Case study sulla regressione lineare multipla

Gruppo n.21 ‘Lubjana’

Mohammed Amin Borqal	**1073928**
Loris Iacoban	**1074130**
Andrea Moressa	**1074124**

Origine dei dati

I dati provenienti dal sito dell'ARPA Lombardia e contenuti nel file "**G21.mat**" riguardanti una stazione di rilevazione in provincia di LeccoSono stati caricati in Matlab

	1 Anno	2 Mese	3 Prodotto	4 Agricoltura	5 Aviazione	6 Aziende_elettriche	7 Bunker_marina	8 Consumatori_finali	9 Ferrovie	10 Forze_armate	11 Industria	12 Merce_SAC	13 Piccola_marina	14 Rete	15 Rivenditori
1	2014	1	G.P.L. Com...	1854	0	0	0	89616	0	364	7132	1115	0	0	77728
2	2014	1	G.P.L. Autot...	0	0	0	0	4303	0	0	0	1003	0	65279	60955
3	2014	1	Virgin Nafta	0	0	0	0	0	0	0	2475	0	0	0	0
4	2014	1	Benzina sen...	45	6	0	0	13444	0	0	966	1726	0	480686	104624
5	2014	1	Benzina jetf...	0	95	0	0	0	0	0	0	0	0	0	0
6	2014	1	Benzina avio	0	44	0	0	0	0	51	0	0	0	0	0
7	2014	1	Benz. Altri usi	0	0	0	0	0	0	0	1322	0	0	0	0
8	2014	1	Petrolio risc...	0	0	0	0	11	0	0	3	0	0	0	648
9	2014	1	Carboturbo ...	0	245282	0	0	0	0	16239	0	0	0	0	97
10	2014	1	Petrolio altri...	0	0	0	0	6	0	0	0	342	0	0	2
11	2014	1	Gasolio mot...	45074	0	1531	24487	73110	1733	6564	7026	29178	24687	1008535	638783
12	2014	1	Gasolio risc...	7	2431	1287	4203	13398	42	2828	803	3343	80	0	148976
13	2014	1	Gasolio uso ...	0	0	5257	0	0	0	0	0	447	437	0	819
14	2014	1	O.C. ATZ	0	0	68085	98376	0	0	0	0	0	0	0	162
15	2014	1	O.C. BTZ	0	0	19012	40381	2769	0	24	2126	7026	0	0	37885
16	2014	1	Lubrif. Motori	435	18	121	3021	5120	19	45	1886	459	156	132	4382
17	2014	1	Lubrif. Indu...	112	38	78	226	2616	1	7	9519	92	49	2	2117
18	2014	1	Lubrificanti ...	0	0	0	0	1188	0	0	351	1035	0	0	7
19	2014	1	Lubrificanti ...	0	0	0	0	3	0	0	55	3	0	0	66
20	2014	1	Lubrificanti ...	0	0	0	0	3	0	0	416	6329	0	0	4
21	2014	1	Lubrificanti ...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	2014	1	Lubrificanti ...	0	0	0	0	0	0	0	0	1007	0	0	33

OBBIETTIVO:

Sviluppare un programma in MatLab attraverso il quale con delle tecniche di regressione lineare multipla si possano ottenere dei modelli che descrivono la relazione di un gruppo di variabili esplicative riguardanti la meteorologia ed il consumo dei carburanti con alcuni inquinanti.

Il nostro studio consiste nel cercare di capire quale relazione intercorre tra gli inquinanti PM10 ed N02 con i 2 pacchetti di dati riguardanti la meteorologia ed il consumo dei carburanti.

Scelta delle variabili per i modelli di PM10

MODELLO 1

VARIABILE DIPENDENTE	
PM10	concentrazione media di particolati fini (PM10) IN MG/M^3 osservati nella stazione in un dato mese
VARIABILI INDIPENDENTI (REGRESSORI)	
Gasolio_riscaldamento	quantità (tonnellate) di gasolio per riscaldamento venduta su rete ordinaria (strade urbane) in un dato mese nella provincia in cui è installata la centralina
Gasolio_motori_rete_ord	quantità (tonnellate) di gasolio venduta su rete ordinaria (strade urbane) in un dato mese nella provincia in cui è installata la centralina
Benzina_vendita_rete_ord	quantità (tonnellate) di benzina venduta su rete ordinaria (strade urbane) in un dato mese nella provincia in cui è installata la centralina

