

110 年度 第一學期  
應用迴歸分析 期末報告

可能影響照護人力需求指數變數  
之複迴歸模型探討

指導教授：鄭宗記 老師

系級：統碩一

學號：110354009

姓名：鄧喻安

# 目錄

摘要 .....	4
壹、資料介紹 .....	4
貳、資料特性 .....	5
一、單變數的內容 .....	5
二、散佈圖 .....	8
參、模型變數選取 .....	14
一、遺失值處理 .....	14
1. 直接剔除 .....	14
二、初步模型 .....	14
1. Added-variable .....	14
2. 簡單線性迴歸 .....	14
3. 變數選取 .....	15
4. 共線性診斷 .....	17
5. Partial residual plot .....	17
三、修改後模型 .....	18
1. 共線性診斷 .....	19
2. Partial residual plot .....	19
肆、殘差分析 .....	21
一、檢定常態性 .....	21
二、檢定常數變異數 .....	21
伍、模型解釋 .....	22
陸、    附錄 .....	24
一、資料來源 .....	24
二、變數定義 .....	24
三、概述統計量 (Summary Statistics) .....	25
四、照護人力需求指數與解釋變數的共變異數矩陣 .....	26
五、程式碼 .....	27

## 圖目錄

圖 2-1 照護人力需求指數長條圖 .....	5
圖 2-2 $\log(\text{照護人力需求指數})$ 長條圖 .....	5
圖 2-3 所有解釋變數的長條圖 .....	7
圖 2-4 所有變項散佈圖 .....	8
圖 2-5 照護人力需求與 65 歲以上人口數散佈圖 .....	9
圖 2-6 照護人力需求與人口密度散佈圖 .....	9
圖 2-7 照護人力需求與扶老比散佈圖 .....	10
圖 2-8 照護人力需求與老化指數散佈圖 .....	10
圖 2-9 照護人力需求與環境便利需求指數散佈圖 .....	11
圖 2-10 照護人力需求與住宅狀況需求指數散佈圖 .....	11
圖 2-11 照護人力需求與低收入戶散佈圖 .....	12
圖 2-12 照護人力需求與社區照顧關懷據點家數數散佈圖 ..	12
圖 2-13 照護人力需求與醫療院所平均每家服務人數散佈圖 .....	13
圖 2-14 照護人力需求與綜合所得稅總額申報中位數散佈圖 .....	13
圖 3-1 模型(1)的 partial residual plot.....	17
圖 3-2 模型(2)的 partial residual plot.....	19
圖 4-1 常態機率圖 .....	21

## 表目錄

表 3-1 模型(1)的參數估計及顯著性 .....	16
表 3-2 模型(1)的 VIF 值 .....	17
表 3-3 模型(1) lack-of-fit tests for the residual plots .....	18
表 3-4 模型(2)的參數估計及顯著性 .....	18
表 3-5 模型(2)的 VIF 值 .....	19
表 3-6 模型(2) lack-of-fit tests for the residual plots .....	20
表 4-1 NCV test 報表 .....	21
表 6-1 變數定義 .....	24
表 6-2 反應變數及解釋變數的概述統計量 .....	25
表 6-3 共變異數矩陣 .....	26

## 摘要

由探索性資料分析可以發現反應變數——照護人力需求指數與 65 歲以上人口數、人口密度、低收入戶、社區照顧關懷據點家數、綜合所得稅總額申報中位數皆有正相關，至於解釋變數間的關係，也可以由散佈圖矩陣看出扶老比與老化指數存在線性相關，還有 65 歲以上人口數與社區關懷據點家數和低收入戶家數也有存在線性相關。

最終選取的模型中，解釋變數放入 65 歲以上人口數、人口密度、扶老比、住宅狀況需求指數、低收入戶與綜合所得稅所得總額申報中位數，可以從這些面向去找出哪些縣市的老人照護需求較高，並針對這些面向去改善老人的生活，模型的 R-squared 為 0.786，表示模型可以解釋資料 78.6% 的變異量。殘差檢定的結果也符合常態分佈及常數變異數，加上 F 檢定也有統計上的顯著性。

## 壹、資料介紹

照護人力需求指數為一個綜合性指標，其內容包含三個面向，分別是考量居住型態（單人戶、老老照顧、其他）、有無與子女同住及有無聘用外籍看護，本報告所選取之此指標類型為居住型態是獨居者，且為有子女但未與子女同住（同縣市），且為未聘用外籍看護者，其數字愈高者表示對照護人力的需求相對較高。

此指標之目的是希望能找出最需要幫助老人，而本篇報告希望能找出有什麼其他的變數可能與照護人力需求有關聯，因此選取了十個可能的解釋變數：65 歲以上人口數、人口密度、扶老比、老化指數、環境便利需求指數、住宅狀況需求指數、低收入戶、綜合所得稅所得總額申報中位數、社區照顧關懷據點家數、醫療院所平均每家服務人數，並從可能有相關的面向中去改善老人的生活。

本篇報告之架構，將在第貳部分中進行探索性資料分析，針對反應變數及每個解釋變數分別做初步的了解，並使用散佈圖觀察變數間的可能關聯，第參部分將討論選擇可以配適模型的各個步驟，第肆部分將針對最終的模型進行殘差分析，檢測模型的適合度，第伍部分會針對最終的模型係數進行解釋，最後，有關本報告分析的細節皆會放在第陸部分。

## 貳、資料特性

### 一、單變數的內容

圖 2-1 顯示照護人力需求指數的分佈為右偏，因此對反應變數取 log 後，長條圖分佈為圖 2-2，取 log 後照護人力需求指數的分佈看起來比較對稱。

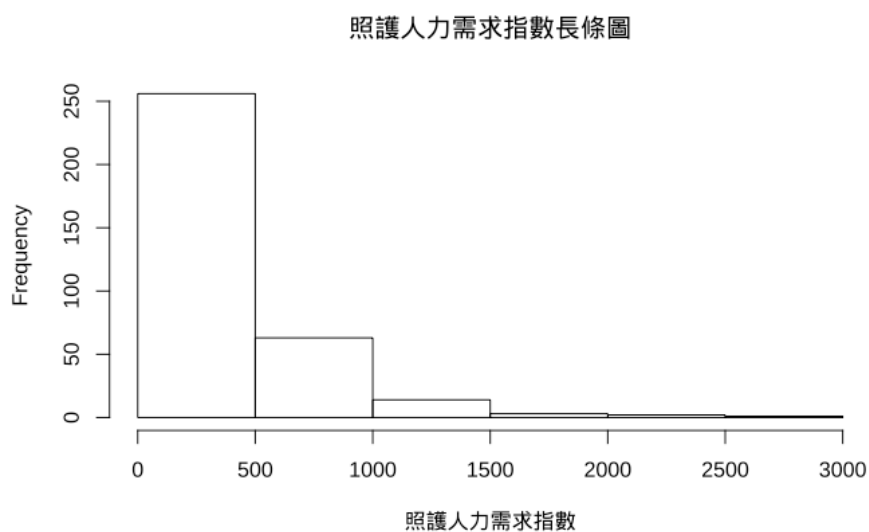


圖 2-1 照護人力需求指數長條圖

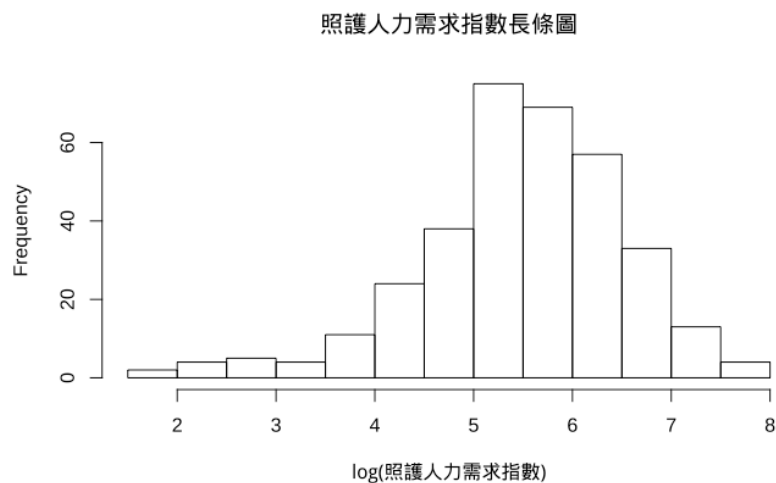
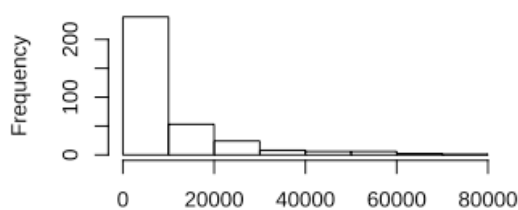


圖 2-2 log(照護人力需求指數)長條圖

觀察圖 2-3，除了扶老比的分佈較對稱外，其餘解釋變數的分佈都相較屬於右偏的資料。65 歲以上人口數最小值為 304 人，發生在高雄市那瑪夏區，最大值為 74819 人，為新北市中和區，平均每鄉鎮市區 10249 人；人口密度最小值

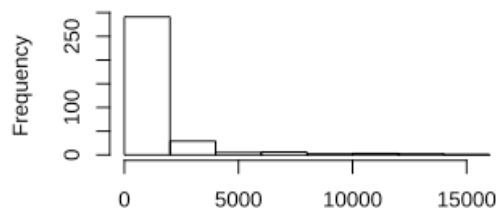
是高雄市桃源區，每單位土地面積內之人口數只有 2.15 人，最大值為新北市蘆洲區，每單位土地面積內之人口數有 14972.7 人；扶老比平均為 26.59，最小為高雄市桃源區 11.75，最大為新北市平溪區；老化指數最高發生在新北市平溪區 722.22，最小值發生在新竹縣竹北市 37.21；環境便利需求指數最小值是花蓮縣卓溪鄉，最大值是台南市後壁區；住宅狀況需求指數最小值是高雄市桃源區，最大值是嘉義市東區；低收入戶家數最小值為連江縣北竿鄉 14 戶，最大值為桃園市桃園區 2559 戶，平均每鄉鎮市區有 377.9 戶；社區照顧關懷據點家數最小值為 0 家，包含很多縣市，最大值是桃園市中壢區有 58 家；醫療院所平均每家服務人數最小值為台北市大安區 338.1 人，最大值為花蓮縣富里鄉 9920 人，平均每鄉鎮市區醫療院所平均每家服務人數為 2086.4 人；綜合所得稅所得總額申報中位數最小值為花蓮縣豐濱鄉 304000 元，最大值為新竹縣竹北市 857000 元，中位數為 386000 元，變數的其他概述統計量細節在附錄的概述統計量章節中。

