112-2 統計學一下 期末專案報告書

第一組

一、組員名單

B11705007 袁詠宸
B11705017 陳芃宇
B11705024 謝友毅
B11705025 林澐稚
B11705034 蔡逸芃

二、分析主題

因應大眾對於注意空氣品質問題的意識增強，我們以台灣環境部提供的各測站在2023年每月每日空氣資料，欲比較不同測站的各項汙染物濃度差異與預測未來之特定汙染物濃度。

這份統計旨在找出不同空氣條件下或不同測站所在區域，台灣可能存在的空氣汙染物問題，以分析和預測結果找出致使空氣品質較差的環境和要素，並提醒大眾在哪些情形下該多留意空氣中的汙染物。

三、分析概要

基本分析

特定時間內，各懸浮粒子、氣體濃度之變化。

多母體比較

比較不同測站間之各項數值是否有顯著差異。

多變量迴歸分析

以風速、溫度、相對濕度作為自變數，各懸浮粒子、污染性氣體濃度作為應變數，嘗試找出關聯性，並預測未來相近日期之空氣品質。應謹慎處理自變數之間之相關性，並考慮引用外部資料（如日雨量、用電量等）豐富模型。

時間序列分析

分析懸浮粒子與污染性氣體濃度於不同季節之循環與長期趨勢。

四、資料說明

歷史資料涵蓋1982年5月至2024年3月，共有8個欄位，數據以區間資料(Interval) 以及日期為主。以下列出用於分析之重要欄位。

siteid
測站代碼。

sitename
測站名稱。

itemid
測項代碼。

itemname, itemengname, itemunit
分別為測項名稱、測項英文名與測項單位。

itemname	itemengname	itemunit
細懸浮微粒	PM2.5	μg/m3
風速	WIND_SPEED	m/sec
小時風速值	WS_HR	m/sec
溫度	AMB_TEMP	℃
相對濕度	RH	%
二氧化硫	SO2	ppb
一氧化碳	CO	ppm
臭氧	O3	ppb
懸浮微粒	PM10	μg/m3
氮氧化物	NOx	ppb
一氧化氮	NO	ppb
二氧化氮	NO2	ppb
甲烷	CH4	ppm
非甲烷碳氫化合物	NMHC	ppm
總碳氫化物	THC	ppm

monitordate
監測日期。格式為「yyyy/mm/dd」。

concentration
監測數值。

五、範例資料

以前五筆資料作為示例。

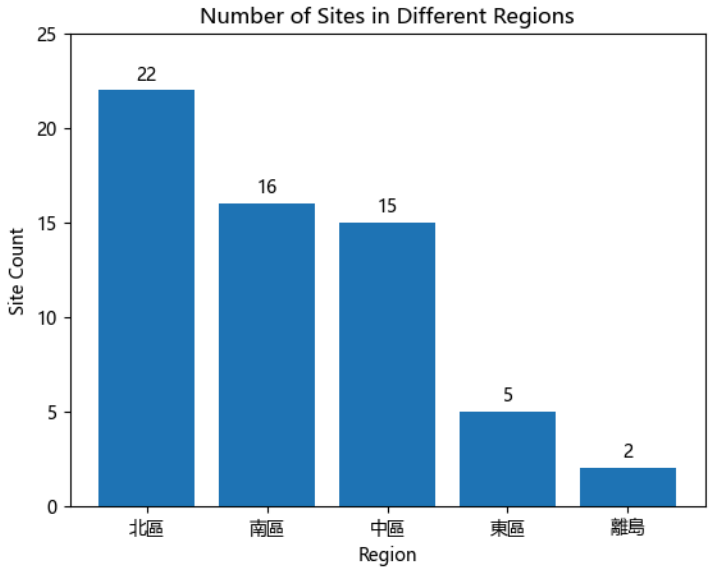
SiteId(測站代碼)、SiteName(測站名稱)、ItemId(測項代碼)、ItemName(測項名稱)、ItemEngName(測項英文名稱)、ItemUnit(測項單位)、MonitorDate(監測日期)、Concentration(數值)。

siteid	sitename	itemid	itemname	itemengname	itemunit	monitordate	concentration
11	士林	10	風速	WIND_SPEED	m/sec	2024/4/1	2
11	士林	14	溫度	AMB_TEMP	℃	2024/4/1	24.9
11	士林	33	細懸浮微粒	PM2.5	μg/m3	2024/4/1	12.4
11	士林	38	相對濕度	RH	%	2024/4/1	60
11	士林	143	小時風速值	WS_HR	m/sec	2024/4/1	1.4

六、分析結果

由於各地的天氣、地理因素不同，我們依照中央氣象局的分類將臺灣分為北區、中區、南區、東區、離島。

區域	測站
北區	三壢、三重、中壢、土城、基隆、富貴角、平鎮、新竹、新莊、松山、板橋、林口、桃園、永和、汐止、淡水、湖口、竹東、苗栗、觀音、頭份、龍潭
中區	二林、南投、嘉義、埔里、大城、大里、崙背、斗六、新港、朴子、竹山、線西、臺西、西屯、豐原
南區	仁武、前鎮、善化、大寮、安南、屏東、左營、復興、恆春、新營、林園、橋頭、潮州、美濃、台南、鳳山
東區	冬山、宜蘭、臺東、花蓮、關山
離島	金門、馬祖



