# Interpolazione Statistica

## Come determinare una funzione che rappresenti la relazione tra due grandezze x e y

a cura di  Roberto Rossi

novembre 2008

# Si parla di *INTERPOLAZIONE* …

quando:

Note alcune coppie di dati ($x,y$) , interpretabili come punti del piano,

ci si propone di costruire una funzione (interpolatrice) che sia in grado di descrivere la relazione esistente tra l'insieme dei valori di $x$ e quello dei valori di $y$.

# Sono possibili due casi:

1. Ci troviamo in presenza di una distribuzione che riteniamo lacunosa (mancante di alcuni dati che non possono più essere rilevati).

   La funzione interpolante viene utilizzata per stimare i dati mancanti.

   Una volta fatta questa stima la distribuzione viene completata con l’insieme dei dati trovati.

   (interpolazione Matematica/per punti)

# … (continua) casi possibili:

2. Ci troviamo in presenza di una distribuzione di dati alcuni dei quali vengono ritenuti affetti da errori.

   La funzione interpolante viene utilizzata per sostituire a quella distribuzione una distribuzione approssimata non affetta

   da errori.

   (interpolazione Statistica/fra punti)

---

In ogni caso la funzione interpolante può essere di tipo diverso: lineare, parabolico, esponenziale …

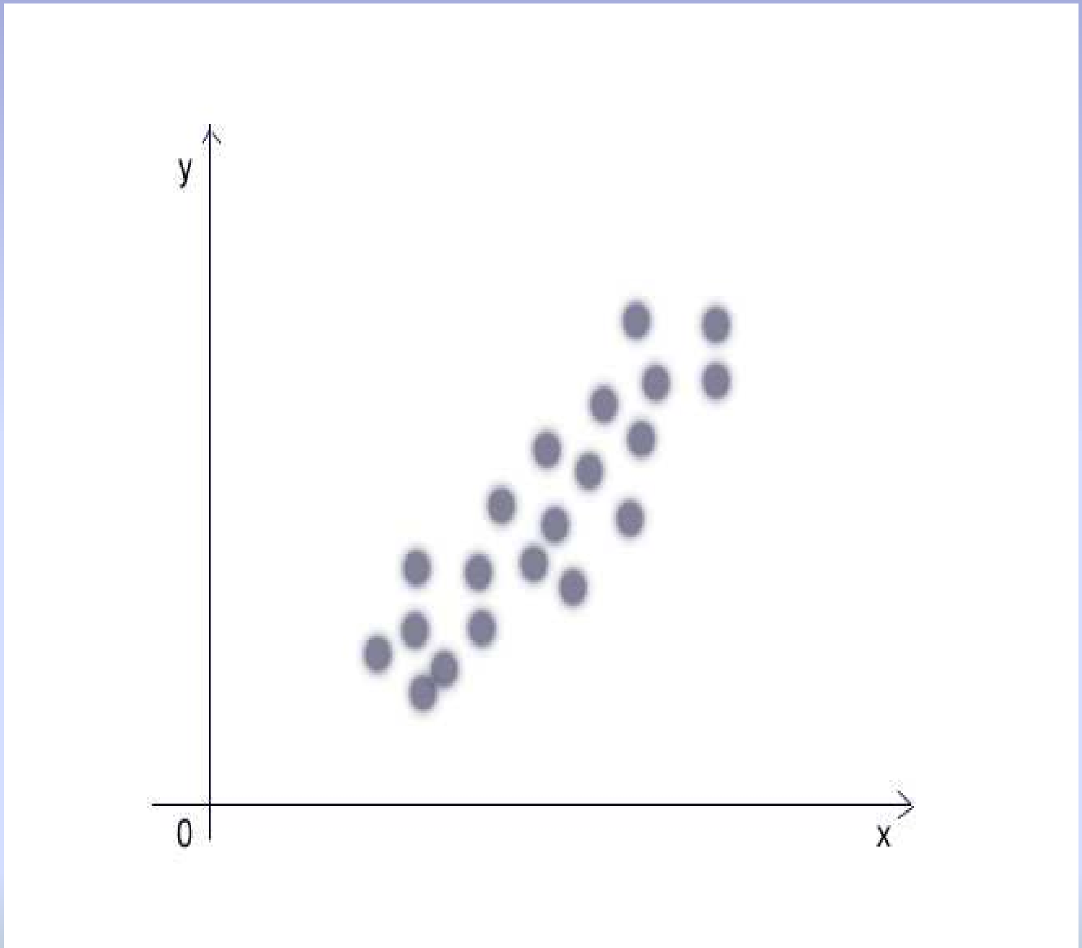
# Interpolazione Statistica
# "metodo dei minimi quadrati"

