	0.5583	60.1306	42.88	<.0001
3	x2	死亡人數	3	0.0759	0.6342	30.1326	26.57	<.0001
4	x5	出國人數	4	0.0634	0.6976	5.4380	26.60	<.0001

步驟 1：系統選取的解釋變數是 X_4 ，F 統計量為 90.88 (p-value<0.0001< α =0.05)。

步驟 2：系統選取的解釋變數是 X_3 ，F 統計量為 81.52 (p-value<0.0001< α =0.05)。

步驟 3：系統選取的解釋變數是 X_2 ，F 統計量為 73.98 (p-value<0.0001< α =0.05)。

步驟 4：系統選取的解釋變數是 X_5 ，F 統計量為 73.23 (p-value<0.0001< α =0.05)。

影響台灣總人口數的因素

之後系統發現未進入模式為 X_1 ，在顯著水準 0.05 下並不顯著，因此不放入模型

中。利用向前選取法選取解釋變數(X_2, X_3, X_4, X_5)的線性模式迴歸完成，其迴歸方程式為：

$$\hat{Y} = 22349615 + 66.98473X_2 + 31.33853X_3 - 2600.49445X_4 + 0.53504X_5$$

第二節：向後消去法 (backward method) 顯著水準=0.05

在 SAS 的 REG 程序中，將所有解釋變數輸入模型中，將不顯著的變數從模型中剔除，原則是參數估計值的 F 統計量最小最不顯著者（即 p-value 最大）優先剔除，直到所有變數加進模式內，其 F 統計量皆顯著時，則選取解釋變數的程序停止。向後消去法一旦變數剔除至程序外，則此變數將不會再次進入此模型。

令 X_1 ：出生人數 X_2 ：死亡人數 X_3 ：社會增加

X_4 ：車禍死亡人數 X_5 ：出國人數

表 9. 向後消去法總結

向後消去的摘要								
步驟	移除的變數	標籤	數字 Vars In	部分 R 平方	模型 R 平方	C(p)	F 值	Pr > F
1	x1	出生人數	4	0.0034	0.6976	5.4380	1.44	0.2327

步驟 1：系統將所有變數放至模式中，而由模式中所得 F 統計量得知解釋變數 X_1 ，F 統計量為 1.44 (p-value=0.2327 > $\alpha=0.05$) 是最不顯著的，因此被剔除。

步驟 2：系統再將剩餘變數 X_2 、 X_3 、 X_4 、 X_5 重新建立模式，而由新的模式得知 F 統計量，發現 X_2 、 X_3 、 X_4 、 X_5 皆具有顯著性，因此不被剔除，其迴歸方程式為：

$$\hat{Y} = 22349615 + 66.98473X_2 + 31.33853X_3 - 2600.49445X_4 + 0.53504X_5$$

第三節：逐步迴歸法(Stepwise method)顯著水準=0.05

此方法結合上述的「前進選擇法」與「向後消去法」，先利用前進選擇法將模式外所剩餘的變數選入，其 F 統計量顯著者（即 $p\text{-value} < \alpha = 0.05$ ）進入，之後在利用向後消去法對已輸入的變數重新檢定，將其中 F 統計量不顯著（即 $p\text{-value} > \alpha = 0.05$ ）的變數移除至模式外，直到模式外的變數皆不顯著或是進入模式中的變數是之前被系統剔除的，此時所留下的變數為最後選取的解釋變數。

步驟 1：輸入變數 X_4 : R 平方 = 0.4114 和 $C(p) = 120.0055$

表 10. 逐步迴歸法步驟 1

變異數分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	1	3.590788E12	3.590788E12	90.88	<.0001
誤差	130	5.136463E12	39511252850		
已校正的總計	131	8.727251E12			

變數	參數估計值	標準誤差	第二型 SS	F 值	Pr > F
Intercept	23761860	92978	2.580609E15	65313.3	<.0001
x4	-4204.72608	441.06565	3.590788E12	90.88	<.0001

步驟 2：輸入變數 X_3 ：R 平方 = 0.5583 和 $C(p) = 60.1306$

表 11. 逐步迴歸法步驟 2

變異數分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	2	4.872285E12	2.436142E12	81.52	<.0001
誤差	129	3.854966E12	29883458910		
已校正的總計	131	8.727251E12			

