Regressione lineare

Nello studio della relazione tra variabili cardinali

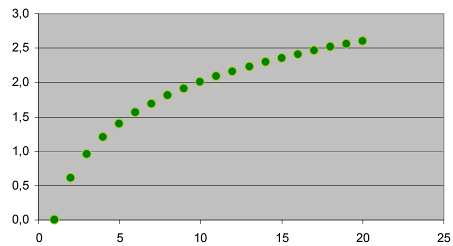
- Significatività
- Forza
- Forma (ricondotta ad una relazione analiticamente nota)

MODELLO STATISTICO

La legge fondamentale della *psicofisica*

R (stimolo)	S (sensazione)
1	0,0
1	0,0
2	0,6
3	1,0
4	1,2
5	1,4
6	1,6
7	1,7
8	1,8
9	1,9
10	2,0
11	2,1
12	2,2
13	2,2
14	2,3
15	2,4
16	2,4
17	2,5
18	2,5
19	2,6
20	2,6

Studio della relazione



$\boxed{S=2 \log R}$

Definizione di una relazione analitica

I modelli lineari esprimono una variabile dipendente (Y) come funzione <u>lineare</u> di almeno una variabile indipendente (X) ed una componente di errore (errore stocastico ε)

$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_i$$

$$Y_i = \alpha + \sum_{k=1}^{K} \beta_{ki} X_{ki} + \varepsilon_i$$

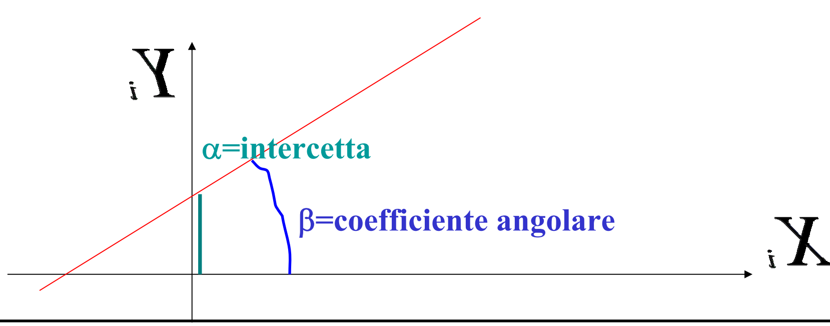
$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_i$$

Ad esempio se la variabile dipendente (Y) fosse il peso, potremmo dire che esso è funzione dell’altezza e dell’età.

Peso dell’individuo= costante alfa +
costante beta_1 * altezza +
costante beta_2 * età +
componente erratica dell’individuo

La regressione lineare mette in relazione una variabile dipendente (Y) misurata a livello di scala cardinale, con una (semplice) o più (multipla) variabili indipendenti (X).

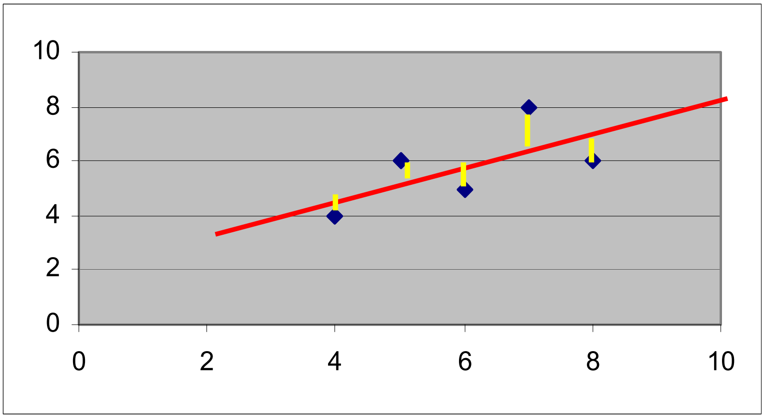
$$Y_i = \alpha + \beta_1 X_{1i} + \varepsilon_i$$



Regressione lineare

Aprile, 2005

Data una distribuzione empirica di due variabili (X,Y), rappresentata con un grafico di dispersione, la retta di regressione è quella che interpola al meglio la nube di punti.

