

逢甲大學學生報告 ePaper

## 美國真的那麼危險嗎？

Is America really that dangerous?

作者：劉育蕙、林虹妤、葉兆鴻、陳啟昌、劉贊隆、唐卓悅、張文薰、陳怡臻

系級：統計學系

學號：D0276721、D0231469、D0289301、D0276675、D0130082、D0262133、D0276704、D0276853

開課老師：陳婉淑

課程名稱：迴歸分析

開課系所：統計學系

開課學年：104 學年度 第一學期

## 摘要

這是一份調查美國 2012 年, 51 州的犯罪率與 10 個因子之間的關係的報告。10 個因子包括, 失業率, 人口數, 人均 GDP, 人均教育花費, 綜合稅, 人均社會福利, 人均醫療花費, 消費者物價指數(CPI), 人均年消費, 警力比例。我們使用了複迴歸分析, 選擇變數, 多重共線性, 殘差分析等方法去分析資料。資料分析發現: 人均 GDP, 人均社會福利以及 CPI 與犯罪率之間呈負相關的關係; 而失業率, 人口數, 人均教育花費, 綜合稅, 人均醫療花費, 人均年花費以及警力比例這 7 個變數與犯罪率呈正相關的關係。換句話說, 當人均 GDP 或人均社會福利或 CPI 上升的時候, 犯罪率會隨著下降, 反之亦然。

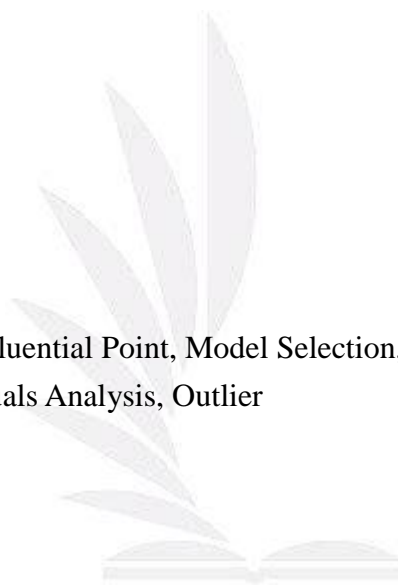


關鍵字：迴歸分析、異常點、影響點分析、多重共線性、殘差分析、選擇重要變數、犯罪率

### **ABSTRACT**

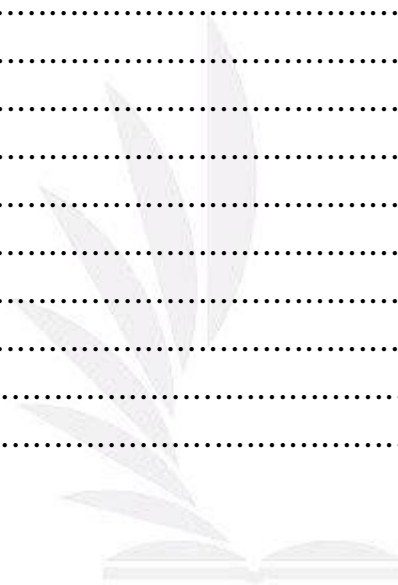
This report examines the relationship between crime rate in 51 states of the USA in 2012 and 10 other factors including: GDP per capita, social welfare per capita, the consumer price index (CPI), unemployment rate, population, education spending per capita, income tax, health spending per capita, overall spending per capita, and the proportion of police. Using the multiple regression analysis, we work on model selection and residual analysis. Our findings suggest: (1) GDP per capita, social welfare per capita and CPI have negative correlation with the crime rate; (2) the unemployment rate, population, education spending per capita, income tax, health spending per capita, overall spending per capita, and the proportion of police have positive correlation with the crime rate. In other words, high GDP per capita, social welfare per capita and CPI lead to the decrease in crime rate while the other factors increase it.

Key Words: Crime Rate, Influential Point, Model Selection, Multicollinearity, Regression Analysis, Residuals Analysis, Outlier



## 目錄

摘要 .....	1
第一章 緒論.....	4
第一節 研究動機.....	4
第二節 研究目的.....	4
第三節 研究方法.....	4
第二章 資料分析.....	5
第一節 一般敘述統計、散佈圖.....	5
第二節 全部變數.....	6
第三節 選取重要變數.....	7
第四節 建立模式.....	8
第三章 診斷分析.....	9
第一節 殘差分析.....	9
第二節 影響點判斷.....	9
第三節 殘差檢定.....	12
第四章 結論.....	13
附錄 .....	14
資料出處 .....	14
參考資料 .....	14
表.....	15
圖.....	16



## 第一章 緒論

這是一份我們針對 2012 年美國犯罪率的研究報告，資料是由美國多個網站收集而來，例如 CityRating.com, 這網站提供我們 CPI 等資料，也有從美國勞動教育部門、美國人口普查局，分別得到費用花費以及人口統計數據。犯罪率是指在一時空範圍內犯罪者與人口總數計算的比率，是一個犯罪的相對指標，以每十萬人口刑事案件發生件數，也就是刑案發生數除以年中(或月中)人口數乘上 100,000。各個國家刑法的規範是不同的，所以這份研究不能套用在各個國家，犯罪率無法衡量各國家的社會治安，也就是說一些在美國是合法的行為，但在其他國家卻是不合法的，例如大麻使用、槍械存在，而我們的研究中，以美國的 51 個州為例，美國是全球的大指標，有些指標甚至會影響全球。現今的社會越來越不穩定，甚至有恐怖攻擊的發生，次數是越來越頻繁，造成全球人民人心惶惶，深怕自己是恐怖攻擊的下一個目標，常見的犯罪行為就是偷拐搶騙，這是每個國家都會發生的，差別在於多還是少，這些犯罪行為人是因為什麼原因而導致他們去做這些事情，不管是犯罪人的心理還是生理，甚至是家裡因素導致這些行為的偏差，在這份研究中，我們探討的是整個社會，當然貧富差距和城鄉差距一定會存在，我們找了十個因素來探討是否影響犯罪率，與犯罪率存在著什麼關係，我們也想知道這些因素影響的程度，是不是適合用來分析犯罪率，更重要的是如何降低犯罪率，讓國家更為和平。

### 第一節 研究動機

很多人，因為美國電影的原因，都有一個美國夢，想要移居美國。但一直以來，很多報章雜誌，電視新聞，都會報導不少美國的犯罪事件，例如連環殺手，校園槍擊事件等等。美國真的那麼危險嗎？到底甚麼原因影響了這犯罪率？這就成為了我們想要去研究美國的犯罪率的動機。

### 第二節 研究目的

利用各州犯罪率的資料，我們想要研究有甚麼因素會影響到犯罪率，從而構思如何改善犯罪率，或者避免潛在的犯罪率。

### 第三節 研究方法

我們首先手檢美國 51 州的各州犯罪率以及我們挑選的覺得會有影響的變數，例如人口數，GDP 等等。再利用變異數分析去檢定我們這些自變數與應變數之間是否存在相關性。在發現有相關性之後，我們再去利用複迴歸去分析各因變數與應變數之間的相關性。因為迴歸分析需要滿足殘差的假設，所以我們也做了殘差的假設檢定，利用這些方法去分析出有那些因素會影響到犯罪率。

## 第二章 資料分析

由於想看出變數之間的關係，在此做了初步統計、介紹變數、選擇變數、決定我們要的變數。

### 第一節 一般敘述統計、散佈圖

圖 1 (接下來報告中所提到的圖片或表格，請參考附錄) 包含了各參數的平均數，標準差，總和及最大最小值，其中  $x_5$  所代表的綜合稅因為有些州的稅率是 0，所以最小值為 0，這些州分別是 Delaware、Montana、New Hampshire、Oregon。由圖 2 可得知失業率越高時，犯罪率相對越高，在失業率最低時有犯罪率最小值，在失業率第二高時，有犯罪率最大值，散佈圖呈現正相關之關係；由圖 3 可得知人口數越高時，犯罪率越高，在人口數第一高的州，有最高的犯罪率，散佈圖呈現正相關之關係；由圖 4 可得知 GDP 越高時，犯罪率越低，在 GDP 最高時，有最低犯罪率，散佈圖呈現負相關之關係；由圖 5 可得知教育花費越高時，犯罪率相對越高，教育花費是一個支出是一個負擔，所以這誘使人為了完成自己的學業或別人的學業而犯罪，散佈圖呈現正相關之關係；由圖 6 可得知綜合稅越高時，犯罪率相對越高，可以推測稅越高人民負擔重，更容易產生犯罪行為，散佈圖呈現正相關之關係；由圖 7 可得知社會福利越高時，犯罪率相對越低，散佈圖呈現負相關之關係；由圖 8 可得知醫療花費越高時，犯罪率卻相對越高，這跟教育花費一樣，是一個支出一個負擔，所以為了應付這支出因而犯罪，散佈圖呈現正相關之關係；由圖 9 可得知 CPI 越高時，犯罪率相對越低，人民較富裕時犯罪率較低，散佈圖呈現負相關之關係；由圖 10 可得知年均花費越高時，犯罪率相對越低，一州人民平均花費越高，表示人民越富裕，犯罪率相對越低，散佈圖呈現負相關之關係；由圖 11 可得知警力的多寡，好像和犯罪率並沒有直接關係。

