

Case Study - Regressione Lineare

Anno accademico 2020-21

Gruppo 19 - LISBONA: Domenico Gaeni e Palazzi Fabio





Obiettivo





Strategia

- Analisi dei grafici dei singoli regressori;
- **Backward Stepwise Regression:**
 - Si considerano tutti i regressori;
 - Ad ogni step si eliminano uno alla volta i regressori non significativi, con p-value > 0.10;
 - Il modello finale sarà composto da tutti regressori significativi.

Analisi dei singoli regressori

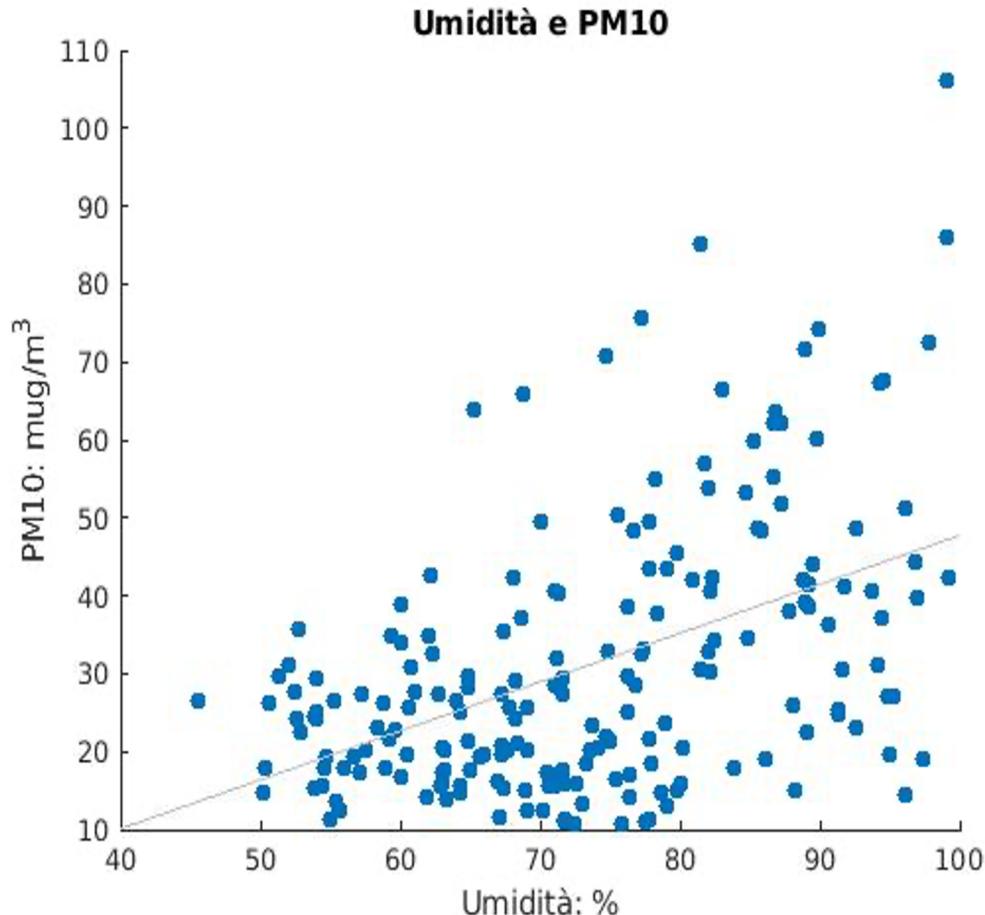


Analisi

Grafici della variabile risposta y (PM10) in funzione di ogni regressore nella stazione di Mantova.