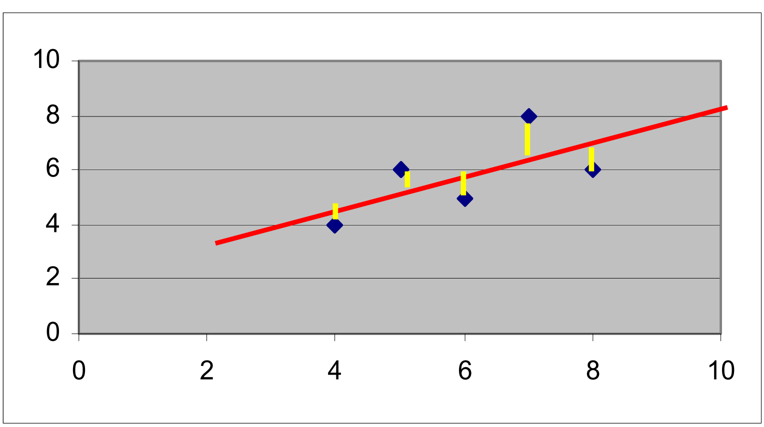


Regressione lineare

Aprile, 2005

$$E(Y_i) = \alpha + \beta_1 X_{1i}$$

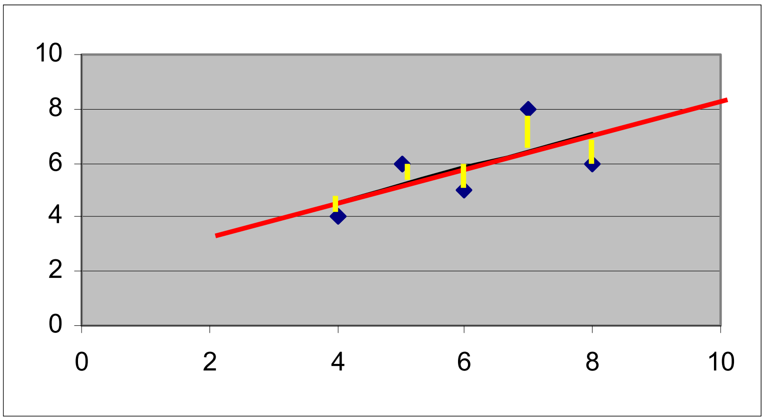
Il valore atteso di Y è funzione di X.



La differenza $\varepsilon_i = Y_i - E(Y_i)$ **è l'errore stocastico**

La retta che interpola al meglio la nuvola di punti è quella che minimizza la somma del quadrato degli errori

Minimi Quadrati Ordinari



Regressione lineare

Aprile, 2005

Assunti

a) $E(\varepsilon_i) = 0$

b) $COV(X, \varepsilon_i) = 0$

c) $COV(\varepsilon_J, \varepsilon_i) = 0$

d) $VAR(\varepsilon_j) = VAR(\varepsilon_i)$ **Omoschedasticità**

f) **Distribuzione normale degli errori**

$$\hat{Y}_i = a + bX_i + e_i$$

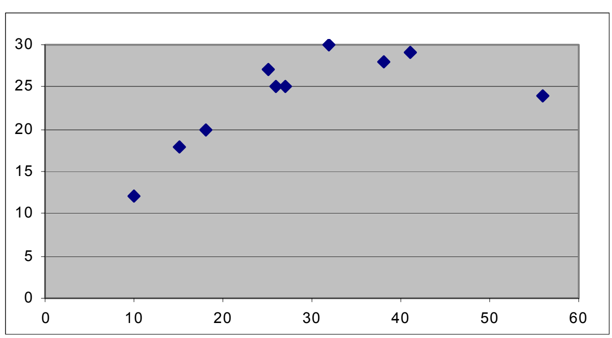
Per stimare i parametri alfa e beta si utilizza il metodo dei minimi quadrati ordinari.

$$a = \overline{Y} - b\overline{X}$$

$$b = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{COV_{x,y}}{VAR_x}$$

Esempio

Ident	Giorni di studio	Votazione
1	25	27
2	38	28
3	41	29
4	27	25
5	56	24
6	32	30
7	15	18
8	10	12
9	26	25
10	18	20



