

迴歸期末報告

臺灣自然增加人數因素分析

目錄

- 主題介紹及資料選取
- 解釋變數選取
- 模型架構
- 殘差分析
- 離群值分析
- 結論

主題介紹及資料選取

- 一. 目的：利用 108 年資料分析全臺各鄉鎮市區的自然增加人數相關因素
- 二. 資料來源：社會經濟資料服務平台(SEGIS)
- 三. 對象：自然增加人數
- 四. 統計區：全臺 368 個鄉鎮市區
- 五. 單位：人數
- 六. 資料表名稱：
 1. 108 年 4 季行政區人口消長統計_鄉鎮市區.csv
 2. 108 年 12 月行政區人口指標_鄉鎮市區.csv
 3. 108 年 12 月行政區人口統計_鄉鎮市區.csv
 4. 108 年 12 月行政區醫療院所統計_鄉鎮市區.csv
 5. 108 年 4 季行政區結婚統計_鄉鎮市區.csv
 6. 108 年 4 季行政區離婚統計_鄉鎮市區.csv
 7. 108 年綜合所得稅所得總額申報統計_鄉鎮市區.csv
 8. 108 年 12 月行政區中低收入戶統計指標_鄉鎮市區.csv
 9. 108 年 12 月行政區低收入戶統計指標_鄉鎮市區.csv

解釋變數選取

一. 變數名稱介紹

| | | |
|---------------------|----------------|------|
| COUNTY | 縣市名稱 | 反應變數 |
| TOWN | 鄉鎮市區名稱 | |
| NATURE_INC_CNT | 自然增加人數 | 解釋變數 |
| M_F_RAT | 性別比 | |
| A0A14_A15A65_RAT | 扶幼比 | |
| A65UP_A15A64_RAT | 扶老比 | |
| P_DEN | 人口密度 | |
| Hospital_H_CNT | 醫療院所家數 | |
| Hospital_H_SRVB | 醫療院所平均每千人擁有病床數 | |
| RLP2_CNT | 中低收入戶戶內人數 | |
| RLP_CNT | 低收入戶戶內人數 | |
| RATE_MARRY_CP_CNT | 結婚率(‰)* | |
| RATE_DIVORCE_CP_CNT | 離婚率(‰)* | |
| FLD03 | 綜合所得總額平均數 | |
| House_P_CNT | 人口數 | |

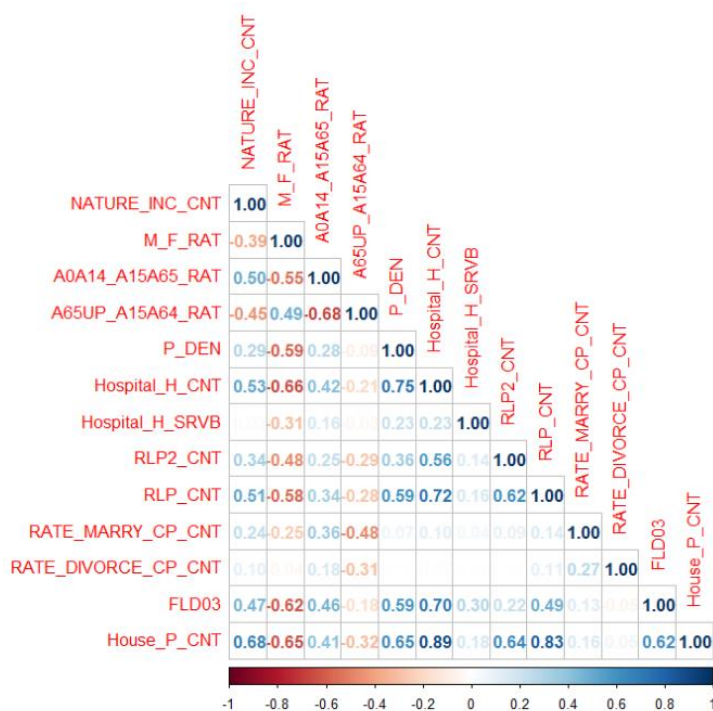
*結婚/離婚率經過公式轉換($\frac{\text{結婚/離婚對數}}{\text{人口數}}$)×1000

```
nat_inc[,12] <- nat_inc[,12]*1000/nat_inc[,15] #結婚對數轉為結婚率(千分比)
nat_inc[,13] <- nat_inc[,13]*1000/nat_inc[,15] #離婚對數轉為離婚率(千分比)
```