Indice di correlazione r per ogni regressore:

- Umidità **0.48**
- Temperatura **-0.61**
- Pioggia **-0.31**
- Ozono **-0.64**
- Ossidi di Azoto **0.76**
- Biossido di Azoto **0.78**



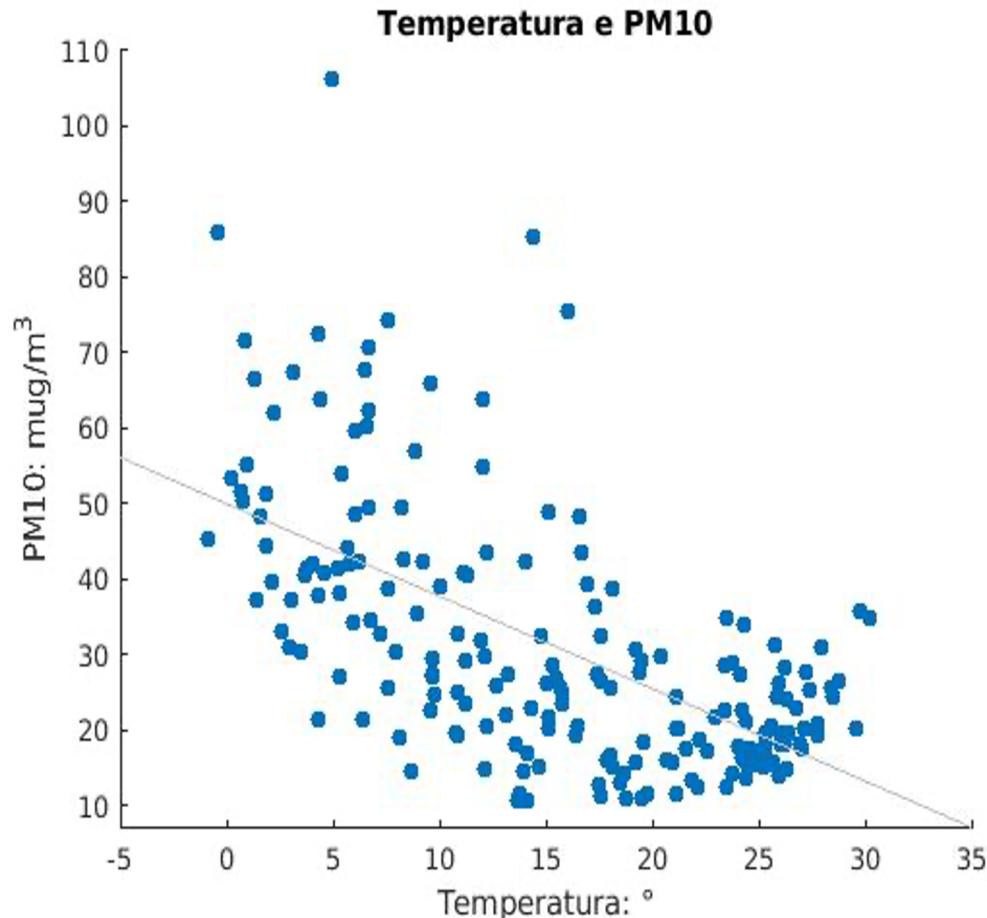


Analisi

Grafici della variabile risposta y (PM10) in funzione di ogni regressore nella stazione di Mantova.

Indice di correlazione r per ogni regressore:

- Umidità **0.48**
- Temperatura **-0.61**
- Pioggia **-0.31**
- Ozono **-0.64**
- Ossidi di Azoto **0.76**
- Biossido di Azoto **0.78**