變數	參數估計值	標準誤差	第二型 SS	F 值	Pr > F
Intercept	23678026	81867	2.499773E15	83650.7	<.0001
x3	48.99414	7.48171	1.281497E12	42.88	<.0001
x4	-4109.07990	383.86002	3.42432E12	114.59	<.0001

步驟 3：輸入變數 X_2 ：R 平方 = 0.6342 和 $C(p) = 30.1326$

表 12. 逐步迴歸法步驟 3

變異數分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	3	5.534997E12	1.844999E12	73.98	<.0001
誤差	128	3.192253E12	24939479250		
已校正的總計	131	8.727251E12			

變數	參數估計值	標準誤差	第二型 SS	F 值	Pr > F
Intercept	22801630	185736	3.758619E14	15071.0	<.0001
x2	64.50223	12.51284	6.627129E11	26.57	<.0001
x3	44.35350	6.89388	1.032322E12	41.39	<.0001
x4	-3507.74684	369.56570	2.24678E12	90.09	<.0001

步驟 4：輸入變數 X_5 ：R 平方 = 0.6976 和 C(p) = 5.4380

表 13. 逐步迴歸法步驟 4

變異數分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	4	6.087872E12	1.521968E12	73.23	<.0001
誤差	127	2.639378E12	20782506913		
已校正的總計	131	8.727251E12			

變數	參數估計值	標準誤差	第二型 SS	F 值	Pr > F
Intercept	22271900	198232	2.62341E14	12623.2	<.0001
x2	64.91584	11.42278	6.712061E11	32.30	<.0001
x3	33.00892	6.66646	5.095296E11	24.52	<.0001
x4	-2771.72344	366.30194	1.189924E12	57.26	<.0001
x5	0.55134	0.10689	5.52875E11	26.60	<.0001

留在模型中的所有變數都是顯著於 0.1500 層級。

沒有其他變數符合輸入至模型中的 0.1500 顯著層級。

因此，系統先選取 X_4 ，其 F=90.88 (p-value<.0001<0.15) 為所有變數最顯著；

接著系統再檢定模式內的唯一解釋變數 X_3 ，結果為顯著的，因此不從模式中剔除。

剩下的四個變數也是如此，最後系統在逐步迴歸法選擇了解釋變數 (x2, x3, x4, x5) 的線性模式迴歸完成迴歸方程式為：

$$\hat{Y} = 22349615 + 66.98473X_2 + 31.33853X_3 - 2600.49445X_4 + 0.53504X_5$$

表 14. 逐步迴歸法總結

逐步選擇的摘要									
步驟	輸入的變數	移除的變數	標籤	變數 Vars In	部分 R 平方	模型 R 平方	C(p)	F 值	Pr > F
1	x4		車禍死亡人數	1	0.4114	0.4114	120.006	90.88	<.0001
2	x3		社會增加	2	0.1468	0.5583	60.1306	42.88	<.0001
3	x2		死亡人數	3	0.0759	0.6342	30.1326	26.57	<.0001
4	x5		出國人數	4	0.0634	0.6976	5.4380	26.60	<.0001

第五章、殘差分析

第一節：檢定殘差加總為零

本組之殘差圖資料，刪除不具代表性之變數後，剩餘變數之殘差圖形，除了「出國人數」有離群值，導致殘差分配不均，其餘皆均勻分布在 $e_i = 0$ 上下側，可推定其殘差加總值為零。

