

應用迴歸分析期末報告

姓名:謝博丞

系級:統計所碩一

學號:110354023

目錄

(一)資料.....	3
資料介紹.....	3
資料視覺化.....	5
(二)建模與處理.....	8
1. 遺失值處理.....	8
Mice	9
直接刪除.....	9
2. 初步模型.....	9
3. 共線性診斷.....	9
4. 挑選變數.....	10
向前選取法	10
向後選取法	10
逐步選取法	10
5. 殘差分析.....	10
變數變換.....	11
Box-Cox 轉換.....	13
Weighted least square.....	15
6. 最終模型.....	17
(三)程式碼.....	19

(一)資料

資料介紹

以年度離婚對數作為應變數，解釋變數則以當年度之國民教育程度、房價，以及前一年之人口消長指標等直覺認為與今年度離婚數有關之變數做挑選；相較於110年，該平台於109年之年度行駐區資料較為完整，故選用109年我行政區離婚對數作為應變數。

空間範圍：全國

空間統計單元：鄉鎮市區

樣本數：358

缺失值：17

變數介紹：

變數名稱	介紹
離婚對數（應變數）	行政區戶籍離婚登記對數
粗結婚率	期間結婚對數對同一期間之總人口數比率
X107 自然增加率	107 年期間中自然增加數對年中人口數的比率，或粗出生率減粗死亡率之差。自然增加率=粗出生率－粗死亡率
X107 社會增加率	107 年期間中社會增加數對年中人口數的比率，或遷入率減遷出率之差。社會增加率=遷入率－遷出率
X107 粗結婚率	107 年期間結婚對數對同一期間之總人口數比率
X107 粗離婚率	107 年期間中社會增加數對年中人口數的比率，或遷入率減遷出率之差。社會增加率=遷入率－遷出率
X108 自然增加率	108 年期間中自然增加數對年中人口數的比率，或粗出生率減粗死亡率之差。自然增加率=粗出生率－粗死亡率
X108 社會增加率	108 年期間中社會增加數對年中人口數的比率，或遷入率減遷出率之差。社會增加率=遷入率－遷出率
X108 粗結婚率	108 年期間結婚對數對同一期間之總人口數比率
X108 粗離婚率	108 年期間離婚對數對同一期間之總人口數比率
性比例	指男性人口對女性人口的比例，亦即每百個女子所對男子數。
扶養比	$(0-14 \text{ 歲人口} + 65 \text{ 歲以上人口}) / (15-64 \text{ 歲人口}) \times 100$

老化指數	$(65 \text{ 歲以上人口}) / (0-14 \text{ 歲人口}) \times 100$
大學院校人口數	戶籍登記實足年齡滿 15 歲以上且受大學院校教育程度 人口數
不含車位中位數房價	不含車位之各建物型態整體中位數房價

表格中未特別標明年份之變數皆為 109 年當年度之資料；

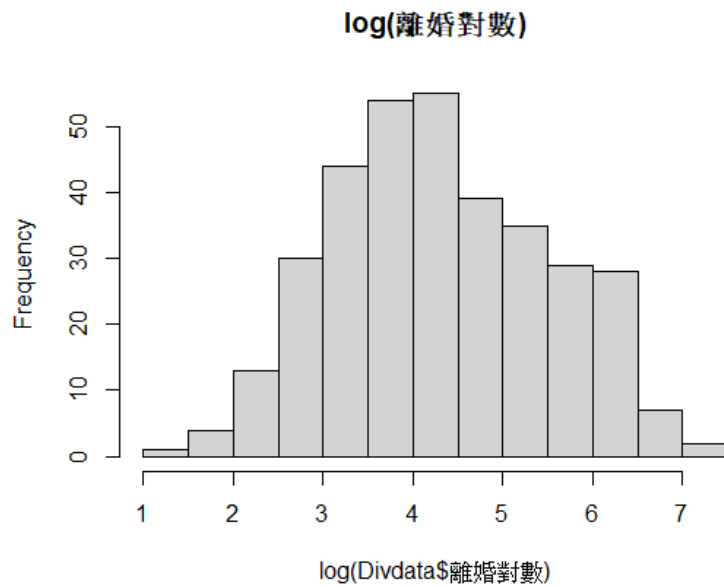
對於離婚數之想法，可能與經濟情況、老年化有關係，房價越高則壓力越大，或是老年化之情況越嚴重，也認為會加劇離婚之情形。

其中自然增加率、社會增加率、結婚率……等認為可代表該地區經濟情況樂觀程度之變數，也就是認為當當地經濟相對繁榮時，這樣指標相對較高，則對應至離婚數，則認為兩方應存在反向關係；相反的，往年粗離婚率、扶養比、老化指數以及房價則是認為會與離婚對數同向變動。

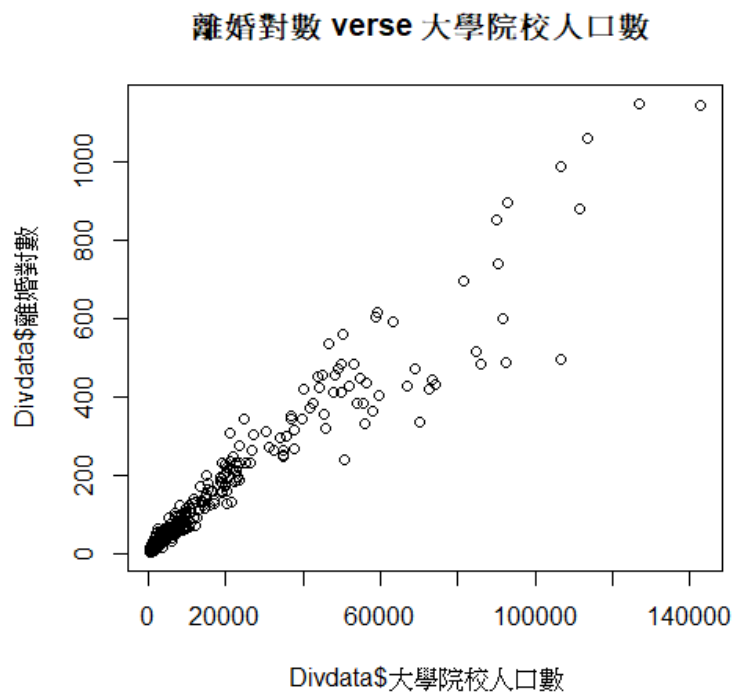
Summary(data)

