

# 逢甲大學學生報告 ePaper

報告題名：

回歸分析：

分析失業率的影響因素

Regression analysis：

Analysis the influence factors of unemployment rate

作者：劉庭汝、張巧穎、鍾宜頻、洪慈玳、黃姿穎、管奕錚、董心成

系級：統計系三年甲班

學號：D0172307、D0130301、D0107356、D0130230、D0172311、D0157694

D0085256

開課老師：高秀蘭 教授

課程名稱：回歸分析

開課系所：統計學系

開課學年： 103 學年度 第 1 學期

## 中文摘要

近年來新聞媒體常常報導有關失業率的問題，而失業率影響著我們未來的就業問題。因而我們討論出幾個可能會影響失業率的因素，分別為出生人數、外籍勞工人數、消費者物價基本指數、景氣同時指標綜合指數、外銷訂單、歇業家數、平均薪資、工業生產指數及所得稅，利用這九個解釋變數來分析哪些變數會影響失業率。

在報告中，主要探討各解釋變數與反應變數之間的關係。首先我們先建立一組回歸模型，然後將模型做適合度檢定，再判斷各解釋變數和反應變數之間是否有線性關係，確定有線性關係後再進行變數的選取。我們進行三種變數選取法，分別為向前選取法、向後選取法、逐次替換法，執行三種選取法後，將我們得到的變數進行殘差分析，來預測此模型是否符合常態性、均齊性及獨立性，因其中均齊性及獨立性不符合假設，所以我們對模型做轉換，再次檢定殘差，進而找出最佳的模型。

最後找出最佳模型中的變數有出生人數、外籍勞工人數、消費者物價基本指數、景氣同時指標綜合指數、外銷訂單、歇業家數及所得稅，客觀地由這七個變數討論影響失業率的因素。



**關鍵字：**失業率、解釋力、殘差分析、變數選取、複回歸模型

## Abstract

In recent years, the news media often reports on the unemployment problem, and the unemployment rate affect our future employment. So we discussed a few factors that may affect the unemployment rate, respectively, The number of births, The number of foreign workers, Basic consumer price index, Composite index, Export orders, Went out of business, Average salary, The index of industrial production and Income tax by nine explanatory variables to analyze which variables will affect the unemployment rate.

In the report, focuses on the relationship between the explanatory variables and the response variable between. First, let's establish a set of regression model, then the model is doing fitness test, and then determine whether there is a linear relationship between the explanatory variables and the response variable, there is a linear relationship then conducted to determine the selection of variables. We conducted three variable selection method, namely Forward selection method, Backward selection method, All-possible selection method, after the implementation of the three selection method, we get residuals variables were analyzed to predict whether this model is consistent with normality, homogeneity and independence, because of which are homogeneity and independence does not comply with the assumption, so we do the conversion model, test residuals again, and then find the best model.

Finally, to find the best model variables have The number of births, The number of foreign workers, Basic consumer price index, Composite index, Export orders, Went out of business and Income tax, objective discussion of these seven variables affected unemployment factor.

**Keyword** : Explanatory power 、 Multiple regression model 、 Residual analysis 、 Unemployment rate 、 Variable selection

## 目 次

第壹章 緒論	7
一、研究動機	7
二、流程圖	8
第貳章 基本敘述統計	9
一、變數名稱介紹	9
二、敘述性統計	10
第參章 原始複回歸模型	11
一、複回歸模型	11
二、模型適合度	12
三、模型解釋力	13
四、參數檢定	14
第肆章 變數選取	18
一、向前選取法(Forward Selection)	18
二、向後選取法(Back Selection)	19
三、逐次替換法(All-Possible Selection)	20
四、選取變數	21
第伍章 選取後新複回歸模型	22
一、複回歸模型	22
二、模型適合度	22
三、模型解釋力	23
第陸章 殘差分析	24
一、常態性	24
二、均齊性	25
三、獨立性	26
第柒章 轉換模型後的殘差分析	27
一、常態性	28
二、均齊性	29
三、獨立性	30
第捌章 轉換後新複回歸模型	31
一、複回歸模型	31
二、模型適合度	31
三、模型解釋力	32
四、參數檢定	32
五、共線性	35
第玖章 結論	36

第壹拾章	附錄	-----	37
一、	原始檔案	-----	37
二、	R 程式碼	-----	40
三、	工作分配	-----	47
四、	參考文獻	-----	47

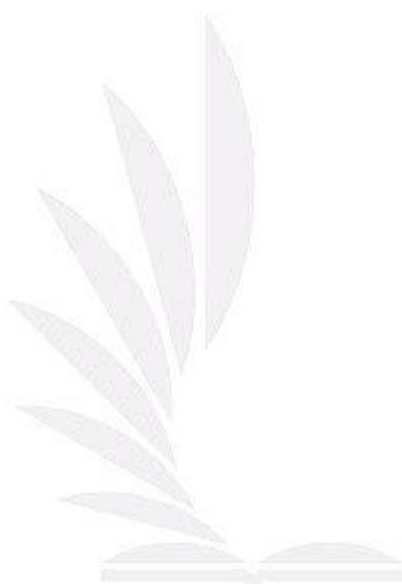


## 表目錄

表 2.1 敘述統計量	10
表 3.1 參數估計表	11
表 3.2 變異數分析表	12
表 3.3 適合度檢定表	12
表 3.4 模型解釋力	13
表 3.5 參數估計表	14
表 4.1 向前選取法結果	18
表 4.2 向後選取法結果	19
表 4.3 逐步替換法結果	20
表 4.4 各選取法統整表	21
表 5.1 選取後新適合度檢定	22
表 5.2 選取後新模型解釋力	23
表 6.1 常態性檢定表	24
表 6.2 均齊性檢定表	25
表 6.3 獨立性檢定表	26
表 7.1 兩種轉換解釋力比較	27
表 7.2 轉換後常態性檢定表	28
表 7.3 轉換後均齊性檢定表	29
表 7.4 轉換後獨立性檢定表	30
表 8.1 轉換後新適合度檢定表	31
表 8.2 轉換後新解釋力	32
表 8.3 轉換後參數檢定表	32
表 8.4 共線性	35

## 圖目錄

圖 6.1 Q-Q Plot	24
圖 6.2 殘差圖	25
圖 7.1 轉換後 Q-Q Plot	28
圖 7.2 轉換後殘差圖	29



## 第壹章 緒論

### 一、研究動機

由於台灣青年失業率已衝破 13%，是平均失業率的 3 倍。與我們息息相關。所以我們想探討失業率幾個相關的影響因素。於是我們經過討論，決定研究的課題為失業率，眾所皆知失業率是指一定時期滿足全部就業條件的就業人口中仍未有工作的勞動力數字旨在衡量閒置中的勞動產能，是反映一個國家或地區失業狀況的主要指標。而從統計學的角度出發，失業率也是一個較為容易受到其餘多種因素影響的變數，就像有關於物價與失業率的菲利普經典曲線，所以從多個角度出發我們選擇了研究失業率，同時我們找了九個可能影響失業率的變數來分析，探討哪些變數對失業率的影響較大。

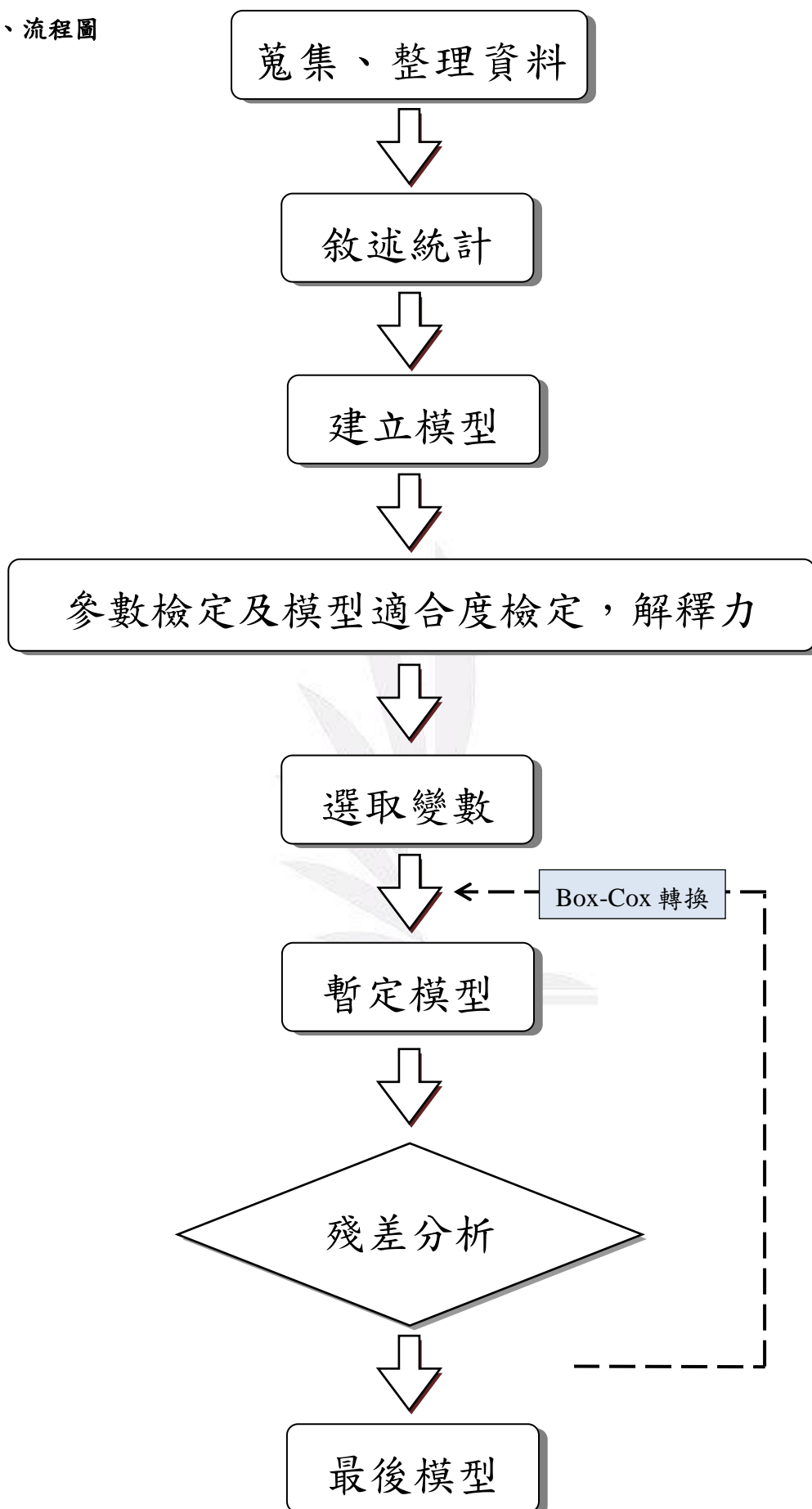
至於研究方法我們則選擇了回歸分析，眾所周知，回歸分析是一種應用極其廣泛的資料分析方法。它是基於觀測資料然後建立變數間適當的依賴關係，從而分析資料其內在規律的一種統計手段，簡單來說回歸分析也是一種處理變數之間統計相關關係的工具。他可以快速的找出解釋變數和反應變數之間的數學表達形式，哪怕原先還未知解釋變數和反應變數之間的確定性關係。