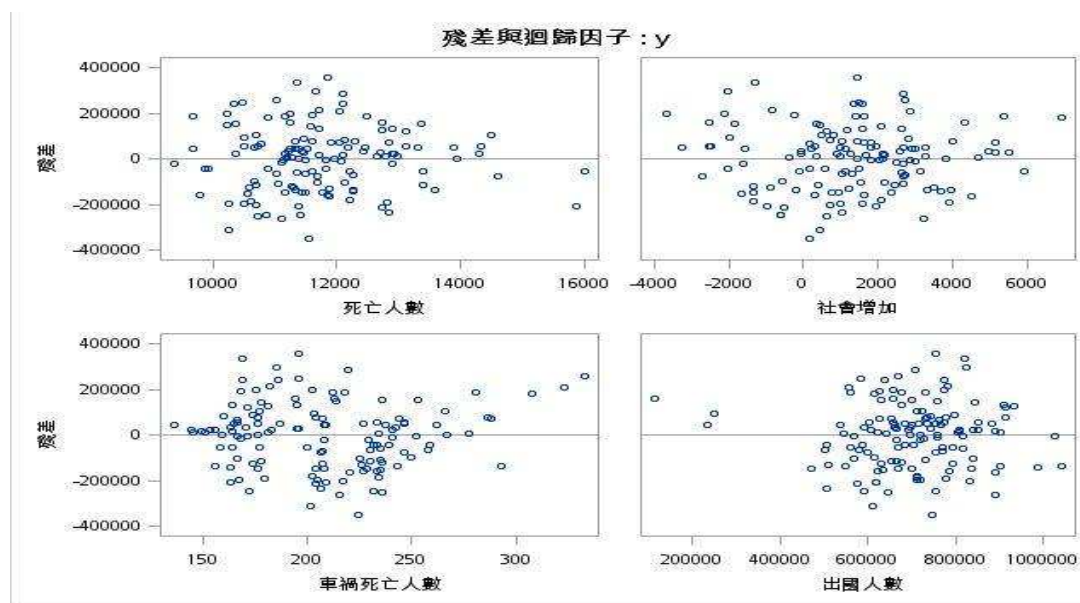


圖 8. 各變數殘差分析圖

為了希望能更加確認誤差期望值為零，做了以下檢定：

H_0 ：殘差加總為零

H_1 ：殘差加總不為零

發現本組之 p-value 值皆大於顯著水準=0.05，代表不拒絕虛無假設，所以殘差加總為零。

表 15. 殘差之 p-value 檢定

位置檢定: $\mu_0=0$			
檢定	統計值	p 值	
Student's t	t	0.050098	Pr > t 0.9601
符號	M	3	Pr >= M 0.6636
符號秩	S	16	Pr >= S 0.9712