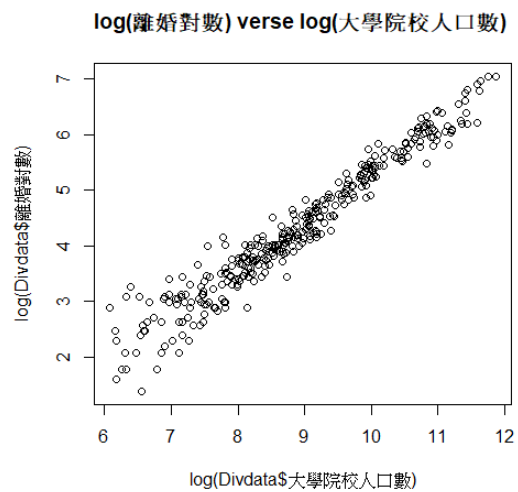
	離婚對數	粗結婚率	107 自然增加率	107 社會增加率	107 粗結婚率
Min	4	1.080	-16.300	-18.470	2.250
1 st Qu.	32	4.100	-6.080	-5.27000	4.620
Median	64	4.740	-2.270	-1.84000	5.330
mean	147	4.672	-2.739	0.00827	5.245
3 rd Qu.	189	5.290	0.790	3.24000	5.890
Max	1148	7.910	7.090	90.94000	8.800
	107 粗離婚率	108 自然增加率	108 社會增加率	108 粗結婚率	108 粗離婚率
Min	1.040	-18.410	-49.910	2.100	0.79
1 st Qu.	1.910	-5.840	-8.490	4.710	1.94
Median	2.220	-2.520	-3.670	5.270	2.22
mean	2.239	-2.857	-3.096	5.222	2.24
3 rd Qu.	2.490	0.650	0.910	5.760	2.49
Max	4.720	6.740	43.730	8.520	5.77
	性比例	扶養比	老化指數	大學院校人口數	房價中位數
Min	86.43	28.63	35.56	438	224811
1 st Qu.	99.50	36.90	113.50	2735	3660000
Median	104.94	40.06	165.97	6738	5750000
mean	105.45	40.73	184.93	16899	6348747
3 rd Qu.	111.59	44.28	235.11	19944	7600000
Max	128.97	58.26	595.00	142771	84000000

資料視覺化

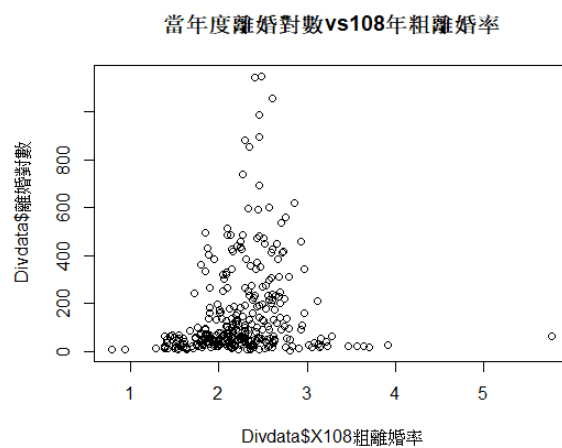
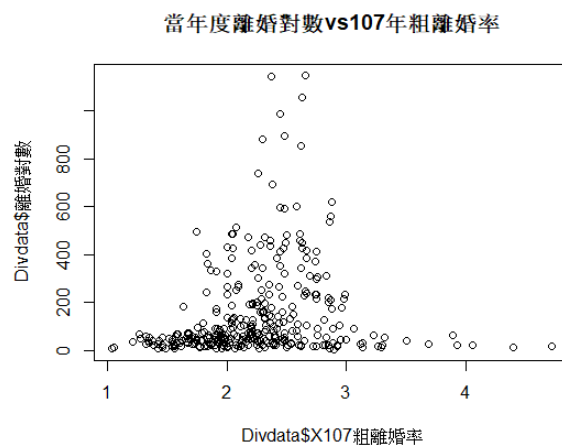


先觀察離婚對數之盒形圖，可了解主要分布在 0 到 400 之間，其中也可發現資料呈現右偏，又因為離婚對數為正數且資料中並無鄉鎮市區為 0 比離婚對數，故初步考慮對此變數取 \log (如上圖 2)，資料呈現則較對稱化。



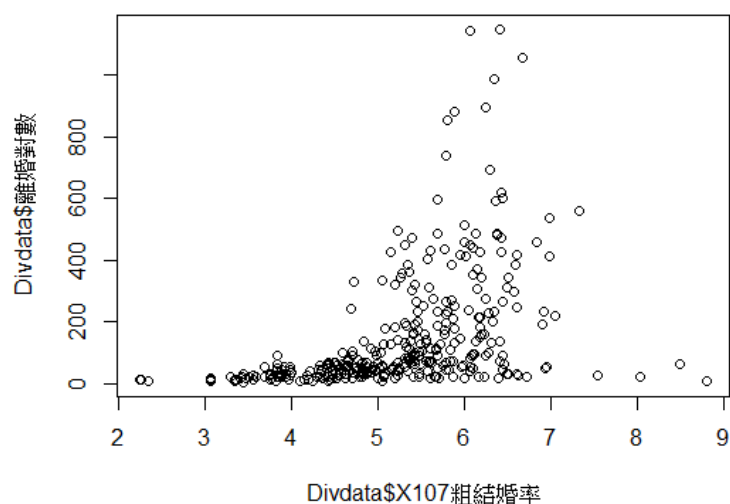


離婚對數(應變數)對大學院校人口數，可看出兩者之間具有高度線性相關，其中資料主要聚集在圖形左下部分，認為是因為離婚對數以及大學院校人口數皆為極右偏分布，故將兩變數取 log 後再做一次 scatterplot 則可發現有更好的線性關係。

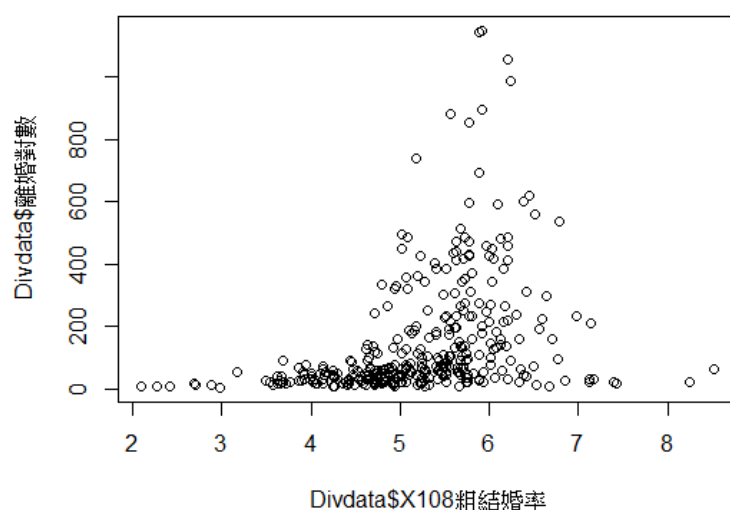


當年度離婚對數對上 107 年以及 108 年粗離婚率之比較，兩者圖形表現差不多，但與預期中會呈現正相關的猜測不一致；而都在離婚率為 2-3 單位時離婚對數分布較分散；而圖中 108 年粗離婚率為 5 單位以上，看似為離群值之點為花蓮縣秀林鄉，但其在今年之離婚對數卻依然是較低者，猜測這些往年離婚率高而離婚對數較低之原因為此點皆為人數不多之地區，離婚率對於離婚對數之影響較為敏感，故才會呈現如此情形，現階段先將其標記。

