

逢甲大學學生報告 ePaper

空氣汙染與氣象因子、其他汙染物之關係

The Relationship between Air Pollution, Meteorological
Factors, and Other Pollutants

作者：陳柏荏、康展碩、鍾秉霖、林逸軒、李怡青、廖家鴻、陳昱誠、蔡秉翰

系級：統計三乙

學號：D1373632、D1242639、D1282736、D1242570、D1223203、D1282629、
D1282719、D1242256

開課老師：劉峰旗

課程名稱：迴歸分析

開課系所：統計學系

開課學年：114 學年度 第 1 學期

摘要

近年來大眾對於空氣品質要求越來越高，而 PM_{2.5} 是影響空氣品質的關鍵因子，所以我們想探討的是氣象因子與汙染物對於 PM_{2.5} 濃度有哪些的顯著影響，分析其顯著性與相關性；至於氣象因素方面，檢視氣溫、濕度、風速、風向與氣壓是否跟 PM_{2.5} 有存在顯著的相關性。在汙染物方面，檢視 PM_{2.5} 與 NO₂、SO₂ 與 CO 等汙染物是否有存在顯著的相關性？此外在不同汙染物間的在特定氣象條件下具有交互作用是否有影響 PM_{2.5} 的增加或減少，可能進一步影響 PM_{2.5} 濃度變化。

我們採用 SAS EG 進行分析，主要運用多元線性回歸來觀測資料。過程包含透過散步圖進行初步觀察，接著進行參數估計檢定，為確保模型的嚴謹性，分析過程中執行了 VIF、Durbin-Watson 以及殘差分析，並針對偏離常態的資料進行剔除，以優化模型的完整性。

PM₁₀(空氣懸浮粒子)是預測 PM_{2.5} 的關鍵指標，而 SO₂、NO₂ 及 NO_x 亦展現正向影響，且風速與風向等氣象因子在線性模型中則呈現隨機分布。最終模型排除 CO 與離群值後，解釋力明顯提升，則證明該變數組合具備良好的環境預測。則未來可應用於環境監測預警甚至是引入其他分析以提升預測精度。

關鍵字：空氣汙染物(PM_{2.5})、氣象因子、細懸浮微粒濃度

Abstract

In recent years, the public has placed increasingly higher demands on air quality, and PM_{2.5} is a key factor affecting air quality. Therefore, we aim to explore the significant impacts of meteorological factors and pollutants on PM_{2.5} concentrations, analyzing their significance and correlation. Regarding meteorological factors, we examine whether temperature, humidity, wind speed, wind direction, and air pressure have significant correlations with PM_{2.5}. As for pollutants, we examine whether PM_{2.5} has significant correlations with NO₂, SO₂, and CO, and whether interactions between different pollutants under specific meteorological conditions affect the increase or decrease of PM_{2.5}, potentially further influencing PM_{2.5} concentration changes.

We used SAS EG for analysis, primarily employing multiple linear regression to observe the data. The process included initial observation through scatter plots, followed by parameter estimation and testing. To ensure model rigor, VIF, Durbin-Watson, and residual analyses were performed during the analysis, and data deviating from normal patterns were removed to optimize model integrity.

PM₁₀ (airborne particulate matter) is a key indicator for predicting PM_{2.5}, while SO₂, NO₂, and NO_x also show a positive impact. Furthermore, meteorological factors such as wind speed and direction exhibit a random distribution in the linear model. After excluding CO and outliers, the explanatory power of the final model significantly improves, demonstrating that this combination of variables possesses good environmental predictive capabilities. Therefore, it can be applied to environmental monitoring and early warning systems, and even incorporated into other analyses to enhance prediction accuracy in the future.

Keyword : PM_{2.5} Concentration, Meteorological Factors, Fine Particulate Matter

目 次

第一章、前言.....	4
第一節、研究架構	4
第二節、研究問題與章節介紹	4
第二章、文獻探討	5
第三章、研究分析與結果	8
第四章、結論與建議	51
第一節、研究結論	51
第二節、研究建議	52
參考文獻	53
附錄一、分工表.....	54

第一章、前言

第一節、研究架構

本研究旨在探討 PM_{2.5}與氣象因子及其他汙染物之間的關聯性，鑑於 PM_{2.5}形成機制複雜，本研究架構主要由「氣象因子」與「空氣汙染物」兩大維度組成，探討其對 PM_{2.5}濃度之顯著性影響與預測能力。

在變數設定上，本研究選取溫度、相對濕度、風速、風向及氣壓作為指標，並且納入二氧化氮、二氧化硫、一氧化碳等汙染物變數。首先透過線性回歸模型量化個因子之影響程度，並驗證高風速對於 PM_{2.5}之稀釋結果，進而分析不同汙染物間之交互作用是否對 PM_{2.5}濃度的影響。在分析之方法應用上，透過迴歸模型的參數估計與模型建構，將參數估計的結果建立預測模型，透過 ANOVA 檢定模型整體的顯著性，確認自變數對 PM_{2.5}是否有顯著的解釋能力。而在變數顯著性檢定與共線性診斷上，針對個別變數的係數估計值進行 t 檢定檢測參數的顯著性，並同時執行 VIF 檢定診斷自變數間是否存在多重共線性問題，以確保參數估計的穩定性。最後，在模型診斷與優化過程中，透過殘差分析檢視模型是否符合迴歸誤差項的假設，如常態性、獨立性與分散一致性等假設，且透過離群值的診斷，將異常的觀察值進行排除，以避免異常的極端值對於此迴歸模型估計的干擾，並確保本研究獲得可信的估計結果。

第二節、研究問題與章節介紹

以下為本研究分析欲透過統計分析方法進行解決的研究問題：

問題 1：氣象因子（如：溫度、相對濕度、風速、風向、氣壓）對 PM_{2.5} 有哪些顯著影響？

問題 2：高風速是否能有效降低 PM_{2.5} 濃度？

問題 3：PM_{2.5} 與 NO₂、SO₂、CO、O₃ 等汙染物是否存在顯著的相關性？

問題 4：不同汙染物之間的交互作用是否會影響 PM_{2.5} 增加或減少？

問題 5：其他汙染物是否能作為 PM_{2.5} 的預測因子？

本研究分析報告的章節安排，第一章前言包含研究動機與研究問題，第二章文獻探討針對三篇相關文獻進行說明，第三章研究分析結果說明本分析的結果發現，第四章結論與建議提供本研究發現與分結結果之建議，最後，在參考文獻中提供本研究所引用的相關文獻。

第二章、文獻探討

一、警惕呼吸道的隱形殺手：PM_{2.5} 與老年 COPD 的急診危機

在台灣高度都市化的環境中，空氣品質對高齡族群的健康威脅日益顯著。根據 Ding 等人於 2017 年發表的研究指出，細懸浮微粒 PM_{2.5} 已成為誘發 65 歲以上老年人慢性阻塞性肺病（COPD）急性發作，進而導致急診就診的首要元兇。

這類微粒之所以如此致命，是因為其粒徑極小，能直接穿透人體呼吸道的過濾系統，深入肺部底層甚至進入血液循環。對於氣道本就脆弱的老年 COPD 患者，PM_{2.5} 會啟動三大致病路徑：首先是微粒攜帶的重金屬與毒素會產生自由基，引發劇烈的炎症反應，導致氣道腫脹與痰液激增；其次是直接損害肺功能，讓患者原本就有限的通氣能力雪上加霜，引發嚴重呼吸困難；最後則是削弱免疫力，透過破壞肺部巨噬細胞，讓病毒與細菌有機可乘，增加肺部感染的風險。

該研究特別提醒，環境對健康的打擊並非僅限於暴露當下，而是具有兩大必須警惕的行為特性。第一是延遲的威脅，即所謂的滯後效應，空污暴露後的影響具有延遲性，急診就診的高峰通常出現在空污發生後的 2 至 7 天，這意味著即便空氣品質轉好後的一週內，仍不可掉以輕心。第二是寒流與空污的雙重夾擊，當高濃度 PM_{2.5} 遇上低溫氣候時，冷空氣引發的支氣管痙攣會讓微粒更容易沉積在肺部，在兩者的協同效應下，患者掛急診的風險將達到最高峰。

針對這些發現，專家建議應採取更精準的防護措施。研究數據顯示 65 歲以上的男性患者風險尤其顯著，因此公共衛生與個人照護應建立雙軌預警機制。照護者不應只關注即時的空氣品質指標，還需結合低溫特報來共同評估健康風險。同時，在空污或寒流發生後的一週內，家屬應加強監測，密切觀察長輩是否有咳嗽加劇或喘不過氣等急性徵兆以便及時就醫。總之，乾淨的空氣對老年 COPD 患者而言是不可或缺的保命符，唯有了解 PM_{2.5} 的滯後性與低溫交互作用，才能在關鍵時刻守護長輩的呼吸健康。

二、台灣大氣的呼吸脈絡：氣象因子對 PM_{2.5} 長期濃度趨勢的調節作用

這篇由 Cheng 等人於 2019 年發表的研究，深度剖析了台灣 PM_{2.5} 濃度的長期演變，並特別強調了自然氣象條件在污染防治過程中扮演的關鍵角色。研究的核心背景建立在台灣過去數十年間為了改善空氣品質所做的減排努力，然而透過長期的大數據分析，作者發現儘管人為排放量已顯著減

少，大氣環境的變異卻在很大程度上決定了汙染物在地面層的最終濃度與分布。

研究指出，台灣的 PM_{2.5} 具有極為顯著的季節性特徵，這主要受到大氣穩定度與邊界層高度的影響。在秋末至隔年春季期間，由於大氣層結較為穩定，邊界層高度降低，如同在城市上方蓋了一個厚重的蓋子，導致汙染物極易在大氣中累積而不易擴散；相較之下，夏季因對流旺盛、降水頻繁的洗刷作用，大氣稀釋能力較強，濃度普遍較低。

在氣象影響機制方面，文獻詳細剖析了東亞冬季季風等大型環流系統如何主導台灣的空污事件。當冷高壓出海並轉為偏東風時，中央山脈的屏障作用使得台灣西半部處於背風側，形成的弱風區與局部環流會讓本土汙染物陷入停滯狀態，進而產生嚴重的空污累積。此外，研究也警示了全球氣候變遷帶來的間接威脅，例如靜穩天氣天數的增加，這解釋了為何即便在排放源持續減量的背景下，極端高濃度的汙染事件在特定氣候條件下依然會發生。

數據分析顯示，台灣北、中、南部的 PM_{2.5} 雖然整體呈現下降態勢，但改善速率卻因地域氣候差異而有所不同。中南部地區受限於地形特徵與氣象因子的交互作用，其改善成效常被不利的擴散條件所抵消。研究總結認為，未來的空氣品質管理策略不能僅停留在單純的排放減量，必須將長期的氣象變異性納入評估框架中。這項研究為台灣制定更具韌性的環境政策提供了科學實證，提醒我們在對抗空污的進程中，必須更精確地掌握大氣運行的規律。

三、台灣大氣的無形推手：天氣分型對空氣汙染物分布的決定性影響

這篇由 Su 等人於 2019 年發表的研究，深入探討了綜觀天氣型態（Synoptic Weather Patterns）與台灣空氣品質之間的動態關聯。研究的核心背景指出，台灣的空氣汙染問題不僅取決於局部的人為排放量，大尺度氣壓系統的移動與演變更是決定汙染物擴散或堆積的關鍵物理機制。透過長期的氣象觀測與汙染物濃度數據比對，作者識別出幾種特定的天氣分型，並量化了每種分型對台灣北、中、南部空氣品質的不同衝擊。

研究指出，台灣空氣品質惡化的主要關鍵在於東亞冷高壓系統的路徑。當冷高壓中心位於中國大陸並帶來強勁的東北季風時，雖然可能伴隨境外傳輸的汙染物，但由於地表風速強、擴散條件佳，台灣整體的空污濃度通常相對較低。然而，一旦冷高壓中心東移出海，台灣附近的盛行風轉為偏東風，情況便會發生劇烈變化。此時，中央山脈對偏東風產生了顯著的屏障作用，導致台灣西半部地區處於背風側的弱風區，形成穩定的大氣

層結與沈降氣流，這使得本地產生的汙染物極易陷入滯留狀態，導致 PM_{2.5} 與臭氧濃度迅速飆升。

在不同天氣型態的演變過程中，研究特別強調了「高壓迴流」與「鋒前暖區」對環境健康的威脅。在高壓迴流型態下，西半部的海陸風環流會變得顯著，汙染物隨季節性環流在沿海與內陸之間往返移動，形成長期的累積效應。而在鋒面抵達前的暖區控制下，大氣穩定度極高，且常伴隨逆溫層的出現，如同在城市上空形成了一層透明的隔絕膜，將車輛與工業排放限制在極淺的近地面層，這解釋了為何在特定季節，西半部常會出現連續數日的嚴重霾害。

數據分析顯示，地理特徵與天氣分型的交互作用具有強烈的地域性。北部地區對境外傳輸較為敏感，而中南部地區則更易受到地形引發的背風弱風區影響。研究總結認為，僅靠傳統的定點排放控制已不足以應對多變的天氣挑戰。這項研究為台灣的空污預報系統提供了重要的科學支撐，建議環保決策應更具彈性，根據綜觀天氣圖預測的系統演變，提前啟動動態的減排降載措施，才能在不利擴散的天氣分型來襲前，有效減緩環境壓力的累積。



第三章、研究分析結果

本章旨在呈現本研究之實證分析流程與主要結果，內容依序包括研究變數之操作性定義、敘述統計分析、變數間關係之初步探索、迴歸模型建構與診斷，以及模型修正與最適模型選取結果。透過系統化之統計分析程序，說明氣象因子與其他空氣汙染物對 PM_{2.5} 濃度之影響機制。

一、變數解釋

本研究以細懸浮微粒濃度 (PM_{2.5}_CONC) 作為應變數，用以衡量空氣汙染程度。解釋變數方面，納入多項主要空氣汙染物指標，包括二氧化硫濃度 (SO₂)、一氧化碳濃度 (CO)、二氧化氮濃度 (NO₂)、氮氧化物濃度 (NO_x)，以及粒徑小於 10 微米之懸浮微粒濃度 (PM₁₀)，以探討其對 PM_{2.5} 濃度之影響。此外，考量氣象條件可能影響汙染物之擴散與累積，本研究亦將風速乘上風向轉換為東西向 (W_SPEED_SIN) 與南北向 (W_SPEED_COS) 之向量分量，作為氣象因子變數，以期更精確描述風場特性。

表 1、變數說明

代號	變數名稱	說明
Y	PM _{2.5} CONC (細懸浮微粒)	應變變數。空氣中粒徑小於 2.5 微米的懸浮微粒濃度。
X1	SO ₂ _CONC (二氧化硫)	解釋變數。空氣中二氧化硫 (SO ₂) 的體積濃度。
X2	CO_CONC (一氧化碳)	解釋變數。空氣中一氧化碳 (CO) 的體積濃度。
X3	NO ₂ _CONC (二氧化氮)	解釋變數。空氣中二氧化氮 (NO ₂) 的體積濃度。
X4	NO _x _CONC (一氧化氮)	解釋變數。空氣中一氧化氮 (NO) 加二氧化氮 (NO ₂) 的體積濃度。
X5	PM ₁₀ _CONC	解釋變數，粒徑小於 10 微米的懸浮微粒濃度。
X6	W_SPEED_SIN	解釋變數。風速乘上 sin 值
X7	W_SPEED_COS	解釋變數。風速乘上 cos 值

二、敘述統計

(一)變數 Y 與自變數 X1 至 X7 的敘述性統計分析

本研究首先針對應變數 Y 與七個自變數 (X1 至 X7) 進行敘述性統計分析，以檢視資料的集中趨勢、離散程度及分佈型態。分析結果彙整於表 (請自行標註您的表格編號)。樣本觀測值總數介於 936 至 1000 筆之間，顯示部分變數存在缺失值，其中變數 X4 的有效樣本數最低 (N=936)。

(二)集中趨勢與離散程度

就應變數 Y 而言，其平均值為 17.39，中位數為 16.00，兩者數值相近，標準差為 7.15，顯示應變數之分佈相對穩定。在自變數方面，各變數間的數值尺度 (Scale) 存在顯著差異。變數 X5 具有最大的平均值 (38.78)，而變數 X2 的平均值最小 (0.33)。特別值得注意的是變數 X4，其平均值為 16.47，但標準差高達 13.26，變異係數 (CV) 約為 0.81，顯示該變數的資料波動程度極大。此外，X4 的最大值達 131.30，遠高於其第三四分位數 (Q3=19.70)，初步判斷該變數內部存在顯著的極端大值 (Outliers)。