65歲以上人口數長條圖



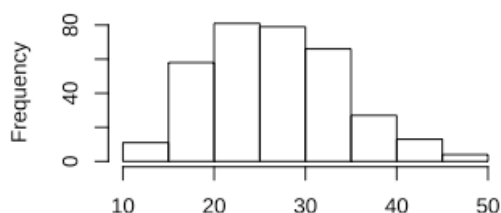
65歲以上人口數

人口密度長條圖



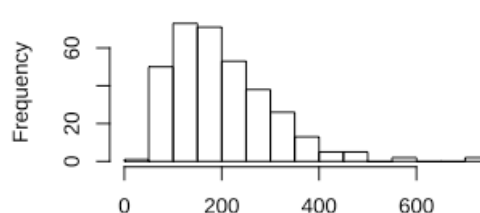
人口密度

扶老比長條圖



扶老比

老化指數長條圖



老化指數

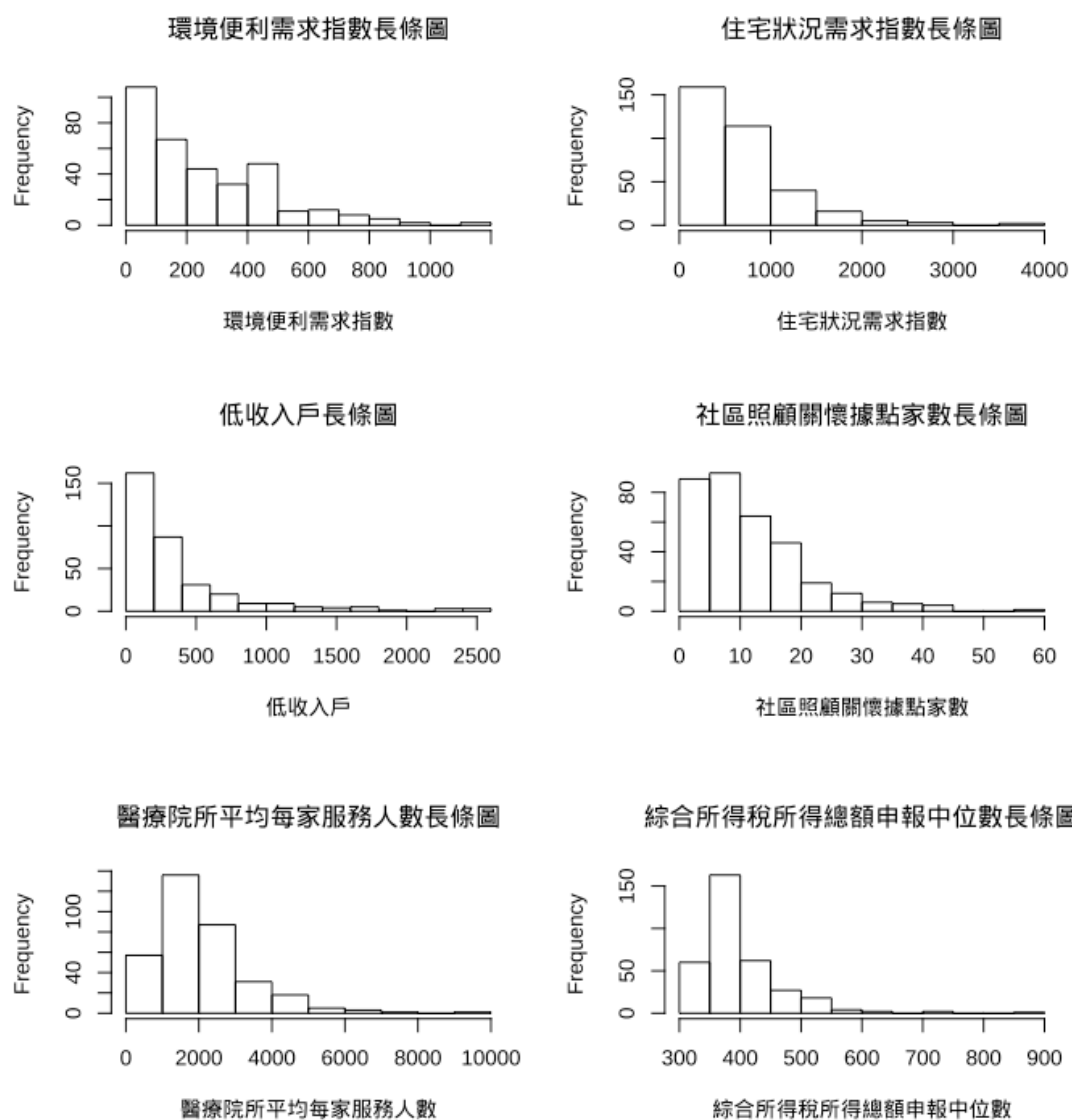


圖 2-3 所有解釋變數的長條圖



## 二、散佈圖

由圖 2-4 可以大致看出所有反應變數與解釋變數間的散佈圖，扶老比與老化指數存在線性相關，兩者間的共變異數高達 0.9057，65 歲以上人口數與社區關懷據點家數也有存在線性相關，兩者間的共變異數為 0.8500，65 歲以上人口數與低收入戶家數也有存在線性相關，兩者間的共變異數為 0.8954，其餘所有變數的共變異數矩陣細節在附錄的概述統計量章節中。

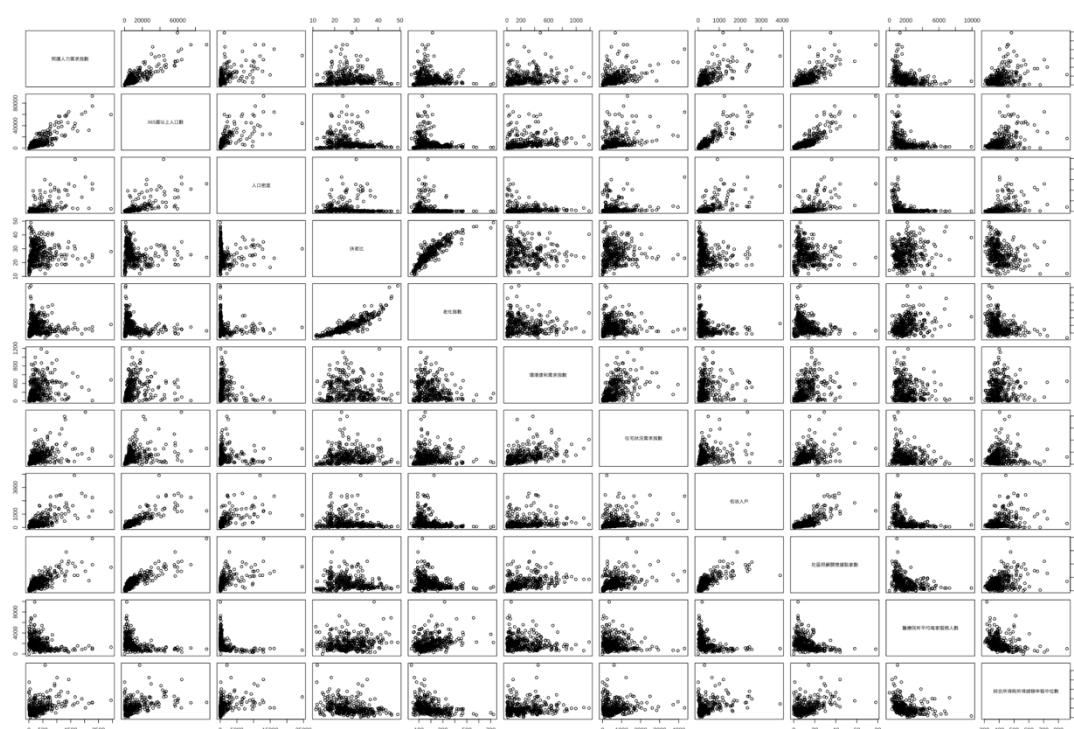


圖 2-4 所有變項散佈圖

以下是各個解釋變數與反應變數的散佈圖，Y 軸為反應變數——照護人力需求指數，X 軸皆為各個解釋變數，黑色線條為資料散佈的趨勢線，紅色虛線為 lowess 配適的 smooth 曲線。