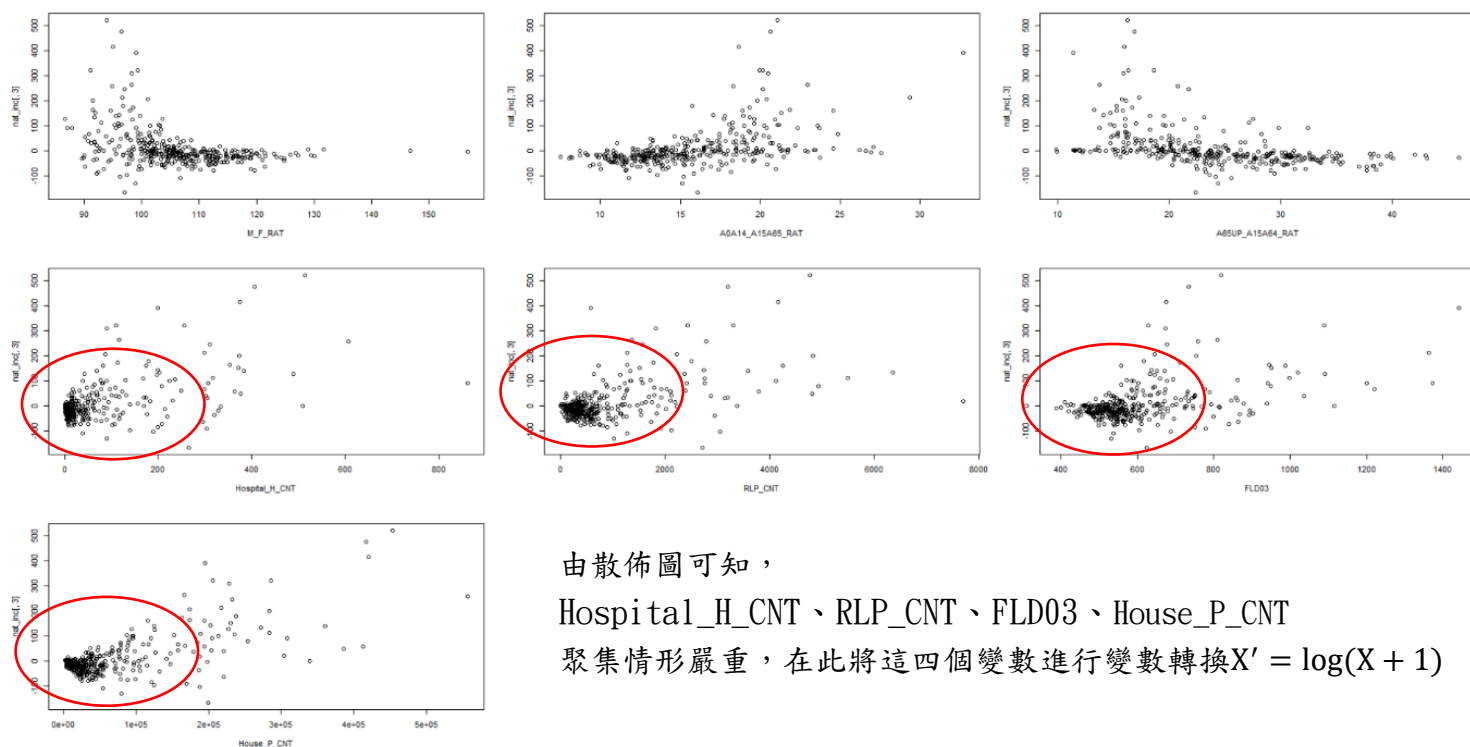
二. 檢查變數相關性

1. 檢查相關係數

先檢查反應變數(Y)與所選之解釋變數(X)是否有相關性，利用 R 套件(corrplot)將相關係數過低或過高之變數剔除，在此選取小於 0.35 之解釋變數 P_DEN、Hospital_H_SRVB、RLP2_CNT、RATE_MARRY_CP_CNT、RATE_DIVORCE_CP_CN