(三)分佈型態與常態性檢定

透過偏態係數 (Skewness) 與峰度係數 (Kurtosis) 檢視資料的分佈形狀。統計結果顯示，除 X6 外，其餘變數之偏態係數皆為正值，呈現不同程度的右偏 (Right-skewed) 分佈。應變數 Y 的偏態係數為 1.07，峰度為 2.12，屬於輕微右偏，大致符合常態分佈假設。在各自變數中，變數 X1、X2 與 X4 呈現嚴重違反常態分佈的特徵。其中 X4 的偏態係數高達 4.20，峰度係數達 25.66；X1 與 X2 的峰度係數亦分別高達 13.99 與 15.01。此種高偏態與高峽峰 (Leptokurtic) 的特徵，意味著資料高度集中於平均數附近，但同時具有「厚尾」特性，即存在少數極端異常值。

表 2、各變數的敘述統計

統計指標	Y	X1	X2	X3	X4	X5	X6	X7
平均值	17.393	1.540	0.326	12.083	16.474	38.782	0.754	1.558
中位數	16	1.4	0.29	11	13.7	36	0.403	1.390
眾數	16	1.3	0.22	7	14	32	0	0
標準差	7.154	0.799	0.148	7.689	13.263	15.245	1.816	2.043
變異數	51.193	0.639	0.021	59.128	175.92	232.435	3.299	4.175

最小值	0	0	0.13	0	1.8	2	3.556	7.513
最大值	58	8	1.65	43	131.3	101	7.447	11.643
第 1 四分位數	12	1.1	0.23	6	9.3	29	0.477	0.216
第 2 四分位數	16	1.4	0.29	11	13.7	36	0.403	1.390
第 3 四分位數	21	1.8	0.37	16	19.7	46	1.737	2.455
偏態	1.070	2.556	3.046	1.036	4.196	0.756	0.899	0.958
峰度	2.121	13.992	15.007	1.258	25.657	0.685	1.048	5.269
個數	975	962	962	959	936	972	1000	1000

三、散布圖的分析

(一)Y、X1 散布圖

整體而言，隨著 SO_2 濃度的上升， $\text{PM}_{2.5}$ 濃度呈現明顯的增加趨勢。此現象顯示兩者間存在正相關性，符合環境科學中 SO_2 作為硫酸鹽氣膠 (Sulfate Aerosols) 前驅物，進而衍生 $\text{PM}_{2.5}$ 的學理機制。數據點呈現「扇形」發散狀。當 SO_2 濃度較低 (< 2.0) 時， $\text{PM}_{2.5}$ 的變異程度較小；然而隨著 SO_2 濃度升高， $\text{PM}_{2.5}$ 的波動範圍顯著擴大。此非均齊變異性暗示在高污染情境下， $\text{PM}_{2.5}$ 的生成可能受到更多複雜因素 (如濕度、風速或其他化學反應) 的干擾。圖形右側 ($\text{SO}_2 > 6.0$) 出現少數高槓桿點 (High Leverage Points)，顯示在極端高 SO_2 排放事件下， $\text{PM}_{2.5}$ 濃度未必隨之呈線性爆發，這可能與氣象擴散條件有關。

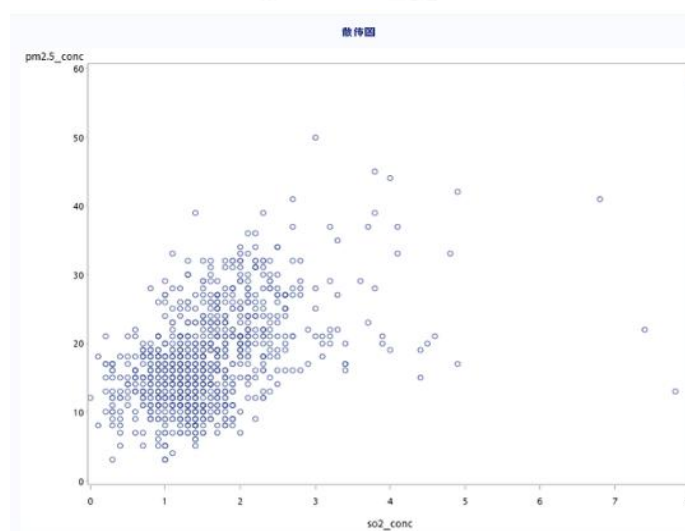


圖 1、Y 與 X1 的散布圖

(二)Y、X2 散布圖

絕大多數觀測值密集分佈於 CO 濃度 0.2 至 0.5 之間，形成一垂直高密度帶。這反映了該研究區域的 CO 背景濃度相對穩定且偏低。雖然整體呈現正向趨勢（CO 越高，PM_{2.5}越高），但數據點較為鬆散。特別是在 CO 濃度較高（>0.8）的區域，數據點數量稀疏且呈現長尾分佈（Long-tail），部分觀測值顯示即使 CO 濃度高達 1.2 以上，PM_{2.5}濃度仍維持在 20 左右的中低水平，顯示 CO 對 PM_{2.5}的解釋力可能存在特定條件限制。

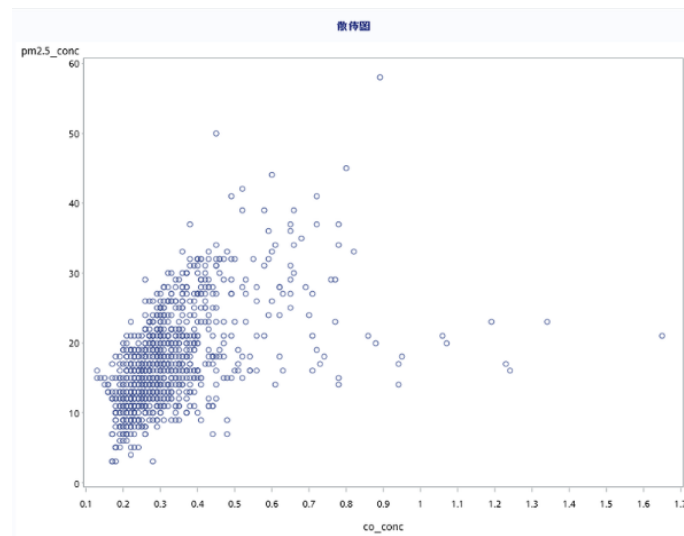


圖 2、Y 與 2 的散布圖

(三)Y、X3 散布圖

數據點呈現清晰的向右上方爬升趨勢，顯示 NO₂ 濃度與 PM_{2.5}濃度之間存在強烈的正向關聯。這印證了交通源排放的氮氧化物（NO_x）轉化為硝酸鹽（Nitrate）是 PM_{2.5}的重要組成來源。與 SO₂ 類似，NO₂ 數據亦表現出顯著的異質變異性。在低濃度區間（NO₂ < 10），PM_{2.5}隨 NO₂ 快速上升且分佈集中；然而當 NO₂ 濃度超過 20 之後，數據點變得極為發散（例如在 NO₂ 約 40 時，PM_{2.5}可從 15 波動至 58），這暗示在高 NO₂ 濃度環境下，大氣擴散條件或其他氣象因子對 PM_{2.5}濃度的干擾效應增強。

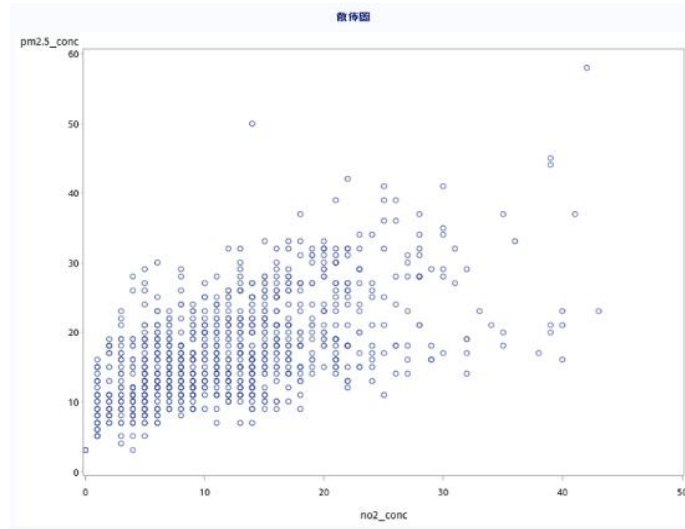


圖 3、Y 與 X3 的散布圖

(四)Y、X4 散布圖

下方散佈圖展示了 PM_{2.5}與 NO_x 的關係。雖然在低濃度區間 (NO_x < 40) 兩者呈現顯著正相關，但在高濃度區間 (NO_x > 60)，數據呈現極度發散且平緩的趨勢。值得注意的是，圖中右側出現數筆 NO_x 高達 80 甚至 100 以上的極端值，但其對應的 PM_{2.5}濃度並未隨之暴增 (維持在 20 左右)。這可能暗示該些極端值來自於距離排放源極近的「新鮮排放 (Fresh Emission)」，此時一次汙染物 (NO_x) 尚未有足夠時間透過光化學反應轉化為二次氣膠 (PM_{2.5})。

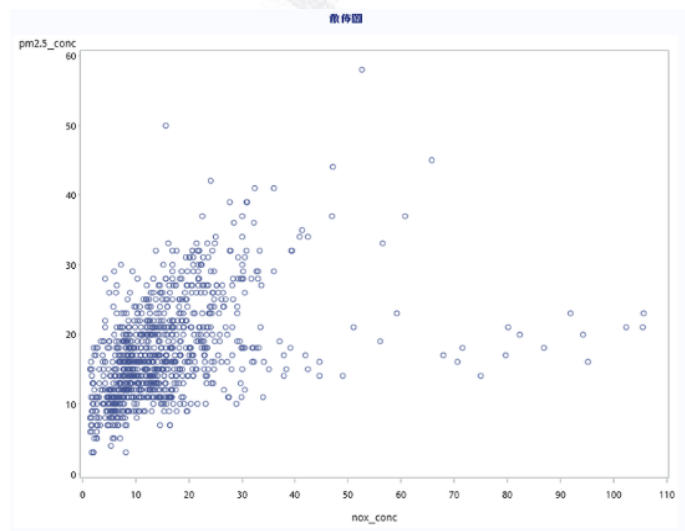


圖 4、Y 與 X4 的散布圖

(五)Y、X5 散布圖

數據點緊密地沿著對角線分佈，顯示兩者具有極高的共變性（Co-variation）。這是由於 PM_{2.5} 本質上是 PM₁₀ 的子集（Subset），兩者在物理傳輸與擴散機制上具有高度同質性。散佈圖的斜率大致反映了 PM_{2.5}/PM₁₀ 的比值。大部分數據集中在斜率較高的區域，顯示該監測點的污染組成以細懸浮微粒為主。值得注意的是，圖形右側（PM₁₀> 80）出現部分偏離主群體的數據點，其 PM_{2.5} 濃度相對較低（約 20-25）。這類「高 PM₁₀、低 PM_{2.5}」的樣本，通常對應於揚塵事件（如營建工程或沙塵暴），此時空氣中主要充斥著大顆粒徑的粗塵，而非燃燒產生的細微粒。

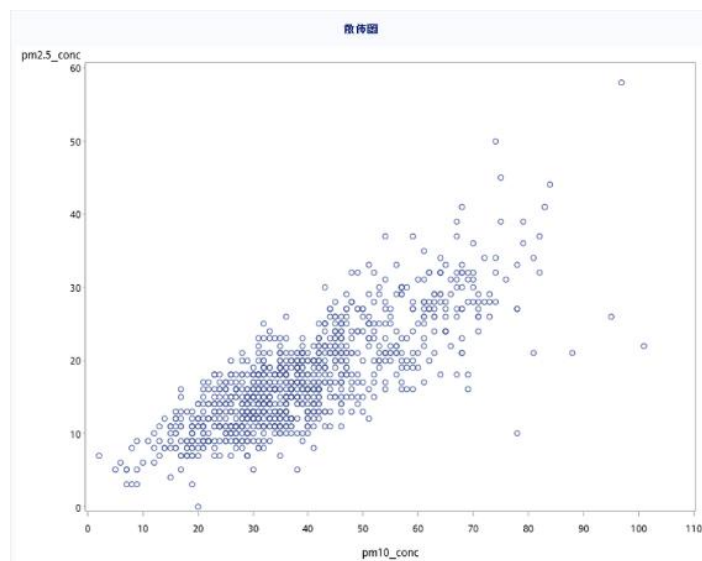


圖 5、Y 與 X5 的散佈圖

(六) Y、X6 散佈圖

當 X6 (W_Sin) 數值較大 (>2.0, 代表該風向分量風速強勁) 時, PM_{2.5} 濃度呈現顯著下降趨勢, 且變異程度極小 (大多低於 20)。這顯示強風提供了良好的大氣擴散條件, 能有效移除或稀釋累積的污染物。PM_{2.5} 的高濃度峰值 (>40) 主要集中在 W_Sin 介於 -2 至 +1 之間的區間。此範圍通常對應於微弱風速或滯留天氣型態, 不利於污染物擴散, 導致本地排放源累積。整體圖形呈現類似三角形或鐘形分佈, 說明風速是控制污染「上限」的關鍵因子—強風下難有高污染, 而弱風則是高污染發生的必要條件 (但非充分條件), 但因為圖形問題不適合做線性迴歸, 故此變數決定刪除不做使用。

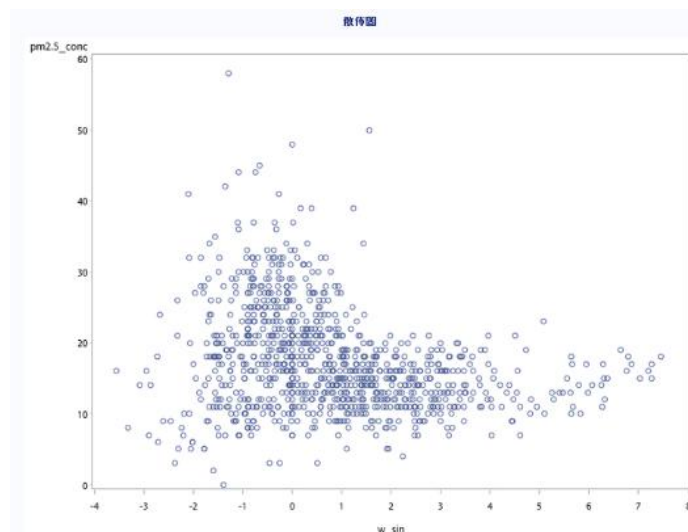


圖 6、Y 與 X6 的散佈圖

(七)Y、X7 散佈圖

數據並未呈現單純的線性關係。隨著 X7 (w_{Cos})數值的變化， $PM_{2.5_CONC}$ 的變異數 (Variance) 呈現明顯的改變。具體而言，數據呈現出一種收斂的三角狀或類似「漏斗形」的分布。絕大多數的觀測值集中在 X 軸數值-1 至+4 之間。在此區間內， $PM_{2.5}$ 濃度的分布跨度最大 (從 0 到近 60)，顯示在 w_{Cos} 數值較低或接近 0 時， $PM_{2.5}$ 濃度具有高度的不確定性與波動性。由於圖形問題也不適合做線性迴歸，故此變數決定刪除不做使用。

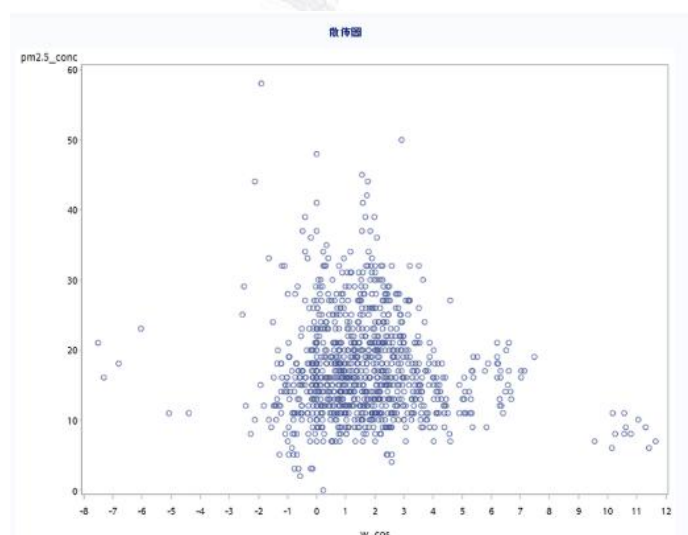


圖 7、Y 與 X7 的散佈圖

四、簡單線性回歸

(一)Y、X1 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F=328.13, P<0.0001$)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R^2 為 0.2563，調整後判定係數 (Adjusted R^2) 為 0.2555。此結果顯示，本迴歸模型中的自變項 (SO₂濃度) 約可解釋依變項 (PM_{2.5}濃度) 25.63% 的變異量。剩餘均方根誤差 (Root MSE) 為 6.06947，反映了模型預測值與觀測值之間的平均離散程度。參數估計結果顯示，SO₂濃度對 PM_{2.5}濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 10.530 (T = 24.69, $P<0.0001$)；SO₂濃度之未標準化迴歸係數為 4.442 (T= 18.11, $P<0.0001$)。