## 第二節 全部變數

在此介紹我們感興趣的變數(犯罪率)，還有可能會影響犯罪率的 10 個變數。

### (一) 反應變數

變數  $Y$ ，在此指犯罪率，(亦稱刑案發生率)，指每州內每十萬人口刑事案件發生件數，其公式為犯罪率等於當期刑案發生數除以當期期中人口數乘以十萬，以萬為單位。這邊以暴力犯罪以及財產犯罪這兩種大類為主。

### (二) 解釋變數

1. 變數  $x_1$ ，在此指失業率，失業意指未獲得任何有薪工作的狀態，而失業率是每州勞動人口裡符合「失業條件」者所佔的比例。
2. 變數  $x_2$ ，在此指人口數，指各州的總人口數目。
3. 變數  $x_3$ ，在此指人均 GDP，GDP 即 gross domestic product 的縮寫，是指經濟社會(即一個州)在一定時期內(多指一年)運用生產要素所生產的全部最終產品(產品和服務)的市場價值，也就是每州生產總值，人均 GDP 是指各州每個人一年的產出，以萬為單位。
4. 變數  $x_4$ ，在此指人均教育花費，該州每人用在教育上的花費，費用以百為單位。
5. 變數  $x_5$ ，在此指綜合稅，是各州的州稅再加上地方稅而成的。
6. 變數  $x_6$ ，在此指人均社會福利，州政府運用在為所有公民普遍提供旨在保證一定生活水平和儘可能提高生活質量的資金和服務的社會保險制度，以千為單位。
7. 變數  $x_7$ ，在此指人均醫療花費，該州每人用在醫療上的花費，費用以百為單位。
8. 變數  $x_8$ ，在此指 CPI，消費者物價指數，Consumer Price Index 的縮寫，亦稱居民消費價格指數，是反映與居民生活有關的產品及勞務價格統計出來的物價變動指標，以百分比變化為表達形式。
9. 變數  $x_9$ ，在此指人均年花費，各州內的每人每年平均花費，費用以千為單位。
10. 變數  $x_{10}$ ，在此指警力比例，是各州的是警察人數和人口的比例。

### 第三節 選取重要變數

在此用了 5 種方法來選取重要變數，分別方法有向前選取法、向後消去法、逐步迴歸法、調整  $R^2$  法、CP 選取法，我們想要在調整  $R^2$  如果沒有差很多的話，想要越多變數越好，我們在最後選出模型為全模型。

#### (一) 向前選取法

向前選取法，開始只放一個變數，選取 p-value 最小的，而且要符合 p-value 值小於 0.15 下的變數，將之選進模型，就算加了別個變數會造成原本 p-value 值改變，也不會將任何變數剔除，可以從圖 12 中看到，最後選進來的變數有  $x_2$  (人口數)、 $x_6$  (人均社會福利)、 $x_9$  (人均年花費)。

#### (二) 向後消去法

向後選取法，一開始加入全部變數，在符合 p-value 值大於 0.15 下的變數，將之剔除，可以從圖 13 中看到，剔除了  $x_1$ 、 $x_3$ 、 $x_5$ 、 $x_9$ 、 $x_{10}$ ，最後選進來的變數有  $x_2$  (人口數)、 $x_4$  (人均教育花費)、 $x_6$  (人均社會福利)、 $x_7$  (人均醫療花費)、 $x_8$  (CPI)。

#### (三) 逐步迴歸法

逐步迴歸法是一綜合「向前選取法」與「向後消去法」的一種方法，開始只放一個變數，選取 p-value 最小的而且符合 p-value 值小於 0.15 下的變數，將之選進模型，但如果加了別個變數會造成原本 p-value 值大於 0.15，則會把 p-value 值大於 0.15 的變數給剔除，可以從圖 14 中看到，最後選進來的變數有  $x_2$  (人口數)、 $x_6$  (人均社會福利)、 $x_9$  (人均年花費)。

#### (四) 調整 $R^2$ 法

調整  $R^2$  法，是對所以可能組合的模型，計算其調整  $R^2$ ，此方法的好處是調整  $R^2$  值不會受到變數個數增加而增加，由於調整  $R^2$  值越高越能解釋  $y_i$  的變異，因為調整  $R^2$  值沒有差很多，所以從圖 15 中，我們挑選的模型為全模型。

#### (五) Mallows' Cp 選取法

Mallows' Cp 是對所以可能組合的模型，計算其 Cp 值，其選擇標準為 Cp 值越小越好，且 Cp 值要接近參數個數越好，從圖 16 中看到，在參數個數為 5 下，最好的模型變數有  $x_2$  (人口數)、 $x_6$  (人均社會福利)、 $x_8$  (CPI)、 $x_9$  (人均年花費)。



#### 第四節 建立模式

一開始，我們是用向前選取法的結果去當作我們的模型，但後來發現殘差的一致性檢定會被拒絕，儘管做了變數的轉換，還是不行；之後再嘗試了向後消去法以及其他方法提供的最佳選模結果，但還是存在殘差的一致性檢定會被拒絕的問題。最後，因為我們覺得全模型的調整  $R^2$  很高，而且和其他最佳模型的調整  $R^2$  相差不到百分之一，所以決定用全模型作為我們的選模結果。

$$\text{Model: } \hat{y}_i = 216999 + 3311.8x_1 + 0.033x_2 - 3639x_3 + 2445x_4 + 2314.3x_5 - 42202x_6 \\ + 102283x_7 - 1298.3x_8 + 762.27x_9 + 2.4 \times 10^{-6}x_{10}$$

從模型中可以看出， $x_3$ 、 $x_6$  以及  $x_8$  都和犯罪率成負相關，也就是說，當其中一個變數的數值增加時，其他不變，犯罪率會下降；其餘七個變數都和犯罪率成正相關，也就是說，當其中一個變數的數值增加時，其他不變，犯罪率會上升。

##### (一)ANOVA

由圖 17 可以得知 F 值為 125.91，p-value < 0.01，而我們用的  $\alpha$  設定為 0.05，所以拒絕虛無假設，因此至少有一個 x 會與 y 有關係。

##### (二)VIF(variance inflation faction)

變異數膨脹系數，此系數是用來檢定有沒有共線性的問題，當 VIF 的值愈小愈好，代表愈沒有共線性問題，當 VIF 大於 10 時，我們可以得知他有嚴重的共線性問題。由圖 17 可以看出每個變數的 VIF 值均沒有大於 10，代表沒有嚴重的共線性，所以變數之間不會互相影響。

##### (三)R-square

由圖 17 知道 R-square 為 0.9692，代表這迴歸模型可以解釋百分之 96.92 的  $y_i$  的變異量。而調整 R 平方(Adjusted R-square)是將誤差平方和和總平方和除以個別的自由度，公式如下：

$$\text{adj } R^2 = 1 - \frac{SSE/df(SSE)}{SST/df(SST)}$$

由圖 17 知道調整 R 平方為 0.9615，不會因為變數增加而數值隨著增加，所以不會影響解釋能力，所以調整 R 平方比 R-square 更能精準的判斷。另外我們在參數估計表中，可以觀察出  $x_3$  (人均 GDP)、 $x_6$  (人均社會福利)、 $x_8$  (CPI) 呈現的是負相關，其他變數是呈現正相關的。

美國真的那麼危險嗎？

### 第三章 診斷分析

在此想知道是否有影響點或異常點，如果有，是否將點去除，去除點前後是否滿足殘差的四大假設，在此雖然有異常點也有影響點，但因為四大假設都沒有拒絕虛無假設，而且有嘗試過把同是異常點也是影響點的第 33 筆資料刪去，但發現刪除後，會拒絕殘差一致性檢定，所以不對異常點和影響點做任何處理。