由圖 2-5 可以看出 65 歲以上人口數與反應變數存在正相關。

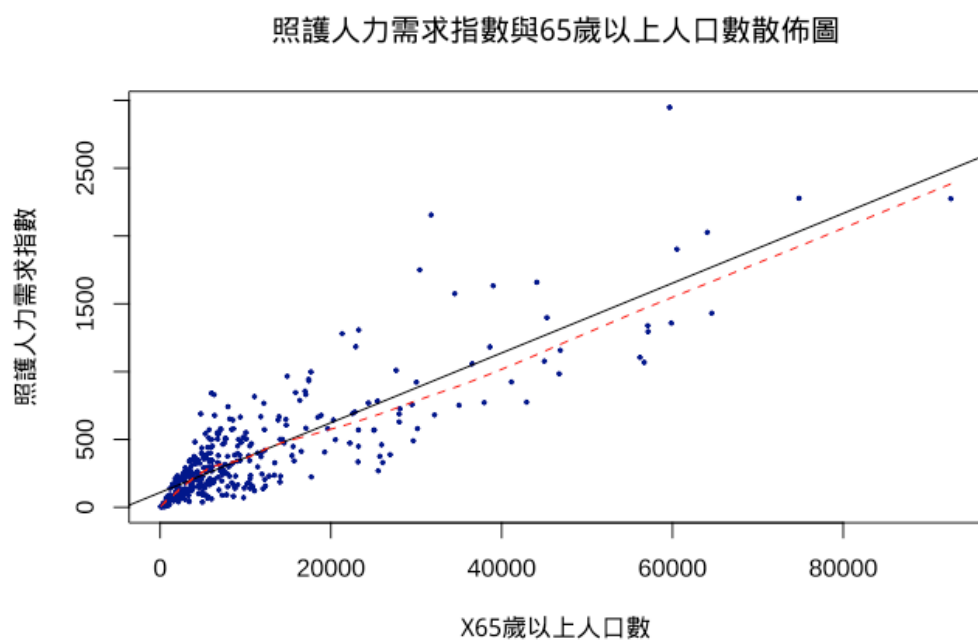


圖 2-5 照護人力需求與 65 歲以上人口數散佈圖

由圖 2-6 可以看出人口密度與反應變數存在些微正相關。

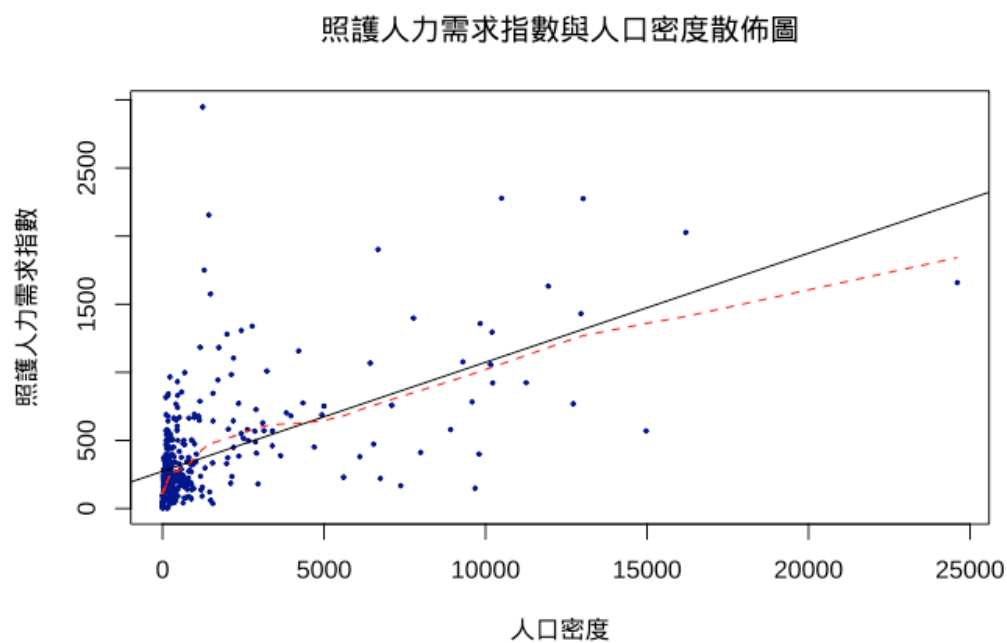


圖 2-6 照護人力需求與人口密度散佈圖

觀察圖 2-7，扶老比與反應變數的關聯性可能不大。

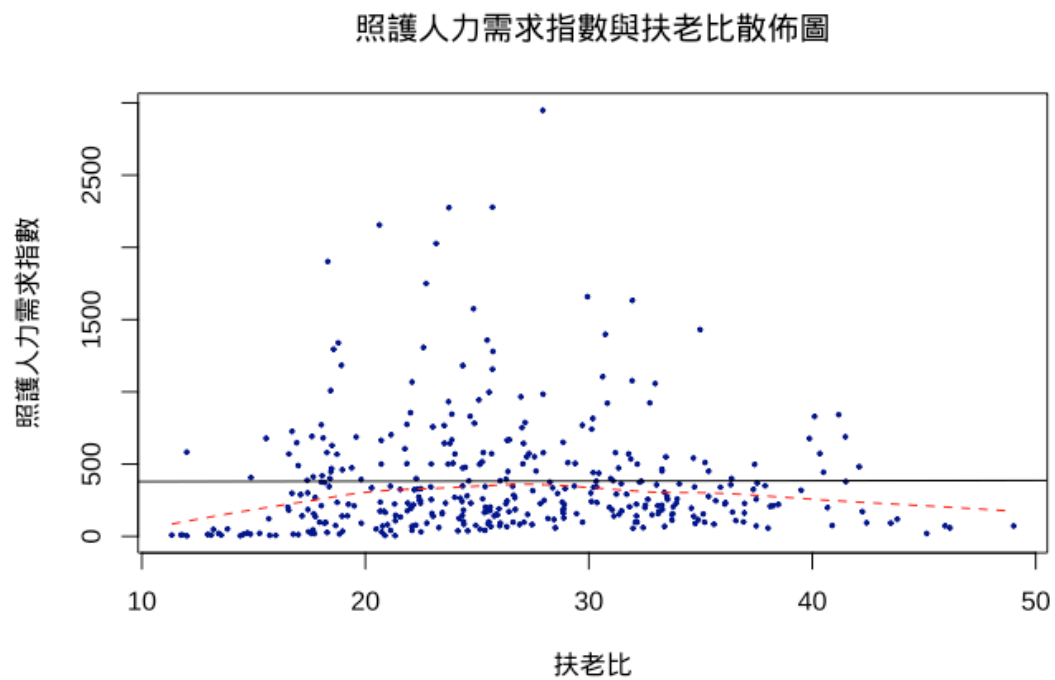


圖 2-7 照護人力需求與扶老比散佈圖

由圖 2-8 可以看出老化指數與反應變數可能存在負相關。

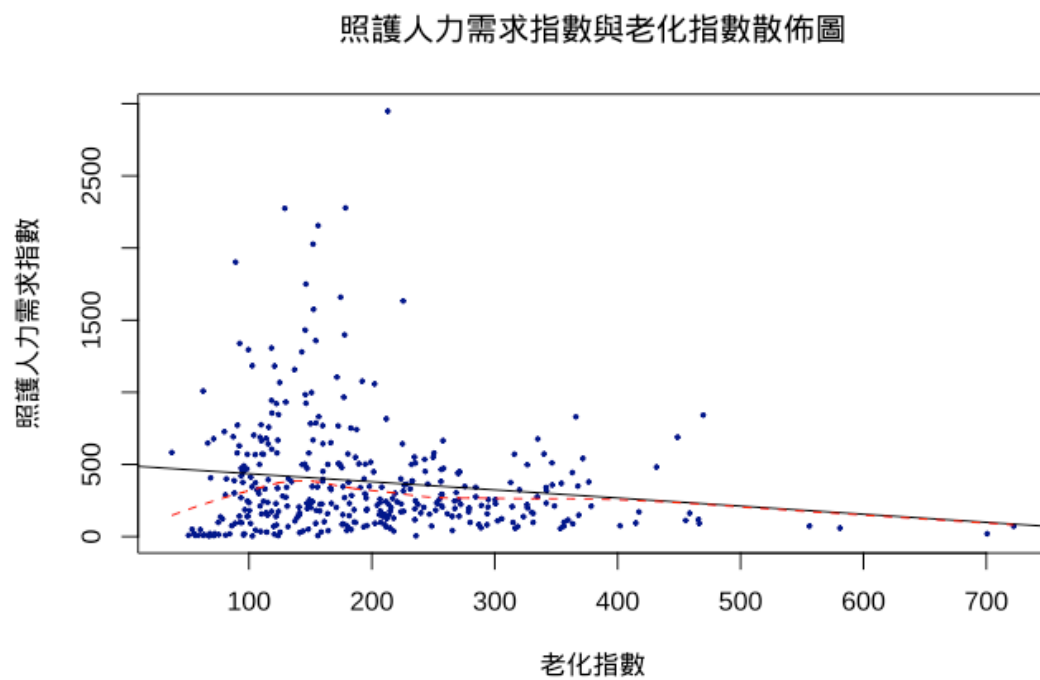


圖 2-8 照護人力需求與老化指數散佈圖

觀察圖 2-9 可知環境便利需求指數與反應變數存在些微正相關。

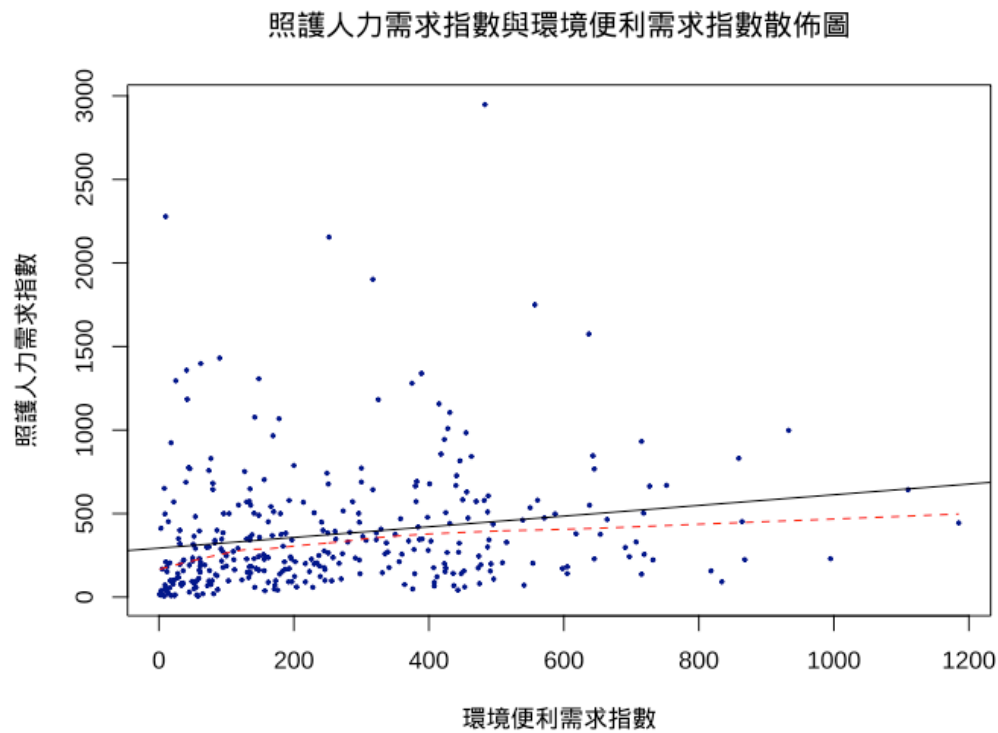


圖 2-9 照護人力需求與環境便利需求指數散佈圖

由圖 2-10 可以看出住宅狀況需求指數與反應變數存在正相關。

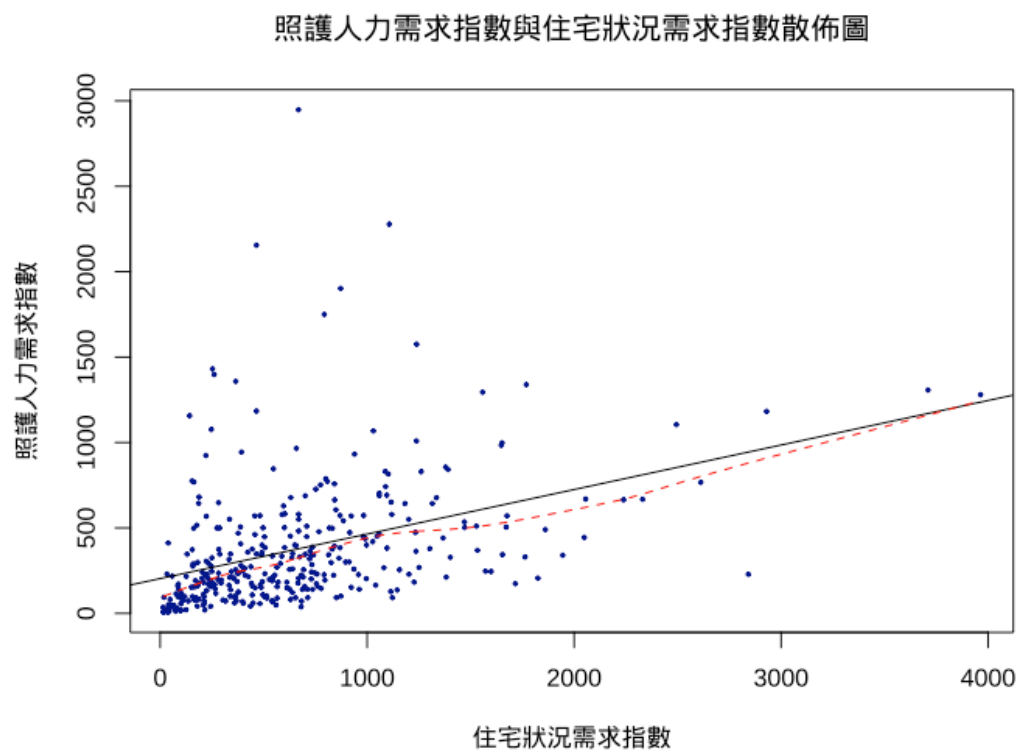


圖 2-10 照護人力需求與住宅狀況需求指數散佈圖

由圖 2-11 可以得知低收入戶家數與反應變數存在正相關。

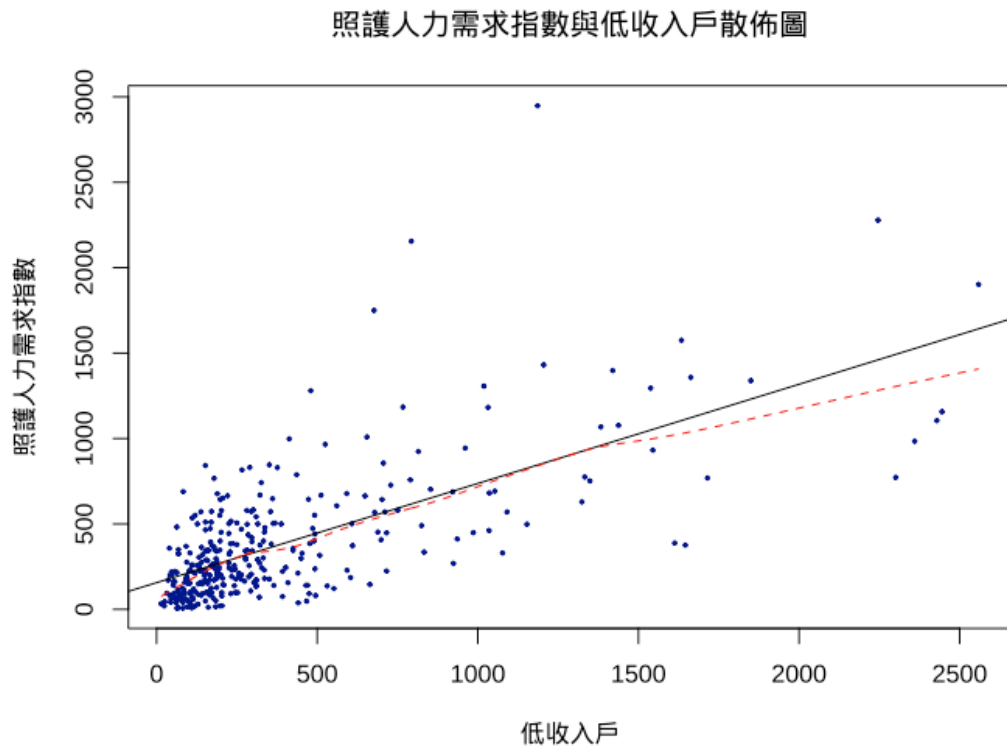


圖 2-11 照護人力需求與低收入戶散佈圖

由圖 2-12 可以看出社區照顧關懷據點家數與反應變數存在正相關。

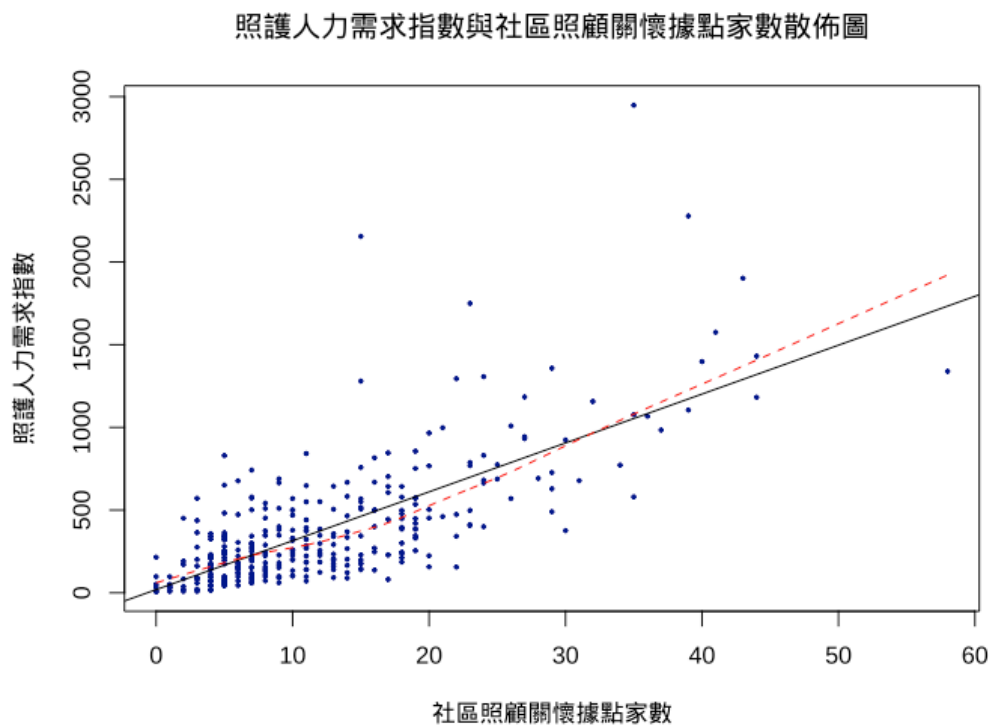


圖 2-12 照護人力需求與社區照顧關懷據點家數散佈圖

由圖 2-13 可以看出醫療院所平均每家服務人數與反應變數可能存在負相關或有平方的關係。

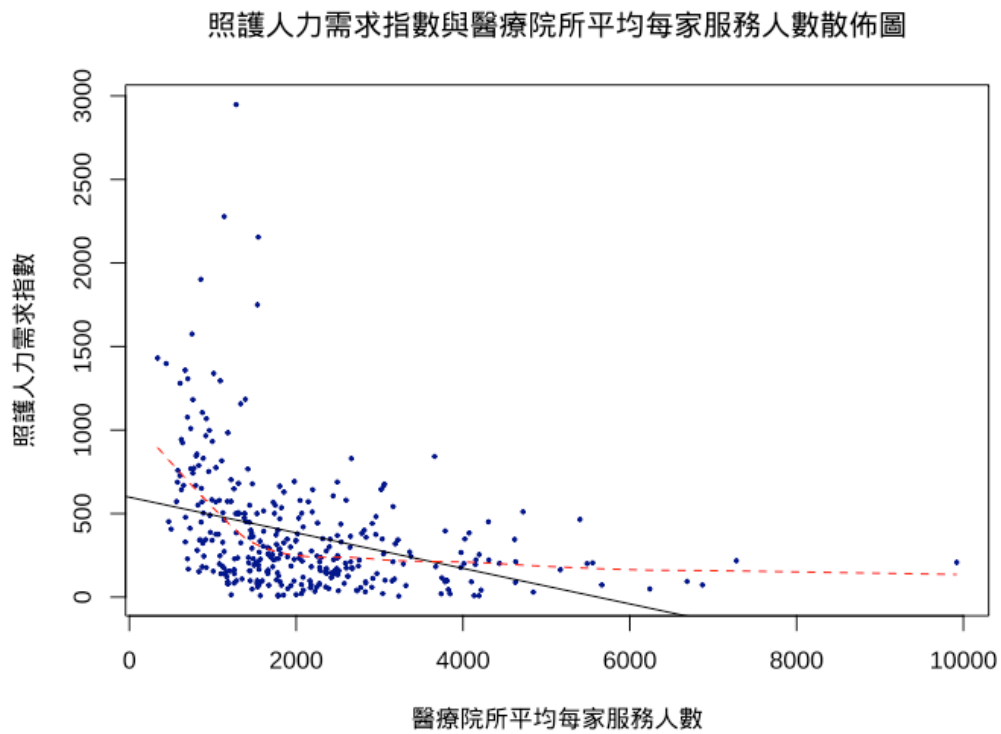


圖 2- 13 照護人力需求與醫療院所平均每家服務人數散佈圖

觀察圖 2-14 得知綜合所得稅總額申報中位數與反應變數可能存在正相關。

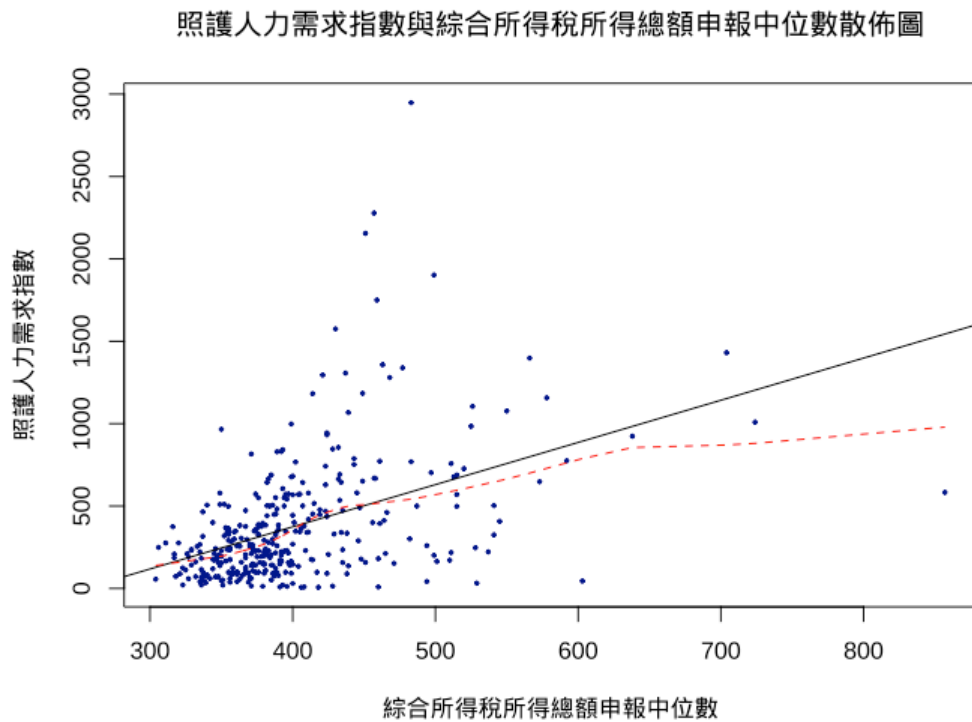


圖 2- 14 照護人力需求與綜合所得稅總額申報中位數散佈圖

## 參、模型變數選取

### 一、遺失值處理

#### 1. 直接剔除

為了瞭解變數與照護人力需求指數的關係，共有 368 筆樣本數，其中，環境便利需求指數這欄有 28 個遺失值，居住狀況需求指數這欄有 1 個遺失值，考量遺失值不多且是隨機性的遺失，所以將此 29 筆資料直接剔除，剔除遺失值後有 339 筆資料。

### 二、初步模型

#### 1. Added-variable

added-variable 決定模型要放入哪些變數，將「人口密度」對「照護人力需求指數」做模型出來的殘差，與「人口密度」對「65 歲以上人口數」做模型出來的殘差，兩個殘差配適模型，檢定結果為顯著，所以在人口密度放入模型中的情況下，決定再多放入 65 歲以上人口數。