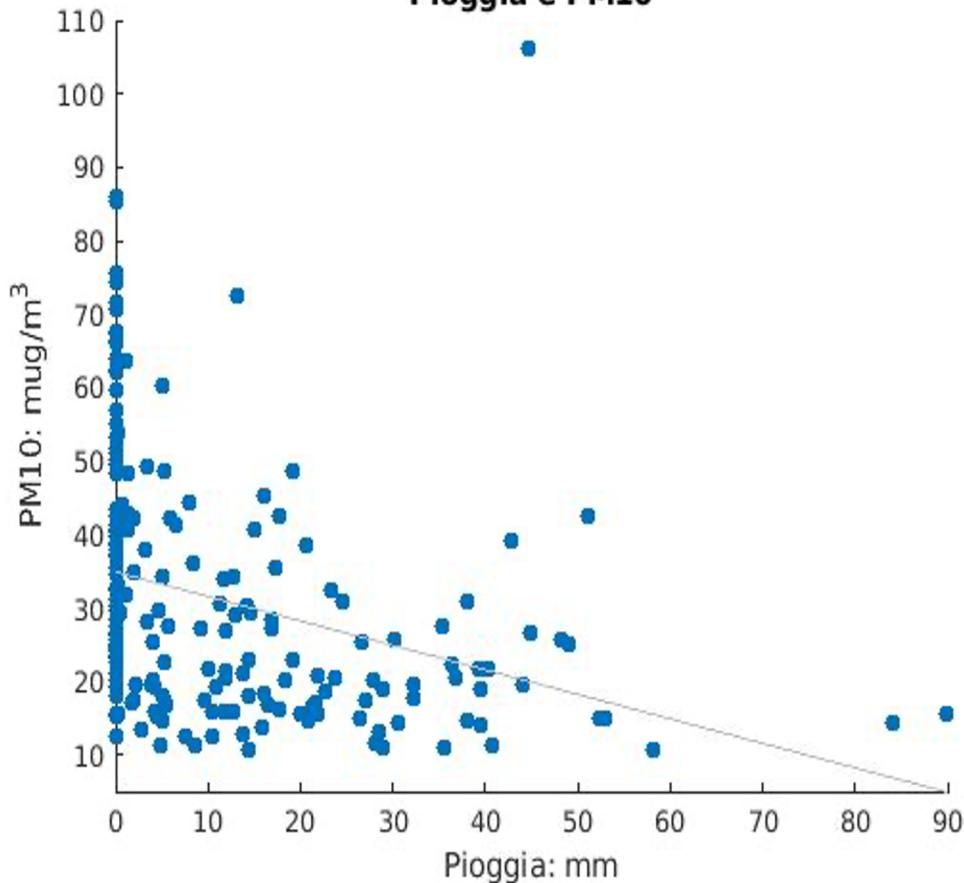
Analisi

Grafici della variabile risposta y (PM10) in funzione di ogni regressore nella stazione di Mantova.

Indice di correlazione r per ogni regressore:

- Umidità **0.48**
- Temperatura **-0.61**
- **Pioggia -0.31**
- Ozono **-0.64**
- Ossidi di Azoto **0.76**
- Biossido di Azoto **0.78**