接下來，利用Multiple Comparison分類出每地區中相似的測站群，並將各地區每個測站群的資料取平均進行後續的分析。

※以下分析結果均有通過檢定，檢定結果在程式HTML檔中呈現

北區

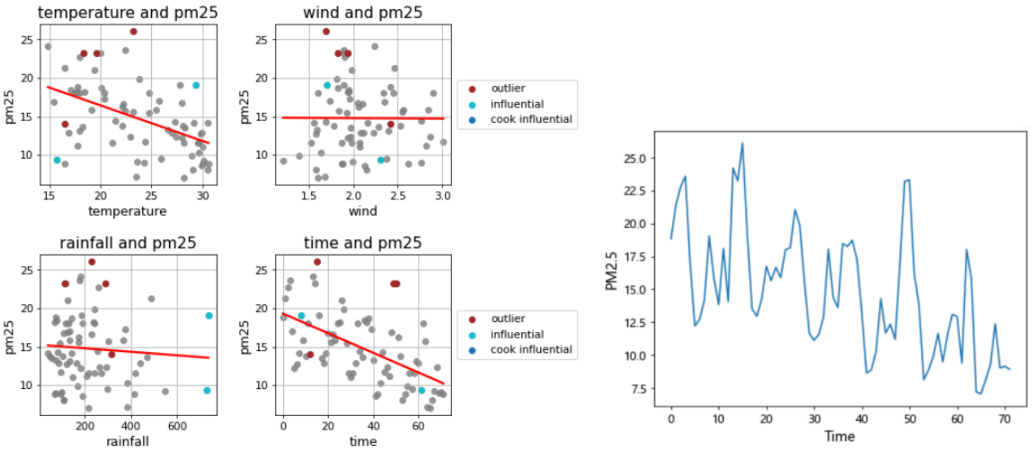
分為兩群測站：雨量皆使用所有北部測站平均

雨量測站：彭佳嶼, 基隆, 臺北, 鞍部, 竹子湖, 淡水, 板橋, 桃園, 新屋, 新竹, 苗栗（共11個測站）

第一區：

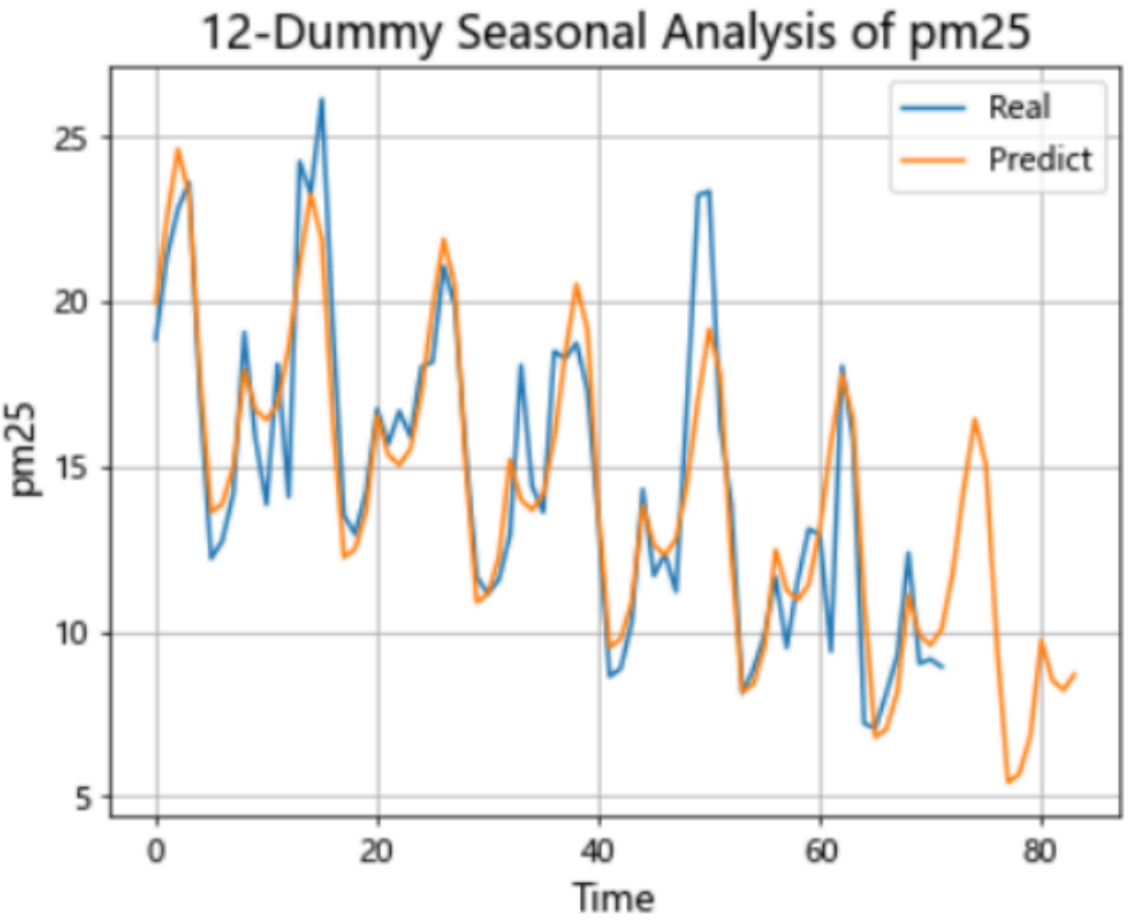
苗栗, 中壢, 觀音, 桃園, 三重, 湖口, 新竹, 平鎮, 三義, 頭份, 新莊, 龍潭, 板橋, 永和, 土城（共15個測站）

基本分析



但由於multiple regression的model沒有過，就沒有放了

能觀察到較明顯的是time的趨勢，因此使用dummy variable來分析



使用Dummy variable作分析但沒有過residual analysis。

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.857
Model:	OLS	Adj. R-squared:	0.822
Method:	Least Squares	F-statistic:	24.37
Date:	Mon, 10 Jun 2024	Prob (F-statistic):	5.27e-19
Time:	01:44:39	Log-Likelihood:	-142.02
No. Observations:	72	AIC:	314.0
Df Residuals:	57	BIC:	348.2
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	23.2462	2.347	9.905	0.000	18.546	27.946
time	-0.1183	0.012	-9.865	0.000	-0.142	-0.094
rainfall	-0.0047	0.002	-2.539	0.014	-0.008	-0.001
wind	-2.5168	1.130	-2.227	0.030	-4.780	-0.254
Month_0	3.2661	0.872	3.744	0.000	1.519	5.013
Month_1	6.4298	0.842	7.634	0.000	4.743	8.116
Month_2	7.1226	0.771	9.241	0.000	5.579	8.666
Month_3	6.3347	0.771	8.215	0.000	4.791	7.879
Month_4	0.2963	0.783	0.378	0.707	-1.273	1.865
Month_5	-3.2653	0.769	-4.246	0.000	-4.805	-1.725
Month_6	-3.4173	0.788	-4.339	0.000	-4.994	-1.840
Month_7	-1.8114	0.827	-2.190	0.033	-3.468	-0.155
Month_8	1.8384	0.795	2.313	0.024	0.247	3.430
Month_9	2.1139	1.169	1.808	0.076	-0.227	4.455
Month_10	1.4075	0.955	1.474	0.146	-0.504	3.319
Month_11	2.9310	1.126	2.602	0.012	0.675	5.187

Omnibus:	5.958	Durbin-Watson:	1.614
Prob(Omnibus):	0.051	Jarque-Bera (JB):	5.271
Skew:	0.521	Prob(JB):	0.0717
Kurtosis:	3.819	Cond. No.	3.78e+18

但在做出最佳模型的時候，仍能使用dummy variable，加入更多變數後，residual analysis就過了。

第二區：

松山, 竹東, 富貴角, 汐止, 林口, 基隆, 淡水（共7個測站）

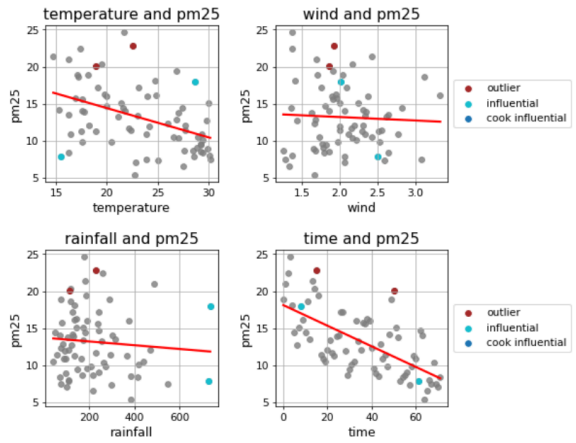
基本分析

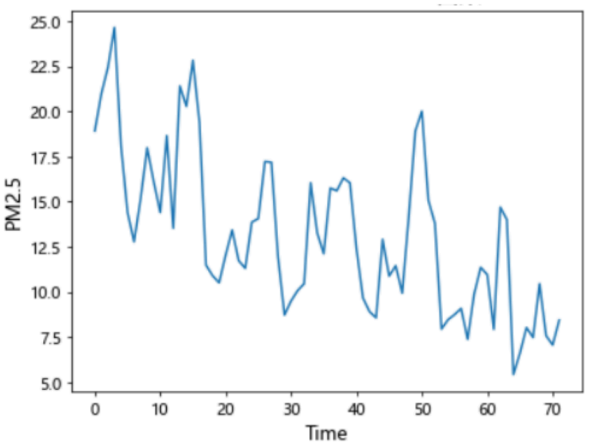
OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.637
Model:	OLS	Adj. R-squared:	0.615
Method:	Least Squares	F-statistic:	29.57
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	4.11e-14
Time:	22:39:39	Log-Likelihood:	-172.21
No. Observations:	72	AIC:	354.4
Df Residuals:	67	BIC:	365.8
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	33.2289	2.832	11.732	0.000	27.575	38.882
temperature	-0.4478	0.073	-6.121	0.000	-0.594	-0.302
wind	-2.2615	0.776	-2.915	0.005	-3.810	-0.713
rainfall	-0.0016	0.002	-0.732	0.467	-0.006	0.003
time	-0.1323	0.016	-8.421	0.000	-0.164	-0.101