#### 2. 簡單線性迴歸

將所有解釋變數放入模型內，報表的結果顯示 65 歲以上人口數、扶老比、人口密度、住宅狀況需求指數、低收入戶皆為顯著的變數，綜合所得稅所得總額申報中位數的 p-value 接近 0.05，因此這裡要決定是否將它放入模型。

將綜合所得稅所得總額申報中位數單獨與反應變數做簡單線性迴歸，報表顯示綜合所得稅所得總額申報中位數與反應變數的關係有統計上的顯著，估計的斜率係數為 2.561， $R\text{-squared} = 0.2046$ ，因此決定將此變數放入模型。

另外，依常理推測社區照顧關懷據點家數可能會與照護人力需求指數有關係，因此將其單獨與反應變數做簡單線性迴歸，報表顯示社區照顧關懷據點家數與反應變數的關係有統計上的顯著，估計的斜率係數為 29.567，R-squared = 0.52，因此將此變數放入模型。因此修正後模型為

$$\begin{aligned} \text{照護人力需求指數} = & 65 \text{ 歲以上人口數} + \text{扶老比} + \\ & \text{人口密度} + \text{住宅狀況需求指數} + \text{低收入戶} + \\ & \text{綜合所得稅所得總額申報中位數} + \text{社區照顧關懷據點家數} \end{aligned}$$

將每個解釋變數與反應變數都單獨做簡單線性迴歸模型後，發現扶老比不顯著，且 R-square 為 0.002381，因此初步決定扶老比從模型中拿掉，模型修正為

$$\begin{aligned} \text{照護人力需求指數} = & 65 \text{ 歲以上人口數} + \\ & \text{人口密度} + \text{住宅狀況需求指數} + \text{低收入戶} + \\ & \text{綜合所得稅所得總額申報中位數} + \text{社區照顧關懷據點家數} \end{aligned}$$

從散佈圖及共變異數矩陣中，可以發現社區照顧關懷據點家數與 65 歲以上人口數有很高的相關性，因此在模型中社區照顧關懷據點家數就變得不顯著，加上從散佈圖中可以發現兩者有高度線性相關，所以決定將社區照顧關懷據點家數從模型中拿掉，因此由簡單線性迴歸方法挑選出來的模型變數為：

$$\begin{aligned} \text{照護人力需求指數} = & 65 \text{ 歲以上人口數} + \text{人口密度} + \text{住宅狀況需求指數} \\ & + \text{低收入戶} + \text{綜合所得稅所得總額申報中位數} \end{aligned}$$

### 3. 變數選取

嘗試完 added-variable 方式與簡單線性迴歸的方式後，再使用程式進行變數選取，由於此報告的目的是要找出有哪些變數會跟照護人力需求指數有關，目的不是在於個體預測，所以不使用「正規化」(regularization)的變數選取方式，此篇目的在尋找解釋變數，所以採用「子集合選取法」(subset selection)，並使用向前選取、向後選取及逐步選取三種方式。

使用向前選取、向後選取及逐步選取三種變數選取的方法，參考三種方式