Pioggia e PM10





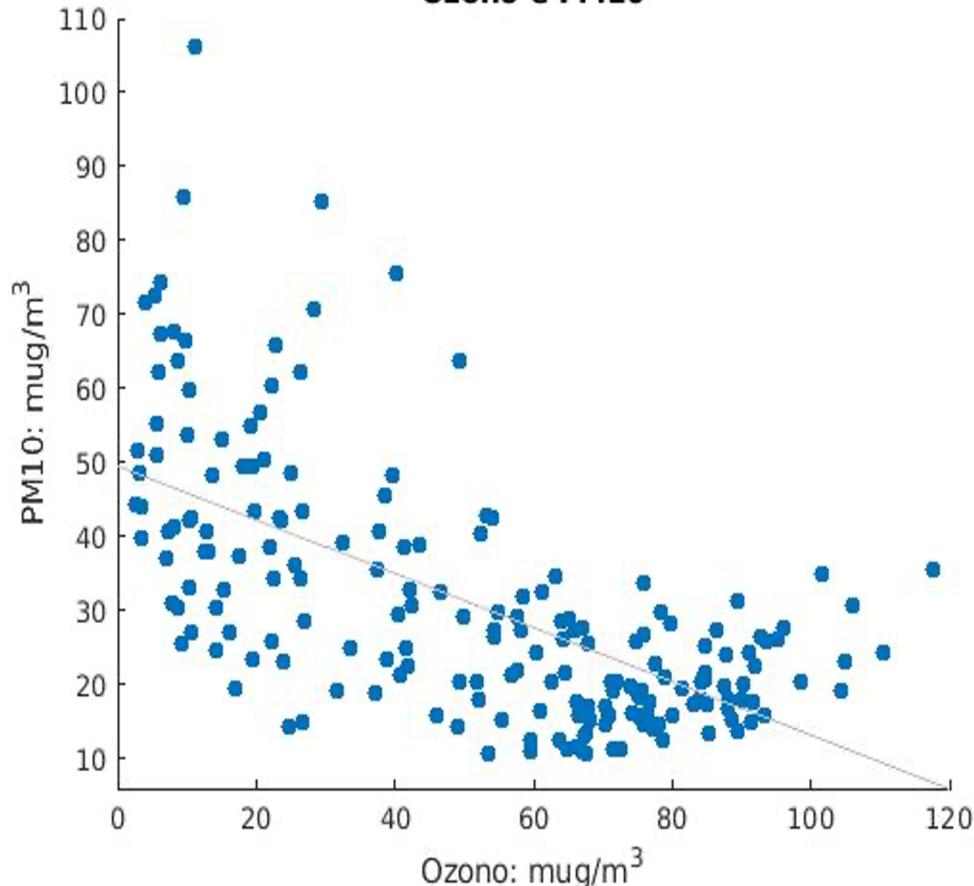
Analisi

Grafici della variabile risposta y (PM10) in funzione di ogni regressore nella stazione di Mantova.

Indice di correlazione r per ogni regressore:

- Umidità **0.48**
- Temperatura **-0.61**
- Pioggia **-0.31**
- Ozono **-0.64**
- Ossidi di Azoto **0.76**
- Biossido di Azoto **0.78**

