

目錄

第一章、緒論	5
第一節、研究背景與動機	5
第二節、研究目的	6
第三節、研究變數	6
第二章、簡單線性迴歸模型統計分析	27
第一節、總生育率和女性勞動力參與率之簡單迴歸模型	27
第一項、原始模型	27
(一)、簡單線性迴歸模型	27
(二)、X 對 Y 的散佈圖	28
(三)、殘差對 X 的散佈圖	29
(四)、X 的時間序列圖	29
(五)、殘差的時間序列圖	30
(六)、殘差的常態機率圖及相關性檢定	30
(七)、X 與 Y 之相關性檢定	31
(八)、BF test 變異數檢定	32
(九)、BP test 變異數檢定	34
(十)、Shapiro-Wilk test 常態性檢定	34
(十一)、F test of Lack of fit 缺適度檢定	35
(十二)、Lowess Curve	36
(十三)、預測線及信賴區間	37
(十四)、BOX-COX 轉換	37
第二項、變數轉換一(log 轉換)模型	38
(一)、簡單線性迴歸模型	38
(二)、X 對 Y 的散佈圖	39
(三)、殘差對 X 的散佈圖	39
(四)、Shapiro-Wilk test 常態性檢定	40

(五)、BF test 變異數檢定.....	41
(六)、Lowess Curve	43
(七)、預測線及信賴區間	43
第三項、變數轉換二(倒數轉換)模型	43
(一)、簡單線性迴歸模型	43
(二)、X 對 Y 的散佈圖.....	44
(三)、殘差對 X 的散佈圖	44
(四)、Shapiro-Wilk test 常態性檢定.....	45
(五)、BF test 變異數檢定.....	46
(六)、Lowess Curve	48
(七)、預測線及信賴區間	48
第四項、變數轉換二(倒數轉換)後刪除離群值模型	49
(一)、簡單線性迴歸模型	49
(二)、X 對 Y 的散佈圖.....	49
(三)、殘差對 X 的散佈圖	50
(四)、Shapiro-Wilk test 常態性檢定.....	50
(五)、BF test 變異數檢定.....	51
(六)、Lowess Curve	53
(七)、預測線及信賴區間	53
第二節、總生育率和扶幼比之簡單迴歸模型	54
第一項、原始模型	54
(一)、簡單線性迴歸模型	54
(二)、X 對 Y 的散佈圖.....	55
(三)、殘差對 X 的散佈圖	56
(四)、X 的時間序列圖.....	56
(五)、殘差的時間序列圖	57
(六)、殘差的常態機率圖及相關性檢定	57

(七)、X 與 Y 之相關性檢定.....	58
(八)、BF test 變異數檢定.....	59
(九)、BP test 變異數檢定.....	61
(十)、Shapiro-Wilk test 常態性檢定.....	61
(十一)、F test of Lack of fit 缺適度檢定.....	62
(十二)、Lowess Curve	63
(十三)、預測線及信賴區間	63
(十四)、BOX-COX 轉換.....	64
第二項、變數轉換(根號轉換)模型	65
(一)、簡單線性迴歸模型	65
(二)、X 對 Y 的散佈圖.....	65
(三)、殘差對 X 的散佈圖	66
(四)、Shapiro-Wilk test 常態性檢定.....	66
(五)、BF test 變異數檢定.....	67
(六)、Lowess Curve	69
(七)、預測線及信賴區間	69
第三項、變數轉換二(根號轉換)後刪除離群值模型	70
(一)、簡單線性迴歸模型	70
(二)、X 對 Y 的散佈圖.....	70
(三)、殘差對 X 的散佈圖	71
(四)、Shapiro-Wilk test 常態性檢定.....	71
(五)、BF test 變異數檢定.....	72
(六)、Lowess Curve	74
(七)、預測線及信賴區間	74
第三章、結論.....	75

第一章、緒論

第一節、研究背景與動機

近年來我國生育率逐年降低，根據行政院主計總處統計，2010 年總生育率 0.895 為歷年最低，近 5 年來因政府提供托兒津貼及廣設公立托兒機構等措施，總生育率才有微幅回升趨勢，至 2015 年上升為 1.175，亦即育齡婦女一生所哺育的兒女僅有 1.175 人，但我國仍屬全球低生育率國家。

生育率下降的主因，多數研究將其歸因於經濟發展過程中，女性勞動力參與率的逐年上升、女性教育程度的提高與醫療水準的改善，如 Becker(1960)、Nerlove(1986)、Murthi et al.(1995)以及 Goni(2008)等學者。。另外，蘇于婷(2012)指出我國女性薪資所得、結婚率對提高生育意願有正向推動力，而失業率、女性勞動力參與率、女性教育程度則有降低生育的趨勢，余清祥(2004)以空間迴歸分析發現，各鄉鎮的人口密度與教育程度的差異對生育率呈負向效果，而結婚率、扶老比和扶幼比為正相關。

婦女的就業可能會提高家庭所得而提升生育子女的意願，但也有可能減少其投入哺育照顧子女的時間，而不論已發開國家或開發中國家，大部分國家婦女勞參率上升皆伴隨著生育率下降，而我國也是如此，惟少數福利制度較完整的歐洲國家，如瑞典及挪威，其該國 1980 年前女性勞參率及生育率之間呈現負向相關，但 1981 年後卻呈現正向相關，係為 1981 年後瑞典由於有全面性的公共政策(如育嬰假、日間托育津貼、彈性工作時間與子女配額補助等)結合成一個系統，刺激生育率與婦女有酬就業，這些政策降低了擁有小孩的機會成本，並同時提供女性更容易二度就業的環境，促使女性出外就業以獲取完全利益，使得 25~54 歲女性勞參率越高，總生育率亦有越高趨勢。

總觀上述研究，本研究將接續探討女性勞動力參與率、男性勞動力參與

率、勞動力參與率、女性失業率、男性失業率、失業率、平均每年工時、扶養比、扶幼比、扶老比、平均每年薪資、名目 GDP 成長率、結婚率及離婚率是否也會影響各國女性總生育率，進行複迴歸分析。

第二節、研究目的

一、研究目的

收集我國及 OECD 各國(包含澳洲、奧地利、比利時、加拿大、智利、捷克、丹麥、愛沙尼亞、芬蘭、法國、德國、希臘、匈牙利、冰島、愛爾蘭、以色列、義大利、日本、韓國、拉脫維亞、盧森堡、墨西哥、荷蘭、紐西蘭、挪威、波蘭、葡萄牙、斯洛伐克、斯洛維尼亞、西班牙、瑞典、瑞士、土耳其、英國、美國、哥倫比亞、哥斯大黎加、立陶宛、巴西、中國大陸、印度、印尼、俄羅斯及我國等)從 1980 年到 2015 年總生育率及各項變數資料，假設每年每個國家資料間獨立，總共有 44 個國家 36 年資料，扣掉缺漏值，剩餘共有 1 千多筆資料利用線性迴歸模型進行實證研究分析，檢視影響各國女性生育意願及決策的各項因子。將反應變數 Y 設為 OECD 各國總生育率，獨立變數(解釋變數)X，包含女性勞動力參與率、男性勞動力參與率、勞動力參與率、女性失業率、男性失業率、失業率、平均每年工時、扶養比、扶幼比、扶老比、平均每年薪資、名目 GDP 成長率、結婚率及離婚率，找出那些因素對提升生育率有正向影響，那些因素有負向影響，並探討不同時間點及不同區域的國家影響效果有何不同，提供政府制定政策的參考。

第三節、研究變數

一、總生育率(反應變數 Y)

育齡婦女總生育率為假定一世代的 1,000 名育齡婦女，依目前年齡別生育率水準渡過其生育年齡期間，一生所可能有的出生數。公式為 15 至未滿 50 歲各育齡婦女年齡組別生育率的總和×5。各育齡婦女年齡組別生育率為

該 5 歲年齡組育齡婦女活產數除以該 5 歲年齡組育齡婦女期中人口數(共分為 15~19 歲、20~24 歲、25~29 歲、30~34 歲、35~39 歲、40~44 歲、45~49 歲共 7 組)。該 5 歲年齡組育齡婦女期中人口數為(當期期末 15 至未滿 50 歲育齡婦女人口數 + 上期期末 15 至未滿 50 歲育齡婦女人口數)÷2。

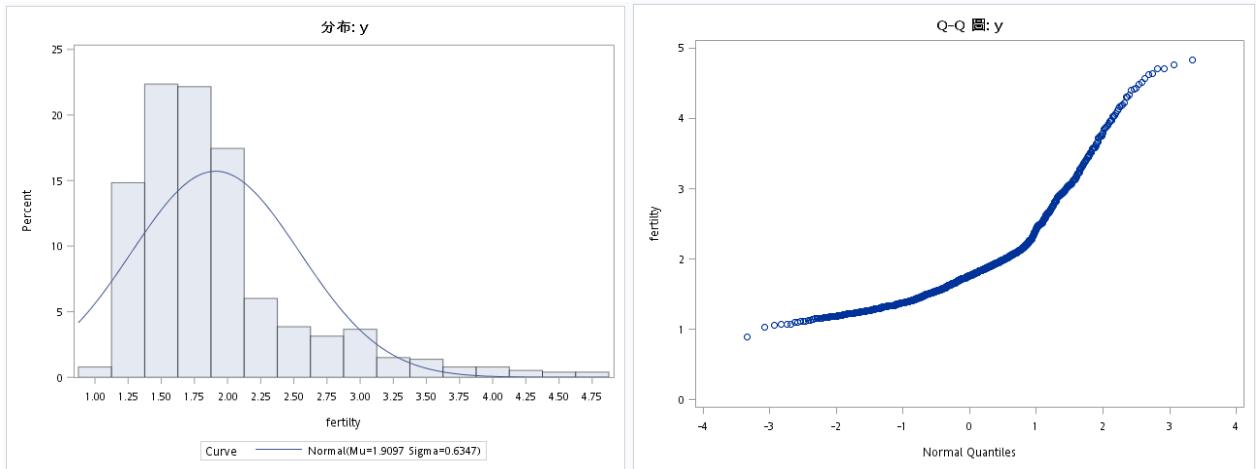
OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,530 筆育齡婦女總生育率資料，平均值為 1.9，標準差為 0.6347，偏態係數=1.73>0，為右偏(skewed right)分配，峰態係數=3.45>0，為高峽峰(Leptokurtic)。由 Shapiro-Wilk 常態性檢定，可看出解釋變數 Y(總生育率)不符合常態分配。

UNIVARIATE 程序

變數: y (fertility)

動差			
N	1530	總和權重	1530
平均值	1.90967536	總和觀測	2921.80329
標準差	0.63468739	變異數	0.40282808
偏態	1.73138188	峰度	3.44919681
未校正平方和	6195.61988	校正平方和	615.924137
變異係數	33.2353553	標準誤差平均值	0.0162261

常態性檢定				
檢定	統計值	p 值		
Shapiro-Wilk	W	0.84342	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.148697	Pr > D	<0.0100
Cramer-von Mises	W-Sq	11.39713	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	66.23917	Pr > A-Sq	<0.0050



二、女性勞動力參與率(解釋變數 X1)

女性勞動力參與率為女性勞動力占 15 歲以上女性民間人口的比率，亦即在 15 歲以上女性民間人口中有參與勞動的比率。因勞動力包含就業者與失業者，故無論是就業者或失業者的增減，都會影響勞參率的升降。公式如下：

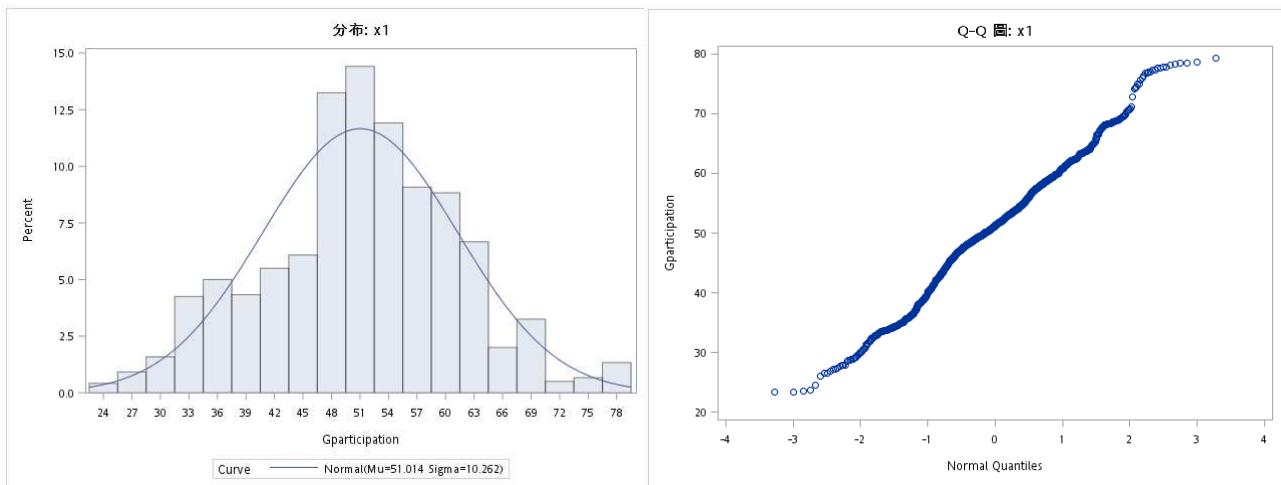
$$\text{女性勞動力參與率}(\%) = \frac{\text{女性勞動力人口}}{15\text{歲以上女性民間人口}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,200 筆女性勞動力參與率資料，平均值為 51.0，標準差為 10.26，偏態係數 $=-0.069 < 0$ ，為微幅左偏(skewed left)分配(接近對稱分配)，峰態係數 $=-0.015 < 0$ ，為低闊峰(Platkurtic)。由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X1 (女性勞動力參與率)不符合常態分配。

UNIVARIATE 程序
變數: x1 (Gparticipation)

動差			
N	1200	總和權重	1200
平均值	51.0140346	總和觀測	61216.8415
標準差	10.2616123	變異數	105.300686
偏態	-0.0694615	峰度	-0.0154868
未校正平方和	3249173.59	校正平方和	126255.523
變異係數	20.1152729	標準誤差平均值	0.29622723

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.991751	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.053566	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.568464	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.438005	Pr > A-Sq	<0.0050



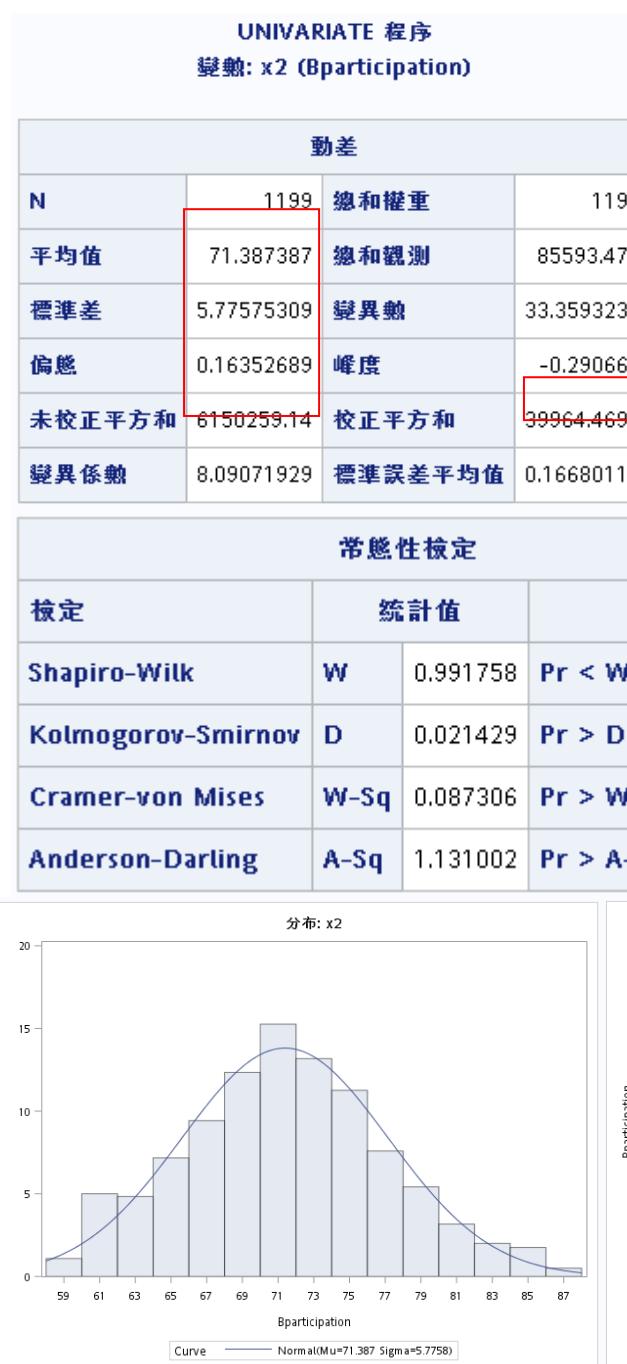
三、男性勞動力參與率(解釋變數 X2)

男性勞動力參與率為男性勞動力占 15 歲以上男性民間人口的比率，亦即在 15 歲以上男性民間人口中有參與勞動的比率。因勞動力包含就業者與失業者，故無論是就業者或失業者的增減，都會影響勞參率的升降。公式如

下：

$$\text{男性勞動力參與率}(\%) = \frac{\text{男性勞動力人口}}{\text{15歲以上男性民間人口}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,199 筆男性勞動力參與率資料，平均值為 71.4，標準差為 5.78，偏態係數=0.1635>0，為右偏(skewed right)分配，峰態係數=-0.29<0，為低闊峰(Platkurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X2 (男性勞動力參與率)不符合常態分配。



四、勞動力參與率(解釋變數 X3)

勞動力參與率為勞動力占 15 歲以上女性民間人口的比率，亦即在 15 歲以上民間人口中有參與勞動的比率。因勞動力包含就業者與失業者，故無論是就業者或失業者的增減，都會影響勞參率的升降。公式如下：

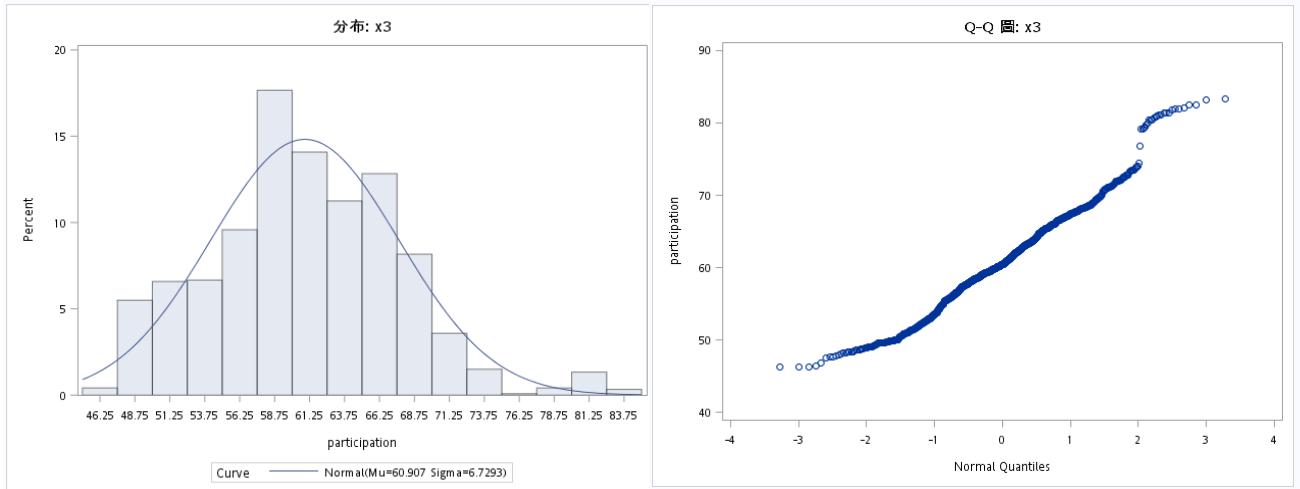
$$\text{勞動力參與率}(\%) = \frac{\text{勞動力人口}}{\text{15歲以上民間人口}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,200 筆勞動力參與率資料，平均值為 60.9，標準差為 6.73，偏態係數 =0.345>0，為右偏(skewed right)分配，峰態係數=0.336>0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X3 (勞動力參與率)不符合常態分配。

UNIVARIATE 程序
變數: x3 (participation)

動差				
N	1200	總和權重	1200	
平均值	60.9066766	總和觀測	73088.0119	
標準差	6.72929124	變異數	45.2833606	
偏態	0.3453492	峰度	0.33598021	
未校正平方和	4505842.65	校正平方和	54294.7494	
變異係數	11.048528	標準誤差平均值	0.19425791	

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.982834	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.03229	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.23954	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.529029	Pr > A-Sq	<0.0050



五、女性失業率(解釋變數 X4)

女性失業率為女性失業者人數占女性勞動力人口的比率，失業者是指年滿 15 歲，具備工作能力，且有工作意願，而在調查當下正在找尋工作者，而就業者為目前有工作的人，包含雇主、自營作業者、受僱者(受私人僱用及受政府僱用)及無酬家屬工作者。公式如下：

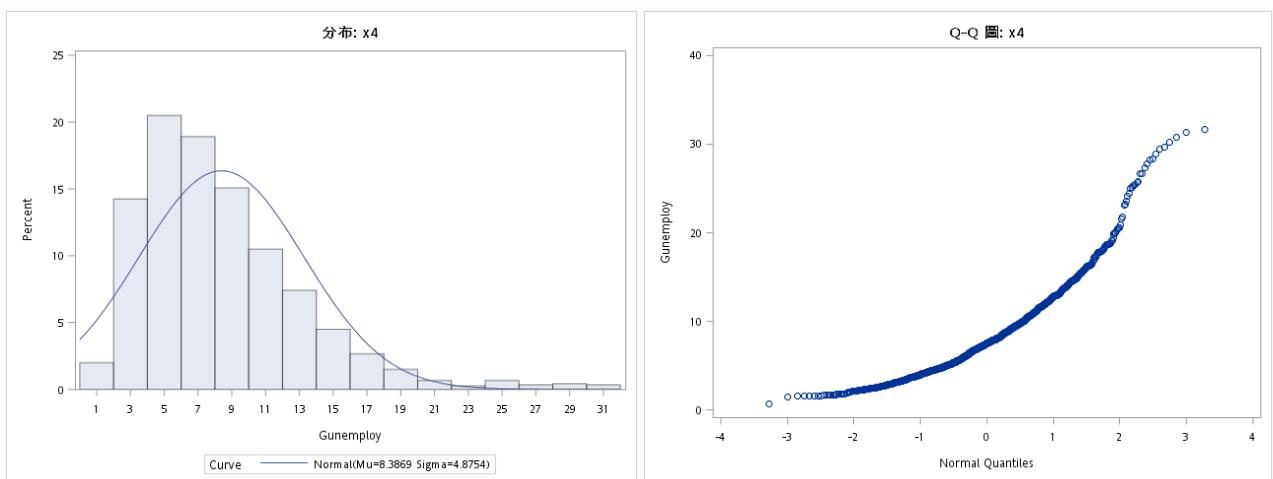
$$\text{女性失業率}(\%) = \frac{\text{女性失業者人數}}{\text{女性勞動力人口}} \times 100\% = \frac{\text{女性失業者人數}}{\text{女性失業者人數} + \text{女性就業者人數}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,200 筆女性失業率資料，平均值為 8.4，標準差為 4.9，偏態係數 = 1.45 > 0，為右偏(skewed right)分配，峰態係數 = 3.17 > 0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X4 (女性失業率)不符合常態分配。

UNIVARIATE 程序
變數: x4 (Gunemploy)

動差			
N	1200	總和權重	1200
平均值	8.38689299	總和觀測	10064.2716
標準差	4.8754492	變異數	23.7700049
偏態	1.45045495	峰度	3.17223514
未校正平方和	112908.205	校正平方和	28500.2359
變異係數	58.1317683	標準誤差平均值	0.1407421

常態性檢定				
檢定		統計值	p 值	
Shapiro-Wilk		W	0.898706	Pr < W <0.0001
Kolmogorov-Smirnov		D	0.092425	Pr > D <0.0100
Cramer-von Mises		W-Sq	3.692333	Pr > W-Sq <0.0050
Anderson-Darling		A-Sq	23.75473	Pr > A-Sq <0.0050



六、男性失業率(解釋變數 X5)

男性失業率為男性失業者人數占男性勞動力人口的比率，失業者是指年滿 15 歲，具備工作能力，且有工作意願，而在調查當下正在找尋工作者，

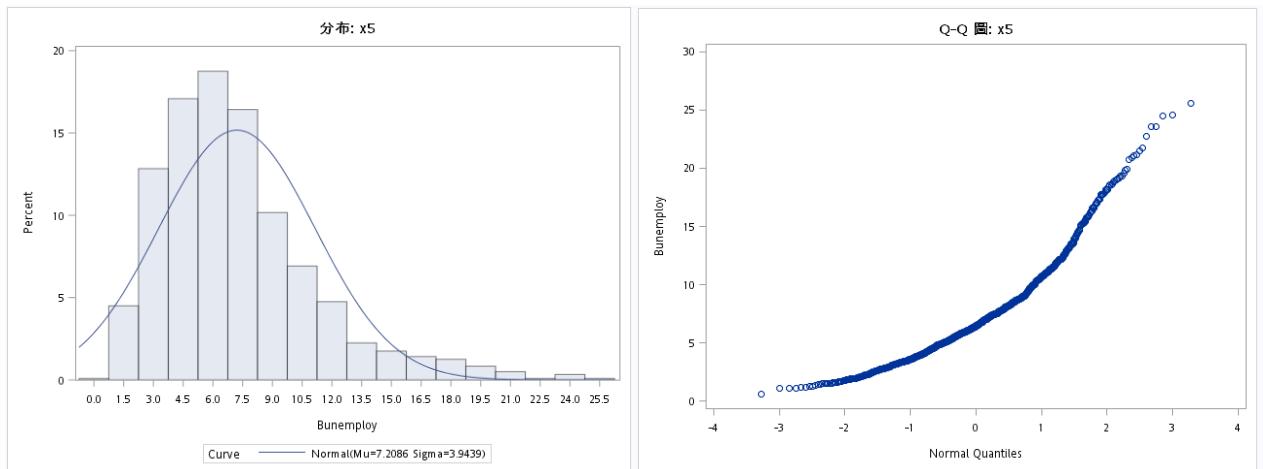
而就業者為目前有工作的人，包含雇主、自營作業者、受僱者(受私人僱用及受政府僱用)及無酬家屬工作者。公式如下：

$$\text{男性失業率}(\%) = \frac{\text{男性失業者人數}}{\text{男性勞動力人口}} \times 100\% = \frac{\text{男性失業者人數}}{\text{男性失業者人數} + \text{男性就業者人數}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,200 筆男性失業率資料，平均值為 7.2，標準差為 3.9，偏態係數=1.33>0，為右偏(skewed right)分配，峰態係數=2.36>0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X5 (男性失業率)不符合常態分配。

UNIVARIATE 程序			
變數: x5 (Bunemploy)			
動差			
N	1200	總和權重	1200
平均值	7.20856603	總和觀測	8650.27923
標準差	3.94387048	變異數	15.5541144
偏態	1.32690753	峰度	2.35657928
未校正平方和	81005.4921	校正平方和	18649.3831
變異係數	54.710888	標準誤差平均值	0.11384973

常態性檢定				
檢定		統計值		p 值
Shapiro-Wilk	W	0.910217	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.096515	Pr > D	<0.0100
Cramer-von Mises	W-Sq	3.916072	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	24.67948	Pr > A-Sq	<0.0050



七、失業率(解釋變數 X6)

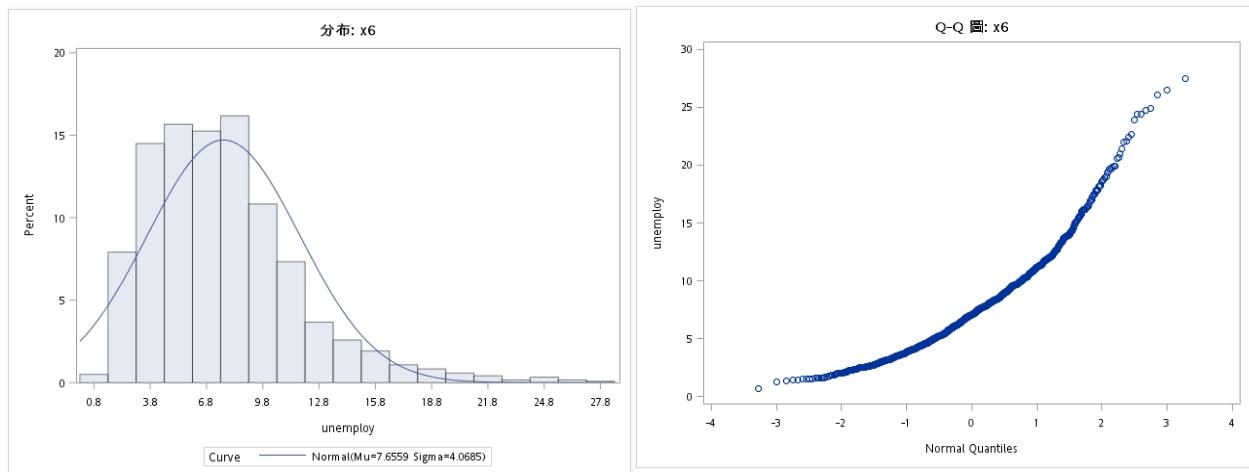
失業率為失業者人數占勞動力人口的比率，失業者是指年滿 15 歲，具備工作能力，且有工作意願，而在調查當下正在找尋工作者，而就業者為目前有工作的人，包含雇主、自營作業者、受僱者(受私人僱用及受政府僱用)及無酬家屬工作者。公式如下：

$$\text{失業率}(\%) = \frac{\text{失業者人數}}{\text{勞動力人口}} \times 100\% = \frac{\text{失業者人數}}{\text{失業者人數} + \text{就業者人數}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,200 筆失業率資料，平均值為 7.66，標準差為 4.07，偏態係數=1.28>0，為右偏(skewed right)分配，峰態係數=2.56>0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X6 (失業率)不符合常態分配。

UNIVARIATE 程序			
變數: x6 (unemploy)			
動差			
N	1200	總和權重	1200
平均值	7.65594679	總和觀測	9187.13615
標準差	4.06845045	變異動	16.5522891
偏態	1.28106758	峰度	2.55838725
未校正平方和	90182.4202	校正平方和	19846.1946
變異係數	53.1410492	標準誤差平均值	0.11744605

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.919873	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.079668	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.623132	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	18.03419	Pr > A-Sq	<0.0050



八、平均每年工時(解釋變數 X7)