Consideriamo un fenomeno statistico per il quale disponiamo della seguente distribuzione di dati:

| $x$ | $y$ |
|---|---|
| $x_1$ | $y_1$ |
| $x_2$ | $y_2$ |
| … | … |
| $x_{n-1}$ | $y_{n-1}$ |
| $x_n$ | $y_n$ |

# (continua) metodo dei minimi quadrati

supponiamo che rappresentando tali dati in un sistema di assi cartesiani, si ottenga il diagramma a dispersione

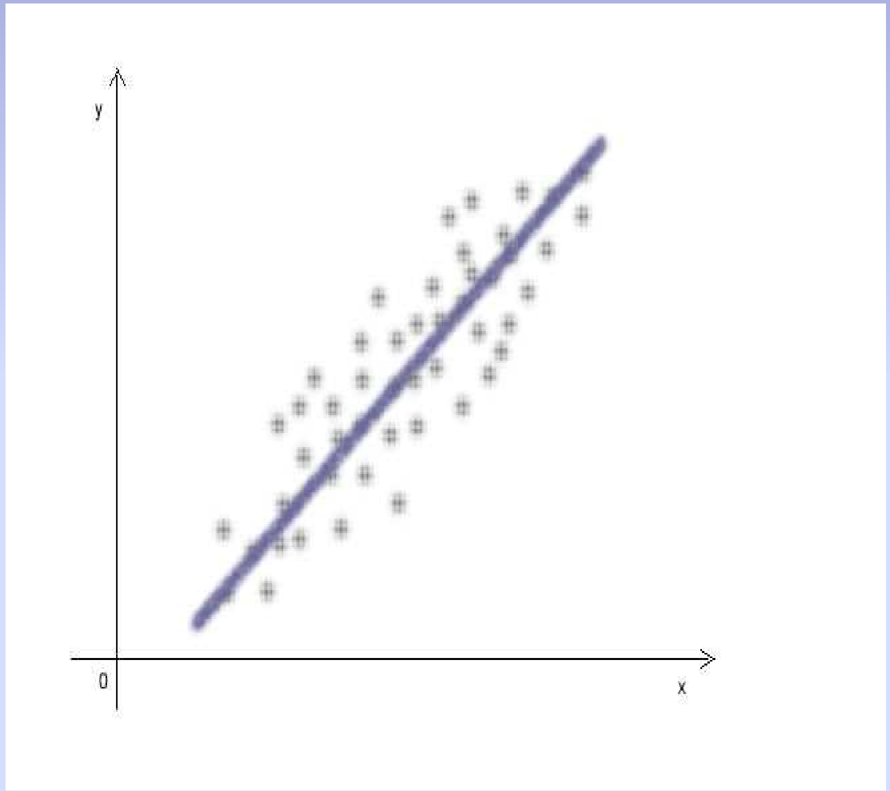


# (continua) metodo dei minimi quadrati

- Poiché riteniamo che i dati osservati siano affetti da errori vogliamo costruire una funzione interpolante che ci permetta di sostituire ai dati $y_i$ osservati i dati yi* interpolati, più regolari anche se approssimati.
- Una volta decisa la funzione interpolante bisognerà fare in modo che l’approssimazione presenti un buon accostamento.