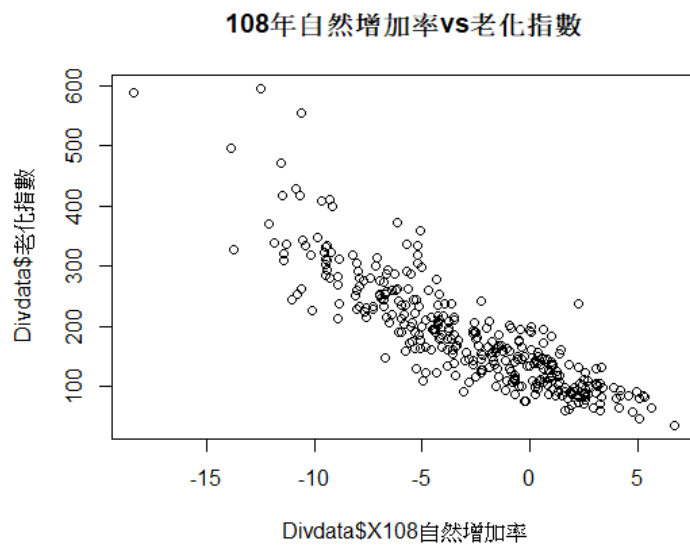
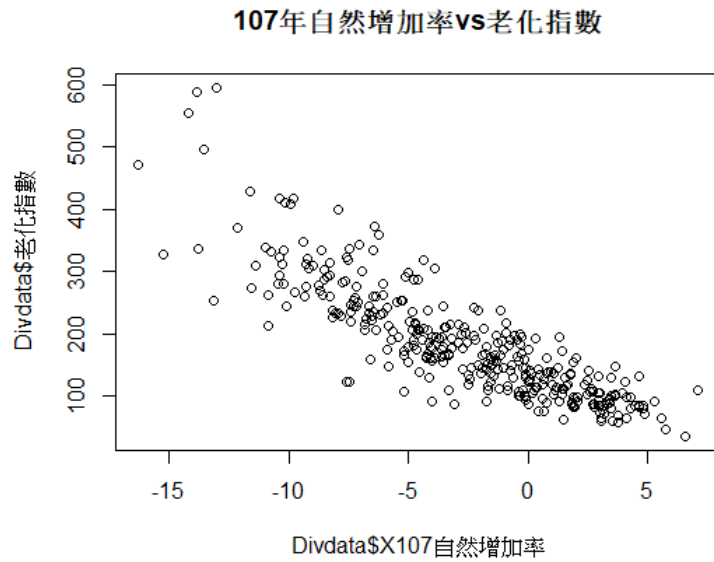
當年度離婚對數vs107年粗結婚率



當年度離婚對數vs108年粗結婚率



對上 107 年以及 108 年粗結婚率，可發現圖形與和往年粗離婚率比較之圖形呈現差不多，則是在結婚率為 5-7 單位時離婚對上分布較分散。



觀察發現 107 年以及 108 年自然增加率與老化指數之間具有高度共線性，故稍後做建模時需要特別注意這三個變數，考慮共線性診斷。

(二)建模與處理

1. 遺失值處理

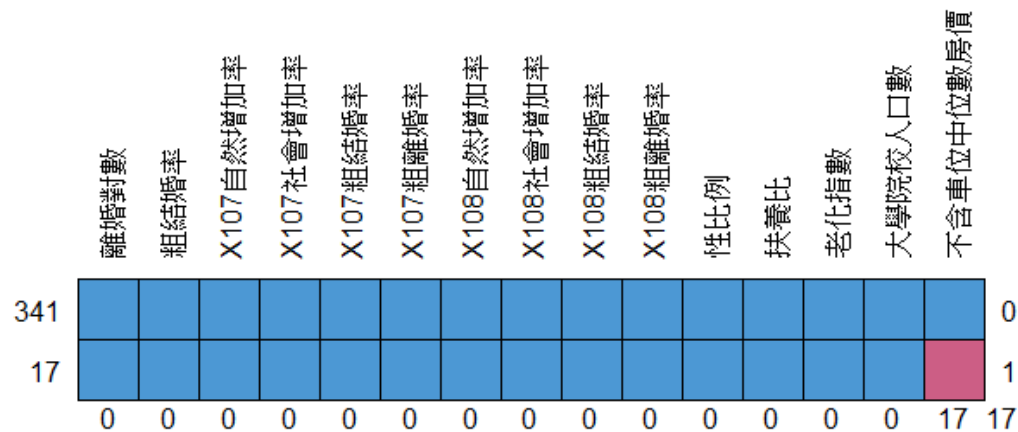
先前提到此筆資料有 17 筆遺失值，故處理資料前須先處理遺失值。

遺失值皆發生於不含車位中位數房價(以下簡稱房價)，地區主要分布在高雄市、新竹縣、彰化縣、台東縣、嘉義縣、花蓮縣以及屏東縣，除了北部資料較完整，其餘中西東部皆有部分缺失，猜測資料之遺失與資料蒐集難度有關，或許當地建立資料庫之系統較不完備，也或是該地區房價之資料建立非以建物之

車位區分，固可認為此缺失情形為完全隨機缺失(missing at random)，可應用 mice 方法補值。

Mice

透過 package mice，可觀察整筆資料之 缺失情形(如下圖)，341 筆觀測值為完整資料，17 筆觀測值具有 1 項缺失值，皆發生在房價中。



接續使用套件下之 mice() 做插補，於此使用 CART 決策數進行遺漏值預測，建立新 data frame : Divdata_mice。

直接刪除

遺失值佔整體資料約 4%，認為其並不影響整體資料分析，故也可考慮將其直接刪除，並在接下來之階段皆以此方法處理後之資料為主。

2. 初步模型

初步模型只先考慮單變數，即

$$\text{離婚對數} = \beta_0 + \beta_1 \times \text{粗結婚率} + \beta_1 \times \text{大學院校人口數} + \dots + \beta_{15} \times \text{不含車位中位數房價}$$

3. 共線性診斷

先前有提到此筆需要考慮共線性問題，於此，使用變異數膨脹因子 VIF 做指標判斷該模型中具有共線性之變數。

粗結婚率	107 自然增加率	107 社會增加率	107 粗結婚率	107 粗離婚率	108 自然增加率	108 社會增加率
2.96783	13.895215	1.326454	2.923324	1.852827	14.936847	1.42561
108 粗結婚率	108 粗離婚率	性比例	扶養比	老化指數	大學院校人口數	房價中位數
2.554044	1.712813	3.086202	1.921292	5.299728	2.025749	1.233151

於此，將 $VIF > 10$ 之變數去除，也就是 107 年自然增加率以及 108 年自然增加率，剩餘變數再判斷一次 VIF，剩餘變數之 VIF 皆小於 5。

4. 挑選變數

向前選取法

Step: AIC = 2585.64

離婚對數 ~ 大學院校人口數 + 扶養比 + X108 社會增加率 + 不含車位中位數房價 + X107 粗離婚率 + 性比例 + X107 粗結婚率 + X108 粗離婚率

向後選取法

Step: AIC = 2585.64

離婚對數 ~ X107 粗結婚率 + X107 粗離婚率 + X108 社會增加率 + X108 粗離婚率 + 性比例 + 扶養比 + 大學院校人口數 + 不含車位中位數房價

逐步選取法

Step: AIC = 2585.64

離婚對數 ~ X107 粗結婚率 + X107 粗離婚率 + X108 社會增加率 + X108 粗離婚率 + 性比例 + 扶養比 + 大學院校人口數 + 不含車位中位數房價