圖 8、Y 與 X1 的簡單線性迴歸

(二)Y、X2 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F= 334.55, P<0.0001$)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R^2 為 0.2598，調整後判定係數

(Adjusted R²) 為 0.2591。此結果顯示，本迴歸模型中的自變項 (CO 濃度) 約可解釋依變項 (PM_{2.5}濃度) 25.98% 的變異量。剩餘均方根誤差 (Root MSE) 為 6.05099，反映了模型預測值與觀測值之間的平均離散程度。參數估計結果顯示，CO 濃度對 PM_{2.5}濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 9.502 (T= 20.07, P <0.0001)；CO 濃度之未標準化迴歸係數為 24.120 (T= 18.29, P <0.0001)。

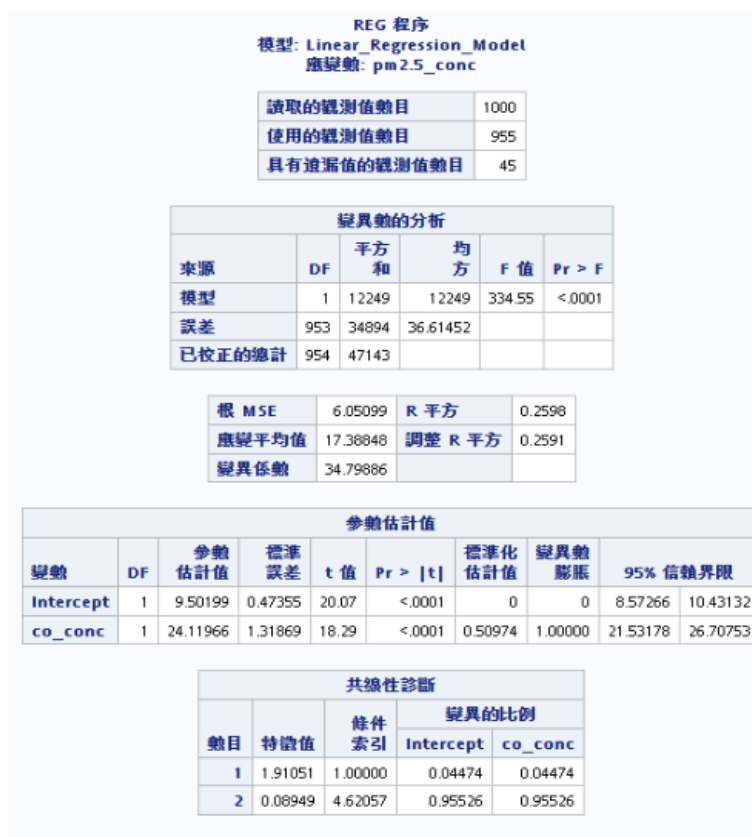


圖 9、Y 與 X2 的簡單線性迴歸

(三)Y、X3 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 (F = 502.92, P <0.0001)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R² 為 0.3464，調整後判定係數 (Adjusted R²) 為 0.3457。此結果顯示，本迴歸模型中的自變項 (NO₂濃度) 約可解釋依變項 (PM_{2.5}濃度) 34.64% 的變異量。剩餘均方根誤差 (Root MSE) 為 5.69296，反映了模型預測值與觀測值之間的平均離散程度。參數估計結果顯示，NO₂濃度對 PM_{2.5}濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 10.857 (T = 31.54, P <0.0001)；NO₂濃度之未標準化迴歸係數為 0.538 (T = 22.43, P <0.0001)。



圖 10、Y 與 X3 的簡單線性迴歸

(四)Y、X4 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F = 1700.31, P < 0.0001$)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R^2 為 0.6401，調整後判定係數 (Adjusted R^2) 為 0.6397。此結果顯示，本迴歸模型中的自變項 (PM₁₀濃度) 約可解釋依變項 (PM_{2.5}濃度) 64.01% 的變異量，為所有單變項模型中解釋力最高者。剩餘均方根誤差 (Root MSE) 為 4.19962，反映了模型預測值與觀測值之間的平均離散程度。參數估計結果顯示，PM₁₀濃度對 PM_{2.5}濃度具有極為顯著的正向影響。截距項 (Intercept) 之估計值為 3.014 ($T = 8.07, P < 0.0001$)；PM₁₀ 濃度之未標準化迴歸係數為 0.370 ($T = 41.23, P < 0.0001$)。



圖 11、Y 與 X4 的簡單線性迴歸

(五)Y、X5 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型之 F 值為 3.85，顯著性機率值 (P 值) 為 0.0501。此結果顯示該模型在 $\alpha = 0.05$ 的顯著水準下剛好未達顯著 (略高於 0.05)，屬於邊緣顯著 (Marginally Significant)，顯示預測變項對依變項變異的解釋能力相當有限。在模型解釋力方面，判定係數 R^2 僅為 0.0042，調整後判定係數 (Adjusted R^2) 為 0.0031。此結果顯示，本迴歸模型中的自變項 (NOx 濃度) 僅能解釋依變項 (PM_{2.5}濃度) 約 0.42% 的變異量，解釋力極低。剩餘均方根誤差 (Root MSE) 為 7.23003。參數估計結果顯示，截距項 (Intercept) 之估計值為 16.848 (T = 44.19, P <0.0001)；NOx 濃度之未標準化迴歸係數為 0.035 (T = 1.96, P = 0.0501)，同樣顯示其統計顯著性處於臨界邊緣。



圖 12、Y 與 X5 的簡單線性迴歸

五、共線性診斷

此處為剔除共線性自變數之分析，以 VIF 數值解讀各變數間的共線性關係。在此模型中，各個自變數的 VIF 值如下：

- SO₂_CONC: 1.416
- CO_CONC: 3.157
- NO₂_CONC: 3.845
- NO_x_CONC: 1.319
- PM₁₀_CONC: 1.728

從 VIF 的角度來看，這個模型沒有嚴重的多重共線性問題。所有的 VIF 值均低於 5（最高為 NO₂_CONC 的 3.845），這表示各個自變數之間的線性相關程度在合理範圍內，不會對迴歸係數的估計造成嚴重的標準誤膨脹。基於 VIF 檢驗，目前模型中的所有變數（SO₂, CO, NO₂, NO_x, PM₁₀）都適合保留在模型中，不需要因為共線性原因進行剔除。

參數估計值							
變數	DF	參數估計值	標準誤差	t 值	Pr > t	變異數膨脹	95% 信賴界限
Intercept	1	2.34221	0.44485	5.27	<.0001	0	1.46911 3.21532
so2_conc	1	0.72684	0.19369	3.75	0.0002	1.41649	0.34669 1.10700
co_conc	1	3.71959	1.55688	2.39	0.0171	3.15698	0.66393 6.77526
no2_conc	1	0.18136	0.03340	5.43	<.0001	3.84451	0.11580 0.24691
nox_conc	1	-0.05223	0.01144	-4.57	<.0001	1.31863	-0.07467 -0.02978
pm10_conc	1	0.29299	0.01146	25.57	<.0001	1.72752	0.27050 0.31548

圖 13、共線性診斷 VIF 值

六、多元迴歸分析

根據變異數分析 (ANOVA) 結果顯示，本研究所建構之多元線性迴歸模型整體達高度統計顯著水準 ($F = 401.22$, $P < .0001$)，顯示模型具備良好的整體解釋能力。在模型解釋力方面，其決定係數 (R^2) 為 0.6970，調整後決定係數為 0.6953，表示所納入之解釋變數，包括 SO_2 、CO、 NO_2 、 NO_x 及 PM_{10} ，可共同解釋約 69.5% 的 $PM_{2.5}$ 濃度變異量，顯示模型具有相當理想之解釋力。此外，模型之根均方誤差 (Root MSE) 為 3.919，變異係數 (CV) 為 22.48，顯示模型預測誤差維持在合理範圍內，整體估計結果具備穩定性與實用性。

而就迴歸參數估計結果而言，多數汙染物變數對 $PM_{2.5}$ 濃度呈現顯著影響。迴歸模型顯示，在控制其他變數不變的情況下，一氧化碳 (CO) 之迴歸係數最大 ($\beta = 3.720$)，顯示 CO 濃度每增加一單位， $PM_{2.5}$ 濃度將顯著上升約 3.72 單位；同時， PM_{10} 為模型中最具顯著性之解釋變數，其 T 值高達 25.57，顯示 PM_{10} 與 $PM_{2.5}$ 之間存在極為強烈之正向關聯。此外，二氧化硫 (SO_2) 與二氧化氮 (NO_2) 之迴歸係數亦皆為正且達統計顯著水準，顯示其濃度上升將伴隨 $PM_{2.5}$ 濃度之增加。相對而言，氮氧化物 (NO_x) 則呈現顯著負向影響 ($\beta = -0.052$, $P < .0001$)，其影響方向與 NO_2 相反，可能反映大氣化學反應過程中不同氮氧化物型態之轉化或在特定環境條件下所產生之抵銷效應，顯示汙染物之交互作用關係仍有進一步深入探討之必要。

REG 程序					
模型: Linear_Regression_Model					
應變數: pm2.5_conc					
讀取的觀測值數目	1000				
使用的觀測值數目	878				
具有遺漏值的觀測值數目	122				
變異數的分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	5	30812	6162.37423	401.22	<.0001
誤差	872	13393	15.35923		
已校正的總計	877	44205			
根 MSE	3.91909	R 平方	0.6970		
應變平均值	17.42938	調整 R 平方	0.6953		
變異係數	22.48551				
參數估計值					
變數	DF	參數估計值	標準誤差	t 值	Pr > t
Intercept	1	2.34221	0.44485	5.27	<.0001
so2_conc	1	0.72684	0.19369	3.75	0.0002
co_conc	1	3.71959	1.55688	2.39	0.0171
no2_conc	1	0.18136	0.03340	5.43	<.0001
nox_conc	1	-0.05223	0.01144	-4.57	<.0001
pm10_conc	1	0.29299	0.01146	25.57	<.0001

圖 14、多元回歸分析

七、殘差分析

(一)PM_{2.5}_CONC 的殘差的直方圖

由殘差分布圖可知，該模型在 PM_{2.5}濃度預測上呈現穩定且具良好表現之特性。首先，就殘差的中心位置而言，其分布均值接近於 0，顯示模型整體而言不存在明顯的系統性偏誤，未出現長期高估或低估 PM_{2.5}濃度之情形。其次，殘差分布在左右兩側呈現大致對稱的型態，顯示模型在預測過程中，預測偏高與偏低的發生機率及幅度相對均衡，符合隨機誤差之假設。

在分布寬度方面，多數殘差集中於-7.5 至+7.5 的區間內，顯示約九成以上的預測誤差落在此範圍，反映模型在誤差控制上具一定穩定性與可接受性。此外，從峰度與尾部特性觀察，殘差分布在中心區域呈現較常態分布更高且尖銳的峰值，顯示誤差值高度集中於 0 附近；然而，在分布兩端仍可觀察到少量極端值，顯示殘差具有輕微重尾現象，意味在少數情況下仍可能產生較大幅度之預測誤差。整體而言，殘差分布結果顯示該模型在 PM_{2.5}濃度預測上具備良好的穩定性與無偏性，並符合多數迴歸模型對誤差分布之基本假設。

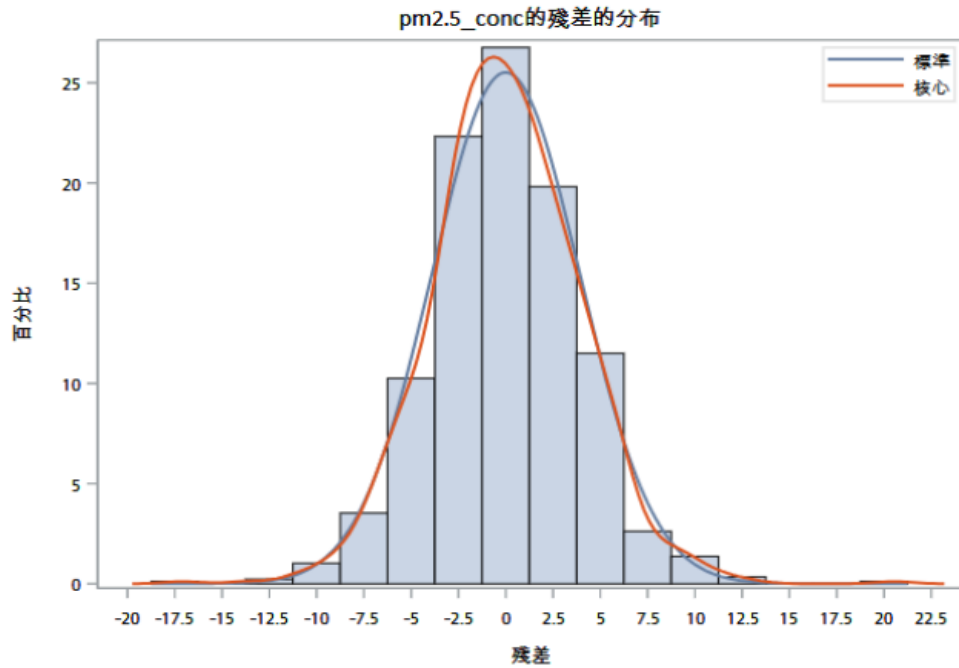


圖 15、殘差直方圖

(二)PM_{2.5}_CONC 的殘差的 Q-Q 圖

為檢驗多元迴歸模型殘差之常態性假設，本研究進一步透過 Q-Q 圖比較實際殘差分布與理論常態分布之吻合程度。由圖中可觀察，在分位數介於-2 至 2 之間時，多數殘差點幾乎緊密分布於 45 度參考線附近，顯示殘差於中心區域之分布型態與常態分布高度一致。然而，在分布兩端之尾部區域，殘差點於低值與高值端明顯偏離參考線，並呈現向外彎曲之趨勢。此現象顯示殘差分布存在重尾 (Heavy Tails) 特性，意即模型中仍存在少數極端預測誤差，其出現頻率高於理想常態分布之預期。整體而言，雖然殘差於中心區域大致符合常態性假設，但尾部偏離現象顯示模型仍可能受到離群值或極端觀測值之影響，值得於後續分析中進一步檢視與修正。

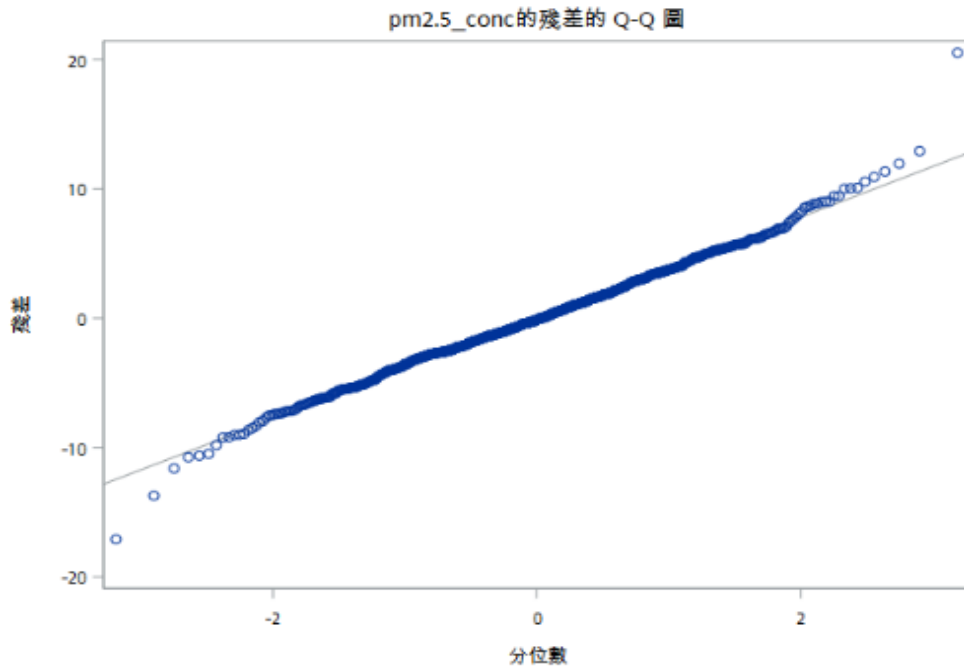


圖 16、PM_{2.5}_CONC 的殘差的 Q-Q 圖

(三)常態性檢定 X1~X5

雖然部分常態性檢定（尤其是 Shapiro–Wilk 檢定）之 P 值小於 0.05，在嚴格統計意義下顯示殘差未完全符合常態分布假設，但結合 Q-Q 圖之視覺化結果可知，殘差於中心區域仍與常態分布高度一致，顯示模型在實務應用上仍具備可接受之穩定性。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.991769	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.030304	Pr > D	0.0485
Cramer-von Mises	W-Sq	0.123149	Pr > W-Sq	0.0563
Anderson-Darling	A-Sq	0.800113	Pr > A-Sq	0.0399