MODELLO 2

VARIABILE DIPENDENTE	
PM10	cioè la concentrazione media di particolati fini (PM10) IN MG/M^3 osservati nella stazione in un dato mese
VARIABILI INDIPENDENTI (REGRESSORI)	
Pioggia_cumulata	quantità di pioggia totale in mm precipitata nella stazione in un dato mese
Umidità_relativa	umidità relativa in % osservata nella stazione in un dato mese
Temperatura	Temperatura media in Celsius osservata nella stazione in un dato mese

Scelta delle variabili per i modelli di NO2

MODELLO 1

VARIABILE DIPENDENTE	
NO2	Concentrazione media di biossido di azoto (NO2) in mg/m^3 osservata nella stazione in un dato mese
VARIABILI INDIPENDENTI (REGRESSORI)	
Gasolio_riscaldamento	quantità (tonnellate) di gasolio per riscaldamento venduta su rete ordinaria (strade urbane) in un dato mese nella provincia in cui è installata la centralina
Gasolio_motori_rete_ord	quantità (tonnellate) di gasolio venduta su rete ordinaria (strade urbane) in un dato mese nella provincia in cui è installata la centralina
Benzina_vendita_rete_ord	quantità (tonnellate) di benzina venduta su rete ordinaria (strade urbane) in un dato mese nella provincia in cui è installata la centralina

MODELLO 2

VARIABILE DIPENDENTE	
NO2	Concentrazione media di biossido di azoto (NO2) in mg/m^3 osservata nella stazione in un dato mese
VARIABILI INDIPENDENTI (REGRESSORI)	
Pioggia_cumulata	quantità di pioggia totale in mm precipitata nella stazione in un dato mese
Umidità_relativa	umidità relativa in % osservata nella stazione in un dato mese
Temperatura	Temperatura media in Celsius osservata nella stazione in un dato mese

Metodo Stepwise Forward

Abbiamo utilizzato il metodo dello Stepwise Forward, un algoritmo che automaticamente rimuove (o aggiunge) una variabile alla volta al modello di regressione. Il modello migliore è quindi scelto in base alla significatività dei vari coefficienti di regressione.

MODELLI PER PM10

```
%trovo il modello di regressione migliore tra le variabili carburanti per il PM10 con lo Stepwise
ModelloPm10Carb=stepwiselm(X_TotCarb,Y_pm10,'constant','Upper','linear')
%trovo il modello di regressione migliore tra le variabili temperatura per il PM10 con lo Stepwise
ModelloPm10Temp=stepwiselm(X_TotTemp,Y_pm10,'constant','Upper','linear')
```

MODELLI PER NO2

```
%trovo il modello di regressione migliore tra le variabili carburanti per il N02 con lo Stepwise
ModelloNO2Carb=stepwiselm(X_TotCarb,Y_no2,'constant','Upper','linear')
%trovo il modello di regressione migliore tra le variabili temperatura per il N02 con lo Stepwise
ModelloNO2Temp=stepwiselm(X_TotTemp,Y_no2,'constant','Upper','linear')
```

Prima di Confrontare i risultati

Condizioni di esistenza:

occorre verificare che il determinante della matrice X'X sia maggiore di zero. per prima cosa si aggiunge una colonna di uni e poi cal- colare il determinante con la funzione "det" fornita da matlab. I determinanti risultano entrambi maggiori di zero perciò vale la condizione di esistenza.Nel caso non lo fossero stati non sarebbe stato possibile ottenere dei buoni modelli.

PM10

```
%Verifica condizioni di esistenza per ModelloPm10Carb
X_varPM10=[[X3]]
uno=ones(66,1)
Z=[uno,X_varPM10]
Z1=Z'
determinantePM10=det(Z1*Z)      %positivo
%Verifica condizioni di esistenza per ModelloPm10Temp
X_varPM101=[X1],[X2],[[X3]]
Z2=[uno,X_varPM101]
Z3=Z2'
determinantePM101=det(Z3*Z2)      %positivo
```

NO2

```
%Verifica condizioni di esistenza per ModelloNO2Carb
X_varNO2=[X1],[X2],[[X3]]
Z4=[uno,X_varNO2]
Z5=Z4'
determinanteNO2=det(Z5*Z4)      %positivo
%Verifica condizioni di esistenza per ModelloNO2Carb
X_varNO21=[[X1]]
Z6=[uno,X_varNO21]
Z7=Z6'
determinanteNO21=det(Z7*Z6)      %positivo
```

Confronto dei modelli per PM10

Una volta ottenuti i 2 modelli cerchiamo di capire quale sia il migliore dei 2 mettendo a confronto i coefficienti elencati in seguito.