Ozono e PM10



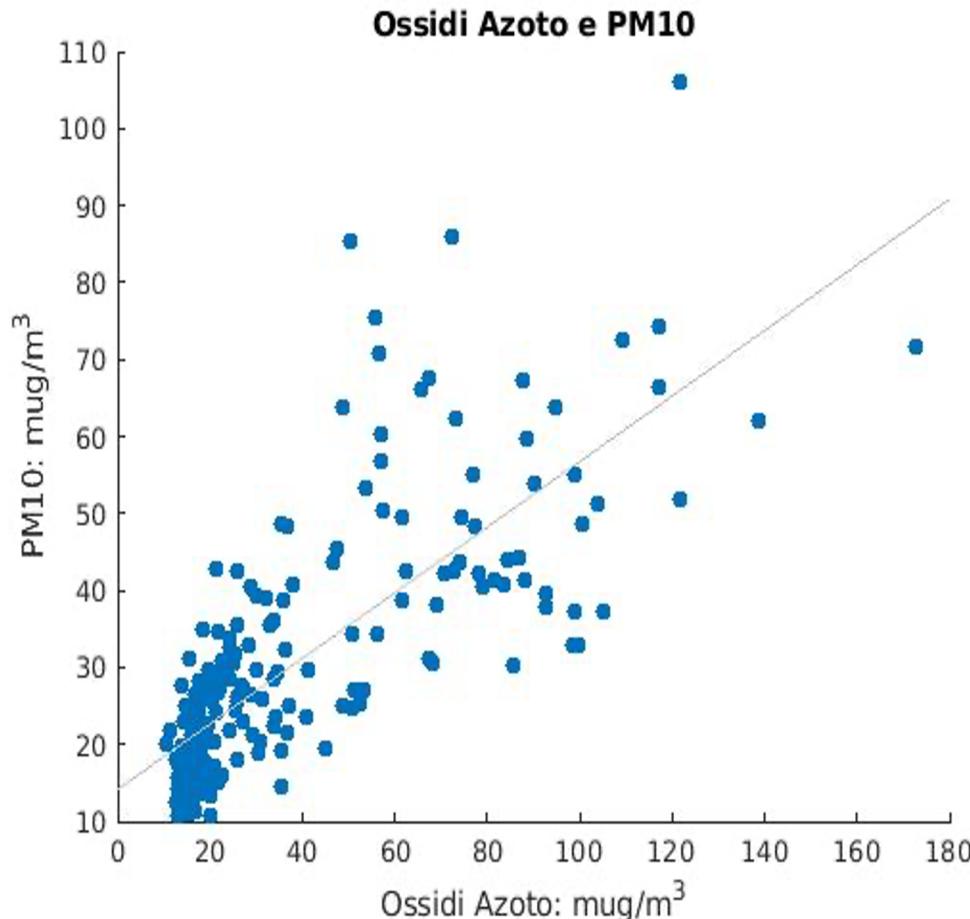


Analisi

Grafici della variabile risposta y (PM10) in funzione di ogni regressore nella stazione di Mantova.

Indice di correlazione r per ogni regressore:

- Umidità **0.48**
- Temperatura **-0.61**
- Pioggia **-0.31**
- Ozono **-0.64**
- **Ossidi di Azoto 0.76**
- Biossido di Azoto **0.78**



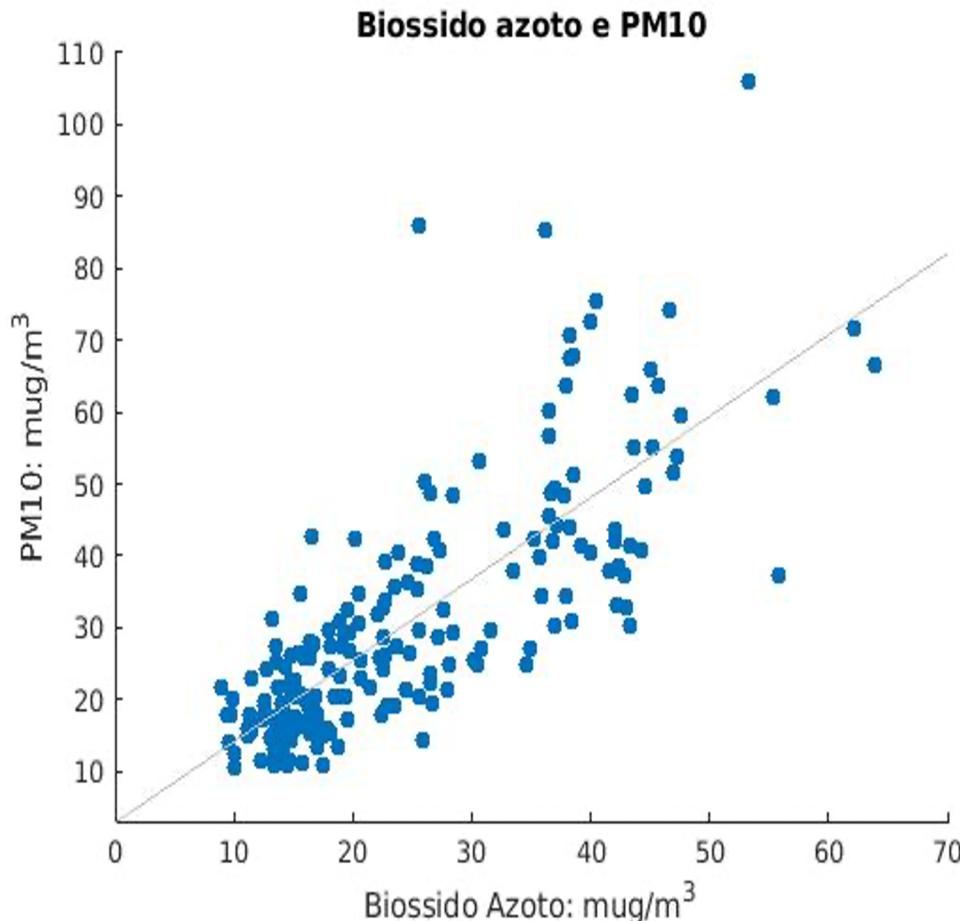


Analisi

Grafici della variabile risposta y (PM10) in funzione di ogni regressore nella stazione di Mantova.