Omnibus:	3.372	Durbin-Watson:	1.245
Prob(Omnibus):	0.185	Jarque-Bera (JB):	3.302
Skew:	0.483	Prob(JB):	0.192
Kurtosis:	2.590	Cond. No.	2.34e+03





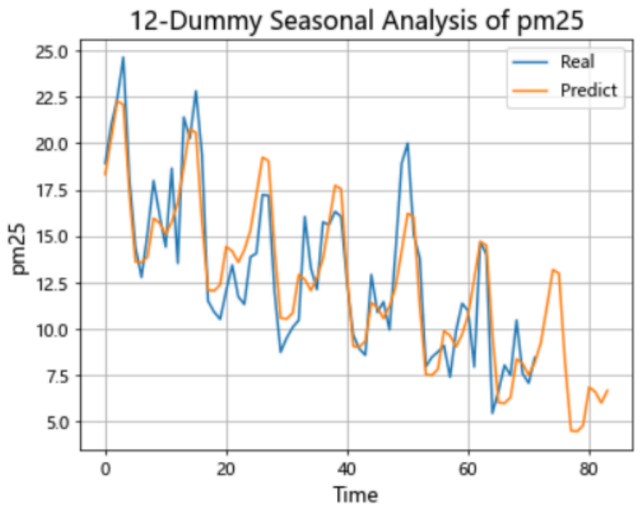
從散佈圖來看，僅時間有較明顯的趨勢。
折線圖中也可看出時間有很重要的影響，因此試著做dummy的時間序列分析

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.814
Model:	OLS	Adj. R-squared:	0.776
Method:	Least Squares	F-statistic:	21.51
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	3.17e-17
Time:	22:39:41	Log-Likelihood:	-148.13
No. Observations:	72	AIC:	322.3
Df Residuals:	59	BIC:	351.9
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	17.1560	0.986	17.398	0.000	15.183	19.129
Time	-0.1265	0.012	-10.520	0.000	-0.151	-0.102
Season_1	1.1759	1.215	0.968	0.337	-1.255	3.607
Season_2	3.2510	1.214	2.679	0.010	0.823	5.679
Season_3	5.3829	1.212	4.440	0.000	2.957	7.809
Season_4	5.3159	1.211	4.388	0.000	2.892	7.740
Season_5	0.6641	1.211	0.549	0.585	-1.758	3.086
Season_6	-2.9198	1.210	-2.414	0.019	-5.341	-0.499
Season_7	-2.8390	1.209	-2.348	0.022	-5.258	-0.420
Season_8	-2.3856	1.209	-1.974	0.053	-4.804	0.033
Season_9	-0.1925	1.208	-0.159	0.874	-2.610	2.225
Season_10	-0.3183	1.208	-0.264	0.793	-2.735	2.099
Season_11	-0.7937	1.208	-0.657	0.514	-3.210	1.623

Omnibus:	0.032	Durbin-Watson:	1.571
Prob(Omnibus):	0.984	Jarque-Bera (JB):	0.032
Skew:	0.016	Prob(JB):	0.984
Kurtosis:	2.901	Cond. No.	516.



Dummy分析有不錯的R-squared值，並且圖中也能看出明顯的季節趨勢

Dep. Variable:	pm25	R-squared:	0.832
Model:	OLS	Adj. R-squared:	0.791
Method:	Least Squares	F-statistic:	20.14
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	4.39e-17
Time:	22:39:42	Log-Likelihood:	-144.50
No. Observations:	72	AIC:	319.0
Df Residuals:	57	BIC:	353.1
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	19.8417	1.699	11.676	0.000	16.439	23.245
time	-0.1263	0.012	-10.702	0.000	-0.150	-0.103
rainfall	-0.0031	0.002	-1.653	0.104	-0.007	0.001
wind	-1.5374	0.825	-1.864	0.067	-3.189	0.114
Month_0	2.2957	0.830	2.766	0.008	0.634	3.958
Month_1	4.8349	0.852	5.675	0.000	3.129	6.541
Month_2	6.0953	0.793	7.683	0.000	4.507	7.684
Month_3	6.2309	0.798	7.807	0.000	4.633	7.829
Month_4	1.1106	0.808	1.375	0.175	-0.507	2.728
Month_5	-2.4981	0.810	-3.082	0.003	-4.121	-0.875
Month_6	-2.5469	0.823	-3.093	0.003	-4.196	-0.898
Month_7	-1.5409	0.819	-1.882	0.065	-3.180	0.099
Month_8	1.2274	0.832	1.476	0.145	-0.438	2.893
Month_9	1.7148	1.071	1.601	0.115	-0.430	3.859
Month_10	0.7725	0.891	0.867	0.389	-1.011	2.556
Month_11	2.1455	0.993	2.160	0.035	0.157	4.134

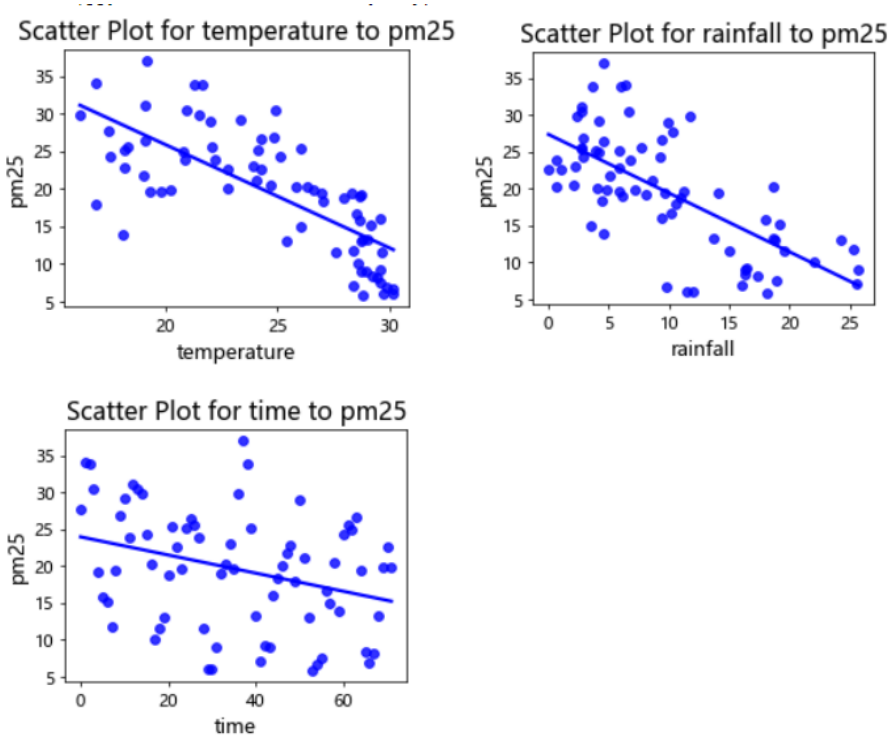
Omnibus:	0.805	Durbin-Watson:	1.387
Prob(Omnibus):	0.669	Jarque-Bera (JB):	0.808
Skew:	0.241	Prob(JB):	0.668
Kurtosis:	2.807	Cond. No.	3.24e+18

經過測試後，最好的regression model如上，有0.832的R-squared值，並且也沒有multicollinearity等問題

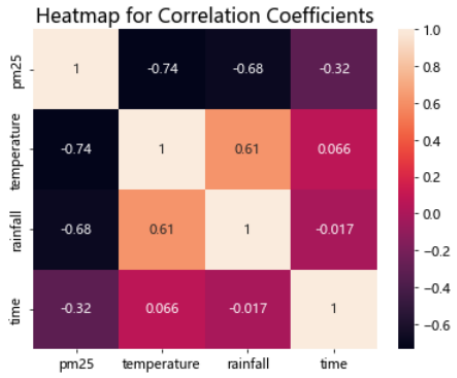
中區

分為兩測站群，其中降雨量使用「分區內被包含測站平均」
雨量測站：臺中、彰化、田中、日月潭、雲林、嘉義、阿里山、玉山
（共8個測站）
第一測站群：
斗六, 竹山, 嘉義, 新港, 崙背, 二林, 朴子, 南投（共8個測站）
對應的雨量測站：彰化、日月潭、雲林、嘉義