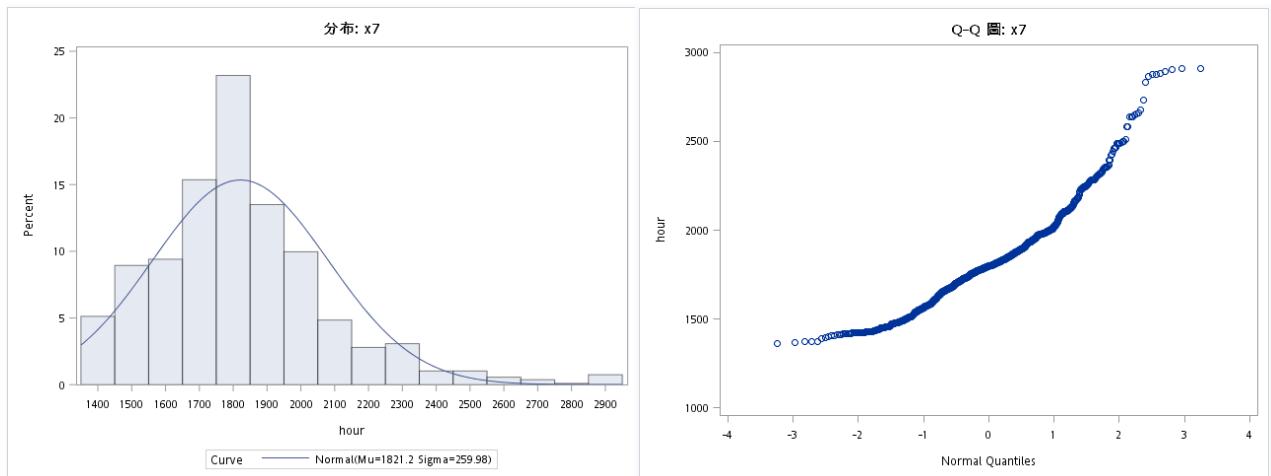
經濟合作發展組織(Organisation for Economic Co-operation and Development, OECD)公佈之各國就業者每年工作總小時數，包含加班時數。OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,074 筆平均每年工時資料，平均值為 1821 小時，標準差為 260，偏態係數 = 1.07 > 0，為右偏(skewed right)分配，峰態係數 = 2.16 > 0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X7 (平均每年工時)不符合常態分配。

UNIVARIATE 程序

變數: x7 (hour)

動差			
N	1074	總和權重	1074
平均值	1821.17358	總和觀測	1955940.42
標準差	259.97518	變異數	67587.0943
偏態	1.06951005	峰度	2.15810462
未校正平方和	3634627960	校正平方和	72520952.1
變異係數	14.2751456	標準誤差平均值	7.93285922

常態性檢定				
檢定	統計值	p 值		
Shapiro-Wilk	W	0.937813	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.084677	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.969043	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	12.41286	Pr > A-Sq	<0.0050



九、扶養比(解釋變數 X8)

扶養比指的是每 100 個工作年齡人口（15 至 64 歲人口）所需負擔依賴人口(即 14 歲以下幼年人口及 65 歲以上老人人口)之比率。扶養比率愈高，表示每位有生產力的成年人所需扶養的無生產力年齡人口愈多，負擔也愈重。

公式如下：

$$\text{扶養比} = \frac{65\text{歲以上人口數} + 0\sim14\text{歲人口數}}{15\sim64\text{歲人口數}} \times 100\%$$

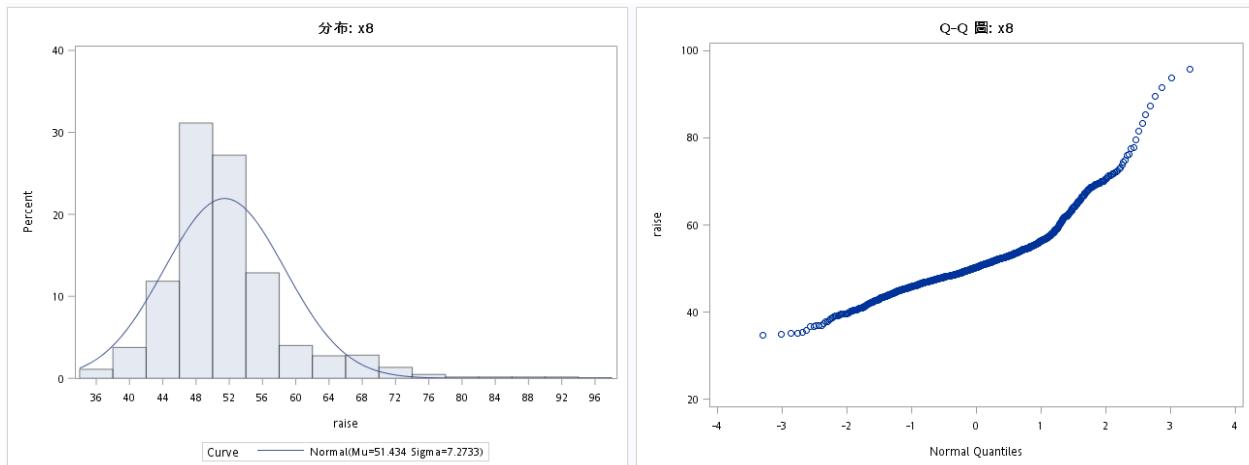
OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,275 筆扶養比資料，平均值為 51.43，標準差為 7.27，偏態係數=1.6>0，為右偏(skewed right)分配，峰態係數=5.06>0，為高峽峰(leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X8 (扶養比)不符合常態分配。

UNIVARIATE 程序

變動: x8 (raise)

動差			
N	1275	總和權重	1275
平均值	51.4341329	總和觀測	65578.5195
標準差	7.27334693	變異數	52.9015756
偏態	1.59888239	峰度	5.06313672
未校正平方和	3440370.9	校正平方和	67396.6073
變異係數	14.1410898	標準誤差平均值	0.20369446

常態性檢定				
檢定	統計值	p 值		
Shapiro-Wilk	W	0.891153	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.118989	Pr > D	<0.0100
Cramer-von Mises	W-Sq	6.014671	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	34.41005	Pr > A-Sq	<0.0050



十、扶幼比 (解釋變數 X9)

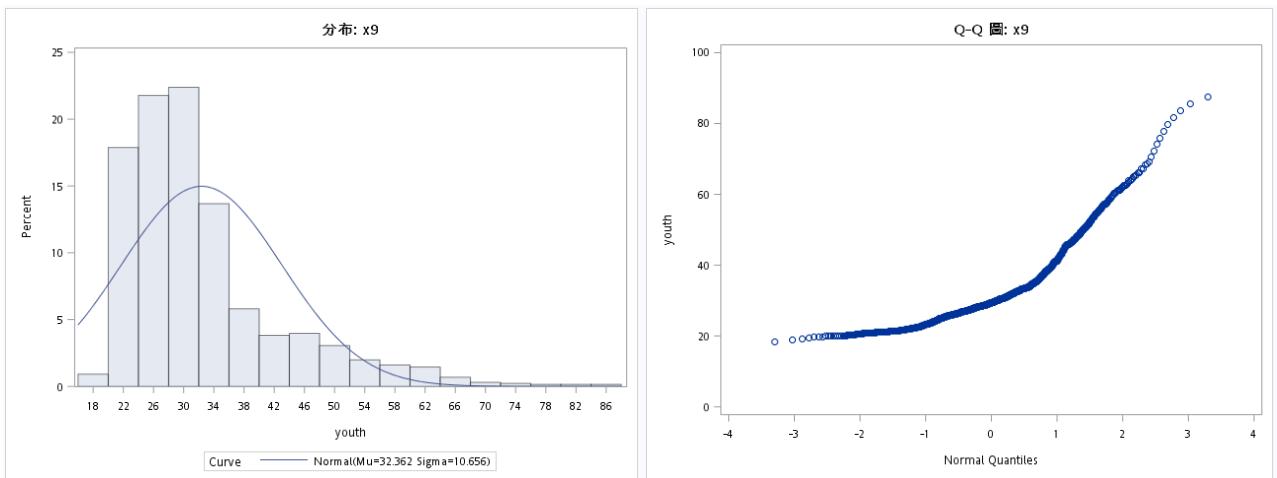
扶幼比定義為 0~14 歲幼年人口占 15~64 歲人口數之比率。公式如下：

$$\text{扶幼比} = \frac{0\sim14\text{歲人口數}}{15\sim64\text{歲人口數}} \times 100\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,309 筆扶幼比資料，平均值為 32.36，標準差為 10.7，偏態係數=1.68>0，為右偏(skewed right)分配，峰態係數=3.3>0，為高峽峰(leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X9 (扶幼比)不符合常態分配。

UNIVARIATE 程序			
變數: x9 (youth)			
動差			
N	1309	總和權重	1309
平均值	32.3624853	總和觀測	42362.4932
標準差	10.6556836	變異動	113.543593
偏態	1.68489042	峰度	3.30193284
未校正平方和	1519470.58	校正平方和	148515.019
變異係數	32.9260361	標準誤差平均值	0.29451776

常態性檢定				
檢定		統計值	p 值	
Shapiro-Wilk	W	0.845129	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.159079	Pr > D	<0.0100
Cramer-von Mises	W-Sq	10.39737	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	58.59005	Pr > A-Sq	<0.0050



十一、扶老比 (解釋變數 X10)

扶老比定義為 65 歲以上老人人口占 15~64 歲人口數之比率。公式如下：

$$\text{扶老比} = \frac{\text{65歲以上人口數}}{\text{15~64歲人口數}} \times 100\%$$

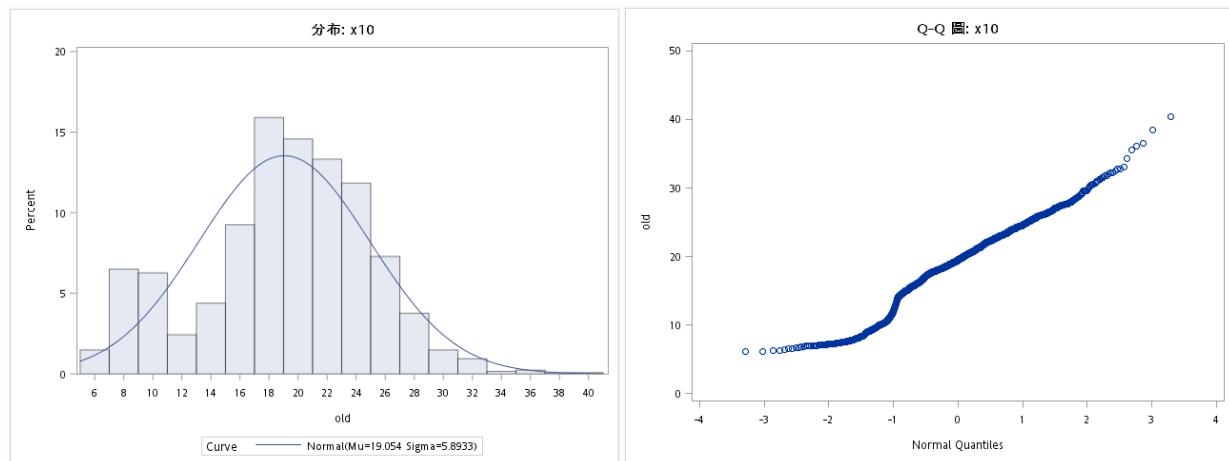
OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,276 筆扶老比資料，平均值為 19.05，標準差為 5.9，偏態係數 = -0.248 < 0，為左偏(skewed left)分配，峰態係數 = -0.135 < 0，為低闊峰(Platkurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X10 (扶老比)不符合常態分配。

UNIVARIATE 程序

變數: x10 (old)

動差			
N	1276	總和權重	1276
平均值	19.0536441	總和觀測	24312.4499
標準差	5.89333777	變異數	34.73143
偏態	-0.2480618	峰度	-0.1348346
未校正平方和	507523.343	校正平方和	44282.5733
變異係數	30.9302395	標準誤差平均值	0.16498178

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.975227	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.067829	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.428493	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	10.34424	Pr > A-Sq	<0.0050



十二、平均每年薪資(解釋變數 X11)

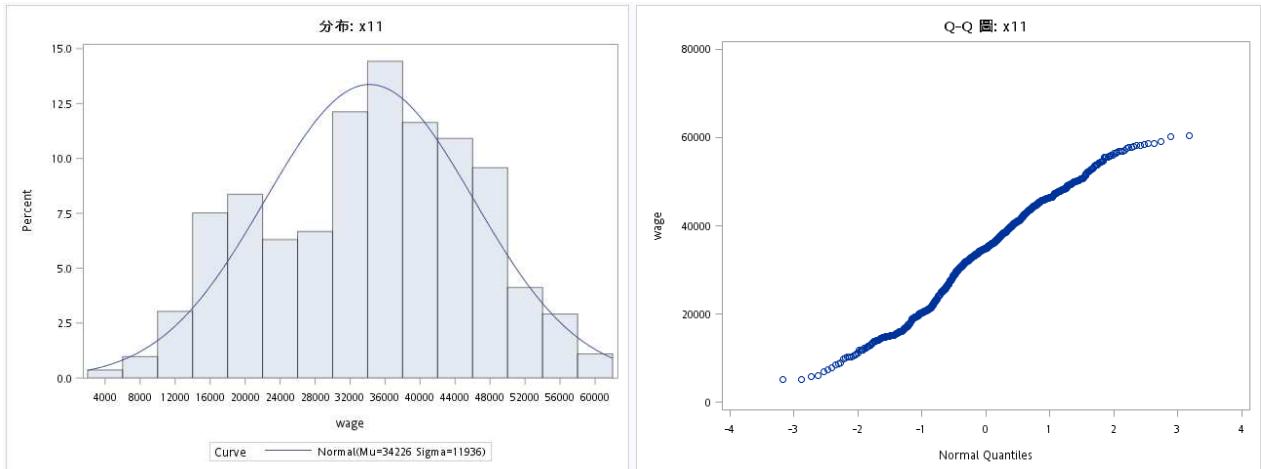
受僱員工每人月平均薪資包括經常性薪資(本薪與按月給付之固定津貼及獎金)、加班費及其他非經常性薪資(如非按月獎金、年節及績效獎金、員工紅利、差旅費、誤餐費及補發調薪差額等)。經過美元購買力平價指數(Purchasing Power Parity conversion rate, PPP)平減後每年平均薪資(美元)。OECD 各國及我國共 44 個國家從 1980 年到 2015 年,扣掉遺漏值總共有 825 筆平均每年薪資資料,平均值為 34,226 元,標準差為 11936,偏態係數 =-0.192<0,為左偏(skewed left)分配,峰態係數=-0.69<0,為低闊峰(Platkurtic)。且由 Shapiro-Wilk 常態性檢定,可看出解釋變數 X11(平均每年薪資)不符合常態分配。

UNIVARIATE 程序

變動: x11 (wage)

動差			
N	825	總和權重	825
平均值	34226.377	總和觀測	28236761
標準差	11936.4861	變異數	142479700
偏態	-0.1920163	峰度	-0.6865151
未校正平方和	1.08385E12	校正平方和	1.17403E11
變異係數	34.8751084	標準誤差平均值	415.575103

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.983507	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.054389	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.602732	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.910664	Pr > A-Sq	<0.0050



十三、名目 GDP 成長率(解釋變數 X12)

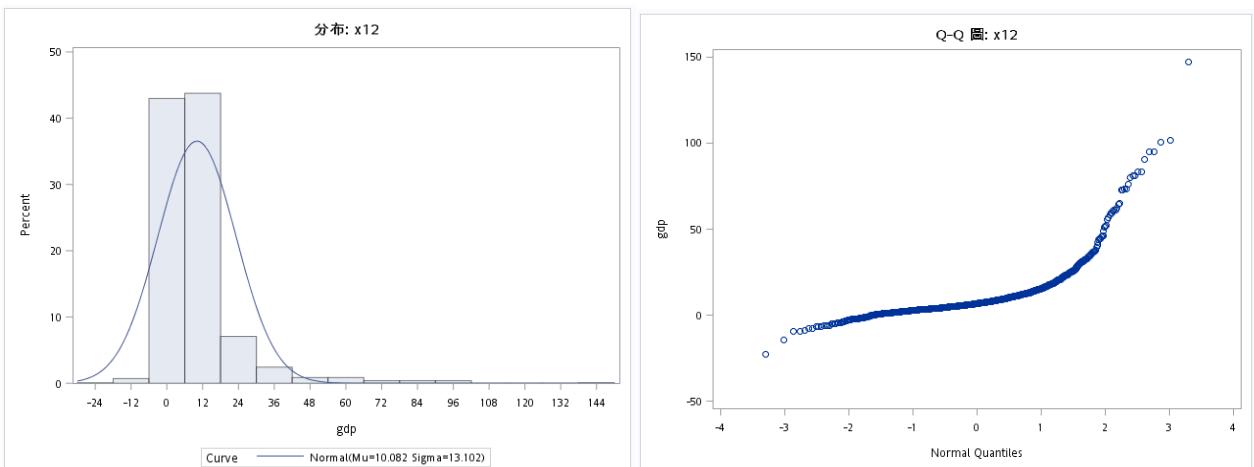
名目 GDP 成長率是用來衡量一個國家或地區，經濟活動暢旺或衰退的最重要指標，各國計算方式不同、取樣也不同，簡單的定義就是每季或每一

年度，換算成美元之名目國內生產毛額(GDP)的年成長率。而 GDP 就是指一個地區或國家在特定時間內，通常 1 季或 1 年，所有生產產品和貨物的總值。OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,275 筆名目 GDP 成長率資料，平均值為 10.08，標準差為 13.1，變異較大，偏態係數=3.938>0，為右偏(skewed right)分配，峰態係數=22.9>0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X12 (名目 GDP 成長率)不符合常態分配。

UNIVARIATE 程序
變數: x12 (gdp)

動差			
N	1275	總和權重	1275
平均值	10.0824514	總和觀測	12855.1255
標準差	13.1021337	變異數	171.665908
偏態	3.93823749	峰度	22.8694625
未校正平方和	348313.545	校正平方和	218702.367
變異係數	129.949883	標準誤差平均值	0.36693314

常態性檢定				
檢定		統計值		p 值
Shapiro-Wilk	W	0.641404	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.204543	Pr > D	<0.0100
Cramer-von Mises	W-Sq	21.51204	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	115.9737	Pr > A-Sq	<0.0050



十四、結婚率(解釋變數 X13)

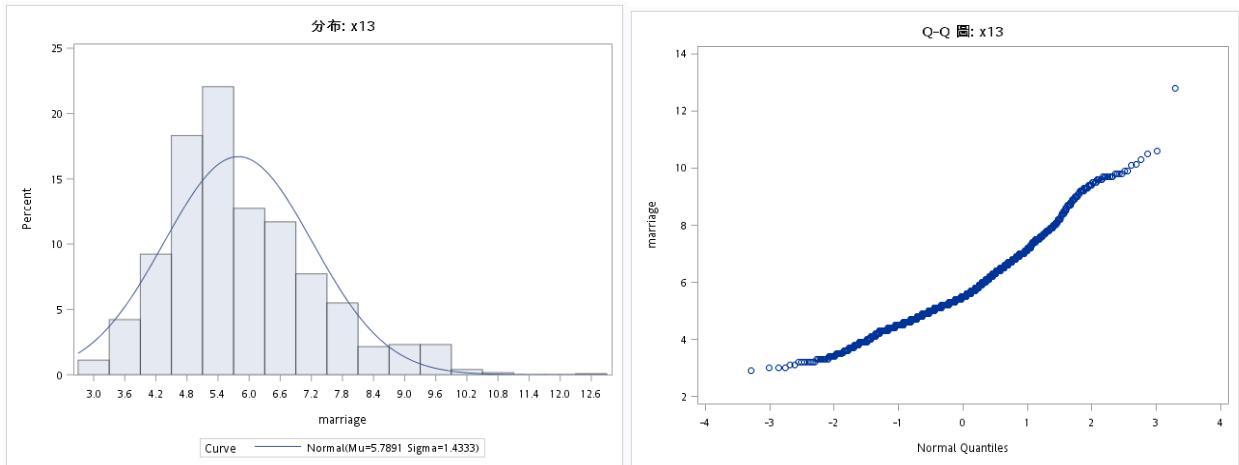
表示 1 年內，每一千人當中有多少人會結婚。粗結婚率為某一特定期間之結婚對數對同一期間之期中人口數之百分比，結婚登記對數為當年戶籍結婚登記對數，年中人口數為(上年年底人口數+本年年底人口數)/2。單位為對/千人。公式如下：

$$\text{結婚率}(\%) = \frac{\text{結婚登記對數}}{\text{年中人口數}} \times 1000\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,256 筆名目結婚率資料，平均值為 5.8，標準差為 1.4，偏態係數=0.833>0，為右偏(skewed right)分配，峰態係數=0.757>0，為高峽峰(Leptokurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X13 (結婚率)不符合常態分配。

UNIVARIATE 程序				
變動: x13 (marriage)				
動差				
N	1256	總和權重	1256	
平均值	5.78907135	總和觀測	7271.07361	
標準差	1.43334385	變異數	2.0544746	
偏態	0.83302155	峰度	0.75709031	
未校正平方和	44671.1296	校正平方和	2578.36562	
變異係數	24.7594781	標準誤差平均值	0.04044414	

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.956319	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.100616	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.67908	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	15.52671	Pr > A-Sq	<0.0050



十五、離婚率(解釋變數 X14)

表示 1 年內，每一千人當中有多少人會離婚。粗離婚率為某一特定期間之離婚對數對同一期間之期中人口數之百分比，離婚登記對數為當年戶籍離婚登記對數，年中人口數為(上年年底人口數+本年年底人口數)/2。單位為對/千人。公式如下：

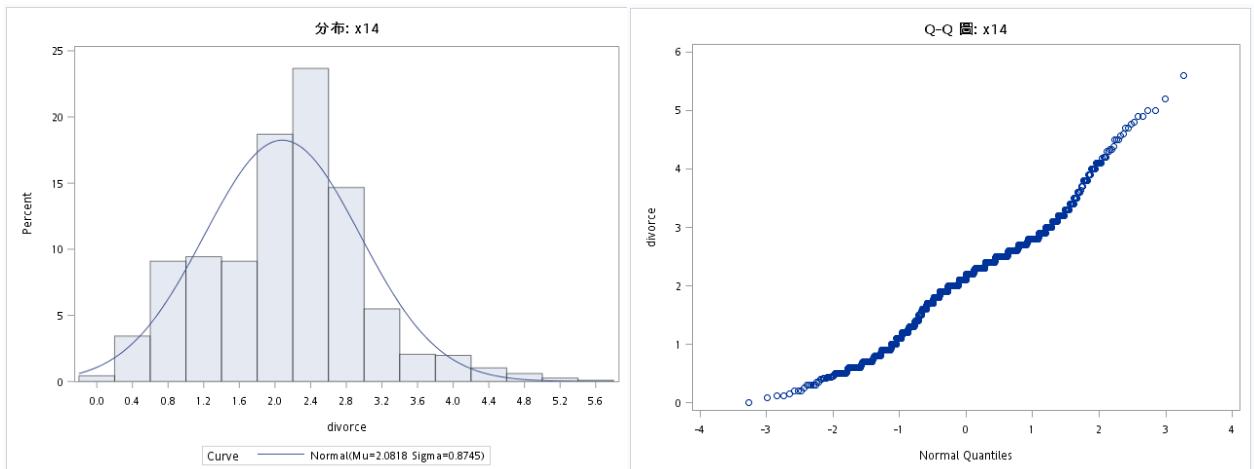
$$\text{離婚率}(\%) = \frac{\text{離婚登記對數}}{\text{年中人口數}} \times 1000\%$$

OECD 各國及我國共 44 個國家從 1980 年到 2015 年，扣掉遺漏值總共有 1,166 筆名目離婚率資料，平均值為 2.1，標準差為 0.9，偏態係數=0.245>0，為右偏(skewed right)分配，峰態係數=0.56<3，為低闊峰(Platkurtic)。且由 Shapiro-Wilk 常態性檢定，可看出解釋變數 X14 (離婚率)不符合常態分配。

UNIVARIATE 程序
變數: x14 (divorce)

動差			
N	1166	總和權重	1166
平均值	2.0817725	總和觀測	2427.34674
標準差	0.87445833	變異動	0.76467737
偏態	0.24504151	峰度	0.55847945
未校正平方和	5944.03283	校正平方和	890.84914
變異係數	42.0054704	標準誤差平均值	0.02560884

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.978083	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.071182	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.194187	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	7.383509	Pr > A-Sq	<0.0050



各解釋變數之簡單敘述統計分析如下：

MEANS 程序

變數	標籤	N	平均值	標準差	最小值	最大值
x1	Gparticipation	1200	51.0	10.3	23.3	79.3
x2	Bparticipation	1199	71.4	5.8	58.9	87.8
x3	participation	1200	60.9	6.7	46.2	83.3
x4	Gunemploy	1200	8.4	4.9	0.7	31.6
x5	Bunemploy	1200	7.2	3.9	0.6	25.6
x6	unemploy	1200	7.7	4.1	0.6	27.5
x7	hour	1074	1821.2	260.0	1361.7	2911.0
x8	raise	1275	51.4	7.3	34.7	95.7
x9	youth	1309	32.4	10.7	18.4	87.3
x10	old	1276	19.1	5.9	6.1	40.4
x11	wage	825	34226.4	11936.5	5051.0	60369.0
x12	gdp	1275	10.1	13.1	-22.6	147.4
x13	marriage	1256	5.8	1.4	2.9	12.8
x14	divorce	1166	2.1	0.9	0.0	5.6
y	fertility	1530	1.9	0.6	0.9	4.8

第二章、簡單線性迴歸模型統計分析

第一節、總生育率和女性勞動力參與率之簡單迴歸模型

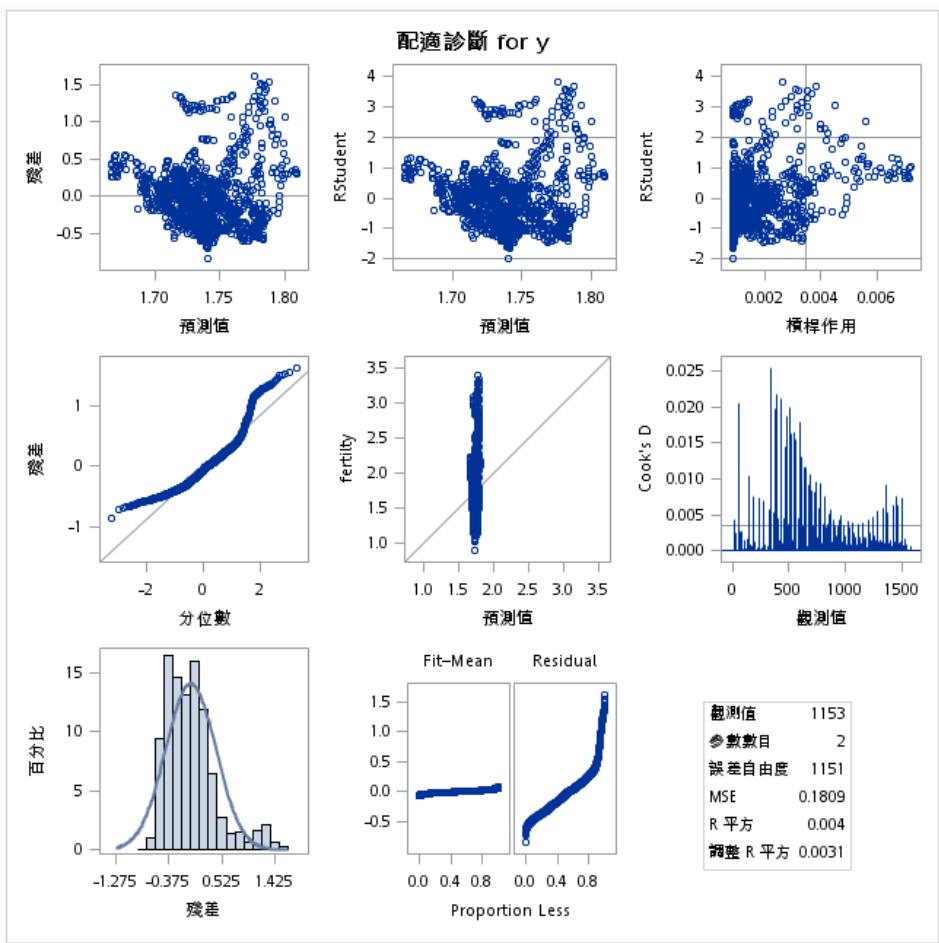
(一)原始模型：

1. 簡單線性迴歸模型

將反應變數 Y 設為 1980 年到 2015 年 OECD 各國及我國總生育率(%)，獨立變數(解釋變數)X 設為 1980 年到 2015 年 OECD 各國及我國女性勞動力參與率(%)，扣除遺漏值後，共有 1,153 筆資料，進行簡單線性迴歸模型統計分析。得到迴歸線為 $Y=1.87038-0.00261X$ ，表示每增加 1%女性勞動力參與率，總生育率減少 0.0261%。另藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0=0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1=0$ 的假設。表示在統計上有證據證明 b_0 和 b_1 參數值都顯著不等於 0。若 $X=50\%$ ，代入公式， $Y=1.7399$ 。

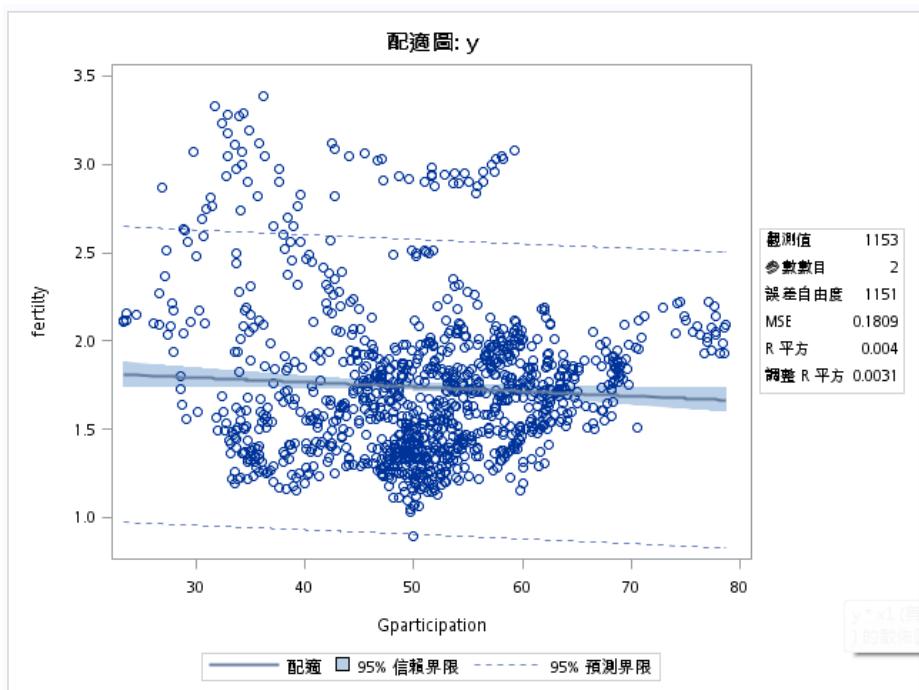
REG 程序		變異數分析				
來源	自由度	平方和	平均值 平方	F 值	Pr > F	
模型	1	0.82912	0.82912	4.58	0.0325	
誤差	1151	208.23087	0.18091			
已校正的總計	1152	209.06000				