Indice di correlazione r per ogni regressore:

- Umidità **0.48**
- Temperatura **-0.61**
- Pioggia **-0.31**
- Ozono **-0.64**
- Ossidi di Azoto **0.76**
- Biassido di Azoto **0.78**



Backward Stepwise Regression



X Stazione di Mantova

Partiamo con il considerare tutti i regressori per estrapolare un modello per il PM10.

I regressori **non significativi** sono:

- Temperatura
- Ozono
- Ossidi di Azoto

Adjusted R-Squared = 0.626

Linear regression model:

PM10 ~ 1 + Temp + Pioggia + Umid + Ozono + Ossidi_Azoto + Biossido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-16.274	13.253	-1.228	0.22097
Temp	0.083453	0.21606	0.38626	0.69974
Pioggia	-0.15412	0.056269	-2.7389	0.0067508
Umid	0.22491	0.126	1.7851	0.075845
Ozono	0.09185	0.088699	1.0355	0.30174
Ossidi_Azoto	0.10977	0.074532	1.4728	0.14245
Biossido_azoto	0.90085	0.20076	4.4872	1.2482e-05



✗ Stazione di Mantova

Eliminiamo la temperatura.

I regressori **non significativi** sono:

- Ossidi di Azoto
- Ozono

Adjusted R-Squared = 0.628

Linear regression model:

PM10 ~ 1 + Pioggia + Umid + Ozono + Ossidi_Azoto + Biossido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-16.543	13.205	-1.2528	0.2118
Pioggia	-0.15497	0.0561	-2.7624	0.0062981
Umid	0.23332	0.12382	1.8843	0.061046
Ozono	0.11337	0.06887	1.6461	0.10139
Ossidi_Azoto	0.10743	0.074118	1.4494	0.14887
Biossido_azoto	0.89846	0.20022	4.4875	1.2435e-05



✗ Stazione di Mantova

Eliminiamo gli ossidi di azoto.

I regressori **non significativi** sono:

- Ozono

Adjusted R-Squared = 0.626

Linear regression model:

PM10 ~ 1 + Pioggia + Umid + Ozono + Biossalido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-20.764	12.916	-1.6076	0.10958
Pioggia	-0.166	0.055741	-2.978	0.0032753
Umid	0.27357	0.12102	2.2606	0.024905
Ozono	0.11304	0.069067	1.6367	0.10333
Biossalido_azoto	1.1266	0.12409	9.079	1.3296e-16



✗ Stazione di Mantova

Eliminiamo l'ozono.

I regressori **non significativi** sono:

- *Umidità*

Adjusted R-Squared = 0.623

Linear regression model:

PM10 ~ 1 + Pioggia + Umid + Biossido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-0.99931	4.6036	-0.21707	0.82838
Pioggia	-0.15192	0.055312	-2.7465	0.0065928
Umid	0.1273	0.081952	1.5533	0.12198
Biossido_azoto	0.98803	0.091113	10.844	1.0613e-21



✓ Stazione di Mantova

Eliminando l'**umidità** si arriva ad un modello con tutti i regressori significativi.

Adjusted R-Squared = 0.62

Linear regression model:

PM10 ~ 1 + Pioggia + Biossalido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	5.4291	2.024	2.6823	0.0079414
Pioggia	-0.11064	0.048688	-2.2725	0.024155
Biossalido_azoto	1.0861	0.065941	16.47	1.0519e-38



✓ Stazione di Mantova

Se si introduce di nuovo gli **ossidi di azoto** visto il forte indice di correlazione, questo diventa significativo e aumenta l'r quadro.