#### 第一節 殘差分析

在此使用 Student 跟 R Student 來分析、判斷我們的資料是否有異常點。異常點是指數據中有一個或多個數值與其他數值相比差異較大，Student 跟 R Student 分析中，檢查我們的資料是否有超過三倍標準差的點，稱為異常點，Student 跟 R Student 的差別為 R Student 有先抽起一筆觀察值，之後再做迴歸分析，然後再把那一點放回模型，檢測是否有超過三倍標準差，所以 R Student 會比較敏感，看圖 18，Student 有一點異常點 New York，看圖 19，R Student 有兩點異常點，但如之前所說，我們並沒有刪除這些點，我們選擇保留。

#### 第二節 影響點判斷

在此影響點的判斷，圖 20，我們用了 5 種方法去判斷，分別有：Cook's Distance、COVRATIO、Hat value、DFBETAS、DFBETAS，雖然有許多影響點，以第 9 筆重複最多，但影響比較大並不代表不好，所以我們沒有將之剷掉(下列 p 為參數共 11 個，n 為資料個數共 51 筆)。

##### (一)Cook's Distance 法

在 Cook's Distance 方法中，準則為  $D_i > F_{0.5}(11, 40)$  時為影響點，即當  $D_i > 0.6614$  時有影響點，得到第 9 筆 District of Columbia 為 6.081 是影響點。

##### (二)COVRATIO 法

表 1，在 Covratio 方法中，準則為當 Covratio 值符合  $COV_i > 1+3*p/n$  或  $COV_i < 1-3*p/n$  時為影響點，即當  $COV_i > 1.6470$  或  $COV_i < 0.3529$  時有影響點，所以第 2 Alaska、5 California、9 District of Columbia、12 Hawaii、22 Massachusetts、27 Montana、30 New Hampshire、33 New York、35 North Dakota、38 Oregon、40 Rhode Island、44 Texas、46 Vermont，是影響點，其中以第 9 筆 District of Columbia 為 3.4884 影響最大。

##### (三)Hat value

表 2，在 hat value 方法中，準則為  $\text{Hat Value} > 2*p/n$ ，即當  $\text{Hat Value} > 0.4314$  是為影響點，所以第 5 California、9 District of Columbia、46 Vermont 比為影響點，其中以第 9 District of Columbia 筆為 0.9246 影響最大。

美國真的那麼危險嗎？

#### (四) DFFITS

表 3，在 DFFITS 中，準則為  $|DFFITS| > 2 * \sqrt{p/n}$ ，即當  $|DFFITS| > 0.9288$  為影響點，所以發現第 5 California、9 District of Columbia、10 Florida、33 New York、44 Texas、51 Wyoming 是影響點，其中以第 33 New York 筆為 -2.0712 影響最大。

#### (五) DFBETAS

表 4，在 DFBETAS 中，準則為  $|DFBETAS| > 2 / \sqrt{n}$ ，即當  $|DFBETAS| > 0.2801$  時表示可能有影響點，Intercept 發現第 8 Delaware、9 District of Columbia、21 Louisiana、31 New Jersey、33 New York、39 Oklahoma、44 Texas 為影響點，其中以第 9 District of Columbia 為 0.8190 影響最大； $x_1$  發現第 9 District of Columbia、33 New York、44 Texas、47 Virginia 為影響點，其中以第 9 District of Columbia 為 0.6691 影響最大； $x_2$  發現第 5 California、10 Florida、33 New York、44 Texas 為影響點，其中以第 44 Texas 為 1.4412 影響最大； $x_3$  發現第 9 District of Columbia、10 Florida、46 Vermont、48 Washington 為影響點，其中以第 9 District of Columbia 為 1.5768 影響最大； $x_4$  發現第 8 Delaware、9 District of Columbia、31 New Jersey、33 New York、39 Pennsylvania、44 Texas 為影響點，其中以第 9 District of Columbia 為 2.4982 影響最大； $x_5$  發現第 8 Delaware、48 Washington 為影響點，其中以第 8 Delaware 為 -0.4919 影響最大； $x_6$  發現第 8 Delaware、9 District of Columbia、20 Kentucky、33 New York、44 Texas、51 Wyoming 為影響點，其中以第 33 New York 為 -1.0952 影響最大； $x_7$  發現第 8 Delaware、9 District of Columbia、10 Florida、20 Kentucky、33 New York、44 Texas、51 Wyoming 為影響點，其中以第 33 New York 為 1.0436 影響最大； $x_8$  發現第 8 Delaware、9 District of Columbia、21 Louisiana、31 New Jersey、33 New York、39 Pennsylvania、44 Texas、47 Virginia、51 Wyoming 為影響點，其中以第 9 District of Columbia 為 -1.0704 影響最大； $x_9$  發現第 5 California、10 Florida、33 New York、46 Vermont、47 Virginia、51 Wyoming 為影響點，其中以第 51 Wyoming 為 -0.7571 影響最大； $x_{10}$  發現第 8 Delaware、9 District of Columbia、10 Florida、31 New Jersey、33 New York、44 Texas、48 Washington 為影響點，其中以第 33 New York 為 -0.6883 影響最大。

#### (六) 影響點的匯總和探討

方法	最大的影響點
Cook's Distance	第 9 筆 District of Columbia
COVRATIO	第 9 筆 District of Columbia
Hat value	第 9 筆 District of Columbia
DFFITS	第 33 筆 New York
DFBETAS	Intercept
	第 9 筆 District of Columbia

$x_1$	第 9 筆 District of Columbia
$x_2$	第 44 筆 Texas
$x_3$	第 9 筆 District of Columbia
$x_4$	第 9 筆 District of Columbia
$x_5$	第 33 筆 New York
$x_6$	第 33 筆 New York
$x_7$	第 33 筆 New York
$x_8$	第 9 筆 District of Columbia
$x_9$	第 51 筆 Wyoming
$x_{10}$	第 33 筆 New York
<p>District of Columbia，美國首都，簡稱 DC，犯罪率排名 40，失業率排名 9，人口數排名 49，人均 GDP 排名 1，人均教育花費排名 1，綜合稅並排 36，人均社會福利排名 1，人均醫療花費排名 3，CPI 排名 1，人均年花費排名 1，警力比例排名 1。根據迴歸模型顯示，人均 GDP、人均社會福利以及 CPI 會與犯罪率呈負相關，而其餘七個變數會與犯罪率呈正相關，在 DC 的數據中，就只有人均人口數、GDP、人均社會福利以及 CPI 這四個變數是符合模型的預期的，其餘六個變數都跟模型的預期不一樣，所以這是 DC 成為影響點的原因。</p>	
<p>New York，世界最大的城市之一，美國的經濟中心，犯罪率排名 4，失業率並排 13，人口數排名 3，人均 GDP 排名 6，人均教育花費排名 4，綜合稅並排 7，人均社會福利排名 8，人均醫療花費排名 14，CPI 排名 6，人均年花費排名 49，警力比例排名 5。在紐約的數據當中，人均 GDP、人均社會福利、CPI 以及人均年花費這四個變數都不符合我們的迴歸模型的預期的，所以才是紐約會成為影響點的原因。</p>	
<p>Texas，簡稱德州，全美土地面積及人口第二大州，犯罪率排名 2，失業率並排 37，人口數排名 2，人均 GDP 排名 13，人均教育花費排名 33，綜合稅並排 11，人均社會福利排名 41，人均醫療花費排名 26，CPI 排名 29，人均年花費排名 50，警力比例排名 30。在德州的數據中，人均 GDP、人均教育花費、人均醫療花費、人均年花費以及警力比例這五個變數都和我們的迴歸模型的預期不一樣的，所以這成為了德州是影響點的原因。</p>	
<p>Wyoming，懷俄明州，美國人口最少的州，犯罪率排名 51，失業率並排 45，人口數排名 51，人均 GDP 排名 4，人均教育花費排名 51，綜合稅並排 43，人均社會福利排名 47，人均醫療花費排名 43，CPI 排名 48，人均年花費排名 2，警力比例排名 6。在懷俄明州的數據中，人均社會福利、CPI、人均年花費以及警力比例這四個變數是和我們的迴歸模型的預期不一樣的，所以這成為了懷俄明州是影響點的原因。</p>	

### 第三節 殘差檢定

對於我們的殘差進行四大假設，分別是：殘差的期望值為 0、殘差的變異數一致性、殘差與殘差之間的共變異數為 0、殘差符合常態分配，檢查資料是否符合假設，在此資料都符合誤差假設。

(一) 殘差的期望值為 0， $E(\varepsilon_i)=0$

$$H_0 : E(\varepsilon_i) = 0$$

$$H_1 : E(\varepsilon_i) \neq 0$$

看圖 21，使用 Student's t，Sign，Signed Rank 三種方法檢定，如果 p-value 值大於  $\alpha=0.05$ ，不會拒絕虛無假設，而從圖 21 中，p-value 值都不拒絕  $H_0$ ，也就是  $E(\varepsilon_i)=0$ ，符合我們的假設。