Modello carburanti

```
Linear regression model:
    y ~ 1 + x3

Estimated Coefficients:
                   Estimate       SE       tStat       pValue
                   ________    ________    ______    __________

    (Intercept)      20.811       2.241    9.2863    1.7947e-13
    x3               0.10243    0.013984   7.3245    5.0019e-10


Number of observations: 66, Error degrees of freedom: 64
Root Mean Squared Error: 11.1
R-squared: 0.456,  Adjusted R-Squared: 0.448
F-statistic vs. constant model: 53.6, p-value = 5e-10
1. Adding x1, FStat = 143.2539, pValue = 5.547928e-18
2. Adding x3, FStat = 18.2565, pValue = 6.6499e-05
3. Adding x2, FStat = 9.3243, pValue = 0.0033296
```

Tipo funzione: y=n+x3
R2 (adjusted): 0.448
P-Value: 5E-10

Modello meteo

```
Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate       SE       tStat       pValue
                   ________    ________    ______    __________

    (Intercept)      24.124       11.684    2.0646    0.043146
    x1               -1.1261      0.16553  -6.8027    4.7141e-09
    x2               0.44603      0.14607   3.0536    0.0033296
    x3               -0.076631    0.01406  -5.4504    9.233e-07


Number of observations: 66, Error degrees of freedom: 62
Root Mean Squared Error: 6.95
R-squared: 0.792,  Adjusted R-Squared: 0.782
F-statistic vs. constant model: 78.6, p-value = 4.19e-21
1. Adding x3, FStat = 30.5919, pValue = 6.30762e-07
2. Adding x2, FStat = 10.2619, pValue = 0.0021313
3. Adding x1, FStat = 10.3543, pValue = 0.0020554
```

Tipo funzione: y=n+x1+x2+x3
R2 (adjusted): 0.782
P-Value: 4.19E-21

Confronto dei modelli per NO2

Una volta ottenuti i 2 modelli cerchiamo di capire quale sia il migliore dei 2 mettendo a confronto i coefficienti elencati in seguito.

Modello carburanti

```
Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate        SE         tStat       pValue
                   ________    ________    _______    ___________

    (Intercept)       8.7031      8.4147     1.0343       0.30502
    x1             -0.023398   0.0072713    -3.2178     0.0020554
    x2              0.018446   0.0043928     4.1991    8.7186e-05
    x3              0.053361    0.014757     3.6161    0.00060164


Number of observations: 66, Error degrees of freedom: 62
Root Mean Squared Error: 9.83
R-squared: 0.501,  Adjusted R-Squared: 0.477
F-statistic vs. constant model: 20.8, p-value = 1.95e-09
1. Adding x1, FStat = 98.5427, pValue = 1.4075e-14
```

Tipo funzione: y=n+x1+x2+x3
R2 (adjusted): 0.477
P-Value: 1.95E-09

Modello meteo

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                   Estimate        SE         tStat       pValue
                   ________    ________    _______    ___________

    (Intercept)       54.943      2.1343     25.743    1.7827e-35
    x1               -1.4062     0.14166    -9.9269    1.4075e-14