跑出來的結果後，決定放入「65 歲以上人口數」、「扶老比」、「人口密度」、「住宅狀況需求指數」、「低收入戶」、「綜合所得稅所得總額申報中位數」六個變數，初次決定模型為

$$\log(\text{照護人力需求指數}) = 65 \text{ 歲以上人口數} + \text{扶老比} + \\ \text{人口密度} + \text{住宅狀況需求指數} + \text{低收入戶} + \\ \text{綜合所得稅所得總額申報中位數}$$

此模型 AIC=3587.82，與簡單線性迴歸方式挑選出來的變數大致相同，只多了扶老比，然扶老比單獨與反應變數做簡單線性迴歸模型並無統計上的不顯著，且 R-square 為 0.002381，在此推測扶老比雖然單獨對照護人力需求指數的解釋能力不佳，但可能在複迴歸模型中與其他變數一起會有加成性的影響，所以在子集合選取法中被判定放入模型中。

此模型中反應變數——照護人力需求指數是取 log，估計出來的模型(1)如下：

$$\widehat{\log(\text{照護人力需求指數})} \\ = 2.82 + 0.000047 \text{ 65 歲以上人口數} - 0.00017 \text{ 人口密度} \\ + 0.038 \text{ 扶老比} + 0.00052 \text{ 住宅狀況需求指數} \\ + 0.000042 \text{ 低收入戶} + 0.0021 \text{ 綜合所得稅所得總和申報中位數}$$

表 3-1 模型(1)的參數估計及顯著性

	Estimate	Std. Error	t value	Pr(> t )
Intercept	2.822687585	0.360970417	7.820	***
65 歲以上人口數	0.000047260	0.000008719	5.421	***
人口密度	-0.000016581	0.000029037	-0.571	0.56837
扶老比	0.037909212	0.005780679	6.558	***
住宅狀況需求指數	0.000517478	0.000075869	6.821	***
低收入戶	0.000041912	0.000202560	0.207	0.83620
綜合所得稅所得總額 申報中位數	0.002059527	0.000735647	2.800	**

Multiple R-squared: 0.5521, Adjusted R-squared: 0.544

F-statistic: 68.22 on 6 and 332 DF, p-value: < 0.000000000000000022

#### 4. 共線性診斷

接著進行共線性診斷，結果如表 3-2，VIF 值皆小於 10，因此判定模型解釋變數間的共線性問題不大。

表 3-2 模型(1)的 VIF 值

變數	VIF 值
65 歲以上人口數	7.705737
人口密度	2.463555
扶老比	1.180372
住宅狀況需求指數	1.24096
低收入戶	5.364571
綜合所得稅所得總額申報中位數	1.600395