因為數據的大量性，同時出於對數據分析準確性的考量所以我們選擇了 R 軟體來進行我們的回歸分析研究，在 R 精密的統計計算環境下我們得到了如下的分析結果與圖表。





二、流程圖



## 第貳章 基本敘述統計

### 一、變數名稱介紹

#### 反應變數 (Y)

失業率：失業率是勞動人口裡符合“失業條件”者所佔的比例。與我們息息相關。所以我們來探討失業率幾個相關的影響因素，以這些解釋變數去分析。

#### 解釋變數 (X)

$X_1$ ：出生人數，最先想到的變數，因為勞動人口最直接與出生率相關。人口的討論，一定是與所有社會活動息息相關。若出生人口是增長趨勢，國家可享有人口紅利，可拉動社會經濟發展，增加就業需求。

$X_2$ ：外籍勞工人數：由於台灣人口趨於老齡化，外籍勞工是台灣失業率所需要討論的因素。由於外籍勞工多為青壯人口，且從事於很多台灣人不願意從事的工作，可以拉低失業率。

$X_3$ ：消費者物價基本指數 (CPI)：放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，也是衡量通貨膨脹的主要指標之一。如果CPI升幅過大，表明通貨膨脹已經成為經濟不穩定因素了，所以，該指數過高升幅不被市場所歡迎。(以民國100年為基期)

$X_4$ ：景氣同時指標綜合指數：根據能夠反映當時景氣狀況的指標編制而成，可用來判斷當時的景氣狀況。目前台灣納入的指標有6項，包括工業生產變動率、製造業生產變動率、國內貨運量、製造業銷售值、製造業平均每月薪資變動率及票據交換金額變動率。若經濟景氣良好，則需要更多工作崗位，勢必會降低失業率。

$X_5$ ：外銷訂單：根據外銷廠商承接國外客戶貨物訂單額度資料所編制之統計，由於台灣大量依賴對外貿易，所以外銷訂單某種意義上會帶動台灣的經濟，增加工作機會。

$X_6$ ：歇業家數：工廠歇業數量，最直觀減少工作的崗位的指標。

$X_7$ ：平均薪資：就業最先考慮的一項指標。當平均薪資增加所需要的的標準會更高。

$X_8$ ：工業生產指數：用加權平均數編制的工業產品實物量指數。該指數是相對指標，衡量製造業、礦業與公共事業的實質產出，反映的是某一時期工業經濟的景氣狀況和發展趨勢。(以民國 100 年為基期)

$X_9$ ：所得稅：指以納稅人的所得額為課稅對象的各種稅收的統稱。稅法規定的所得額，是指納稅人在一定時期內，由於生產、經營等取得的可用貨幣計量的額收入，扣除為取得這些收入所需各種耗費後的淨額。

## 二、敘述性統計

表 2.1 敘述統計量

敘述統計量							
	標籤	樣本數	平均數	標準差	總和	最小值	最大值
Y	失業率(%)	96	4.47708	0.68238	429.8	3.78	6.13
$X_1$	出生人數(千)	96	16.57316	2.29609	1591.023	10.597	23.398
$X_2$	外籍勞工人數(千)	96	384.732	44.96188	36934.272	327.311	489.134
$X_3$	消費者物價基本指數	96	98.50979	3.15064	9456.94	91.94	104.06
$X_4$	景氣同時指標綜合指數(點)	96	90.51854	10.42325	8689.78	66.76	103.93
$X_5$	外銷訂單(千)	96	9.83729	1.35283	944.38	5.87	12.56
$X_6$	歇業家數(千)	96	2.84524	1.29472	273.143	1.307	7.81
$X_7$	平均薪資(千)	96	44.4688	9.67392	4269.005	38.138	83.392
$X_8$	工業生產指數	96	90.00698	11.90233	8640.67	54.21	107.42
$X_9$	所得稅(百萬)	96	58927.19417	69778.29919	5657010.64	-26803.73	332543.84

## 第參章 原始複回歸模型

### 一、複回歸模型

影響失業率 (Y) 的變數有出生人數 (X<sub>1</sub>)、外籍勞工人數 (X<sub>2</sub>)、消費者物價基本指數 (X<sub>3</sub>)、景氣同時指標綜合指數 (X<sub>4</sub>)、外銷訂單 (X<sub>5</sub>)、歇業家數 (X<sub>6</sub>)、平均薪資 (X<sub>7</sub>)、工業生產指數 (X<sub>8</sub>)、所得稅 (X<sub>9</sub>)，利用這九個變數組成複回歸模型。

表 3.1 參數估計表

參數估計表					
變數	標籤	參數估計值	標準誤差	t 值	Pr(> t )
截距	截距	-8.4094	2.3969	-3.508	0.000720
X <sub>1</sub>	出生人數(千)	-0.0634	0.0198	-3.202	0.001916
X <sub>2</sub>	外籍勞工人數(千)	-0.0137	0.0025	-5.455	4.64e-07
X <sub>3</sub>	消費者物價基本指數	0.2220	0.0308	7.202	2.11e-10
X <sub>4</sub>	景氣同時指標綜合指數(點)	-0.0523	0.0130	-4.027	0.000121
X <sub>5</sub>	外銷訂單(千)	0.3515	0.0889	3.955	0.000157
X <sub>6</sub>	歇業家數(千)	-0.2094	0.0356	-5.887	7.41e-08
X <sub>7</sub>	平均薪資(千)	0.0040	0.0052	0.774	0.440969
X <sub>8</sub>	工業生產指數	-0.0099	0.0124	-0.798	0.426926
X <sub>9</sub>	所得稅(百萬)	-1.311e-06	5.976e-07	-2.193	0.030975

由上表參數估計值初步建立的預測模型為：

$$\hat{Y} = -8.4094 - 0.0634X_1 - 0.0137X_2 + 0.2220X_3 - 0.0523X_4 + 0.3515X_5 - 0.2094X_6 + 0.0040X_7 - 0.0099X_8 - 0.000001311X_9$$

## 二、模型適合度

我們利用 R 做出變異數分析表如下：

表 3.2 變異數分析表

anova						
變數	標籤	自由度	平方和	均方和	F 值	Pr(>F)
X <sub>1</sub>	出生人數(千)	1	3.9854	3.9854	25.6634	2.295e-06
X <sub>2</sub>	外籍勞工人數(千)	1	1.0915	1.0915	7.0283	0.0095499
X <sub>3</sub>	消費者物價指數	1	16.1593	16.1593	104.0562	< 2.2e-16
X <sub>4</sub>	景氣同時指標綜合指數(點)	1	1.8968	1.8968	12.2145	0.0007521
X <sub>5</sub>	外銷訂單(千)	1	1.3895	1.3895	8.9477	0.0036239
X <sub>6</sub>	歇業家數(千)	1	5.1927	5.1927	33.4378	1.160e-07
X <sub>7</sub>	平均薪資(千)	1	0.2721	0.2721	1.7523	0.1890979
X <sub>8</sub>	工業生產指數(%)	1	0.1468	0.1468	0.9452	0.3336661
X <sub>9</sub>	所得稅(百萬)	1	0.7471	0.7471	4.8111	0.0309750
	殘差	86	13.3552	0.1553		

經過我們整理得到適合度檢定表，如下：

表 3.3 適合度檢定表

來源	df	平方和	均方	F 值	Pr>F
Model	9	30.88142	3.431269	22.09584	< 2.2e-16
Error	86	13.35496	0.1552903		
SST	95	44.23638			