Number of observations: 66, Error degrees of freedom: 64
Root Mean Squared Error: 8.6
R-squared: 0.606,  Adjusted R-Squared: 0.6
F-statistic vs. constant model: 98.5, p-value = 1.41e-14
```

Tipo funzione: y=n+x1
R2 (adjusted): 0.6
P-Value:1.41E-14

Confronto dei risultati ottenuti

PM10	
Modello carburanti	**Modello meteo**
R2 = 0.448	R2 = 0.782
P_value = 5E-10	P_value = 4.19E-21

R2_MODELLO CARBURANTI < R2_MODELLO METEO

ENTRAMBI P_VALUE PROSSIMI ALLO ZERO -> RIFIUTANO H0

NO2	
Modello carburanti	**Modello meteo**
R2 = 0.477	R2 = 0.6
P_value = 1.95E-09	P_value = 1.41E-14

R2_MODELLO CARBURANTI < R2_MODELLO METEO

ENTRAMBI P_VALUE PROSSIMI ALLO ZERO -> RIFIUTANO H0

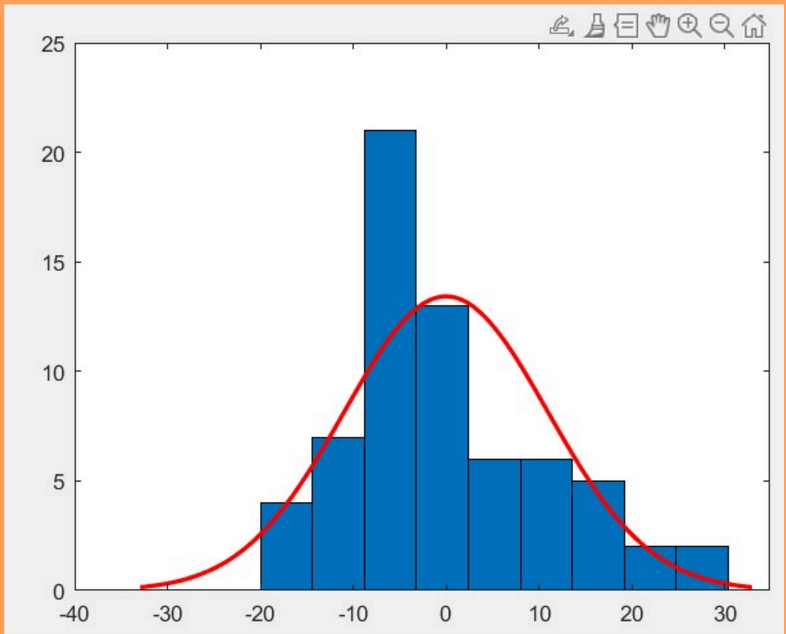
F - TEST

Devo verificare l'ipotesi H0: $\beta0=\beta1=\beta2=\beta3=0$