參數估計值						
變動	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t
Intercept	Intercept	1	1.87038	0.06314	29.62	<.0001
x	participation	1	-0.00261	0.00122	-2.14	0.0325



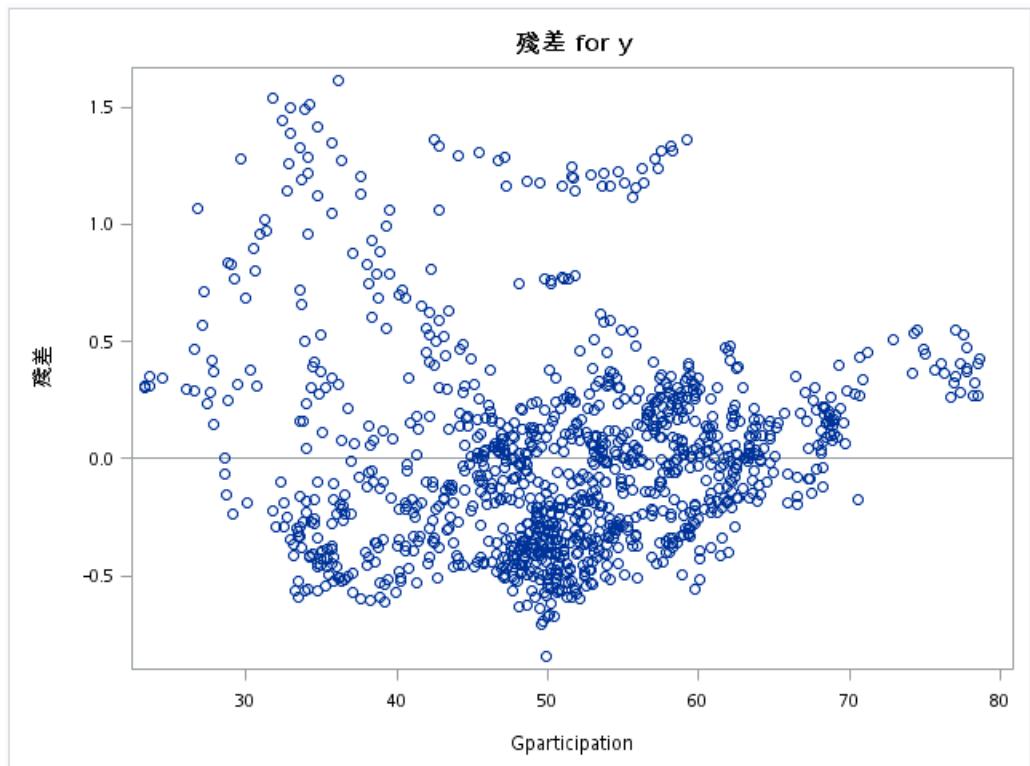
2.X 對 Y 的散佈圖

由 X 對 Y 散佈圖可看出有很多值都落在信賴區間外面，且殘差的變異數不一致情形。



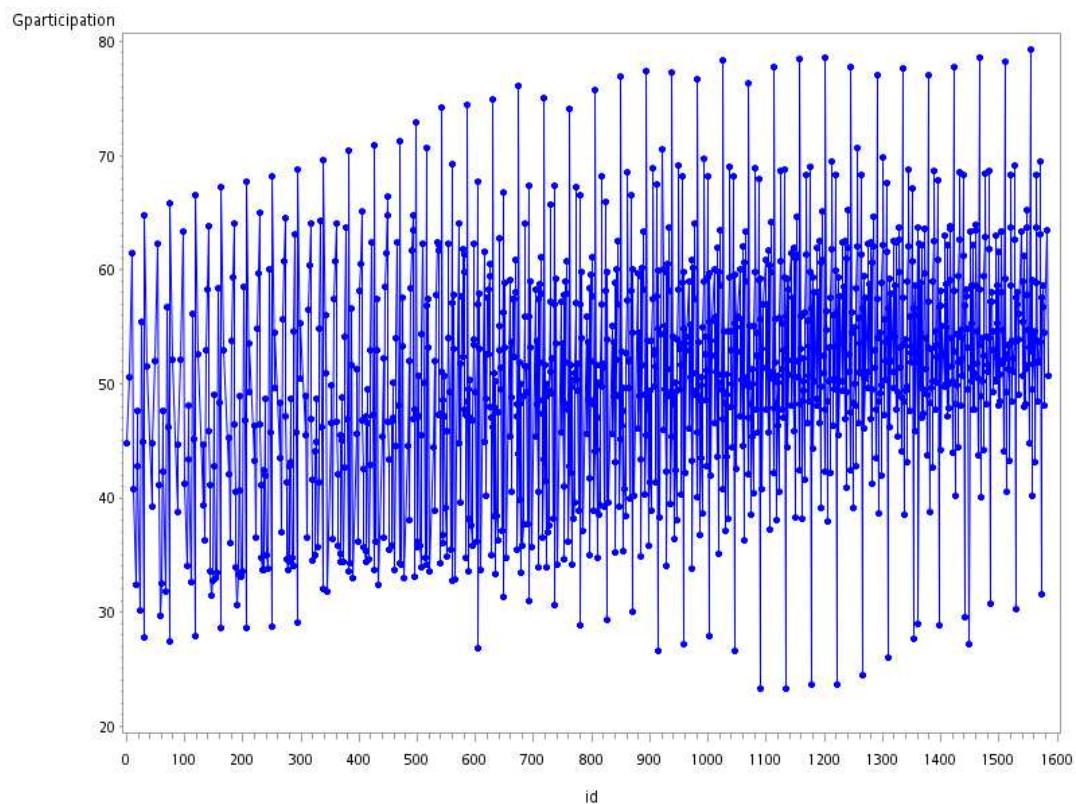
3. 殘差對 X 的散佈圖

由殘差對 X 散佈圖可看出殘差的變異數不一致情形。



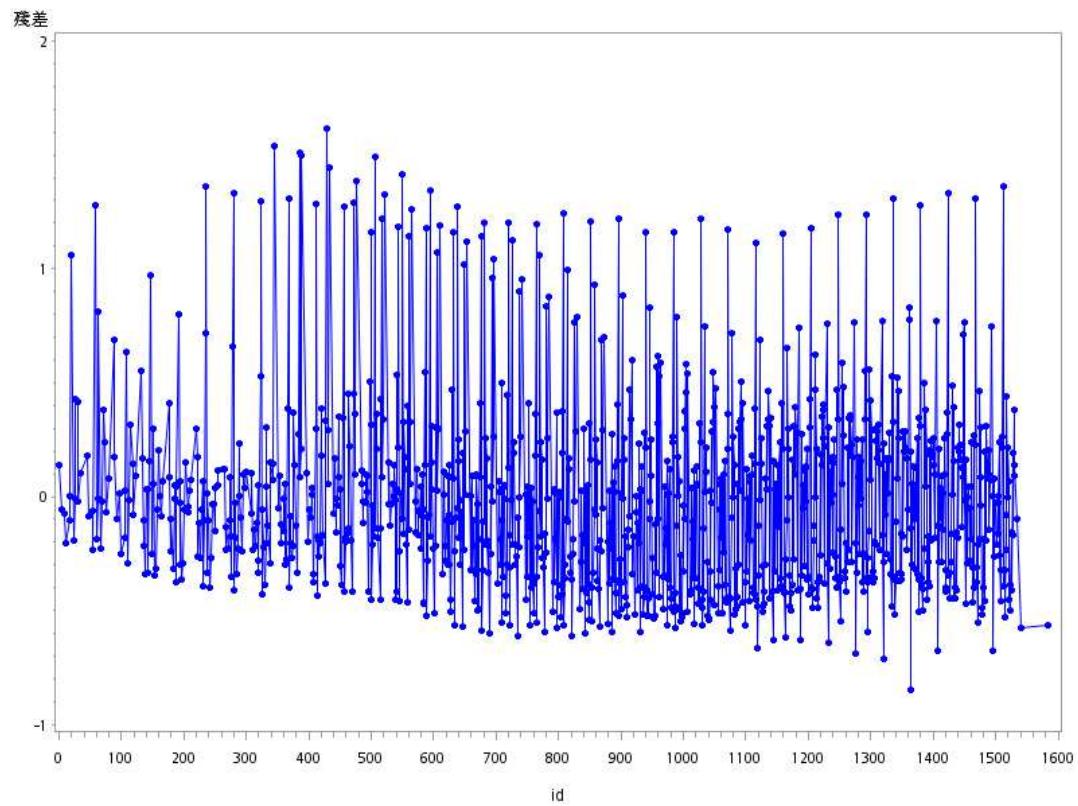
4. X 的時間序列圖

由 X 的時間序列圖可知，X 為隨機連續變數。



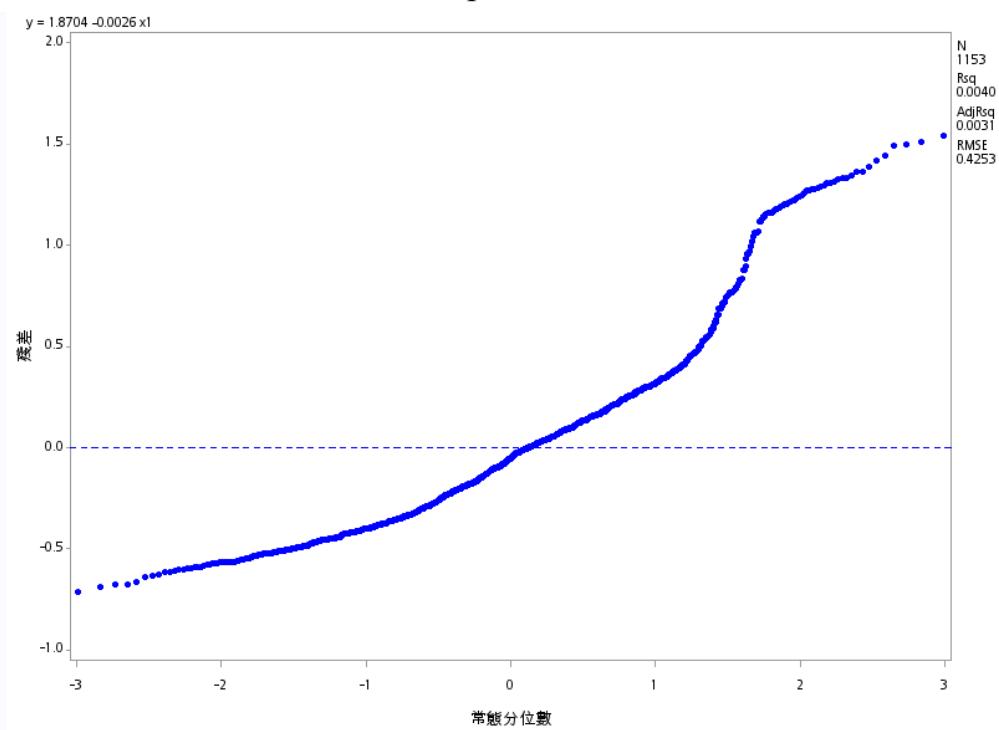
5. 殘差的時間序列圖

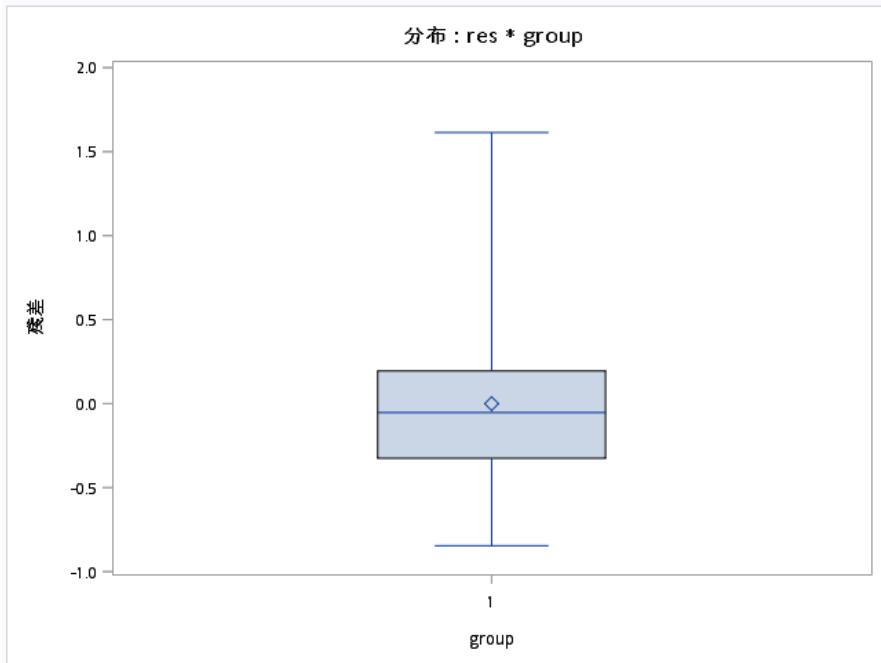
由殘差的時間序列圖可知，殘差為隨機連續變數。



6. 殘差的常態機率圖、箱型圖及與殘差期望值之相關性檢定

由常態機率圖可知殘差為右偏分配，且檢定結果為兩者有顯著相關，Pearson 相關係數為 0.9849，Spearman 相關係數為 1.00。





Pearson 相關係數 Prob > r (位於 H0 底下): Rho=0 觀測值數目		
	res	expec
res	1.00000	0.98490 <.0001
殘差	1153	1153
expec	0.98490 <.0001 1153	1.00000 1584

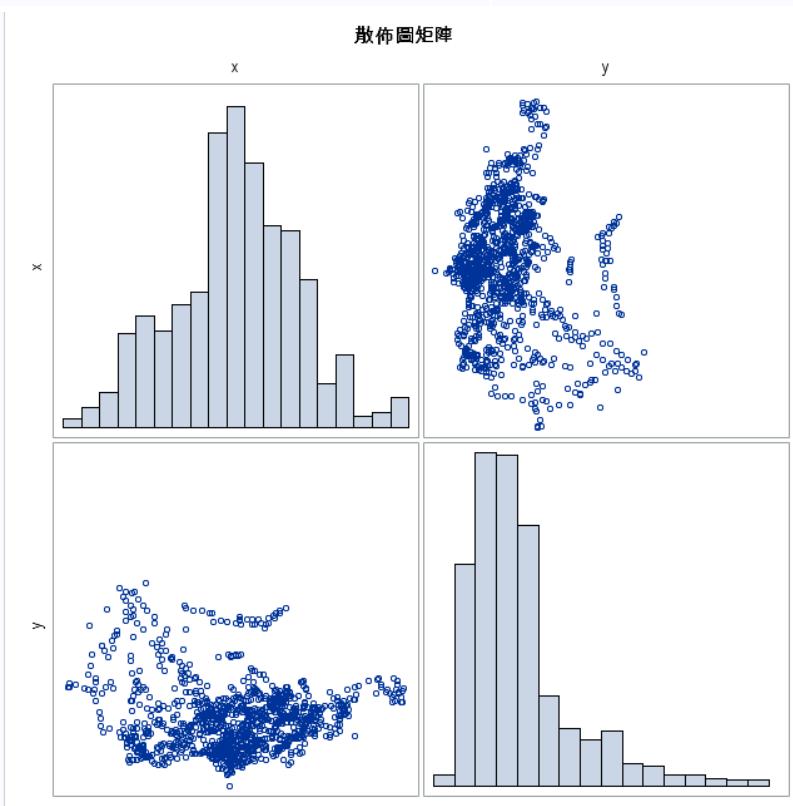
Spearman 相關係數 Prob > r (位於 H0 底下): Rho=0 觀測值數目		
	res	expec
res	1.00000	1.00000 <.0001
殘差	1153	1153
expec	1.00000 <.0001 1153	1.00000 1584

7.X 與 Y 之相關性檢定

因為 X 及 Y 不符合常態性分配，可藉由 Spearman 相關係數 0.09075 知道兩者為微幅正相關，且 P-value 為 $0.002 < 0.05$ ，拒絕 H_0 ，表示統計上有足夠證據證明兩者相關係數不為 0。

簡單統計值							
變數	N	平均值	標準差	中位數	最小值	最大值	標籤
x	1200	51.01403	10.26161	51.25263	23.29849	79.28249	participation
y	1530	1.90968	0.63469	1.76000	0.89500	4.82800	fertility

Pearson 相關係數			Spearman 相關係數		
			觀測值數目		
	x	y		x	y
x	1.00000	-0.06298	x	1.00000	0.09075
participation		0.0325	participation		0.0020
	1200	1153		1200	1153
y	-0.06298	1.00000	y	0.09075	1.00000
fertility	0.0325		fertility	0.0020	
	1153	1530		1153	1530



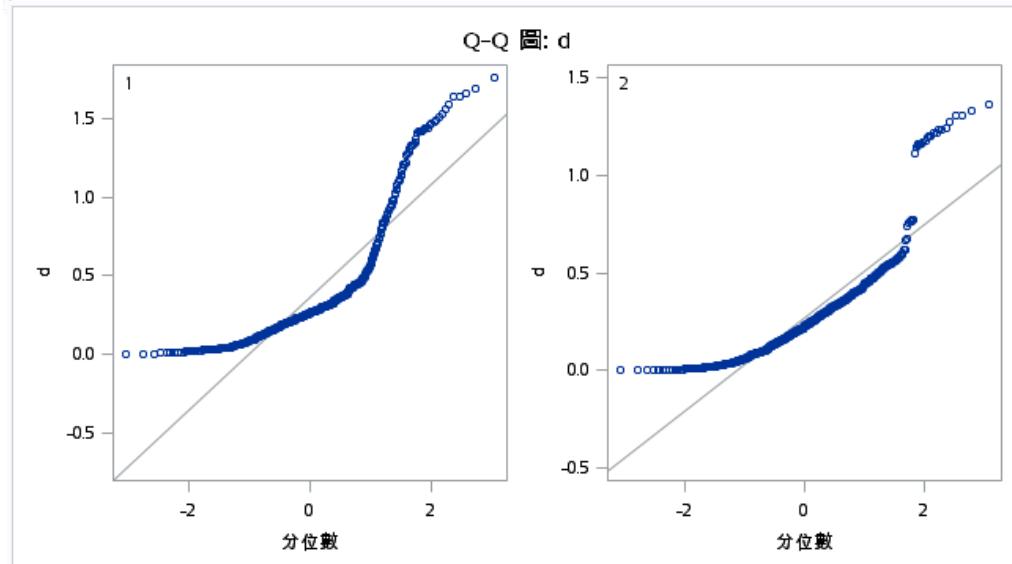
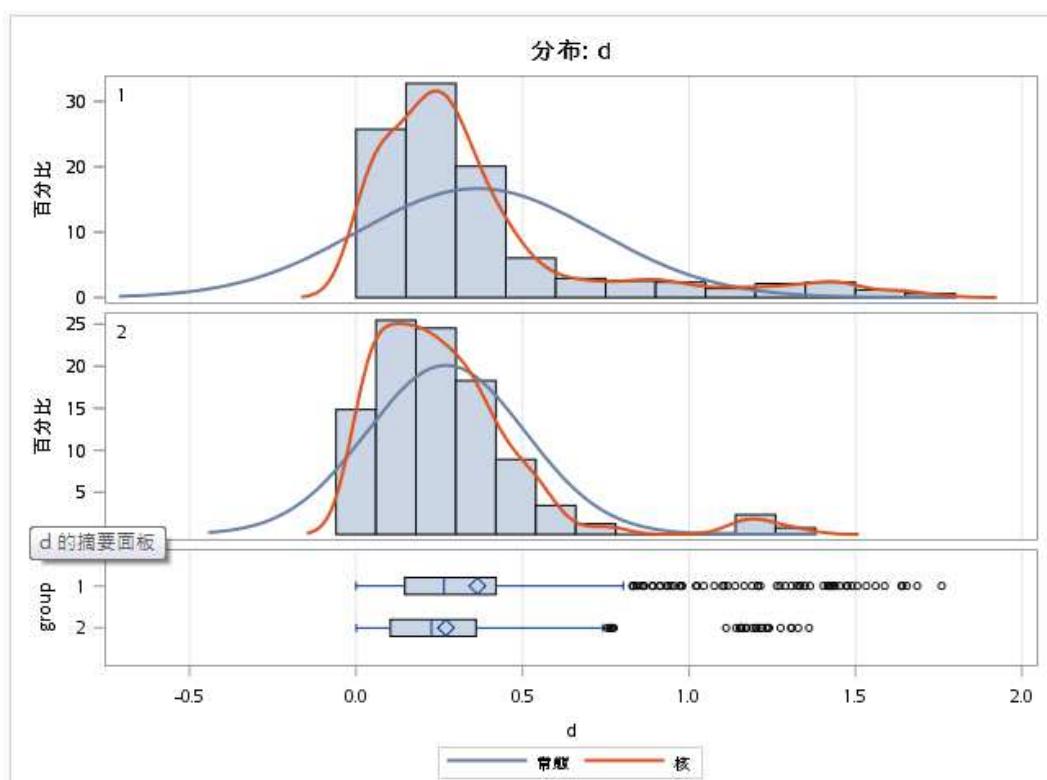
8. Brown-Forsythe test 變異數檢定

利用 Brown-Forsythe test 作檢定，檢定結果 P-value < 0.0001，拒絕 H0，表示統計上有足夠的證據證明殘差的變異數不相等。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	513	0.3645	0.3588	0.0158	0	1.7589
2	640	0.2702	0.2380	0.00941	0.000901	1.3609
Diff (1-2)		0.0942	0.2979	0.0177		

group	方法	平均值	95% CL 平均值	標準差	95% CL 標準差
1		0.3645	0.3333	0.3956	0.3588 0.3382 0.3823
2		0.2702	0.2517	0.2887	0.2380 0.2257 0.2519
Diff (1-2)	集區	0.0942	0.0596	0.1289	0.2979 0.2862 0.3106
Diff (1-2)	Satterthwaite	0.0942	0.0581	0.1304	

方法	變異數	自由度	t 值	Pr > t	變異數相等性		
方法	分子自由度	分母自由度	F 值	Pr > F			
集區	均等	1151	5.34	<.0001			
Satterthwaite	不均等	851.97	5.11	<.0001	Folded F	512	639 2.27 <.0001



9. Brown-Pagan test 變異數檢定

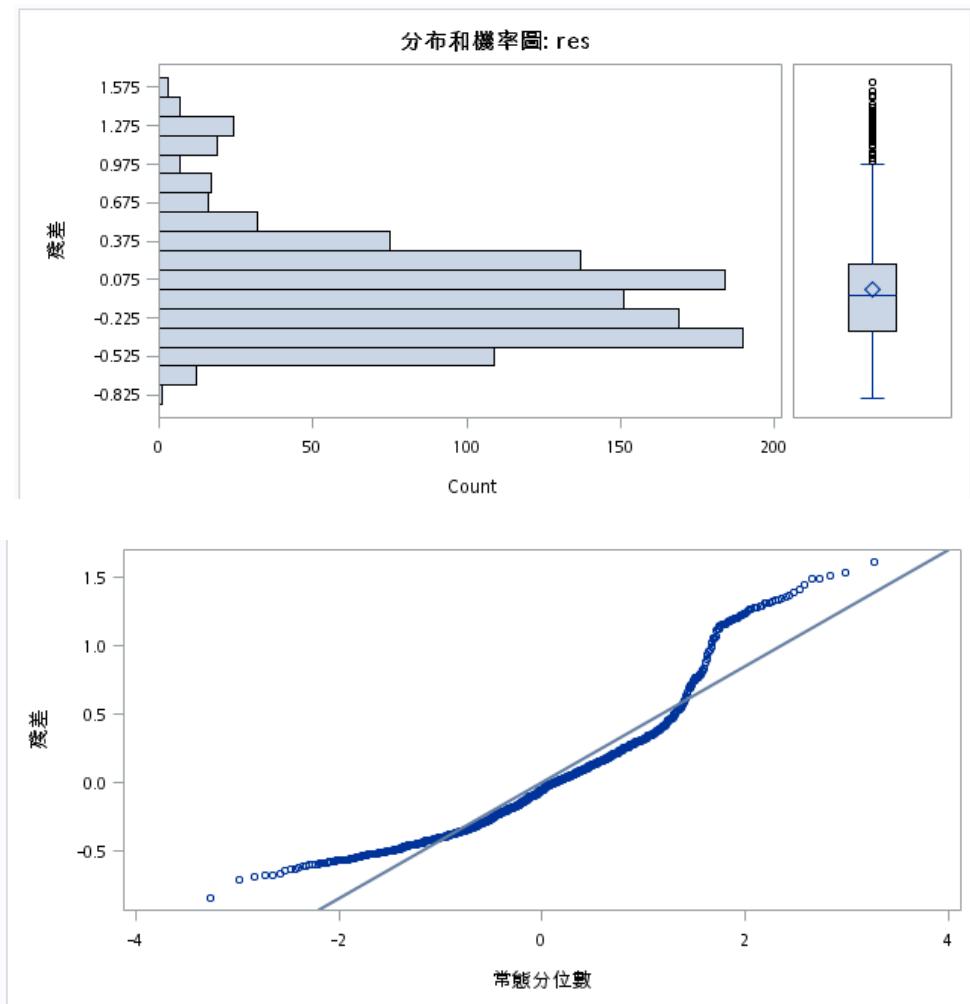
利用 Brown-Pagan test 作檢定，檢定結果 $X_{BP}^2 = 11227.53 > c$ ，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的變異數不相等。

Obs	ssrs	sse	nobs	tests	pv
1	388.056	208.231	1584	11227.53	1

10. Shapiro-Wilk test 常態性檢定

利用 Shapiro-Wilk 對殘差作常態性檢定，檢定結果 P-value < 0.0001，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的分布不符合常態分配。

UNIVARIATE 程序				
變動: res (殘差)				
動差				
N	1153	總和權重	1153	
平均值	0	總和觀測	0	
標準差	0.42515405	變異數	0.18075597	
偏態	1.22720079	峰度	1.76577206	
未校正平方和	208.230874	校正平方和	208.230874	
變異係數	.	標準誤差平均值	0.01252079	
常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.909773	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.078261	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.972767	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	23.03131	Pr > A-Sq	<0.0050



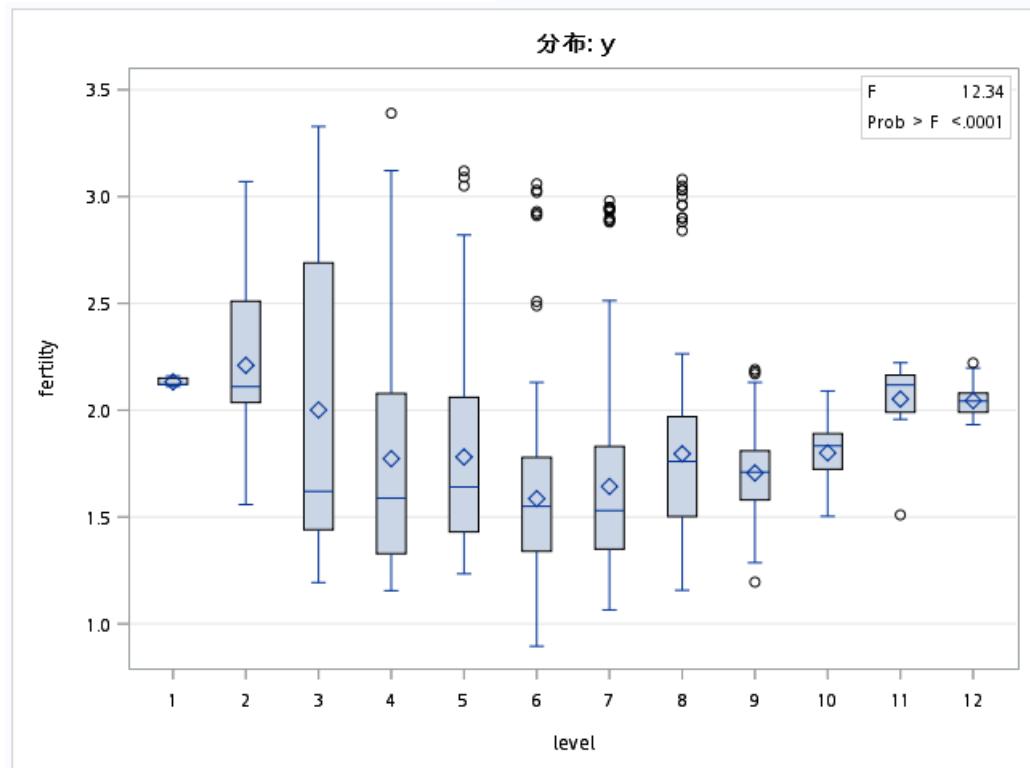
11. F Test for Lack of Fit 缺適度檢定

因為 X_1 沒有重覆觀察值，故只好將附近觀察值是為同一群，進行 F Test for Lack of Fit。因此我們假設 X_1 大於等於 20 小於 25 設為第一組， X_1 大於等於 25 小於 30 設為第二組， X_1 大於等於 30 小於 35 設為第三組，以此類推到， X_1 大於等於 75 小於 80 設為第十二組，總共分為 12 群組。檢定結果 $P\text{-value} < 0.0001$ ，拒絕 H_0 ，表示統計上有足夠的證據證明此迴歸模型不適用於簡單線性迴歸模型。