第二節：檢定常態性

方法(1)、一般檢定常態皆利用常態檢定圖，當以下假設成立，常態機率圖也應呈現 45 度角。

設立假設 H_0 ：誤差項遵循常態分配

H_1 ：誤差項未遵循常態分配

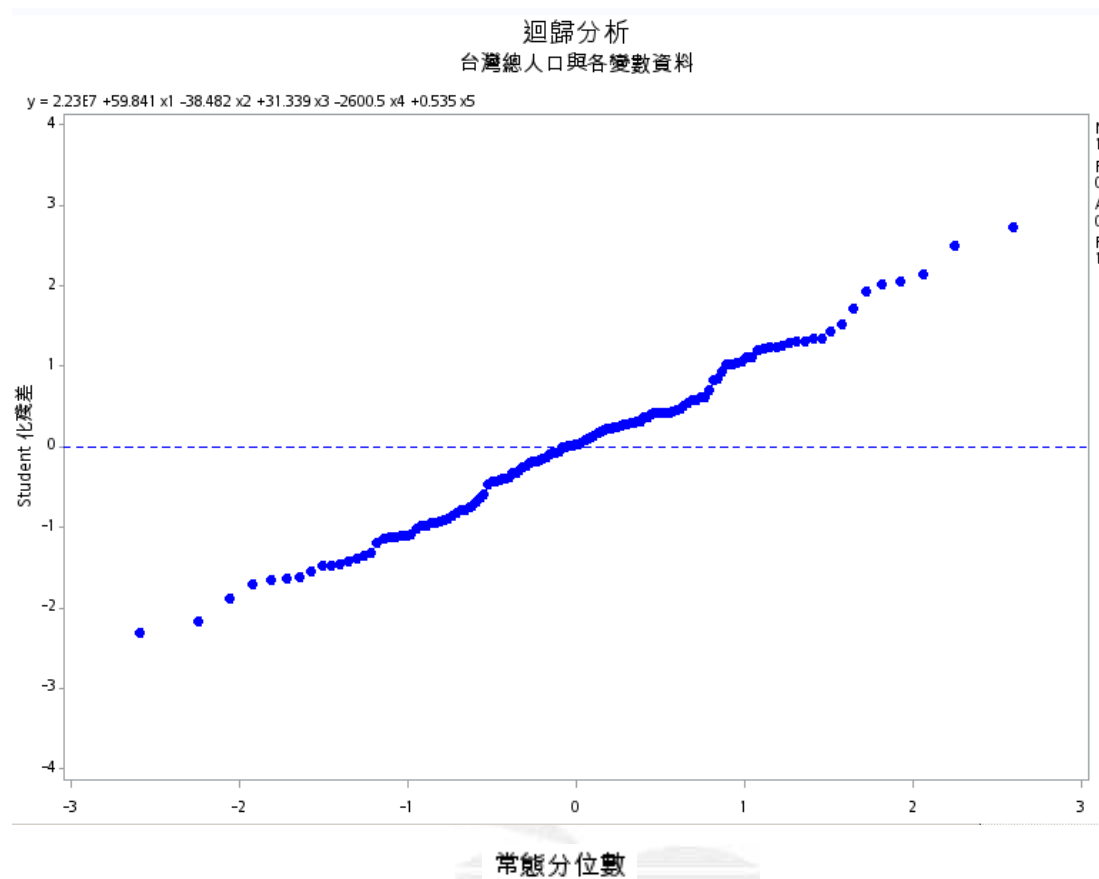


圖 9. 常態機率圖之一

根據 Shapiro-Wilk test, $P\text{-value} = 0.5435 > \text{信心水準} = 0.05$ ，所以符合常態分配

表 16. 常態性檢定表

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.990869	Pr < W	0.5435
Kolmogorov-Smirnov	D	0.053792	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.063209	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.389173	Pr > A-Sq	>0.2500

方法(2)由常態分配圖來看，所有的數值在圖上呈現 45 度的圖形，可得知為誤差項的假設服從常態分配。

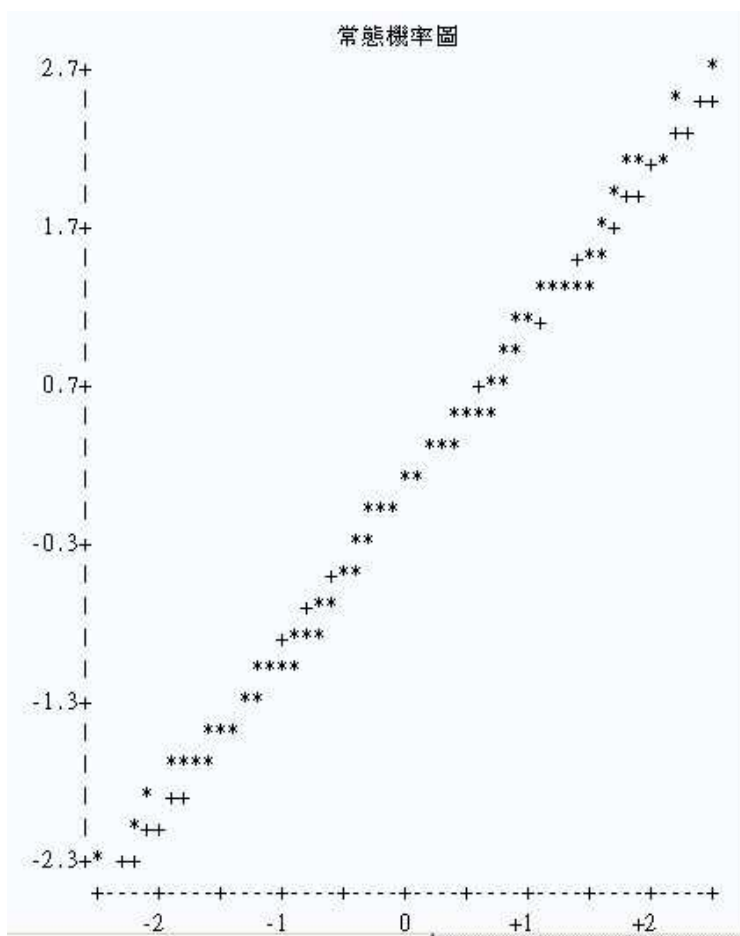


圖 10. 常態機率圖之二

第陸章、結論

這份報告主要是探討「出生人數」、「死亡人數」、「自然增加人數」、「總增加人數」、「社會增加人數」、「車禍死亡人數」、「出國人數」此七個變數對台灣總人口的影響，我們使用前進選擇法、後退消去法和逐步迴歸法來選擇最佳的變數。在檢定過後刪除