2. 檢查 Y(反應變數)與 X(解釋變數)散佈程度

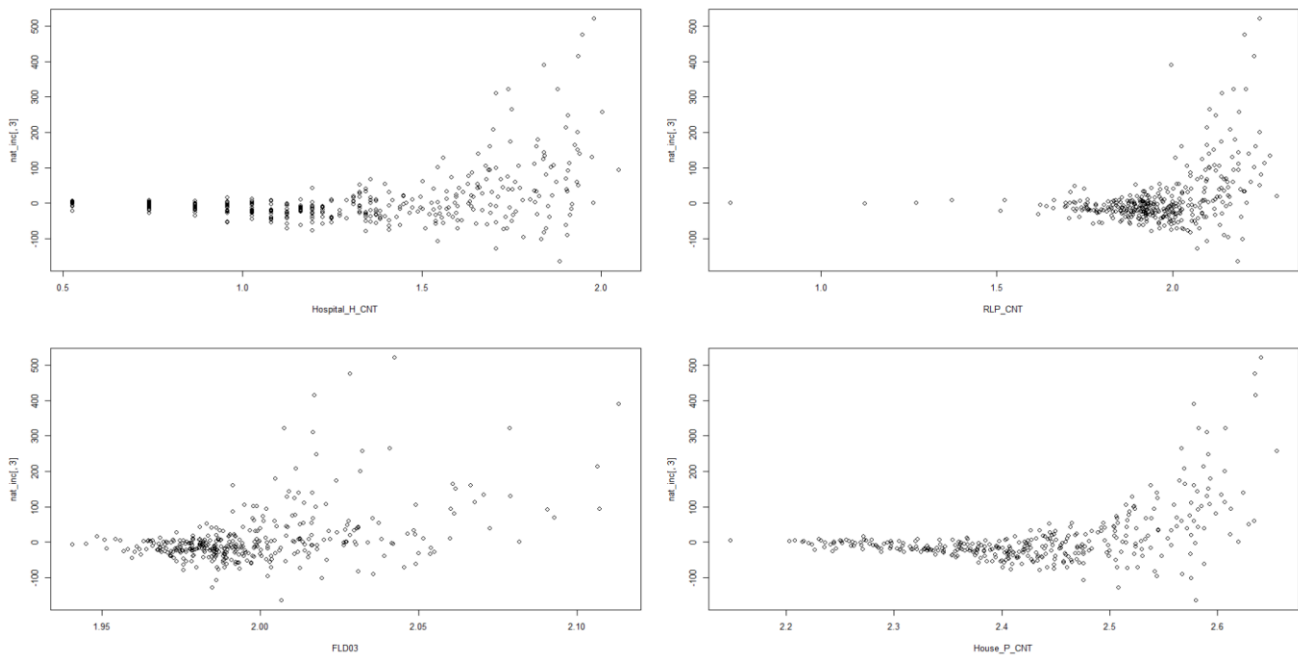


由散佈圖可知，

Hospital_H_CNT、RLP_CNT、FLD03、House_P_CNT

聚集情形嚴重，在此將這四個變數進行變數轉換 $X' = \log(X + 1)$

轉換後變數尾端雖擴散嚴重，但聚集情形已有緩解。



3. 檢查共線性

檢查所選擇的解釋變數(X)是否有共線性，因此使用 VIF 函數判斷。

| | | | | | |
|-------------|------------------|------------------|----------------|---------|--------|
| M_F_RAT | A0A14_A15A65_RAT | A65UP_A15A64_RAT | Hospital_H_CNT | RLP_CNT | FLD03 |
| 3.7704 | 2.6171 | 2.5030 | 14.9790 | 3.6193 | 2.5954 |
| House_P_CNT | | | | | |
| 11.9300 | | | | | |

可看出 House_H_CNT 與 House_P_CNT 的 VIF 值 > 10，因此先刪去 VIF 值較大的 House_H_CNT 後，再檢查一次是否仍有共線性。

| | | | | | |
|---------|------------------|------------------|---------|--------|-------------|
| M_F_RAT | A0A14_A15A65_RAT | A65UP_A15A64_RAT | RLP_CNT | FLD03 | House_P_CNT |
| 3.2119 | 2.4282 | 2.2278 | 3.5296 | 2.5279 | 4.7887 |

去除 House_H_CNT 後所有的 VIF 值均 < 10，因此選定這六個變數進行分析。

4. 選取變數

在此先列出現有的六個變數及兩兩交互作用，再運用 AIC 選取 AIC 值最小的 Linear Model。

```
#AIC
lm_1 <- lm(NATURE_INC_CNT ~ (.-TOWN -COUNTY)^2 -TOWN -COUNTY, data = nat_inc)
step.model <- stepAIC(lm_1, direction = "both", trace = TRUE)
summary(step.model)
```

Stepwise selection

```

Step: AIC=2730.31
NATURE_INC_CNT ~ M_F_RAT + A0A14_A15A65_RAT + A65UP_A15A64_RAT +
  RLP_CNT + FLD03 + House_P_CNT + M_F_RAT:RLP_CNT + M_F_RAT:House_P_CNT +
  A0A14_A15A65_RAT:RLP_CNT + A0A14_A15A65_RAT:FLD03 + A0A14_A15A65_RAT:House_P_CNT +
  A65UP_A15A64_RAT:House_P_CNT + RLP_CNT:House_P_CNT + FLD03:House_P_CNT

              Df Sum of Sq    RSS    AIC
<none>                        675567 2730.3
- A0A14_A15A65_RAT:RLP_CNT      1      3960 679527 2730.4
+ M_F_RAT:FLD03                  1       413 675153 2732.1
+ M_F_RAT:A65UP_A15A64_RAT      1       320 675246 2732.1
+ A65UP_A15A64_RAT:FLD03        1       311 675256 2732.2
+ A65UP_A15A64_RAT:RLP_CNT      1       132 675434 2732.2
+ RLP_CNT:FLD03                  1       130 675436 2732.2
+ M_F_RAT:A0A14_A15A65_RAT      1         8 675559 2732.3
+ A0A14_A15A65_RAT:A65UP_A15A64_RAT 1         3 675563 2732.3