圖 17、常態性檢定

(四)均質性 X1~X5

為檢視多元迴歸模型是否符合變異數同質性假設，本研究透過殘差散佈圖作為視覺化診斷工具，將學生化殘差（RStudent）與模型預測值進行比較分析。由圖中可觀察，當預測值介於約 10 至 20 之間時，殘差點分布相對集中，且大多落於 ±2 的合理範圍內，顯示在較低預測區間中，模型誤差之變異程度相對穩定。然而，隨著預測值提升至 25 以上，殘差點分布逐漸趨於分散，並開始出現

學生化殘差超出 ± 2 ，甚至接近 ± 4 之觀測值，顯示誤差變異幅度隨預測值增加而擴大。此一分布型態暗示模型可能存在異質變異數(heteroscedasticity)之情形，表示殘差變異並未在各預測水準下維持一致。

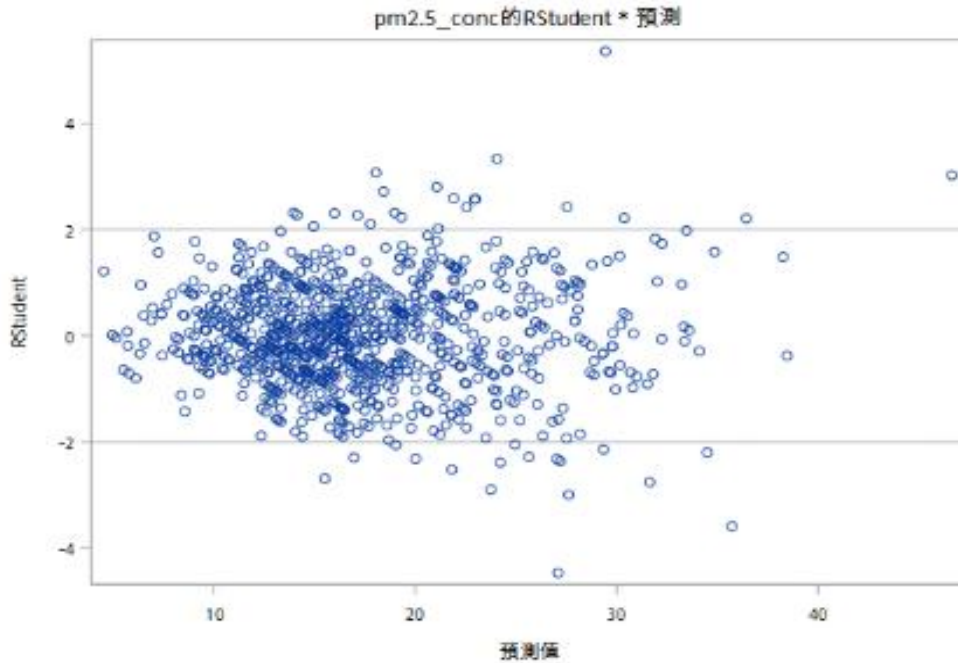


圖 18、學生化殘差散佈圖



圖 19、變異數均質性檢定

(五)獨立性檢定 X1~X5

為檢驗多元迴歸模型殘差是否符合獨立性假設，本研究同時透過殘差散佈圖與 Durbin-Watson (DW) 檢定進行診斷。由殘差散佈圖可觀察，數據點於 0 軸上下呈現隨機分布，未出現明顯之曲線型態、週期性波動或規律性趨勢（如鋸齒狀或緩慢遞增現象），顯示殘差之變動不具系統性依賴，視覺化結果支持誤差項彼此獨立之假設。進一步就 Durbin-Watson 檢定結果而言，其統計量為 1.585，雖略低於理想值 2，代表殘差間可能存在極輕微之正相關，但該數值仍落於可接受區間內，顯示模型殘差整體上仍符合獨立性假設。此外，一階自相關係數為

0.207，顯示觀測值之間僅存在微弱的連續性，此程度之相關性在統計實務上通常不足以破壞迴歸模型之有效性與推論可靠度。

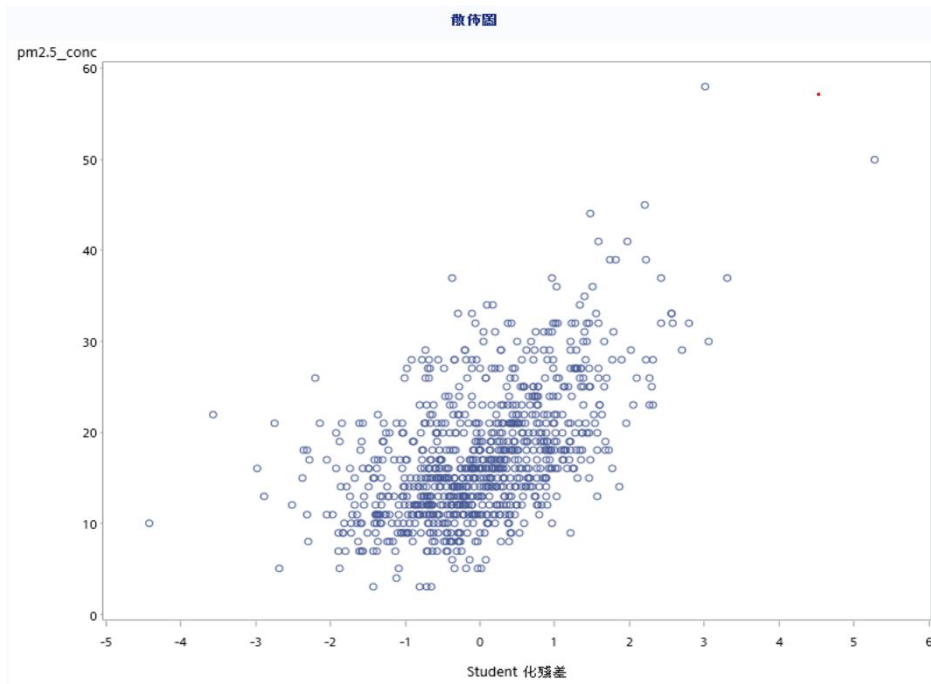


圖 20、學生化殘差與反應變數 Y 的散佈圖

Durbin-Watson D	1.585
觀測值數目	878
一階自相關	0.207

圖 21、Durbin-Watson (DW) 檢定

八、離群值及影響點的診斷

(一)學生化殘差散布圖

本圖展示了針對依變數「PM_{2.5}_CONC」（PM_{2.5}濃度）進行迴歸分析後的殘差診斷結果，具體呈現為學生化殘差（RStudent）與預測值（Predicted Value）之間的散佈圖。此圖表的主要目的在於檢驗迴歸模型是否符合普通最小平方法（OLS）的基本統計假設，特別是關於線性關係、變異數同質性(Homoscedasticity)以及極端值（Outliers）的偵測。

首先，就線性假設與變異數同質性而言，觀察圖中數據點的分佈情形，可以發現殘差點隨機散佈在以零為中心的水平軸上下，並未呈現明顯的拋物線或曲線型態，這暗示了模型設定的線性關係大致成立，解釋變數與應變數之間的關係

已受到模型的合理捕捉，並無明顯的系統性偏差。此外，數據點在水平方向上的分佈寬度雖然在預測值的中段略為密集，但整體而言，殘差的垂直散佈範圍並未隨著預測值的增加而呈現出明顯的擴大或縮小趨勢（即未出現顯著的漏斗狀分佈），這表明誤差項的變異數大致保持恆定，符合變異數同質性的假設，意即模型對於不同濃度範圍的預測穩定性相當。

然而，在極端值的檢測方面，圖表揭示了值得注意的現象。一般而言，學生化殘差值若落在 ± 2 之外即視為異常，若超過 ± 3 則視為顯著的極端值。細看此圖，雖然絕大多數的觀測值均落在 ± 2 的區間內，顯示模型對大部分數據具有良好的解釋力，但仍有部分數據點的殘差值顯著偏離零軸，甚至延伸至 ± 4 乃至接近5的位置。這些遠離中心群體的離群值，暗示了樣本中存在特定的觀測對象，其PM_{2.5}濃度的實際值與模型的預測值存在巨大落差。

綜上所述，該迴歸模型在結構上並未違反線性和變異數同質性的核心假設，模型設定尚屬得宜。然而，鑑於圖中存在數個顯著的離群值，後續研究建議應針對這些高殘差的特定樣本進行深入檢視，確認其數據是否由測量誤差、特殊環境事件或其他未納入模型的變數所導致，並評估是否需要透過強韌迴歸（Robust Regression）或剔除異常值的方式來優化模型的預測精確度。

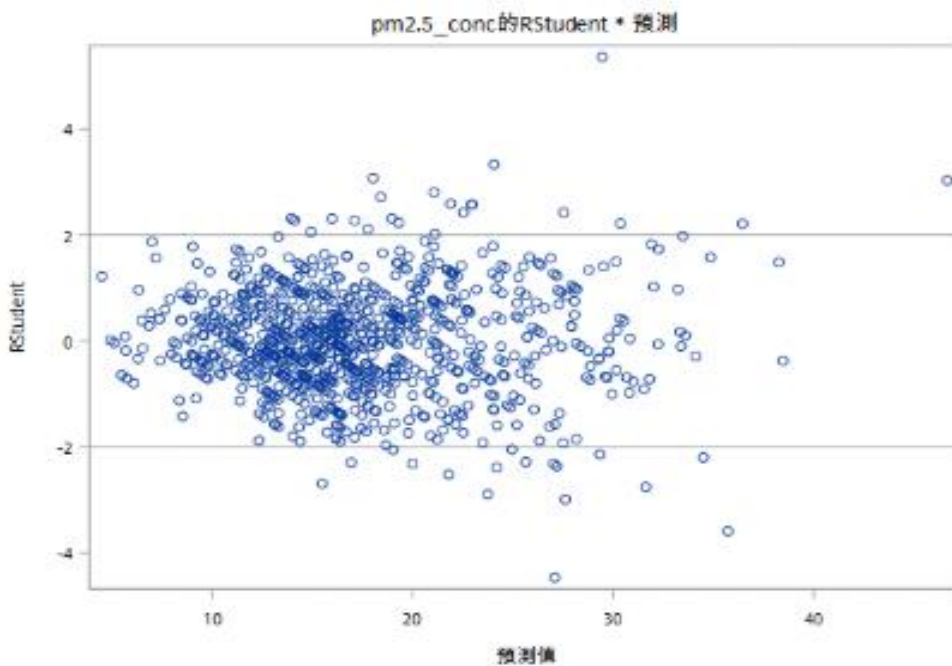


圖 22、學生化殘差與預測值的散佈圖

(二)Cook's Distance 長條圖

本研究採用庫克距離 (Cook's Distance, Cook's D) 作為診斷統計量，旨在評估個別觀測值對 PM_{2.5} 濃度 (PM_{2.5}_CONC) 迴歸模型參數估計的影響程度。庫克距離綜合考量了殘差大小與槓桿值 (Leverage)，藉以識別出若移除該觀測值後，整體迴歸係數將產生顯著變動的強影響點 (Influential Points)。

根據模型診斷結果圖示，橫軸代表觀測值序列 (Observation Index)，縱軸則為庫克距離數值。整體而言，絕大多數觀測值的庫克距離數值極低，緊貼於水平軸線附近，且低於圖中設定的基準閾值線，這顯示樣本中大部分數據點對於迴歸模型的建構具有一致性，並未對模型參數造成過度且不成比例的拉扯與扭曲，暗示模型在整體數據結構上的適配度尚稱穩定。

然而，圖中明顯呈現出若干離群的突波 (Spikes)，顯示特定觀測值具有異常高的庫克距離，其影響力遠超其餘樣本。具體檢視圖譜，可發現位於觀測值序列第 50 筆附近的數據點呈現出最高的峰值，其 Cook's D 值逼近 0.150，為該資料集中最具影響力的單一觀測值。此外，在觀測值序列約第 900 筆附近亦出現另一顯著峰值，數值超過 0.125；其餘如序列第 200、400、420 以及 870 筆附近的觀測值，亦呈現出顯著高於背景雜訊的突起，數值介於 0.050 至 0.080 之間。這些高庫克距離的觀測點意味著它們在決定迴歸平面的斜率與截距時扮演了關鍵角色，若將其移除，模型預測結果可能會發生較大幅度的改變。

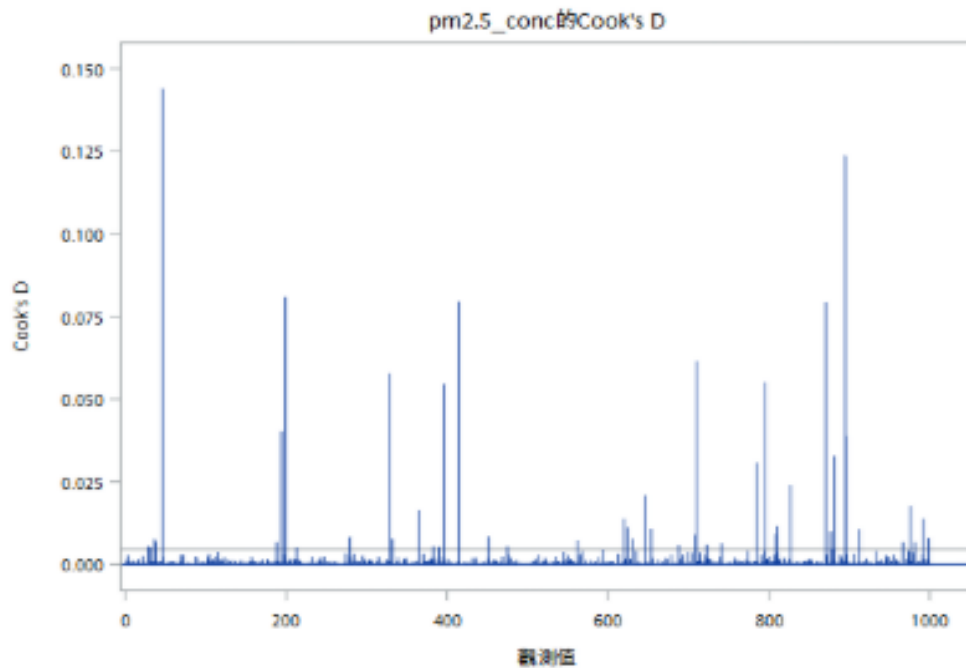


圖 23、Cook's Distance 長條圖

(三)離群值與槓桿作用診斷

下圖展示了針對變數 $PM_{2.5_CONC}$ 進行的離群值與槓桿作用診斷分析結果，旨在檢視迴歸模型中數據點的適配情形與潛在影響力。圖中橫軸代表槓桿作用 (Leverage)，用以衡量自變數數值的極端程度；縱軸則為學生化殘差 (RStudent)，用於評估因變數預測值的偏差程度。分析中設定了特定的統計閾值，包括殘差絕對值大於 2 的水平界線以及槓桿值的垂直界線，藉此將數據點劃分為不同的診斷類別以利判讀。

觀察整體數據分佈情形，絕大多數的觀測值均密集集中於左側中心區域(即藍色圓點群)，顯示大部分數據點既無過高的槓桿值，其殘差亦控制在合理範圍內，暗示模型對這些主要樣本的解釋能力尚佳且結構穩定。然而，圖中亦揭示了三類須特別關注的異常觀測值。首先，紅色空心圓點代表單純的「離群值」，這些樣本雖然在自變數空間中並不獨特，但其預測殘差顯著超出 ± 2 的標準範圍，顯示模型未能準確捕捉這些點的變異。其次，綠色圓點代表高「槓桿作用」點，這類樣本具有較極端的自變數數值，但由於其殘差極小，意味著它們雖然數值特殊，但仍大致符合整體的迴歸趨勢線，尚未構成嚴重的預測偏差。