(二) 殘差的變異數為  $\sigma^2$

$$H_0 : \text{Var}(\varepsilon_i) = \sigma^2$$

$$H_1 : \text{Var}(\varepsilon_i) \neq \sigma^2$$

看圖 22，主要使用 Breusch-Pagan 檢定，變異數是否一致，如果 p-value 值大於  $\alpha=0.05$ ，不會拒絕虛無假設，而從圖 22 中，p-value 值不拒絕  $H_0$ ，所以殘差的變異數有一致性。

(三) 殘差與殘差之間的共變異數為 0

正自我相關      負自我相關

$$H_0 : \rho = 0 \quad H_0 : \rho = 0$$

$$H_1 : \rho > 0 \quad H_1 : \rho < 0$$

看圖 23，主要使用 Durbin-Watson 檢測是否有自我相關，如果 p-value 值大於  $\alpha=0.05$ ，不會拒絕虛無假設，從圖 23 中看出，沒有正自我相關，也沒有負自我相關。

(四) 殘差符合常態

$$H_0 : \text{Residuals follow normally distributed.}$$

$$H_1 : \text{Residuals do not follow normally distributed.}$$

看圖 21、24，使用 Shapiro-Wilk，Kolmogorov-Smirnov，Cramer-von Mises，Anderson-Darling 四種方法檢定，如果 p-value 值大於  $\alpha=0.05$ ，不會拒絕虛無假設，而從圖 21 和 24 中，p-value 值都不拒絕  $H_0$ ，也就是殘差符合常態，符合我們的假設。

#### 第四章 結論

我們發現如果人均 GDP、人均社會福利、CPI 呈現負相關，也就是說當這三個變數的數值越高，該州的犯罪率會越低，這也很符合經濟學的理论。在經濟學的角度，GDP 越高，相對失業率越低，也就是說普遍人都有工作。有工作就少了時間去犯罪，也減低了犯罪的誘因之一，金錢的影響。人均社會福利也降低了人們因為金錢而犯罪的誘因。而且在犯罪率中，盜竊罪佔了超過一半的比例，說明金錢是犯罪的一個很大的誘因。至於 CPI 是受到了 GDP 的影響比較多，因為根據經濟學的理论，GDP 越高，CPI 也會越高。

而失業率、人口數、人均教育花費、綜合稅、人均醫療花費、人均年花費、警力比例這幾個是呈現正相關，換句話說失業率高、花費高，會使得犯罪率跟著升高。教育花費，綜合稅，人均醫療花費，人均年花費以及失業率，這些都是跟金錢有關係的，支出大，收入低（失業率越高，薪水越低，因為勞力資源供過於求），或完全沒收入（失業人士），這些都容易因為金錢問題而犯罪。

我們這份報告所得出的結論，只能應用在美國各州上，而無法應用在其他國家的城市，而且也只能應用在 2012 年，這一年，畢竟每年的數據不一樣，分析出來的結果也會不一樣。這就是我們這份報告的限制。

其實最根本的原因就是錢，除非你的生活超過金錢所能支付，或是你無法支付現在的生活，或是心理狀態不穩定，有人說錢是萬惡之源，但是沒有錢又萬萬不能，做人要懂得知足，如果不能滿足，那就需要用正當的方式去爭取，而不是造成社會的不穩定和人心惶惶，維持社會上的穩定，也是靠每個人民所去維持的，光靠一個人是很難改變的，互相扶持，犯罪率也會降低的喔。

美國真的那麼危險嗎？

## 附錄

### (一) 資料出處

1. [http://www.npa.gov.tw/NPAGip/wSite/lp?ctNode=12599&xq\\_xCat=02&mp=1](http://www.npa.gov.tw/NPAGip/wSite/lp?ctNode=12599&xq_xCat=02&mp=1)(中華民國內政部警政署全球資訊網)
2. [http://ap6.pccu.edu.tw/Encyclopedia\\_media/main-soc.asp?id=7135](http://ap6.pccu.edu.tw/Encyclopedia_media/main-soc.asp?id=7135) (中華百科全書)
3. <http://www.bea.gov/regional/downloadzip.cfm> (Bureau of Economic Analysis)
4. <http://www.cityrating.com/crime-statistics/alaska/> (CityRating.com)
5. <http://taxfoundation.org/article/state-and-local-sales-taxes-2012> (Tax Foundation)
6. <https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/resource-pages/about-cius> (The FBI Federal Bureau Of Investigation)
7. [https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/77tabledatadecpdf/table\\_77\\_full\\_time\\_law\\_enforcement\\_employess\\_by\\_state\\_2012.xls](https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/77tabledatadecpdf/table_77_full_time_law_enforcement_employess_by_state_2012.xls) (The FBI Federal Bureau Of Investigation)

### (三) 參考用書及資料

1. Abraham, B. and Ledolter, J. (2006) Introduction to regression modeling, Belmont, CA, Duxbury.
2. Krugman, P. and Wells, R. (2012) Economics, third edition, Worth Publishers.
3. <http://gate.sinovision.net:82/gate/big5/bbs.sinovision.net/forum.php?mod=viewthread&tid=127791> (美國中文網-前十大失業率和收入與犯罪率的關聯)
4. <http://vhouse.163.com/15/0108/16/AFEVNMH300294MCG.html> (海外房產資訊-教育水準影響收入而導致犯罪率高)
5. <http://www.epochtimes.com/b5/13/10/11/n3984138.htm%E7%94%B0%E7%B4%8D%E8%A5%BF%E5%B7%9E%E6%9A%B4%E5%8A%9B%E7%8A%AF%E7%BD%AA%E7%8E%87%E5%85%A8%E7%BE%8E%E6%9C%80%E9%AB%98.html> (大紀元-犯罪率和各州的經濟、教育、收入有關聯)

(二)表

第 i 點	Covratio 值
2	2.2457
5	2.1462
9	3.4884
12	1.675
22	1.7293
27	1.8309
30	2.1122
33	0.0255
35	1.9968
38	1.9609
40	1.9907
44	0.1661
46	2.2515

表 1

第 i 點	Hat value 值
5	0.5353
9	0.9246
46	0.4471

表 2

第 i 點	DFFITS
5	-1.0787
9	8.6905
10	0.9562
33	-2.0712
44	1.9835
51	-1.4217

表 3

	Intercept	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
5			-0.9204							-0.3662	
8	-0.3181				-0.4919	-0.4919	0.3592	-0.3506	0.3125		0.2956
9	0.8190	0.6691		1.5768	2.4982		-0.4276	-0.3123	-1.0704		0.4385
10			0.6302	-0.5286				0.3299		0.3016	0.4044
20							0.4166	-0.3395			
21	-0.2811								0.3569		
31	0.6768				0.4397				-0.4370		-0.4278
33	0.5478	0.3414	-0.8721		0.5889		-1.0952	1.0436	-0.5166	0.5668	-0.6883
39	-0.3518				-0.3802				0.34790		
44	0.5578	-0.5153	1.4412		0.3413		-0.6309	0.4581	-0.4009		-0.3690
46				0.3038						-0.3216	
47		0.4235							-0.3663	0.3110	
48				0.4420		0.3678					-0.4824
51							0.6030	-0.4685	0.3286	-0.7571	

表 4



美國真的那麼危險嗎？

(三) 圖

**The SAS System**  
**The CORR Procedure**

11 Variables: y x1 x2 x3 x4 x5 x6 x7 x8 x9 x10

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
y	51	199802	231590	10189902	14383	1210409
x1	51	7.37255	1.71185	376.00000	3.10000	11.20000
x2	51	6155445	6974481	313927717	576412	38041430
x3	51	5.24330	1.96292	267.40810	3.40520	17.05220
x4	51	6.07596	7.46762	309.87420	1.49200	52.04330
x5	51	6.33020	2.30003	322.84000	0	9.45000
x6	51	3.77770	0.97566	192.66290	2.44260	7.52010
x7	51	1.72640	0.27898	88.04660	1.22390	2.38330
x8	51	227.84314	17.46239	11620	202.00000	283.00000
x9	51	20.95248	24.93076	1069	1.50130	128.56660
x10	51	0.00303	0.0008969	0.15450	0.00170	0.00770