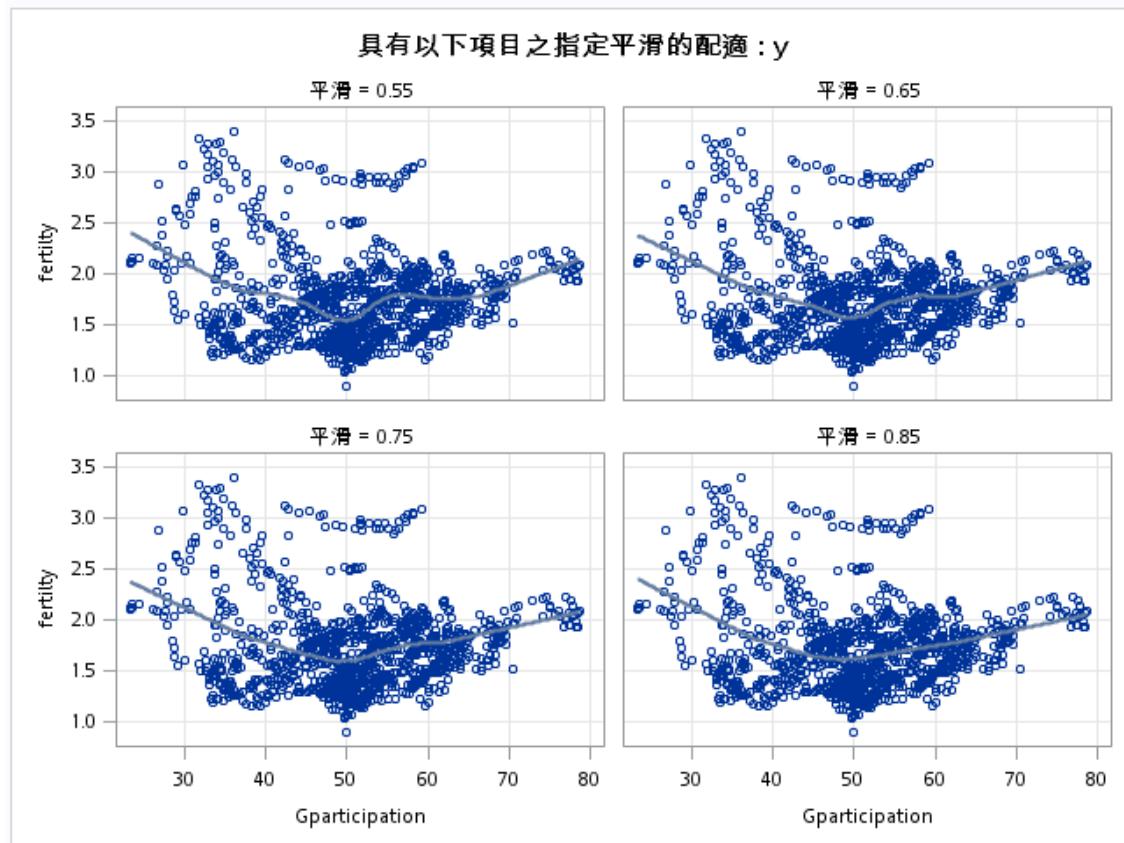
應變數: y fertility

來源	自由度	平方和	均方	F 值	Pr > F
模型	11	22.2235998	2.0203273	12.34	<.0001
誤差	1141	186.8363969	0.1637479		
已校正的總計	1152	209.0599966			

R 平方	變異係數	根 MSE	y 平均值	來源	自由度	ANOVA SS	均方	F 值	Pr > F
0.106302	23.28421	0.404658	1.737907	level	11	22.22359977	2.02032725	12.34	<.0001

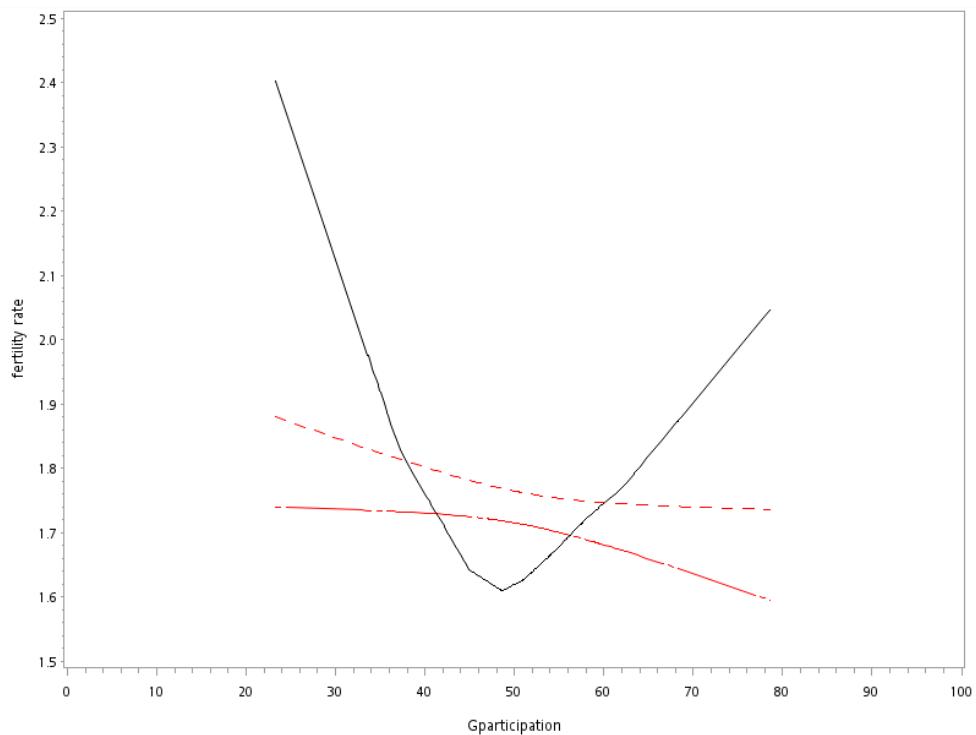


12. Lowess Curve 觀察 Lowess Curve 看趨勢線。



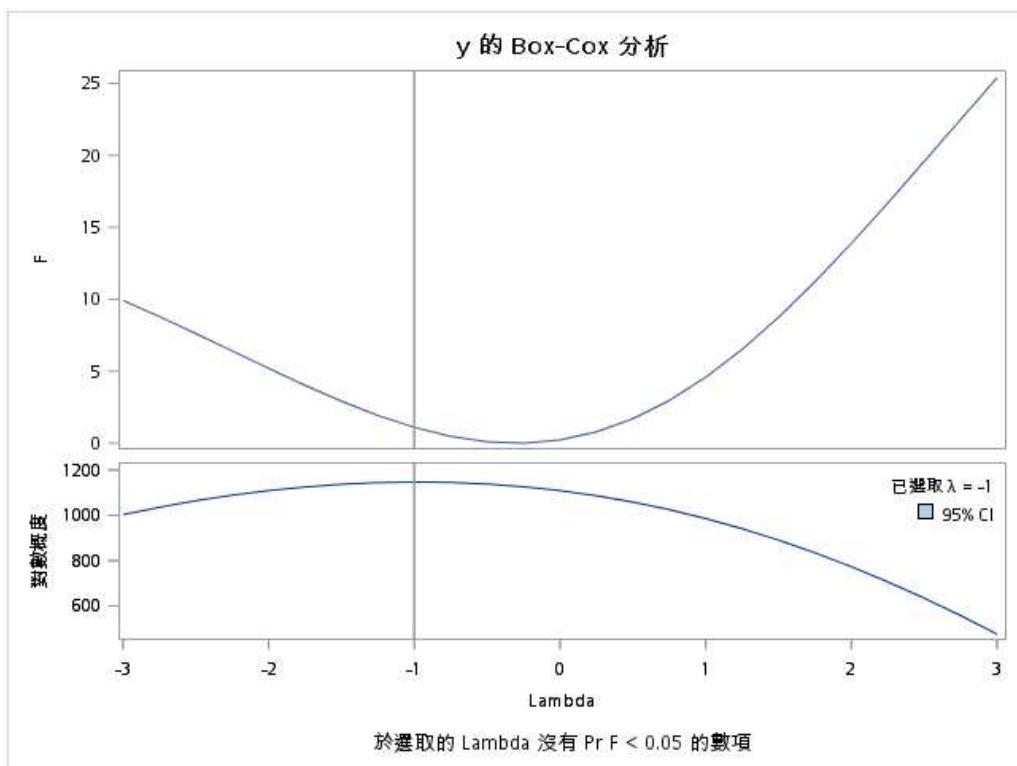
13. 預測線及信賴區間

預測線都沒有包含在信賴區間內，表示此模型配適不恰當。



14. Box-Cox 轉換

綜觀以上，我們發現此迴歸模型殘差不符合常態分配且殘差變異數不為常數值，且由 F test for Lack of Fit 檢定可知，此迴歸模型不適用於簡單線性迴歸。故我們要進行變數轉換，因此利用 Box-Cox 找出最適合的轉換值。



根 MSE	0.12698	R 平方	0.0335
應變平均值	0.39449	調整 R 平	0.0326
變異係數	32.18941	Lambda	-1.0000

根據一般自由度的調整多變量 ANOVA 表格					
應變數計分參數=48 S=1 M=23 N=574.5					
統計值	值	F 值	分子自由度	分母自由度	p
Wilks' Lambda	0.966545	0.83	48	1151	0.7894
Pillai 的追蹤	0.033455	0.83	48	1151	0.7894
Hotelling-Lawley 追蹤	0.034613	0.83	48	1151	0.7894
Roy 最大根	0.034613	0.83	48	1151	0.7894

(二) 變數轉換一(log 轉換)模型：

1. 簡單迴歸模型：

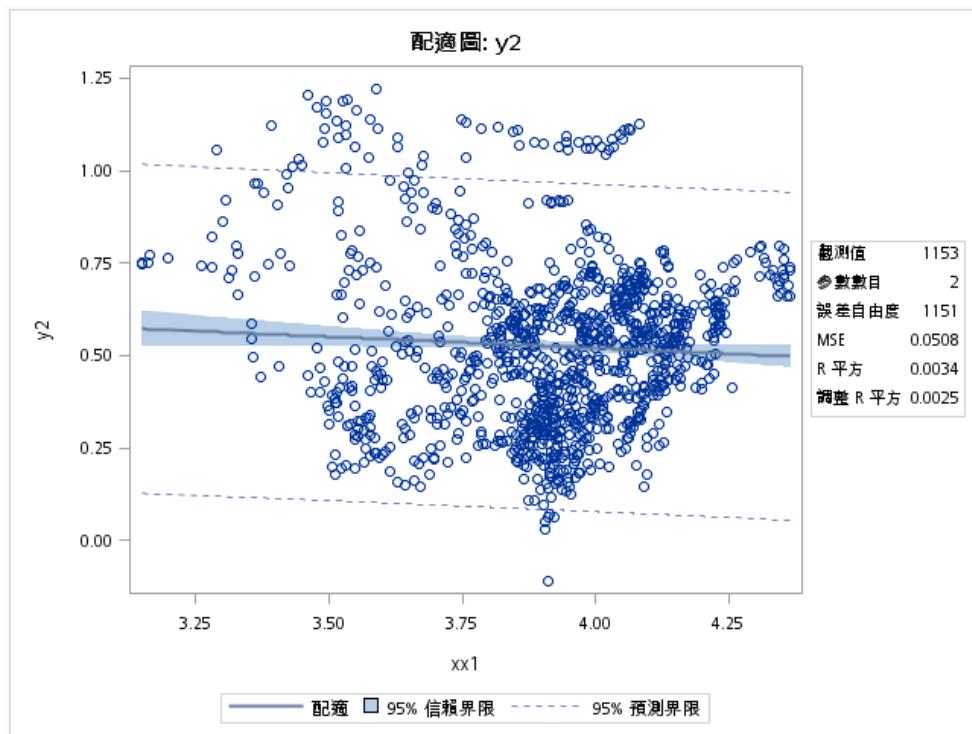
可發現當 lamda 接近 0 時，可得到最小變異數，故我們將 Y 進行 log 轉換，X 也進行 log 轉換。亦即 $X' = \log(X)$ 、 $Y' = \log(Y)$ 。得到迴歸線為 $Y' = 0.76448 - 0.06104 X'$ ，表示每增加 1% 女性勞動力參與率，總生育率減少 0.6104%。藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0 = 0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1 = 0$ 的假設。表示在統計上有證據證明 b_0 和 b_1 參數值都顯著不等於 0。

根 MSE	0.22548	R 平方	0.0034
應變平均值	0.52601	調整 R 平方	0.0025
變異係數	42.86608		

參數估計值					
變數	自由度	參數 估計值	標準 誤差	t 值	Pr > t
Intercept	1	0.76448	0.12085	6.33	<.0001
xx1	1	-0.06104	0.03089	-1.98	0.0484

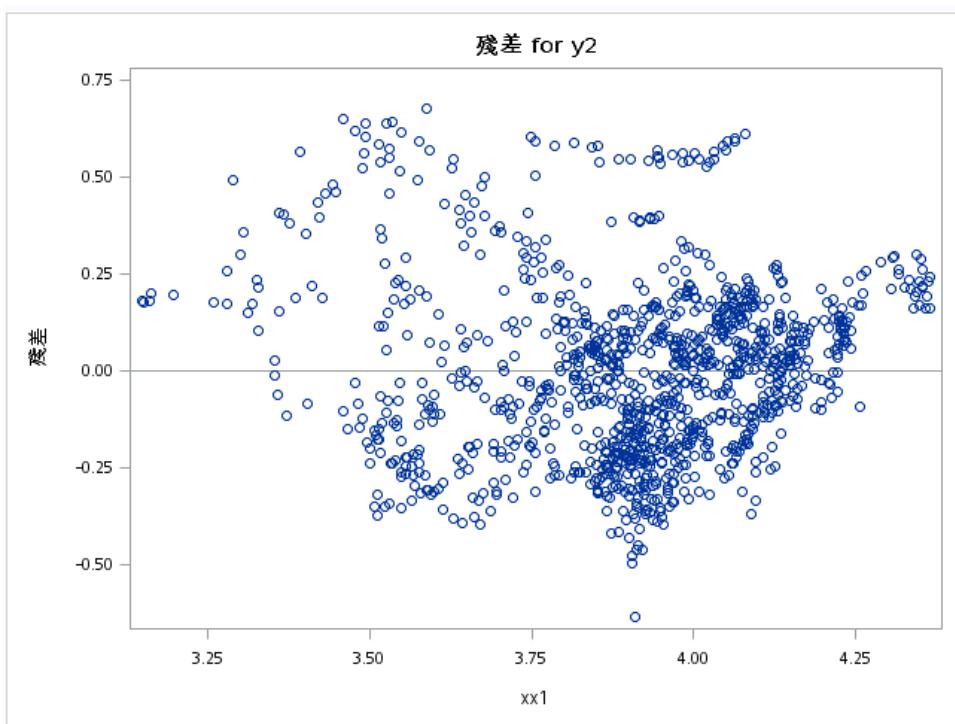
2.X 對 Y 的散佈圖

由 X 對 Y 散佈圖可看出有很多值都落在信賴區間外面，且殘差的變異數不一致情形。



3.殘差對 X 的散佈圖

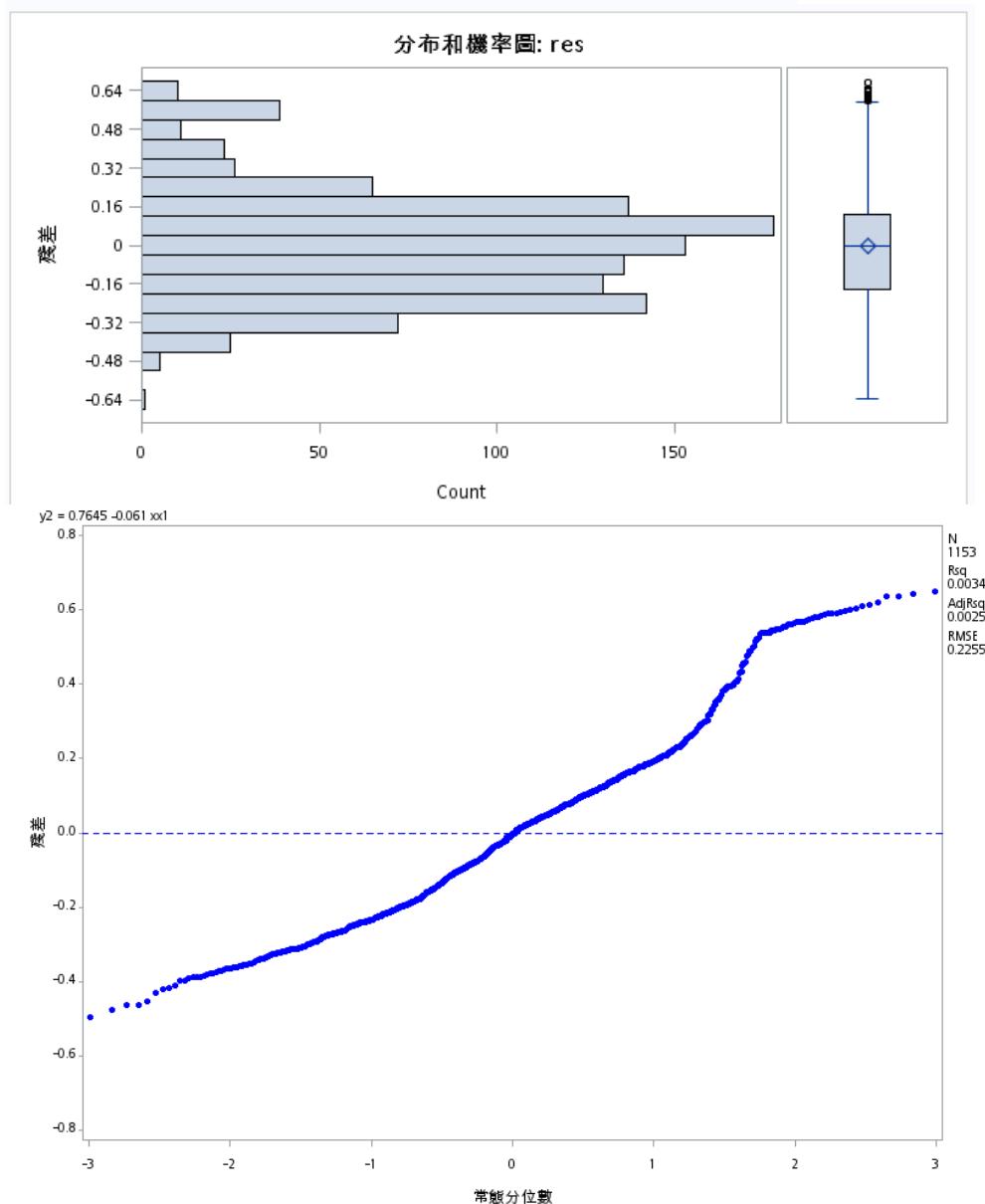
由殘差對 X 散佈圖可看出殘差的變異數不一致情形。



4. Shapiro-Wilk test 常態性檢定

轉換後檢定殘差常態性分配之檢定結果 P-value < 0.0001，拒絕 H₀，仍不符合常態分配。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.972519	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.040011	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.653973	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	6.472424	Pr > A-Sq	<0.0050



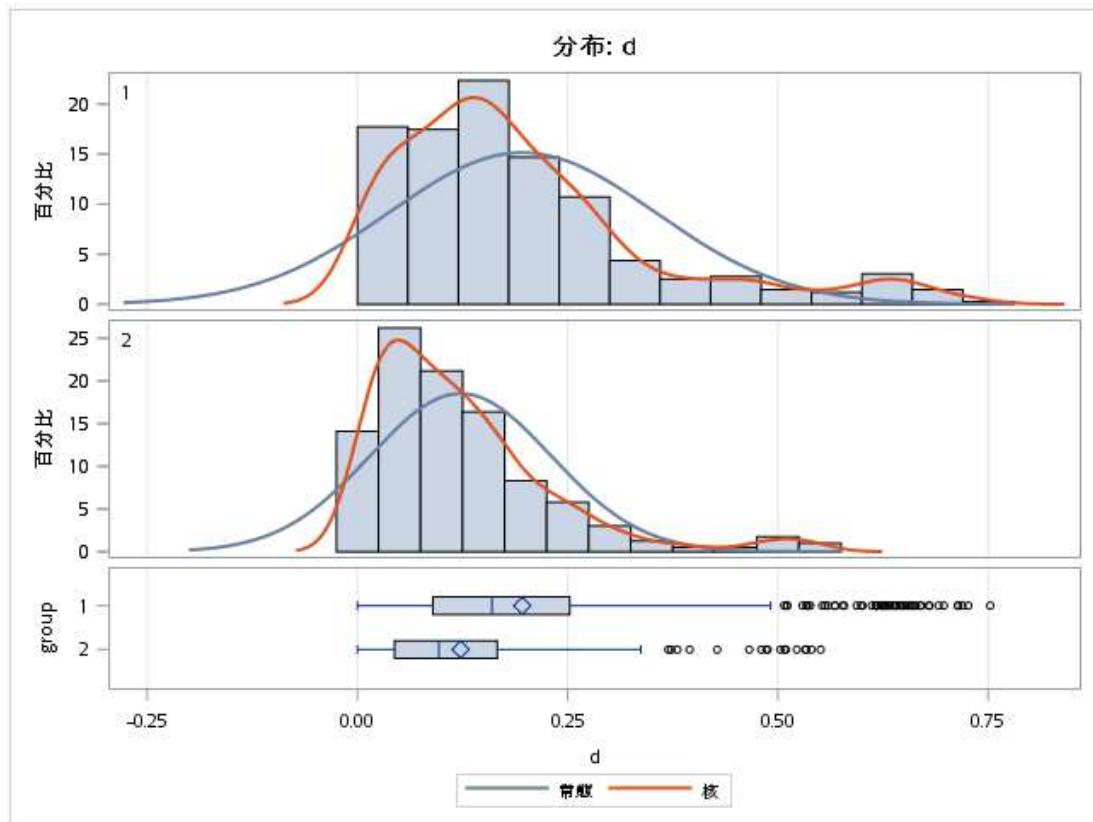
5. Brown-Forsythe test 變異數檢定

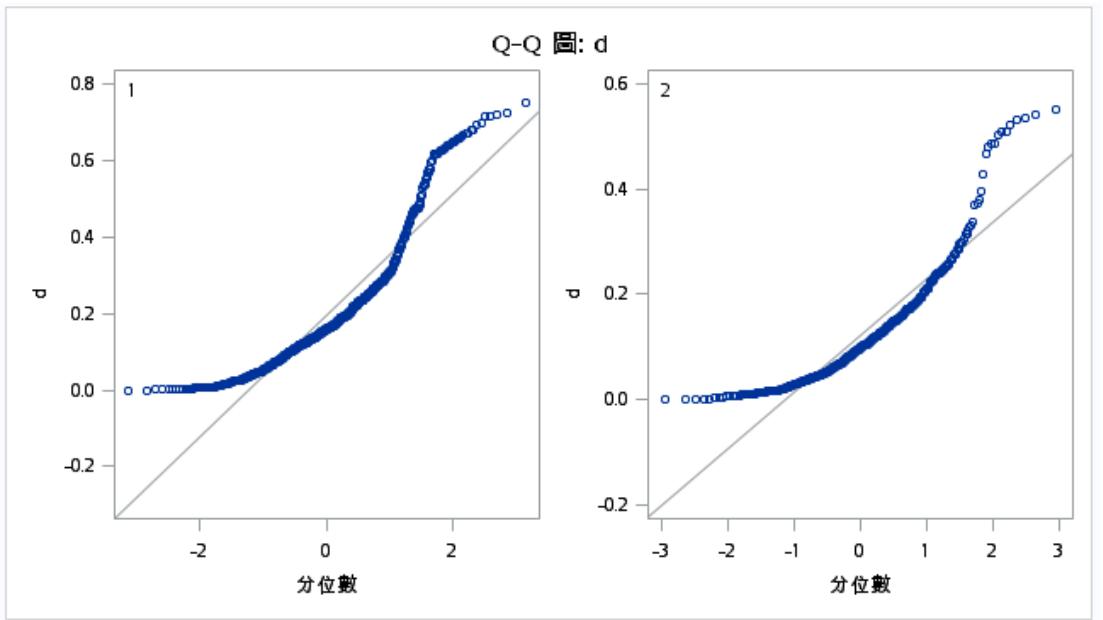
轉換後變數，利用 Brown-Forsythe test 作檢定，檢定結果 $P\text{-value} < 0.0001$ ，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的變異數不相等。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	756	0.1958	0.1581	0.00575	3.11E-6	0.7526
2	397	0.1226	0.1077	0.00541	0	0.5509
Diff (1-2)		0.0732	0.1428	0.00885		

方法	變異數	自由度	t 值	Pr > t
集區	均等	1151	8.28	<.0001
Satterthwaite	不均等	1076.1	9.28	<.0001

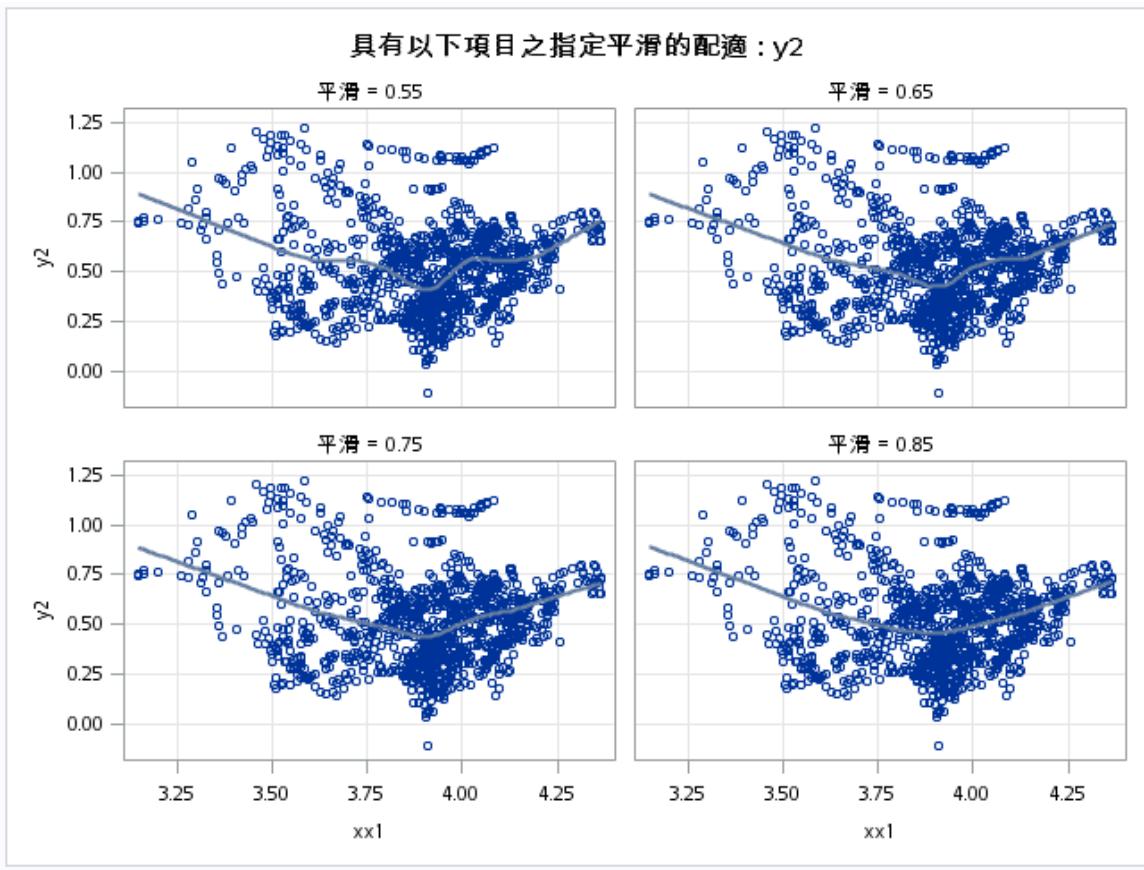
變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	755	396	2.15	<.0001





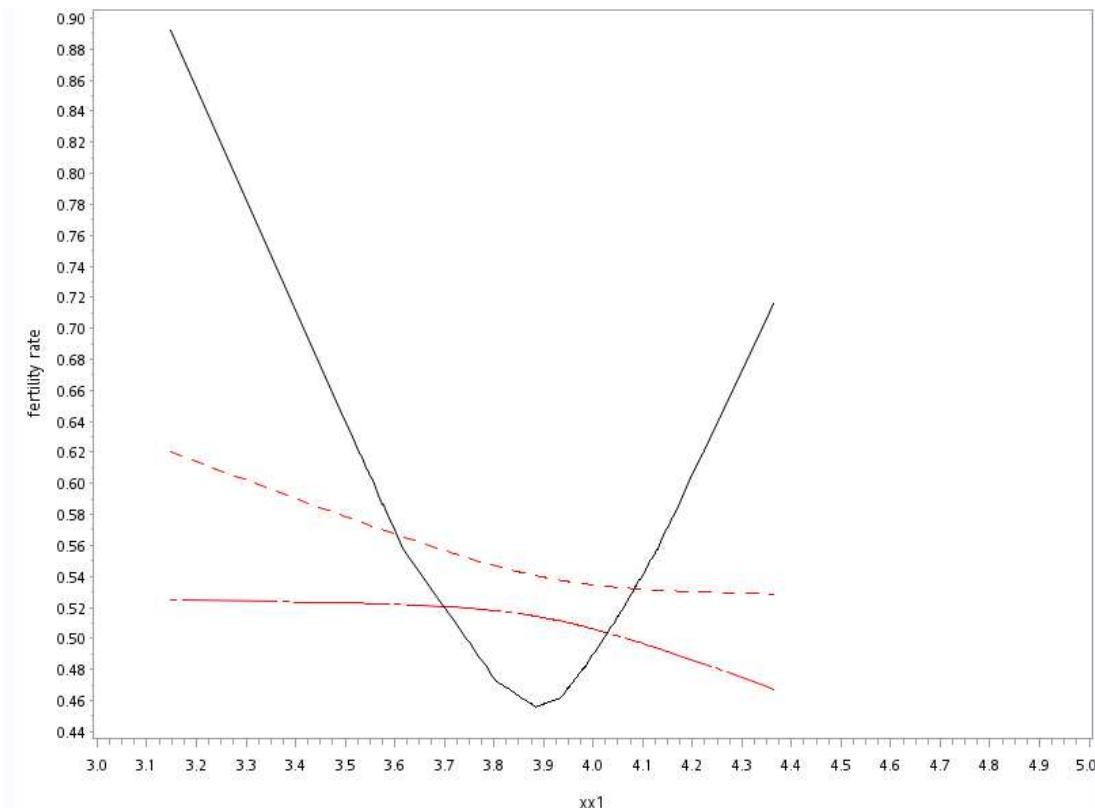
6. Lowess Curve

觀察 Lowess Curve 看趨勢線。



7. 預測線及信賴區間

預測線都沒有包含在信賴區間內，表示此模型配適不恰當。



(三) 變數轉換二(倒數轉換)模型：

1. 簡單迴歸模型：

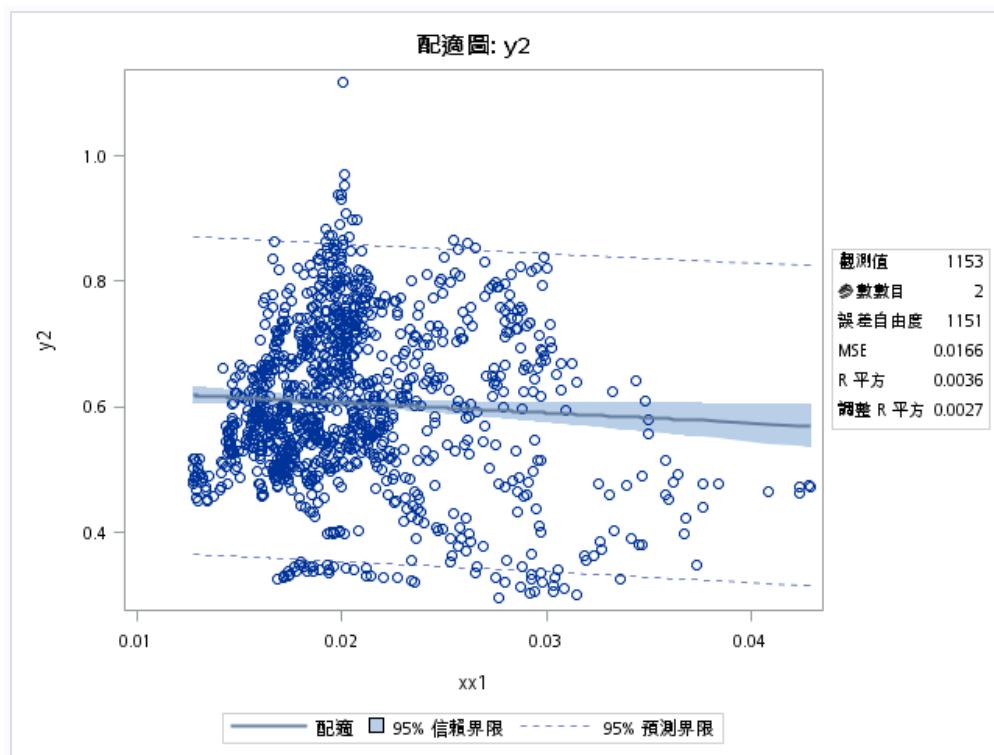
將 X 及 Y 均進行 log 轉換後，殘差仍不符合常態分配，且殘差變異數仍不為常數。故我們嘗試用 $\lambda = -1$ 進行轉換(亦即 $Y' = 1/Y$ 、 $X' = 1/X$)。得到迴歸線為 $Y' = 0.63865 - 1.60874 X'$ ，藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0 = 0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1 = 0$ 的假設。表示在統計上有證據證明 b_0 和 b_1 參數值都顯著不等於 0。