整體模式檢定

假設檢定

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$$

$$H_1: \text{至少有一 } \beta_i \neq 0, \quad i = 1, 2, 3, 4, 5, 6, 9$$

檢定統計量  $F^* = 22.09584$

$$P - \text{value} = 2.2e - 16 < \alpha = 0.05$$

由  $P - \text{value} < \alpha = 0.05$ ，我們得出拒絕虛無假設  $H_0$ ，表示解釋變數與反應變數有線性關係。

### 三、模型解釋力

表 3.4 模型解釋力

解釋力			
標準化殘差值	0.3941	調整 R 平方	0.6665
自由度	86	F 統計量	22.1
判定係數 R 平方	0.6981	P 值	< 2.2e-16

$$R^2 = 0.6981$$

$$\text{Adj } R^2 = 0.6665$$

全模型的  $R^2$  為 0.6981，而經過校正的  $R^2$  為 0.6665，減輕因樣本估計所帶來的  $R^2$  膨脹效果。表示解釋變數的變異可以被新模型所解釋的比例有 66.65%。在下一章我們將進行選取變數來選出最佳模型。

#### 四、參數檢定

建立模型之後並不代表所有變數都為顯著的解釋變數，因此，我們接下來要判斷各個變數對Y是否有線性相關，即是否為顯著的變數。

表 3.5 參數估計表

參數估計表					
變數	標籤	參數估計值	標準誤差	t 值	Pr(> t )
截距	截距	-8.4094	2.3969	-3.508	0.000720
$X_1$	出生人數(千)	-0.0634	0.0198	-3.202	0.00191
$X_2$	外籍勞工人數(千)	-0.0137	0.0025	-5.455	4.64e-07
$X_3$	消費者物價指數	0.2220	0.0308	7.202	2.11e-10
$X_4$	景氣同時指標綜合指數(點)	-0.0523	0.0130	-4.027	0.000121
$X_5$	外銷訂單(千)	0.3515	0.0889	3.955	0.000157
$X_6$	歇業家數(千)	-0.2094	0.0356	-5.887	7.41e-08
$X_7$	平均薪資(千)	0.0040	0.0052	0.774	0.440969
$X_8$	工業生產指數	-0.0099	0.0124	-0.798	0.426926
$X_9$	所得稅(百萬)	-1.311e-06	5.976e-07	-2.193	0.030975

- (1) 想要判斷 $X_1$ (出生人數)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=0.00191 <  $\alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_1 \neq 0$ 。因此， $X_1$ (出生人數)是一個顯著的解釋變數和Y(失業率)存在線性相關。

- (2) 想要判斷 $X_2$ (外籍勞工人數)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=4.64e-07 <  $\alpha=0.05$  所以拒絕虛無假設  $H_0$ ，表示我們有充分的證據顯示  $\beta_2 \neq 0$ 。因此， $X_2$ (外籍勞工人數) 是一個顯著的解釋變數和Y(失業率)存在線性相關。

- (3) 想要判斷 $X_3$ (消費者物價指數)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=2.11e-10 <  $\alpha=0.05$  所以不拒絕虛無假設  $H_0$ ，表示我們沒有充分的證據顯示  $\beta_3 \neq 0$ 。因此， $X_3$ (消費者物價指數)是一個顯著的解釋變數和Y(失業率)存在線性相關。

- (4) 想要判斷 $X_4$ (景氣同時指標綜合指數)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=0.000121 <  $\alpha=0.05$  所以不拒絕虛無假設  $H_0$ ，表示我們沒有充分的證據顯示  $\beta_4 \neq 0$ 。因此， $X_4$ (景氣同時指標綜合指數)



是一個顯著的解釋變數和Y(失業率)存在線性相關。

- (5) 想要判斷 $X_5$ (外銷訂單)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha = 0.05$

檢定如下：因為  $P\text{-value} = 0.000157 < \alpha = 0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_5 \neq 0$ 。因此， $X_5$ (外銷訂單)是一個顯著的解釋變數和Y(失業率)存在線性相關。

- (6) 想要判斷 $X_6$ (歇業家數)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha = 0.05$

檢定如下：因為  $P\text{-value} = 7.41e-08 < \alpha = 0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_6 \neq 0$ 。因此， $X_6$ (歇業家數)是一個顯著的解釋變數和Y(失業率)存在線性相關。

- (7) 想要判斷 $X_7$ (平均薪資)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_7 = 0$$

$$H_1 : \beta_7 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha = 0.05$

檢定如下：因為  $P\text{-value} = 0.440969 > \alpha = 0.05$  所以不拒絕虛無假設 $H_0$ ，

表示我們沒有充分的證據顯示 $\beta_7 \neq 0$ 。因此， $X_7$ (平均薪資)不是一個顯著的解釋變數和 $Y$ (失業率)不存在線性相關。

- (8) 想要判斷 $X_8$ (工業生產指數)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_8 = 0$$

$$H_1 : \beta_8 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha = 0.05$

檢定如下：因為  $P\text{-value} = 0.426926 > \alpha = 0.05$  所以拒絕虛無假設 $H_0$ ，

表示我們有充分的證據顯示 $\beta_8 \neq 0$ 。因此， $X_8$ (工業生產指數)不是一個顯著的解釋變數和 $Y$ (失業率)不存在線性相關。

- (9) 想要判斷 $X_9$ (所得稅)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_9 = 0$$

$$H_1 : \beta_9 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha = 0.05$

檢定如下：因為  $P\text{-value} = 0.030975 < \alpha = 0.05$  所以不拒絕虛無假設 $H_0$ ，表示我們沒有充分的證據顯示 $\beta_9 \neq 0$ 。因此， $X_9$ (所得稅)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關。

## 第肆章 變數選取

在選取變數時，我們都希望能找到最準確也最簡單的模型，即挑選變數來精簡模型，所以我們利用三種選取法來選出一組最佳的組合，能讓模型為最佳的模型，以下為我們所使用的三種選取法：

### 一、向前選取法(Forward Selection)：

是以“Step-by-step”方式，將最有區別能力的變數選入，也就是在每一步驟將評估選取最有區別能力的變數進入模型內，然後再進行下一步驟，如此反覆選取變數中最有區別能力者，如此依序進行，但其解釋能力也要通過事先訂好的門檻。

● 選取方法：

1. 選出最大的調整 $R^2$
2. 選出 $p + 1$ 最接近  $C_p$  值 (p：模型中被選取的變數總數)
3. 選出 BIC 最小者

表 4.1 向前選取法結果

No	(Intercept)	x1	x2	x3	x4	x5	x6	x7	x8	x9	rsq	adj $R^2$	rss	Cp	BIC
1	1	1	0	0	0	0	0	0	0	0	0.0901	0.0804	40.2510	167.1931	0.0651
2	1	1	0	0	1	0	0	0	0	0	0.1563	0.1381	37.3243	150.3469	-2.6176
3	1	1	1	1	0	0	0	0	0	0	0.4801	0.4631	23.0003	60.1086	-44.5304
4	1	1	1	1	1	0	0	0	0	0	0.5229	0.5020	21.1035	49.8941	-48.2287
5	1	1	1	1	1	0	1	0	0	0	0.6025	0.5805	17.5819	29.2170	-61.1911
6	1	1	1	1	1	1	1	0	0	0	0.6717	0.6496	14.5213	11.5086	-74.9870
7	1	1	1	1	1	1	1	0	0	1	0.6928	0.6683	13.5907	7.5164	-76.7805
8	1	1	1	1	1	1	1	0	1	1	0.6960	0.6680	13.4483	8.5993	-73.2275

由上表，依據選取方法

1. 從調整 $R^2$ 找出最大值為 0.6683 (模型編號 7)
2.  $p + 1$ 最接近  $C_p$  值得為 8.5993(模型編號 8)
3. 最小的 BIC 值為-76.7805(模型編號 7)

因此我們在模型 7 與模型 8 之間繼續挑選最佳模型，由調整 $R^2$ 和 BIC 值我們更傾向於模型 7，且 $p + 1$ 與  $C_p$  值之間的差值，兩組模型相差不大，所以向後選取法的最佳模型為模型編號 7，解釋變數包括 $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_9$ 。

## 二、向後選取法(Back Selection)：

與向前選取法相反，開始時將所有變項都選入模型內，接著第一步驟將最沒有區別能力的變項刪除，再依序將沒有區別能力的變項一一去除。

表 4.2 向後選取法結果

No	(Intercept)	x1	x2	x3	x4	x5	x6	x7	x8	x9	rsq	adjR <sup>2</sup>	rss	Cp	BIC
1	1	1	0	0	0	0	0	0	0	0	0.0901	0.0804	40.2510	167.1931	0.0651
2	1	0	1	1	0	0	0	0	0	0	0.4545	0.4427	24.1319	65.3956	-44.4841
3	1	0	1	1	0	0	1	0	0	0	0.5412	0.5263	20.2939	42.6810	-56.5483
4	1	0	1	1	1	0	1	0	0	0	0.5862	0.5680	18.3058	31.8784	-61.8820
5	1	0	1	1	1	1	1	0	0	0	0.6397	0.6196	15.9403	18.6463	-70.6007
6	1	1	1	1	1	1	1	0	0	0	0.6717	0.6496	14.5213	11.5086	-74.9870
7	1	1	1	1	1	1	1	0	0	1	0.6928	0.6683	13.5907	7.5164	-76.7805
8	1	1	1	1	1	1	1	0	1	1	0.6960	0.6680	13.4483	8.5993	-73.2275