分析中最值得注意的為圖中散佈於右側上下兩端的褐色圓點，即同時具備「離群值和槓桿作用」特性的觀測值。這類數據點位於圖形的右側且遠離水平零軸，意味著它們不僅在自變數特徵上極為極端，且嚴重偏離預測模型。在統計推論上，這類觀測值通常被視為具高度影響力的點 (Influential Points)，極可能對迴歸係數的估計產生不成比例的拉扯與偏差，進而影響模型的整體效度。因此，在後續的分析階段，必須優先針對這些褐色標記的樣本進行深入審查，確認其是否為數據登錄錯誤或真實發生的極端環境事件；若確認為真實數據，則建議考慮採用穩健迴歸 (Robust Regression) 方法或進行敏感度分析，以確保模型分析結果的穩健性與準確度。

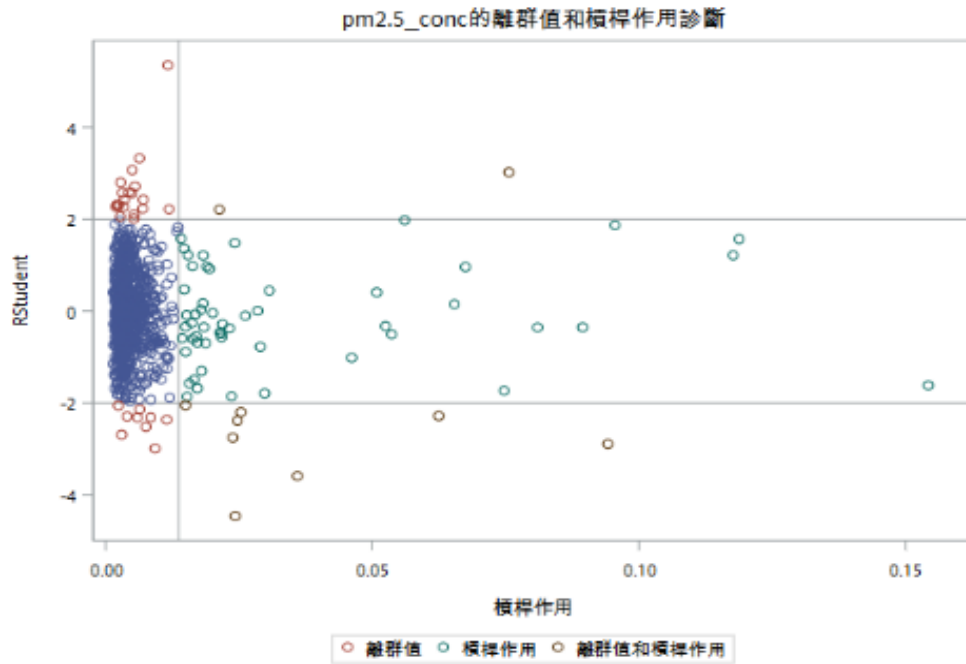


圖 24、離群值與槓桿作用診斷

(四)排除掉離群值、影響點

針對殘差分佈的檢測，該分析採用了學生化殘差 (Studentized Residuals, 介面中變數標記為 $R_{student_PM_{2.5}_CONC}$) 作為核心篩選指標。有別於一般原始殘差，學生化殘差將每個殘差除以其標準差的估計值，從而消除了不同觀測值變異數不均的影響，使其在統計上具備標準化的可比較性。介面中設定的篩選條件為保留數值「大於 -2 且小於 2」的觀測值。此閾值的設定具有堅實的統計學理基礎，依據常態分佈的經驗法則，約有 95% 的觀測數據應落在平均值正負兩個標準差的範圍內。凡是學生化殘差絕對值超過 2 的樣本，通常意味著該點的實際觀測值與模型預測值存在顯著偏差。在空氣品質研究的情境下，移除這些殘差過大的樣本，能有效排除因偶發性因素導致的極端讀數，確保模型擬合的是整體環境變數的普遍規律，而非特例雜訊。

其次，除了單純檢視殘差大小外，該分析進一步引入了庫克距離 (Cook's Distance, 介面中變數標記為 $Cookd_PM_{2.5}_CONC$) 作為衡量觀測值影響力的關鍵指標。庫克距離綜合考量了殘差大小與槓桿值 (Leverage)，用以量化單一觀測值的移除對整體迴歸係數估計的變動程度。圖中顯示篩選條件設定為庫克距離必須「小於 $0.006((自變數+應變數)/1000)$ 」。這一極低的閾值顯示出研究者採取了相當嚴格的品質控制標準。一般而言，庫克距離越大，代表該樣本點對迴歸線的斜率與截距具有強大的拉扯作用。透過限制庫克距離在極低的水準，研究者能確保最終的模型參數不會被少數幾個具有高槓桿效應的特殊樣本所主導 (例如特

定氣象條件下的極端高污染日），從而提升模型在解釋一般性空氣品質變化時的穩定度與代表性。

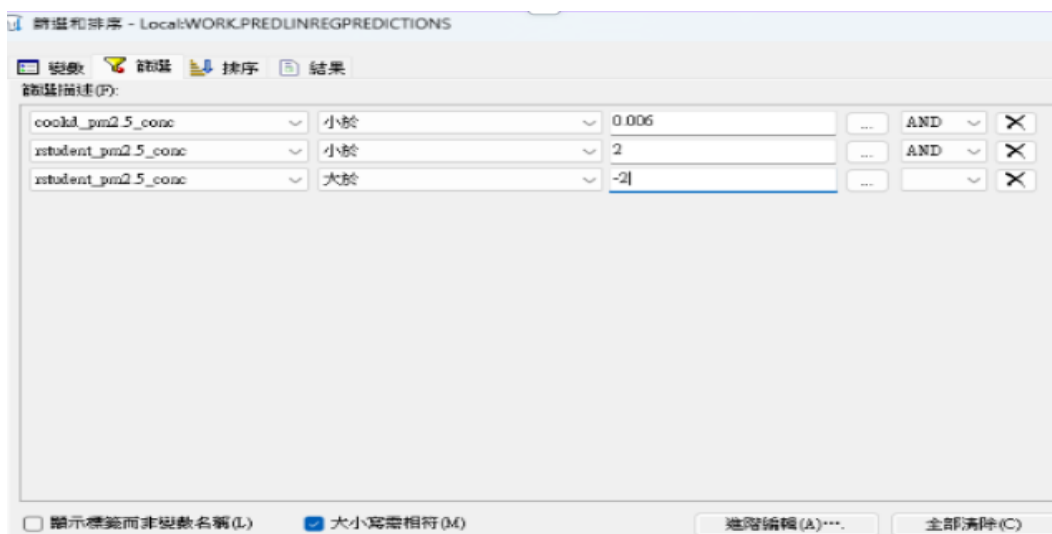


圖 25、篩選與排序

(五)刪除離群值與影響點

在環境科學與空氣品質監測的統計建模中，數據的純淨度與模型的穩健性是確保普通最小平方法（OLS）估計有效性的關鍵前提。由於 PM_{2.5}濃度數據本質上具有高度波動性，且易受極端氣象條件或局部污染源干擾，原始數據集中常潛藏著偏離常態分佈的異常值（Outliers）以及對迴歸係數具有不成比例影響力的觀測值（Influential Points）。為了解決此一問題，本研究針對 PM_{2.5}線性迴歸預測模型採取了一套結合巨觀圖形診斷與微觀數據篩選的雙重過濾機制，並依據實際數據分佈特徵，將庫克距離（Cook's Distance）的篩選閾值精確設定為 0.006，旨在於保留有效樣本與剔除極端影響之間取得最佳平衡。

首先，針對模型殘差結構的診斷，分析採用了學生化殘差（Studentized Residuals）作為衡量單一觀測值偏離程度的標準化指標。由殘差對預測值的散佈圖可見，數據點主要分佈於零軸兩側，顯示出隨機分佈的特性，然而仍有部分樣本點落在±2 個標準差的邊界之外。根據統計學理，約 95% 的常態分佈數據應落在平均值正負兩個標準差範圍內，超出此範圍的觀測值往往代表模型無法解釋的極端誤差。因此，本研究設定了明確的篩選規則，僅保留學生化殘差介於 -2 與 2 之間的樣本。此一操作有效地移除了縱軸方向上的離群值，修正了因儀器誤差或突發性環境事件所導致的極端讀數，確保了殘差分佈滿足常態性的假設要求。

除了單純的誤差檢定，本研究更進一步引入庫克距離以識別高影響點，防止模型參數被少數具有高槓桿值（Leverage）的樣本所主導。依據庫克距離的長

條圖 (Bar Chart) 診斷結果顯示，雖然絕大多數樣本的影響力數值極低，但仍有少數突出的峰值超過了 0.006 的水準，顯示這些特定觀測值對迴歸線的斜率具有強大的拉扯作用。為此，篩選程序將庫克距離的閾值設定為「小於 0.006」。此閾值的選擇具有高度的實證意義：它精確地切除了圖表中那些顯著突出的極端峰值，同時保留了絕大部分的常態樣本。進一步檢視詳細的數據列表 (Data Table)，可以發現編號 93、528 與 567 的觀測值，其庫克距離分別約為 0.0005 至 0.0006 之間。在調整後的 0.006 閾值標準下，這些樣本點被判定為遠低於危險邊界的安全觀測值，因而被正確地保留在模型中。這證實了該篩選策略的精準度，既能有效移除那些真正具破壞力的高影響點 (即圖表中高於 0.006 的部分)，又避免了因過度篩選而誤刪僅具輕微影響力的正常數據 (如數據表中所示之樣本)。

綜上所述，透過整合殘差散佈圖的視覺化診斷與數據列表的數值驗證，本研究確立了一套嚴謹且具針對性的數據清理程序。利用學生化殘差限制在 ± 2 區間內以剔除異常誤差，並將庫克距離嚴格控制在 0.006 以下以排除不當影響點，有效地提升了資料集的同質性與代表性。這種「去蕪存菁」的過程，顯著降低了模型估計的變異數並修正了潛在偏差。經過此程序優化後的線性迴歸模型，其預測結果將更能反映 PM_{2.5} 濃度與環境變數間的一般性規律，為後續的空氣品質管理與決策提供更具科學信度與穩定性的實證依據。

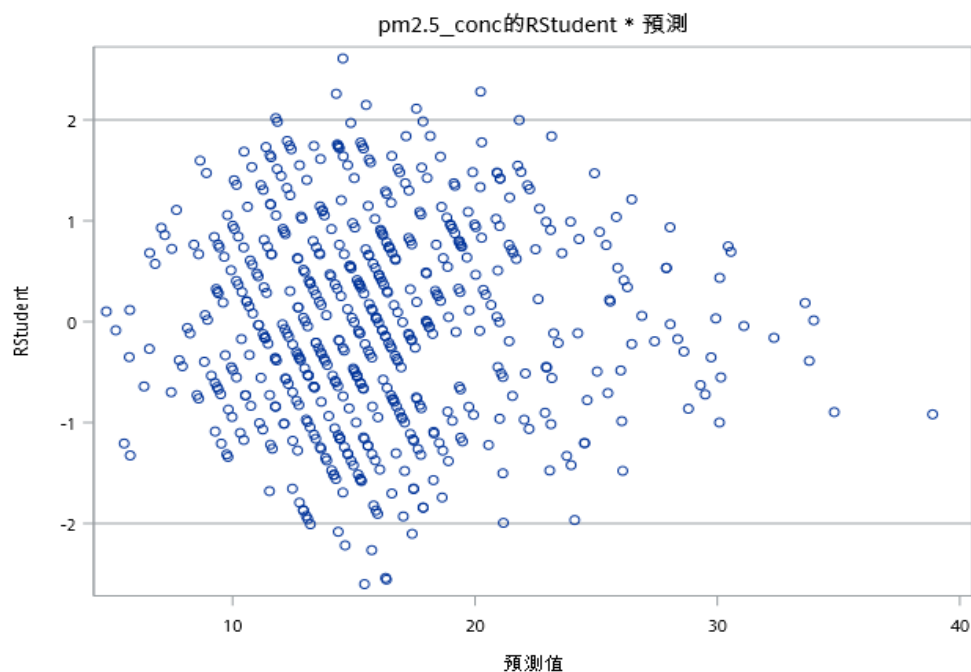


圖 26、刪除離群值後學生化殘差與預測值的散佈圖

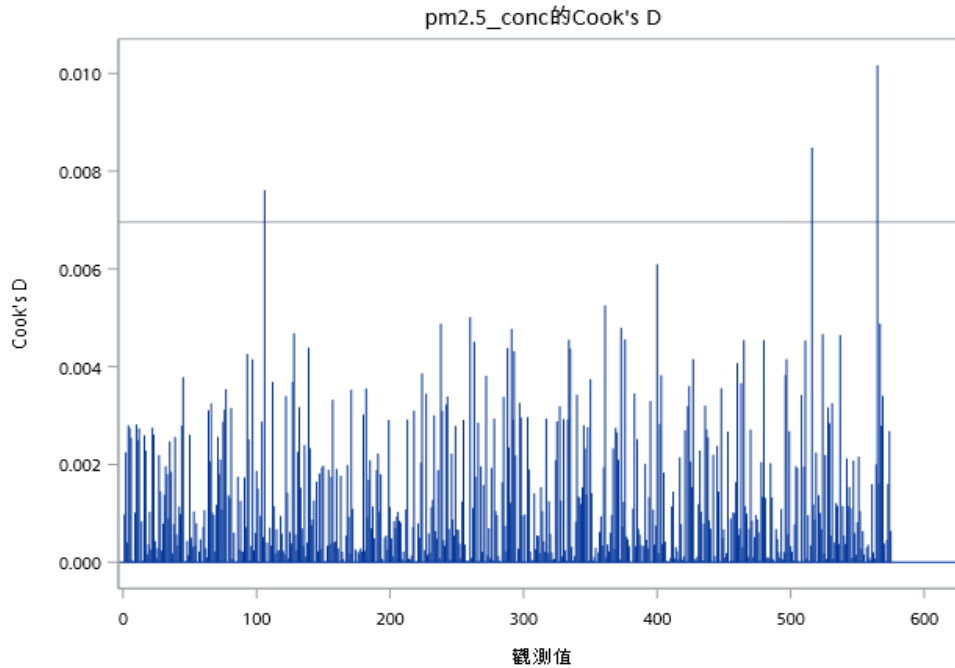


圖 27、刪除離群值後 Cook's Distance 長條圖

九、刪除離群值後之簡單線性迴歸、多元迴歸、殘差分析

(一)刪除後 Y、X1 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F = 276.34, P < 0.0001$)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R^2 為 0.3254，調整後判定係數 (Adjusted R^2) 為 0.3242。此結果顯示，在排除離群值後，本迴歸模型中的自變項 (SO_2 濃度) 約可解釋依變項 ($PM_{2.5}$ 濃度) 32.54% 的變異量。相較於未刪除離群值之模型 (R^2 約 0.2563)，解釋力有顯著提升。剩餘均方根誤差 (Root MSE) 為 4.62411，數值較原始模型下降，反映了模型預測精確度提高，預測值與觀測值之間的平均離散程度降低。

參數估計結果顯示， SO_2 濃度對 $PM_{2.5}$ 濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 8.483 ($T = 17.14, P < 0.0001$)； SO_2 濃度之未標準化迴歸係數為 5.315 ($T = 16.62, P < 0.0001$)。值得注意的是，刪除離群值後，迴歸係數由原本的 4.442 上升至 5.315，顯示在核心數據群中，二氧化硫對 $PM_{2.5}$ 的邊際影響力實際上更為強烈。



圖 28、刪除後 Y、X1 簡單線性迴歸

(二)刪除後 Y、X2 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F = 414.04, P < 0.0001$)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R^2 為 0.4195，調整後判定係數 (Adjusted R^2) 為 0.4185。此結果顯示，在排除離群值後，本迴歸模型中的自變項 (CO 濃度) 約可解釋依變項 (PM_{2.5}濃度) 41.95% 的變異量。相較於未刪除離群值之模型 (R^2 約 0.2598)，解釋力有大幅度的提升 (增加了約 16 個百分點)，這表明離群值嚴重干擾了原始數據的線性關係。剩餘均方根誤差 (Root MSE) 為 4.28944，較原始模型顯著下降，顯示模型對核心數據群的預測準確度大幅提高。

參數估計結果顯示，CO 濃度對 PM_{2.5}濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 5.724 ($T = 10.63, P < 0.0001$)；CO 濃度之未標準化迴歸係數為 34.559 ($T = 20.35, P < 0.0001$)。與原始模型相比 (係數約 24.12)，刪除離群值後的迴歸係數顯著增加，這意味著在一般情況下 (排除極端事件後)，一氧化碳濃度對 PM_{2.5}的邊際貢獻實際上比原始模型評估的更為強烈。