三種變數選取之方法 AIC 以及選出之變數皆相同，以此作為新模型繼續分析。

Call:

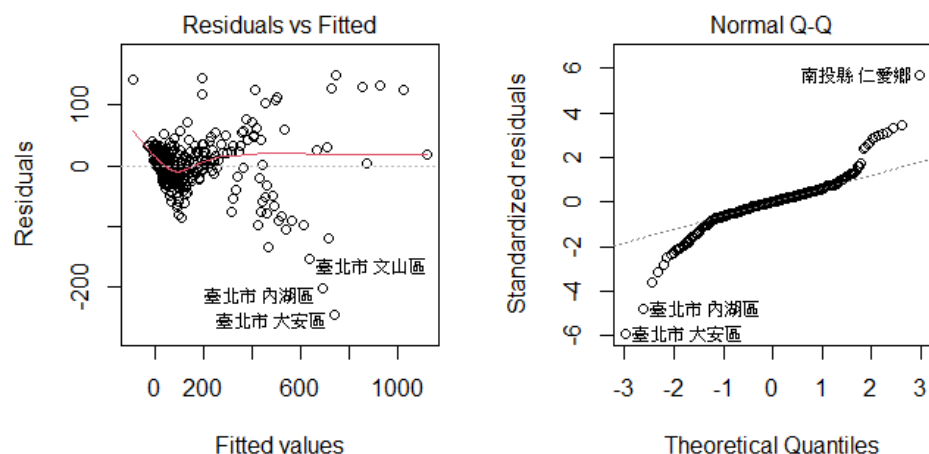
lm(formula = Divdata\$離婚對數 ~ X107 粗結婚率 + X107 粗離婚率 + X108 社會增加率 + X108 粗離婚率 + 性比例 + 扶養比 + 大學院校人口數 + 不含車位中位數房價)

coefficients:

	estimate	Std. error	T value	P value
Intercept	-1.126e+02	6.005e+01	-1.874	0.062
X107 粗結婚率	5.281e+00	3.645e+00	1.449	0.148
X107 粗離婚率	1.331e+01	5.796e+00	2.296	*
X108 社會增加率	1.353e+00	3.043e-01	4.446	***
X108 粗離婚率	9.068e+00	6.325e+00	1.434	0.153
性比例	1.547e+00	4.082e-01	3.791	***
扶養比	-2.379e+00	5.948e-01	-4.000	***
大學院校人口數	7.809e-03	1.359e-04	57.453	***
中位數房價	-1.852e-06	4.633e-07	-3.997	***
Residual standard error 43.74 on 332 degrees of freedom				
Multiple R-square: 0.9491, adjusted R-square: 0.9479				
F-statistic: 773.6 on 8 and 332 DF, p-value:<2.2e-16				

5. 殘差分析

殘差分析之部分主要觀察所配適得到之模型，其殘差是否服從原 OLS 之假設，檢測其殘差是否異質以及是否符合常態假設，以下對於模型之殘差作圖：



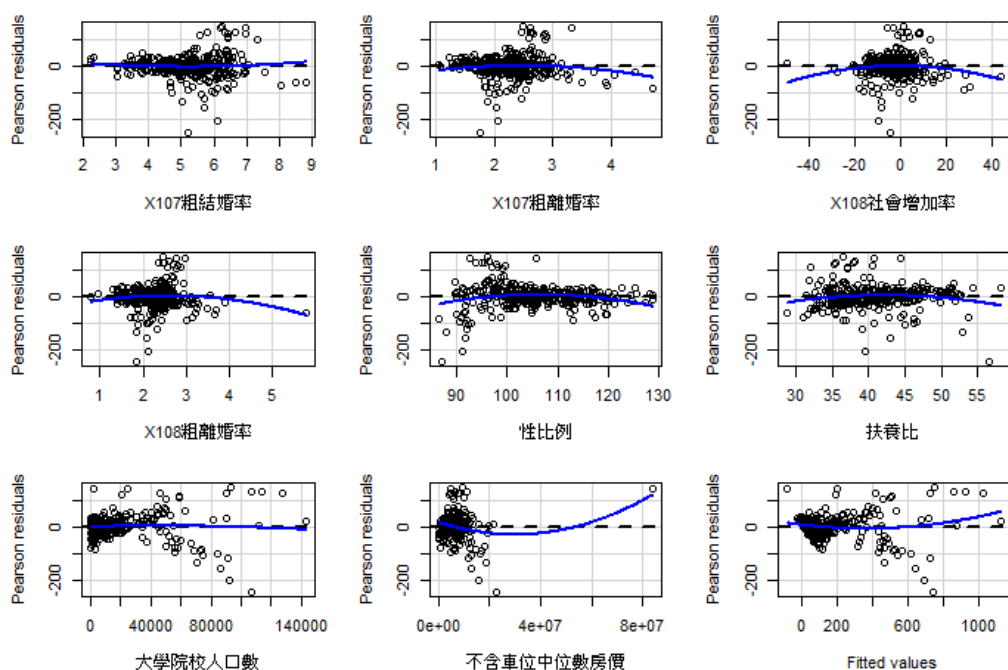
殘差圖往右發散(喇叭型)的分布，非隨機化，具異質變異數問題，且常態機率圖也表現其並非呈現常態，故須討論 WLS 以及變數變換等議題。

變數變換

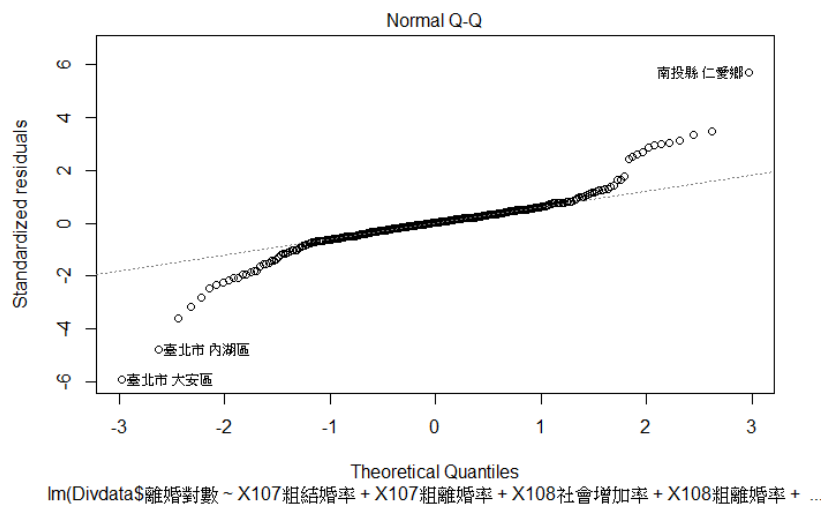
觀察解釋變數是否需要做變數變換

Tukey test :p-value = 0.001703

其 partial residual plot:



Tukey test p-value = 0.001703 < 0.05，代表此模型配適不佳，解釋變數對於殘差有特定 pattern，並非 null plot；其中，先前提及大學院校人口數為右偏分布，考慮做 log 轉換，而中位數房價和 X108 粗離婚率則被該筆資料之離群值影響。