由上表，依據選取方法

1. 從調整R<sup>2</sup>找出最大值為 0.6683 (模型編號 7)
2.  $p + 1$ 最接近 Cp 值得為 8.5993(模型編號 8)
3. 最小的 BIC 值為-76.7805(模型編號 7)

因此我們在模型 7 與模型 8 之間繼續挑選最佳模型，由調整R<sup>2</sup>和 BIC 值我們更傾向於模型 7，且 $p + 1$ 與 Cp 值之間的差值，兩組模型相差不大，所以向後選取法的最佳模型為模型編號 7，解釋變數包括X<sub>1</sub>、X<sub>2</sub>、X<sub>3</sub>、X<sub>4</sub>、X<sub>5</sub>、X<sub>6</sub>、X<sub>9</sub>。

### 三、逐步替換法(All-Possible Selection)：

結合「向前選取法」與「向後選取法」而成，開始時以向前選取法選入一個預測變數；而後每當選入一個新的預測變數時，就利用向後選取法，看看在模式中已存在的預測變數有無偏 F 值小於  $F_{out}$  的變數。如果有，則偏 F 值最小的預測變數就會被排除在模型之外，接著再進行向前選取；如果沒有，則繼續向前選取，重複輪流使用向前、向後選取的步驟，直到沒有預測變數可以再被選進來，也沒有預測變數應該被剔除掉時即停止。

表 4.3 逐步替換法結果

No	(Intercept)	x1	x2	x3	x4	x5	x6	x7	x8	x9	rsq	adjR <sup>2</sup>	rss	Cp	BIC
1	1	1	0	0	0	0	0	0	0	0	0.0901	0.0804	40.2510	167.1931	0.0651
2	1	0	1	1	0	0	0	0	0	0	0.4545	0.4427	24.1319	65.3956	-44.4841
3	1	0	1	1	0	0	1	0	0	0	0.5412	0.5263	20.2939	42.6810	-56.5483
4	1	0	1	1	1	0	1	0	0	0	0.5862	0.5680	18.3058	31.8785	-61.8820
5	1	0	1	1	1	1	1	0	0	0	0.6397	0.6196	15.9403	18.6463	-70.6007
6	1	1	1	1	1	1	1	0	0	0	0.6717	0.6496	14.5213	11.5086	-74.9870
7	1	1	1	1	1	1	1	0	0	1	0.6928	0.6683	13.5907	7.5164	-76.7805
8	1	1	1	1	1	1	1	0	1	1	0.6960	0.6680	13.4483	8.5993	-73.2275

由上表，依據選取方法

1. 從調整  $R^2$  找出最大值為 0.6683 (模型編號 7)
2.  $p + 1$  最接近  $C_p$  值得為 8.5993 (模型編號 8)
3. 最小的 BIC 值為 -76.7805 (模型編號 7)

因此我們在模型 7 與模型 8 之間繼續挑選最佳模型，由調整  $R^2$  和 BIC 值我們更傾向於模型 7，且  $p + 1$  與  $C_p$  值之間的差值，兩組模型相差不大，所以向後選取法的最佳模型為模型編號 7，解釋變數包括  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_9$ 。

#### 四、選取變數

表 4.4 各選取法統整表

選取出的變數		
	模型	BIC
向前選取法	$Y \sim X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_9$	-76.7805
向後選取法	$Y \sim X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_9$	-76.7805
逐次替換法	$Y \sim X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_9$	-76.7805

分析三種選取法後，選出調整 $R^2$ 找出最大值、Cp 值最接近、BIC 值最小的最佳模型，為第七組模型，變數分別為 $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_9$ 。



## 第五章 選取後新複回歸模型

### 一、複回歸模型

利用前一章三種選取法選出的變數有 $X_1$ (出生人數)、 $X_2$ (外籍勞工人數)、 $X_3$ (消費者物價基本指數)、 $X_4$ (景氣同時指標綜合指數)、 $X_5$ (外銷訂單)、 $X_6$ (歇業家數)、 $X_9$ (所得稅)，其模型為：

$$\hat{Y} = -8.787 - 0.06489X_1 - 0.01381X_2 + 0.2295X_3 - 0.05492X_4 + 0.2724X_5 - 0.2042X_6 - 0.000001439X_9$$

### 二、模型適合度

經整理我們得到適合度檢定表，如下

表 5.1 選取後新適合度檢定表

來源	df	平方和	均方	F 值	Pr(>F)
Model	7	30.64697	4.378138	28.35119	< 2.2e-16
Error	88	13.58942	0.1544252		
SST	95	44.23638			

假設檢定：

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_9 = 0$$

$$H_1: \text{至少有一 } \beta_i \neq 0, \quad i = 1, 2, 3, 4, 5, 6, 9$$

檢定統計量  $F^* = 28.35119$

$$P\text{-value} = 2.2e - 16 < \alpha = 0.05$$

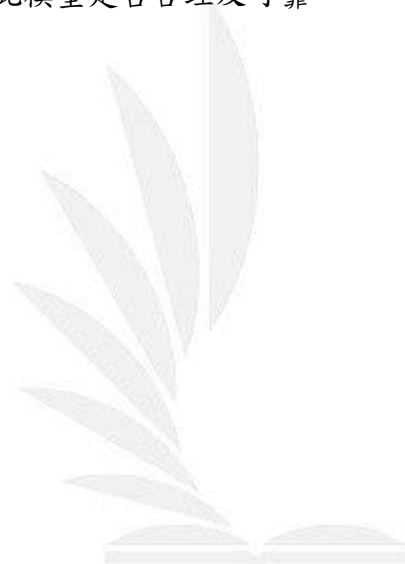
由  $P\text{-value} < \alpha = 0.05$ ，我們得出拒絕虛無假設 $H_0$ ，表示解釋變數與反應變數有線性關係。

### 三、模型解釋力

表 5.2 選取後新模型解釋力

解釋力			
標準化殘差值	0.393	調整 R 平方	0.6683
自由度	88	F 統計量	28.35119
判定係數 R 平方	0.6928	P 值	< 2.2e-16

模型的 $R^2$ 為 0.6928，而經過調整的 $R^2$ 為 0.6683，減輕因樣本估計所帶來的 $R^2$ 膨脹效果。表示解釋變數的變異可以被新模型所解釋的比例有 66.83%。下一步我們將進行殘差檢定，判斷此模型是否合理及可靠。