- A0A14_A15A65_RAT:House_P_CNT  1      8401 683967 2732.7
- A0A14_A15A65_RAT:FLD03        1     12066 687633 2734.7
- M_F_RAT:House_P_CNT           1     21598 697165 2739.6
- M_F_RAT:RLP_CNT                1     22717 698284 2740.2
- FLD03:House_P_CNT              1     75009 750576 2766.0
- A65UP_A15A64_RAT:House_P_CNT  1     84966 760533 2770.7
- RLP_CNT:House_P_CNT            1    115115 790681 2784.6

Call:
lm(formula = NATURE_INC_CNT ~ M_F_RAT + A0A14_A15A65_RAT + A65UP_A15A64_RAT +
  RLP_CNT + FLD03 + House_P_CNT + M_F_RAT:RLP_CNT + M_F_RAT:House_P_CNT +
  A0A14_A15A65_RAT:RLP_CNT + A0A14_A15A65_RAT:FLD03 + A0A14_A15A65_RAT:House_P_CNT +
  A65UP_A15A64_RAT:House_P_CNT + RLP_CNT:House_P_CNT + FLD03:House_P_CNT,
  data = nat_inc)

Residuals:
    Min       1Q   Median       3Q      Max
-197.884  -16.704    3.021   18.559   224.972

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   11835.043   1454.8532    8.135 7.65e-15 ***
M_F_RAT        -35.8152     5.1490   -6.956 1.78e-11 ***
A0A14_A15A65_RAT  25.5620    20.2602    1.262 0.207919
A65UP_A15A64_RAT  26.9168     4.6960    5.732 2.18e-08 ***
RLP_CNT       -620.6774    122.4446   -5.069 6.55e-07 ***
FLD03        -1021.2083    172.7517   -5.911 8.19e-09 ***
House_P_CNT   -1095.9400    169.2506   -6.475 3.28e-10 ***
M_F_RAT:RLP_CNT    2.7791     0.8183    3.396 0.000764 ***
M_F_RAT:House_P_CNT  2.1205     0.6403    3.311 0.001027 **
A0A14_A15A65_RAT:RLP_CNT  1.8063     1.2738    1.418 0.157095
A0A14_A15A65_RAT:FLD03  -9.4209     3.8062   -2.475 0.013801 *
A0A14_A15A65_RAT:House_P_CNT  2.6448     1.2806    2.065 0.039649 *
A65UP_A15A64_RAT:House_P_CNT  -3.0296     0.4613   -6.568 1.89e-10 ***
RLP_CNT:House_P_CNT  28.6634     3.7493    7.645 2.12e-13 ***
FLD03:House_P_CNT  116.6165    18.8968    6.171 1.91e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44.38 on 343 degrees of freedom
(因為不存在，10 個觀察量被刪除了)
Multiple R-squared:  0.7022,    Adjusted R-squared:  0.6901
F-statistic: 57.77 on 14 and 343 DF,  p-value: < 2.2e-16

