

Esempio

/*Un campione di 34 negozi di una catena viene selezionato per una ricerca di mercato su un prodotto.

Si prendono in considerazione il numero Y di confezioni vendute in un mese, il prezzo (in centesimi) di vendita X1

e la spesa mensile (in centesimi) X2 per le attività promozionali*/

```
data a;
input id_negozio Y X1 X2 @@;
cards;
1 4141 59 200 18 2730 79 400
2 3842 59 200 19 2618 79 400
3 3056 59 200 20 4421 79 400
4 3519 59 200 21 4113 79 600
5 4226 59 400 22 3746 79 600
6 4630 59 400 23 3532 79 600
7 3507 59 400 24 3825 79 600
8 3754 59 400 25 1096 99 200
9 5000 59 600 26 761 99 200
10 5120 59 600 27 2088 99 200
11 4011 59 600 28 820 99 200
12 5015 59 600 29 2114 99 400
13 1916 79 200 30 1882 99 400
14 675 79 200 31 2159 99 400
15 3636 79 200 32 1602 99 400
16 3224 79 200 33 3354 99 600
17 2295 79 400 34 2927 99 600
;
proc print;

/*scatter di Y contro X1 e contro X2*/
proc gplot;
plot Y*X1;
plot Y*X2;

/*matrice di correlazione*/
proc corr;
var Y X1 X2;

/*verifiche dell'ipotesi di normalità e omoschedasticità della Y*/
proc sort;
by X1;
proc univariate normal;
var Y;
by X1;
proc glm;
class X1;
model Y=X1;
means X1 /hovtest;

proc sort;
by X2;
proc univariate normal;
var Y;
by X2;
proc glm;
class X2;
model Y=X2;
means X2 /hovtest;
```

```

/*regressione*/
proc reg;
model y=x1 x2/influence tol vif clb;
output out=b
      rstudent=res_stud p=stime;
/*La stima del coefficiente di regressione beta1 ci dice come varia Y in
corrispondenza di una variazione unitaria
della variabile X1 si tiene conto anche degli effetti della variabile X2.
La stima -53.22 ci dice che, per un dato ammontare della spesa per l'attività
promozionale, si dovrebbero vendere
circa 53.21 barrette in meno per ogni aumento unitario del prezzo. La stima
3.6131) ci dice che, per un dato
prezzo, si dovrebbero vendere circa 3.61 barrette in più per ogni centesimo
speso in più in attività promozionali.
Tali stime permettono alla divisione di marketing di prevedere l'effetto che
eventuali decisioni in merito
al prezzo e all'attività promozionale possono avere sulle vendite del prodotto*/

proc print;

run;

```

Una volta ottenuta la tabella con i diagnostici, questa può essere copiata sull'Editor e letta da programma per individuare i probabili outliers, come nell'esempio riportato qui di seguito

```

data a;
input osserv res res_stud diag_hat covratio dffits dfbeta_int dfbeta_x1
dfbeta_x2;
cards;
1 720.6905 1.2125 0.1190 1.0850 0.4457 0.4030 -
0.2864 -0.2864
2 421.6905 0.6983 0.1190 1.1933 0.2567 0.2321 -
0.1649 -0.1649
3 -364.3095 -0.6020 0.1190 1.2082 -0.2213 -0.2001
0.1422 0.1422
4 98.6905 0.1622 0.1190 1.2492 0.0596 0.0539 -
0.0383 -0.0383
5 -439.9628 -0.7092 0.0699 1.1286 -0.1945 -0.0889
0.0051 0.1478
6 -1681 -3.0840 0.0699 0.5193 -0.8457 -0.3865 0.0222
0.6425
7 1280 2.2061 0.0699 0.7557 0.6050 0.2765 -0.0158
-0.4596
8 868.0372 1.4345 0.0699 0.9722 0.3934 0.1798 -
0.0103 -0.2989
9 -195.6161 -0.3208 0.1131 1.2314 -0.1146 0.0239 -
0.0708 0.0614
10 -530.6161 -0.8798 0.1131 1.1525 -0.3142 0.0657 -
0.1942 0.1683
11 796.3839 1.3423 0.1131 1.0444 0.4793 -0.1002
0.2963 -0.2568
12 -471.6161 -0.7799 0.1131 1.1714 -0.2785 0.0582 -
0.1721 0.1492
13 83.0789 0.1329 0.0699 1.1843 0.0364 0.0280 -
0.0277 -0.0010
14 487.0789 0.7867 0.0699 1.1159 0.2157 0.1659 -
0.1639 -0.0057
15 -635.9211 -1.0346 0.0699 1.0679 -0.2837 -0.2182
0.2156 0.0074
16 -388.9211 -0.6258 0.0699 1.1411 -0.1716 -0.1320
0.1304 0.0045