「死亡人數」、「自然增加人數」、「總增加人數」這四項假設的解釋變數。

接著利用檢定通過的四個變數來完成複迴歸方程式，其迴歸方程式為

$$\hat{Y}=22349615+66.98473X_2+31.33853 X_3-2600.49445 X_4+0.53504 X_5$$

在此線性迴歸模式，各參數代表的意義為：

β_0 ：當不考慮 X_2 、 X_3 、 X_4 、 X_5 影響的情況下，即 X_2 、 X_3 、 X_4 、 $X_5=0$ 時，Y 平均的值為 22349615

β_2 ：當其他變數 X_3 、 X_4 、 X_5 不變時，每增加一單位的 X_2 ，Y 的平均值會增加 66.98473
即不考慮社會增加人數、車禍死亡人數、出國人數的情況，當死亡人數每增加一單位，台灣總人口之平均值則會增加 66.98473 人。

β_3 ：當其他變數 X_2 、 X_4 、 X_5 不變時，每增加一單位的 X_3 ，Y 的平均值會增加 31.33853
即不考慮死亡人數、車禍死亡人數、出國人數的情況，當社會增加人數每增加一單位，台灣總人口之平均值則會增加 31.33853 人。

β_4 ：當其他變數 X_2 、 X_3 、 X_5 不變時，每增加一單位的 X_4 ，Y 的平均值會減少 2600.49445
即不考慮死亡人數、社會增加人數、出國人數的情況，當車禍死亡人數每增加一單位，台灣總人口之平均值則會減少 2600.49445 人。

β_5 ：當其他變數 X_2 、 X_3 、 X_4 不變時，每增加一單位的 X_5 ，Y 的平均值會增加 0.53504
即不考慮死亡人數、社會增加人數、車禍死亡人數的情況，當出國人數每增加一單位，台灣總人口之平均值則會增加 0.53504 人。

當死亡人數、社會增加人數、出國人數越高，在相對的情況下，台灣的總人口數也會相對增加。

第柒章、附錄

第一節：原始檔案

總人口數	出生人數	死亡人數	社會增加-增加人數	車禍死亡人數(人)	出國人數
22414047	20931	12252	-200	293	547259
22421793	17551	10264	459	201	610394
22429972	20185	11413	-593	232	589178
22440435	20565	10713	611	236	645551
22448817	20603	10566	-1655	235	635046
22457488	18137	9775	309	233	623611
22467176	21093	11551	146	224	745914
22475585	20365	10613	-1343	234	710749
22484364	20735	9915	-2041	232	586296
22496285	23175	10676	-578	250	615504
22507693	21120	9868	156	233	506730
22520776	23070	11030	1043	258	503234
22528673	19533	11264	-372	277	542979
22533951	17626	11369	-979	209	619199
22540155	19312	11765	-1343	204	470929
22545429	18092	11242	-1576	208	232836
22549292	17634	11251	-2520	213	112983
22554253	17448	10493	-1994	203	249789
22560996	19368	11269	-1356	236	678036
22567203	17196	9373	-1616	229	718522
22573965	19813	10507	-2544	246	601265
22584281	20646	10703	373	230	651386
22593641	18815	9680	225	237	536481
22604550	21587	11885	1207	226	508667
22610665	16531	11168	752	228	668148
22615997	17128	12849	1053	206	506105
22621478	18735	12742	-512	204	576992
22626589	17274	11297	-866	207	630184
22632661	16527	10369	-86	241	605731
22640250	16524	10504	1569	205	720843
22647160	16897	10695	708	217	831624