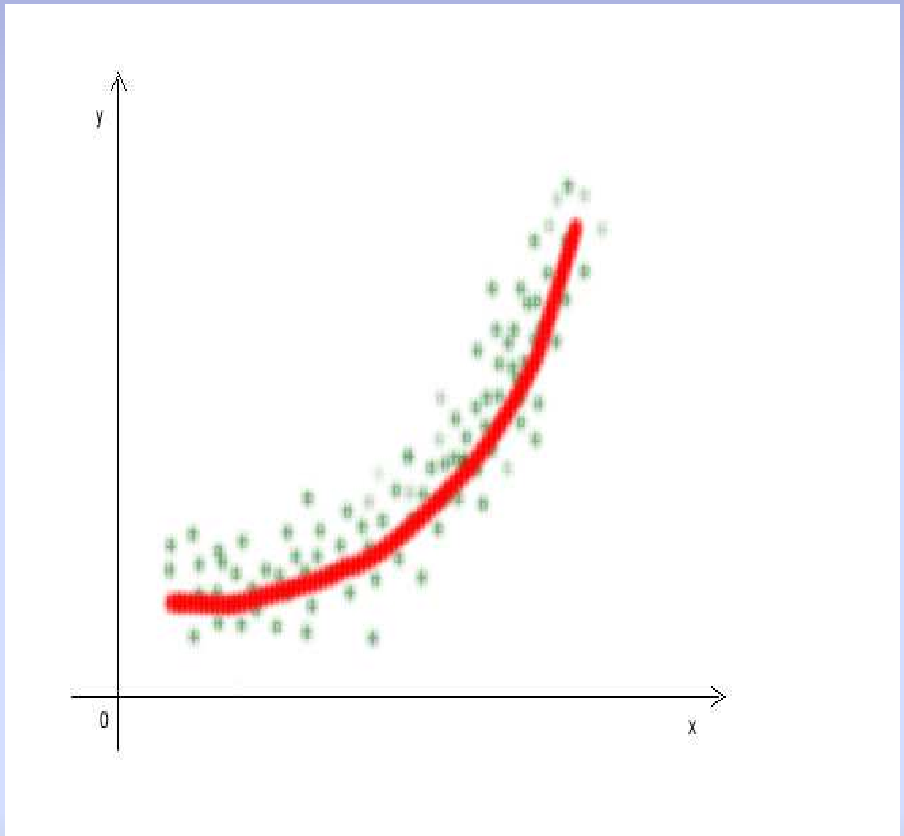
# Scelta del tipo di funzione interpolante

- la funzione può essere di tipo diverso: lineare (primo grado), parabolica (secondo grado), esponenziale etc.
- in generale la scelta può essere suggerita dall'esame del diagramma a dispersione.
- In questo primo esempio potrebbe essere logica l'adozione di una funzione lineare (una certa tendenza dei dati di disporsi attorno a una retta)



# Scelta del tipo di funzione interpolante

- In questo secondo esempio sembra più logica l'adozione di una funzione esponenziale.
- Osservazione: in concreto viene spesso adottata una funzione lineare, ciò dipende dalla sua semplicità e dalla facilità di manipolazione che consegue.



# Condizioni per un 'buon accostamento'

- dopo aver scelto il tipo di funzione interpolante, come possiamo ottenere un buon accostamento tra la distribuzione dei valori osservati e quella dei valori teorici ottenuti interpolando?
- Consideriamo i valori $y_i$ osservati e quelli teorici $y_i^*$ forniti dalla funzione interpolante.
- Gli errori commessi sostituendo a ciascun $y_i$ il corrispondente $y_i^*$ sono dati dalle differenze:

$$y_i - y_i^*$$

- al fine di ottenere un buon accostamento occorre minimizzare questi errori.

# (continua) buon accostamento

- Per raggiungere l'obiettivo non possiamo prendere in considerazione la somma delle differenze: essendo alcune positive, altre negative altre ancora nulle esse potrebbero anche compensarsi.
- Prenderemo in considerazione la somma dei quadrati delle differenze e poniamo come condizione di accostamento:

$$\Sigma_i (y_i - y_i^*)^2 = \text{minimo}$$

(metodo dei minimi quadrati)

# Interpolazione statistica metodo dei minimi quadrati

- qualunque sia il tipo di funzione interpolante adottato, la condizione di accostamento prevede che sia minima la somma dei quadrati delle differenze fra valori osservati e valori teorici
- Osservazione: precisiamo che nelle coppie (x, y) i valori di x indicano i tempi (le date) in cui i vengono rilevati i valori di y. Ecco perché si parla di evoluzione temporale di un fenomeno statistico. Le coppie mostrano come il fenomeno evolve nel tempo.