Adjusted R-Squared = 0.625

Linear regression model:

PM10 ~ 1 + Pioggia + Ossidi_Azoto + Biossalido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	8.1488	2.4834	3.2814	0.0012261
Pioggia	-0.11324	0.048399	-2.3396	0.020326
Ossidi_Azoto	0.13139	0.070381	1.8669	0.063434
Biossalido_azoto	0.76824	0.18243	4.2113	3.8922e-05



✓ Stazione di Bergamo

Eseguendo lo stesso algoritmo per la stazione di Bergamo si trova il seguente modello.

Adjusted R-Squared = 0.663

Linear regression model:

PM10 ~ 1 + Pioggia + Ossidi_azoto + Biossido_azoto

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	12.842	1.9418	6.6134	3.5942e-10
Pioggia	-0.12328	0.024133	-5.1084	7.7712e-07
Ossidi_azoto	0.10715	0.039039	2.7448	0.0066266
Biossido_azoto	0.47129	0.1141	4.1304	5.3894e-05

Osservazioni e conclusioni



Stima dei coefficienti

📍 Mantova

Estimated Coefficients:

Estimate

(Intercept)	8.1488
Pioggia	-0.11324
Ossidi_Azoto	0.13139
Biossido_azoto	0.76824

- ✓ Presentano lo **stesso segno**, in accordo con il coefficiente di correlazione;
- ✓ In generale sono **simili** e differiscono di alcuni centesimi (*salvo biossido di azoto*)

📍 Bergamo

Estimated Coefficients:

Estimate

(Intercept)	12.842
Pioggia	-0.12328
Ossidi_azoto	0.10715
Biossido_azoto	0.47129



Regressori significativi

📍 Mantova

tStat	pValue
3.2814	0.0012261
-2.3396	0.020326
1.8669	0.063434
4.2113	3.8922e-05

- ✓ I regressori sono tutti **significativi**, p-value <<0.10
- ✓ **Coerente** con la strategia utilizzata.

📍 Bergamo

tStat	pValue
6.6134	3.5942e-10
-5.1084	7.7712e-07
2.7448	0.0066266
4.1304	5.3894e-05



Performances di adattamento

📍 Mantova

62,5 %

- ✓ Differiscono di poco;
- ✓ Percentuale che la variabilità complessiva di PM10 è spiegata dalla relazione lineare con la **pioggia, ossidi di azoto, biossido di azoto.**

📍 Bergamo

66,3 %

Verifica dei risultati



✓ Verifica con il comando: stepwisefit

📍 Mantova

{'Coeff' }	{'Std.Err.'}	{'Status'}	{'P' }	
{[-0.1106]}	{[0.0487]}	{'In'}	{[0.0242]}	Pioggia
{[0.1273]}	{[0.0820]}	{'Out'}	{[0.1220]}	Umidità
{[-0.0023]}	{[0.0471]}	{'Out'}	{[0.9618]}	Ozono
{[0.0607]}	{[0.1516]}	{'Out'}	{[0.6894]}	Temperatura
{[1.0861]}	{[0.0659]}	{'In'}	{[1.0519e-38]}	NO2
{[0.1314]}	{[0.0704]}	{'Out'}	{[0.0634]}	NOX

📍 Bergamo

{'Coeff' }	{'Std.Err.'}	{'Status'}	{'P' }	
{[-0.1233]}	{[0.0241]}	{'In'}	{[7.7712e-07]}	Pioggia
{[0.0661]}	{[0.0731]}	{'Out'}	{[0.3668]}	Umidità
{[-0.0125]}	{[0.0343]}	{'Out'}	{[0.7168]}	Ozono
{[-0.0262]}	{[0.1417]}	{'Out'}	{[0.8533]}	Temperatura
{[0.4713]}	{[0.1141]}	{'In'}	{[5.3894e-05]}	NO2
{[0.1072]}	{[0.0390]}	{'In'}	{[0.0066]}	NOX