影響台灣總人口數的因素

22653642	17932	10853	-597	172	755375
22659341	18905	10734	-2472	233	655359
22668274	19223	10874	584	239	684425
22678940	20858	11502	1310	230	579489
22689122	19885	11505	1802	252	566377
22696349	17844	12138	1521	225	568456
22701627	14567	10258	969	167	712463
22703295	18956	14592	-2696	246	598709
22708280	15473	11442	954	208	700031
22715030	17085	11833	1498	259	701569
22722559	15920	11063	2672	235	729041
22730819	16119	11097	3238	215	890098
22738559	17933	11840	1647	227	785337
22744839	17559	11231	-48	242	696855
22752547	17584	10684	808	266	710221
22761464	18699	11167	1385	281	560080
22770383	18115	12053	2857	323	555265
22778581	16477	11021	2742	333	667004
22784720	16264	12214	2089	202	710619
22791052	17524	12832	1640	288	651368
22797314	15126	10658	1794	246	732699
22805647	16757	11314	2890	262	699865
22814636	16558	11376	3807	267	757014
22823604	16340	11320	3948	243	902710
22832173	17311	11354	2612	235	816489
22839043	16864	10357	363	236	698682
22850661	18927	11316	4007	286	778136
22863122	18957	11192	4696	234	641668
22876527	17354	10885	6936	308	615121
22879510	18224	13374	-1867	253	631110
22879132	13496	10218	-3656	202	772835
22880454	18490	13890	-3278	227	715988
22886902	15448	11679	2679	206	752006
22894283	16463	12083	3001	241	757205
22901897	14687	10583	3510	174	792506
22911292	16737	11870	4528	220	891395
22920250	18109	11493	2342	176	834790

影響台灣總人口數的因素

22925311	14863	10238	436	214	758748
22934997	19233	12260	2713	207	769592
22945782	19311	11510	2984	209	661099
22958360	19353	11913	5138	244	626438
22966459	18323	12897	2673	208	662847
22973622	16065	12805	3903	179	776687
22978800	16114	13378	2442	178	663305
22983286	15549	12195	1132	177	769832
22988428	15309	11594	1427	204	733000
22994262	14429	11104	2509	168	726563
23000827	16063	11696	2198	166	838607
23005067	14866	10782	156	166	761432
23007737	16048	11245	-2133	175	657988
23016257	18740	12040	1820	207	732191
23027672	20388	10487	1514	196	583775
23037031	16839	13401	5921	200	558945
23046177	13469	9689	5366	212	674519
23049407	17159	15877	1948	163	618982
23058550	20773	12635	1005	151	620797
23059339	10597	11953	2145	159	695011
23063027	12646	10327	1369	186	635635
23069345	16157	12422	2583	163	591187
23077191	16297	11609	3158	158	794198
23082125	14699	11824	2059	171	817785
23086441	16170	11610	-244	168	625834
23098049	17574	11476	5510	195	742345
23109141	17556	11450	4986	170	661050
23119772	18213	12710	5128	196	665603
23124514	15603	12496	1635	218	731609
23127845	12364	11150	2117	153	837946
23131093	15750	13943	1441	176	717854
23133074	13371	12125	735	160	743940
23135715	12902	11379	1118	136	767294
23138381	12722	11714	1658	149	806972
23142460	12668	12259	3670	163	987584
23145020	13761	12184	983	176	883581
23146090	13637	11722	-845	182	782978