# Metodo dei minimi quadrati: funzione lineare

- Se come funzione interpolante viene scelta la funzione lineare $y^*=a+bx_i$
- La condizione di accostamento diventa:

$$\Sigma_i\,[y_i - (a+bx_i)\,]^2 = \text{minimo}$$

- Si tratta quindi di determinare i valori dei parametri $a$ e $b$ in modo tale che la funzione: $f(a, b) = \Sigma_i\,[y_i - (a+bx_i)\,]^2$ sia minima.

(minimizzare una funzione di due variabili)

# Valori di a e b nel caso di funzione lineare (minimi quadrati)

…che soddisfano la condizione di accostamento sono:

$$b = \frac{\sum_{i} [x_i - M(x)][(y_i - M(y)]}{\sum_{i} [x_i - M(x)]^2}$$

$$a = M(y) - b \cdot M(x)$$

Dove:

- M(x) la media dei valori di $x_i$,
- M(y)la media dei valori di $y_i$

# Es. produzione di frumento (milioni di quintali) anni dal 1993 al 2001.

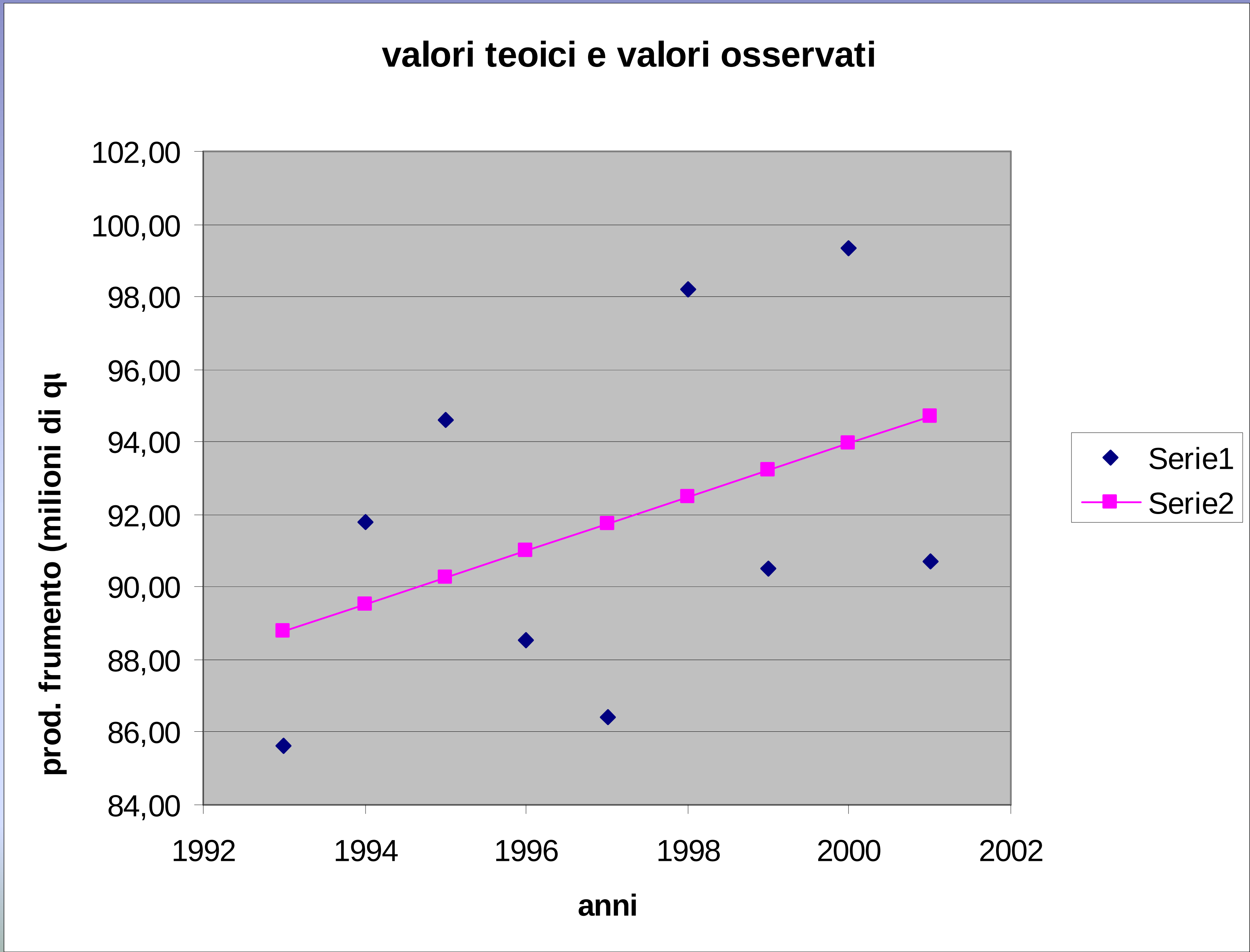
- I valori $x_i$ rappresentano gli anni che per semplicità vengono contati assumendo il 1993 come anno 0, il ’94 come anno 1 …il 2001 come l’anno 8.
- i valori $y_i$ rappresentano la quantità di frumento prodotta nei singoli anni.