```

在此選定模型： $y \sim +M_F_RAT + A65UP_A15A64_RAT + RLP_CNT + FLD03 + House_P_CNT + M_F_RAT:RLP_CNT + M_F_RAT:House_P_CNT + A65UP_A15A64_RAT:House_P_CNT + A65UP_A15A64_RAT:House_P_CNT + RLP_CNT:House_P_CNT + FLD03:House_P_CNT$

模型架構

初步模型分析

一. 模型係數分析

```
#依照AIC值選取變數
y <- nat_inc$NATURE_INC_CNT
lm_2 <- lm(y ~ +M_F_RAT +A65UP_A15A64_RAT +RLP_CNT +FLD03 +House_P_CNT +M_F_RAT:RLP_CNT +M_F_RAT:House_P_CNT
+A65UP_A15A64_RAT:House_P_CNT +A65UP_A15A64_RAT:House_P_CNT +RLP_CNT:House_P_CNT +FLD03:House_P_CNT, data = nat_inc)
summary(lm_2)
anova(lm_2)
autoplot(lm_2)
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------------------|------------|------------|---------|--------------|
| (Intercept) | 11569.5833 | 1354.2551 | 8.543 | 4.19e-16 *** |
| M_F_RAT | -31.9977 | 4.9622 | -6.448 | 3.79e-10 *** |
| A65UP_A15A64_RAT | 32.6897 | 3.6710 | 8.905 | < 2e-16 *** |
| RLP_CNT | -549.2719 | 112.8557 | -4.867 | 1.72e-06 *** |
| FLD03 | -1144.2679 | 157.7105 | -7.255 | 2.64e-12 *** |
| House_P_CNT | -1023.0517 | 134.8081 | -7.589 | 3.00e-13 *** |
| M_F_RAT:RLP_CNT | 2.3589 | 0.7745 | 3.046 | 0.00250 ** |
| M_F_RAT:House_P_CNT | 1.9765 | 0.6087 | 3.247 | 0.00128 ** |
| A65UP_A15A64_RAT:House_P_CNT | -3.7185 | 0.3658 | -10.165 | < 2e-16 *** |
| RLP_CNT:House_P_CNT | 28.8580 | 3.8132 | 7.568 | 3.45e-13 *** |
| FLD03:House_P_CNT | 116.1096 | 14.3259 | 8.105 | 9.14e-15 *** |

所選變數
均有顯著差異

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 46.13 on 347 degrees of freedom

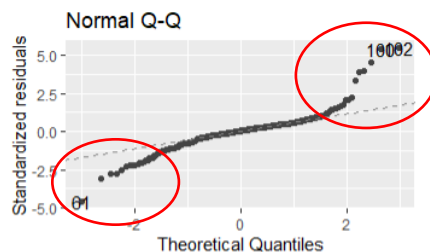
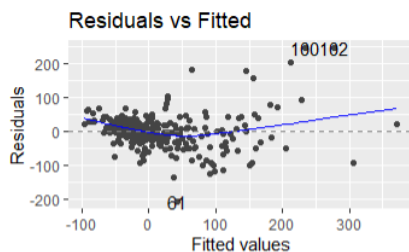
(因為不存在, 10 個觀察量被刪除了)

Multiple R-squared: 0.6746, Adjusted R-squared: 0.6652

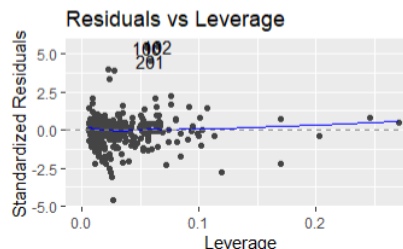
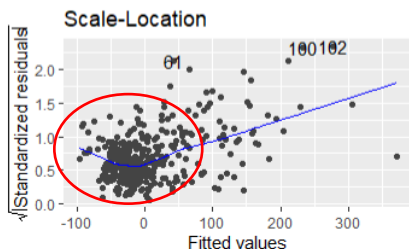
F-statistic: 71.92 on 10 and 347 DF, p-value: < 2.2e-16

$\bar{R}^2 = 0.6652$ 解釋能力不錯

整體模型顯著



尾端翹起, 偏離常態假設



標準化殘差有集中趨勢且
其與 \hat{Y} 之迴歸線相當傾斜

二. 模型變異數分析

Analysis of Variance Table

Response: y

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|------------------------------|-----|--------|---------|----------|-----------|-----|
| M_F_RAT | 1 | 345307 | 345307 | 162.2905 | < 2.2e-16 | *** |
| A65UP_A15A64_RAT | 1 | 194015 | 194015 | 91.1849 | < 2.2e-16 | *** |
| RLP_CNT | 1 | 111970 | 111970 | 52.6245 | 2.657e-12 | *** |
| FLD03 | 1 | 215352 | 215352 | 101.2130 | < 2.2e-16 | *** |
| House_P_CNT | 1 | 38460 | 38460 | 18.0758 | 2.732e-05 | *** |
| M_F_RAT:RLP_CNT | 1 | 99564 | 99564 | 46.7938 | 3.575e-11 | *** |
| M_F_RAT:House_P_CNT | 1 | 17857 | 17857 | 8.3924 | 0.004007 | ** |
| A65UP_A15A64_RAT:House_P_CNT | 1 | 226346 | 226346 | 106.3802 | < 2.2e-16 | *** |
| RLP_CNT:House_P_CNT | 1 | 141697 | 141697 | 66.5961 | 6.208e-15 | *** |
| FLD03:House_P_CNT | 1 | 139767 | 139767 | 65.6889 | 9.137e-15 | *** |
| Residuals | 347 | 738316 | 2128 | | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

由 ANOVA Table 可知性別比、扶老比、綜合所得總額平均數、扶老比與人口數的交互作用的解釋能力較高。

殘差分析

一. 模型殘差假設分析

檢查殘差是否滿足三大假設－常態性、獨立性、同質性。

```
#殘差檢定
e_2 = residuals(lm_2)
shapiro.test(e_2) #常態性檢定
dwtest(lm_2)     #獨立性檢定
ncvTest(lm_2)    #同質變異數檢定
```

Shapiro-Wilk normality test

```
data: e_2
W = 0.86825, p-value < 2.2e-16
```

p 值過小
不滿足常態假設

Durbin-Watson test

```
data: lm_2
DW = 1.5804, p-value = 2.117e-05
alternative hypothesis: true autocorrelation is greater than 0
```

p 值過小
不滿足獨立假設

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 350.3871, Df = 1, p = < 2.22e-16
```

p 值過小
不滿足同質變異數假設

三項假設均不滿足，在此我們嘗試滿足常態假設及同質變異數假設。

二. 常態性處理

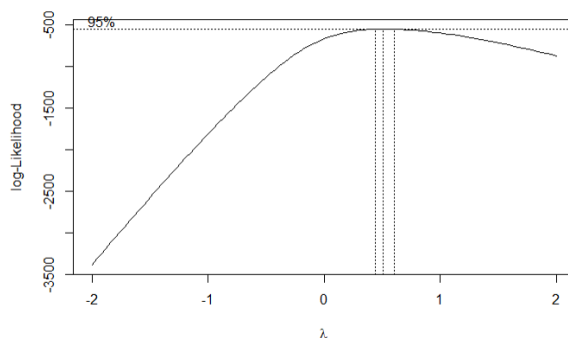
1. 變數轉換－ Box-Cox Method

為了解決不滿足常態假設的問題，將使用 boxcox 函數將Y進行變數轉換，在此之前，先將Y轉為正數Y'，因此我們先將Y進行變數轉換， $Y' = Y - \min(Y) + 1 = Y - 164$ （調整後最小值為1）。

```
#先把y轉為positive
y_prime <- y-min(y)+1 #y'=y-min(y)+1
```

轉為正數Y'後，進行 Box-Cox 轉換

```
#使用Box-Cox轉換
lm_3 <- lm(y_prime ~ +M_F_RAT +A65UP_A15A64_RAT +RLP_CNT +FLD03 +House_P_CNT +M_F_RAT:RLP_CNT +M_F_RAT:House_P_CNT
+A65UP_A15A64_RAT:House_P_CNT +A65UP_A15A64_RAT:House_P_CNT +RLP_CNT:House_P_CNT +FLD03:House_P_CNT, data = nat_inc)
b=boxcox(lm_3)
lambda<-b$x[which(b$y==max(b$y))]
lambda → [1] 0.5050505
```