根 MSE	0.12893	R 平方	0.0036
應變平均值	0.60551	調整 R 平方	0.0027
變異係數	21.29327		

參數估計值							
變動	自由度	參數 估計值	標準 誤差	t 值	Pr > t	95% 信賴界限	
Intercept	1	0.63865	0.01667	38.31	<.0001	0.60594	0.67136
xx1	1	-1.60874	0.78810	-2.04	0.0415	-3.15502	-0.06246

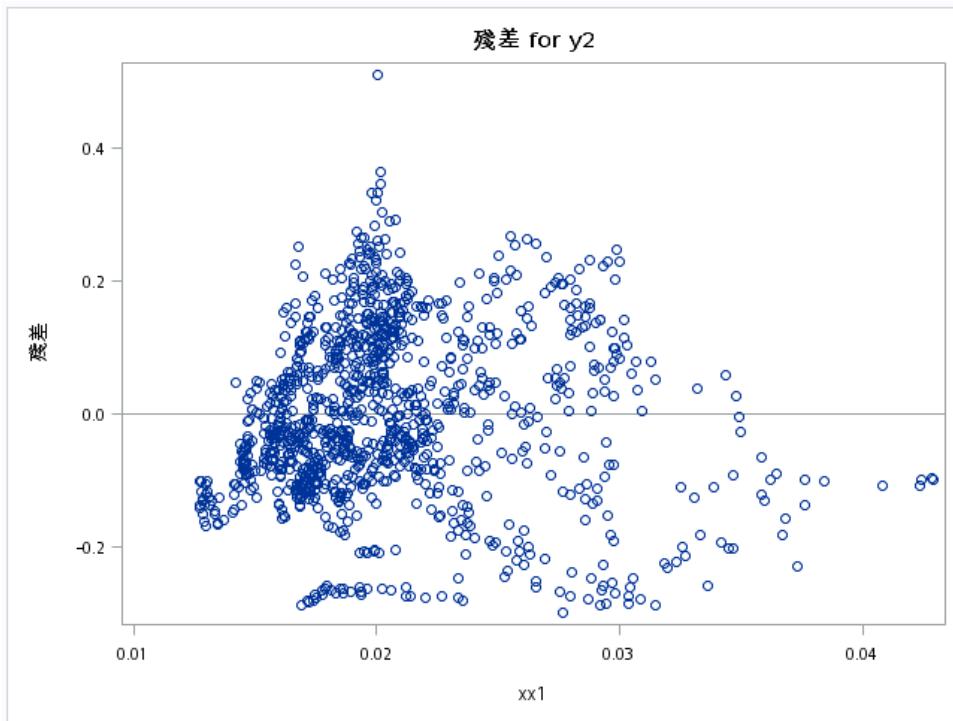
2.X 對 Y 的散佈圖

由 X 對 Y 散佈圖可看出有很多值都落在信賴區間外面，且殘差的變異數不一致情形。



3.殘差對 X 的散佈圖

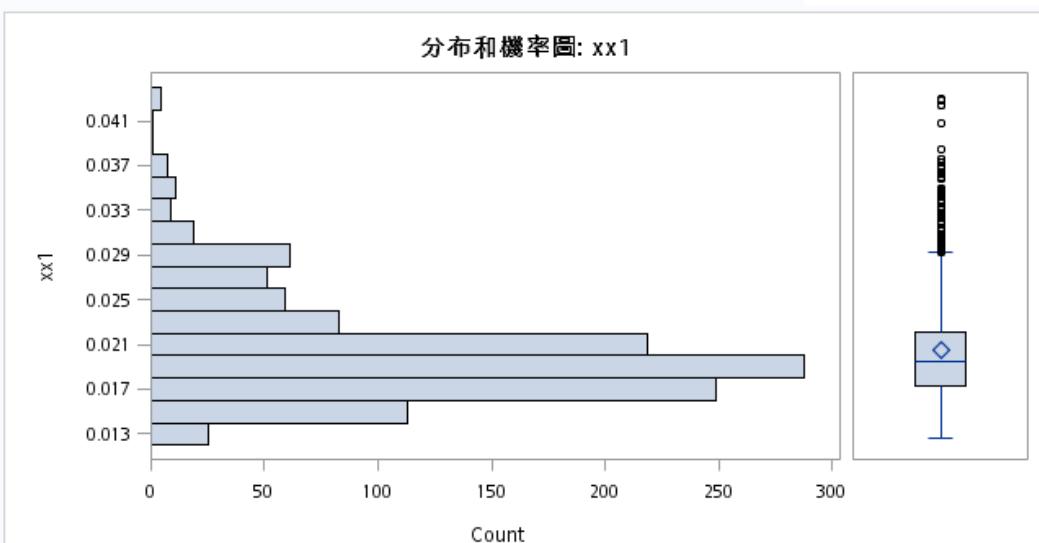
由殘差對 X 散佈圖可看出殘差的變異數不一致情形，且有極端值情形。

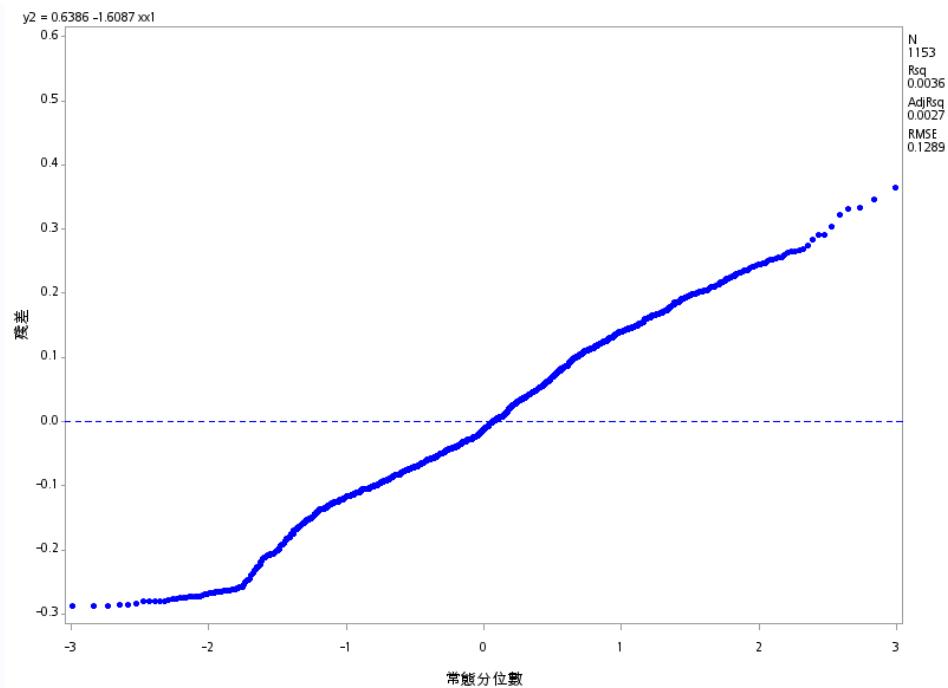


4. Shapiro-Wilk test 常態性檢定

轉換後檢定殘差常態性分配之檢定結果 P-value < 0.0001，拒絕 H0，仍不符合常態分配。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.990806	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.050299	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.537084	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.08607	Pr > A-Sq	<0.0050





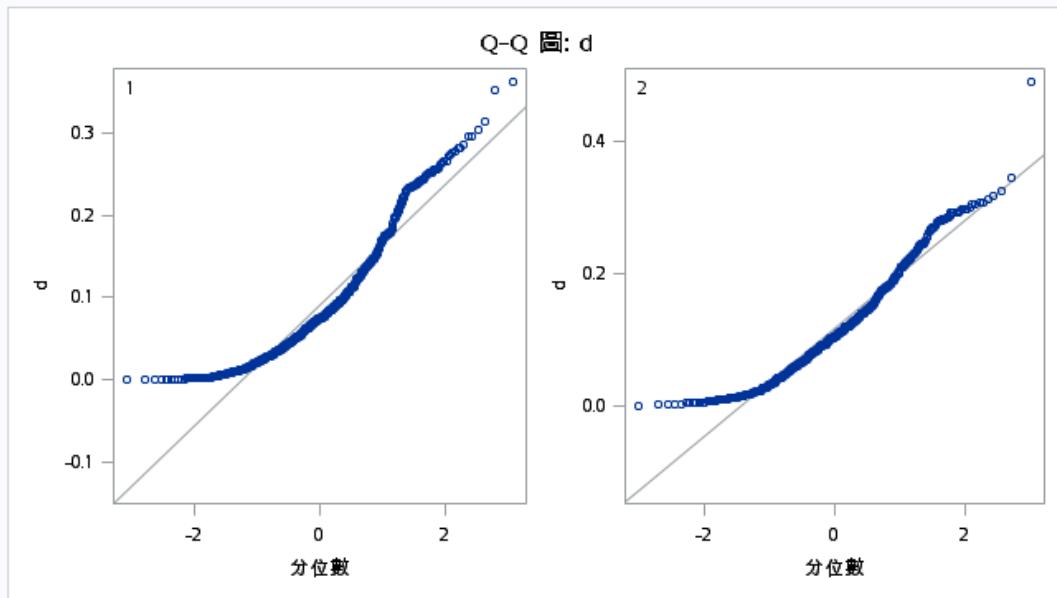
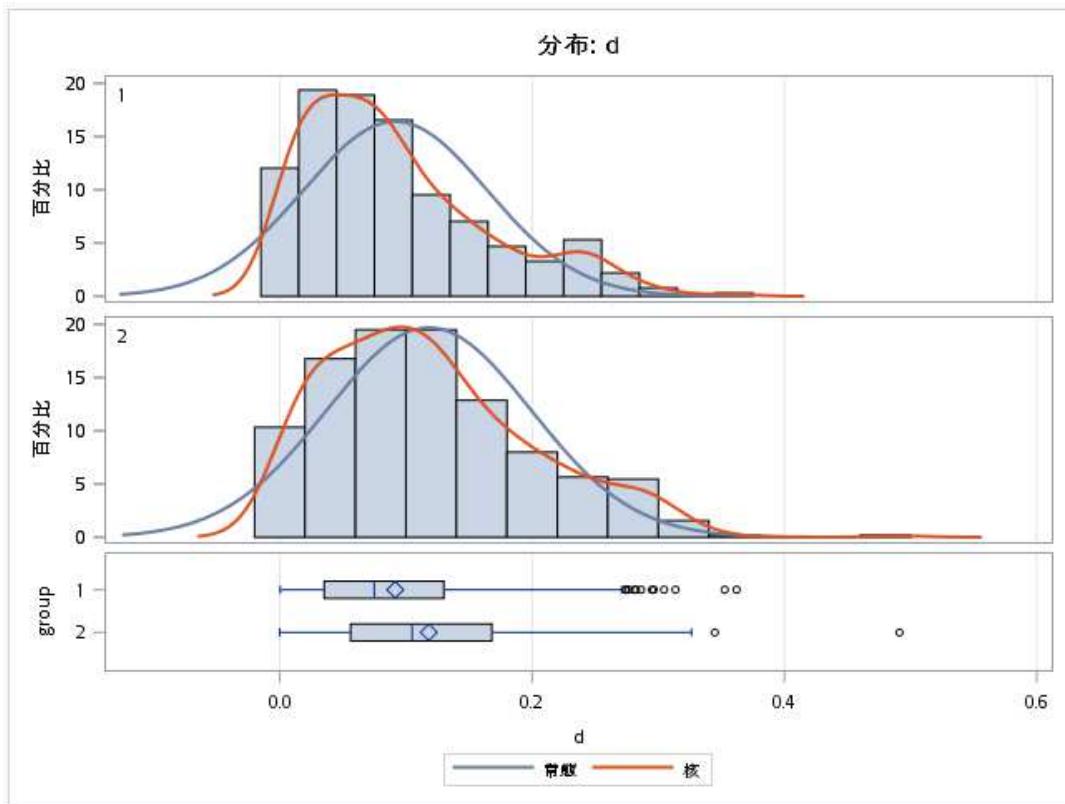
5. Brown-Forsythe test 變異數檢定

轉換後變數，利用 Brown-Forsythe test 作檢定，檢定結果 $P\text{-value} < 0.0001$ ，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的變異數不相等。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	640	0.0914	0.0729	0.00288	0.000296	0.3620
2	513	0.1181	0.0810	0.00358	0	0.4911
Diff (1-2)		-0.0266	0.0766	0.00454		

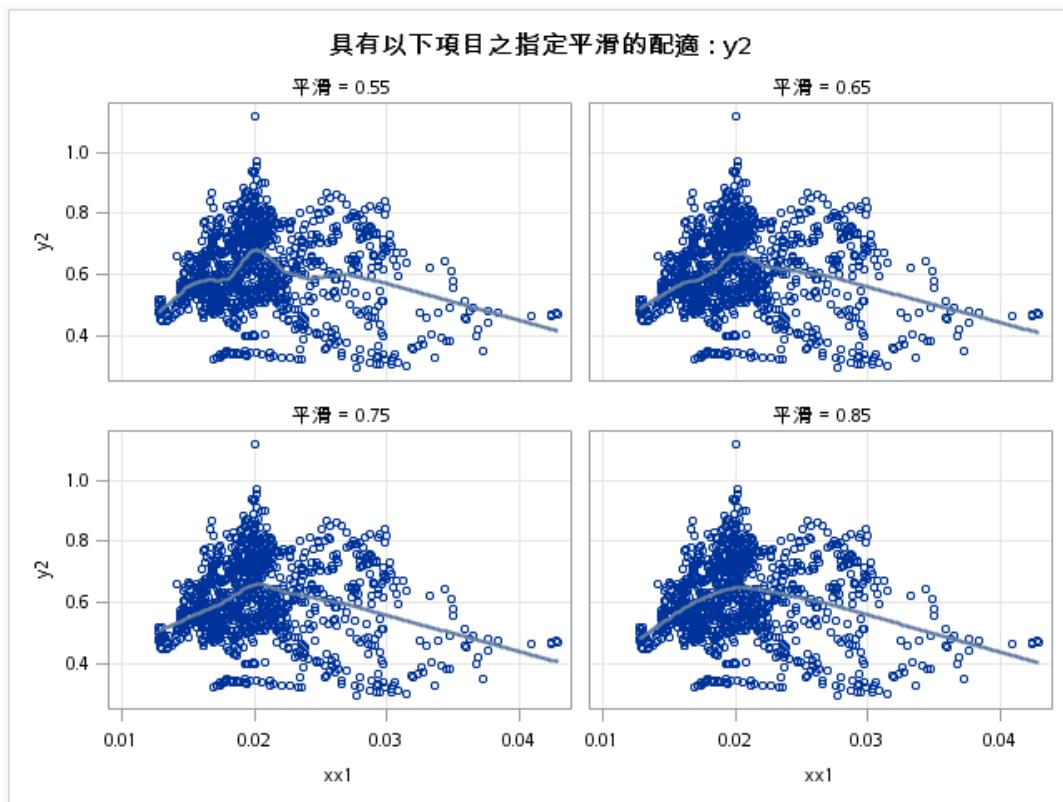
方法	變異動	自由度	t 值	Pr > t
集區	均等	1151	-5.86	<.0001
Satterthwaite	不均等	1040.9	-5.79	<.0001

變異動相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	512	639	1.24	0.0115



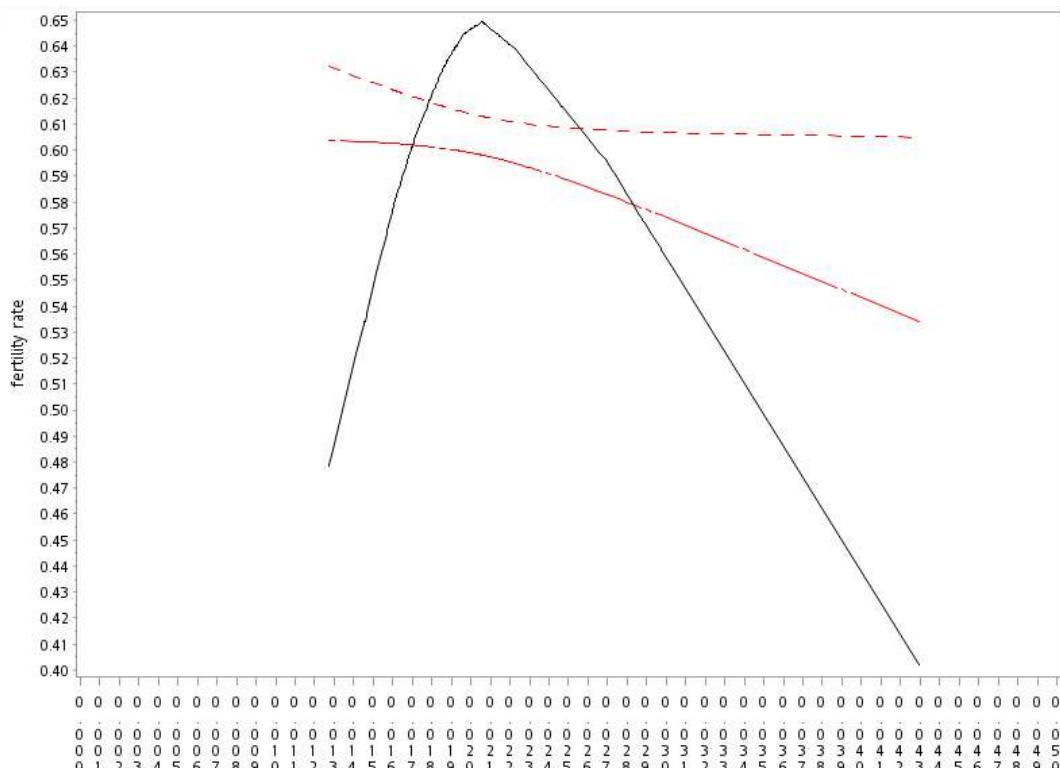
6. Lowess Curve

觀察 Lowess Curve 看趨勢線。



7. 預測線及信賴區間

預測線都沒有包含在信賴區間內，表示此模型配適仍不恰當。



(四) 變數轉換二(倒數轉換)後刪除離群值模型：

1. 簡單迴歸模型：

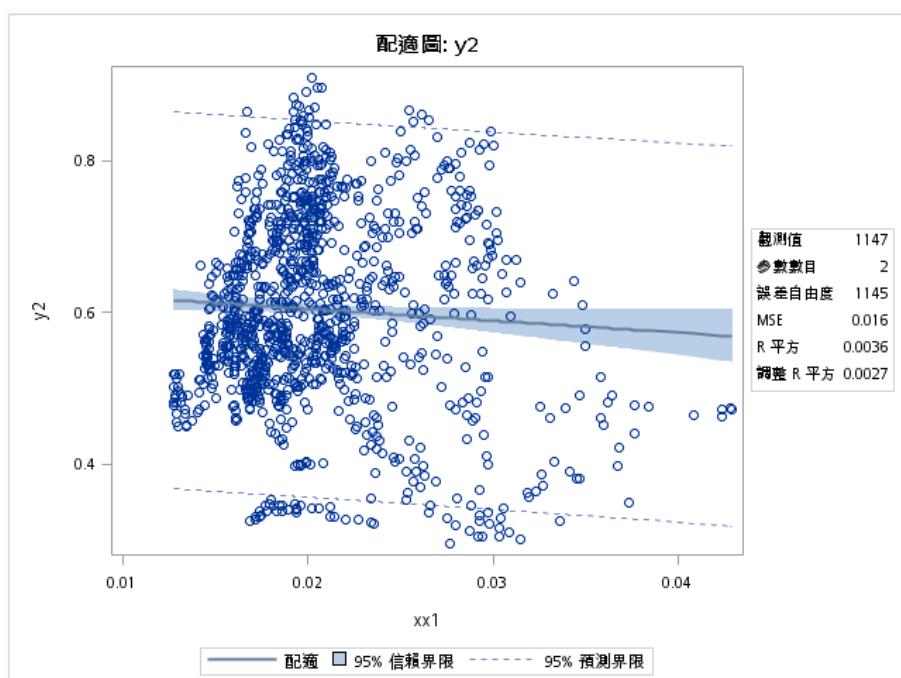
因為 X 及 Y 變數轉換後仍未達到最適的迴歸線，因此我們刪除離群值，亦即刪除半 T 化殘差絕對值 $\left| \frac{e}{\sqrt{MSE}} \right|$ 大於 2.5 的值，共 6 筆資料。殘差仍不符合常態分配，但殘差圖看來很接近常態分配，且殘差變異數仍然不為常數。得到迴歸線為 $Y' = 0.63574 - 1.56114 X'$ ，表示每增加 1% 女性勞動力參與率，總生育率減少 15.6114%。藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0=0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1=0$ 的假設。表示在統計上有證據證明 b_0 和 b_1 參數值都顯著不等於 0。

根 MSE	0.12639	R 平方	0.0036
應變平均值	0.60358	調整 R 平方	0.0027
變異係數	20.94002		

參數估計值					
變動	自由度	參數 估計值	標準 誤差	t 值	Pr > t
Intercept	1	0.63574	0.01635	38.89	<.0001
xx1	1	-1.56114	0.77259	-2.02	0.0435

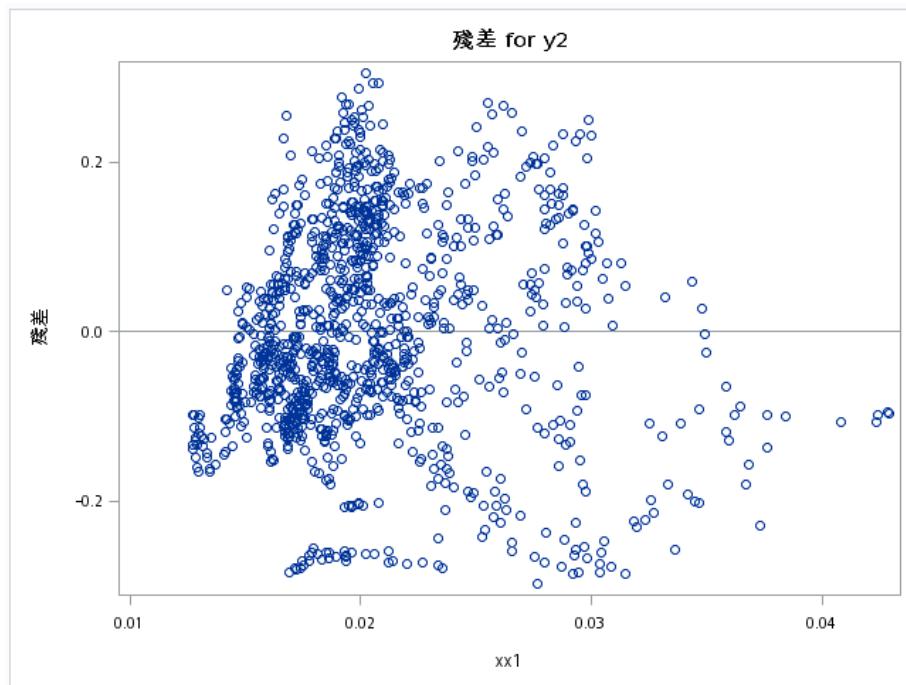
2. X 對 Y 的散佈圖

由 X 對 Y 散佈圖可看出有很多值都落在信賴區間外面，且殘差的變異數不一致情形。



3. 殘差對 X 的散佈圖

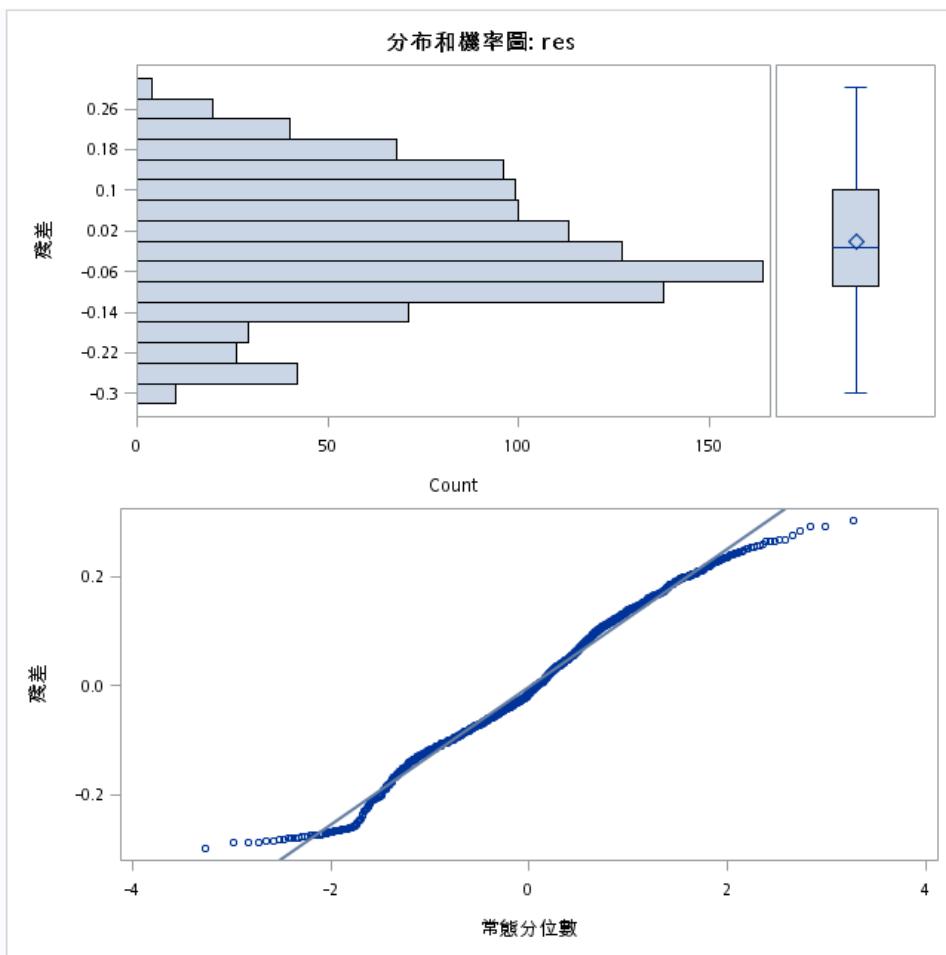
由殘差對 X 散佈圖可看出殘差的變異數不一致情形。



4. Shapiro-Wilk test 常態性檢定

轉換後檢定殘差常態性分配之檢定結果 P-value < 0.0001，拒絕 H₀，仍不符合常態分配。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.987981	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.048254	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.555276	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.426415	Pr > A-Sq	<0.0050



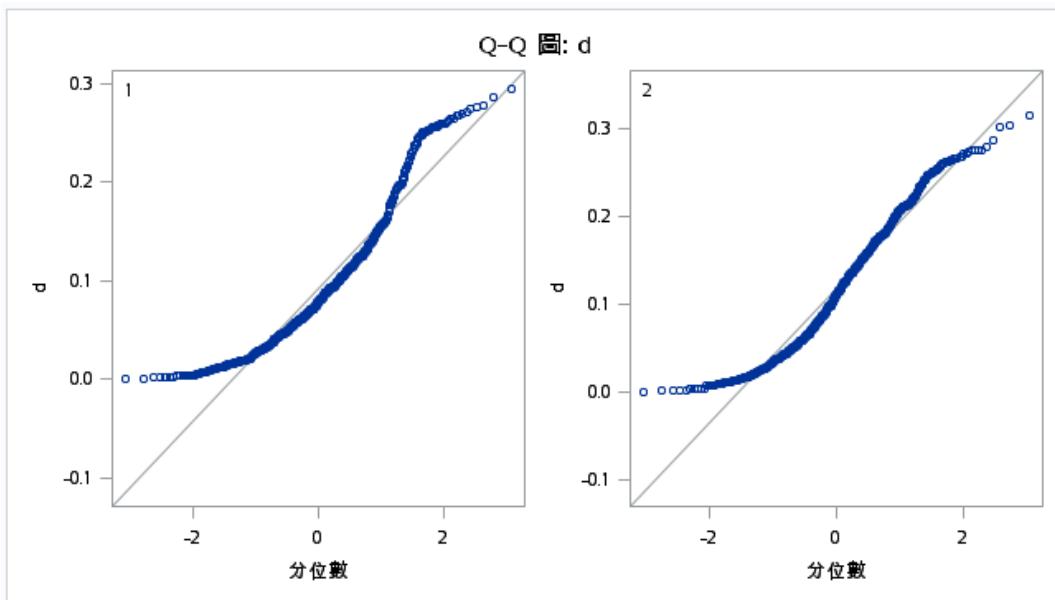
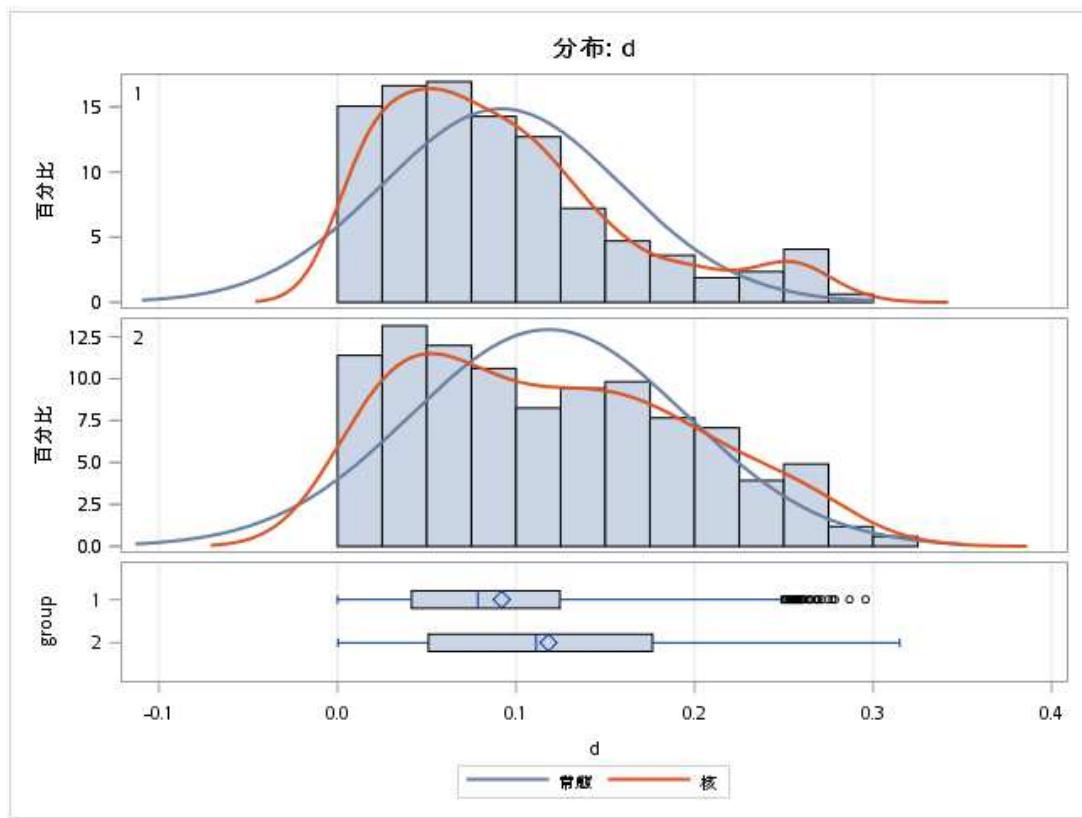
5. Brown-Forsythe test 變異數檢定