## 第陸章 殘差分析

在模型的基本假設之中，殘差富含了重要的信息。利用殘差的特性來檢測模型假設的合理性及數據的可靠性。

而再建立迴歸分析前，須先檢查資料是否符合迴歸分析的基本統計假設；倘若資料沒有符合迴歸分析的基本統計假設，則會造成統計推論偏誤的發生。

迴歸分析基本假設檢定分別有：

- 常態性
- 均齊性
- 獨立性

利用上述假設檢定來檢驗模型假設是否合理及可靠。

### 一、常態性

表 6.1 常態性檢定表

常態性檢定	
W 統計量	0.9885
P 值	0.5788

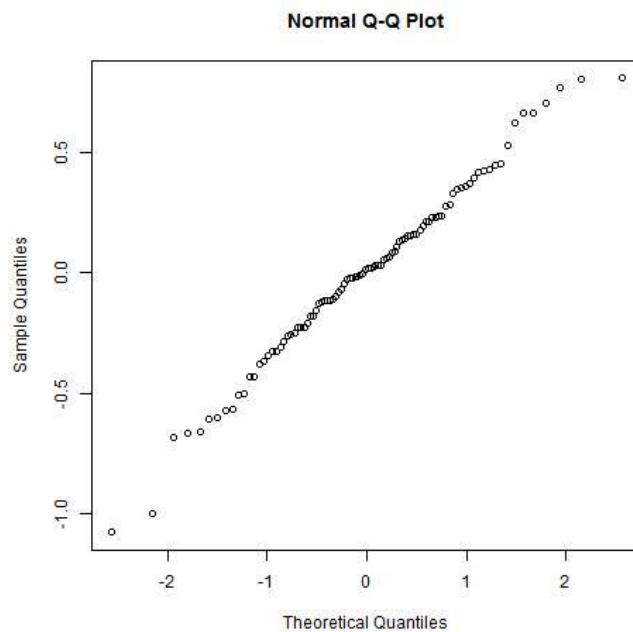


圖 6.1 Q-Q Plot

### 基本假設

$H_0$ ：殘差服從常態分配

$H_1$ ：殘差不服從常態分配

檢定統計量  $W = 0.9885$

P-Value檢定  $P - Value = 0.5788 > \alpha = 0.05$

因此不拒絕  $H_0$ ，顯示資料服從常態分配

且由 Q-Q Plot 中(圖 1)中，看出殘差和常態分位數的分布，散落在 45 度的直線上。所以由此推斷殘差也具有常態性。

### 二、均齊性

表 6.2 均齊性檢定表

均齊性檢定	
卡方值	5.132734
自由值	1
P 值	0.02347872

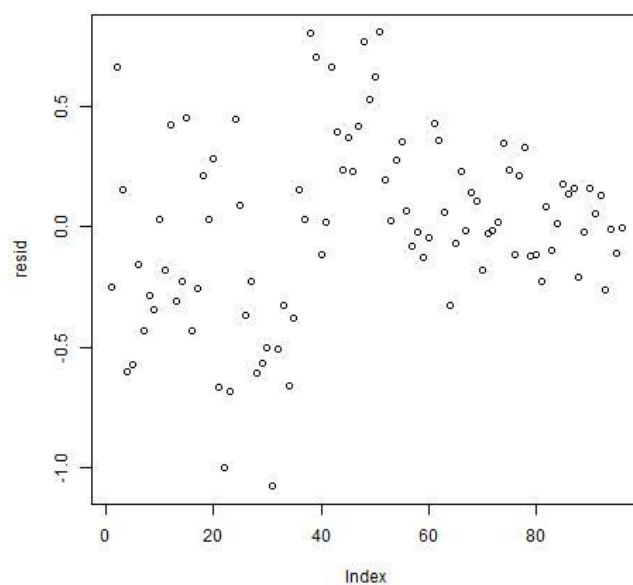


圖 6.2 殘差圖

假設

$H_0$ ：殘差變異數有均齊性

$H_1$ ：殘差變異數沒有均齊性

P-Value 檢定  $P - Value = 0.02347872 < \alpha = 0.05$

所以拒絕  $H_0$ ，表示資料顯示殘差變異數並不符合均齊性

但由殘差圖(圖 6.2)中，可發現殘差均勻分布在-1.0~0.5 之間，故我們可以推論殘差具有均齊性。

### 三、獨立性

表 6.3 獨立性檢定表

獨立性檢定	
D-W 統計量	0.9926741
P 值	0

基本假設

$H_0$ ：殘差之間互相獨立

$H_1$ ：殘差之間互相不獨立

獨立性的 P-Value 等於 0 時表示獨立性假定被違反。

由於上述的均齊性及獨立性皆不符合基本統計假設，因此需重新轉換模型。

### 第七章 轉換模型後的殘差分析

由於殘差分析的均齊性及獨立性沒有服從假設，因此我們將模型做轉換。我們通常對 Y 的轉換有兩種方法，一種是對 Y 取倒數另一種是對 Y 取對數 log，而我們這兩種方法都執行，然後比較兩種的調整R<sup>2</sup>來選擇使用哪一種轉換，如下：

表 7.1 兩種轉換解釋力比較

兩種轉換解釋力比較		
Y 轉換	轉換後R <sup>2</sup>	轉換後調整R <sup>2</sup>
$\frac{\hat{1}}{Y}$	0.686	0.661
$\widehat{\log(Y)}$	0.6902	0.6656

由上表(表 7.1)，我們可得知轉換後的R<sup>2</sup>及轉換後調整R<sup>2</sup>皆是 $\widehat{\log(Y)}$ 比較大，所以我們選用對 Y 取對數 log 來做轉換，得到我們轉換後的模型，如下：

$$\widehat{\log(Y)} = -1.435 - 0.01395X_1 - 0.02903X_2 + 0.04939X_3 - 0.01109X_4 + 0.05567X_5 - 0.04189X_6 - 0.0000003046X_9$$

經轉換模型後我們再檢定一次殘差分析，檢查是否符合殘差檢定。

### 一、常態性

表 7.2 轉換後常態性檢定表

常態性檢定	
W 統計量	0.9867
P 值	0.4489

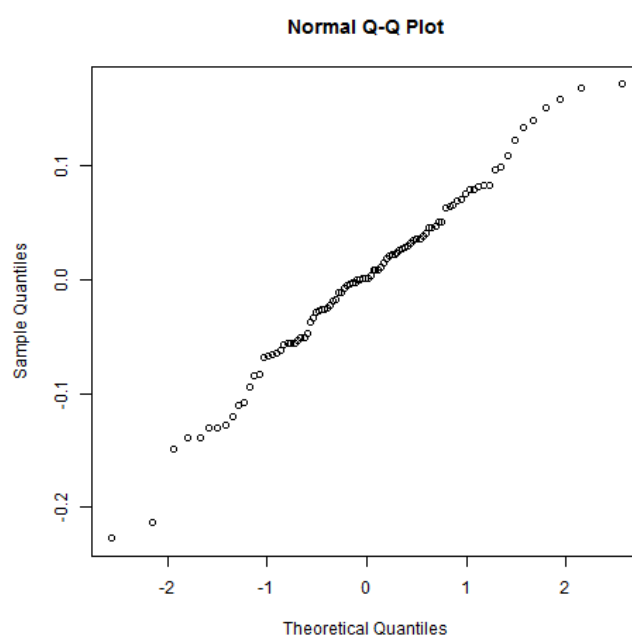


圖 7.1 轉換後 Q-Q Plot

#### 基本假設

$H_0$ ：殘差服從常態分配

$H_1$ ：殘差不服從常態分配

檢定統計量  $W = 0.9867$

P-Value檢定  $P - Value = 0.4489 > \alpha = 0.05$

因此不拒絕 $H_0$ ，顯示資料服從常態分配。

且由 Q-Q Plot 中(圖 7.1)，看出殘差和常態分位數的分布，散落在 45 度的直線上。所以由此推斷殘差也具有常態性。

## 二、均齊性

表 7.3 轉換後均齊性檢定

均齊性檢定	
卡方值	3.609447
自由值	1
P 值	0.05745221

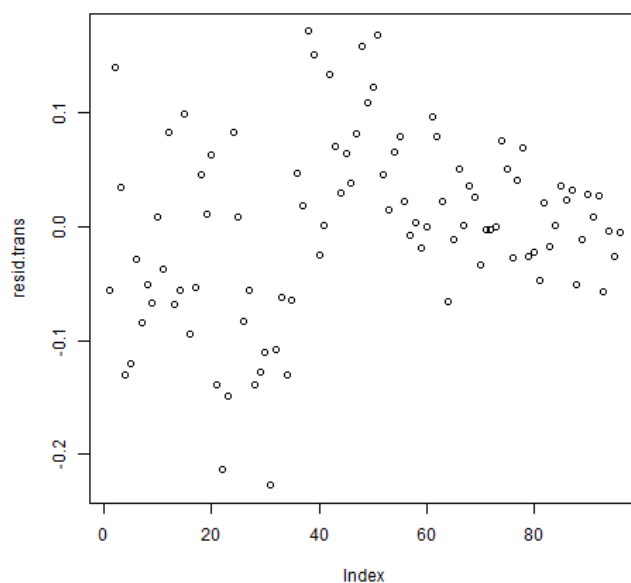


圖 7.2 轉換後殘差圖

基本假設

$H_0$ ：殘差變異數有均齊性

$H_1$ ：殘差變異數沒有均齊性

P-Value 檢定  $P - Value = 0.05745221 > \alpha = 0.05$

所以不拒絕 $H_0$ ，表示資料顯示殘差變異數有均齊性。

且由殘差圖中(圖 7.2)，可發現殘差均勻分布在-0.2~0.1 之間，故我們可以推論殘差具有均齊性。

### 三、獨立性

表 7.4 轉換後獨立性檢定

獨立性檢定	
D-W 統計量	1.005478
P 值	0

#### 基本假設

$H_0$ ：殘差之間互相獨立

$H_1$ ：殘差之間互相不獨立

獨立性的 p-value = 0 → 獨立性假定被違反。

當獨立性假設被違反時，則表示殘差值彼此俱有時間性的關聯。

由於我們所擁有的資料與時間有相關性，以及知識技術的局限性，我們無法將獨立性假設的 p-value 值調整至 0.05 以上。但排除獨立性假設的問題，我們所做的迴歸分析過程皆符合本課程的上課內容，藉由上課所學的專業知識，我們利用對 Y 的轉換在進行模型擬合的過程縮小了不可觀測的誤差和預測變量的相關性，所以，此新模型即為我們能力範圍內所可以完成的最佳模型。

## 第捌章 轉換後新複回歸模型

### 一、複回歸模型

利用第四章選出的變數有 $X_1$ (出生人數)、 $X_2$ (外籍勞工人數)、 $X_3$ (消費者物價基本指數)、 $X_4$ (景氣同時指標綜合指數)、 $X_5$ (外銷訂單)、 $X_6$ (歇業家數)、 $X_9$ (所得稅)，經過殘差分析後發現模型需要轉換，經轉換後又再次執行殘差分析，得到最後模型為：

$$\widehat{\log(Y)} = -1.435 - 0.01395X_1 - 0.02903X_2 + 0.04939X_3 - 0.01109X_4 + 0.05567X_5 - 0.04189X_6 - 0.0000003046X_9$$

### 二、模型適合度

經過我們整理得到適合度檢定表，如下：

表 8.1 轉換後新適合度檢定表

來源	df	平方和	均方	F 值	Pr(>F)
Model	7	30.53195	4.361707	28.00775	< 2.2e-16
Error	88	13.70443	0.1557322		
SST	95	44.23638			

假設檢定：

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_9 = 0$$

$$H_1: \text{至少有一 } -\beta_i \neq 0, \quad i = 1, 2, 3, 4, 5, 6, 9$$

檢定統計量  $F^* = 28.00775$

$$P\text{-value} = 2.2e - 16 < \alpha = 0.05$$

由  $P\text{-value} < \alpha = 0.05$ ，我們得出拒絕虛無假設 $H_0$ ，表示解釋變數與反應變數有線性關係。



### 三、模型解釋力

表 8.2 轉換後新模型解釋力

解釋力			
標準化殘差值	0.08237	調整 R 平方	0.6656
自由度	88	F 統計量	28.00775
判定係數 R 平方	0.6902	P 值	< 2.2e-16

全模型的 $R^2$ 為 0.6902，而經過校正的 $R^2$ 為 0.6656，減輕因樣本估計所帶來的 $R^2$ 膨脹效果。表示解釋變數的變異可以被新模型所解釋的比例有 66.56%。因此我們最後模型的解釋力為 66.56%。