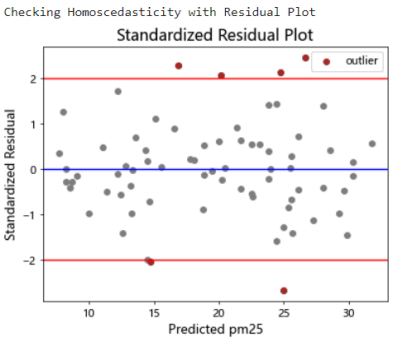
Multiple Rgression:



中部測站-圖一、二、三 (pm2.5對溫度、降雨量及時間的散佈圖)



中部測站-圖四 多元共線性檢測



	Chi2	p value
Value	1.472222	0.479

Runs Test for Standardized Residuals

	# >= Median	# < Median	Runs	Z stat	p value
Value	36	36	29	-1.899039	0.0576

中部測站-圖五、六、七 常態分佈、同質性、隨機性檢測

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.717
Model:	OLS	Adj. R-squared:	0.704
Method:	Least Squares	F-statistic:	57.30
Date:	Mon, 10 Jun 2024	Prob (F-statistic):	1.37e-18
Time:	20:03:07	Log-Likelihood:	-206.01
No. Observations:	72	AIC:	420.0
Df Residuals:	68	BIC:	429.1
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	49.7533	3.346	14.871	0.000	43.077	56.430
temperature	-0.8801	0.153	-5.752	0.000	-1.185	-0.575
rainfall	-0.4641	0.096	-4.855	0.000	-0.655	-0.273
time	-0.1137	0.025	-4.582	0.000	-0.163	-0.064

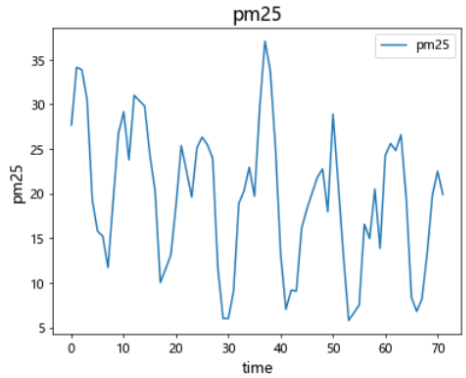
Omnibus:	1.185	Durbin-Watson:	1.208
Prob(Omnibus):	0.553	Jarque-Bera (JB):	0.600
Skew:	0.145	Prob(JB):	0.741
Kurtosis:	3.341	Cond. No.	311.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

中部測站-圖八 多變數回歸模型

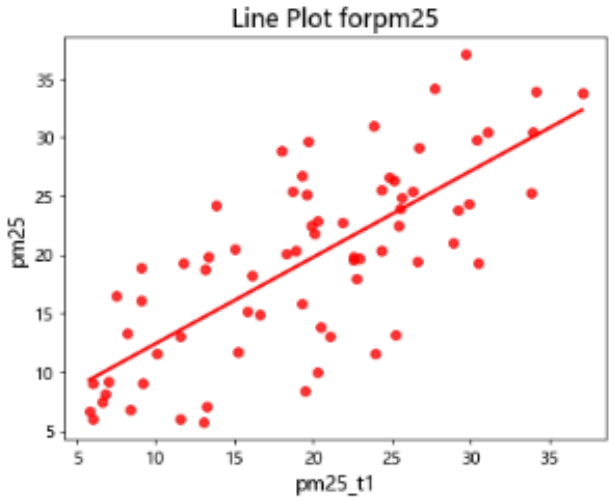
1. 解釋力為71.7%，且R-Square與adjusted R-Square之間的差值小於0.06，沒有過適問題
2. 通過F檢定，此模型有效
3. 各自變數皆通過t檢定

Time Series:



中部測站-圖九 第一測站群中Pm2.5每月數值折線圖

由敘述統計，可見此測站群中的季節性效應並不明顯，因此採用自回歸模型



OLS Regression Results

Dep. Variable:	pm25	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	1.849e+30
Date:	Tue, 11 Jun 2024	Prob (F-statistic):	0.00
Time:	09:29:00	Log-Likelihood:	2115.9
No. Observations:	71	AIC:	-4224.
Df Residuals:	67	BIC:	-4215.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-7.949e-14	1.34e-14	-5.924	0.000	-1.06e-13	-5.27e-14
pm25	1.0000	6.35e-16	1.57e+15	0.000	1.000	1.000
time	3.816e-16	1.75e-16	2.179	0.033	3.21e-17	7.31e-16
pm25_t1	1.277e-15	6.36e-16	2.008	0.049	7.41e-18	2.55e-15

Omnibus:	2.532	Durbin-Watson:	0.140
Prob(Omnibus):	0.282	Jarque-Bera (JB):	1.607
Skew:	0.096	Prob(JB):	0.448
Kurtosis:	2.289	Cond. No.	191.

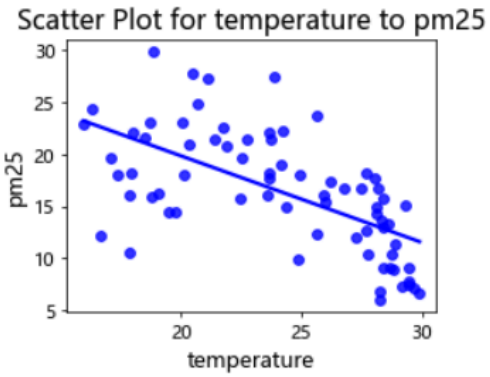
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

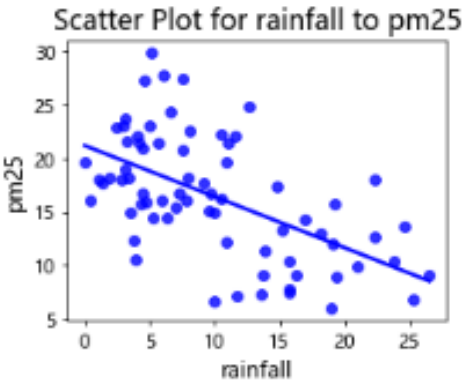
中部測站-圖十、十一 shift一個月後的資料和當月資料散佈圖及自回歸模型

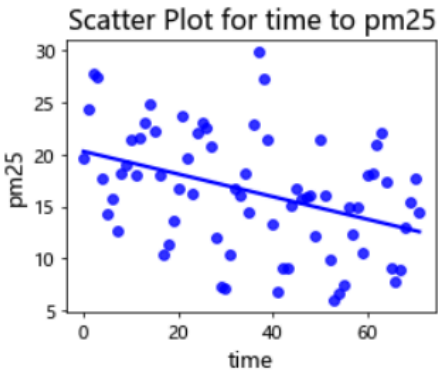
1. 解釋力為100%，且R-Square與adjusted R-Square之間的差值小於0.06，沒有過適問題
2. 通過F檢定，此模型有效
3. 各自變數皆通過t檢定