#### 5. Partial residual plot

此模型對所有 X 的殘差如圖 3-1，可以看出解釋變數的殘差並非隨機分佈，

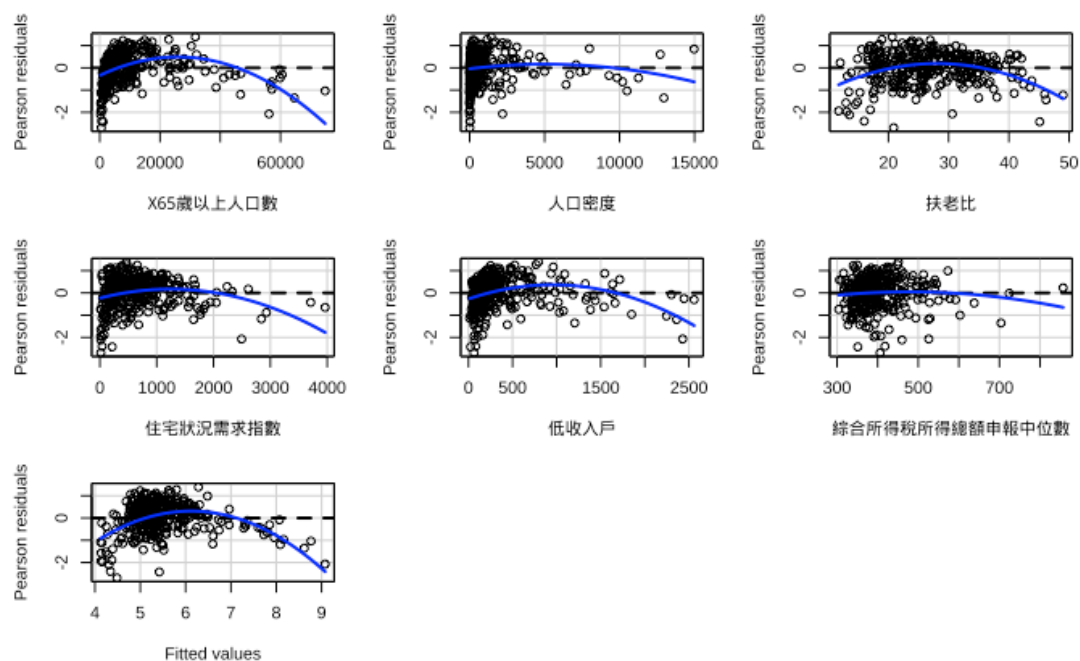


圖 3-1 模型(1)的 partial residual plot

另外再看殘差圖的 lack-of-fit 檢定如表 3-3，發現有許多變數的顯著水準都很小，表示有足夠證據顯示多個變數的殘差圖皆非常數。

表 3-3 模型(1) lack-of-fit tests for the residual plots

	Test stat	Pr(> Test stat )
65 歲以上人口數	-10.3099	< 0.00000000000000022 ***
人口密度	-2.1412	0.03299 *
扶老比	-6.6503	0.00000000012099 ***
住宅狀況需求指數	-4.8477	0.00000192445407 ***
低收入戶	-6.9698	0.00000000001725 ***
綜合所得稅所得總額申報中位數	-1.3897	0.16556
Tukey test	-11.6077	< 0.00000000000000022 ***

### 三、修改後模型

從檢定結果得知，先對 65 歲以上人口數、人口密度、扶老比做變數轉換，各自都取 log 後再配適一次模型，估計出來的模型(2)如下：

$$\begin{aligned} \widehat{\log(\text{照護人力需求指數})} &= -6.598 + 0.876 \log(65 \text{ 歲以上人口數}) - 0.000050 \text{ 人口密度} \\ &+ 1.156 \log(\text{扶老比}) + 0.000041 \text{ 住宅狀況需求指數} \\ &- 0.000056 \text{ 低收入戶} + 0.00188 \text{ 綜合所得稅所得總和申報中位數} \end{aligned}$$

表 3-4 模型(2)的參數估計及顯著性

	Estimate	Std. Error	t value	Pr(> t )
Intercept	-6.59826313	0.52635854	-12.536	***
log(65 歲以上人口數)	0.87628265	0.04441994	19.727	***
人口密度	-0.00004966	0.00001883	-2.638	**
log(扶老比)	1.15593064	0.10434060	11.078	***
住宅狀況需求指數	0.00004106	0.00005795	0.709	0.479108
低收入戶	-0.00005648	0.00009976	-0.566	0.571670
綜合所得稅所得總額申報中位數	0.00187959	0.00049493	3.798	***

Multiple R-squared: 0.786, Adjusted R-squared: 0.7821

F-statistic: 203.2 on 6 and 332 DF, p-value: < 0.00000000000000022

## 1. 共線性診斷

接著進行共線性診斷，結果如表 3-5，VIF 值皆小於 10，因此判定模型變數間的共線性問題不大。

表 3-5 模型(2)的 VIF 值

變數	VIF 值
log(65 歲以上人口數)	3.160575
人口密度	2.167833
log(扶老比)	1.187520
住宅狀況需求指數	1.515352
低收入戶	2.723276
綜合所得稅所得總額申報中位數	1.516008

## 2. Partial residual plot

此模型對所有 X 的殘差如圖 3-2，可以看出解釋變數的殘差圖大致都有呈現隨機性及一致性。

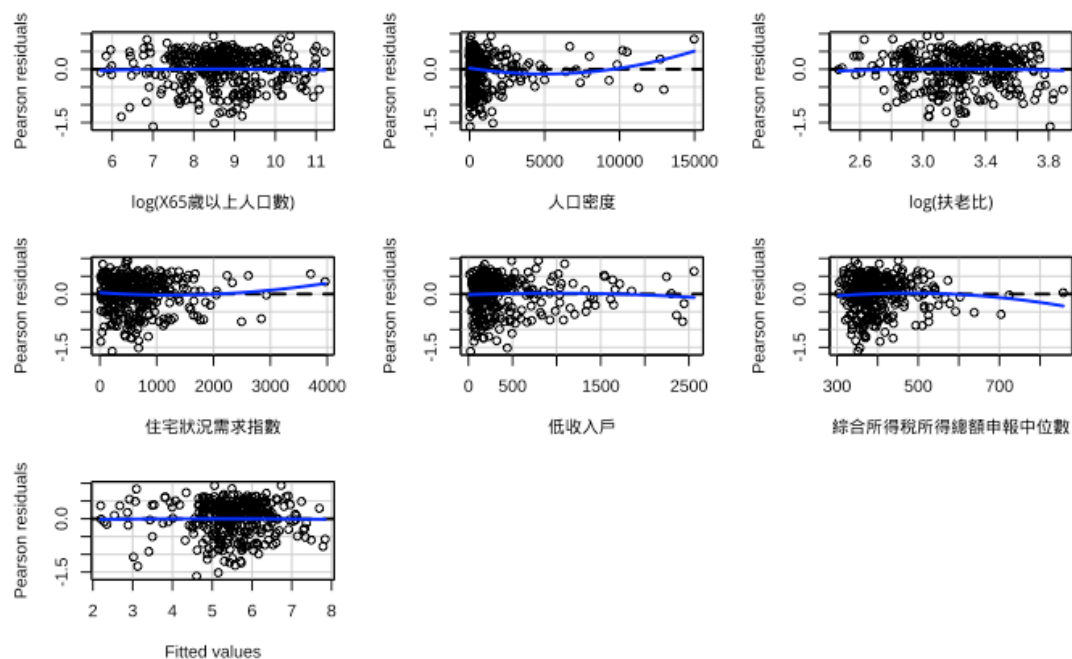


圖 3-2 模型(2)的 partial residual plot

殘差圖的 lack-of-fit 檢定的檢定結果如表 3-6，只剩人口密度的殘差圖是非 null plot，不過整體模型的檢定結果顯著性不大。

表 3-6 模型(2) lack-of-fit tests for the residual plots

	Test stat	Pr(> Test stat )
log(65 歲以上人口數)	-0.5250	0.599917
人口密度	2.6376	0.008743 **
log(扶老比)	-0.4027	0.687417
住宅狀況需求指數	1.2497	0.212293
低收入戶	-0.7139	0.475806
綜合所得稅所得總額申報中位數	-1.0317	0.302953
Tukey test	-0.1820	0.855554

## 肆、殘差分析

### 一、檢定常態性

從圖 3-3 常態機率圖看起來，殘差有服從常態性，因此判定模型是合適的。

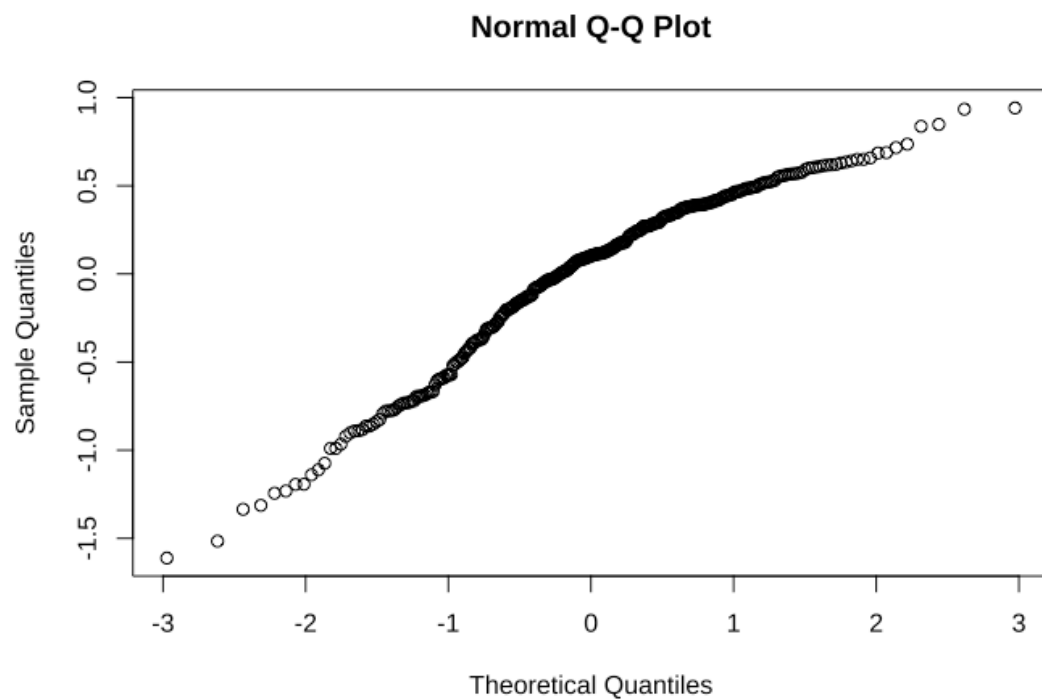


圖 4-1 常態機率圖

### 二、檢定常數變異數

從表 4-1 檢定結果看起來，顯著性不小，表示無足夠證據顯示殘差的變異數不服從常數，因此常態變異數的檢定通過，因此判定模型為合適的模型。

表 4-1 NCV test 報表

Non-constant Variance Score Test		
Variance formula: ~ fitted.values		
Chisquare = 2.451265	Df = 1	p = 0.11743