轉換後變數，利用 Brown-Forsythe test 作檢定，檢定結果 $P\text{-value} < 0.0001$ ，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的變異數不相等。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	638	0.0920	0.0672	0.00266	0	0.2958
2	509	0.1182	0.0771	0.00342	0.000246	0.3149
Diff (1-2)		-0.0262	0.0718	0.00426		

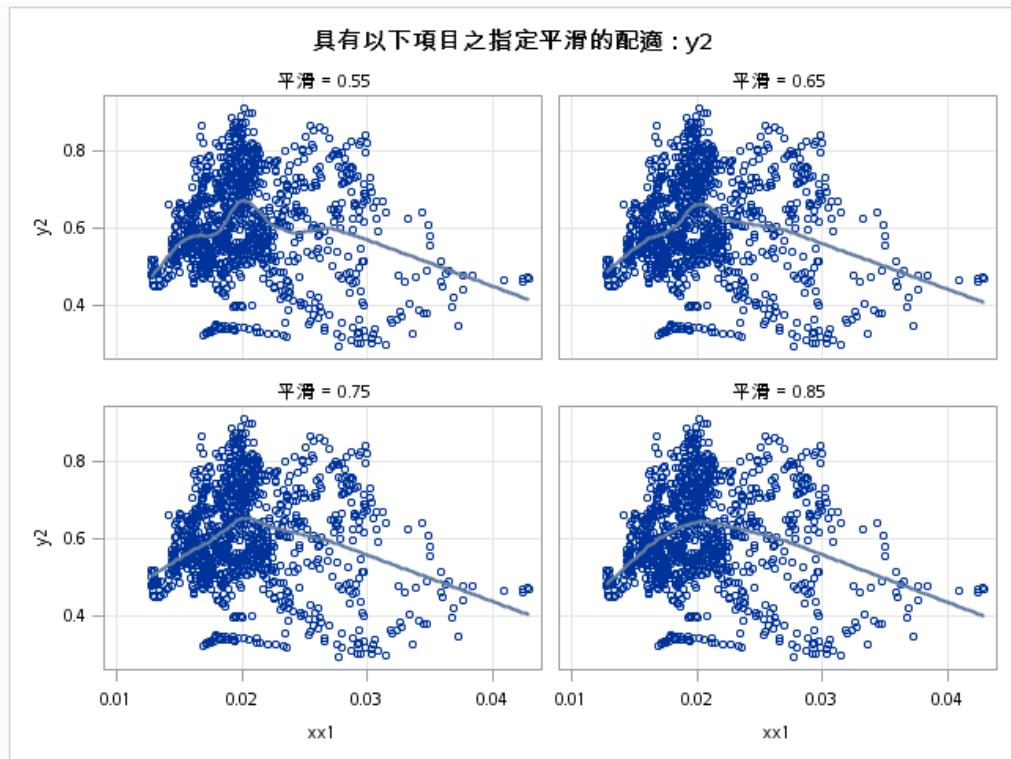
方法	變異數	自由度	t 值	Pr > t
集區	均等	1145	-6.14	<.0001
Satterthwaite	不均等	1013.2	-6.04	<.0001

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	508	637	1.32	0.0010



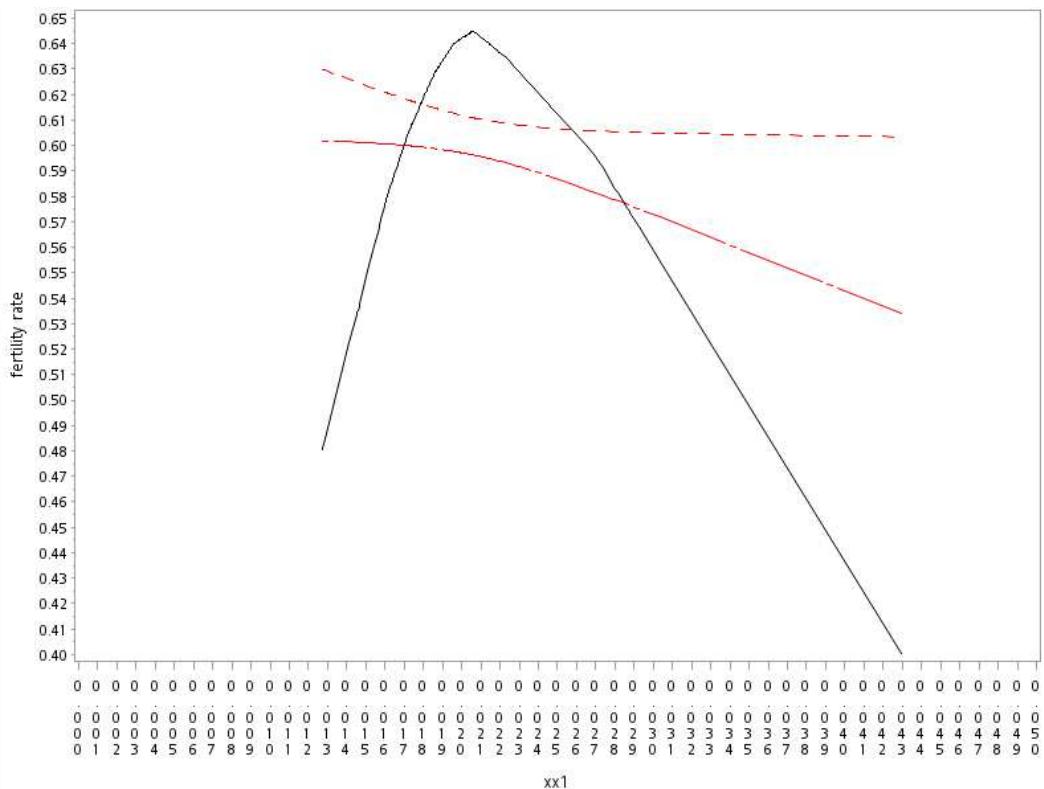
6. Lowess Curve

觀察 Lowess Curve 看趨勢線。



7. 預測線及信賴區間

預測線都沒有包含在信賴區間內，表示此模型配適仍不恰當。



第二節、總生育率和扶幼比之簡單迴歸模型

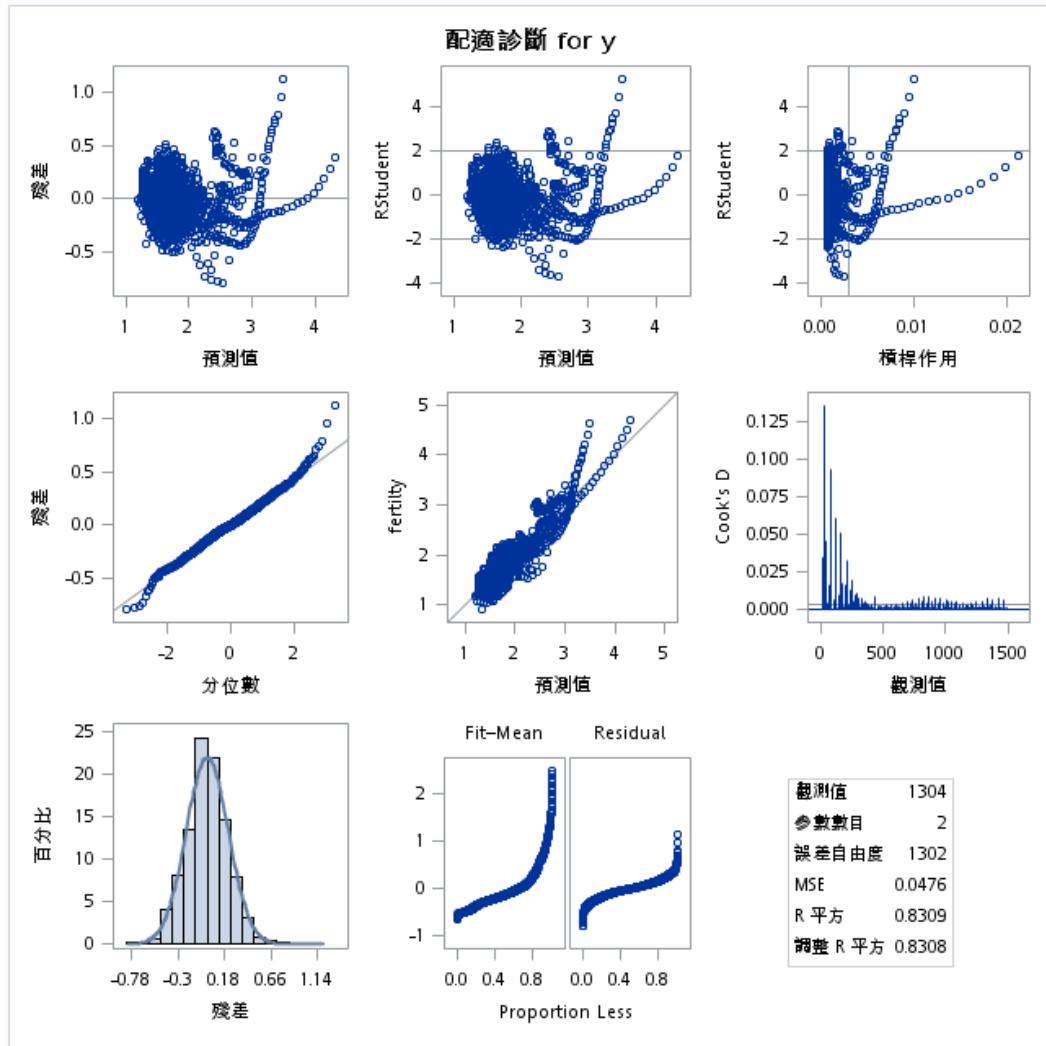
(一)原始模型：

1. 簡單線性迴歸模型

將反應變數 Y 設為 1980 年到 2015 年 OECD 各國及我國總生育率(%)，獨立變數(解釋變數)X 設為 1980 年到 2015 年 OECD 各國及我國扶幼比(%)，扣除遺漏值後，共有 1,304 筆資料，進行簡單線性迴歸模型統計分析。得到迴歸線為 $Y=0.35819+0.04535X$ ，表示每增加 1% 扶幼比，總生育率增加 0.4535%。另藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0=0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1=0$ 的假設。表示 b_0 和 b_1 參數值都顯著不等於 0。若 $X=50$ ，代入公式， $Y=2.62569$ 。因 X 及 Y 均符合常態性分配，可藉由 Pearson 相關係數 0.91155 知道兩者為高度正相關，且 P-value 為 < 0.0001，拒絕 H_0 ，表示統計上有足夠證據證明兩者相關係數不為 0。

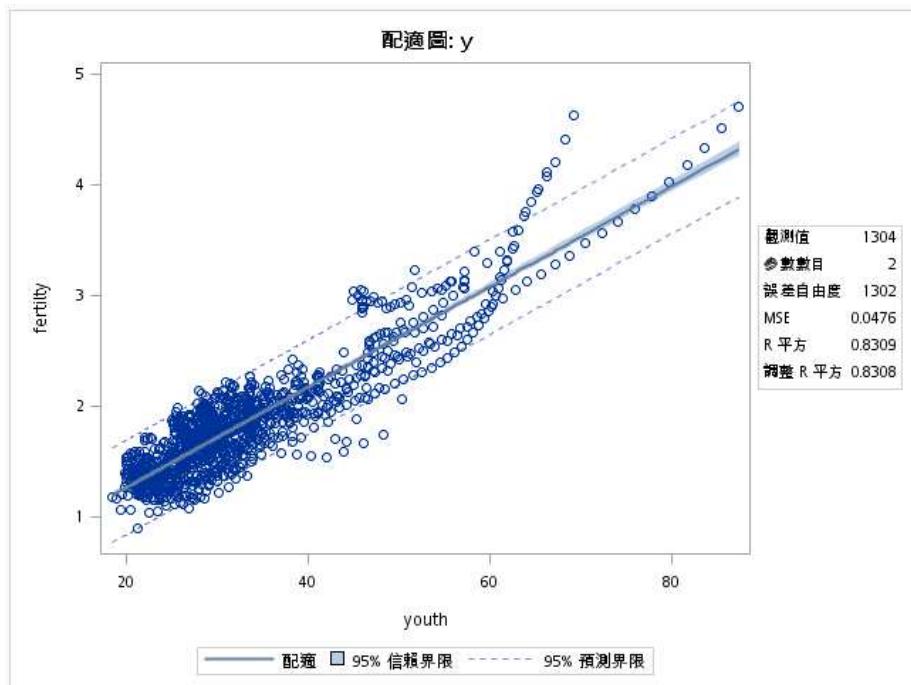
REG 程序		變異數分析					
		來源	自由度	平方和	平均值 平方	F 值	Pr > F
模型		模型	1	304.31952	304.31952	6398.25	<.0001
誤差		誤差	1302	61.92689	0.04756		
含有遺漏值的觀測值數目	280	已校正的總計	1303	366.24641			

參數估計值							
			參數 估計值		標準 誤差	t 值	Pr > t
根 MSE	0.21809	R 平方	0.8309				
應變平均值	1.82572	調整 R 平方	0.8308	Intercept	Intercept	1 0.35819	0.01932 18.54 <.0001
變異係數	11.94537			x9	youth	1 0.04535	0.00056700 79.99 <.0001



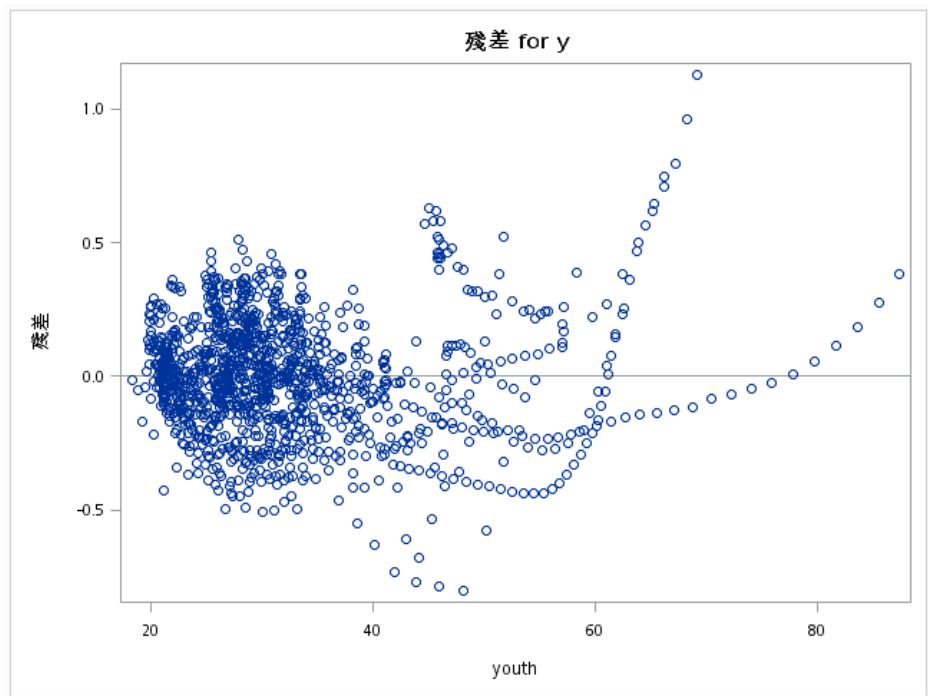
2. X 對 Y 散佈圖

由 X 對 Y 散佈圖可看出有很多值落在信賴區間外面，且殘差的變異數不一致情形。



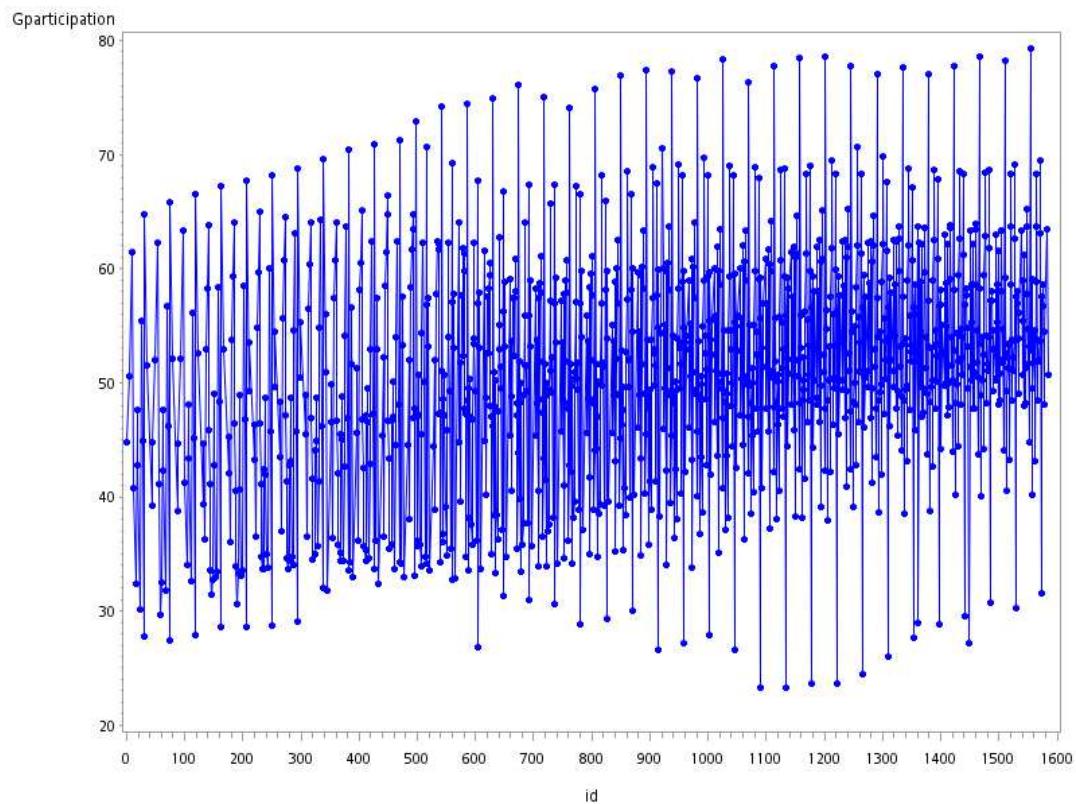
3. X 對殘差的散佈圖

由 X 對殘差的散佈圖可看出有部分極端值，且殘差的變異數不一致情形。



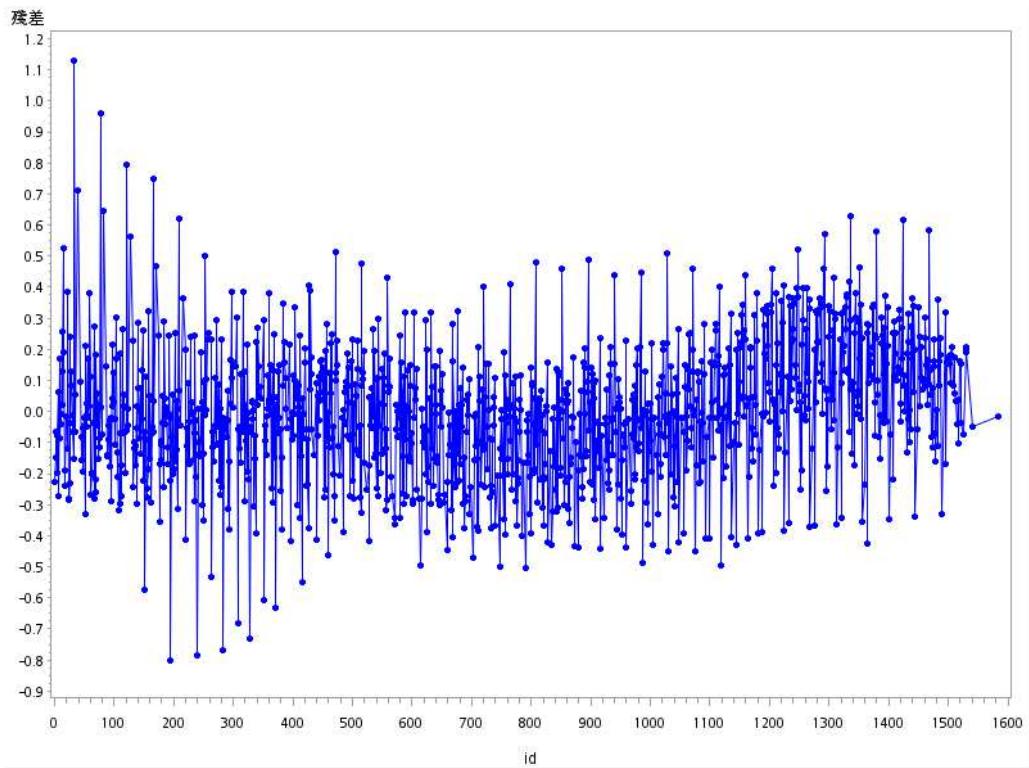
4.X 的時間序列圖

由 X 的時間序列圖可知，X 為隨機連續變數。



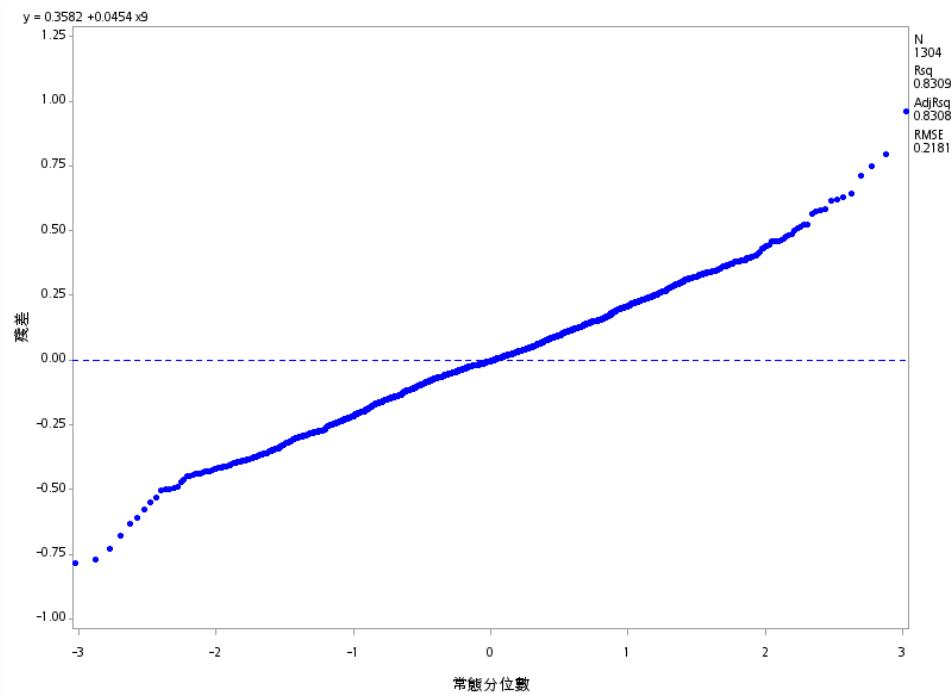
5. 殘差的時間序列圖

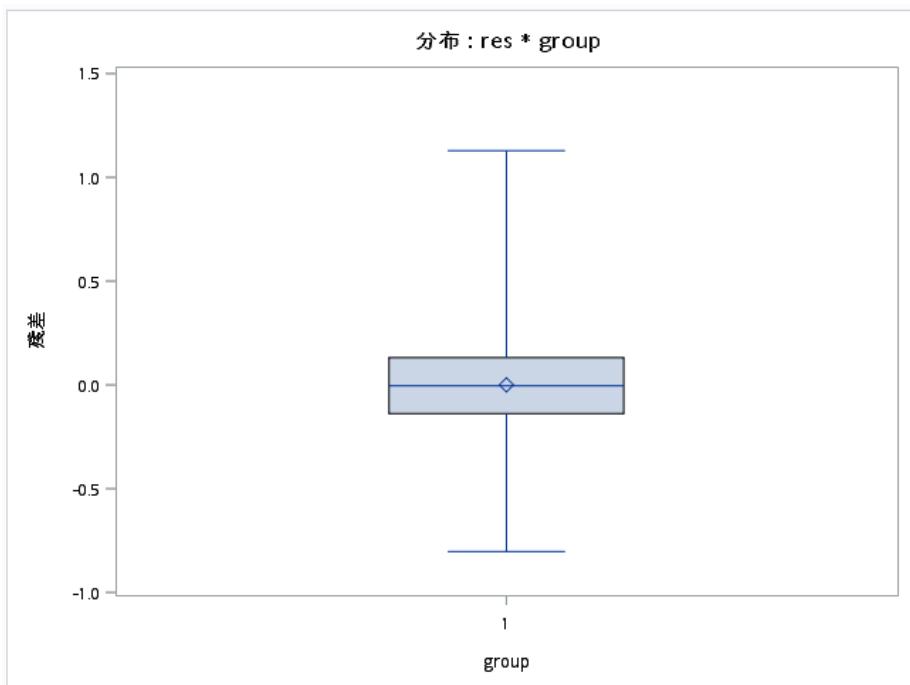
由殘差的時間序列圖可知，殘差為隨機連續變數。



6. 殘差的常態機率圖、箱型圖及與殘差期望值之相關性檢定

由常態機率圖可知殘差接近常態分配，且檢定結果為兩者有顯著相關，Pearson 相關係數為 0.978，Spearman 相關係數為 1.00。





Pearson 相關係數		
Prob > r (位於 H0 底下): Rho=0 觀測值數目		
	res	expec
res	1.00000	0.97802
殘差	1304	<.0001 1304
expec	0.97802 <.0001 1304	1.00000 1584

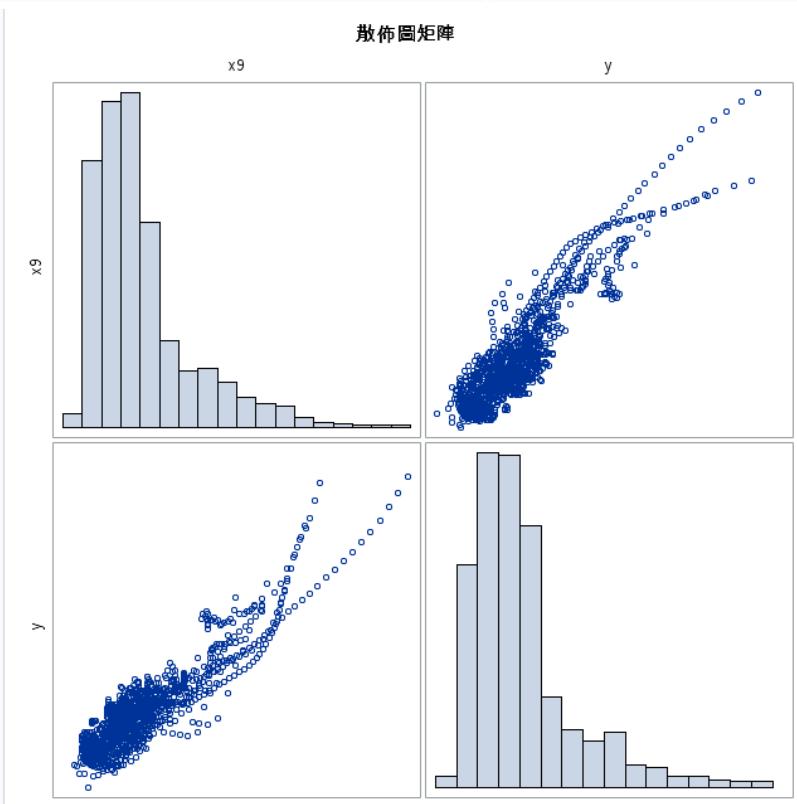
Spearman 相關係數		
Prob > r (位於 H0 底下): Rho=0 觀測值數目		
	res	expec
res	1.00000	1.00000
殘差	1304	<.0001 1304
expec	1.00000 <.0001 1304	1.00000 1584

7.X 與 Y 之相關性檢定

因為 X 及 Y 不符合常態分配，可藉由 Spearman 相關係數 0.836 知道兩者為正相關，且 P-value < 0.001，拒絕 H0，表示統計上有足夠證據證明兩者相關係數不為 0。

簡單統計值							
變動	N	平均值	標準差	中位數	最小值	最大值	標籤
x	1200	51.01403	10.26161	51.25263	23.29849	79.28249	participation
y	1530	1.90968	0.63469	1.76000	0.89500	4.82800	fertility

Pearson 相關係數			Spearman 相關係數		
Prob > r (位於 H0 底下): Rho=0 觀測值數目			Prob > r (位於 H0 底下): Rho=0 觀測值數目		
	x9	y		x9	y
x9	1.00000	0.91155	x9	1.00000	0.83614
youth		<.0001	youth		<.0001
	1309	1304		1309	1304
y	0.91155	1.00000	y	0.83614	1.00000
fertility	<.0001		fertility	<.0001	
	1304	1530		1304	1530



