

Analisi di regressione con SAS

PROC CORR

PROC GPLOT

PROC REG

Esempio: due test su 31 individui

```
title1 'Risultato di 2 test su 31 soggetti';  
data dati;  
infile 'V:\Didat\Carla\regressione\tab11_1.txt';  
input id test1 test2;  
proc print;  
run;
```

N. di variabili: $p = 2$

N. di osservazioni: $n=31$

dati in tab11_1.txt

Risultato di 2 test su 31 soggetti			
Obs	id	test1	test2
1	1	50	69
2	2	66	85
3	3	73	88
4	4	84	70
5	5	57	84
...			
30	30	65	73
31	31	61	52

Ci interessa...

- ◆ Valutare la correlazione tra i due test
- ◆ Noto il risultato del 1° test, possiamo prevedere il risultato del 2° test?
- ◆ La previsione che facciamo è valida?

Matrice di covarianza

$$\mathbf{S} = \{ \text{cov}(\mathbf{X}_s, \mathbf{X}_v) \} \quad s, v = 1, 2, \dots, p$$

Matrice di correlazione

$$\mathbf{R} = \{ \rho_{sv} \} = (1/n) \mathbf{Z}' \mathbf{Z}$$

Coefficiente di correlazione

$$\rho_{sv} = [\text{cov}(\mathbf{X}_s, \mathbf{X}_v)] / [V(\mathbf{X}_s)V(\mathbf{X}_v)]^{1/2}$$

MEDIE E CORRELAZIONI

```
PROC CORR DATA= dati;  
VAR test1 test2;  
RUN;
```

Calcola la matrice di correlazione e di covarianza (opzione COV) per le variabili indicate nell'istruzione VAR

Output di PROC CORR

Risultato di 2 test su 31 soggetti

The CORR Procedure

2 Variables: test1 test2

Simple Statistics

Variable	N	Mean	StdDev	Sum	Minimum	Maximum
test1	31	70.65	14.83	2190	42.00	95.00
test2	31	75.77	15.92	2349	44.00	98.00

Pearson Correlation Coefficients, N = 31

	test1	test2
test1	1.00000	0.70307
test2	0.70307	1.00000

Prob > |r| under H0: Rho=0
<.0001
<.0001

Correlazione e covarianza

$$\mathit{corr}(x, y) = \frac{\mathit{cov}(x, y)}{\sqrt{\mathit{var}(x)} \sqrt{\mathit{var}(y)}}$$

Calcolo delle covarianze in SAS

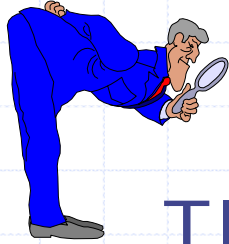
```
proc corr cov;  
var test1 test2;  
run;
```

L'opzione **COV** consente di ottenere sia la matrice di correlazione che la matrice di covarianza

Covariance Matrix, DF = 30

	test1	test2
test1	219.9698925	165.9838710
test2	165.9838710	253.3806452

Scatterplot dei punti



```
TITLE 'SCATTERPLOT di test1*test2';  
PROC GPLOT;  
PLOT test2 * test1;  
RUN;
```

Programma SAS per scatterplot: tab11_1.sas

Scatterplot dei punti con retta di regressione

```
symbol1 v=dot c=blue i=rl ci=red;  
proc gplot data=dati;  
plot test2*test1;  
run;
```