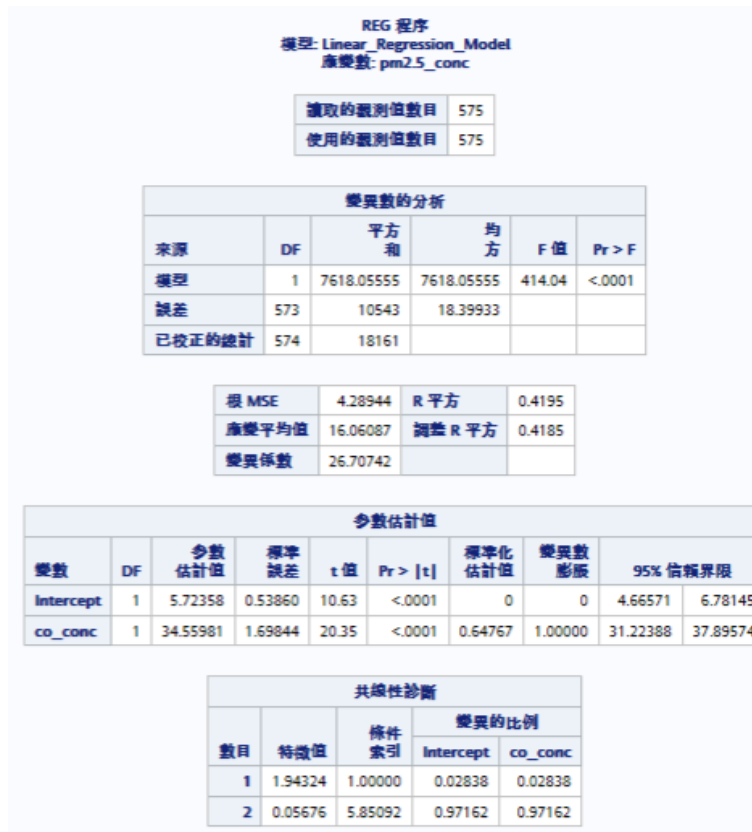


圖 29、刪除後 Y、X2 簡單線性迴歸

(三)刪除後 Y、X3 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F = 542.87, P < 0.0001$)，顯示預測變項能有效解釋依變項之變異。在模型解釋力方面，判定係數 R^2 為 0.4865，調整後判定係數 (Adjusted R^2) 為 0.4856。此結果顯示，在排除離群值後，本迴歸模型中的自變項 (NO₂濃度) 約可解釋依變項 (PM_{2.5}濃度) 48.65% 的變異量。相較於未刪除離群值之模型 (R^2 約 0.3464)，解釋力有著顯著的提升 (增加約 14 個百分點)，且為所有氣態汙染物模型中解釋力最高者。剩餘均方根誤差 (Root MSE) 為 4.03423，顯示模型擬合度極佳。

參數估計結果顯示，NO₂濃度對 PM_{2.5}濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 9.661 ($T = 29.99, P < 0.0001$)；NO₂濃度之未標準化迴歸係數為 0.583 ($T = 23.30, P < 0.0001$)。其標準化迴歸係數高達 0.69750，這進一步證實了在核心數據群中，二氧化氮是影響 PM_{2.5}濃度變化的主導性關鍵因子。



圖 30、刪除後 Y、X3 簡單線性迴歸

(四)刪除後 Y、X4 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F = 26.55, P < 0.0001$)。這一點與未刪除離群值前的模型 ($F = 3.85, P = 0.0501$ ，僅達邊緣顯著) 形成強烈對比，顯示在排除極端值干擾後，氮氧化物濃度與 PM_{2.5}濃度之間確實存在穩定的統計關聯。

在模型解釋力方面，判定係數 R^2 為 0.0443，調整後判定係數 (Adjusted R^2) 為 0.0426。雖然此數值顯示自變項 (NO_x 濃度) 僅能解釋依變項 (PM_{2.5}濃度) 約 4.43% 的變異量，屬於低度解釋力，但相較於原始模型 (R^2 僅 0.0042)，其解釋能力已提升了十倍以上。剩餘均方根誤差 (Root MSE) 為 5.50370，較原始模型 (7.23) 有顯著改善。

參數估計結果顯示，NO_x 濃度對 PM_{2.5}濃度具有顯著的正向影響。截距項 (Intercept) 之估計值為 14.044 ($T = 30.96, P < 0.0001$)；NO_x 濃度之未標準化迴歸係數為 0.135 ($T = 5.15, P < 0.0001$)。此結果修正了原始模型中 NO_x 係數不顯著 ($P = 0.0501$) 的結論，證實了在常態大氣條件下，氮氧化物總量仍是 PM_{2.5}的顯著預測因子，唯其預測效力遠低於單獨的二氧化氮 (NO₂) 指標。

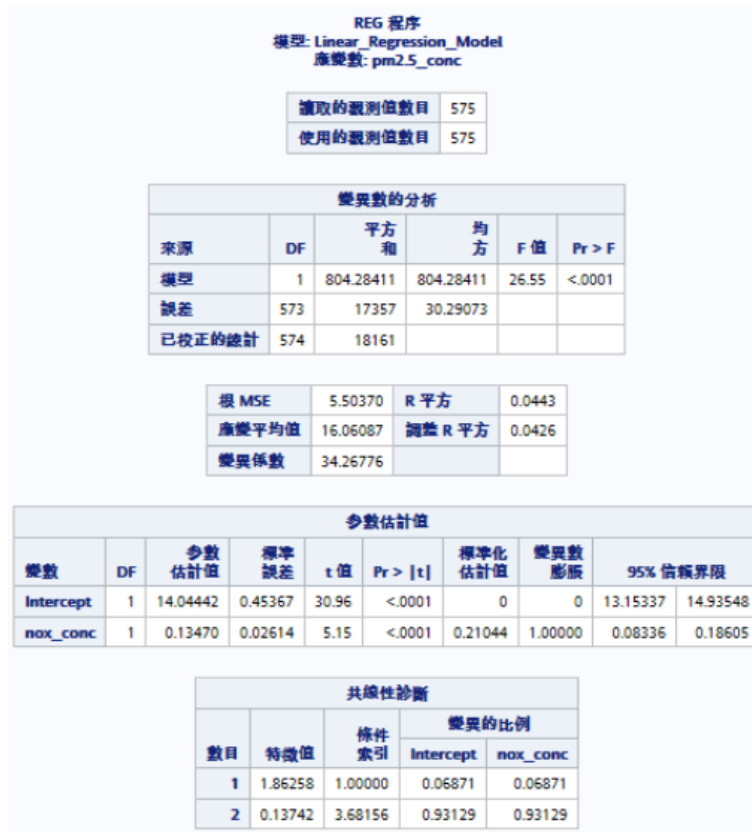


圖 31、刪除後 Y、X4 簡單線性迴歸

(五)刪除後 Y、X5 簡單線性迴歸

根據變異數分析 (Analysis of Variance, ANOVA) 結果顯示，整體迴歸模型達到統計上之高度顯著水準 ($F = 2407.61, P < 0.0001$)，顯示預測變項能有效解釋依變項之變異。此 F 值為所有已分析模型中最高者，顯示模型適配度極佳。

在模型解釋力方面，判定係數 R^2 高達 0.8078，調整後判定係數 (Adjusted R^2) 為 0.8074。此結果顯示，在排除離群值後，本迴歸模型中的自變項 (PM_{10} 濃度) 可解釋依變項 ($PM_{2.5}$ 濃度) 約 80.78% 的變異量。相較於未刪除離群值之模型 (R^2 約 0.6401)，解釋力大幅提升 (增加約 16.7 個百分點)。剩餘均方根誤差 (Root MSE) 降至 2.46840，為所有模型中最低，反映出 PM_{10} 與 $PM_{2.5}$ 之間存在極為緊密且穩定的線性關係。

參數估計結果顯示， PM_{10} 濃度對 $PM_{2.5}$ 濃度具有極為顯著的正向影響。截距項 (Intercept) 之估計值為 2.721 ($T = 9.36, P < 0.0001$)； PM_{10} 濃度之未標準化迴歸係數為 0.374 ($T = 49.07, P < 0.0001$)。其標準化迴歸係數高達 0.89875，接近 0.9 的數值顯示兩者之間具有極強的共變關係，這符合物理預期，即 $PM_{2.5}$ 本身即為

PM₁₀的一部分（粒徑小於 2.5 微米的懸浮微粒），因此兩者在濃度消長上呈現高度同步。

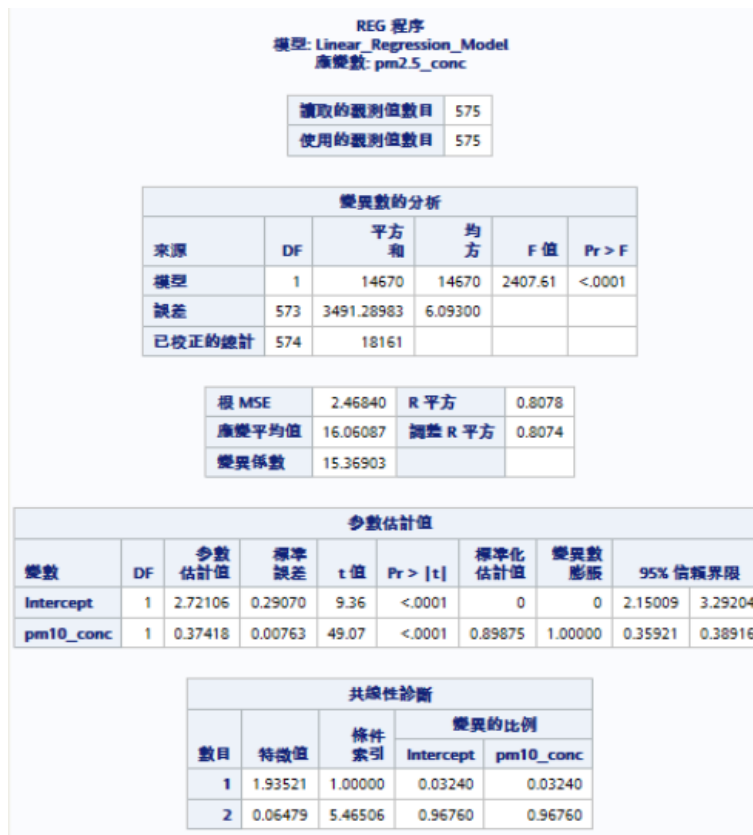


圖 32、刪除後 Y、X5 簡單線性迴歸

(六)刪除後多元迴歸分析

根據左側變異數分析 (ANOVA) 結果顯示，本研究所建構之原始多元線性迴歸模型整體達高度統計顯著水準 ($F = 401.22$, $P < 0.0001$)，顯示模型具備良好的整體解釋能力。在模型解釋力方面，其決定係數 (R^2) 為 0.6970，調整後決定係數為 0.6953，表示在包含離群值的情況下，所納入之解釋變數可共同解釋約 69.5% 的 PM_{2.5}濃度變異量。此外，模型之根均方誤差 (Root MSE) 為 3.919，變異係數 (CV) 為 22.41。

就參數估計結果而言，多數污染物變數對 PM_{2.5}濃度呈現顯著影響。其中，PM₁₀為模型中最具顯著性之解釋變數，其 T 值高達 25.57，顯示 PM₁₀與 PM_{2.5}之間存在極為強烈之正向關聯。一氧化碳 CO 之迴歸係數為 3.720 ($T = 2.39$, $P = 0.017$)，雖然顯著但統計強度相對較弱。值得注意的是，氮氧化物 (NO_x) 呈現顯著負向影響 ($\beta = -0.052$)，這可能反映了 NO_x 與 NO₂之間的光化學轉換機制在統計上的抵銷效應。

據右側變異數分析結果顯示，在排除離群值後，優化模型之統計顯著性大幅提升，F 值躍升至 706.67 (P <0.0001)，顯示模型架構更為穩健。在模型解釋力方面，決定係數 (R²) 顯著提升至 0.8613，調整後決定係數為 0.8601。此結果表明，在剔除極端值干擾後，五種汙染物變數能解釋高達 86% 的 PM_{2.5} 濃度變異，較原始模型提升了約 16.5 個百分點。同時，根均方誤差 (Root MSE) 大幅下降至 2.104，顯示模型的預測精確度獲得顯著改善，預測值更貼近觀測平均值。

參數估計結果顯示，PM₁₀ 依然是最強的預測因子，其 T 值由 25.57 上升至 31.72，標準化係數達 0.691，證實粒狀汙染物之間的共變性是 PM_{2.5} 濃度變化的核心驅動力。一氧化碳 (CO) 的迴歸係數由 3.720 上升至 5.306 (T = 3.39, P = 0.0007)。這顯示在排除離群值後，CO 對 PM_{2.5} 的邊際影響力實際上更為強烈，原始模型可能受到部分低相關極端值的稀釋。NO₂ 維持穩定的正向影響 (T = 6.55)，而 NO_x 仍呈現負向係數 ($\beta = -0.065$, T = -5.55)，確認了此變數間的統計關係並非由離群值所導致，而是反映了穩定的化學反應特徵或共線性結構。

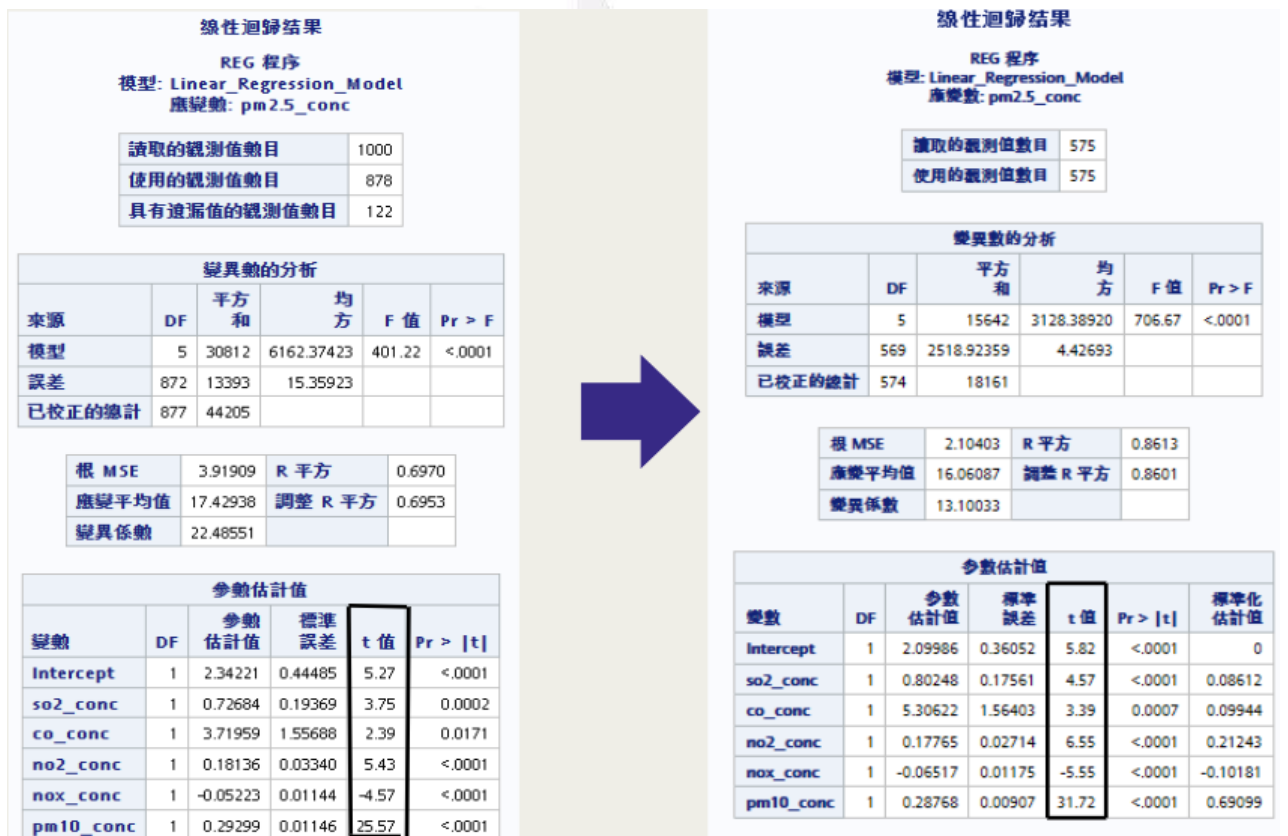


圖 33、刪除離群值後多元迴歸分析比較

(七)刪除離群值後 PM_{2.5}_CONC 的殘差的直方圖

為了確保上述多元線性迴歸模型之推論統計（如 T 檢定與 F 檢定）的有效性，本研究針對模型殘差（Residuals）進行了常態性檢定。線性迴歸模型的基本假設之一為誤差項需服從平均值為 0 的常態分配（Normal Distribution）。觀察殘差分布圖，最顯著的改善在於誤差範圍的顯著收縮。橫軸（殘差值）的分布範圍目前精確地控制在約 -6.0 至 +6.0 之間。