最大值 λ 為 0.5050505

2. 轉換後模型係數分析

經轉換後， $Y^* = \frac{Y'^{\lambda}-1}{\lambda}$ ， $\lambda = 0.5050505$

```
y_star <- (y_prime^lambda-1)/lambda #y*=(y'^lambda-1)/lambda
lm_4<-lm(y_star ~ +M_F_RAT +A65UP_A15A64_RAT +RLP_CNT +FLD03 +House_P_CNT +M_F_RAT:RLP_CNT +M_F_RAT:House_P_CNT
+A65UP_A15A64_RAT:House_P_CNT +A65UP_A15A64_RAT:House_P_CNT +RLP_CNT:House_P_CNT +FLD03:House_P_CNT, data = nat_inc)
summary(lm_4)
autoplot(lm_4)
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------------------|-----------|------------|---------|--------------|
| (Intercept) | 732.35237 | 95.36379 | 7.680 | 1.64e-13 *** |
| M_F_RAT | -1.87753 | 0.34942 | -5.373 | 1.42e-07 *** |
| A65UP_A15A64_RAT | 2.22056 | 0.25850 | 8.590 | 2.99e-16 *** |
| RLP_CNT | -27.57090 | 7.94706 | -3.469 | 0.000588 *** |
| FLD03 | -73.48417 | 11.10564 | -6.617 | 1.39e-10 *** |
| House_P_CNT | -65.37662 | 9.49290 | -6.887 | 2.69e-11 *** |
| M_F_RAT:RLP_CNT | 0.10439 | 0.05454 | 1.914 | 0.056445 . |
| M_F_RAT:House_P_CNT | 0.13612 | 0.04286 | 3.176 | 0.001627 ** |
| A65UP_A15A64_RAT:House_P_CNT | -0.25620 | 0.02576 | -9.946 | < 2e-16 *** |
| RLP_CNT:House_P_CNT | 1.59003 | 0.26852 | 5.922 | 7.67e-09 *** |
| FLD03:House_P_CNT | 7.51321 | 1.00880 | 7.448 | 7.60e-13 *** |

該項 p 值略大於 0.05
仍可接受

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.248 on 347 degrees of freedom

(因為不存在，10 個觀察量被刪除了)

Multiple R-squared: 0.6416, Adjusted R-squared: 0.6313

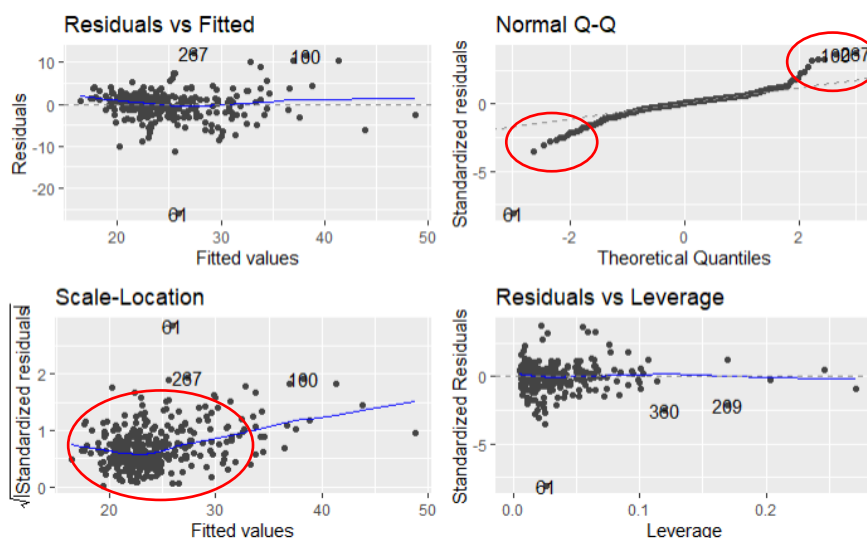
F-statistic: 62.13 on 10 and 347 DF, p-value: < 2.2e-16

$\bar{R}^2 = 0.6313$

解釋能力些許下降

但仍不錯

整體模型顯著



尾端已稍微下降

殘差雖仍集中，但已稍微分散

3. 轉換後模型殘差分析

```
#殘差檢定
e_4 = residuals(lm_4)
shapiro.test(e_4) #常態性檢定
dwtest(lm_4) #獨立性檢定
ncvTest(lm_4) #同質變異數檢定
```

Shapiro-Wilk normality test

data: e_4
w = 0.85848, p-value < 2.2e-16

Durbin-Watson test

data: lm_4
DW = 1.5807, p-value = 2.145e-05
alternative hypothesis: true autocorrelation is greater than 0

三項假設仍不滿足

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 100.011, Df = 1, p = < 2.22e-16

三. 同質變異數處理