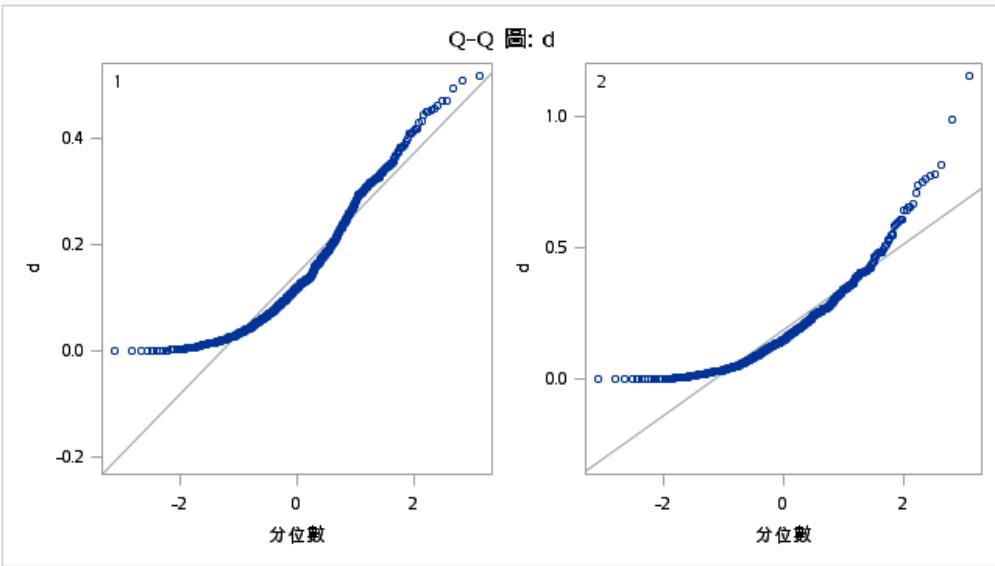
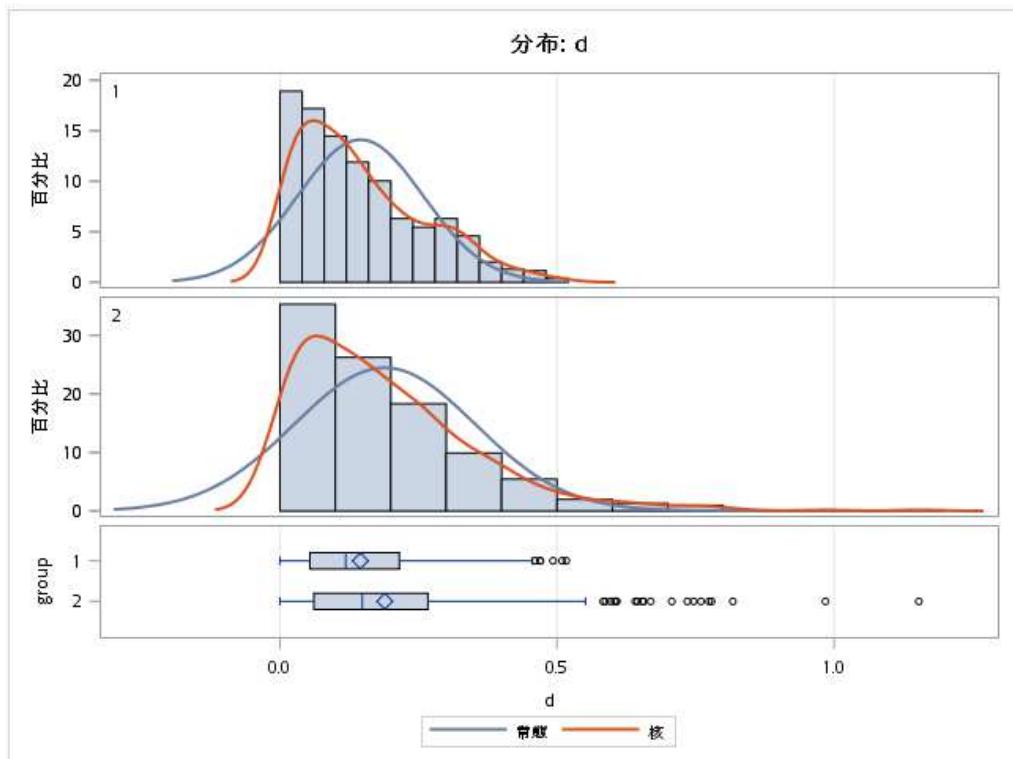
8. Brown-Forsythe test 變異數檢定

利用 Brown-Forsythe test 作檢定，檢定結果 P-value < 0.0001，拒絕 H0，表示統計上有足夠的證據證明殘差的變異數不相等。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	698	0.1451	0.1131	0.00428	0.000048	0.5166
2	606	0.1891	0.1631	0.00663	0.000048	1.1527
Diff (1-2)		-0.0440	0.1386	0.00769		

方法	變異數	自由度	t 值	Pr > t
集區	均等	1302	-5.72	<.0001
Satterthwaite	不均等	1055.7	-5.58	<.0001

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	605	697	2.08	<.0001



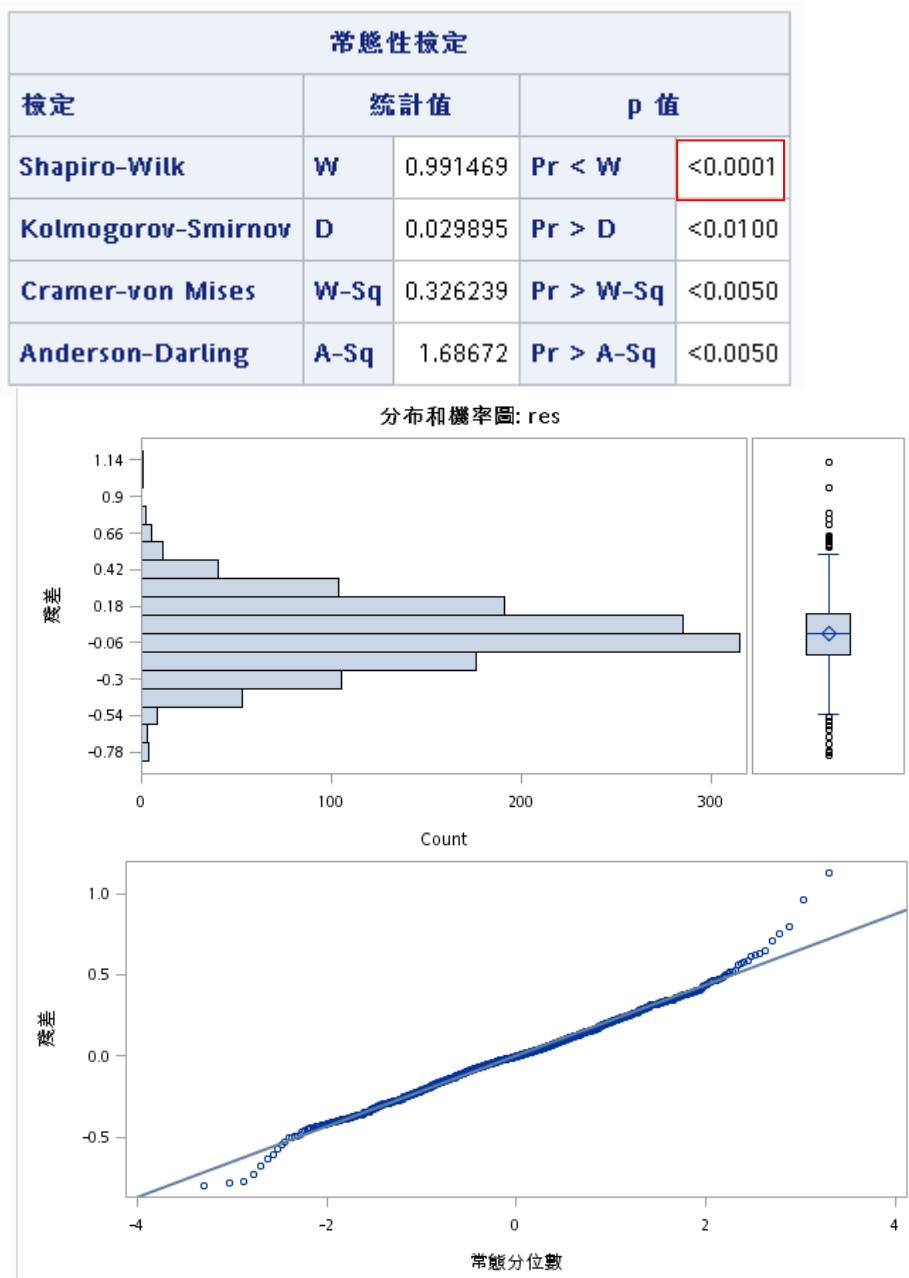
9. Brown-Pagan test 變異數檢定

利用 Brown-Pagan test 作檢定，檢定結果 $X_{BP}^2 = 289 > c$ ，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的變異數不相等。

Obs	ssrs	sse	nobs	tests	pr
1	0.88439	61.9269	1584	289.311	1

10. Shapiro-Wilk test 常態性檢定

利用 Shapiro-Wilk 對殘差作常態性檢定，檢定結果 P-value < 0.0001，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的分布不符合常態分配。

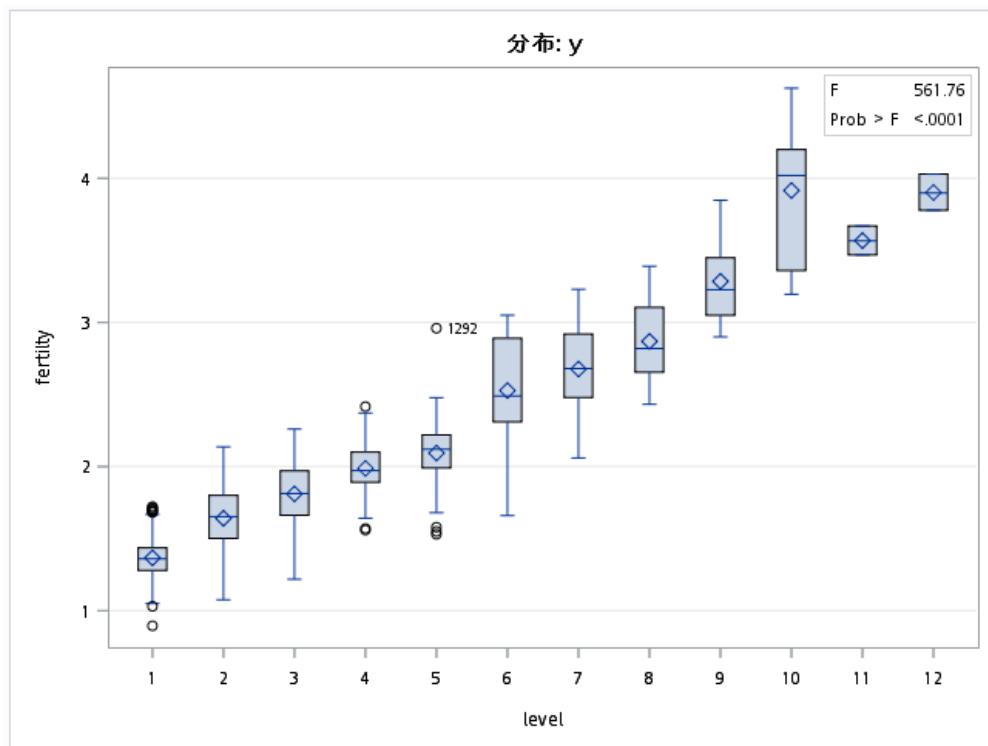


11. F Test for Lack of Fit 缺適度檢定

因為 X9 沒有重覆觀察值，故只好將附近觀察值是為同一群，進行 F Test for Lack of Fit。因此我們假設 X9 大於等於 20 小於 25 設為第一組，X9 大於等於 25 小於 30 設為第二組，X9 大於等於 30 小於 35 設為第三組，以此類推到，X9 大於等於 75 小於 80 設為第十二組，總共分為 12 群組。檢定結果 P-value < 0.0001，拒絕 H0，表示統計上有足夠的證據證明此迴歸模型不適用於簡單線性迴歸模型。

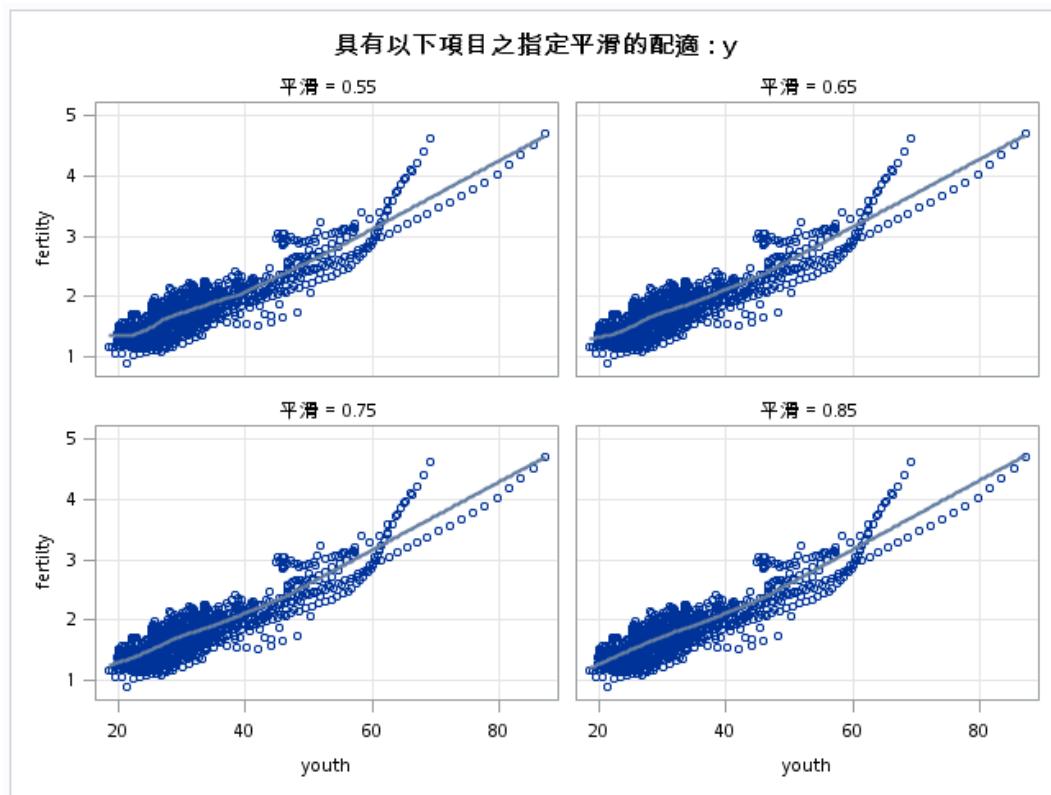
來源	自由度	平方和	均方	F 值	Pr > F
模型	11	278.1650798	25.2877345	561.76	<.0001
誤差	1276	57.4398152	0.0450155		
已校正的總計	1287	335.6048950			

來源	自由度	ANOVA SS	均方	F 值	Pr > F
level	11	278.1650798	25.2877345	561.76	<.0001



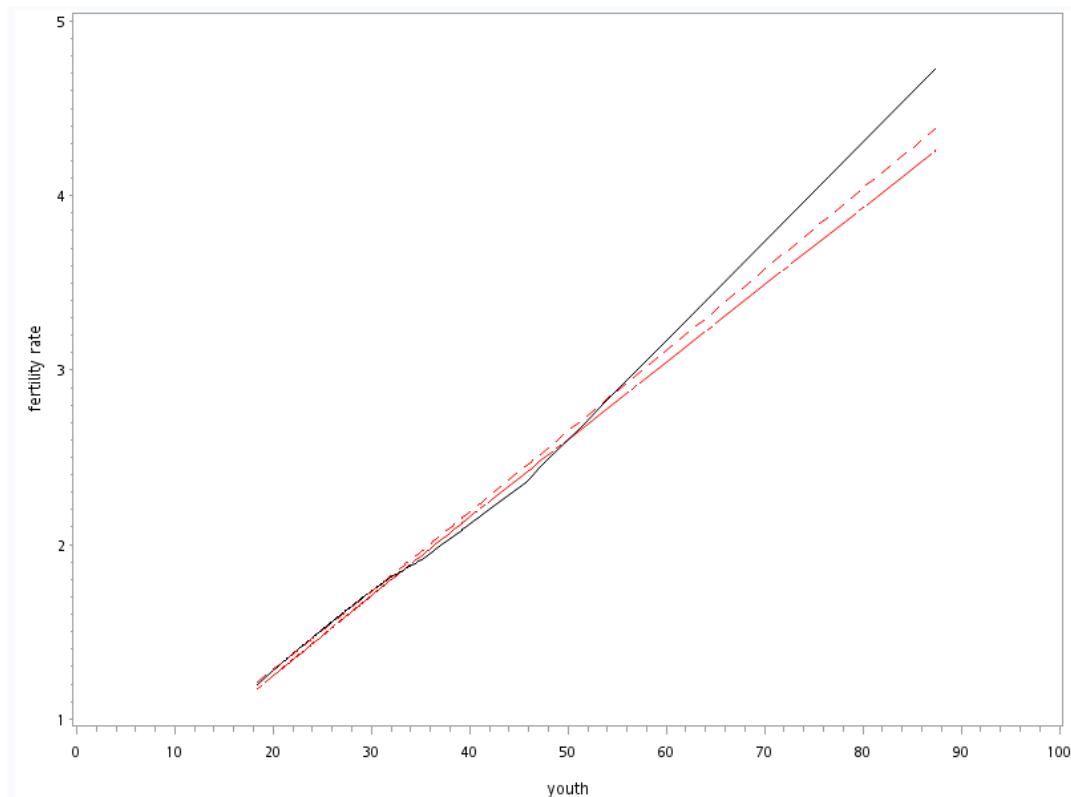
12. Lowess Curve

觀察 Lowess Curve 看趨勢線。



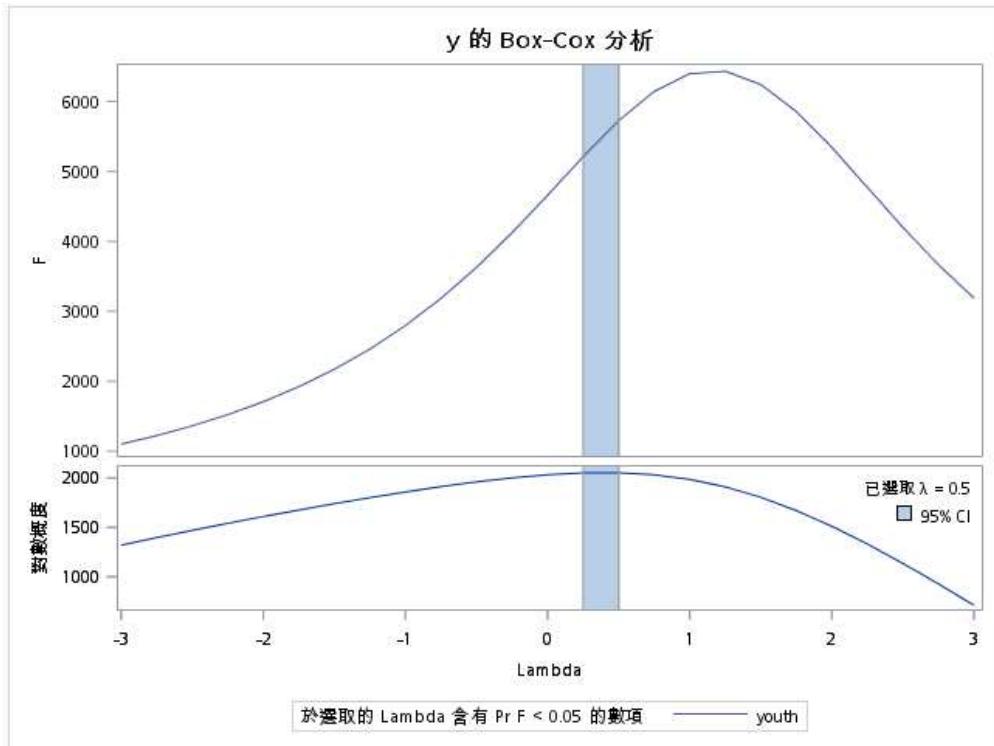
13. 預測線及信賴區間

預測線都沒有包含在信賴區間內，表示此模型配適不恰當。



14. Box-Cox 轉換

綜觀以上，我們發現此迴歸模型殘差不符合常態分配且殘差變異數不為常數值，且由 F test for Lack of Fit 檢定可知，此迴歸模型不適用於簡單線性迴歸。故我們要進行變數轉換，因此利用 Box-Cox 找出最適合的轉換值。



根 MSE	0.15571	R 平方	0.8155
應變平均值	0.67800	調整 R 平	0.8153
變異係數	22.96572	Lambda	0.5000

根據一般自由度的調整多變量 ANOVA 表格					
應變數計分參數=6 S=1 M=2 N=650					
統計值	值	F 值	分子自由度	分母自由度	p
Wilks' Lambda	0.184536	958.92	6	1302	<.0001
Pillai 的追蹤	0.815464	958.92	6	1302	<.0001
Hotelling-Lawley 追蹤	4.419002	958.92	6	1302	<.0001
Roy 最大根	4.419002	958.92	6	1302	<.0001

(二) 變數轉換(根號轉換)模型：

1. 簡單迴歸模型：

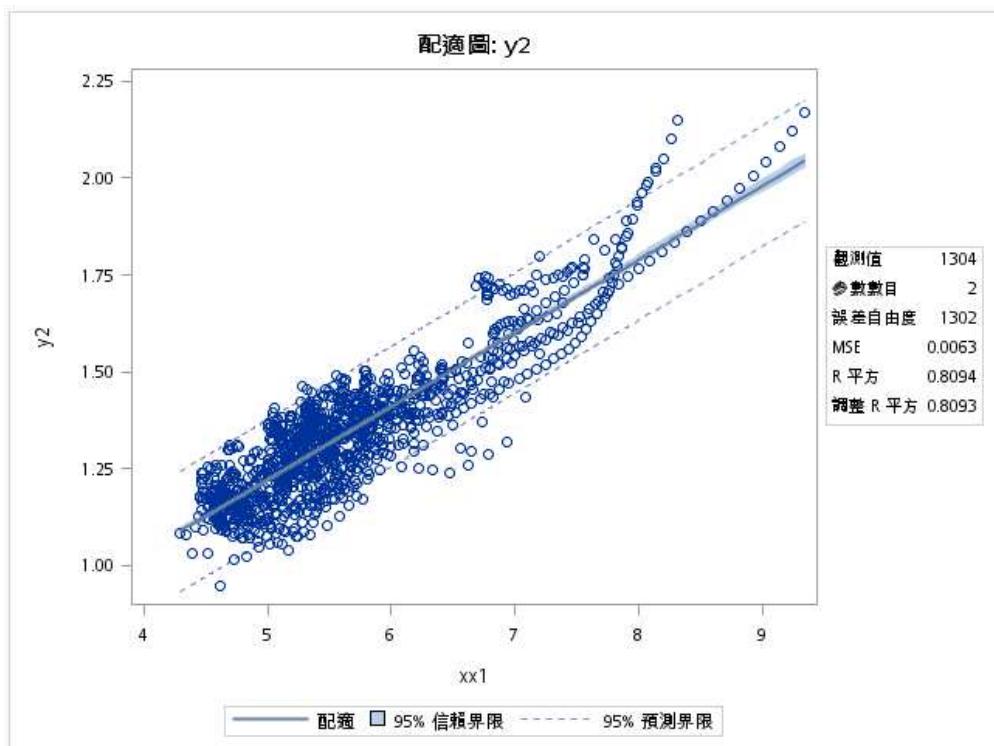
因此我們嘗試將 X 及 Y 均進行開根號轉換(亦即 $Y' = \sqrt{Y}$ 、 $X' = \sqrt{X}$)。實證結果殘差仍不符合常態分配，且殘差變異數亦不為常數。但由殘差圖可看出殘差越來越接近常態分配。得到迴歸線為 $Y' = 0.27381 + 0.18944 X'$ ，表示每增加 1% 扶幼比，總生育率增加 1.8944%。藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0=0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1=0$ 的假設。表示在統計上有證據證明 b_0 和 b_1 參數值都顯著不等於 0。

根 MSE	0.07911	R 平方	0.8094
應變平均值	1.33900	調整 R 平方	0.8093
變異係數	5.90845		

參數估計值					
變數	自由度	參數估計值	標準誤差	t 值	Pr > t
Intercept	1	0.27381	0.01449	18.90	<.0001
xx1	1	0.18944	0.00255	74.37	<.0001

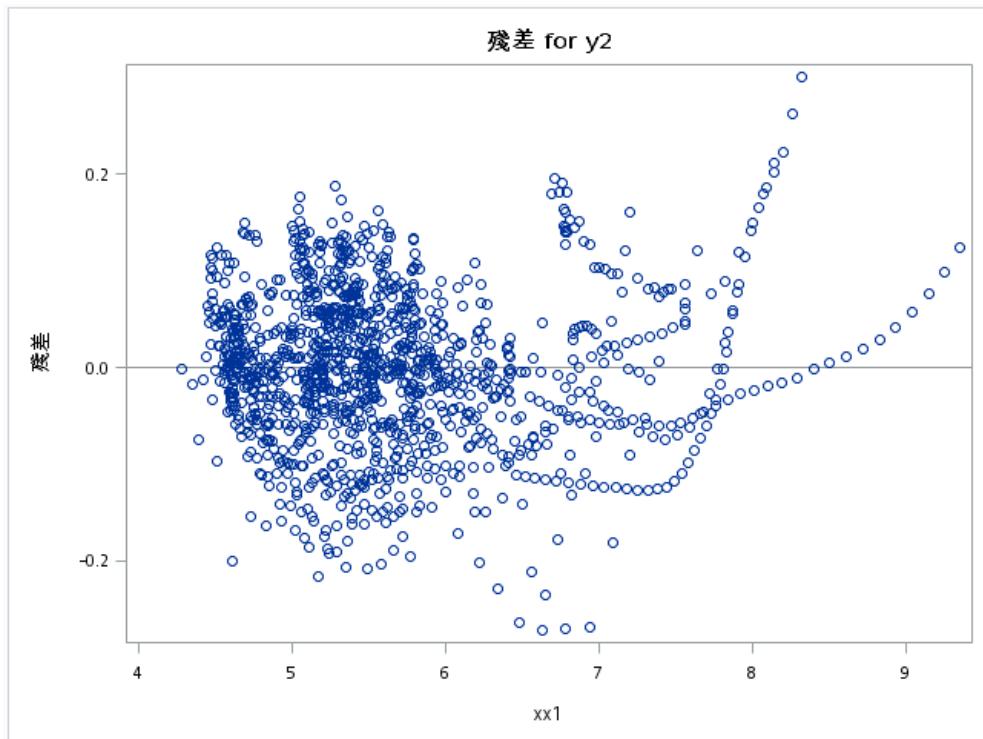
2. X 對 Y 的散佈圖

由 X 對 Y 散佈圖可看出有很多值都落在信賴區間外面，且殘差的變異數不一致情形。



3. 殘差對 X 的散佈圖

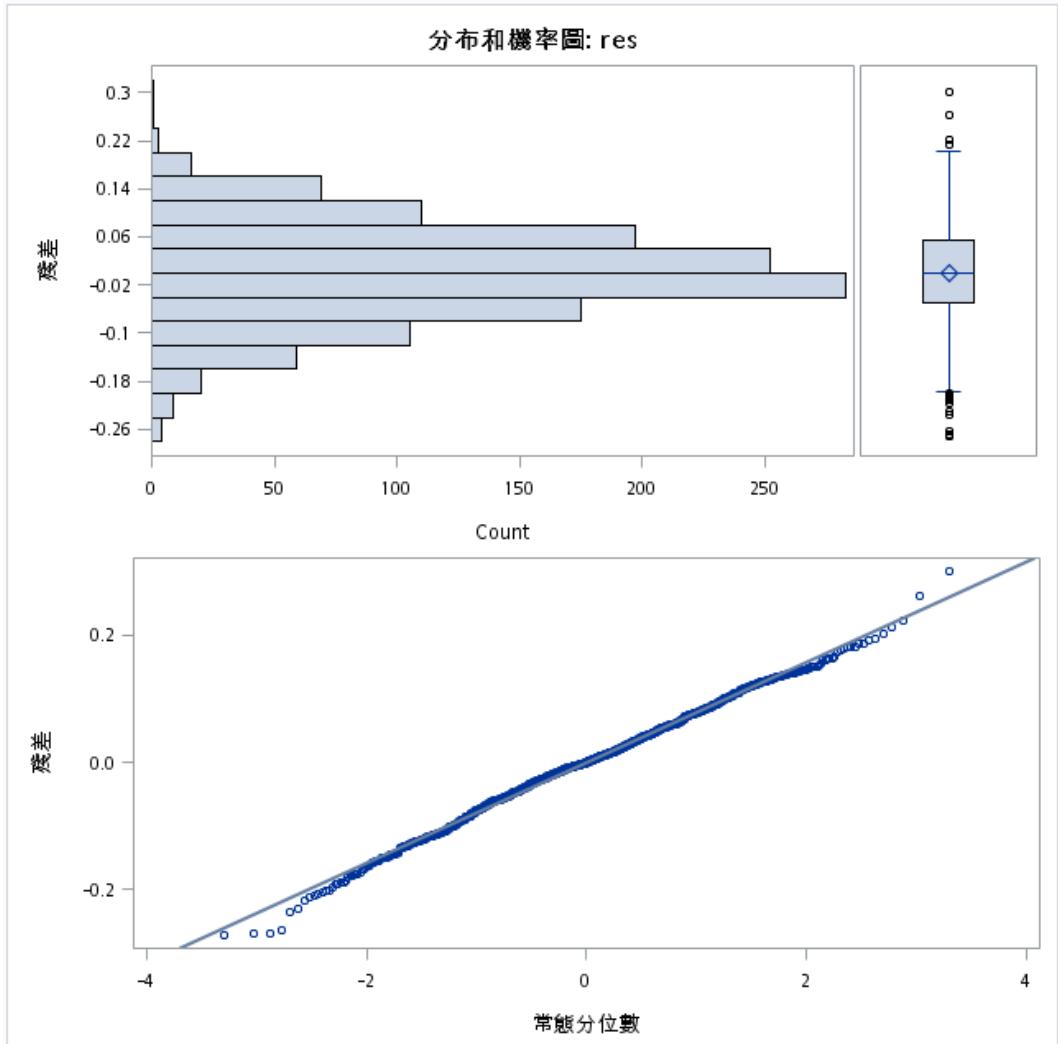
由殘差對 X 散佈圖可看出殘差的變異數不一致情形，且有極端值情形。



4. Shapiro-Wilk test 常態性檢定

利用 Shapiro-Wilk 對殘差作常態性檢定，檢定結果 $P\text{-value} < 0.0001$ ，拒絕 H_0 ，表示統計上有足夠的證據證明殘差的分布不符合常態分配。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.996626	Pr < W	0.0063
Kolmogorov-Smirnov	D	0.030748	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.219566	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.228387	Pr > A-Sq	<0.0050