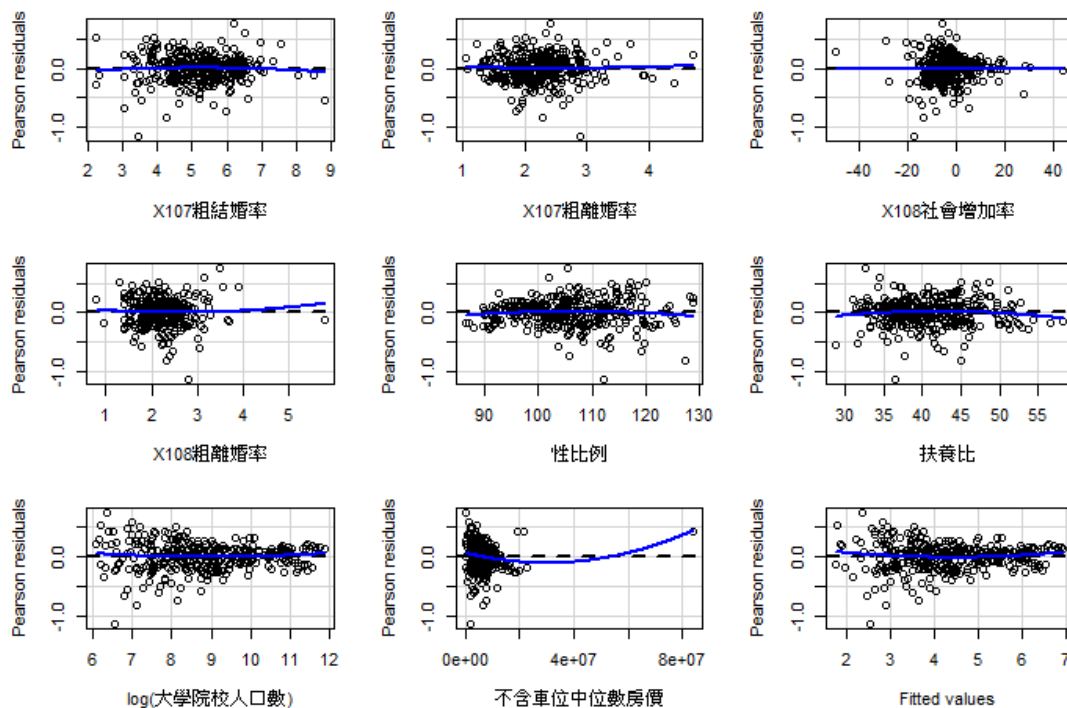


另外，常態機率圖也表示此模型不符合常態假設。

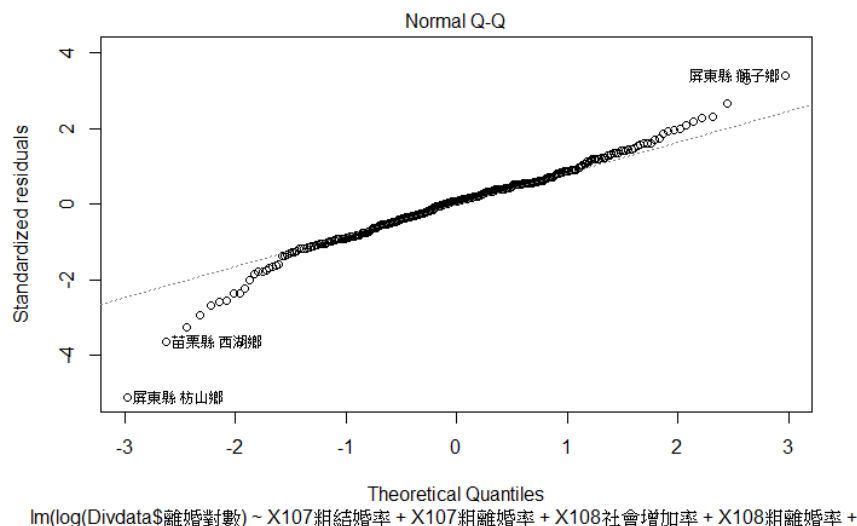
因應前段視覺資料化得知離婚對數之分布情形呈現右偏，考慮可將取 log 轉換，故，考慮將離婚對數以及大學院校人口數取 log 後做配適。

Tukey test : p value = 0.1027056

partial residual plot:



除了房價對應殘差之圖之外，可看出模型配適優化許多，大部分殘差皆呈現帶狀，大致具有隨機分布之特性，tukey test 之 p 值也提高許多達到 0.10。



常態機率圖看似與原模型之圖差別不大，故做檢定檢測。

One-sample Kolmogorov-smirnov test

D = 0.053463, p-value = 0.2839

Alternative hypothesis: two-sided

由檢定之 p-value 值可知無足夠證據拒絕虛無假設，即殘差呈現常態。

模型：

	estimate	Std. error	T value	P value
Intercept	-4.813e+00	4.484e-01	-10.735	***
X107 粗結婚率	2.013e-02	1.983e-02	1.015	0.3108
X107 粗離婚率	2.189e-01	3.090e-02	7.084	***
X108 社會增加率	-1.598e-03	1.652e-03	-0.967	0.3343
X108 粗離婚率	1.387e-01	3.342e-02	4.150	***
性比例	5.130e-03	2.550e-03	2.012	0.0451*
扶養比	-1.189e-03	3.093e-03	-0.385	0.7008
Log(大學院校人口數)	8.691e-01	1.779e-02	48.860	***
中位數房價	7.022e-10	2.408e-09	0.292	0.7708
Residual standard error 0.2291 on 332 degrees of freedom				
Multiple R-square: 0.9366, adjusted R-square: 0.9624				
F-statistic: 1090 on 8 and 332 DF, p-value:<2.2e-16				

Box-Cox 轉換

應用 power transform 和 box cox 對解釋變數以及應變數做轉換。

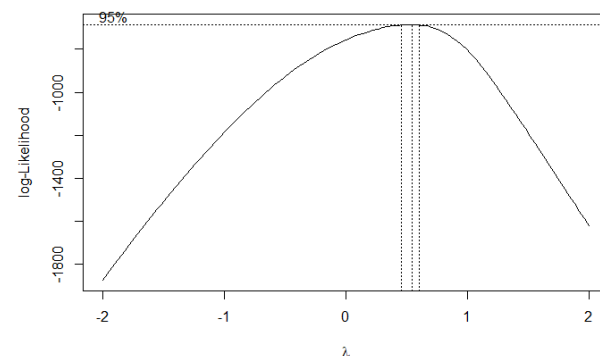
對 X 做 power transformation 得

Estimated transformation parameters:

X107 粗結婚 率	X107 粗 離婚率	X108 社 會增加 率	X108 粗 離婚率	性比例	扶養比	大學院 校人口 數	不含車位中 位數房價
0.950	-0.048	0.912	0.185	-0.435	-0.944	-0.022	0.174