圖 1

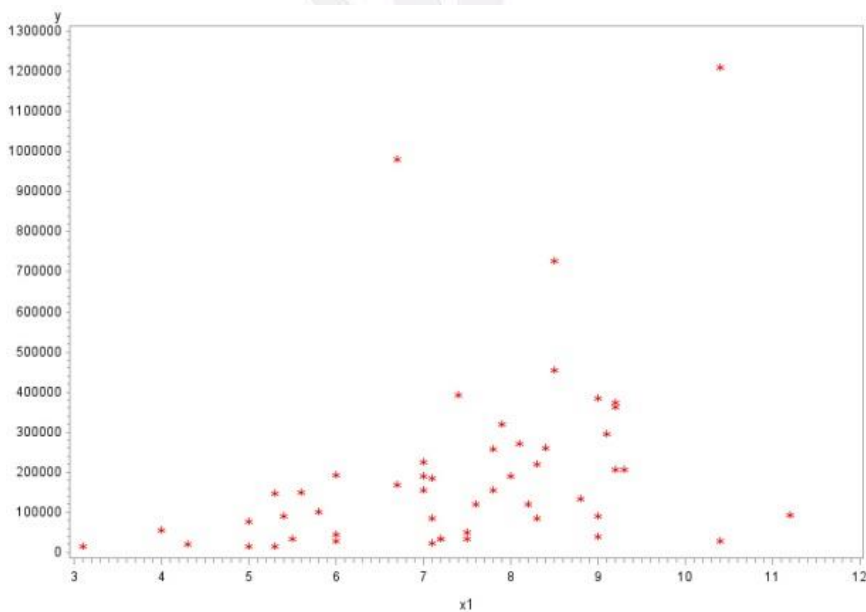


圖 2

美國真的那麼危險嗎？

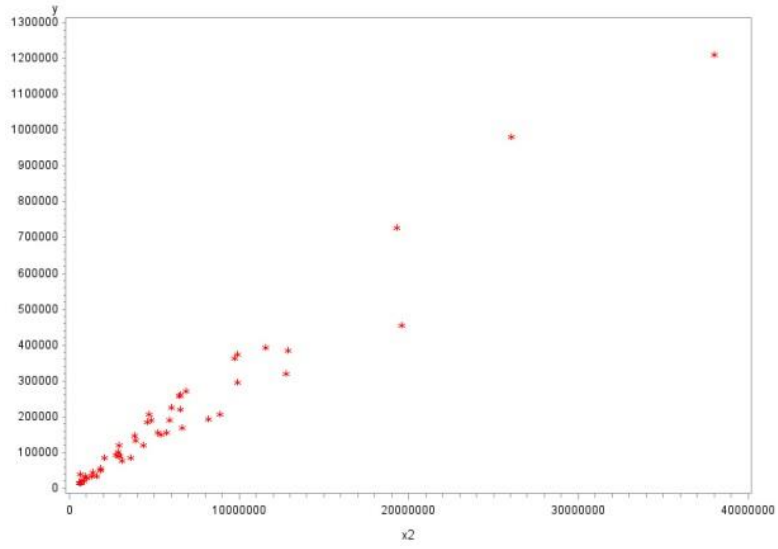


圖 3

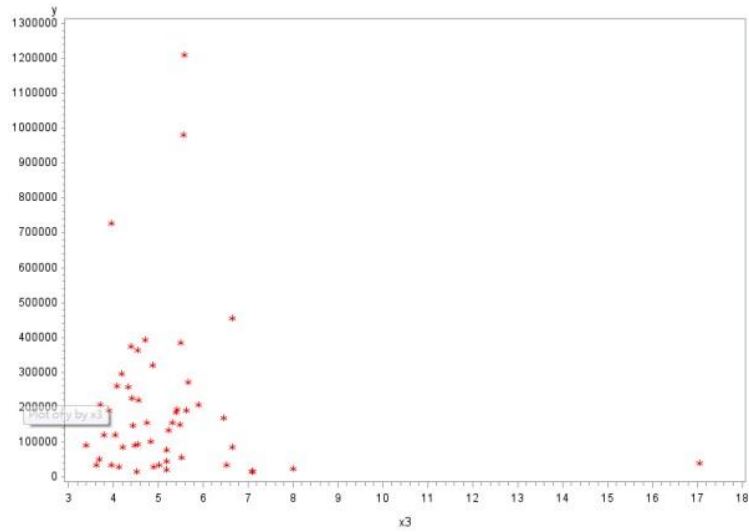


圖 4

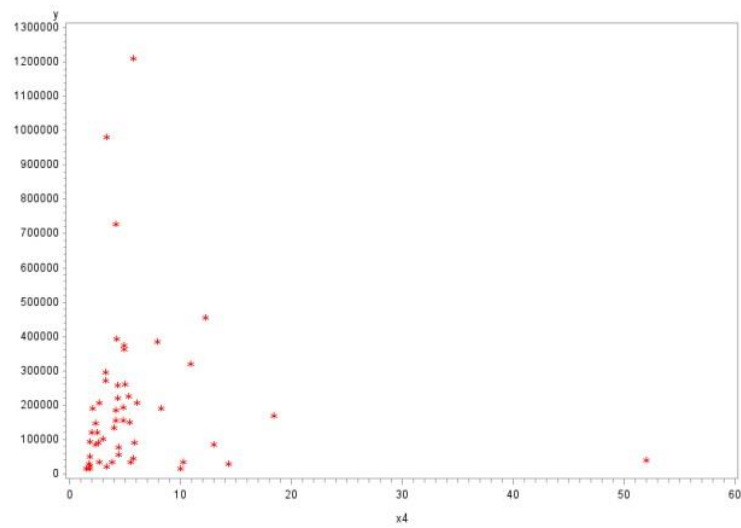


圖 5

美國真的那麼危險嗎？

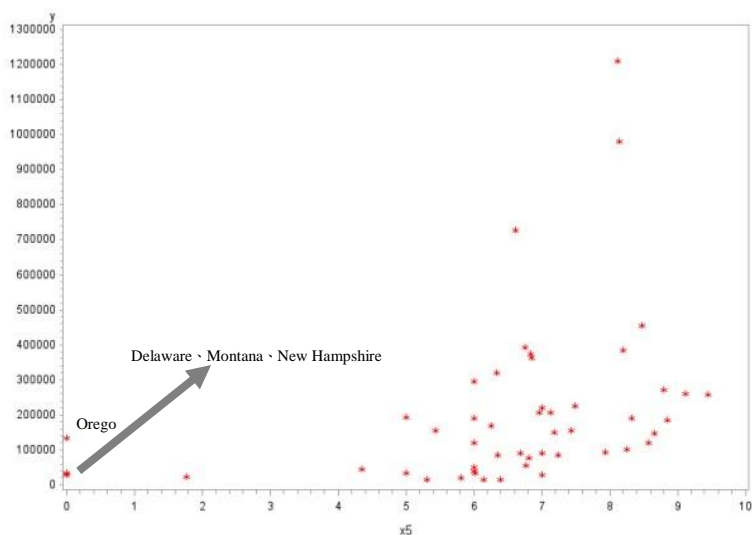


圖 6

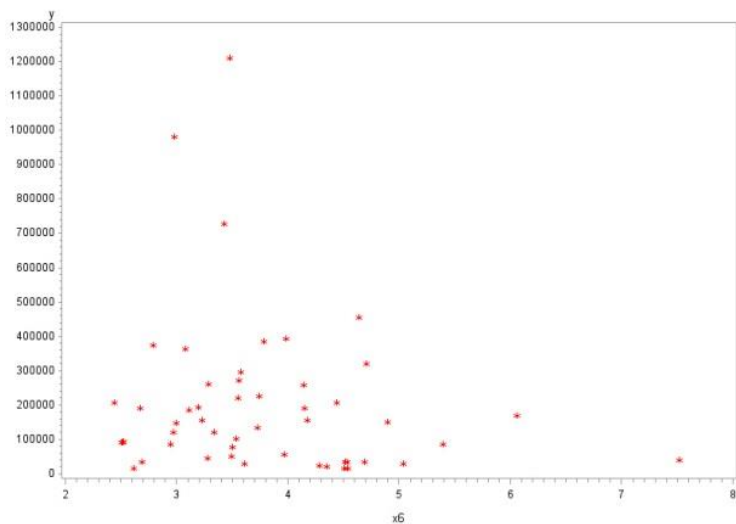


圖 7

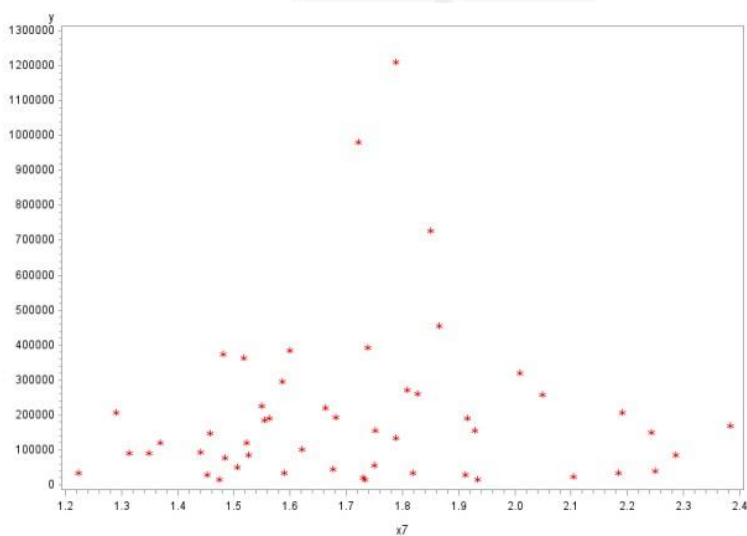


圖 8

美國真的那麼危險嗎？

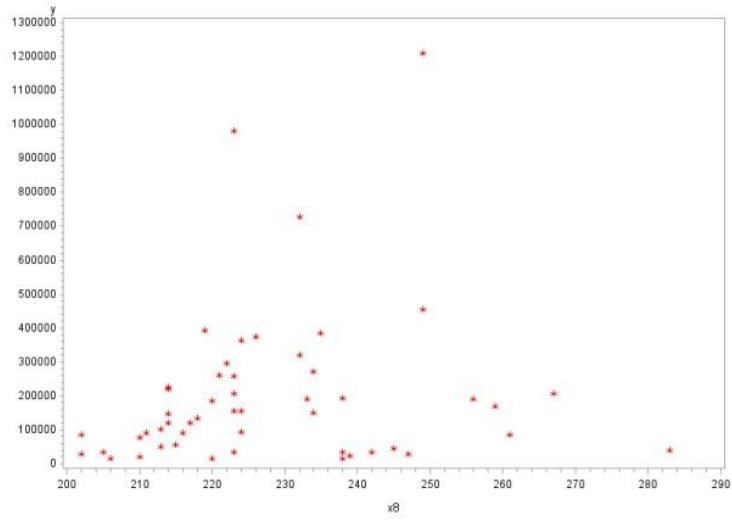


圖 9

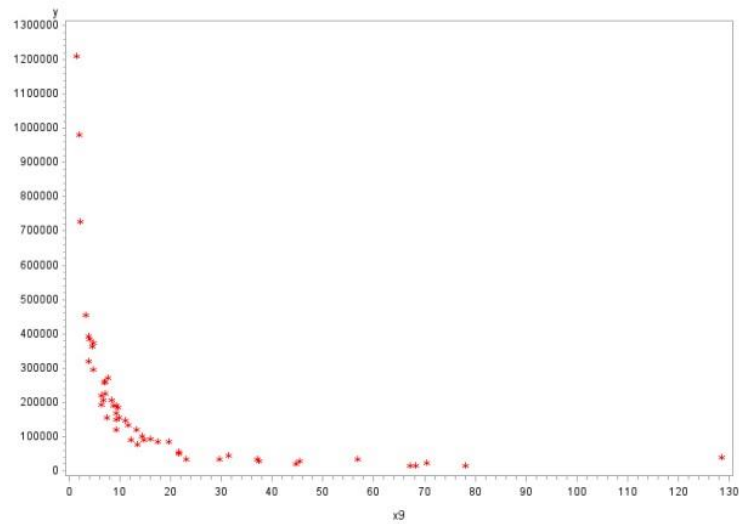


圖 10

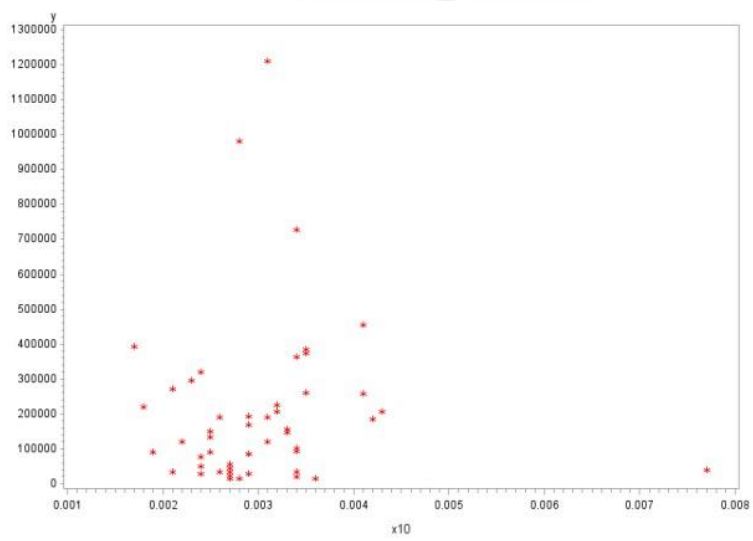


圖 11

No other variable met the 0.1500 significance level for entry into the model.

Summary of Forward Selection							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x2	1	0.9586	0.9586	6.8072	1134.07	<.0001
2	x6	2	0.0048	0.9633	2.6196	6.24	0.0160
3	x9	3	0.0016	0.9650	2.5127	2.18	0.1469

圖 12

All variables left in the model are significant at the 0.1500 level.

Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x10	9	0.0000	0.9692	9.0418	0.04	0.8390
2	x3	8	0.0001	0.9691	7.1856	0.15	0.7032
3	x5	7	0.0004	0.9686	5.7284	0.57	0.4555
4	x1	6	0.0004	0.9682	4.2523	0.55	0.4611
5	x9	5	0.0013	0.9670	3.9069	1.76	0.1909

圖 13

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x2		1	0.9586	0.9586	6.8072	1134.07	<.0001
2	x6		2	0.0048	0.9633	2.6196	6.24	0.0160
3	x9		3	0.0016	0.9650	2.5127	2.18	0.1469

圖 14

美國真的那麼危險嗎？

**The SAS System**

The REG Procedure  
Model: MODEL4  
Dependent Variable: y

Adjusted R-Square Selection Method

Number of Observations Read	51