Regressione lineare

Aprile, 2005

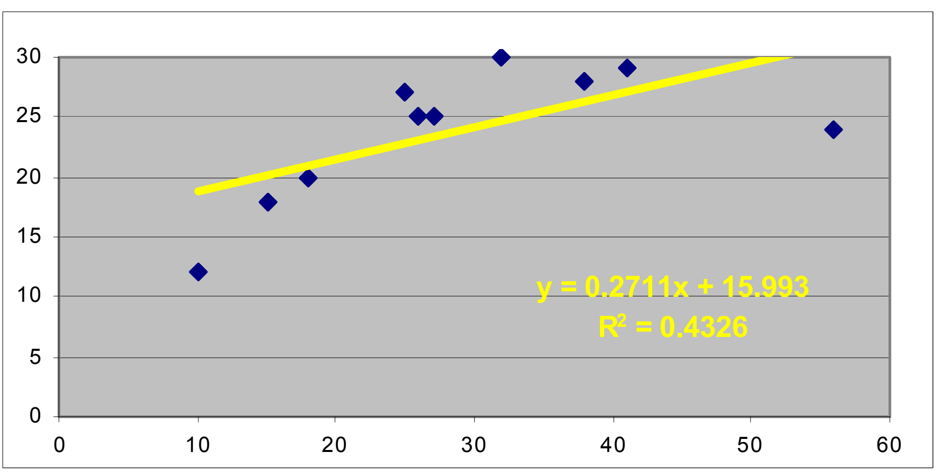
Esempio

Ident	Giorni di studio	Votazione	x-xmedio	y-ymedio	$(x\text{-}xmedio)^2$	(x-xmedio) * (y-ymedio)
1	25	27	-3,8	3,2	14,44	-12,16
2	38	28	9,2	4,2	84,64	38,64
3	41	29	12,2	5,2	148,84	63,44
4	27	25	-1,8	1,2	3,24	-2,16
5	56	24	27,2	0,2	739,84	5,44
6	32	30	3,2	6,2	10,24	19,84
7	15	18	-13,8	-5,8	190,44	80,04
8	10	12	-18,8	-11,8	353,44	221,84
9	26	25	-2,8	1,2	7,84	-3,36
10	18	20	-10,8	-3,8	116,64	41,04
media	28,8	23,8		sommatoria	1669,6	452,6
				b	0,27	
				a	15,99	

Regressione lineare

Aprile, 2005

Esempio



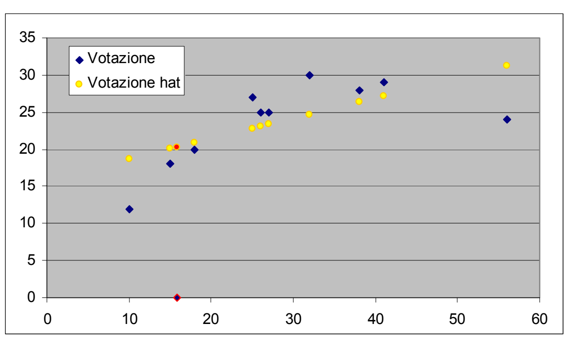
Bontà adattamento della retta

Attraverso il coefficiente di determinazione R2 otteniamo un valore interpretabile come varianza della variabile Y riprodotta dalla variabile X

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = \frac{b^2 \, VAR(X)}{VAR(Y)}$$

Applicando la funzione definita possiamo ricostruire le votazioni dei soggetti (anche per prevedere il voto in base alle ore di studio)

Ident	Giorni di studio	Votazione	Votazione hat
1	25	27	22,8
2	38	28	26,3
3	41	29	27,1
4	27	25	23,3
5	56	24	31,2
6	32	30	24,7
7	15	18	20,1
8	10	12	18,7
9	26	25	23,0
10	18	20	20,9
11	15,9	mancante	20,3



Analisi della Varianza - ANOVA

Aprile, 2005

esercitazione

ident	ore lettura settimana	genere lettura	ORE TV	TITOLO DI STUDIO
2	20	1	6	2
5	8	1	14	3
6	5	1	22	2
8	3	1	36	2
10	13	1	11	3
15	1	1	41	3
20	6	1	23	3
1	15	2	7	1
9	2	2	53	1
11	9	2	15	1
13	3	2	38	3
14	1	2	29	1
16	5	2	25	1
17	2	2	17	1
3	5	3	21	1
4	4	3	26	3
7	9	3	13	2
12	15	3	11	2
18	2	3	56	2
19	11	3	15	2
21	5	3	27	3
		1=giallo		1=obbligo
		2=classico		2=m.sup
		3=fant.		3=laurea