## 伍、模型解釋

最終配適之模型 R-squared 為 0.786，表示模型可以解釋資料 78.6% 的變異量，另外，殘差檢定也符合常態及變異數常數性，加上 F 檢定也有統計上的顯著性，因此判定此模型為合適的模型。針對各解釋變數與反應變數的關係，因為有對反應變數及一些解釋變數進行 log 的轉換因此係數的解釋會稍微有所差異，而模型的係數解釋如下：

### 一、65 歲以上人口數

當其他變數固定的情況下，65 歲以上人口數與照護人力需求指數有正相關，65 歲以上人口數增加 10%，照護人力需求指數平均乘以  $\exp(0.1 \times 0.876) = 1.092$ ，表示照護人力需求指數會增加  $100\% (1.092 - 1) = 9.2\%$ ，因此 65 歲以上人口數愈多之鄉鎮市，其對於照護人力的需求就愈高。

### 二、人口密度

當其他變數不變的情況下，人口密度與照護人力需求指數之模型係數為負，表示人口密度增加一單位，照護人力需求指數平均減少  $100(\exp(-0.00005) - 1) = -0.005\%$ ，推測可能是因為較需要照護的老人可能較偏向住在人口密度較低之非都市集中區域。

### 三、扶老比

在其他變數固定的情況下，扶老比增加 10%，照護人力需求指數平均乘以  $\exp(0.1 \times 1.156) = 1.123$ ，亦即照護人力需求指數會增加  $100\% (1.123 - 1) = 12.3\%$ ，表示扶老比愈高之區域，愈需要照護人力，此結果也符合常理。

### 四、住宅狀況需求指數

在其他變數固定的情況下，住宅狀況需求指數增加一單位，照護人力需求指數平均增加  $100(\exp(0.000041) - 1) = 0.0041\%$ ，雖兩者為正相關，但影響不大。

## 五、低收入戶

在其他變數固定的情況下，低收入戶與照護人力需求指數之模型係數為負，表示低收入戶數量增加一單位，照護人力需求指數平均減少  $100(\exp(-0.000056) - 1) = -0.0056\%$ ，表示該地區低收入戶數量愈多，照護人力需求可能愈高。

## 六、綜合所得稅所得總額申報中位數

在其他變數固定的情況下，綜合所得稅所得總額申報中位數與照護人力需求指數為正相關，表示綜合所得稅所得總額申報中位數增加一千元，照護人力需求指數平均增加  $100(\exp(0.00188) - 1) = 0.188\%$ ，表示該地區綜合所得稅所得總額申報中位數愈高，照護人力需求可能愈高。



## 陸、附錄

### 一、資料來源

SEGIS 社會經濟資料服務平台，網址：

[https://segis.moi.gov.tw/STAT/Web/Platform/QueryInterface/STAT\\_QueryInterface.aspx?Type=0](https://segis.moi.gov.tw/STAT/Web/Platform/QueryInterface/STAT_QueryInterface.aspx?Type=0)，擷取日期為 2021 年 12 月 15 日。

### 二、變數定義

表 6-1 變數定義

變數	定義
照護人力需求指數	來自「108 年行政區銀髮安居資料之照護人力需求指數_鄉鎮市區」資料檔，取英文欄位名為 N13N22N32 變項，係為居住型態是獨居者，且係為有子女未與子女同住（同縣市），且係為未聘用外籍看護。
65 歲以上人口數	「110 年 9 月行政區三段年齡組性別人口統計_鄉鎮市區」資料檔中取 65 歲以上人口數。
人口密度	「110 年 9 月行政區人口指標_鄉鎮市區」資料檔中取人口密度，係為每單位土地面積內之人口數。
扶老比	「110 年 9 月行政區人口指標_鄉鎮市區」資料檔中取扶老比， $\text{扶老比} = \frac{65 \text{ 歲以上人口}}{15-64 \text{ 歲人口}} \times 100$ 。
老化指數	「110 年 9 月行政區人口指標_鄉鎮市區」資料檔中取老化指數， $\text{老化指數} = \frac{65 \text{ 歲以上人口}}{0-14 \text{ 歲人口}} \times 100$ 。
環境便利需求指數	「108 年行政區銀髮安居資料之環境便利需求指數_鄉鎮市區」資料檔中取英文欄位名為 L12L23L32，係為公車站牌距離大於或等於 100 公尺且小於 500 公尺者，且便利超商距離大於或等於 500 公尺者，且醫院診所距離大於或等於 200 公尺且小於 1,000

變數	定義
	公尺者。
住宅狀況需求指數	「108 年行政區銀髮安居資料之住宅狀況需求指數_鄉鎮市區」資料檔中取英文欄位名為 E12E23E32，係為屋齡大於或等於 30 年，且不知是否是有電梯公寓者，且房屋結構為非鋼骨或非鋼筋混凝土結構。
低收入戶	「110 年 6 月行政區低收入戶統計指標_鄉鎮市區」資料檔中取低收入戶變項，係為該單位內的低收入戶數量。
社區照顧關懷據點家數	「110 年 6 月行政區社區照顧關懷據點統計_鄉鎮市區」資料檔，為該單位內的社區照顧關懷據點家數，指社會福利服務機構資料中，該年(月)底之社區照顧關懷據點家數。
醫療院所平均每家服務人數	「110 年 6 月行政區醫療院所統計_鄉鎮市區」資料檔中取醫療院所平均每家服務人數變項，指年(月)底之醫療院所平均每家服務之人口數。
綜合所得稅所得總額申報中位數	「108 年綜合所得稅所得總額申報統計_鄉鎮市區」資料檔中的中位數，單位為千元。

### 三、概述統計量 (Summary Statistics)

表 6-2 反應變數及解釋變數的概述統計量

	平均數	中位數	標準差	最小值	最大值
照護人力需求指數	375.7	256	374.3	6	2948
65 歲以上人口數	10249	5919	12240.4	304	74819
人口密度	1026.3	300.68	2078.1	2.15	14972.7
扶老比	26.6	25.71	7.2	11.8	49
老化指數	199.7	179.3	103.6	37.2	722.2
環境便利需求指數	258.1	193	222	1	1185
住宅狀況需求指數	660.9	542	564.5	14	3964
低收入戶	377.9	206	439.6	14	2559
社區照顧關懷據點家數	12.1	10	9.1	0	58

	平均數	中位數	標準差	最小值	最大值
醫療院所平均每家服務人數	2086.4	1785.2	1251.7	338.1	9920
綜合所得稅所得總額申報中位數	400.6	386	66.1	304	857

#### 四、照護人力需求指數與解釋變數的共變異數矩陣

表 6-3 共變異數矩陣

	照護人力需求指數	65 歲以上人口數	人口密度	扶老比	老化指數	環境便利需求指數
照護人力需求指數	1.0000	0.8172	0.5122	-0.0488	-0.1723	0.1895
65 歲以上人口數	0.8172	1.0000	0.7351	-0.2409	-0.3500	0.1627
人口密度	0.5122	0.7351	1.0000	-0.1696	-0.2658	-0.1026
扶老比	-0.0488	-0.2409	-0.1696	1.0000	0.9057	-0.0924
老化指數	-0.1723	-0.3500	-0.2658	0.9057	1.0000	-0.1437
環境便利需求指數	0.1895	0.1627	-0.1026	-0.0924	-0.1437	1.0000
住宅狀況需求指數	0.3928	0.3111	0.0308	0.0180	-0.0948	0.4554
低收入戶	0.6820	0.8954	0.6541	-0.2740	-0.3631	0.1448
社區照顧關懷據點家數	0.7211	0.8500	0.5679	-0.2314	-0.3456	0.3329
醫療院所平均每家服務人數	-0.3564	-0.4200	-0.3600	0.1952	0.3337	-0.1833
綜合所得稅所得總額申報中位數	0.4523	0.5390	0.4830	-0.3286	-0.3852	0.0568

	住宅狀況需求指數	低收入戶	社區照顧關懷據點家數	醫療院所平均每家服務人數	綜合所得稅所得總額申報中位數
照護人力需求指數	0.3928	0.6820	0.7211	-0.3564	0.4523
65 歲以上人口數	0.3111	0.8954	0.8500	-0.4200	0.5390
人口密度	0.0308	0.6541	0.5679	-0.3600	0.4830
扶老比	0.0180	-0.2740	-0.2314	0.1952	-0.3286
老化指數	-0.0948	-0.3631	-0.3456	0.3337	-0.3852

環境便利需求指數	0.4554	0.1448	0.3329	-0.1833	0.0568
住宅狀況需求指數	1.0000	0.2531	0.3820	-0.2409	0.1097
低收入戶	0.2531	1.0000	0.8004	-0.3965	0.4235
社區照顧關懷據點 家數	0.3820	0.8004	1.0000	-0.4147	0.4490
醫療院所平均每 家服務人數	-0.2409	-0.3965	-0.4147	1.0000	-0.2881
綜合所得稅所得總 額申報中位數	0.1097	0.4235	0.4490	-0.2881	1.0000

## 五、程式碼

```
library(plyr)
data1 <- read.csv("1.csv",header = T,sep = ",")
data2 <- read.csv("2.csv",header = T,sep = ",")
data2 <- data2[,c(-1,-2)]
data3 <- read.csv("3.csv",header = T,sep = ",")
data4 <- read.csv("4.csv",header = T,sep = ",")
data5 <- read.csv("5.csv",header = T,sep = ",")
newdata <- join(data1,data2,by=c("鄉鎮市區代碼","鄉鎮市區名稱"),type="inner")
newdata <- join(newdata,data3,by=c("鄉鎮市區代碼","鄉鎮市區名稱"),type="inner")
newdata <- join(newdata,data4,by=c("鄉鎮市區代碼","鄉鎮市區名稱"),type="inner")
newdata <- join(newdata,data5,by=c("鄉鎮市區代碼","鄉鎮市區名稱"),type="inner")
data_r <- newdata[,c(-1,-2,-3,-4)] ##去掉縣市代碼那些
str(data_r)
summary(data_r)

rm.data <- data_r[complete.cases(data_r), ] #去除有遺失值的資料 #剩 339 筆
str(rm.data)
summary(rm.data)
```

```

apply(rm.data,2,sd)

#install.packages("mice")
#library(mice)
#md.pattern(data_r)

#讓圖顯示中文字
library(showtext)
showtext_auto(enable = TRUE)

#正式開始處理資料
install.packages("car")
installed.packages("carData")
install.packages("effests")
library(alr4)

## step1-1:探索性分析__單變數
hist#####
summary(rm.data)
hist(rm.data$照護人力需求指數,main = "照護人力需求指數長條圖",xlab = "照護
人力需求指數")
hist(log(rm.data$照護人力需求指數),main = "照護人力需求指數長條圖",xlab =
"log(照護人力需求指數)")
qqnorm(log(rm.data$照護人力需求指數))

par(mfrow=c(2,2))
hist(rm.data$X65 歲以上人口數,main = "65 歲以上人口數長條圖",xlab = "65 歲以
上人口數")
hist(rm.data$人口密度,main = "人口密度長條圖",xlab = "人口密度")
hist(rm.data$扶老比,main = "扶老比長條圖",xlab = "扶老比")
hist(rm.data$老化指數,main = "老化指數長條圖",xlab = "老化指數")
hist(rm.data$環境便利需求指數,main = "環境便利需求指數長條圖",xlab = "環境
便利需求指數")
hist(rm.data$住宅狀況需求指數,main = "住宅狀況需求指數長條圖",xlab = "住宅
狀況需求指數")