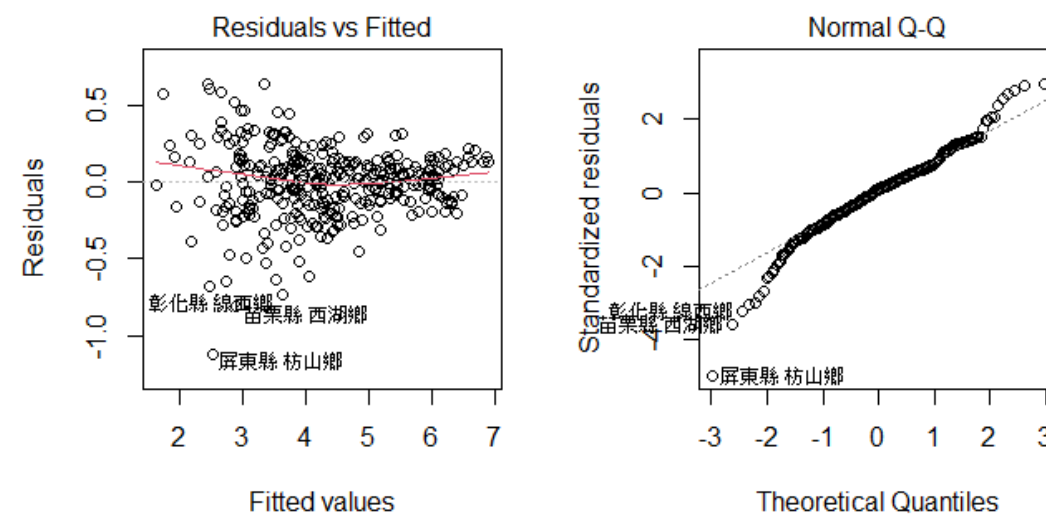
考慮對 X107 粗離婚率、X108 粗離婚率、性比例大學院校人口數、以及房價取 log，以及對 X108 社會增加率取倒數。

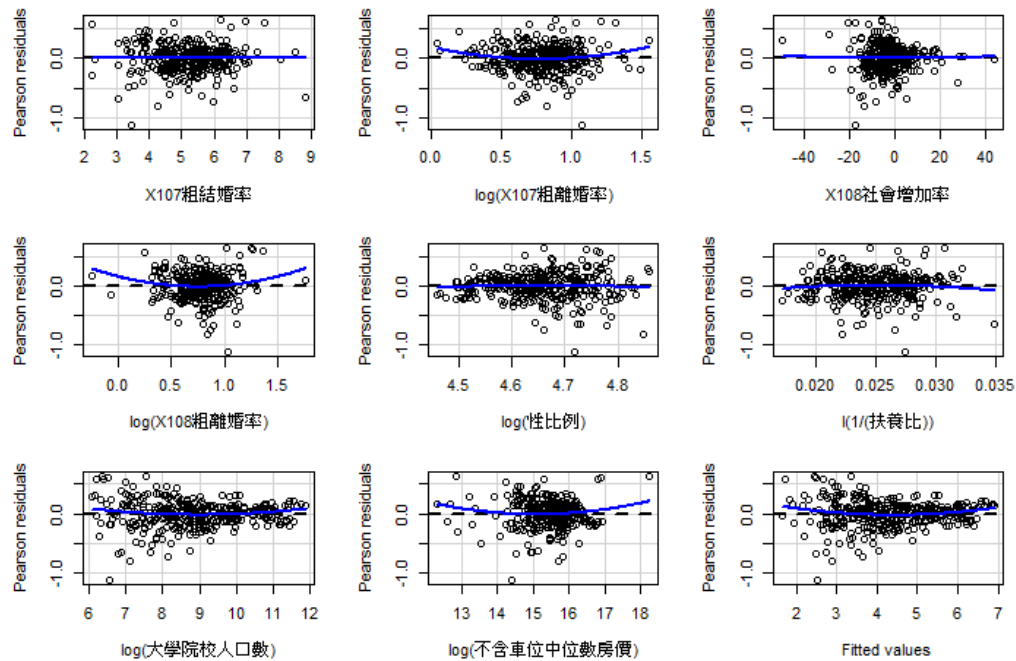
對應變數做 box cox，取使 likelihood 達最大之 lambda



Lambda = 0.5454，取 lambda = 0，即對應變數取 log 轉換。

接著一樣看 partial residual plot、常態機率圖以及做 ks-test





One-sample Kolmogorov-smirnov test

D = 0.06455, p-value = 0.1166

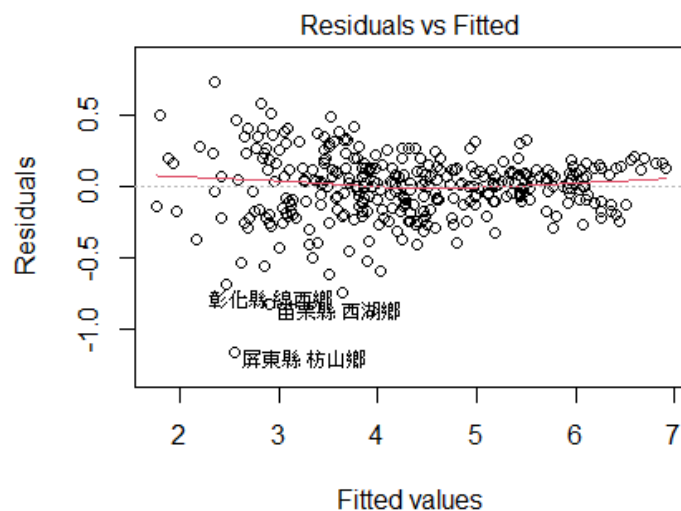
其也無足夠證據拒絕虛無假設，即呈常態分布，比較此用 box cox 及 power transformation 轉換過之模型以及前段判斷資料值觀做變數變換後之模型，兩模型於應變數皆取 log，而後者只將其一解釋變數做變數變換，兩者之殘差圖看似相差不遠，其中，前者與後者之調整判定 R 平方比較，前者為 0.9621 後者為 0.9624，且因後者之解釋變數幾乎皆做過變數變換調整，較難以依據模型配適做解釋，故以下分析仍以後者為主，即只對於應變數及大學院校人口數做 log 轉換之模型。

Weighted least square

上段解決常態假設之問題，但發現殘差圖仍看似有異質變異數之問題，利用 non-constant variance score test：ncvTest() 對模型之 fitted value 作檢定得到之 p-value: $\leq 2.22e-16$ ，即有足夠證據說明此模型具異質變異數。故考慮使用 WLS。

權重之估計使用 HC3: $\widehat{w}_i = \frac{\hat{\sigma}^2}{1-h_{ii}}$ ，代入 lm() 中之 weights 參數。

其檢定結果以即圖形分別為



離婚對數) ~ X107粗結婚率 + X107粗離婚率 + X108社會增加率 ·

ncvTest: p-value = 0.5425

ks.test: p-value = 0.2932

兩檢定結果皆不拒絕虛無假設，即具常態假設以及同質變異之假設，但取過 weight 後再做一次共線性診斷，即檢測模型中變數之 VIF，發現許多變數之值提升許多，如下表：