### 四、參數估計

模型轉換後，再次判斷被選取到的變數對 Y(失業率)是否有線性相關。

表 8.3 轉換後參數估計表

參數估計表					
變數	標籤	參數估計值	標準誤差	t 值	Pr(> t )
截距	截距	-1.435	0.4968	-2.888	0.00487
$X_1$	出生人數(千)	-0.01395	0.004128	-3.380	0.00108
$X_2$	外籍勞工人數(千)	-0.002903	0.0005177	-5.607	2.34e-07
$X_3$	消費者物價基本指數	0.04939	0.006308	7.829	1.03e-12
$X_4$	景氣同時指標綜合指數(點)	-0.01109	0.001893	-5.859	7.92e-07
$X_5$	外銷訂單(千)	0.05567	0.01219	4.568	1.59e-05
$X_6$	歇業家數(千)	-0.04189	0.007366	-5.688	1.66e-07
$X_9$	所得稅(百萬)	-3.046e-07	1.229e-07	-2.479	0.01509

- (1) 想要判斷 $X_1$ (出生人數)和Y(失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=0.000934 <  $\alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_1 \neq 0$ 。因此， $X_1$ (出生人數)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關。

- (2) 想要判斷 $X_2$ (外籍勞工人數)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=2.62e-07 <  $\alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_2 \neq 0$ 。因此， $X_2$ (外籍勞工人數)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關。

- (3) 想要判斷 $X_3$ (消費者物價基本指數)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

在虛無假設之下的拒絕域為：P-value <  $\alpha=0.05$

檢定如下：因為 P-value=4.74e-12 <  $\alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_3 \neq 0$ 。因此， $X_3$ (消費者物價基本指數)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關。

- (4) 想要判斷 $X_4$ (景氣同時指標綜合指數)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha=0.05$

檢定如下：因為  $P\text{-value}=2.28e-07 < \alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_4 \neq 0$ 。因此， $X_4$ (景氣同時指標綜合指數)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關

- (5) 想要判斷 $X_5$ (外銷訂單)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha=0.05$

檢定如下：因為  $P\text{-value}=0.006181 < \alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_5 \neq 0$ 。因此， $X_5$ (外銷訂單)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關

- (6) 想要判斷 $X_6$ (歇業家數)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha=0.05$

檢定如下：因為  $P\text{-value}=3.18e-07 < \alpha=0.05$  所以拒絕虛無假設 $H_0$ ，表

示我們有充分的證據顯示 $\beta_6 \neq 0$ 。因此， $X_6$ (歇業家數)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關

(7) 想要判斷 $X_9$ (所得稅)和 $Y$ (失業率)之間是否存在線性相關，首先我們先假設其他變數為固定的情況下

統計基本假設

$$H_0 : \beta_9 = 0$$

$$H_1 : \beta_9 \neq 0$$

在虛無假設之下的拒絕域為： $P\text{-value} < \alpha = 0.05$

檢定如下：因為  $P\text{-value} = 0.015107 < \alpha = 0.05$  所以拒絕虛無假設 $H_0$ ，表示我們有充分的證據顯示 $\beta_9 \neq 0$ 。因此， $X_9$ (所得稅)是一個顯著的解釋變數和 $Y$ (失業率)存在線性相關

## 五、共線性

最後我們想解釋變數間是否有嚴重的共線性問題

表 8.4 共線性

	標籤	VIF
$X_1$	出生人數	1.258136
$X_2$	外籍勞工人數	7.587294
$X_3$	消費者物價基本指數	5.531432
$X_4$	景氣同時指標綜合指數	5.453839
$X_5$	外銷訂單	3.806348
$X_6$	歇業家數	1.273499
$X_9$	所得稅	1.029317
	平均	3.705695

- 判斷共線性的標準： $VIF > 10$ 表示可能有共線性問題

依據上表，以整體來看這個模型的共線性問題不是很嚴重。

### 第玖章 結論

我們這個專題主要是想瞭解Y(失業率)和X<sub>1</sub>(出生人數)、X<sub>2</sub>(外籍勞工人數)、X<sub>3</sub>(消費者物價基本指數)、X<sub>4</sub>(景氣同時指標綜合指數)、X<sub>5</sub>(外銷訂單)、X<sub>6</sub>(歇業家數)、X<sub>9</sub>(所得稅)之間的相關程度影響，我們採用了模型選取，模型檢定等方法進行了相關分析與殘差和適合度檢定，再使用向前選取法、向後選取法、逐次替換法選取出了一個適當的複回歸模型

$$\widehat{\log(Y)} = -1.435 - 0.01395X_1 - 0.02903X_2 + 0.04939X_3 - 0.01109X_4 + 0.05567X_5 - 0.04189X_6 - 0.0000003046X_9$$

通過這個模型我們可以明確的看出X<sub>1</sub>(出生人數)、X<sub>2</sub>(外籍勞工人數)、X<sub>3</sub>(消費者物價基本指數)、X<sub>4</sub>(景氣同時指標綜合指數)、X<sub>5</sub>(外銷訂單)、X<sub>6</sub>(歇業家數)、X<sub>9</sub>(所得稅)和Y(失業率)存在顯著相關的關係，其中X<sub>1</sub>(出生人數)、X<sub>2</sub>(外籍勞工人數)、X<sub>4</sub>(景氣同時指標綜合指數)、X<sub>6</sub>(歇業家數)、X<sub>9</sub>(所得稅)和Y(失業率)存在負相關關係，表示每增加一單位的出生人數，就會減少0.01395單位的失業率，每增加一單位的外籍勞工人數，就會減少0.02903單位的失業率，每增加一單位的景氣同時指標綜合指數，就會減少0.01109單位的失業率，每增加一單位的歇業家數，就會減少0.04189單位的失業率，每增加一單位的所得稅，就會減少0.0000003046單位的失業率。其中X<sub>3</sub>(消費者物價基本指數)、X<sub>5</sub>(外銷訂單)和Y(失業率)存在正相關，表示每增加一單位的消費者物價基本指數，就會增加0.04939單位的失業率；每增加一單位的外銷訂單，就會增加0.05567單位的失業率。而我們七個變數的解釋力 66.56%雖然數值並不高，但近年的失業率更多影響取決於人們是否想工作的心理，不同家境和不同教育程度的人有著不同的想法，從而造成主觀性失業，而我們的變數只能討論客觀性失業，所以從現實生活中，若要改善失業率，需要所有年輕人共同努力，提升自我競爭力及端正工作態度，而政府方面需要多給一些工作資訊及資源，為整個社會提供一個良好的就業環境。

第壹拾章 附錄

一、原始檔案

Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
3.8	16.477	327.311	93.12	79.57	7.09	2.326	82.087	76.37	24823.01
3.92	16.264	331.741	92.2	79.91	6.549	4.132	39.972	68.32	25084.54
3.87	17.524	333.189	91.94	80.11	8.09	3.099	39.07	82.43	23647.34
3.78	15.126	333.593	93.15	80.21	7.962	1.772	39.422	78.33	16436.05
3.84	16.757	333.477	94	80.24	7.8	1.381	40.349	81.06	84102.16
3.98	16.558	336.945	94.43	80.24	8.08	2.386	39.563	78.99	228589.6
4.05	16.34	336.779	94.64	80.19	7.841	2.295	39.78	78.7	2304.53
4.09	17.311	336.462	94.07	80.24	8.582	2.564	41.255	81.44	44529.37
3.96	16.864	336.864	93.56	80.35	8.751	2.409	39.93	79.31	79183.55
3.9	18.927	336.985	93.53	80.56	9.05	4.224	39.578	79.57	73769.77
3.86	18.957	338.041	93.45	80.86	8.919	3.359	38.812	79.76	19891.2
3.81	17.354	338.755	93.65	81.18	8.568	6.671	42.476	78.95	23856.71
3.79	18.224	341.871	93.45	81.56	8.483	2.262	48.929	80.5	19758.87
3.78	13.496	341.623	93.81	82.02	7.215	2.841	74.807	65.86	18510.22
3.94	18.49	345.812	92.72	82.6	9.239	4.718	39.619	82.36	34755.81
3.83	15.448	347.172	93.79	83.31	9.09	2.691	39.455	81.25	20884.16
3.87	16.463	348.67	93.98	84.12	9.145	2.131	40.504	86.94	146361.1
3.96	14.687	350.955	94.55	84.96	9.442	4.677	40.835	84.97	221383.4
4.03	16.737	351.215	94.33	85.84	9.734	4.251	41.524	88.7	118.94
4.09	18.109	352.297	95.58	86.68	10.034	5.945	41.463	89.05	43036.5
3.99	14.863	355.756	96.47	87.52	10.188	3.182	42.668	85.21	96125.79
3.92	19.233	356.51	98.52	88.39	10.465	3.027	39.728	92.23	77785.18
3.87	19.311	357.705	97.94	89.24	10.295	3.996	39.587	89.93	24968.34
3.83	19.353	357.937	96.77	90.09	10.042	7.81	43.951	89.45	26471.68
3.8	18.323	359.103	96.2	90.8	9.81	5.298	65.74	90.89	28852.59
3.94	16.065	362.782	97.43	91.24	8.165	1.907	59.608	76.63	26607.41
3.86	16.114	365.229	96.38	91.37	9.656	3.577	40.273	90.71	26823.12
3.81	15.549	367.119	97.42	91.17	9.623	3.239	40.192	90.71	20114.51
3.84	15.309	369.406	97.47	90.47	9.623	2.877	40.554	92.06	101831.2
3.95	14.429	372.097	99.25	89.15	9.506	3.342	41.18	89.37	332543.8
4.06	16.063	374.147	99.81	87.08	9.516	2.962	42.69	90.6	-14459.8
4.14	14.866	373.336	100.1	84.3	10.009	7.064	41.415	90.46	45160.89