第二測站群：
西屯, 線西, 臺西, 埔里, 大里, 豐原（共6個測站）
對應的雨量測站：臺中、彰化

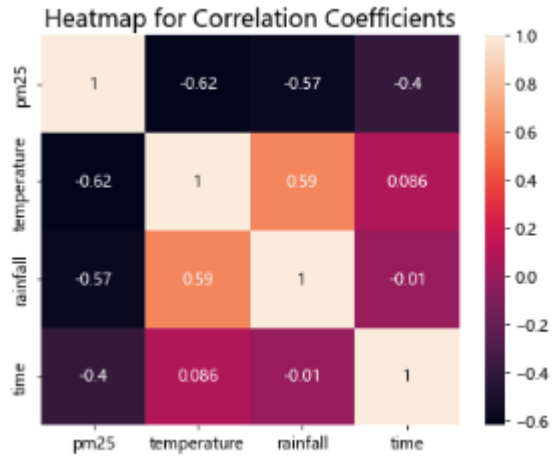
Multiple Rgression:





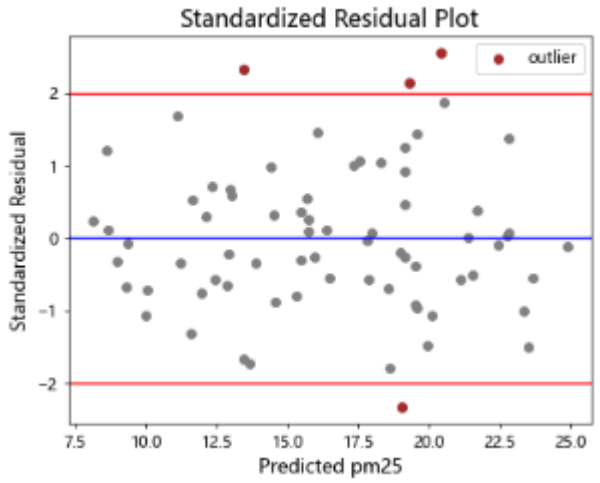


中部測站-圖十二、十三、十四 (pm2.5對溫度、降雨量及時間的散佈圖)



中部測站-圖十五 多元共線性檢測

Checking Homoscedasticity with Residual Plot



	Chi2	p value
Value	2.444444	0.2946

	# >= Median	# < Median	Runs	Z stat	p value
Value	36	36	29	-1.899039	0.0576

中部測站-圖十六、十七、十八 常態分佈、同質性、隨機性檢測

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.579
Model:	OLS	Adj. R-squared:	0.560
Method:	Least Squares	F-statistic:	31.18
Date:	Tue, 11 Jun 2024	Prob (F-statistic):	8.55e-13
Time:	13:33:30	Log-Likelihood:	-196.14
No. Observations:	72	AIC:	400.3
Df Residuals:	68	BIC:	409.4
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	35.2463	2.841	12.405	0.000	29.576	40.916
temperature	-0.5125	0.133	-3.865	0.000	-0.777	-0.248
rainfall	-0.2903	0.082	-3.534	0.001	-0.454	-0.126
time	-0.1007	0.022	-4.646	0.000	-0.144	-0.057

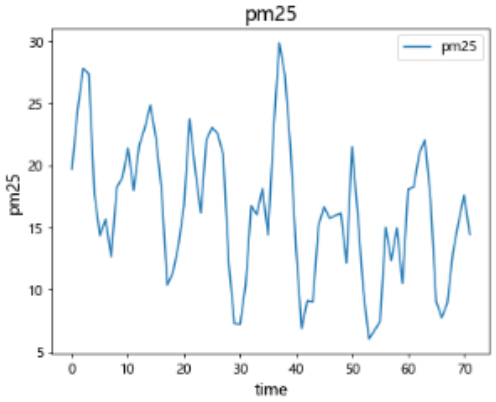
Omnibus:	1.687	Durbin-Watson:	1.145
Prob(Omnibus):	0.430	Jarque-Bera (JB):	1.390
Skew:	0.340	Prob(JB):	0.499
Kurtosis:	2.987	Cond. No.	302.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

中部測站-圖十九 多變數回歸模型

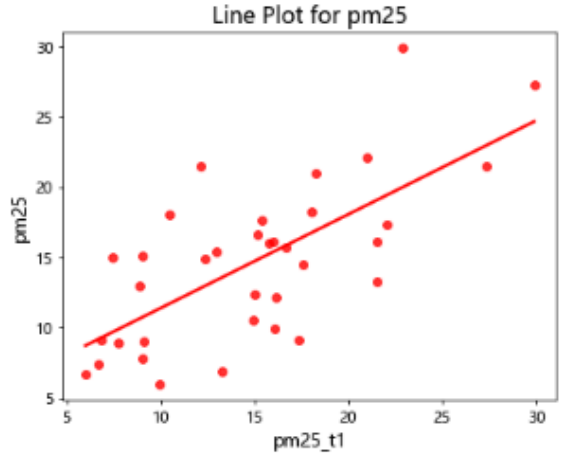
4. 解釋力為57.1%，且R-Square與adjusted R-Square之間的差值小於0.06，沒有過適問題
5. 通過F檢定，此模型有效
6. 各自變數皆通過t檢定

Time Series:



中部測站-圖二十 第二測站群中Pm2.5每月數值折線圖

由敘述統計，可見此測站群中的季節性效應並不明顯，因此採用自回歸模型



OLS Regression Results

Dep. Variable:	pm25	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	5.436e+30
Date:	Tue, 11 Jun 2024	Prob (F-statistic):	0.00
Time:	13:23:34	Log-Likelihood:	1087.4
No. Observations:	35	AIC:	-2167.
Df Residuals:	31	BIC:	-2161.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.132e-14	5.49e-15	2.061	0.048	1.17e-16	2.25e-14
pm25	1.0000	3.41e-16	2.93e+15	0.000	1.000	1.000
time	-3.053e-16	1.43e-16	-2.130	0.041	-5.98e-16	-1.3e-17
pm25_t1	9.992e-16	3.37e-16	2.969	0.006	3.13e-16	1.69e-15

Omnibus:	0.808	Durbin-Watson:	0.574
Prob(Omnibus):	0.668	Jarque-Bera (JB):	0.794
Skew:	-0.327	Prob(JB):	0.672
Kurtosis:	2.657	Cond. No.	113.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

中部測站-圖二十一、二十二 shift一個月後的資料和當月資料散佈圖及自回歸模型(取近四年)

7. 解釋力為100%，且R-Square與adjusted R-Square之間的差值小於0.06，沒有過適問題
8. 通過F檢定，此模型有效
9. 各自變數皆通過t檢定

南區

分為A、B、C三個測站群：

A：美濃[降雨量測站：高雄]

Multiple Regression:

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.777
Model:	OLS	Adj. R-squared:	0.768
Method:	Least Squares	F-statistic:	88.24
Date:	Tue, 28 May 2024	Prob (F-statistic):	1.08e-24
Time:	21:10:32	Log-Likelihood:	-224.32
No. Observations:	80	AIC:	456.6
Df Residuals:	76	BIC:	466.2
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	39.4915	3.070	12.864	0.000	33.377	45.606
wind	-6.8643	3.318	-2.069	0.042	-13.473	-0.255
rainfall$^{0.3}$	-2.9707	0.223	-13.300	0.000	-3.416	-2.526
time	-0.1423	0.021	-6.791	0.000	-0.184	-0.101

Omnibus:	2.926	Durbin-Watson:	1.522
Prob(Omnibus):	0.232	Jarque-Bera (JB):	2.894
Skew:	-0.443	Prob(JB):	0.235
Kurtosis:	2.711	Cond. No.	452.