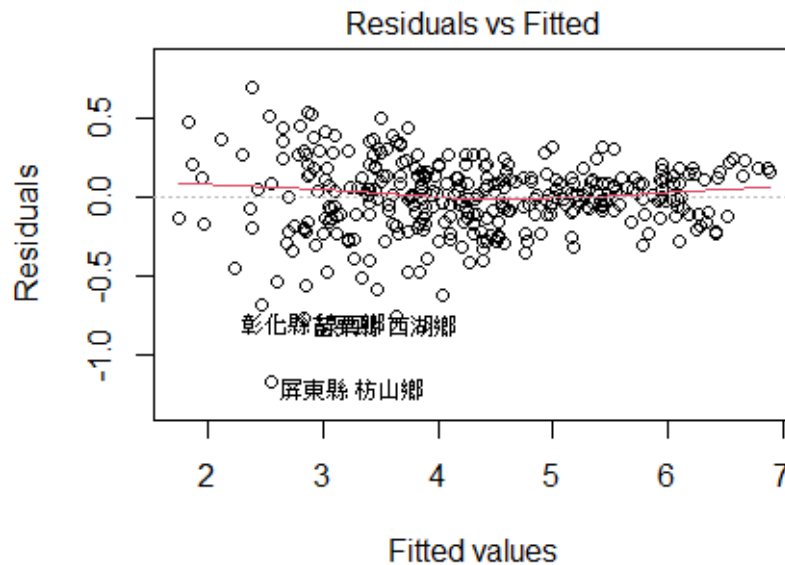
107 粗 結婚率	107 粗 離婚 率	108 社 會增 加率	108 粗 離婚 率	性比例	扶養比	Log(大 學院校 人口 數)	中位數 房價
10.066	5.007	9.419	3.617	73.933	11.713	69.274	11.176

依序剔除 VIF 最大值，直到模型內變數 VIF 皆小於 10，得到剩下變數為 X107 粗結婚率、X107 粗離婚率、X108 粗離婚率、扶養比、log(大學院校人口數)以及不含車位中位數房價

各變數之 VIF 為：

107 粗 結婚 率	107 粗 離婚 率	108 粗 離婚 率	扶養比	Log(大 學院校 人口 數)	中位數 房價
2.436	2.689	1.928	3.142	2.425	3.876

而此 WLS 模型之殘差圖與兩相關假設檢定結果分別為：



`data$離婚對數) ~ X107粗結婚率 + X107粗離婚率 + X108粗離婚率`

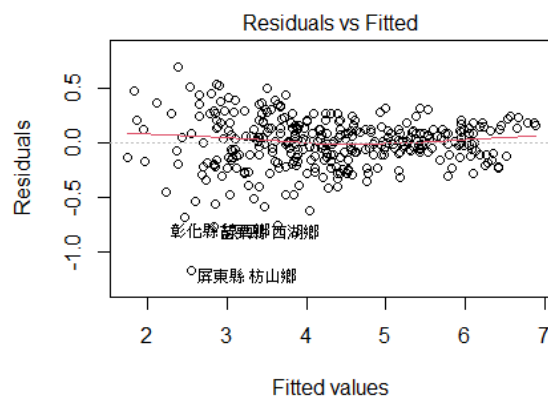
ncvTest: p-value = 0.74924

ks.test: p-value = 0.2835

皆不拒絕虛無假設，因此就以此為最終之模型。

6. 最終模型

殘差對配適值圖形：



Call :

`Lm(log(離婚對數) ~ X107 粗結婚率 + X107 粗離婚率 + X108 粗離婚率 + 扶養比 + log(大學院校人口數) + 不含車位中位數房價`

	estimate	Std. error	T value	P value
Intercept	-3.941e+00	3.671e-02	-107.346	< 2e-16 ***

X107 粗結婚率	1.447e-02	3.696e-03	3.915	0.00011 ***
X107 粗離婚率	2.185e-01	5.869e-03	37.240	< 2e-16 ***
X108 粗離婚率	1.282e-01	6.699e-03	19.143	< 2e-16 ***
扶養比	-7.691e-04	4.792e-04	-1.605	0.10947
Log(大學院校人口數)	8.365e-01	1.478e-03	565.925	< 2e-16 ***
中位數房價	9.450e-10	6.663e-10	1.418	0.15706
Residual standard error 1.031 on 334 degrees of freedom				
Multiple R-square: 0.9996, adjusted R-square: 0.9995				
F-statistic: 1.251e+05 on 6 and 334 DF, p-value:<2.2e-16				

調整判定係數達到 0.9995，代表起模型呈極佳之解釋能力。

比起最初假設，社會增加率以及自然增加率對於離婚對數最終並不需要在此模型中，即可達到不錯之解釋能力；

而往年之離婚率以及結婚率皆對離婚對數有些許正向影響，離如當固定其他變數之下，前一年(即 108 年)之離婚率每提高 1 單位，則今天之離婚對數會提高 0.1282 單位。因為自然需要先有結婚，才有可能會有離婚發生，故皆呈向正向關係；

扶養比則可看出扶養比越大，則離婚對數越少，這是有悖於原先所猜測可能扶養比大代表壓力過大，則離婚可能性提高之假設。

而具有大學院校學歷之人口數以及中位數房價則對於離婚對數也具有正向關係。

(三)程式碼

```
library(plyr)
library(tidyverse)
library(effects)
library(alr4)
getwd()
setwd("D:/ALR/data")
#####資料整理#####
div109 <- read.csv("109 年行政區離婚統計_鄉鎮市區.csv")
#取離婚對數 設為應變數 y
name <- div109[1,]
name = as.factor(name)
colnames(div109) = name
div109 <- div109[-1,c(-1,-6,-7,-8)]

h_ct <- read.csv("109 年行政區家戶數統計_鄉鎮市區.csv")
#取總戶數&共同生活戶數
name <- h_ct[1,]
name = as.factor(name)
colnames(h_ct) = name
h_ct <- h_ct[-1,c(-1,-9,-7,-8)]

b_ct <- read.csv("109 年行政區不動產實價登錄建物成交資訊(交易日)—按建物型態分_鄉鎮市區.csv")
#取整體中位數房價(不含車位)
b_ct <- b_ct[,c(2,3,4,31,70)]
b_ct[1,4] <- "不含車位中位數房價"
b_ct[1,5] <- "含車位中位數房價"
name <- b_ct[1,]
name = as.factor(name)
colnames(b_ct) = name
b_ct <- b_ct[-1,]
b_ct <- b_ct[,-5]

GD <- read.csv("109 年行政區人口消長指標統計_鄉鎮市區.csv")
#取出生死亡率、自然&社會增加率和粗結婚率
name <- GD[1,]
```

```

name = as.factor(name)
colnames(GD) = name
GD <- GD[-1,c(-1,-5,-6,-7,-8,-10,-11,-12)]

GD108 <- read.csv("108 年行政區人口消長指標統計_鄉鎮市區.csv")
#取出生死亡率、自然&社會增加率和粗結婚率
name <- GD108[1,]
name = as.factor(name)
colnames(GD108) = name
GD108 <- GD108[-1,c(-1,-11,-12)]

GD107 <- read.csv("107 年行政區人口消長指標統計_鄉鎮市區.csv")
#取出生死亡率、自然&社會增加率和粗結婚率
GD107[1,c(5:10)] <- c("107 粗出生率","107 粗死亡率","107 自然增加率","107 社會增加率","107 粗結婚率","107 粗離婚率")
name <- GD107[1,]
name = as.factor(name)
colnames(GD107) = name
GD107 <- GD107[-1,c(-1,-11,-12)]

population <- read.csv("109 年 6 月行政區人口指標_鄉鎮市區.csv")
#取扶養比、扶幼比和扶老比
name <- population[1,]
name = as.factor(name)
colnames(population) = name
population <- population[-1,c(2:5,8,11)]

educaiton <- read.csv("109 年行政區 15 歲以上人口教育程度統計_鄉鎮市區.csv")
#取大學院校人數
name <- educaiton[1,]
name = as.factor(name)
colnames(educaiton) = name
educaiton <- educaiton[-1,c(2:4,7)]
Divdata <- Reduce(function(x,y) join(x,y ,by = c("縣市名稱",
                                                "鄉鎮市區代碼",
                                                "鄉鎮市區名稱"),type =
"inner"),
list(div109,GD,GD107,GD108,population,educaiton,b_ct))