continua…

# … produzione di frumento (milioni di quintali) anni dal 1993 al 2001.

| anni | x | y | X-M(x) | Y-M(y) | [X-M(x)][Y-M(y)] | [X-M(x)]^2 | y* | a | b |
|---|---|---|---|---|---|---|---|---|---|
| 1993 | 0,00 | 85,65 | -4,00 | -6,10 | 24,40 | 16,00 | 88,80 | 0,74 | 88,80 |
| 1994 | 1,00 | 91,80 | -3,00 | 0,05 | -0,15 | 9,00 | 89,54 | | |
| 1995 | 2,00 | 94,60 | -2,00 | 2,85 | -5,70 | 4,00 | 90,27 | | |
| 1996 | 3,00 | 88,55 | -1,00 | -3,20 | 3,20 | 1,00 | 91,01 | | |
| 1997 | 4,00 | 86,40 | 0,00 | -5,35 | 0,00 | 0,00 | 91,75 | | |
| 1998 | 5,00 | 98,20 | 1,00 | 6,45 | 6,45 | 1,00 | 92,49 | | |
| 1999 | 6,00 | 90,50 | 2,00 | -1,25 | -2,50 | 4,00 | 93,23 | | |
| 2000 | 7,00 | 99,35 | 3,00 | 7,60 | 22,80 | 9,00 | 93,97 | | |
| 2001 | 8,00 | 90,70 | 4,00 | -1,05 | -4,20 | 16,00 | 94,70 | | |
| | 36,00 | 825,75 | 0,00 | 0,00 | 44,30 | 60,00 | 825,75 | | |

| anni | X | Y | X-M(x) | Y-M(y) | x'y' | X-M(x) ^2 | y* | b | a |
|---|---|---|---|---|---|---|---|---|---|
| 1993 | 0 | 85,65 | -4 | -6,10 | 24,4 | 16 | 88,79667 | 0,738333 | 88,79667 |
| 1994 | 1 | 91,80 | -3 | 0,05 | -0,15 | 9 | 89,535 | | |
| 1995 | 2 | 94,60 | -2 | 2,85 | -5,7 | 4 | 90,27333 | | |
| 1996 | 3 | 88,55 | -1 | -3,20 | 3,2 | 1 | 91,01167 | | |
| 1997 | 4 | 86,40 | 0 | -5,35 | 0 | 0 | 91,75 | | |
| 1998 | 5 | 98,20 | 1 | 6,45 | 6,45 | 1 | 92,48833 | | |
| 1999 | 6 | 90,50 | 2 | -1,25 | -2,5 | 4 | 93,22667 | | |
| 2000 | 7 | 99,35 | 3 | 7,60 | 22,8 | 9 | 93,965 | | |
| 2001 | 8 | 90,70 | 4 | -1,05 | -4,2 | 16 | 94,70333 | | |
| | 36 | 825,75 | 0 | 0,00 | 44,3 | 60 | 825,75 | | |
| M(x) = | 4 | M(y) =91,75 | | | S | S* | | | |


valori teoici e valori osservati
prod. frumento (milioni di q
102,00
100,00
98,00
96,00
94,00
92,00
90,00
88,00
86,00
84,00
1992
1994
1996
1998
2000
2002
anni
Serie1
Serie2