Number of Observations Used	51

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
6	0.9639	0.9682	4.2523	x2 x4 x6 x7 x8 x9
7	0.9635	0.9686	5.7284	x1 x2 x4 x6 x7 x8 x9
7	0.9635	0.9686	5.7838	x2 x4 x5 x6 x7 x8 x9
5	0.9633	0.9670	3.9069	x2 x4 x6 x7 x8
7	0.9632	0.9684	6.0671	x2 x3 x4 x6 x7 x8 x9
8	0.9632	0.9691	7.1856	x1 x2 x4 x5 x6 x7 x8 x9
7	0.9632	0.9683	6.1519	x2 x4 x6 x7 x8 x9 x10
8	0.9628	0.9688	7.5426	x2 x3 x4 x5 x6 x7 x8 x9
7	0.9628	0.9680	6.5107	x1 x2 x5 x6 x7 x8 x9
6	0.9628	0.9673	5.4994	x2 x4 x6 x7 x8 x10
8	0.9628	0.9687	7.6248	x1 x2 x3 x4 x6 x7 x8 x9
8	0.9627	0.9687	7.6519	x1 x2 x4 x6 x7 x8 x9 x10
6	0.9627	0.9672	5.5899	x2 x5 x6 x7 x8 x9
3	0.9627	0.9650	2.5127	x2 x6 x9
6	0.9627	0.9672	5.6398	x2 x6 x7 x8 x9 x10
6	0.9627	0.9672	5.6422	x2 x3 x4 x6 x7 x8
5	0.9616	0.9654	5.9405	x2 x4 x6 x9 x10
6	0.9615	0.9662	6.9630	x1 x2 x6 x7 x8 x10
5	0.9615	0.9654	5.9680	x2 x3 x6 x9 x10
5	0.9615	0.9654	5.9746	x2 x3 x6 x8 x9
4	0.9615	0.9646	4.9749	x2 x4 x6 x10
6	0.9615	0.9661	6.9781	x2 x5 x6 x8 x9 x10
5	0.9615	0.9654	5.9794	x2 x3 x4 x6 x9
10	0.9615	0.9692	11.0000	x1 x2 x3 x4 x5 x6 x7 x8 x9 x10
6	0.9615	0.9661	7.0158	x2 x4 x6 x7 x9 x10
6	0.9615	0.9661	7.0185	x2 x4 x6 x8 x9 x10
5	0.9615	0.9653	6.0374	x1 x2 x6 x9 x10
4	0.9615	0.9646	5.0384	x2 x3 x6 x10
4	0.9615	0.9646	5.0460	x2 x3 x6 x7
6	0.9615	0.9661	7.0562	x1 x2 x4 x6 x7 x9
5	0.9615	0.9653	6.0604	x1 x2 x4 x6 x9
5	0.9615	0.9653	6.0612	x2 x3 x4 x6 x8
5	0.9615	0.9653	6.0675	x2 x3 x6 x8 x10
6	0.9614	0.9661	7.0738	x1 x2 x6 x8 x9 x10