Contro l'ipotesi H1: $\beta j \neq 0$ per al massimo un elemento

In tutti e 4 i modelli tende a zero

Confronto dei residui per PM10

Modello carburanti



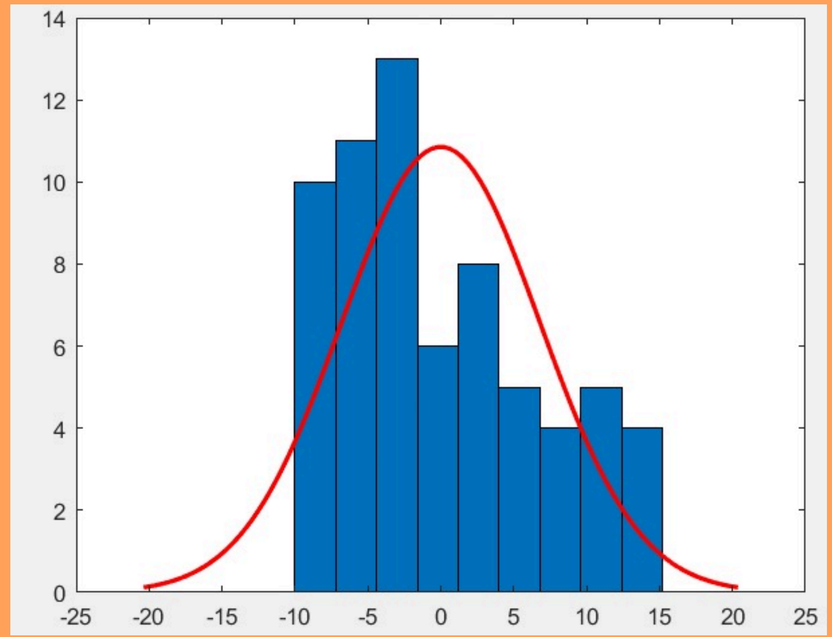
MSE:

122.46

MEDIA RESIDUI:

~0

Modello meteo



MSE :

48.4

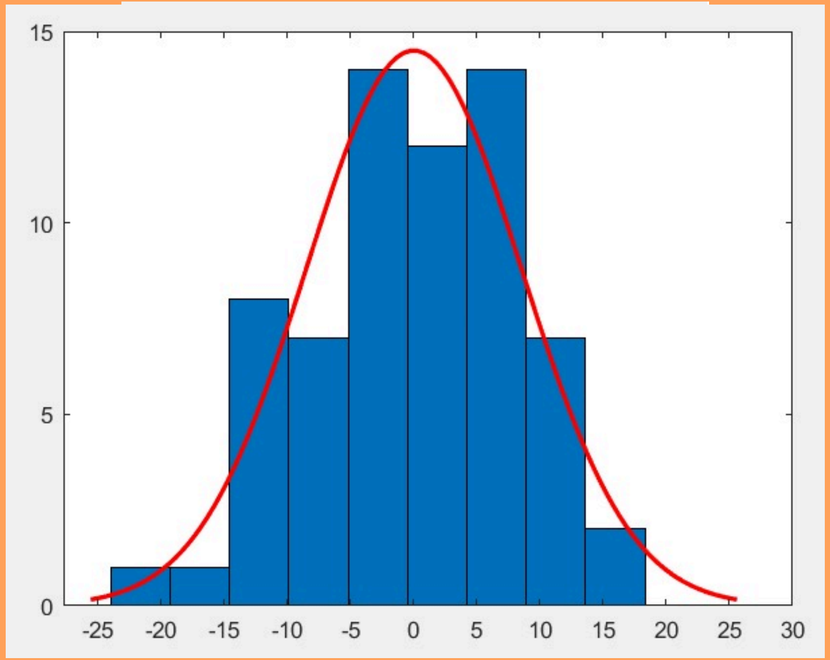
MEDIA RESIDUI:

~0

Analizzando i vari modelli per l’inquinante PM10 abbiamo riscontrato che il modello con i risultati migliori è il modello PM10 Meteorologia

Confronto dei residui NO2

Modello carburanti



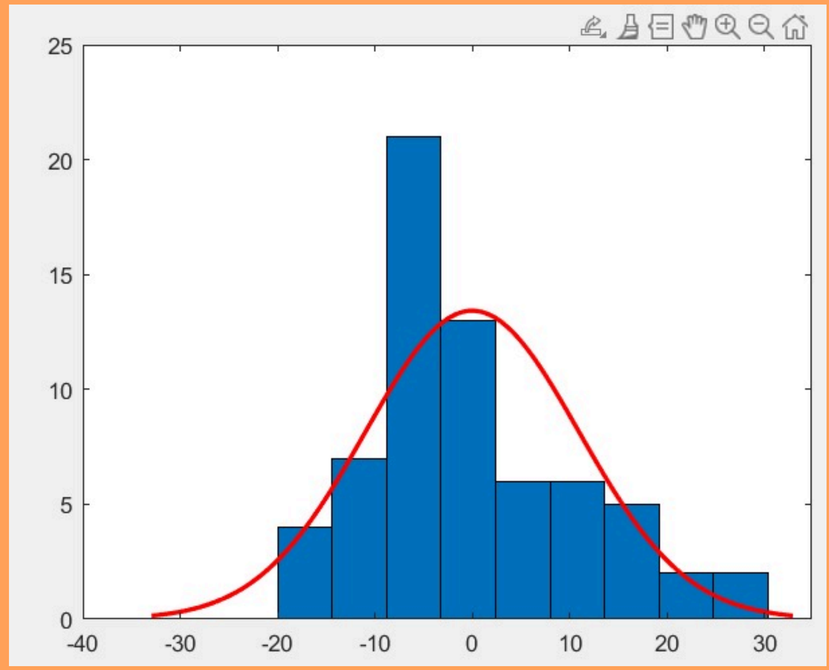
MSE:

97

MEDIA RESIDUI:

~0

Modello meteo



MSE:

74

MEDIA RESIDUI:

~0

Analizzando i vari modelli per l'inquinante NO2 abbiamo riscontrato che il modello con i risultati migliori è il modello NO2 Meteorologia