回歸分析-分析失業率的影響因素

4.27	16.048	369.73	99.46	80.72	10.148	7.106	41.313	85.09	125594.7
4.37	18.74	373.19	100.9	76.49	9.921	6.564	39.538	81.02	85714.68
4.64	20.388	373.276	99.83	72.22	7.535	3.321	38.138	65.38	26514.21
5.03	16.839	365.06	97.99	68.83	6.882	3.643	41.868	61.86	29690.97
5.31	13.469	355.743	97.62	66.96	5.87	2.174	68.084	54.21	23624.65
5.75	17.159	349.258	96.13	66.76	6.882	2.763	41.151	58.45	29482.64
5.81	20.773	343.227	96.24	67.74	8.205	2.932	38.58	69.83	18816.41
5.76	10.597	344.372	96.97	69.59	8.454	2.968	38.904	73.76	18210.86
5.82	12.646	341.943	97.38	71.96	8.265	2.016	39.722	75.41	74114.74
5.94	16.157	341.484	97.29	74.62	9.144	2.584	38.21	80.45	258005
6.07	16.297	343.228	97.48	77.22	9.403	2.313	39.644	84	-24561.6
6.13	14.699	344.599	99.24	79.58	9.253	2.354	39.522	81.25	36660.39
6.04	16.17	345.755	98.58	81.75	10.038	2.074	40.534	85.89	117071.3
5.96	17.574	347.618	98.96	83.66	10.244	1.931	39.816	86.37	42563.52
5.86	17.556	349.433	98.22	85.39	10.09	1.877	39.035	84.6	19576.3
5.74	18.213	351.016	97.75	86.99	10.225	3.022	42.768	91.13	27402.57
5.68	15.603	353.805	97.87	88.49	9.667	1.811	53.739	88.22	21514.65
5.76	12.364	355.136	98.39	89.96	8.775	1.811	67.177	76.16	23122.91
5.67	15.75	359.583	97.45	91.29	10.939	3.203	40.701	94.97	30842.41
5.39	13.371	362.118	98.27	92.45	10.68	2.513	40.207	95.05	17669.77
5.14	12.902	364.789	98.12	93.49	10.733	1.742	41.074	96.88	122835.7
5.16	12.722	367.72	98.45	94.45	11.009	2.847	41.519	97.42	175501.9
5.2	12.668	370.703	98.76	95.41	10.853	4.121	41.485	100.57	-21116.3
5.17	13.761	372.146	98.77	96.39	11.112	2.025	40.322	99.45	48848.13
5.05	13.637	372.785	98.87	97.42	11.405	1.969	42.981	96.17	68734.46
4.92	13.487	376.008	99.52	98.51	10.976	2.657	39.954	98.15	57371.28
4.73	15.406	378.083	99.71	99.62	10.822	2.295	39.553	100.15	20677.52
4.67	15.215	379.653	98.96	100.64	10.922	1.937	44.274	105.83	24385.16
4.64	14.703	383.164	98.95	101.39	10.014	2.919	82.7	103.68	21583.66
4.69	12.27	387.106	99.69	101.84	8.451	1.432	45.723	85.43	28040.14
4.48	17.524	390.423	98.82	101.97	11.485	2.202	40.747	107.42	19090.8
4.29	15.089	395.944	99.56	101.83	10.814	1.627	40.632	102.95	15834.51
4.27	16.206	399.749	99.75	101.52	10.788	1.89	41.948	104.78	121331.7
4.35	15.866	403.492	100.4	101.08	10.749	2.644	41.583	102.28	253079.9
4.41	15.887	409.434	100.1	100.56	10.814	2.601	42.471	103.72	-26803.7
4.45	16.797	412.933	100.1	99.99	10.615	1.923	42.194	102.84	67418.85
4.28	18.003	417.844	100.2	99.45	10.958	2.473	43.105	96.76	88528.66

回歸分析-分析失業率的影響因素

4.3	17.136	420.931	100.8	99.08	11.24	2.223	40.445	98.56	73900.27
4.28	18.443	423.338	100.7	98.84	11.058	2.644	39.7	95.12	19617.53
4.18	18.703	425.66	101	98.77	10.967	3.055	45.446	96.46	28568.75
4.18	15.296	426.378	101.3	98.9	9.442	2.583	83.392	86.29	19995.16
4.25	17.5	429.824	99.94	99.18	10.013	2.025	41.932	92.63	32861.52
4.17	19.283	433.956	100.1	99.64	11.308	2.823	41.161	104.19	22096.91
4.1	17.405	434.325	101	100.04	10.617	2.058	41.262	100.49	17900.89
4.12	18.542	436.593	101.5	100.31	10.731	2.846	41.706	104.6	138170.6
4.21	19.482	439.98	102.1	100.44	10.867	2.359	42.653	99.88	278422.2
4.31	16.897	441.507	102.5	100.6	10.756	2.684	42.177	103.68	-16157.7
4.4	19.615	440.382	103.5	100.74	10.813	2.12	41.515	104.4	59393.21
4.32	18.908	442.581	103.2	100.85	11.103	1.871	43.557	99.12	78632.81
4.33	23.398	443.809	103.1	100.93	11.211	1.892	40.685	102.45	79013.94
4.27	21.953	444.604	102.3	100.96	11.843	2.046	39.805	100.79	21925.77
4.18	21.202	445.579	102.6	101	11.428	2.892	47.444	98.49	28554.31
4.16	21.982	445.301	102.4	101.1	10.784	2.523	60.005	102.74	21613.45
4.24	15.666	448.017	102.9	101.16	8.581	1.307	62.283	82.23	27219.92
4.17	16.88	451.202	101.4	101.23	10.641	2.272	41.394	100.84	23555.16
4.07	15.506	454.171	102.1	101.39	10.628	1.852	41.695	99.56	18970.11
4.06	16.223	459.424	102.2	101.63	10.797	1.839	42.166	104.15	128789.9
4.14	13.963	462.658	102.8	101.87	10.493	1.964	42.619	99.36	267684.5
4.25	16.221	465.581	102.6	102.04	10.806	2.159	43.04	105.83	-14658.5
4.33	15.318	469.199	102.7	102.26	10.863	1.907	42.241	103.39	49378.72
4.24	15.75	475.006	104.1	102.6	11.384	1.754	43.106	98.91	85745.55
4.24	17.521	478.331	103.8	103.01	11.624	2.117	41.952	103.29	80763.54
4.16	16.632	484.367	103	103.48	12.096	1.978	41.362	100.57	23532.94
4.08	17.451	489.134	102.9	103.93	12.559	2.651	46.33	103.96	30694.29



## 二、R 程式碼

```
#輸入數據
data=read.csv("C:/Users/Ting-Ru Liu/Desktop/回歸分析
/newdata.csv",head=T)
data

#檢視資料類型
str(data)

#敘述性統計
#基本統計
y=c(length(data$y),"失業率
(%)",round(mean(data$y),5),round(sd(data$y),5),sum(data$y),min(da
ta$y),max(data$y))
x1=c(length(data$x1),"出生人數
(千)",round(mean(data$x1),5),round(sd(data$x1),5),sum(data$x1),min(da
ta$x1),max(data$x1))
x2=c(length(data$x2),"外籍勞工人數
(千)",round(mean(data$x2),5),round(sd(data$x2),5),sum(data$x2),min(da
ta$x2),max(data$x2))
x3=c(length(data$x3),"消費者物價基本指數
",round(mean(data$x3),5),round(sd(data$x3),5),sum(data$x3),min(data$x
3),max(data$x3))
x4=c(length(data$x4),"景氣同時指標綜合指數
(點)",round(mean(data$x4),5),round(sd(data$x4),5),sum(data$x4),min(da
ta$x4),max(data$x4))
x5=c(length(data$x5),"外銷訂單
(千)",round(mean(data$x5),5),round(sd(data$x5),5),sum(data$x5),min(da
ta$x5),max(data$x5))
x6=c(length(data$x6),"歇業家數
(千)",round(mean(data$x6),5),round(sd(data$x6),5),sum(data$x6),min(da
ta$x6),max(data$x6))
x7=c(length(data$x7),"平均薪資
(千)",round(mean(data$x7),5),round(sd(data$x7),5),sum(data$x7),min(da
ta$x7),max(data$x7))
x8=c(length(data$x8),"工業生產指數
",round(mean(data$x8),5),round(sd(data$x8),5),sum(data$x8),min(data$x
```

```
8), max(data$x8))
x9=c(length(data$x9), "所得稅(百
萬)", round(mean(data$x9), 5), round(sd(data$x9), 5), sum(data$x9), min(dat
a$x9), max(data$x9))
a=data.frame(rbind(y, x1, x2, x3, x4, x5, x6, x7, x8, x9))
colnames(a)=c("樣本數", "標籤", "平均數", "標準差", "總和", "最小值", "最大
值");a
library(xtable)
print(xtable(a), type="html")

#建立模型
#複回歸模型(丟入所有變數)
model=lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9, data=data, x=T)
model

library(xtable)
#ANOVA 表
anova(model)
print(xtable(anova(model)), type="HTML")

#適合度表
fit=matrix(c("來源", "模型", "誤差", "Residual standard
error", "df", 9, 86, 95, "平方和", 30.88118, 13.35525, 44.23638, "均方
", 3.431243, 0.1553148, "", "F 值
", 22.09218, "", "", "Pr>F", 2.2e-16, "", ""), nrow=4, ncol=6)
print(xtable(fit), type="HTML")

#參數估計表
summary(model)
print(xtable(summary(model)), type="HTML")
aov=matrix(c("Residual standard error", "df", "Multiple
R-squared", 0.3941, 86, 0.6981, "Adjusted
R-squared", "F-statistic", "p-value", 0.6665, 22.1, "<2.2e-16"), nrow=3, nco
l=4)
print(xtable(aov), type="HTML")

#選取變數
library(leaps)
```