istruzione SYMBOL

V= definisce simbolo dei punti

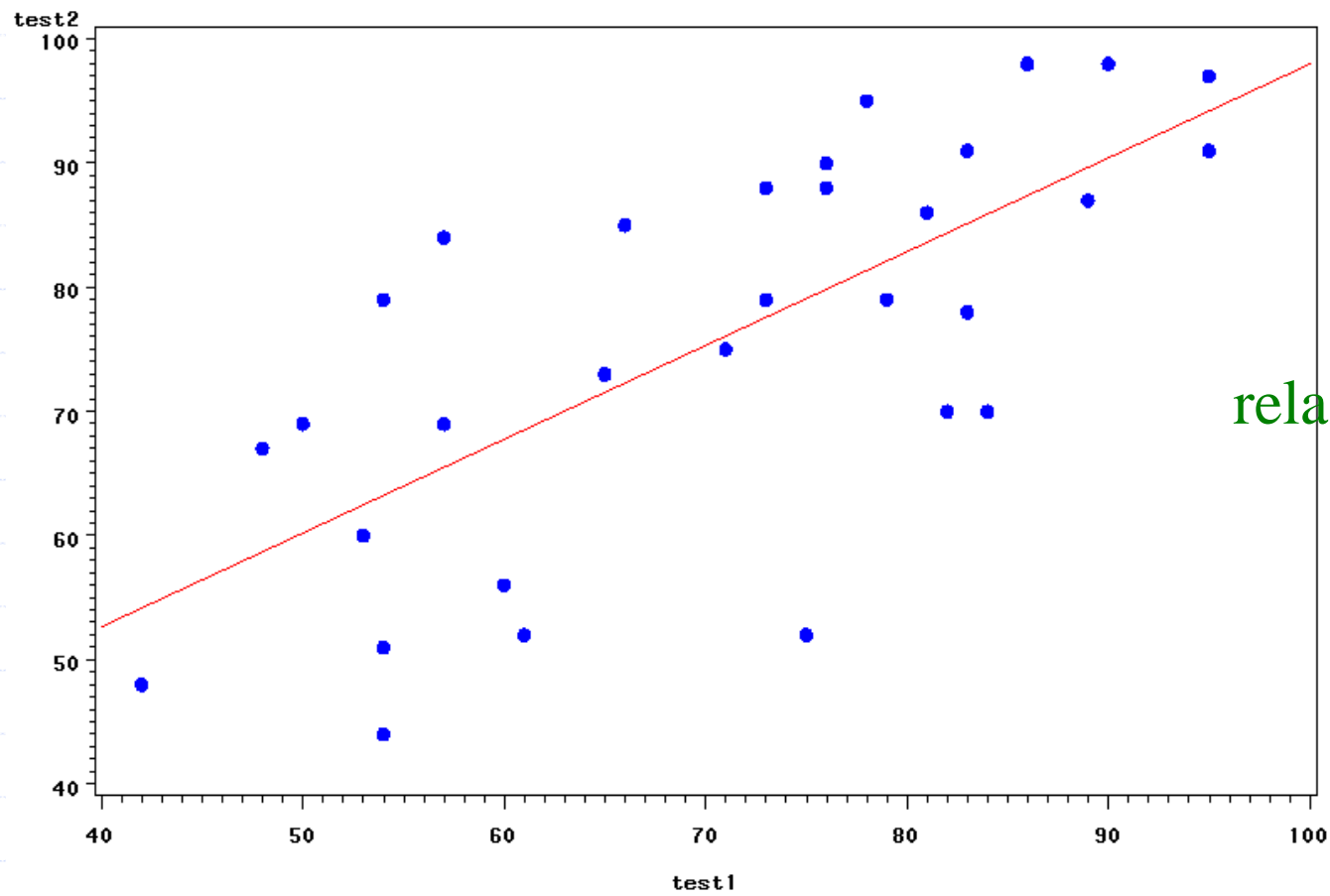
C= definisce colore punti

I= definisce tipo di interpolazione (RL: lineare)

CI= definisce colore linea interpolante

OUTPUT DI PROC GPLOT

Risultato di 2 test su 31 soggetti



relazione lineare

Modello di regressione lineare

```
proc reg data=dati;  
model test2=test1;  
run;
```

$$Y = a + bX + e$$

Stima con metodo dei M.Q.

$$\hat{a} = \bar{y} - \hat{b}\bar{x}, \quad \hat{b} = \frac{\text{cov}(X, Y)}{\text{var}(X)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

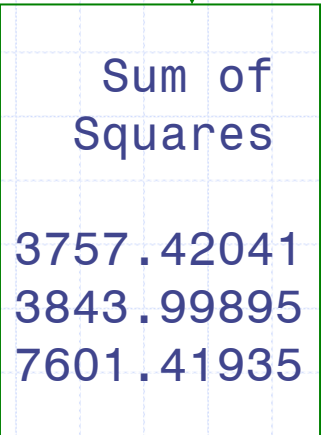
Output di PROC REG (1)

Scomposizione della varianza totale:

$$V(\text{total}) = V(\text{Model}) + V(\text{Error})$$

Analysis of Variance

devianze



Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3757.42041	3757.42041	28.35	<.0001
Error	29	3843.99895	132.55169		
Corrected Total	30	7601.41935			

Output di PROC REG (2)

Root MSE	11.51311
Dependent Mean	75.77419
Coeff Var	15.19397

Std(e)

R-Square	0.4943
Adj R-Sq	0.4769

R²

Media e coefficiente di variazione di Y

Output di PROC REG (3)

Parameter Estimates

	Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
a	Intercept	1	22.46709	10.22358	2.20	0.0361
b	test1	1	0.75458	0.14173	5.32	<.0001

↑
parametri stimati

$$Y^* = 22.467 + 0.755 X$$

residui e valori stimati

Opzioni p e r:
calcola e stampa valori stimati e residui

```
proc reg data=dati;  
model test2=test1/p r;  
output out=stime p=stime r=residui;  
run;
```

output out=

crea un sas data set che contiene:

p= i valori stimati per Y

r= i residui

Analisi dei residui

```
PLOT < y*x > < =symbol >  
      < ... y*x > < =symbol > < / options >;
```

Grafico residui vs valori stimati

```
plot r.* p.;
```

Grafico residui vs variabile indipendente

```
plot r.*test1;
```

Calcolo di Y stimati

```
DATA new;  
test1=70;  
DATA new;  
SET dati new;  
RUN;
```

Aggiunge una obs per X al data set

```
PROC REG DATA=new;  
MODEL test2=test1/p;  
run;
```

Stima modello di regressione

Calcola valori previsti dal modello

Dependent Variable: test2

Output Statistics

Obs	Dep Var test2	Predicted Value	Residual
1	69.0000	60.1959	8.8041
2	85.0000	72.2691	12.7309
5	84.0000	65.4779	18.5221
...			
29	51.0000	63.2142	-12.2142
30	73.0000	71.5145	1.4855
31	52.0000	68.4962	-16.4962
32	.	75.2874	.

$$Y^* = 22.467 + 0.755 * 70$$

Valore previsto