圖 15

美國真的那麼危險嗎？

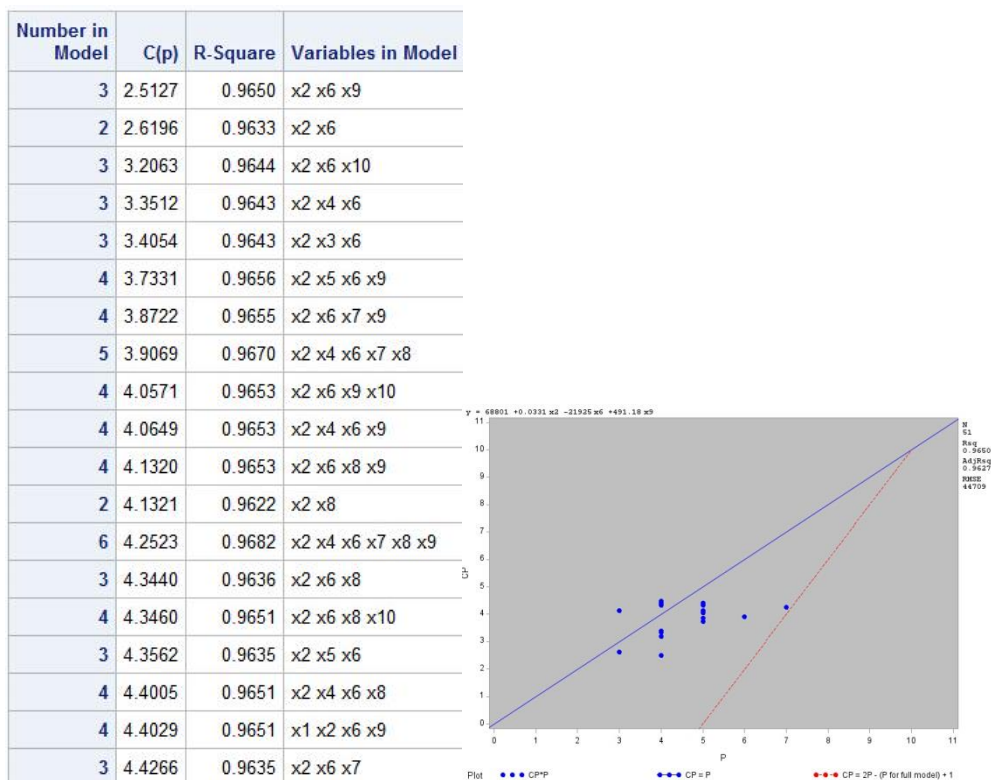


圖 16

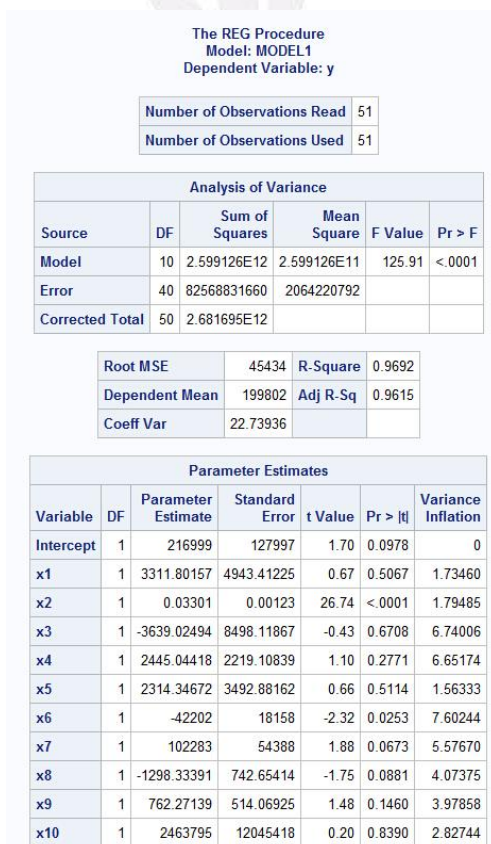


圖 17

美國真的那麼危險嗎？

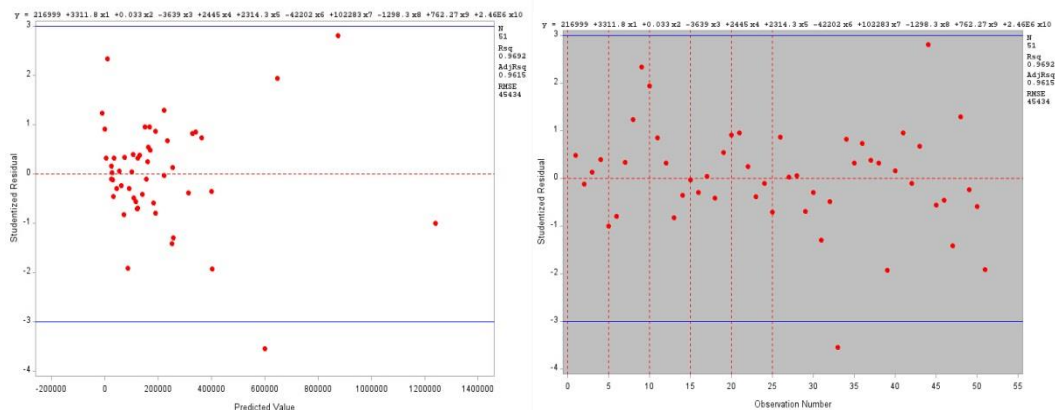


圖 18

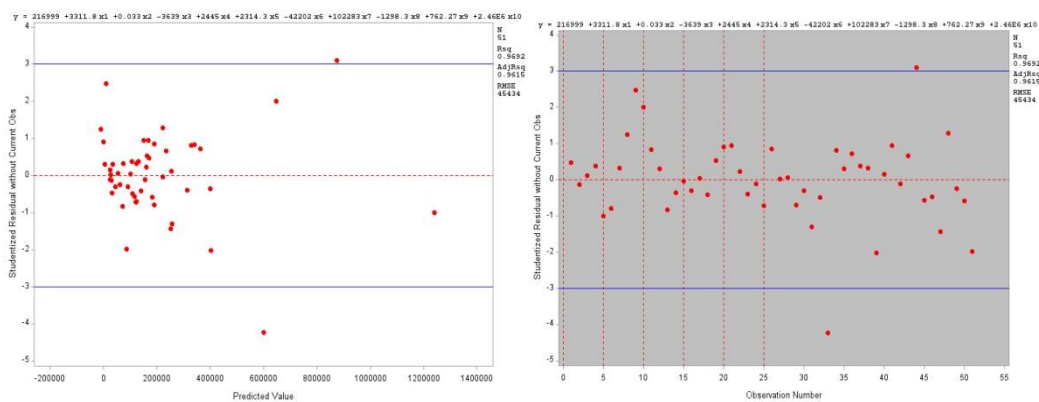


圖 19







美國真的那麼危險嗎？

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
y	11	40	8.257E10	2.0642E9	45433.7	0.9692	0.9615

Nonlinear OLS Parameter Estimates				
Parameter	Estimate	Approx Std Err	t Value	Approx Pr >  t
a1	216998.6	127997	1.70	0.0978
b1	3311.802	4943.4	0.67	0.5067
b2	0.033009	0.00123	26.74	<.0001
b3	-3639.02	8498.1	-0.43	0.6708
b4	2445.044	2219.1	1.10	0.2771
b5	2314.347	3492.9	0.66	0.5114
b6	-42201.6	18158.1	-2.32	0.0253
b7	102282.9	54388.0	1.88	0.0673
b8	-1298.33	742.7	-1.75	0.0881
b9	762.2714	514.1	1.48	0.1460
b10	2463795	12045418	0.20	0.8390

Number of Observations		Statistics for System	
Used	51	Objective	1.619E9
Missing	0	Objective*N	8.2569E10

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
y	White's Test	51.00	50	0.4341	Cross of all vars
	Breusch-Pagan	16.01	10	0.0994	1, x1, x2, x3, x4, x5, x6, x7, x8, x9, x10

圖 22

**Durbin-Watson test**  
The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	8.25688E10	DFE	40
MSE	2064220792	Root MSE	45434
SBC	1269.44051	AIC	1248.19043
MAE	29314.7381	AICC	1254.95966
MAPE	34.2548617	HQC	1256.31071
		Regress R-Square	0.9692
		Total R-Square	0.9692