影響台灣總人口數的因素

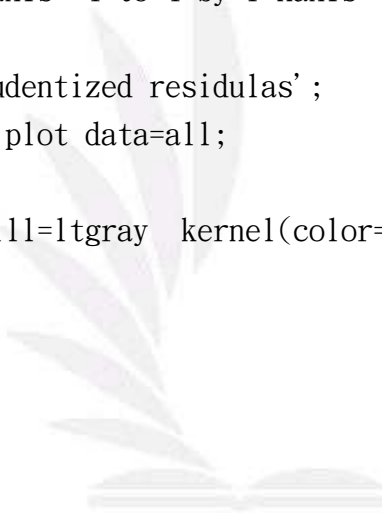
23150923	13487	11469	2815	174	795114
23157178	15406	12444	3293	165	682970
23162123	15215	12887	2617	195	677232
23164628	14703	14313	2115	183	690743
23164457	12270	12783	342	145	804856
23165878	17524	16007	-96	164	667160
23170321	15089	13306	2660	187	752558
23174528	16206	12905	906	156	809015
23180477	15866	12098	2181	164	890332
23188078	15887	11141	2855	166	1026829
23193638	16797	12287	1050	176	911458
23197947	18003	11660	-2034	185	823272
23205605	17136	11581	2103	178	844814
23214620	18443	12109	2681	219	706108
23224912	18703	12725	4314	194	656728
23230506	15296	12991	3289	181	901293
23234003	17500	14318	315	166	772495
23239268	19283	14487	469	177	723643
23245018	17405	12919	1264	144	851486
23252392	18542	13120	1952	166	849354
23261747	19482	11719	1592	164	908600
23268372	16897	13576	3304	156	1041365
23276441	19615	12726	1180	181	933988
23282670	18908	11359	-1320	169	818593
23293593	23398	13107	632	171	912063
23305021	21953	12086	1561	169	773440
23315822	21202	11843	1442	196	753440

第二節：SAS 程式碼

以下程式碼為原始資料(變數未做刪減前)

```
dm "output;clear;log;clear;program;recall;graph;cler;";
options ps=55;
title '迴歸分析';
title2 '台灣總人口與各變數資料';
data price;
input y x1 x2 x3 x4 x5;
label
y='總人口數'
x1='出生人數'
x2='死亡人數'
x3='社會增加'
x4='車禍死亡人數'
x5='出國人數'
cards;
*ods rtf;
ods graphics on;
proc corr;
var y x1 - x5;
proc gplot;
plot y*(x1 x2 x3 x4 x5 );
symbol1 v=star c=red;
proc reg;
model y=x1 -x5 /p r clm cli dw vif;
symbol2 v=dot c=blue;
run;
proc reg data=price;
model y=x1 x2 x3 x4 x5 /selection=backward slstay=0.05;
model y=x1 x2 x3 x4 x5 /selection=forward slentry=0.05;
model y=x1 x2 x3 x4 x5 /selection=stepwise;
model y=x1 x2 x3 x4 x5 /selection=adjrsq;
model y=x1 x2 x3 x4 x5 /selection=cp;
plot cp.*np.
/ chocking=red cmallows=blue
vaxis= 0 to 12 by 2 cframe=ligr;
symbol1 v=dot c=blue;
plot y*(x1 x2 x3 x4 x5) / conf;
```

```
run;
proc reg outest=est;
model y=x1 x2 x3 x4 x5 /selection=stepwise;
model y=x1 x2 x3 x4 x5 /selection=adjrsq cp best=5;
symbol1 v=dot c=blue;
plot y*(x1 x2 x3 x4 x5 ) / conf;
proc gplot data=est;
plot _cp*_p_/vaxis=0 to 10 by 0.5 haxis=0 to 10 by 0.5;
proc reg data=price;
model y=x1-x5/p r influence clm cli dw;
output out=all student=student rstudent=rstudent;
plot y* (x1 x2 x3 x4 x5);
plot y* (x1 x2 x3 x4 x5) / pred;
plot (student. rstudent. )*(x1 x2 x3 x4 x5 predicted.);
plot student. * nqq./vaxis=-4 to 4 by 1 haxis=-3 to 3 by 1.0;
run;
title 'Histogram of studentized residulas';
proc univariate normal plot data=all;
var student;
histogram student / cfill=ltgray kernel(color=red) name='MyHist';
run;
*ods rtf close;
ods graphics off;
ods html close;
run;
quit;
```



參考文獻

1. 中華民國統計資訊網。
<http://ebas1.ebas.gov.tw/pxweb/Dialog/statfile9L.asp>
2. 行政院經濟建設委員會「中華民國 2012 年至 2060 年人口推計」簡報。
<http://www.cepd.gov.tw/m1.aspx?sNo=0000455>
3. 高雄市發展銀髮產業之可行性研究。
<http://c034.wtuc.edu.tw/ezcatfiles/c034/img/img/938/785696794.pdf>