#向前選取法

```
out.forward=regsubsets(as.matrix(data[-1]),  
                        y=data$y,  
                        nbest=1,  
                        method="forward")
```

out.forward

```
s.forward=summary(out.forward)
```

s.forward

```
forward=round(cbind(s.forward$which,  
                    rsq=s.forward$rsq,  
                    adjr2=s.forward$adjr2,  
                    rss=s.forward$rss,  
                    cp=s.forward$cp,  
                    bic=s.forward$bic),4)
```

forward

```
print(xtable(forward), type="html")
```

#向後選取法

```
out.backward=regsubsets(as.matrix(data[-1]),  
                        y=data$y,  
                        nbest=1,  
                        method="backward")
```

out.backward

```
s.back=summary(out.backward)
```

s.back

```
backward=round(cbind(s.back$which,  
                     rsq=s.back$rsq,  
                     adjr2=s.back$adjr2,  
                     rss=s.back$rss,  
                     cp=s.back$cp,  
                     bic=s.back$bic),4)
```

backward

```
print(xtable(backward), type="html")
```

#逐次替換法

```
out.step=regsubsets(as.matrix(data[-1]),  
                    y=data$y,  
                    nbest=1,
```

```
method=" seqrep" )  
out. step  
s. step=summary(out. step)  
s. step  
step=round(cbind(s. step$which,  
                rsq=s. step$rsq,  
                adjr2=s. step$adjr2,  
                rss=s. step$rss,  
                cp=s. step$cp,  
                bic=s. step$bic), 4)  
step  
print(xtable(step), type="html" )  
  
library(xtable)  
method=matrix(c("y~x1+x2+x3+x4+x5+x6+x9", "y~x1+x2+x3+x4+x5+x6+x9", "y~  
x1+x2+x3+x4+x5+x6+x9", "y~x1+x2+x3+x4+x5+x6+x9", "-76.7805", "-76.7805",  
"-76.7805", "-76.7805 " ), nrow=4, ncol=2)  
rownames(method)=c("向前選取法", "向後選取法", "所有可能模型選取法", "逐  
次替換法")  
colnames(method)=c("模型", "BIC" )  
print(xtable(method), type="html" )  
  
#挑出選出的變數→新模型  
model. new=lm(y~x1+x2+x3+x4+x5+x6+x9, data=data)  
model. new  
  
#ANOVA 表  
anova(model. new)  
print(xtable(anova(model)), type="HTML" )  
  
#適合度表  
fit. new=matrix(c("來源", "模型", "誤差", "Residual standard  
error", "df", 7, 88, 95, "平方和", 30.64697, 13.58942, 44.23638, "均方  
, 4.378138, 0.1544252, "", "F 值  
, 28.35119, "", "", "Pr(>F)", 2.2e-16, "", "" ), nrow=4, ncol=6)  
print(xtable(fit. new), type="HTML" )  
  
#參數估計表
```

```
summary(model.new)
print(xtable(summary(model.new)), type="HTML")
aov.new=matrix(c("Residual standard error", "df", "Multiple
R-squared", 0.393, 88, 0.6928, "Adjusted
R-squared", "F-statistic", "p-value", 0.6683, 28.35119, "<2.2e-16"), nrow=3
, ncol=4)
print(xtable(aov.new), type="HTML")

#殘差分析
library(car)
resid=model.new$residuals
#常態性(H0:假設殘差服從常態分配)
shapiro.test(resid)
shap.t_new=matrix(c(0.9885, 0.5788), nrow=2, ncol=1)
rownames(shap.t_new)=c("W", "p-value")
colnames(shap.t_new)=c("new Shapiro-Wilk normality test")
print(xtable(shap.t_new), type="html")
#均齊性(H0:假設殘差變異數有均齊性)
ncvTest(model.new)
ncv.t_new=matrix(c(5.132734, 1, 0.02347872), nrow=3, ncol=1)
rownames(ncv.t_new)=c("Chisquare", "df", "p-value")
colnames(ncv.t_new)=c("new ncv test")
print(xtable(ncv.t_new), type="html")
#獨立性(H0:殘差之間互獨立)
durbinWatsonTest(model.new)
dw.t_new=matrix(c(0.9926741, 0), nrow=2, ncol=1)
rownames(dw.t_new)=c("D-W statistic", "p-value")
colnames(dw.t_new)=c("new D-W test")
print(xtable(dw.t_new), type="html")

#殘差圖形
path="C:/Users/Ting-Ru Liu/Desktop/回歸分析/殘差圖形/"
#殘差圖
png(paste(path, "殘差圖(原).png", sep=""))
plot(resid)
dev.off()
#qq圖
png(paste(path, "qq圖(原).png", sep=""))
```

## 回歸分析-分析失業率的影響因素

```
qqnorm(resid)
dev.off()
#盒形圖
png(paste(path, "盒型圖(原).png", sep=""))
boxplot(resid)
dev.off()
#莖葉圖
stem(resid)

#模型轉換
model.trans=lm(log(y)~x1+x2+x3+x4+x5+x6+x9, data=data, x=T)

#轉換後殘差分析
resid.trans=model.trans$residuals
#常態性(H0:假設殘差服從常態分配)
shapiro.test(resid.trans)
shap.t_new=matrix(c(0.9867, 0.4489), nrow=2, ncol=1)
rownames(shap.t_new)=c("W", "p-value")
colnames(shap.t_new)=c("new Shapiro-Wilk normality test")
print(xtable(shap.t_new), type="html")
#均齊性(H0:假設殘差變異數有均齊性)
ncvTest(model.trans)
ncv.t_new=matrix(c(3.609447, 1, 0.05745221), nrow=3, ncol=1)
rownames(ncv.t_new)=c("Chisquare", "df", "p-value")
colnames(ncv.t_new)=c("new ncv test")
print(xtable(ncv.t_new), type="html")
#獨立性(H0:殘差之間互獨立)
durbinWatsonTest(model.trans)
dw.t_new=matrix(c(1.005478, 0), nrow=2, ncol=1)
rownames(dw.t_new)=c("D-W statistic", "p-value")
colnames(dw.t_new)=c("new D-W test")
print(xtable(dw.t_new), type="html")

#殘差圖形
path="C:/Users/Ting-Ru Liu/Desktop/回歸分析/殘差圖形/"
#殘差圖
png(paste(path, "殘差圖.png", sep=""))
plot(resid.trans)
```

```
dev.off()
#qq 圖
png(paste(path, "qq 圖.png", sep=""))
qqnorm(resid.trans)
dev.off()
#盒形圖
png(paste(path, "盒型圖.png", sep=""))
boxplot(resid.trans)
dev.off()
#莖葉圖
stem(resid.trans)

#新模型適合度檢定&解釋力
anova(model.trans)
fit=matrix(c("來源", "模型", "誤差", "Residual standard
error", "df", 7, 88, 95, "平方和", 30.53195, 13.70443, 44.23638, "均方
", 4.361707, 0.1557322, "", "F 值
", 28.00775, "", "", "Pr>F", "2.2e-16", "", ""), nrow=4, ncol=6)
print(xtable(fit), type="HTML")
summary(model.trans)
aov=matrix(c("Residual standard error", "Multiple R-squared", "Adjusted
R-squared", 0.08237, 0.6902, 0.6656, "F-statistic", "p-value", "", 28.00775,
"<2.2e-16", ""), nrow=3, ncol=4)
print(xtable(summary(model.trans)), type="HTML")

#共線性
vif(model.trans)
mean(vif(model.trans))
v=matrix(c("出生人數", "外籍勞工人數", "消費者物價基本指數", "景氣同時指
標綜合指數", "外銷訂單", "歇業家數", "所得稅", "平均
", 1.258136, 70587294, 5.531432, 5.453839, 3.806348, 1.273499, 1.029317, 3.70
5695), nrow=8, ncol=2)
colnames(v)=c("標籤", "VIF")
rownames(v)=c("x1", "x2", "x3", "x4", "x5", "x6", "x9", "")
print(xtable(v), type="html")
```

### 三、工作分配

報告內容	組員
摘要	張巧穎 劉庭汝
流程圖、基本敘述統計	董心成
複回歸模型 (原始、選取後、轉換後)	管奕錚 劉庭汝
變數選取法 (向前、向後、逐次替換)	洪慈炆 黃姿穎
殘差分析 (常態性、均齊性、獨立性)	張巧穎 鍾宜頻
動機、結論	管奕錚
R 程式、報告統整	劉庭汝

### 四、參考文獻

1. R 軟體-應用統計方法
2. 逢甲大學優質報告-<http://epaper.lib.fcu.edu.tw/event.htm>
3. 總體統計資料庫-<http://ebas1.ebas.gov.tw/pxweb/Dialog/statfile9L.asp>