空氣品質大致逐年變好
和風速、雨量呈負相關

Dummy Variables:

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.939
Model:	OLS	Adj. R-squared:	0.907
Method:	Least Squares	F-statistic:	29.42
Date:	Tue, 28 May 2024	Prob (F-statistic):	4.32e-11
Time:	21:16:02	Log-Likelihood:	-72.092
No. Observations:	36	AIC:	170.2
Df Residuals:	23	BIC:	190.8
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	21.9051	1.564	14.006	0.000	18.670	25.141
Time	-0.0802	0.038	-2.102	0.047	-0.159	-0.001
Season_1	4.6162	1.879	2.457	0.022	0.730	8.502
Season_2	2.8198	1.870	1.508	0.145	-1.049	6.689
Season_3	2.0174	1.863	1.083	0.290	-1.836	5.871
Season_4	-1.0135	1.856	-0.546	0.590	-4.854	2.827
Season_5	-8.7727	1.850	-4.741	0.000	-12.601	-4.945
Season_6	-14.7620	1.845	-8.000	0.000	-18.579	-10.945
Season_7	-14.5672	1.841	-7.913	0.000	-18.376	-10.759
Season_8	-13.9892	1.837	-7.614	0.000	-17.790	-10.188
Season_9	-8.7727	1.835	-4.782	0.000	-12.568	-4.978
Season_10	-4.0078	1.833	-2.187	0.039	-7.799	-0.217
Season_11	0.5383	1.831	0.294	0.771	-3.250	4.327

Omnibus:	0.402	Durbin-Watson:	1.159
Prob(Omnibus):	0.818	Jarque-Bera (JB):	0.552
Skew:	0.191	Prob(JB):	0.759
Kurtosis:	2.529	Cond. No.	259.

空氣品質大致逐年變好

夏天可能被西南季風影響，使空氣品質特別好

B：恆春[降雨量測站：恆春]
Multiple Regression:

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.586
Model:	OLS	Adj. R-squared:	0.564
Method:	Least Squares	F-statistic:	26.47
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	8.60e-11
Time:	16:59:17	Log-Likelihood:	-116.73
No. Observations:	60	AIC:	241.5
Df Residuals:	56	BIC:	249.8
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	10.2431	1.389	7.377	0.000	7.461	13.025
wind	0.1375	0.172	0.800	0.427	-0.207	0.482
rainfall$^{0.3}$	-0.4560	0.148	-3.087	0.003	-0.752	-0.160
time	-0.1027	0.013	-7.838	0.000	-0.129	-0.076

Omnibus:	0.374	Durbin-Watson:	1.479
Prob(Omnibus):	0.830	Jarque-Bera (JB):	0.163
Skew:	0.127	Prob(JB):	0.922
Kurtosis:	3.017	Cond. No.	215.

空氣品質大致逐年變好
和風速、雨量呈負相關

CMA:

OLS Regression Results

Dep. Variable:	CMA_deseasoned	R-squared:	0.543
Model:	OLS	Adj. R-squared:	0.535
Method:	Least Squares	F-statistic:	68.81
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	1.96e-11
Time:	17:22:27	Log-Likelihood:	-112.02
No. Observations:	60	AIC:	228.0
Df Residuals:	58	BIC:	232.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	9.2582	0.406	22.806	0.000	8.446	10.071
Time	-0.0984	0.012	-8.295	0.000	-0.122	-0.075

Omnibus:	2.680	Durbin-Watson:	1.294
Prob(Omnibus):	0.262	Jarque-Bera (JB):	2.591
Skew:	0.484	Prob(JB):	0.274
Kurtosis:	2.683	Cond. No.	67.6

空氣品質大致逐年變好

C：大寮、橋頭、仁武、鳳山、屏東、復興、林園、潮州、前鎮、左營、臺南、安南、新營、善化[降雨量測站：高雄、臺南、屏東]
Multiple Regression:

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.825
Model:	OLS	Adj. R-squared:	0.819
Method:	Least Squares	F-statistic:	134.4
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	2.65e-22
Time:	21:52:43	Log-Likelihood:	-164.94
No. Observations:	60	AIC:	335.9
Df Residuals:	57	BIC:	342.2
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	31.1978	1.108	28.148	0.000	28.978	33.417
temperature	-3.928e-14	2.46e-15	-15.975	0.000	-4.42e-14	-3.44e-14
time	-0.0701	0.029	-2.416	0.019	-0.128	-0.012

Omnibus:	0.463	Durbin-Watson:	1.631
Prob(Omnibus):	0.793	Jarque-Bera (JB):	0.372
Skew:	0.187	Prob(JB):	0.830
Kurtosis:	2.904	Cond. No.	6.82e+14

空氣品質大致逐年變好
和溫度呈負相關

Dummy Variables:

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.921
Model:	OLS	Adj. R-squared:	0.901
Method:	Least Squares	F-statistic:	45.90
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	8.10e-22
Time:	16:24:37	Log-Likelihood:	-140.94
No. Observations:	60	AIC:	307.9
Df Residuals:	47	BIC:	335.1
Df Model:	12		
Covariance Type:	nonrobust		

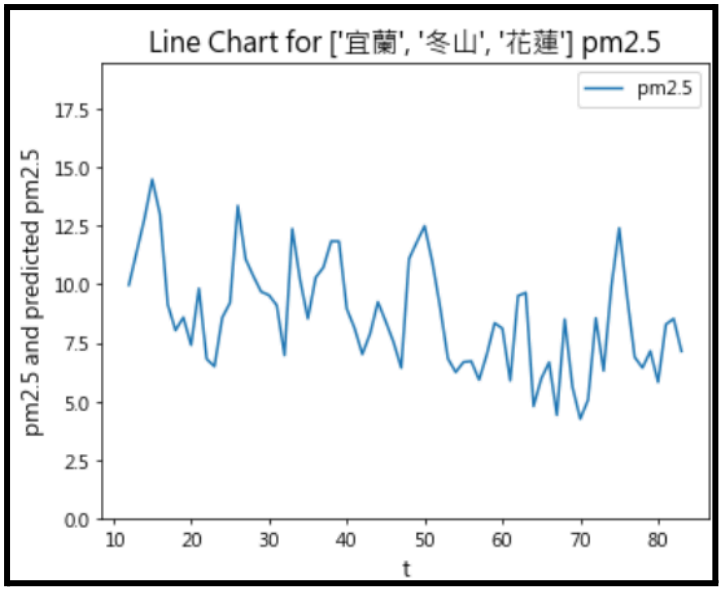
	coef	std err	t	P>|t|	[0.025	0.975]
const	28.6162	1.491	19.198	0.000	25.618	31.615
Time	-0.0808	0.022	-3.709	0.001	-0.125	-0.037
Season_1	0.1000	1.827	5.575	0.001	2.491	9.842
Season_2	2.9253	1.824	1.604	0.116	-0.745	6.595
Season_3	1.6199	1.822	0.889	0.378	-2.045	5.285
Season_4	-3.2738	1.820	-1.799	0.078	-6.934	0.387
Season_5	-12.0716	1.818	-6.641	0.000	-15.728	-8.415
Season_6	-19.2273	1.816	-10.588	0.000	-22.881	-15.574
Season_7	-18.5735	1.815	-10.236	0.000	-22.224	-14.923
Season_8	-16.6747	1.813	-9.195	0.000	-20.323	-13.027
Season_9	-10.1270	1.812	-5.587	0.000	-13.773	-6.481
Season_10	-4.5215	1.812	-2.496	0.016	-8.166	-0.877
Season_11	-0.2496	1.811	-0.138	0.891	-3.894	3.395

Omnibus:	1.668	Durbin-Watson:	1.219
Prob(Omnibus):	0.434	Jarque-Bera (JB):	1.235
Skew:	0.350	Prob(JB):	0.539
Kurtosis:	3.060	Cond. No.	430.