```

17	-348.5744	-0.5483	0.0298	1.1037	-0.0960	-0.0068	-
	0.0077	-0.0077					
18	-460.5744	-0.7272	0.0298	1.0791	-0.1274	-0.0090	-
	0.0102	-0.0102					
19	1342	2.2753	0.0298	0.7054	0.3985	0.0283	0.0320
		0.0320					
20	-783.5744	-1.2584	0.0298	0.9746	-0.2204	-0.0156	-
	0.0177	-0.0177					
21	99.7723	0.1606	0.0818	1.1986	0.0480	-0.0300	
		0.0383	0.0058				
22	-132.2277	-0.2129	0.0818	1.1963	-0.0636	0.0398	-
	0.0508	-0.0077					
23	144.7723	0.2331	0.0818	1.1952	0.0696	-0.0435	
		0.0556	0.0084				
24	-412.2277	-0.6682	0.0818	1.1496	-0.1995	0.1248	-
	0.1594	-0.0242					
25	134.4673	0.2203	0.1131	1.2380	0.0787	0.0222	-
	0.0421	0.0486					
26	254.4673	0.4178	0.1131	1.2226	0.1492	0.0420	-
	0.0799	0.0922					
27	-854.5327	-1.4470	0.1131	1.0161	-0.5167	-0.1456	
		0.2768	-0.3194				
28	149.4673	0.2449	0.1131	1.2366	0.0875	0.0247	-
	0.0469	0.0541					
29	311.8140	0.5038	0.0818	1.1717	0.1504	-0.0507	
		0.0182	0.1202				
30	-55.1860	-0.0888	0.0818	1.2008	-0.0265	0.0089	-
	0.0032	-0.0212					
31	-269.1860	-0.4345	0.0818	1.1793	-0.1297	0.0437	-
	0.0157	-0.1037					
32	23.8140	0.0383	0.0818	1.2015	0.0114	-0.0039	
		0.0014	0.0091				
33	617.1607	1.0463	0.1429	1.1560	0.4272	-0.3249	
		0.2819	0.2819				
34	190.1607	0.3172	0.1429	1.2744	0.1295	-0.0985	
		0.0855	0.0855				

```

;
data a1; set a;
h=3;
n=34;
drop res diag_hat;
proc print;

data b1; set a1;
if abs(res_stud)>=2;
proc print;
title1 '|res_stud|>=2';

data b2; set a1;
soglia=3*h/n;
if abs(covratio-1)>=soglia;
proc print;
title1 'abs(covratio-1)>=valore soglia';

data b3; set a1;
soglia=2*(h/n)**0.5;
if abs(dffits)>=soglia;
proc print;
title1 '|dffits|>=valore soglia';

data b4; set a1;
soglia=2*(1/n)**0.5;
if abs(dfbeta_int)>=soglia;

```

```

proc print;
title1 '|dfbeta_int|>=valore soglia';

data b5; set a1;
soglia=2*(1/n)**0.5;
if abs(dfbeta_x1)>=soglia;
proc print;
title1 '|dfbeta_x1|>=valore soglia';

data b6; set a1;
soglia=2*(1/n)**0.5;
if abs(dfbeta_x2)>=soglia;
proc print;
title1 '|dfbeta_x2|>=valore soglia';

run;

```

La sesta osservazione è un probabile outlier, pertanto si può eliminare dal set iniziale e si può ripetere da capo tutta l'analisi.

Per eliminare l'osservazione dal dataset iniziale si possono usare le seguenti istruzioni

```

data a;
input id_negozio Y X1 X2 @@;
cards;
1 4141 59 200 18 2730 79 400
.
.
.
17 2295 79 400 34 2927 99 600
;
data a1;
set a;
if id_negozio ne 6;

```