```

```

area <- paste(Divdata[,1],Divdata[,3])
Divdata <- Divdata[,c(-1,-2,-3)]
row.names(Divdata) <- area
i <- 1
for (i in 1:19){
  Divdata[,i] <- as.numeric(Divdata[,i])
}
str(Divdata)
Divdata <- Divdata[,c(-3,-4,-9,-10)]
write.csv(Divdata,file = "D:/ALR/data/DivData.csv",row.names = T)

Divdata <- read.csv("Divdata.csv")
rownames <- Divdata[,1]
row.names(Divdata) <- rownames
Divdata <- Divdata[,-1]
plot(Divdata)
#####
#mice 補值
length(which(is.na(Divdata$不含車位中位數房價) == TRUE))
narrow <- rownames(Divdata[c(156,157,170,182,183,
                           188,211,278,305,306,307,
                           309,323,324,325,327,339),])
library(mice)
require(mice)
mice <- data.frame(md.pattern(Divdata, rotate.names = TRUE))
mice.data <- mice(Divdata,
                  m=3,
                  maxit = 50,
                  methods = "cart",
                  seed = 188)
Divdata_mice <- complete(mice.data , 1)
#遺失值直接刪除
Divdata <- na.omit(Divdata)
summary(Divdata)
#####
#####

##畫圖

```

```

boxplot(Divdata$離婚對數,
        main = "離婚對數")
plot(Divdata$離婚對數~Divdata$大學院校人口數,
     main = "離婚對數 verse 大學院校人口數 ")
hist(Divdata$離婚對數,
     main = "log(離婚對數)")
plot(Divdata$離婚對數~ Divdata$X108 粗結婚率,
     main = "當年度離婚對數 vs108 年粗結婚率")
plot(Divdata$離婚對數~ Divdata$X108 粗離婚率,
     main = "當年度離婚對數 vs108 年粗離婚率")
which(Divdata$X108 粗離婚率 >5)
Divdata[322,]
plot(Divdata$X107 自然增加率,Divdata$老化指數,
     main = "107 年自然增加率 vs 老化指數")
#####
##共線性診斷
lm <- lm(Divdata$離婚對數~. ,data = Divdata)
summary(lm)
plot(lm)
pairs <- pairs(Divdata)
plot(predict(lm), residuals(lm))
par(mfrow = c(2,2))
residualPlots(lm)
library(car)
car::vif(lm)
lm <- lm(Divdata$離婚對數 ~. - X107 自然增加率 - X108 自然增加率,data =
Divdata)
car::vif(lm)
#####
#變數選取
#向前選取
full <- formula(lm)
m0 <- lm(Divdata$離婚對數 ~ 1, Divdata) # the base model
m.forward <- step(m0, scope=full, direction="forward")
#向後選取
m1 <- update(m0, full) #full model
m.backward <- step(m1, direction="backward")
#stepwise

```

```

m.stepwise <- step(m1, direction="both")
lm2 <- lm(Divdata$離婚對數 ~ X107 粗結婚率 +
          X107 粗離婚率 + X108 社會增加率 +
          X108 粗離婚率 + 性比例 + 扶養比 +
          大學院校人口數 + 不含車位中位數房價,data = Divdata)

summary(lm2)
plot(lm2)
residualPlots(lm2)
#####
lmm <-lm(log(Divdata$離婚對數) ~ X107 粗結婚率 +
          X107 粗離婚率 + X108 社會增加率 +
          X108 粗離婚率 + 性比例 + 扶養比 +
          log(大學院校人口數) + 不含車位中位數房價,data = Divdata)
car::vif(lmm)
summary(lmm)
plot(lmm)
residualPlots(lmm)
ks.test(scale(lmm$residuals),pnorm)
ncvTest(lmm)
#####
#####powertransorm#####
library(MASS)
boxcox <- boxcox(lm)
lambda <- boxcox$x[which.max(boxcox$y)]

y <- Divdata[1]
x <- Divdata[c(5,6,8,10,11,12,14,15)]
powerTransform(x,family = "yjPower")
lmlog <- lm(log(Divdata$離婚對數) ~ X107 粗結婚率 +
            log(X107 粗離婚率) + X108 社會增加率 +
            log(X108 粗離婚率) + log(性比例) + I(1/(扶養比)) +
            log(大學院校人口數) + log(不含車位中位數房價),data = Divdata)
residualPlots(lmlog)
plot(lmlog)
ks.test(scale(lmlog$residuals),pnorm)
summary(lmlog)
ncvTest(lmlog)

```

```

lmm$residuals
#WLS#####
lev <- lm.influence(lmm)$hat
hc0wt <- 1/lmm$residuals^2
ss <- hccm(lmm,typr = "hc3")
hc3wt <- 1/lmm$residuals^2/(1-lev)^2
lmm_wt <- lm(log(Divdata$離婚對數) ~ X107 粗結婚率 +
              X107 粗離婚率 + X108 社會增加率 +
              X108 粗離婚率 + 性比例 + 扶養比 +
              log(大學院校人口數) + 不含車位中位數房價,data = Divdata,
              weights = ss)
ncvTest(lmm_wt)
ks.test(scale(lmm_wt$residuals),pnorm)
residualPlots(lmm_wt)
car::vif(lmm_wt)
plot(lmm_wt)

summary(lmm_wt)
car::vif(lmm_wt)
lm <- lm(log(Divdata$離婚對數) ~ X107 粗結婚率 +
          X107 粗離婚率 +
          X108 粗離婚率 + 扶養比 +
          log(大學院校人口數) + 不含車位中位數房價,data = Divdata)
levn <- lm.influence(lm)$hat
hc3wtn <- 1/lm$residuals^2/(1-levn)^2
lmnew_wt <- lm(log(Divdata$離婚對數) ~ X107 粗結婚率 +
               X107 粗離婚率 +
               X108 粗離婚率 + 扶養比 +
               log(大學院校人口數) + 不含車位中位數房價,data = Divdata,
               weights = hc3wtn)
par(mfrow=c(2,2))
plot(lmnew_wt)
summary(lmnew_wt)
ncvTest(lmnew_wt)
ks.test(scale(lmnew_wt$residuals),pnorm)
car::vif(lmnew_wt)

```