```
#加權最小平方方法
wt <- 1 / lm(abs(lm_4$residuals) ~ lm_4$fitted.values)$fitted.values^2
wls_lm_5 <- lm(lm_4, data = nat_inc, weight=wt)
summary(wls_lm_5)
autoplot(wls_lm_5)
```

選取權重 $\hat{w}_i = \frac{1}{\hat{e}_i'^2}$, $\tilde{Y} = WY^*$

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|------------------------------|-----------|------------|---------|----------|-----|
| (Intercept) | 732.35237 | 95.36379 | 7.680 | 1.64e-13 | *** |
| M_F_RAT | -1.87753 | 0.34942 | -5.373 | 1.42e-07 | *** |
| A65UP_A15A64_RAT | 2.22056 | 0.25850 | 8.590 | 2.99e-16 | *** |
| RLP_CNT | -27.57090 | 7.94706 | -3.469 | 0.000588 | *** |
| FLD03 | -73.48417 | 11.10564 | -6.617 | 1.39e-10 | *** |
| House_P_CNT | -65.37662 | 9.49290 | -6.887 | 2.69e-11 | *** |
| M_F_RAT:RLP_CNT | 0.10439 | 0.05454 | 1.914 | 0.056445 | . |
| M_F_RAT:House_P_CNT | 0.13612 | 0.04286 | 3.176 | 0.001627 | ** |
| A65UP_A15A64_RAT:House_P_CNT | -0.25620 | 0.02576 | -9.946 | < 2e-16 | *** |
| RLP_CNT:House_P_CNT | 1.59003 | 0.26852 | 5.922 | 7.67e-09 | *** |
| FLD03:House_P_CNT | 7.51321 | 1.00880 | 7.448 | 7.60e-13 | *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.248 on 347 degrees of freedom

(因為不存在，10 個觀察量被刪除了)

Multiple R-squared: 0.6416, Adjusted R-squared: 0.6313

F-statistic: 62.13 on 10 and 347 DF, p-value: < 2.2e-16

模型無顯著變化

```
#殘差檢定
e_5= residuals(wls_lm_5)
shapiro.test(e_5) #常態性檢定
dwtest(wls_lm_5) #獨立性檢定
bptest(wls_lm_5) #同質變異數檢定(Breusch-Pagan test)
```

Shapiro-wilk normality test

data: e_5

W = 0.85848, p-value < 2.2e-16

Durbin-Watson test

data: wls_lm_5

DW = 1.5807, p-value = 2.145e-05

alternative hypothesis: true autocorrelation is greater than 0

studentized Breusch-Pagan test

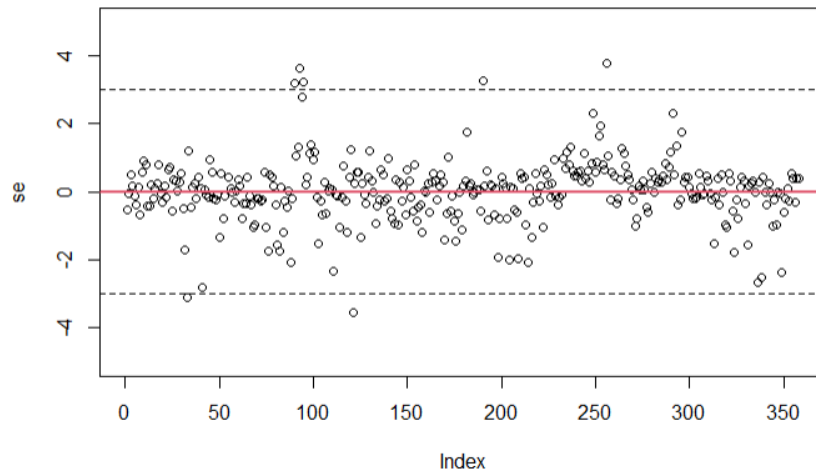
data: wls_lm_5

BP = 44.487, df = 10, p-value = 2.69e-06

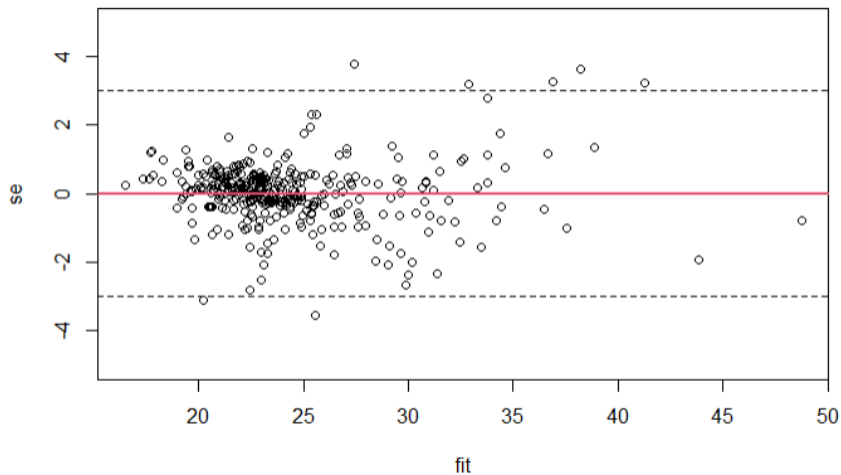
三項假設仍不滿足

離群值分析

一. 一般殘差、標準化殘差