相較於刪除離群值前的模型（其誤差範圍往往擴及+-10 甚至+-20），此結果證實了刪除離群值策略的有效性。極端汙染事件或異常觀測值已被排除，使殘差不再受到少數極端值的拉扯，表明模型現在的預測結果更加集中且穩定，更能反映一般大氣條件下的預測能力。

進一步觀察分布的核心區域，可發現數據在中間區段（約 -1.2 至 +1.2）的分布呈現相對平均的態勢。與理論上的標準常態分布曲線（藍色曲線）相比，實際數據（紅色曲線）並未在 0 點處形成非常突出的單一尖銳高峰，而是呈現出頂部較為平緩的特徵。這說明在刪除離群值後，模型誤差在小範圍內的分布變得更加均勻（Uniform），顯示大部分的預測誤差都均勻落在一個可接受的低誤差區間內，而非過度集中於某一點。

在檢視對稱性時，觀察到分布的最高峰值並非精確位於 0，而是略微偏向左側（約在 -0.4 附近）；同時，在正值區間 1.2 至 2.0 之間的頻率亦稍高於理論值。殘差為負值意味著預測值大於觀測值。因此，峰值微幅左偏（-0.4）暗示了模型存在非常微小的系統性偏移，即模型在預測時，出現「輕微高估（Overestimation）」的頻率略多於「輕微低估」。然而，此偏移量極小，並不影響整體常態分布的鐘形架構。

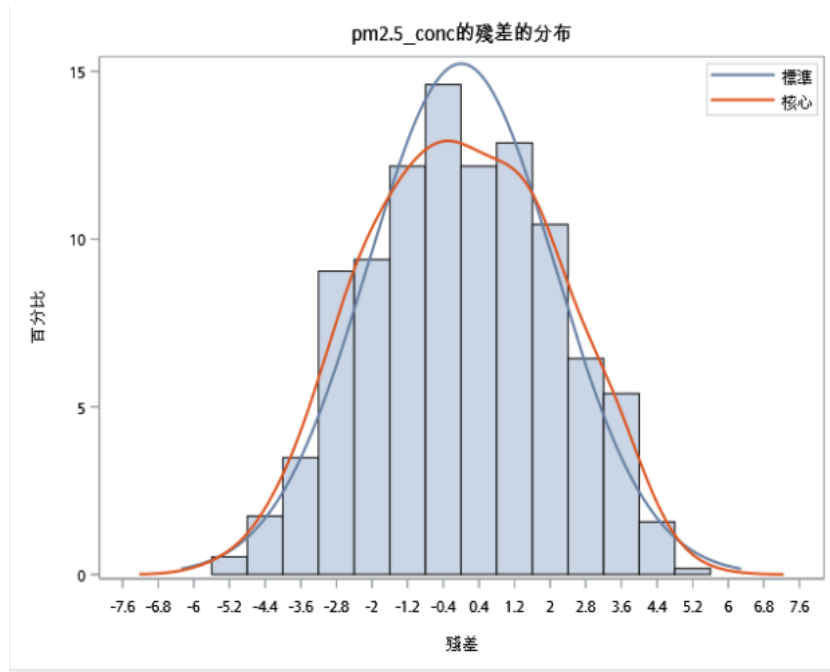


圖 34、刪除離群值後殘差直方圖

(八)刪除離群值後 Q-Q 圖

為了彌補直方圖在判斷尾端分布 (Tail Behavior) 上的不足，本研究進一步採用常態分位數圖 (Normal Q-Q Plot) 來檢視優化後多元迴歸模型殘差的分布情形。Q-Q 圖將觀測到的殘差數值 (Y 軸) 與標準常態分布下的理論分位數 (X 軸) 進行對照，若數據服從常態分配，則散佈點應緊密貼合於 45 度的對角直線 (Reference Line)。Q-Q 圖可以歸納出以下幾個特徵：(1) 高度線性趨勢 (High Linearity)：圖中絕大多數的藍色圓圈 (殘差觀測點) 極為緊密地沿著灰色對角線分布。特別是在標準差 -2 到 +2 的核心區間內，數據點幾乎與參考線重疊。這顯示殘差的中段分布完全符合常態分布的理論預期，進一步佐證了前節直方圖中「對稱且集中」的觀察結果。(2) 尾端行為 (Tail Behavior)：在分位數小於 -2 的區域，數據點依然貼合直線，僅在極端值處 (約 -2.5 以下) 出現非常微小的向下偏離，顯示左側極端誤差的發生機率與理論值相當吻合。在分位數大於 +2 的區域，數據點呈現些微的向下偏離趨勢 (Drifting below the line)。這意味著實際觀測到的極端正殘差值，略小於理論常態分布的預期值。換言之，該模型的右尾屬於「輕尾 (Light-tailed)」特徵，表示模型極少出現預測值遠小於觀測值的極端低估情況。

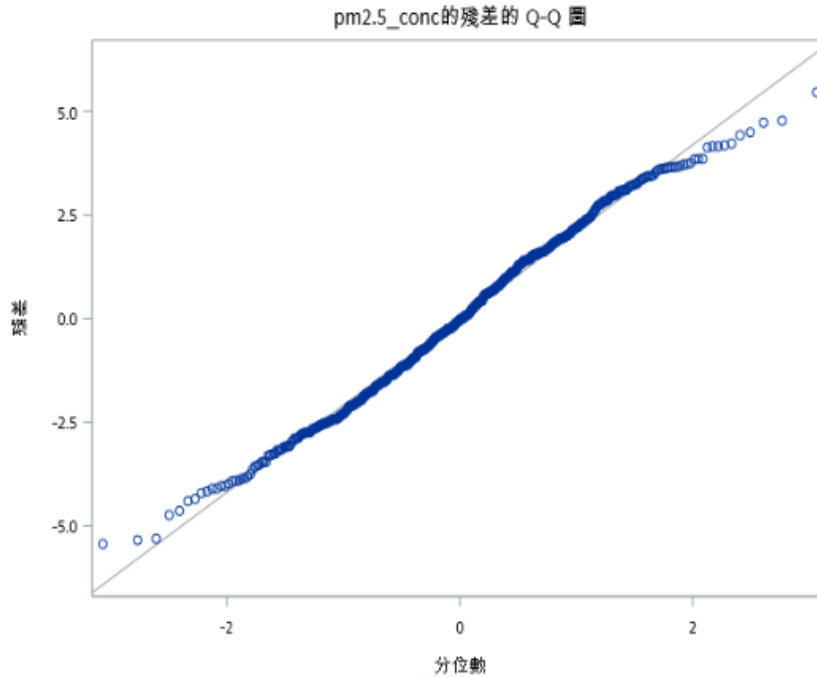


圖 35、刪除離群值後 Q-Q 圖

(九)刪除離群值後常態

在前述直方圖與 Q-Q 圖顯示殘差高度符合常態分布特徵的基礎上，本研究進一步採用 Shapiro-Wilk、Kolmogorov-Smirnov、Cramer-von Mises 及 Anderson-Darling 等四種統計檢定方法進行嚴謹的量化驗證。檢定結果顯示，Shapiro-Wilk 檢定之統計量 W 值高達 0.992239，但其對應之 P 值為 0.0043；Kolmogorov-Smirnov 檢定之統計量 D 為 0.03888，P 值為 0.0340。此外，Cramer-von Mises 與 Anderson-Darling 檢定之 P 值分別為 0.0157 與 0.0076。由於上述四項檢定之 P 值均小於傳統顯著水準 0.05，在嚴格的統計推論下，形式上必須拒絕「殘差服從常態分布」之虛無假設。

然而，詮釋此一統計結果時必須審慎考量「大樣本矛盾」現象。由於本研究排除離群值後之有效樣本數仍多達 575 筆，統計檢定對於微小的分布偏差將變得極為敏感，即使是實務上可忽略的極微幅偏差，也容易導致檢定結果呈現統計顯著差異。值得注意的是，Shapiro-Wilk 的 W 值極度接近 1 (0.992)，這提供了強有力的證據，顯示殘差分布型態與理論常態分布在實質上具有極高的相似度。此外，根據中央極限定理 (Central Limit Theorem)，當樣本數足夠大時，線性迴歸模型對於誤差項常態性假設的違背具有高度的容忍力 (Robustness)。因此，儘管統計檢定拒絕了虛無假設，但綜合考量視覺診斷的高度擬合以及 W 值所反映的分布特徵，此微小的常態性偏離並不會影響迴歸係數估計的不偏性或推論統計的有效性。本研究最終認定該多元迴歸模型之殘差在實務上已充分滿足常態性要求，模型具備良好的統計適配性質，可用於後續之分析與預測。

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.992239	Pr < W	0.0043
Kolmogorov-Smirnov	D	0.03888	Pr > D	0.0340
Cramer-von Mises	W-Sq	0.166392	Pr > W-Sq	0.0157
Anderson-Darling	A-Sq	1.092372	Pr > A-Sq	0.0076

圖 36、刪除離群值後常態性檢定

(十)刪除離群值後均勻

為了進一步檢驗迴歸模型是否符合「變異數齊一性 (Homoscedasticity)」與「線性關係」之基本假設，本研究繪製了學生化殘差 (RStudent) 與預測值 (Predicted Value) 之散佈圖進行視覺化診斷。在此圖中，橫軸代表多元迴歸模型預測之 PM_{2.5} 濃度，縱軸則為經學生化處理後的殘差值。依據高斯-馬可夫定理 (Gauss-Markov Theorem)，若模型設定正確且誤差項變異數固定，散佈點應呈現無規律的隨機分布 (Random Cloud)，且均勻落在 0 軸上下，不應隨預測值大小而呈現擴散、收斂或特定幾何圖形。

觀察下圖之散佈型態，數據點大致均勻地散佈在以 0 為中心的水平帶狀區域內。在預測值較低 (約 10-20) 與較高 (約 20-30) 的區間中，殘差的垂直分布範圍並無顯著差異，並未出現典型違反假設的「扇形 (Fan shape)」或「漏斗形」擴散特徵。此一結果顯示，殘差之變異程度並未隨著預測值的增加而產生系統性的改變，證實本模型符合變異數齊一性之假設，意即模型在不同濃度區間下的預測精確度具有一致性。此外，散佈圖中數據點並未排列成拋物線 (U 型或倒 U 型) 或其他非線性趨勢，這進一步支持了應變數與自變數之間存在線性關係的假設，說明無須進行額外的高次項轉換。

值得注意的是，圖中設有標準差±2 的水平參考界線。檢視圖表可見，絕大多數的觀測點均緊密地落在這兩條界線所構成的信賴區間之內，僅極少數點位 (Outliers) 略微超出此範圍，且並無極端遠離界線的異常點。這再次印證了前述章節中進行離群值剔除程序的成效，顯示模型在排除極端干擾後，其預測誤差已獲得有效控制，殘差結構穩定且無明顯的異質性 (Heteroscedasticity)，模型具備良好的預測穩健性。

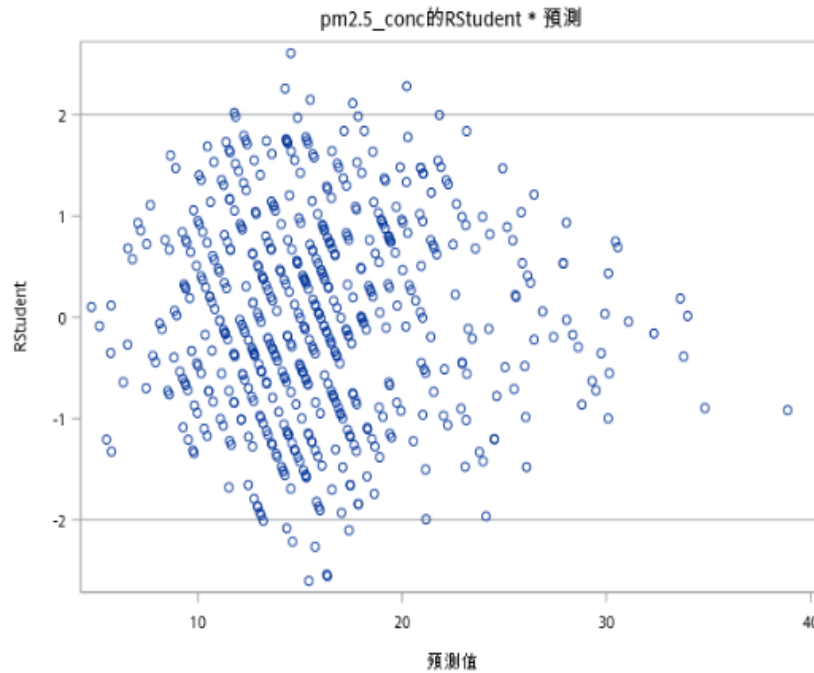


圖 37、刪除離群值後學生化殘差與預測值的散佈圖

(十一)刪除離群值後均勻

儘管前述殘差與預測值之散佈圖在視覺上並未呈現明顯的變異數不齊一特徵，為求統計推論之嚴謹性，本研究進一步採用 White 檢定概念下之「第一與第二動差規格檢定 (Test of First and Second Moment Specification)」進行量化驗證。此檢定之虛無假設 (H_0) 為模型設定正確且誤差項具有同質變異性 (Homoscedasticity)，對立假設 (H_1) 則為模型存在異質變異性或函數形式設定誤誤 (如遺漏非線性項)。

根據 SAS 分析報表顯示，該檢定之自由度 (DF) 為 20，卡方統計量 (Chi-Square) 為 73.63，其對應之顯著性機率 (P 值) 小於 0.0001。由於 p 值遠小於顯著水準 0.05，在嚴格的統計定義下，我們必須拒絕虛無假設。此統計結果暗示模型殘差可能存在異質變異性，或者模型中可能仍遺漏了部分解釋變數的高次項 (如平方項或交互作用項)。

然而，詮釋此一檢定結果時，必須審慎考量樣本規模的影響與統計檢定的敏感度。本研究在刪除離群值後之有效樣本數仍達 575 筆，屬於大樣本迴歸分析。在大型樣本中，規格檢定 (Specification Tests) 往往具有極高的檢定力 (Power)，即使殘差變異數僅存在極微小的波動，或數據分布與理論假設有極些微的偏離，檢定結果均極易呈現顯著差異。對照前節圖 14 之學生化殘差散佈圖，數據點均勻分布於水平帶狀區間內，並無明顯擴散或系統性圖形，顯示其異質性在「實務上」並不嚴重。因此，雖然統計檢定提示潛在的規格問題，但綜合

考量視覺證據與大樣本特性，本研究認為該模型的變異數結構在可接受範圍內，並未嚴重違背線性迴歸之基本假設，若需進一步提升參數推論的穩健性，未來研究可考慮採用異質變異性穩健標準誤(Heteroscedasticity-Consistent Standard Errors, HC SE) 進行修正。

DF	卡方	Pr > ChiSq
20	73.63	<.0001

圖 38、刪除離群值後均檢定

(十二)刪除離群值後獨立

為了進一步剖析模型誤差的結構特性，並檢視殘差是否獨立於觀測數據的分布型態，本研究繪製了 PM_{2.5}觀測值 (Y 軸) 與學生化殘差 (X 軸) 之散佈圖進行診斷。觀察圖表數據分布，數據點呈現出一種明顯的正向線性趨勢。需釐清的是，此正向相關性在線性迴歸模型中是數學上的必然結果，因為根據迴歸方程式定義，觀測值等於預測值加上殘差，因此當判定係數 R 平方未達 1.0 時，較高的觀測值往往伴隨著較大的正殘差，而較低的觀測值則伴隨著負殘差。因此，圖中呈現的分布型態是符合理論預期的，並不代表違反模型假設。

儘管存在預期的正相關，此圖表在檢驗樣本獨立性與誤差結構上仍提供了關鍵資訊。圖中數據點呈現單且連續的雲狀分布，並未分裂成兩個或多個截然不同的聚類，這暗示了樣本數據很可能來自同一個母體，並未受到隱藏的分類變數干擾，間接支持了樣本之間的獨立性假設。此外，觀察橫軸的學生化殘差，絕大多數點位均勻地落在 -2 至 +2 的標準差區間內。無論是在低濃度區或高濃度區，殘差的擴散範圍大致保持一致，這再次佐證了模型符合變異數齊一性，且誤差項的產生並未受到觀測值規模的劇烈扭曲。