空氣品質大致逐年變好
夏天可能被西南季風影響，使空氣品質特別好

東區

A: 宜蘭、冬山、花蓮 **[**雨量測站**:** 宜蘭、花蓮、成功**]**
Line Chart



Auto Regression

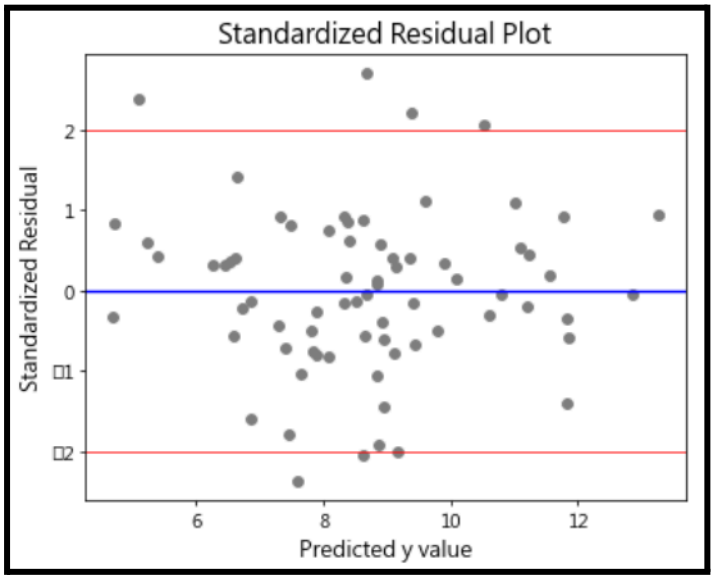
OLS Regression Results

Dep. Variable:	y	R-squared:	0.662
Model:	OLS	Adj. R-squared:	0.636
Method:	Least Squares	F-statistic:	25.82
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	2.44e-14
Time:	21:38:19	Log-Likelihood:	-123.00
No. Observations:	72	AIC:	258.0
Df Residuals:	66	BIC:	271.7
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	20.8198	4.726	4.406	0.000	11.385	30.255
pm25_1	0.4207	0.077	5.432	0.000	0.266	0.575
pm25_12	0.2298	0.070	3.306	0.002	0.091	0.369
temperature	-0.1387	0.047	-2.970	0.004	-0.232	-0.045
rainfall	-0.0032	0.001	-2.128	0.037	-0.006	-0.000
rh	-0.1794	0.051	-3.498	0.001	-0.282	-0.077

Omnibus:	1.805	Durbin-Watson:	2.181
Prob(Omnibus):	0.405	Jarque-Bera (JB):	1.136
Skew:	0.145	Prob(JB):	0.567
Kurtosis:	3.543	Cond. No.	6.45e+03

Variation Test

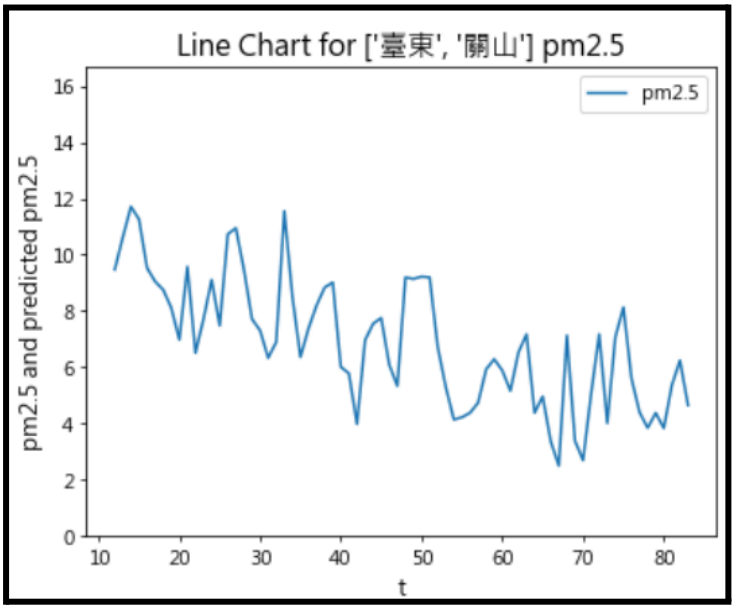


Normality Test and Run Test

```
Shapiro Test
Statistics=0.979, p=0.265
runs =  41
n1 =  36
n2 =  37
runs_exp = 37.493151
stan_dev = 4.241424
z = 0.826809
pval_z = 0.408345
```

B: 臺東、關山 [雨量測站:臺東、大武、蘭嶼]

Line Chart



Auto Regresstion

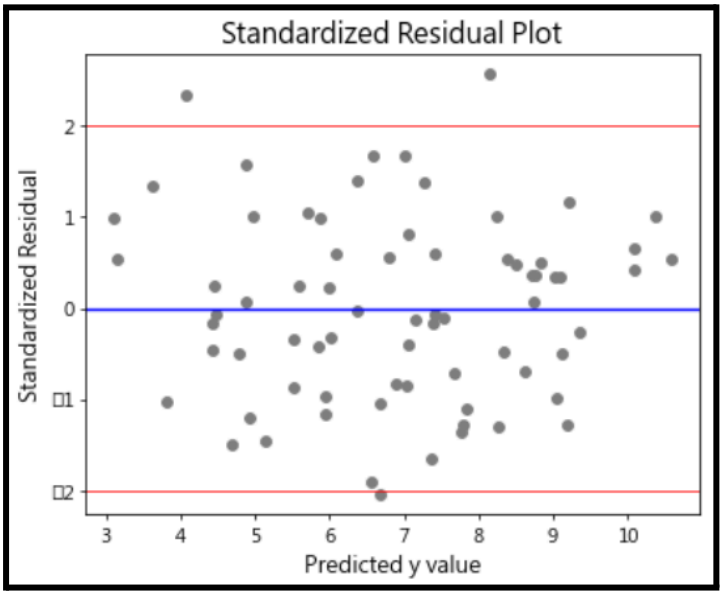
OLS Regression Results

Dep. Variable:	y	R-squared:	0.657
Model:	OLS	Adj. R-squared:	0.637
Method:	Least Squares	F-statistic:	32.14
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	6.01e-15
Time:	21:47:24	Log-Likelihood:	-121.70
No. Observations:	72	AIC:	253.4
Df Residuals:	67	BIC:	264.8
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.0068	1.052	0.957	0.342	-1.094	3.107
pm25_1	0.3716	0.087	4.266	0.000	0.198	0.545
pm25_12	0.3746	0.089	4.194	0.000	0.196	0.553
rainfall	-0.0036	0.001	-2.697	0.009	-0.006	-0.001
wind	0.8853	0.873	1.014	0.314	-0.857	2.627

Omnibus:	0.733	Durbin-Watson:	1.882
Prob(Omnibus):	0.693	Jarque-Bera (JB):	0.853
Skew:	0.181	Prob(JB):	0.653
Kurtosis:	2.609	Cond. No.	1.71e+03

Variation Test

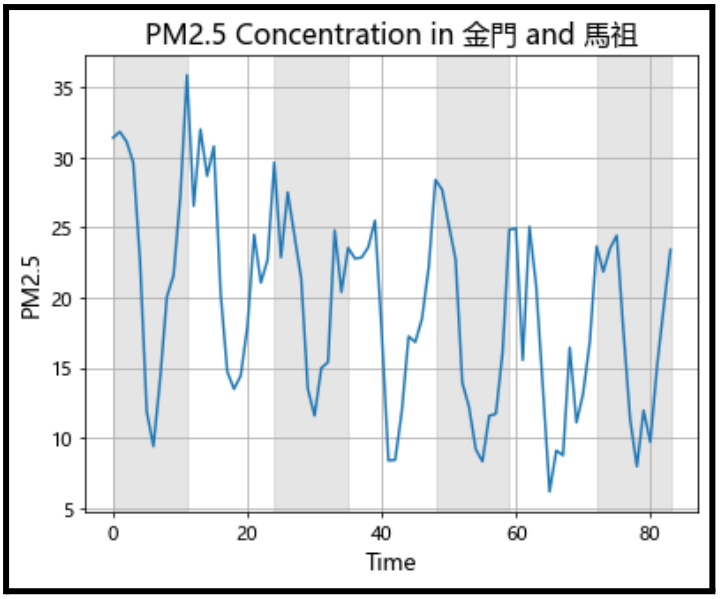


Normality Test and Run Test

```
Shapiro Test
Statistics=0.988, p=0.699
runs =  31
n1 =  37
n2 =  36
runs_exp = 37.493151
stan_dev = 4.241424
z = -1.530889
pval_z = 0.125797
```