```
#標準化殘差
e = residuals(wls_lm_5)
se = e/sd(e)
fit = fitted(wls_lm_5)
plot(se,ylim=c(-5,5))
abline(h=0,col=2,lwd=2)
abline(h=-3,lty=2)
abline(h=3,lty=2)
```



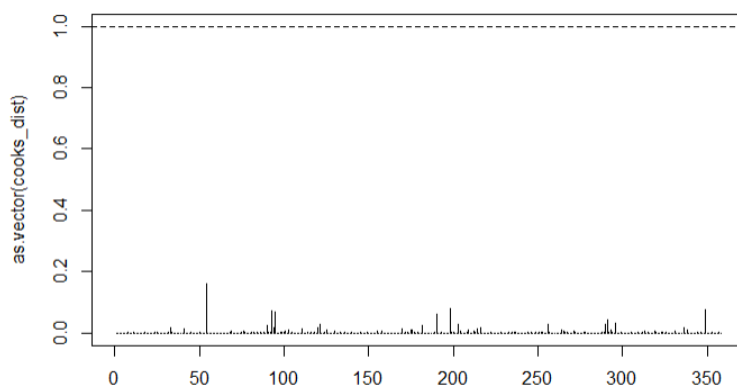
```
#一般殘差
plot(fit,se,ylim=c(-5,5))
abline(h=0,col=2,lwd=2)
abline(h=-3,lty=2)
abline(h=3,lty=2)
```

可看出超過三個標準差的離群值不多，影響不大在此將不去除。

二. Cook's Distance

亦可使用 Cook's Distance 找出離群值

```
#用Cook distance 尋找離群值
cooks_dist<-cooks.distance(wls_lm_5)
which(cooks_dist>1) #選取準則 Cook distance >1
plot(as.vector(cooks_dist),type="h",ylim=c(0,1))
abline(h=1,lty=2)
```



選取準則：Cook's Distance > 1
所有 Cook's Distance 均小於 1，因此亦不去除離群值。

結論

一. 最終模型選取變數

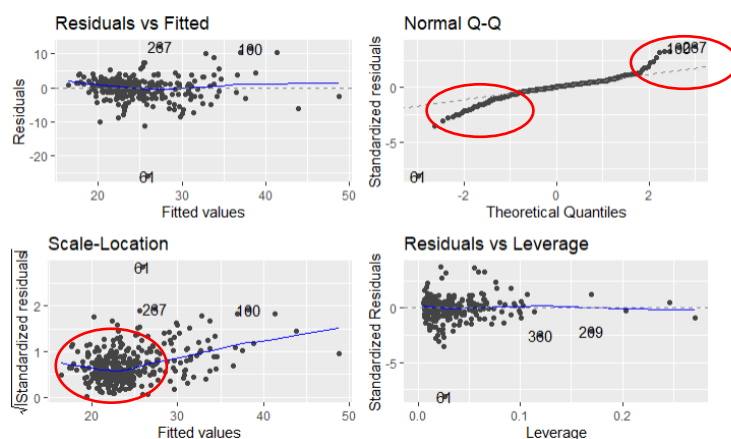
| 估計參數 | 解釋變數 |
|------------------------------|-------------------|
| $\hat{\beta}_0 = 732.3524$ | (截距項) |
| $\hat{\beta}_1 = -1.8775$ | 性別比 |
| $\hat{\beta}_2 = 2.2056$ | 扶老比 |
| $\hat{\beta}_3 = -27.5709$ | 低收入戶戶內人數 |
| $\hat{\beta}_4 = -73.4842$ | 綜合所得總額平均數 |
| $\hat{\beta}_5 = -65.3766$ | 人口數 |
| $\hat{\beta}_{14} = 0.1044$ | 性別比與低收入戶戶內人數交互作用 |
| $\hat{\beta}_{15} = 0.13612$ | 性別比與人口數交互作用 |
| $\hat{\beta}_{25} = -0.2562$ | 扶老比與人口數交互作用 |
| $\hat{\beta}_{35} = 1.5900$ | 低收入戶戶內人數人口數交互作用 |
| $\hat{\beta}_{45} = 7.5132$ | 綜合所得總額平均數與人口數交互作用 |

二. 最終模型估計式

$$X' = \begin{pmatrix} 1 \\ X_1 \\ X_2 \\ \log(X_3 + 1) \\ \log(X_4 + 1) \\ \log(X_5 + 1) \\ X_1 \log(X_4 + 1) \\ X_1 \log(X_5 + 1) \\ X_2 \log(X_5 + 1) \\ X_3 \log(X_5 + 1) \\ X_4 \log(X_5 + 1) \end{pmatrix}, W^{\frac{1}{2}} Y^* \hat{\beta} = W^{\frac{1}{2}} \frac{(Y-164)^{0.505}-1}{0.505} \hat{\beta} = W^{\frac{1}{2}} X' \hat{\beta}$$

三. 殘差分析

1. 殘差過大或過小時偏離常態假設
2. 資料過於集中，表示 Fitted value 更加隨機，預測能力差



四. 結語

常態假設、獨立性、同質變異數假設經變數轉換後仍不符合，因此該模型參考價值不高，應是變數選取不周全，此外自然增加人數與出生人數及死亡人數有直接關係，考慮出生人數及死亡人數後，實際情形應更加複雜。