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	216999	127997	1.70	0.0978
x1	1	3312	4943	0.67	0.5067
x2	1	0.0330	0.001234	26.74	<.0001
x3	1	-3639	8498	-0.43	0.6708
x4	1	2445	2219	1.10	0.2771
x5	1	2314	3493	0.66	0.5114
x6	1	-42202	18158	-2.32	0.0253
x7	1	102283	54388	1.88	0.0673
x8	1	-1298	742.6541	-1.75	0.0881
x9	1	762.2714	514.0693	1.48	0.1460
x10	1	2463795	12045418	0.20	0.8390

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.7261	0.1793	0.8207
2	1.8110	0.2962	0.7038
3	2.3091	0.9251	0.0749
4	1.8702	0.5160	0.4840

圖 23

美國真的那麼危險嗎？

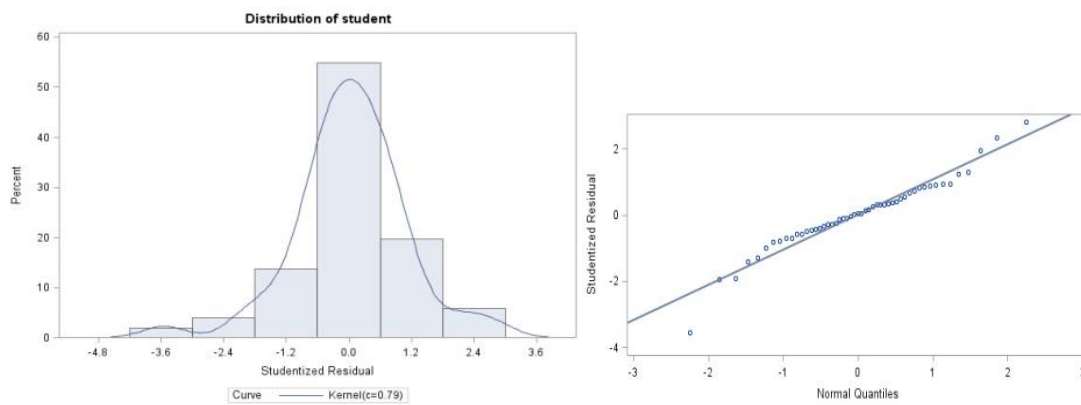


圖 24



美國真的那麼危險嗎？

2012年	犯罪率	失業率	人口數	人均GDP	人均教育花費	綜合稅	人均社會福利	人均醫療花費	CPI	人均年花費	警力比例
1	190571	8	4822023	3.8891	2.0655	8.33	2.6767	1.5630	233	8.7673	0.0026
2	24449	7.1	731449	7.9958	1.8457	1.77	4.2805	2.1040	239	70.4984	0.0027
3	260038	8.4	6553255	4.0772	4.9929	9.12	3.2857	1.8273	221	7.0074	0.0035
4	121776	7.6	2949131	3.7960	1.9836	8.58	2.9714	1.3682	214	13.2042	0.0031
5	1210409	10.4	38041430	5.5838	5.7540	8.11	3.4737	1.7887	249	1.5013	0.0031
6	155293	7.8	5187582	5.3212	4.1908	7.44	3.2258	1.7521	224	9.8067	0.0033
7	86994	8.3	3590347	6.6550	12.9987	6.35	5.3903	2.2867	261	17.3680	0.0029
8	35659	7.2	917092	6.5193	3.8491	0	4.5164	1.8177	242	56.7446	0.0034
9	38600	9	646000	17.0522	52.0433	6	7.5201	2.2492	283	128.5666	0.0077
10	727075	8.5	19317568	3.9672	4.1853	6.62	3.4252	1.8496	232	2.2595	0.0034
11	375920	9.2	9919945	4.4019	4.9426	6.84	2.7885	1.4804	226	4.7137	0.0035
12	46147	6	1392313	5.1933	5.7315	4.35	3.2780	1.6756	245	31.4908	0.0027
13	34969	7.2	1595728	3.6366	2.6446	6.02	2.6884	1.2239	205	23.0841	0.0027
14	385416	9	12875255	5.5087	7.8857	8.2	3.7824	1.5999	235	4.0846	0.0035
15	220634	8.3	6537334	4.5706	4.3672	7	3.5530	1.6623	214	6.3726	0.0018
16	77951	5	3074186	5.1816	4.4532	6.81	3.5027	1.4846	210	13.3717	0.0024
17	100942	5.8	2885905	4.8397	2.9939	8.26	3.5393	1.6206	213	14.3969	0.0034
18	121578	8.2	4380415	4.0485	2.4929	6	3.3358	1.5234	217	9.3126	0.0022
19	185804	7.1	4601893	5.4076	4.2070	8.85	3.1157	1.5546	220	9.5630	0.0042
20	34992	7.5	1329192	3.9614	5.5222	5	4.6871	1.5897	223	29.5510	0.0021
21	190086	7	5884563	5.6180	8.2300	6	4.1527	1.9148	256	9.1854	0.0031
22	170042	6.7	6646144	6.4675	18.4227	6.25	6.0595	2.3833	259	9.2971	0.0029
23	295023	9.1	9883360	4.1894	3.2519	6	3.5795	1.5854	222	4.7687	0.0023
24	150571	5.6	5379139	5.4791	5.4637	7.18	4.8993	2.2433	234	9.3167	0.0025
25	91692	9	2984926	3.4052	2.6198	7	2.5069	1.3481	216	12.2130	0.0019
26	226745	7	6021988	4.4089	5.3703	7.49	3.7453	1.5495	214	7.1515	0.0032
27	28706	6	1005141	4.1287	1.7609	0	3.6134	1.4525	202	37.3828	0.0024
28	55932	4	1855525	5.5303	4.4138	6.77	3.9633	1.7499	215	21.5384	0.0027
29	94273	11.2	2758931	4.5502	1.8558	7.93	2.5195	1.4411	224	15.9913	0.0034
30	33174	5.5	1320718	5.0296	10.2974	0	4.5278	2.1837	238	37.0730	0.0026
31	207208	9.3	8864590	5.8914	6.1221	6.97	4.4404	2.1914	267	6.7084	0.0043
32	86754	7.1	2085538	4.2074	2.3064	7.24	2.9455	1.5253	202	19.5676	0.0029
33	455750	8.5	19570261	6.6400	12.2359	8.48	4.6410	1.8645	249	3.2237	0.0041
34	363058	9.2	9752073	4.5495	4.9097	6.85	3.0785	1.5176	224	4.4908	0.0034
35	15775	3.1	699628	7.1136	1.8295	6.39	4.5081	1.7338	220	68.2920	0.0028
36	394478	7.4	11544225	4.7200	4.2861	6.75	3.9854	1.7379	219	3.8696	0.0017
37	147645	5.3	3814820	4.4468	2.3802	8.66	2.9980	1.4572	214	11.1295	0.0033
38	135376	8.8	3899353	5.2274	4.0340	0	3.7237	1.7888	218	11.5452	0.0025
39	320999	7.9	12763536	4.8685	10.8959	6.34	4.7076	2.0091	232	3.8451	0.0024
40	29668	10.4	1050292	4.8942	14.3674	7	5.0357	1.9118	247	45.4464	0.0029
41	206947	9.2	4723723	3.7192	2.6801	7.13	2.4426	1.2905	223	8.4239	0.0032
42	19850	4.3	833354	5.1948	3.2999	5.81	4.3499	1.7292	210	44.6689	0.0034
43	259214	7.8	6456243	4.3314	4.3570	9.45	4.1394	2.0484	223	6.8292	0.0041
44	982535	6.7	26059203	5.5624	3.3105	8.14	2.9826	1.7207	223	1.9648	0.0028
45	91300	5.4	2855287	4.4742	5.8278	6.68	2.5111	1.3137	211	14.6367	0.0025
46	15909	5	626011	4.5260	9.9998	6.14	4.5383	1.9345	238	67.1602	0.0027
47	192549	6	8185867	5.4287	4.8449	5	3.1945	1.6812	238	6.3424	0.0029
48	272719	8.1	6897012	5.6649	3.2405	8.8	3.5595	1.8076	234	7.6917	0.0021
49	49747	7.5	1855413	3.6973	1.8109	6	3.4914	1.5064	213	21.6669	0.0024
50	156577	7	5726398	4.7601	4.8076	5.43	4.1731	1.9293	223	7.4701	0.0033
51	14383	5.3	576412	7.0767	1.4920	5.31	2.6127	1.4746	206	78.0206	0.0036

圖 總數據