離島

將金門、馬祖兩測站之觀測值平均。



Multiple Regression

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.825
Model:	OLS	Adj. R-squared:	0.816
Method:	Least Squares	F-statistic:	93.03
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	4.34e-29
Time:	17:34:17	Log-Likelihood:	-209.85
No. Observations:	84	AIC:	429.7
Df Residuals:	79	BIC:	441.9
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	51.4607	2.246	22.909	0.000	46.989	55.932
temperature	-0.7352	0.064	-11.491	0.000	-0.863	-0.608
wind	-3.5455	0.812	-4.368	0.000	-5.161	-1.930
rainfall	-0.0216	0.005	-4.243	0.000	-0.032	-0.011
time	-0.1523	0.017	-8.834	0.000	-0.187	-0.118

Omnibus:	2.140	Durbin-Watson:	1.639
Prob(Omnibus):	0.343	Jarque-Bera (JB):	1.646
Skew:	0.071	Prob(JB):	0.439
Kurtosis:	3.671	Cond. No.	767.

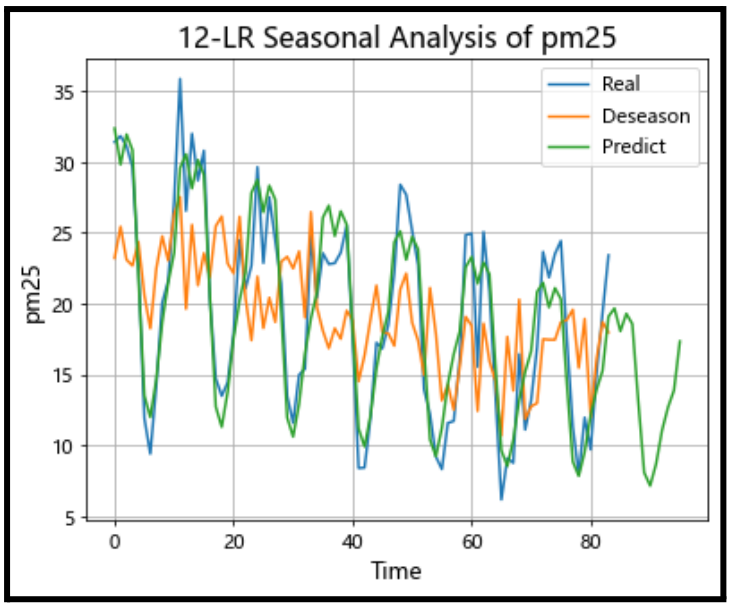
Seasonal Index - Linear Regression

OLS Regression Results

Dep. Variable:	LR_deseasoned	R-squared:	0.489
Model:	OLS	Adj. R-squared:	0.483
Method:	Least Squares	F-statistic:	78.56
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	1.34e-13
Time:	17:42:21	Log-Likelihood:	-204.74
No. Observations:	84	AIC:	413.5
Df Residuals:	82	BIC:	418.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	23.9390	0.606	39.493	0.000	22.733	25.145
Time	-0.1118	0.013	-8.864	0.000	-0.137	-0.087

Omnibus:	4.706	Durbin-Watson:	1.632
Prob(Omnibus):	0.095	Jarque-Bera (JB):	2.279
Skew:	-0.044	Prob(JB):	0.320
Kurtosis:	2.198	Cond. No.	95.3

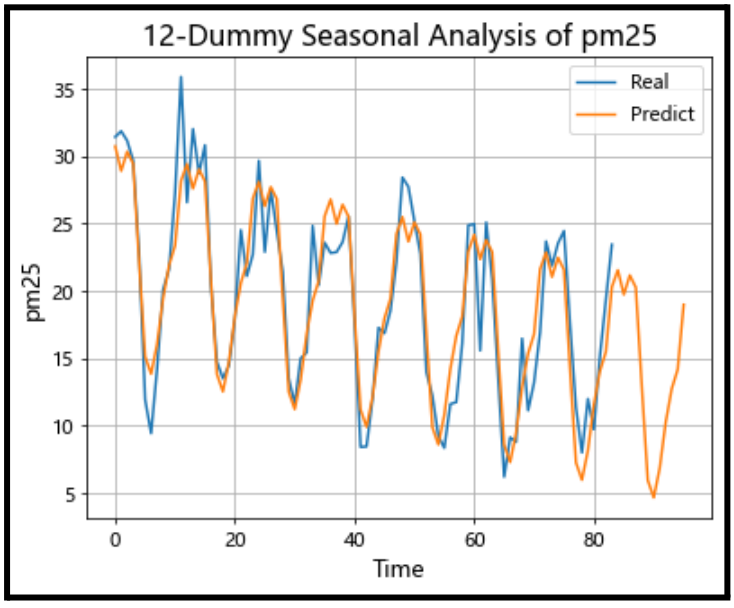


Seasonal Index - Dummy Variables

OLS Regression Results

Dep. Variable:	pm25	R-squared:	0.855
Model:	OLS	Adj. R-squared:	0.836
Method:	Least Squares	F-statistic:	34.79
Date:	Sun, 09 Jun 2024	Prob (F-statistic):	6.05e-25
Time:	17:43:11	Log-Likelihood:	-202.03
No. Observations:	84	AIC:	430.1
Df Residuals:	71	BIC:	461.7
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	29.3508	1.266	23.182	0.000	26.826	31.875
Time	-0.1093	0.013	-8.245	0.000	-0.136	-0.083
Season_1	1.3588	1.566	0.868	0.388	-1.763	4.480
Season_2	-0.3511	1.564	-0.224	0.823	-3.470	2.768
Season_3	1.1897	1.563	0.761	0.449	-1.927	4.307
Season_4	0.4017	1.562	0.257	0.798	-2.713	3.517
Season_5	-6.8057	1.561	-4.359	0.000	-9.919	-3.692
Season_6	-13.6893	1.561	-8.771	0.000	-16.801	-10.577
Season_7	-14.8680	1.560	-9.530	0.000	-17.979	-11.757
Season_8	-12.5606	1.560	-8.054	0.000	-15.670	-9.451
Season_9	-9.0523	1.559	-5.806	0.000	-12.161	-5.943
Season_10	-6.4769	1.559	-4.155	0.000	-9.585	-3.368
Season_11	-4.9194	1.559	-3.156	0.002	-8.027	-1.811



七、研究發現

- 空氣品質(PM2.5)在選定測站大致上有逐年變好的趨勢
- 冬天的空氣品質(PM2.5)相對較差
- 空氣品質(PM2.5)在不同月份之間有規律的變化
- 自變數重要性: Time, Rainfall > Temperature > Wind
- 各地區空氣品質(PM2.5)具解釋力的影響變因不同
- 還有其他沒被納入的重要變因，例如：相對濕度、其他化學汙染物

八、結論與建議

隨著企業與政府的環保意識增強，空氣品質雖有改善，但仍需持續監測與加強措施。冬季空氣品質較差，建議大家注意防護，並根據季節變化加強排放管控和促進空氣流通。制定全面的空氣品質改善措施應考慮時間、降雨、溫度、風速、相對濕度和其他化學污染物等多種因素。通過加強排放管控、促進空氣流通、提升公眾意識，我們能確保更清新的空氣和健康的生活環境。

九、資料來源

環境部監測資訊司

https://data.moenv.gov.tw/dataset/detail/AQX_P_19