```

```
hist(rm.data$低收入戶,main = "低收入戶長條圖",xlab = "低收入戶")
hist(rm.data$社區照顧關懷據點家數,main = "社區照顧關懷據點家數長條圖",xlab = "社區照顧關懷據點家數")
hist(rm.data$醫療院所平均每家服務人數,main = "醫療院所平均每家服務人數長條圖",xlab = "醫療院所平均每家服務人數")
hist(rm.data$綜合所得稅所得總額申報中位數,main = "綜合所得稅所得總額申報中位數長條圖",xlab = "綜合所得稅所得總額申報中位數")
```

```
## step1-2:散佈圖 #####
```

```
#每個變項散佈圖
```

```
pairs(照護人力需求指數~,data = data_r)
```

```
#每個 x 和 y 的散佈圖
```

```
plot(照護人力需求指數~X65 歲以上人口數,data=rm.data,
      main = "照護人力需求指數與 65 歲以上人口數散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~X65 歲以上人口數,data=rm.data), lty=1) #abline 看清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~X65 歲以上人口數, f=6/10, iter=1),
lty=2,col=2)) #lowess 配適 smooth 線
```

```
plot(照護人力需求指數~人口密度,data=rm.data,
      main = "照護人力需求指數與人口密度散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~人口密度,data=rm.data), lty=1) #abline 看清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~人口密度, f=6/10, iter=1),
lty=2,col=2)) #lowess 配適 smooth 線
```

```
plot(照護人力需求指數~扶老比,data=rm.data,
      main = "照護人力需求指數與扶老比散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~扶老比,data=rm.data), lty=1) #abline 看清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~扶老比, f=6/10, iter=1), lty=2,col=2))
#lowess 配適 smooth 線
```

```
plot(照護人力需求指數~老化指數,data=rm.data,
      main = "照護人力需求指數與老化指數散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~老化指數,data=rm.data), lty=1) #abline 看清楚資料
分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~老化指數, f=6/10, iter=1),
lty=2,col=2)) #lowess 配適 smooth 線
```

```
plot(照護人力需求指數~環境便利需求指數,data=rm.data,
      main = "照護人力需求指數與環境便利需求指數散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~環境便利需求指數,data=rm.data), lty=1) #abline 看
清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~環境便利需求指數, f=6/10, iter=1),
lty=2,col=2)) #lowess 配適 smooth 線
```

```
plot(照護人力需求指數~住宅狀況需求指數,data=rm.data,
      main = "照護人力需求指數與住宅狀況需求指數散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~住宅狀況需求指數,data=rm.data), lty=1) #abline 看
清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~住宅狀況需求指數, f=6/10, iter=1),
lty=2,col=2)) #lowess 配適 smooth 線
```

```
plot(照護人力需求指數~低收入戶,data=rm.data,
      main = "照護人力需求指數與低收入戶散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~低收入戶,data=rm.data), lty=1) #abline 看清楚資料
分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~低收入戶, f=6/10, iter=1),
lty=2,col=2))
```

```
plot(照護人力需求指數~社區照顧關懷據點家數,data=rm.data,
      main = "照護人力需求指數與社區照顧關懷據點家數散佈圖",
```

```

col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~社區照顧關懷據點家數,data=rm.data), lty=1)
#abline 看清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~社區照顧關懷據點家數, f=6/10,
iter=1), lty=2,col=2))

plot(照護人力需求指數~醫療院所平均每家服務人數,data=rm.data,
      main = "照護人力需求指數與醫療院所平均每家服務人數散佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~醫療院所平均每家服務人數,data=rm.data), lty=1)
#abline 看清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~醫療院所平均每家服務人數,
f=6/10, iter=1), lty=2,col=2))

plot(照護人力需求指數~綜合所得稅所得總額申報中位數,data=rm.data,
      main = "照護人力需求指數與綜合所得稅所得總額申報中位數散
佈圖",
      col="dark blue",cex=0.5,pch=20)
abline(lm(照護人力需求指數~綜合所得稅所得總額申報中位數,data=rm.data),
lty=1) #abline 看清楚資料分布趨勢
with(rm.data, lines(lowess(照護人力需求指數~綜合所得稅所得總額申報中位數,
f=6/10, iter=1), lty=2,col=2))

## step2-1:配適迴歸線 #####
mfull <- lm(照護人力需求指數 ~ ., data=rm.data)
summary(mfull)
#variance covariance matrix
as.data.frame(cor(rm.data))
as.data.frame(round(cor(rm.data),4))    ##共變異數矩陣
#residual Plots
plot(mfull)
residualPlots(mfull)

## step2-2:added-variable plots
#####

```



```

m1 <- lm(照護人力需求指數 ~ 人口密度, data=rm.data)
r1 <- residuals(m1)
m2 <- lm(X65 歲以上人口數 ~ 人口密度, data=rm.data)
r2 <- residuals(m2)
m4 <- lm(resid(m1) ~ resid(m2))
summary(m4)
plot(resid(m2), m4$fitted.values)

## effect plot
m5<- lm(formula = 照護人力需求指數 ~ X65 歲以上人口數 + 人口密度 + 扶老
比 + 住宅狀況需求指數 +低收入戶+綜合所得稅所得總額申報中位數,
        data = rm.data)
plot(Effect("人口密度", m5), grid=TRUE, rug=TRUE)
plot(Effect("X65 歲以上人口數", m4), grid=TRUE, rug=TRUE)
plot(Effect("扶老比", m4), grid=TRUE, rug=TRUE)
plot(Effect("住宅狀況需求指數", m4), grid=TRUE, rug=TRUE)
plot(Effect("低收入戶", m4), grid=TRUE, rug=TRUE)
plot(Effect("綜合所得稅所得總額申報中位數", m4), grid=TRUE, rug=TRUE)

## step2-3:簡單線性迴歸
#####
#y 和每個 x 單獨做模型
m_community_care <- lm(照護人力需求指數 ~ 社區照顧關懷據點家數,
data=rm.data)
summary(m_community_care)

m_income_tax <- lm(照護人力需求指數 ~ 綜合所得稅所得總額申報中位數,
data=rm.data)
summary(m_income_tax)

m_65 <- lm(照護人力需求指數 ~ X65 歲以上人口數, data=rm.data)
summary(m_65)

m_pop_density <- lm(照護人力需求指數 ~ 人口密度, data=rm.data)
summary(m_pop_density)

```

```

m_old_ratio <- lm(照護人力需求指數 ~ 扶老比, data=rm.data) #不顯著
summary(m_old_ratio)

m_residential <- lm(照護人力需求指數 ~ 住宅狀況需求指數, data=rm.data)
summary(m_residential)

m_low_income <- lm(照護人力需求指數 ~ 低收入戶, data=rm.data)
summary(m_low_income)

m_environment <- lm(照護人力需求指數 ~ 環境便利需求指數, data=rm.data)
summary(m_environment)

m_hospital <- lm(照護人力需求指數 ~ 醫療院所平均每家服務人數,
data=rm.data)
summary(m_hospital)

m_aging_index <- lm(照護人力需求指數 ~ 老化指數, data=rm.data)
summary(m_aging_index)

## step2-4:變數選取#####
f <- ~ X65 歲以上人口數 + 人口密度 + 扶老比 +老化指數+環境便利需求指數+
  住宅狀況需求指數 +低收入戶+綜合所得稅所得總額申報中位數+
  社區照顧關懷據點家數+醫療院所平均每家服務人數
#向前選取
m0 <- lm(照護人力需求指數 ~ 1, rm.data) # the base model
m.forward <- step(m0, scope=f, direction="forward")
#向後選取
m1 <- update(m0, f) #full model
m.backward <- step(m1, direction="backward")
#stepwise
m.stepwise <- step(m1, direction="both")

## step3:殘差分析#####
#####修正的模型#####

```

```

#model1
model1log <- lm(log(照護人力需求指數) ~ X65 歲以上人口數 + 人口密度 +
               扶老比 + 住宅狀況需求指數 +
               低收入戶 + 綜合所得稅所得總額申報中位數,

data=rm.data)
summary(model1log)
install.packages("car")
library(car)
vif(model1log) ###共線性診斷
plot(model1log)
residualPlots(model1log) #partial residual plot
ks.test(scale(model1log$residuals),pnorm) #檢定殘差是否常態
qqnorm(model1log$residuals) #常態機率圖

#model2
model2 <- lm(log(照護人力需求指數) ~ log(X65 歲以上人口數) + 人口密度 +
             log(扶老比) + 住宅狀況需求指數 +
             低收入戶 + 綜合所得稅所得總額申報中位數,

data=rm.data)
summary(model2)
vif(model2)
residualPlots(model2)#Tukey test 檢定是檢定殘差及所有 x 殘差是否是 null
plot , 不拒絕 ( p-value>0.05 ) 才是對的
ks.test(scale(model2$residuals),pnorm) #檢定殘差是否常態
qqnorm(model2$residuals) #常態機率圖
plot(model2)
ncvTest(model2) #檢定變異數常態

```