必須指出的是，雖然本散佈圖顯示出殘差在觀測值空間中並無特殊的非線性構造或異常極端值，但針對空氣品質這類可能具有時間序列特性的數據，最嚴謹的獨立性檢定仍建議搭配 Durbin-Watson (DW) 統計量來偵測殘差是否存在序列相關。然而，結合前述殘差常態性與變異數齊一性的檢定結果，以及本圖顯示無明顯異常分群或異質擴散的現象，我們可以推論本模型之誤差機制在處理不同濃度區間的樣本時是穩定且一致的，未發現明顯違反獨立抽樣假設的視覺證據，模型整體具備良好的統計穩健性。

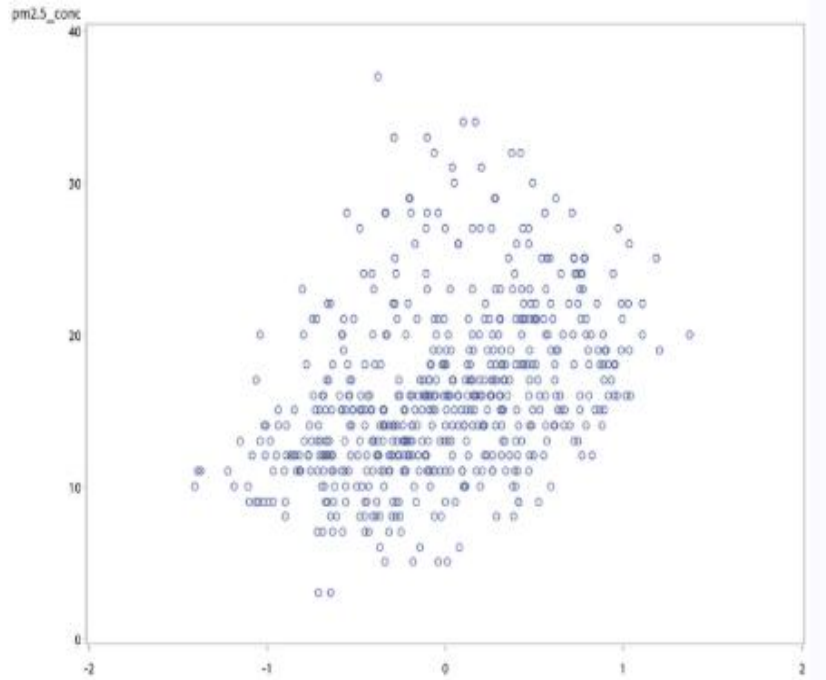


圖 39、刪除離群值後觀測值與殘差散佈圖

如前述針對觀測值與殘差散佈圖之視覺診斷，本研究進一步採用 Durbin-Watson (DW) 統計量，以更嚴謹的量化標準檢測模型殘差是否存在一階序列相關 (First-order Serial Correlation)。在涉及環境監測數據的迴歸分析中，由於樣本採集往往具有時間順序性，誤差項之間容易產生前後期干擾(自相關)，進而導致參數估計之標準誤失真。Durbin-Watson 統計量的數值範圍介於 0 至 4 之間，理論上數值越接近 2，代表殘差之間越不存在相關性；數值趨近於 0 則暗示正自相關，趨近於 4 則暗示負自相關。

根據 SAS 分析報表顯示，本研究優化後模型之 Durbin-Watson D 統計量為 1.827，其對應之一階自相關係數 (1st Order Autocorrelation) 僅為 0.085。一般統計實務上，若 DW 值落在 1.5 至 2.5 的區間內，即可視為殘差符合獨立性假設。本模型之 DW 值 (1.827) 不僅落在此安全區間內，且高度趨近於代表完全獨立的理論值 2.0，顯示誤差項之間並無顯著的序列相關問題。此一量化結果驗證了先前的視覺推論，證實本研究所建構之多元迴歸模型成功過濾了時間序列上的噪音干擾，其殘差項滿足獨立常態分配 (i.i.d.) 之基本假設，確保了模型參數估計與顯著性檢定的統計穩健性。

Durbin-Watson D	1.827
觀測值數目	575
一階自相關	0.085

圖 40、刪除離群值後 Durbin-Watson (DW)檢定

九、模式選取

為了從眾多候選汙染物變數中篩選出最具預測效率的變數子集，並確保最終模型在最大化解釋能力與最小化模型複雜度之間取得最佳平衡，本研究採用了多種統計變數選取策略進行交叉驗證，包括 Mallows'C(P)準則、前進選取法 (Forward Selection)、向後消去法 (Backward Elimination) 以及逐步選取法 (Stepwise Selection)。

首先，依據 Mallows'C(P)選取方法的分析結果顯示，包含全數五個預測變數 (SO₂,CO,NO₂,NO_x,PM₁₀) 的模型表現最佳。根據統計理論，當 C(P)值趨近於模型參數個數 P (即解釋變數個數加上截距項) 時，該模型被視為無偏誤且最適配的。觀察報表數據，包含五個變數的完整模型其 C(P) 值恰好為 6.0000。由於本模型參數個數 P 為 6 (5 個自變數 + 1 個截距項)，此結果顯示 C(P) = P，代表該模型在變異數與偏差之間達到了完美的平衡，且該組合之判定係數 R 平方達到 0.8613，為所有可能模型中最高者。

其次，前進選取法 (Forward Selection) 的演算過程進一步揭示了各變數進入模型的相對重要性順序。分析顯示，PM₁₀為第一個進入模型的變數，單獨解釋了 80.78%的變異量，顯示其為影響 PM_{2.5}最關鍵的主導因子。隨後依序進入的變數為 NO₂、NO_x、SO₂，最後為 CO。值得注意的是，即便 CO 是最後進入模型的變數，其偏 F 檢定 (Partial F Test) 結果顯示 F 值為 11.51 (P=0.0007)，仍達顯著水準，證實其對提升模型解釋力具有獨特的邊際貢獻。該程序最終在步驟 5 停止，選取了包含所有五個變數的模型。

此外，逐步選取法 (Stepwise Selection) 與向後消去法 (Backward Elimination) 的執行結果亦呈現完全的一致性。逐步選取法在每一步驟加入變數後會重新檢視既有變數的顯著性，結果顯示最終模型依然保留了所有五個變數，且最終 C(P) 值維持在 6.0000，確認模型中並無多餘或不顯著的變數需被移除。同樣地，向後消去法從完整模型開始檢測，結果顯示所有變數在 0.1000 的顯著水準下皆具統計意義，因此程序在步驟 0 即停止，未移除任何變數。

綜上所述，無論是基於訊息準則的 C(P)方法，或是基於統計檢定的三種逐步演算法，四種不同的選取策略均一致收斂至同一個結果，即包含 SO₂、CO、

NO₂、NO_x 與 PM₁₀ 的完整模型為本研究數據下的最佳預測模型。此高度的一致性提供了強而有力的證據，證實這五個空氣汙染物指標在統計上皆不可或缺，且該多元迴歸模型具備極佳的結構穩健性與解釋效力。



圖 41、向前選取法模型選擇結果



圖 42、向候選取法模型選擇結果



圖 43、逐步選取法模型選擇結果

空氣汙染與氣象因子、其他汙染物之關係

REG 程序
模型: Linear_Regression_Model
應變數: pm2.5_conc

C(p) 選擇方法

讀取的觀測值數目	575
使用的觀測值數目	575

模型索引	模型中的數目	C(p)	R 平方	調整的 R 平方	SSE	模型中的變數
1	5	6.0000	0.8613	0.8601	2518.92359	so2_conc co_conc no2_conc nox_conc pm10_conc
2	4	15.5102	0.8585	0.8575	2569.87824	so2_conc no2_conc nox_conc pm10_conc
3	4	24.8819	0.8562	0.8552	2611.36625	co_conc no2_conc nox_conc pm10_conc
4	4	34.7631	0.8538	0.8528	2655.10970	so2_conc co_conc no2_conc pm10_conc
5	3	39.3458	0.8522	0.8514	2684.25076	no2_conc nox_conc pm10_conc
6	3	43.6313	0.8512	0.8504	2703.22253	so2_conc no2_conc pm10_conc
7	4	46.8597	0.8509	0.8498	2708.66038	so2_conc co_conc nox_conc pm10_conc
8	3	57.8132	0.8477	0.8469	2766.00488	co_conc no2_conc pm10_conc
9	3	57.8183	0.8477	0.8469	2766.02715	so2_conc co_conc pm10_conc
10	3	64.1569	0.8461	0.8453	2794.08783	co_conc nox_conc pm10_conc
11	2	71.9404	0.8438	0.8432	2837.39888	no2_conc pm10_conc
12	2	78.2413	0.8422	0.8417	2865.29244	co_conc pm10_conc
13	2	178.4155	0.8178	0.8172	3308.75671	so2_conc pm10_conc
14	3	179.5681	0.8180	0.8171	3305.00524	so2_conc nox_conc pm10_conc
15	1	217.6479	0.8078	0.8074	3491.28983	pm10_conc
16	2	219.0957	0.8079	0.8072	3488.84497	nox_conc pm10_conc
17	4	1010.044	0.6161	0.6134	6972.60869	so2_conc co_conc no2_conc nox_conc
18	3	1024.160	0.6121	0.6101	7043.95362	so2_conc no2_conc nox_conc
19	3	1077.918	0.5990	0.5969	7281.93660	so2_conc co_conc no2_conc
20	2	1090.999	0.5954	0.5939	7348.69846	so2_conc no2_conc
21	3	1331.934	0.5371	0.5347	8406.44895	so2_conc co_conc nox_conc
22	2	1337.692	0.5352	0.5336	8440.79503	so2_conc co_conc
23	3	1371.986	0.5273	0.5249	8583.75823	co_conc no2_conc nox_conc
24	2	1420.869	0.5149	0.5132	8809.01349	no2_conc nox_conc
25	2	1486.093	0.4990	0.4973	9097.75235	co_conc no2_conc
26	1	1535.557	0.4865	0.4856	9325.58181	no2_conc
27	2	1790.405	0.4249	0.4229	10445	co_conc nox_conc
28	1	1810.518	0.4195	0.4185	10543	co_conc
29	2	2099.134	0.3496	0.3473	11812	so2_conc nox_conc
30	1	2196.626	0.3254	0.3242	12252	so2_conc
31	1	3349.682	0.0443	0.0426	17357	nox_conc

圖 44、Mallows' Cp 選取法模型選擇結果

第四章、結論與建議

第一節、研究結論

(一) 氣象因子與 PM_{2.5}之連動性分析

本研究針對氣象因子中風速對 PM_{2.5}濃度之影響進行進一步探討，將風速分解為東西向 (X6) 與南北向 (X7) 之向量分量，以其更精確描述風向與風速的綜合效果。相關文獻普遍指出，當風速減弱時，局部大氣擴散條件不佳，容易導致汙染物於近地層累積，進而提高 PM_{2.5}濃度。然而，本研究之實證結果顯示，在線性迴歸架構下，風速向量分量與 PM_{2.5}之間僅呈現極弱相關，其散佈圖分布趨近隨機，未能顯示明確的線性趨勢。基於上述結果，風速向量分量無法有效建立對 PM_{2.5}的線性解釋關係，亦未對模型預測力提供實質貢獻，故最終於多元預測模型建構過程中予以排除。

(二) 汙染物間之共伴效應分析

在汙染物相關性分析方面，本研究結果顯示，PM_{2.5}分別與 SO₂、CO、NO₂、NO_x 及 PM₁₀ 均呈現正相關關係，顯示細懸浮微粒濃度與其他汙染物之間具有高度共伴變動特性。其中，PM_{2.5}與 PM₁₀之線性關係最為顯著，散佈圖呈現狹長且明確的上升趨勢，顯示兩者在時間變化上具有高度同步性。此結果不僅反映 PM₁₀ 為 PM_{2.5}的重要來源指標之一，亦支持細懸浮微粒與可吸入懸浮微粒在生成機制與排放來源上具有高度關聯性，進一步突顯 PM₁₀在 PM_{2.5}預測模型中的關鍵角色。

(三) 最適多元預測模型之建立與解釋

在考量汙染物交互影響之情況下，本研究首先建立包含多項汙染物的多元線性迴歸模型，其初步模型解釋力為 69.70%。為提升模型穩定性與估計準確度，進一步針對顯著離群值與高影響點進行檢視與剔除，結果顯示模型解釋力顯著提升至 86.13%，顯示離群值對模型估計具有實質影響。隨後，本研究運用 Mallows' Cp 準則進行模型篩選，以平衡模型複雜度與預測誤差。分析結果顯示，包含 SO₂、CO、NO₂、NO_x 及 PM₁₀ 共五個解釋變數之模型，其 C(P) 值為 6.0000。依統計理論，當 C(P) 值約等於模型參數個數 P (本研究中為 5 個自變數加 1 個截距項) 時，表示模型預測偏差最小，且未出現過度擬合現象，故可視為最優模型結構。在該最終模型中，各解釋變數皆達極顯著水準，顯示其對 PM_{2.5}濃度具有穩定且顯著的解釋能力。其中，一氧化碳 (CO) 之邊際影響效果最為顯著，

在其他變數維持不變的情況下，CO 每增加一單位濃度，PM_{2.5}平均增加約 5.31 單位，顯示 CO 可能反映燃燒來源活動，對 PM_{2.5}變動具有高度代表性。

第二節、研究建議

雖然本研究在模型解釋力與變數篩選上已取得一定成果，仍有若干限制與改進空間，提出以下建議供後續研究參考。

第一、針對統計假設未完全通過之問題，雖然透過排除離群值可提升模型解釋力，但殘差分析結果顯示，模型仍存在異質變異數與殘差非常態性之情形。未來研究可考慮採用變數轉換、加權最小平方法（WLS）或穩健迴歸等方法，以改善模型對統計假設的符合程度。

第二、關於氣象因子之影響，本研究僅以線性關係與散佈圖進行初步判斷，結果顯示風速與風向呈現近似隨機分布。然而，氣象因子對空氣汙染之影響可能具有非線性特性或交互作用，建議後續研究可嘗試非線性模型、分群分析或機器學習方法，以更全面評估氣象條件對 PM_{2.5}的潛在影響。

第三、空氣品質資料本身具有明顯的時間序列特性，前一時點之汙染濃度可能對下一時點產生影響，此亦可能為 Durbin-Watson 值未達理想值 2 的主要原因之一。未來研究可引入時間序列模型（如 ARIMA、SARIMA）或動態迴歸模型，以更妥善處理資料的自相關問題。

最後，在研究背景中曾提及境外汙染傳輸與季節性特徵對空氣品質之影響，惟本研究之定量分析尚未進一步區分不同季節或風向來源。建議後續研究可依季節分層，或結合氣流軌跡分析，以強化研究結論與研究動機之連結性，提升整體研究之完整性與政策參考價值。

參考文獻

1. P.H. Ding, G.S. Wang, Y.L. Guo, S.C. Chang, G.H. Wan (2017) Urban air pollution and meteorological factors affect emergency department visits of elderly patients with chronic obstructive pulmonary disease in Taiwan. *Environ Pollut.*, 224, 751-758. doi: 10.1016/j.envpol.2016.12.035.
2. F.Y. Cheng, C.H. Hsu (2019) Long-term variations in PM_{2.5} concentrations under changing meteorological conditions in Taiwan. *Scientific Reports*, 9, 6635. <https://doi.org/10.1038/s41598-019-43104-x>
3. C.H. Hsu and F.Y. Cheng (2019). Synoptic weather patterns and associated air pollution in Taiwan. *Aerosol and Air Quality Research*, 19, 1139-1151. <https://doi.org/10.4209/aaqr.2018.09.0348>
4. 數據來源:空氣品質指標(AQI) 2025/11/19 8:00~19:00 - 政府資料開放平台 <https://data.gov.tw/dataset/151824>



附錄一、分工表

成員姓名	工作內容	佔比
陳柏荏	研究背景動機文獻找尋、敘述統計、散步圖製圖分析、迴歸分析製圖分析、VIF 製圖分析、殘差分析(性質檢定)製圖、離群值製圖與排除離群值、模型製圖、結論建議、修正書面報告	23%
林逸軒	散步圖分析、迴歸分析分析、書面報告書寫	13%
鍾秉霖	上台報告、多元迴歸分析、書面報告書寫	12%
李怡青	上台報告、美編、排版	10%
廖家鴻	殘差分析、部分內容分析	10%
陳昱誠	上台報告、書面報告書寫	10%
蔡秉翰	上台報告、書面報告書寫	10%
康展碩	書面報告書寫	12%