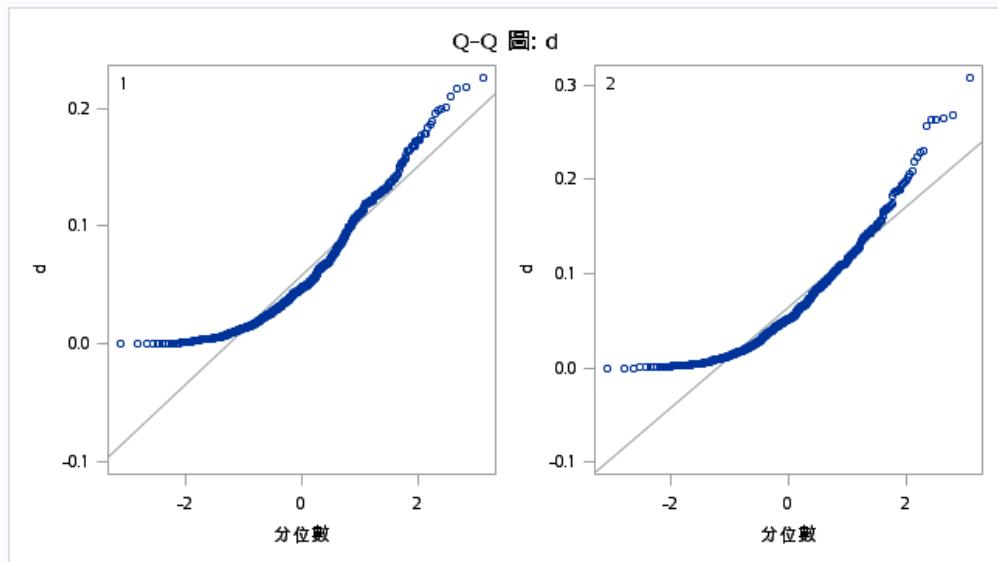
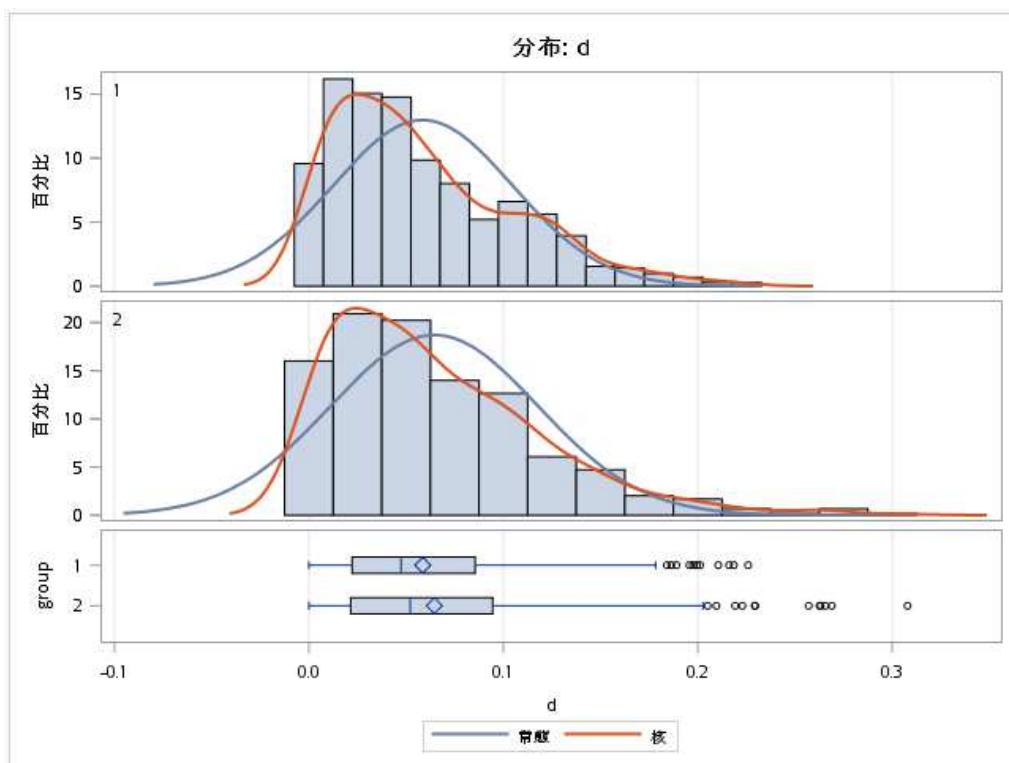
5. Brown-Forsythe test 變異數檢定

利用 Brown-Forsythe test 作檢定，檢定結果 P-value = 0.03 < 0.05，拒絕 H₀，表示統計上有足夠的證據證明殘差的變異數不相等。但 P-value 較轉換前變大，越來越接近接受 H₀ 的值了。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	711	0.0585	0.0461	0.00173	0	0.2258
2	593	0.0645	0.0533	0.00219	0	0.3078
Diff (1-2)		-0.00596	0.0495	0.00275		

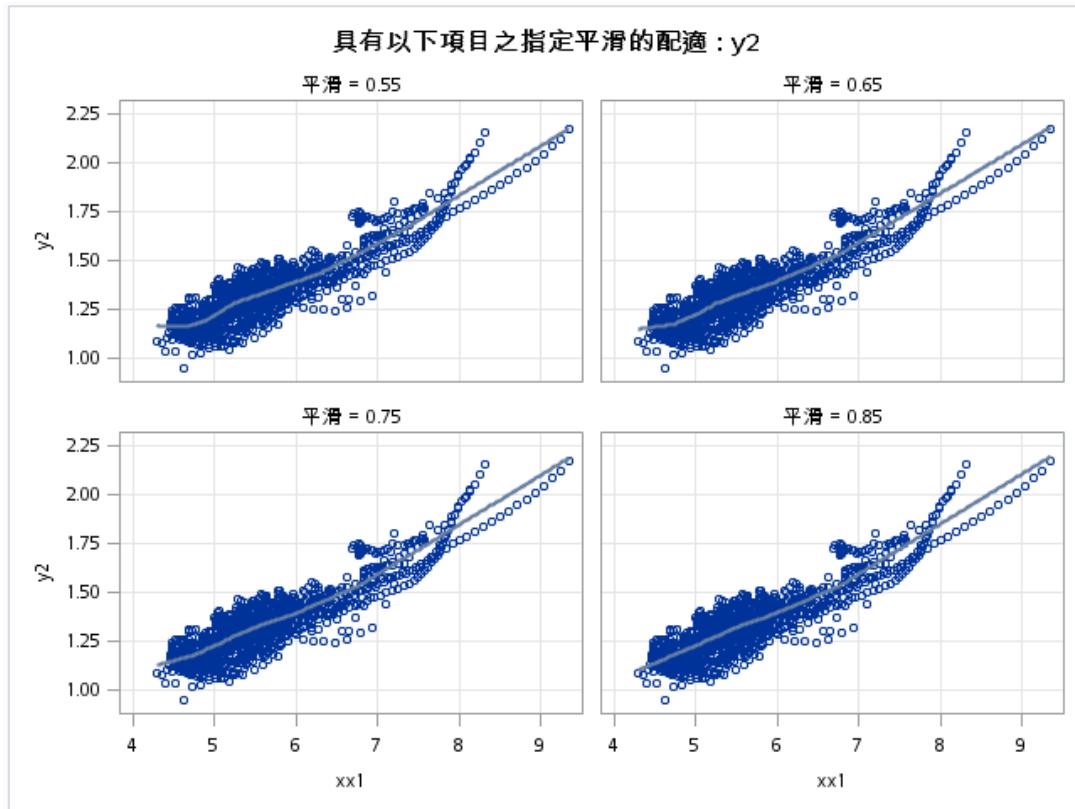
方法	變異數	自由度	t 值	Pr > t
集區	均等	1302	-2.16	0.0308
Satterthwaite	不均等	1179.3	-2.13	0.0330

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	592	710	1.33	0.0002



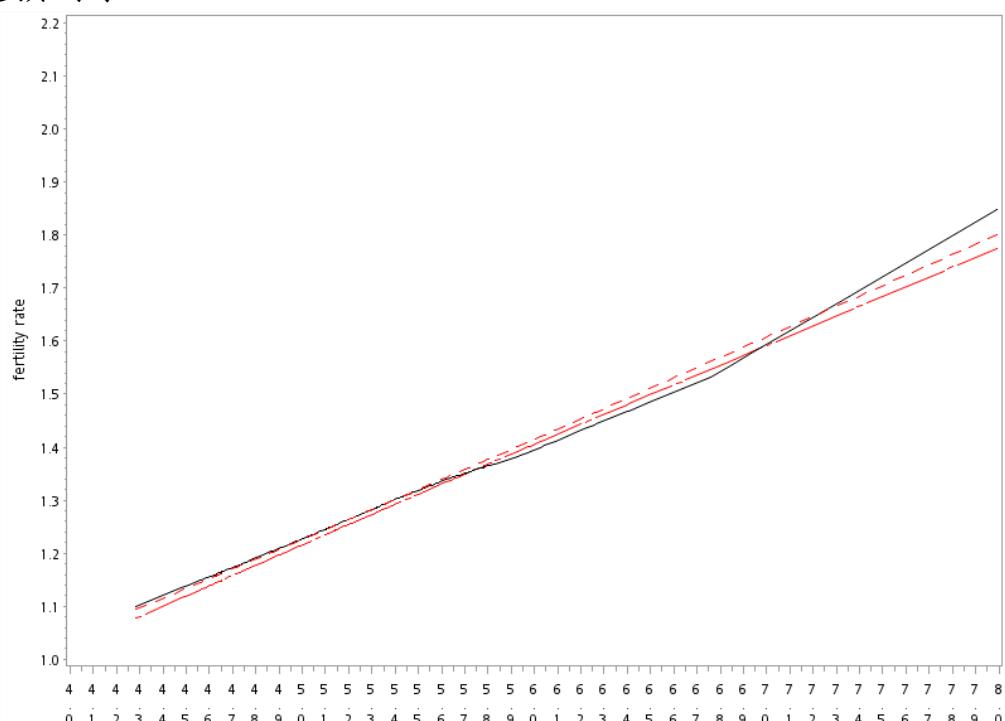
6. Lowess Curve

觀察 Lowess Curve 看趨勢線。



7. 預測線及信賴區間

此模型預測線雖然仍沒有全部包含在信賴區間內，但較轉換之前模型更接近預測線。



(三) 變數轉換後刪除離群值模型：

1. 簡單迴歸模型：

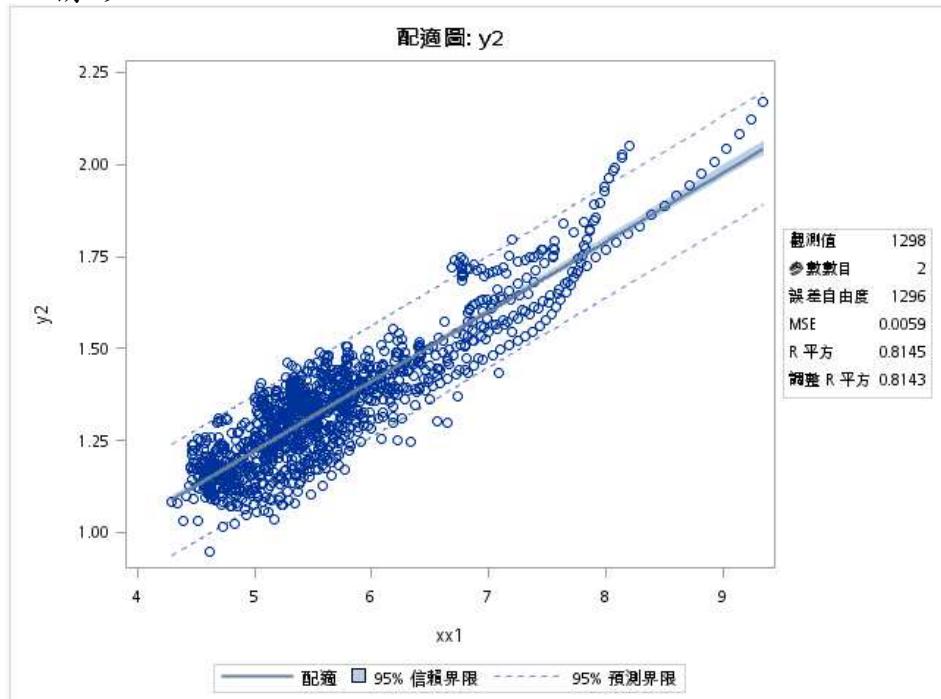
因為 X 及 Y 變數轉換後仍未達到最適的迴歸線，因此我們刪除離群值，亦即刪除半 T 化殘差絕對值 $\left| \frac{e}{\sqrt{MSE}} \right|$ 大於 3 的值，共 6 筆資料。殘差仍不符合常態分配，但殘差圖看來很接近常態分配，且殘差變異數轉變為常數。得到迴歸線為 $Y' = 0.27791 + 0.18866 X'$ ，表示每增加 1% 扶幼比，總生育率增加 1.8866‰。藉由 T 檢定可知， b_0 的 P-value 值很小，拒絕 $H_0 : b_0=0$ 的假設，同理 b_1 的 P-value 值也小於 0.05，拒絕 $H_1 : b_1=0$ 的假設。表示在統計上有證據證明 b_0 和 b_1 參數值都顯著不等於 0。

根 MSE	0.07925	R 平方	0.8023
應變平均值	1.33927	調整 R 平方	0.8022
變異係數	5.91704		

2. X 對 Y 的散佈圖

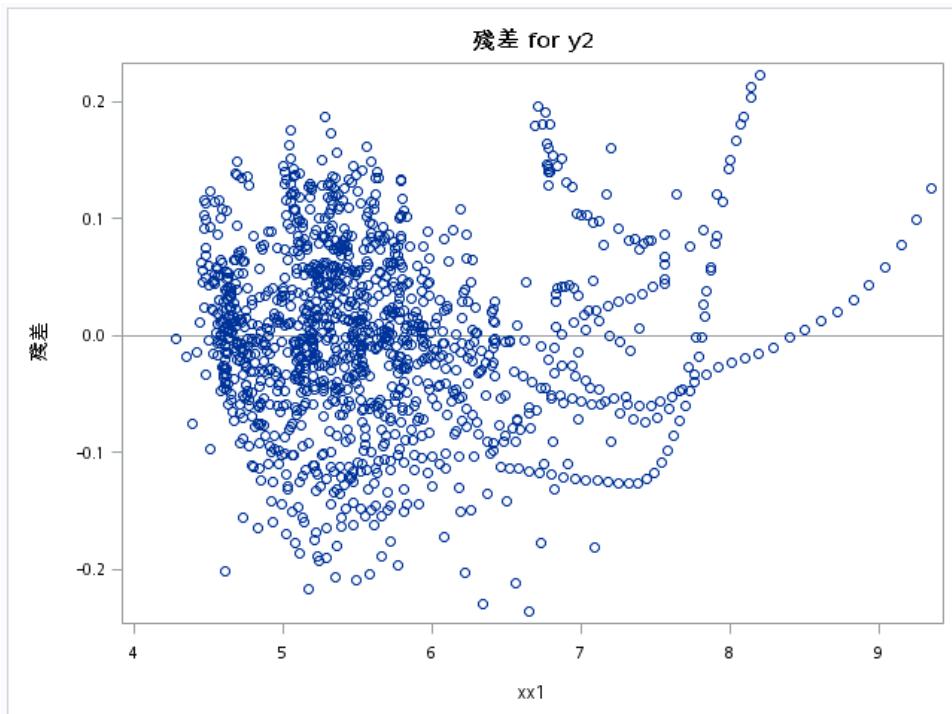
參數估計值					
變數	自由度	參數估計值	標準誤差	t 值	Pr > t
Intercept	1	0.27791	0.01483	18.74	<.0001
xx1	1	0.18866	0.00261	72.38	<.0001

由 X 對 Y 散佈圖可看出有部分值都落在信賴區間外面，且殘差的變異數不一致情形。



3. 殘差對 X 的散佈圖

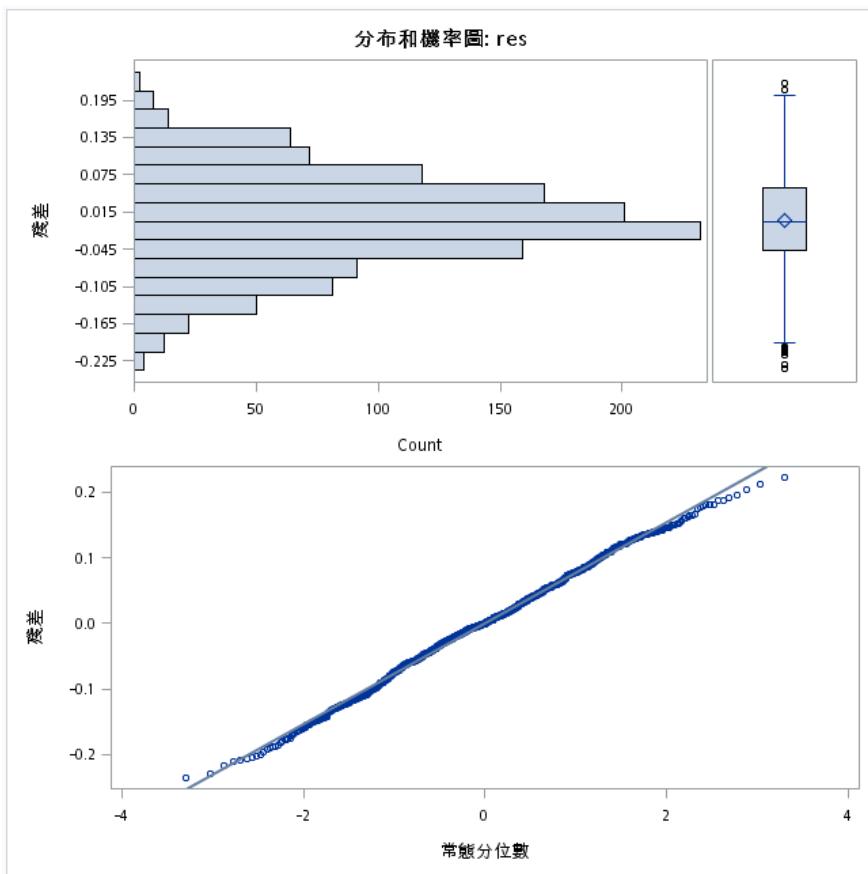
由殘差對 X 散佈圖可看出殘差的變異數不一致情形，且有極端值情形。



4. Shapiro-Wilk test 常態性檢定

利用 Shapiro-Wilk 對殘差作常態性檢定，檢定結果 P-value < 0.0001，拒絕 H₀，表示統計上有足夠的證據證明殘差的分布不符合常態分配。但看常態機率圖發現此模型已經很近似常態分配。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.997111	Pr < W	0.0180
Kolmogorov-Smirnov	D	0.026915	Pr > D	0.0222
Cramer-von Mises	W-Sq	0.163284	Pr > W-Sq	0.0173
Anderson-Darling	A-Sq	1.002473	Pr > A-Sq	0.0130



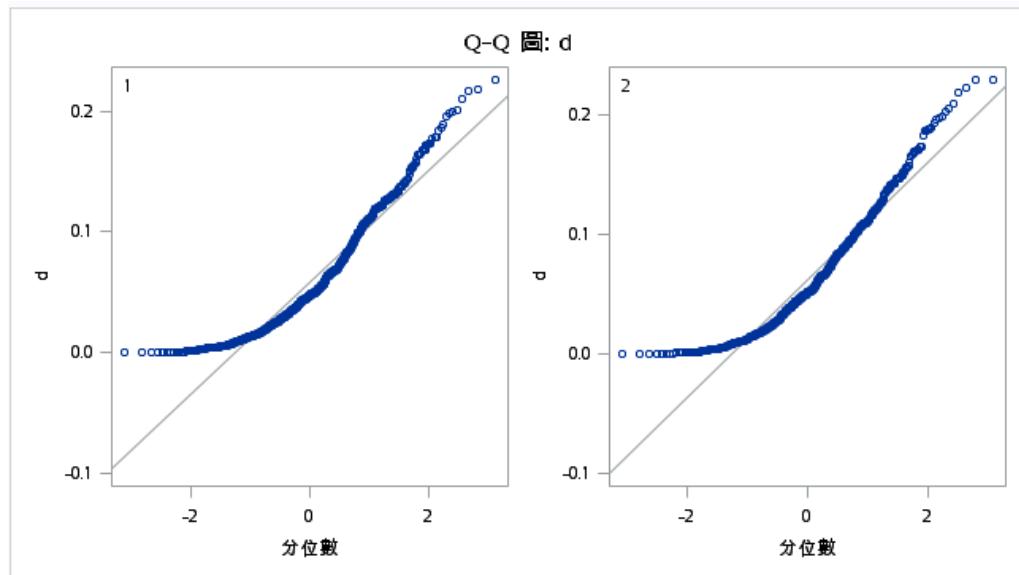
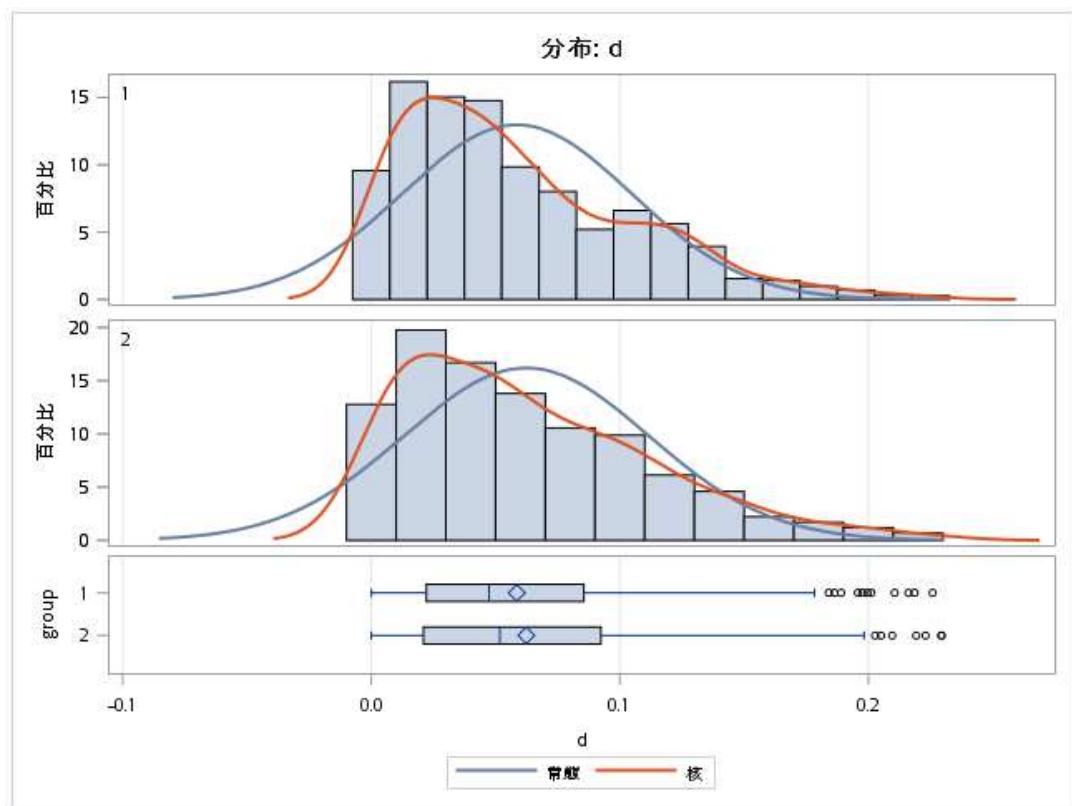
5. Brown-Forsythe test 變異數檢定

利用 Brown-Forsythe test 作檢定，檢定結果 P-value = 0.15 > 0.05，不拒絕 H₀，表示統計上沒有足夠的證據證明殘差的變異數不相等(亦即殘差變數相等)。

group	N	平均值	標準差	標準誤差	最小值	最大值
1	711	0.0585	0.0461	0.00173	0	0.2258
2	587	0.0624	0.0493	0.00203	0	0.2294
Diff (1-2)		-0.00385	0.0476	0.00265		

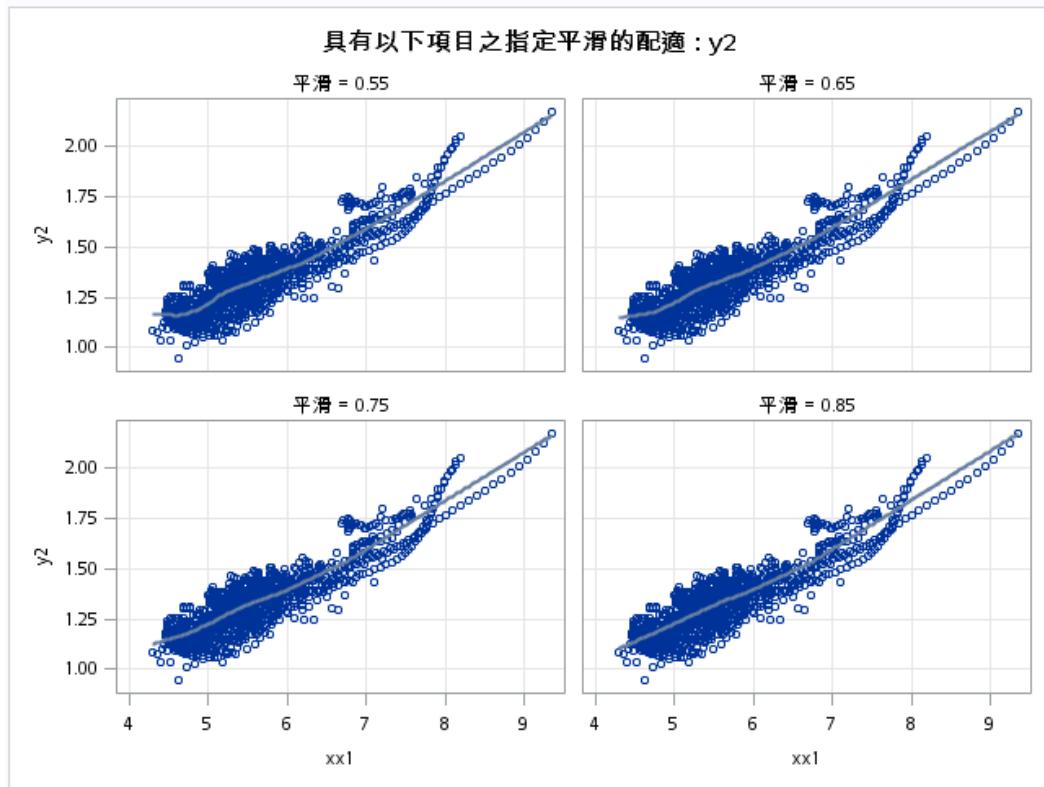
方法	變異數	自由度	t 值	Pr > t
集區	均等	1296	-1.45	0.1474
Satterthwaite	不均等	1216	-1.44	0.1499

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	586	710	1.14	0.0972



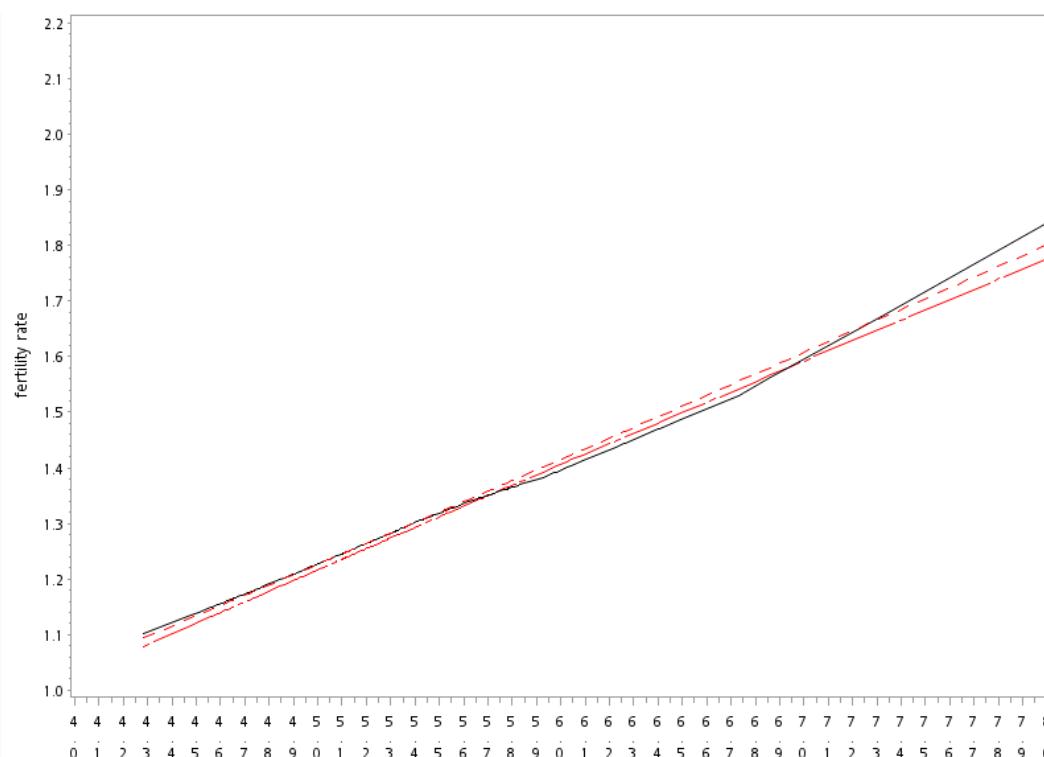
6. Lowess Curve

觀察 Lowess Curve 看趨勢線。



7. 預測線及信賴區間

此模型預測線雖然仍沒有全部包含在信賴區間內，但較前述模型更接近預測線。



第三章、結論

首先，我們先對總生育率 Y 和女性勞動力參與率 X1 做簡單線性迴歸模型，發現 R^2 僅有 0.004 的解釋力，且殘差不符合常態性檢定及殘差變異數不為常數值，且迴歸預測線不包含在信賴區間內。故我們進行轉換，根據 BOX-COX 可知道 lamda 接近 0 有最佳解，因此先針對 X 及 Y 進行 log 轉換，但跑出來的結果仍然殘差不符合常態性檢定及殘差變異數不為常數值，且迴歸預測線不包含在信賴區間內。因此我們再嘗試用 lamda=-1 進行倒數轉換，結果仍然殘差不符合常態性檢定及殘差變異數不為常數值，且迴歸預測線不包含在信賴區間內。故我們接著刪除離群值，也就是刪除半 T 化殘差絕對值大於 2.5 的值，總共刪除 6 筆資料。結果仍然殘差不符合常態性檢定及殘差變異數不為常數值，且迴歸預測線不包含在信賴區間內， R^2 依然僅有 0.0036 的解釋力，因此我們決定捨棄此組資料。

接著我們選擇對總生育率 Y 和扶幼比 X9 做簡單線性迴歸模型，發現 R^2 有 0.8309 的解釋力，但殘差不符合常態性檢定及殘差變異數不為常數值，且迴歸預測線不包含在信賴區間內。故我們進行轉換，根據 BOX-COX 可知道 lamda 接近 0.5 有最佳解，因此針對 X 及 Y 進開根號轉換，但跑出來的結果仍然殘差不符合常態性檢定及殘差變異數不為常數值，且迴歸預測線不包含在信賴區間內。故我們接著刪除離群值，也就是刪除半 T 化殘差絕對值大於 3 的值，總共刪除 6 筆資料。結果雖然殘差仍不符合常態性檢定及殘差變異數已經符合常數，且迴歸預測線已很接近在信賴區間內， R^2 也有 0.8023 的解釋力。