Università degli Studi di Napoli
Federico II
Dipartimento di Scienze Economiche e Statistiche

Anno accademico 2015-'16

Corso di **Statistica**

Germana Scepi

scepi@unina.it

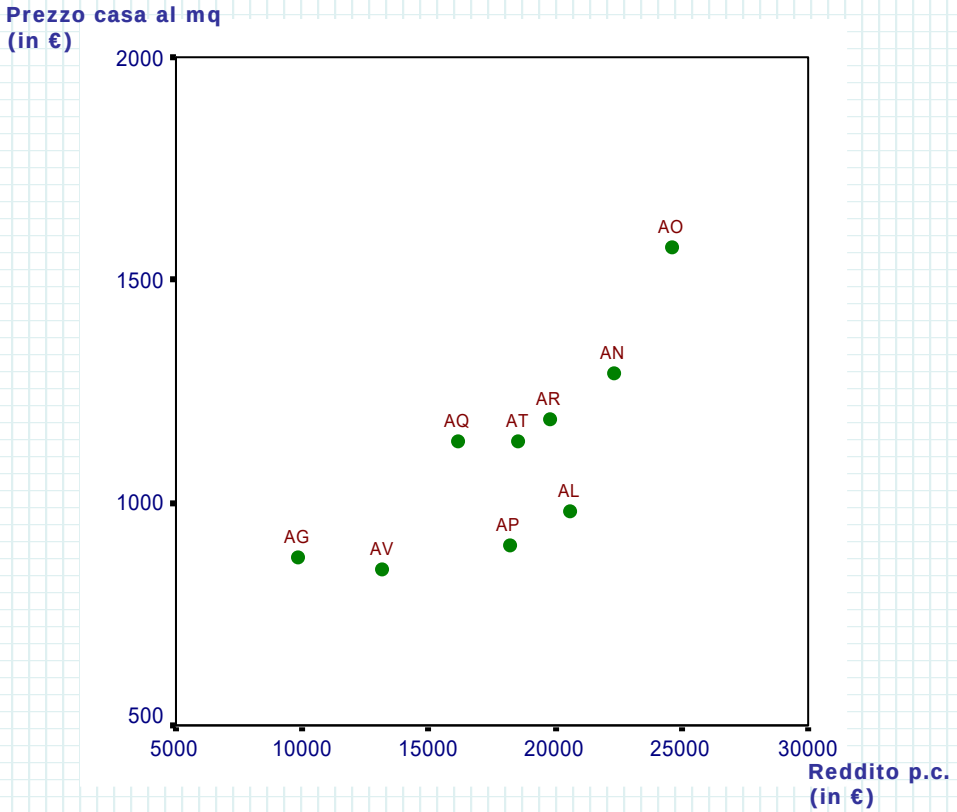
Lezione: **11** Argomento: ***Interpolazione e Regressione***

L'interpolazione

Date due variabili, **X** e **Y**, rappresentabili come assi di un piano cartesiano, e data una nuvola di punti sul piano, costituita dalle *n* coppie di valori osservati sulle unità statistiche, il problema dell'***interpolazione*** consiste nel trovare l'equazione di una curva passante per alcuni punti del piano, oppure "vicino" ai punti stessi.

L'interpolazione può essere di due tipi:

- **Interpolazione matematica**
- **Interpolazione statistica**



L'interpolazione

Interpolazione matematica

Data una successione di n coppie di numeri x_i,y_i, che nel piano corrispondono ad altrettanti punti P_1, P_2, ..., P_n, e scelta una **funzione di *X* contenente *n* parametri a_0, a_1, ..., a_{n-1}**, determinare il valore di questi parametri in modo che la curva passi ***per*** i punti dati.

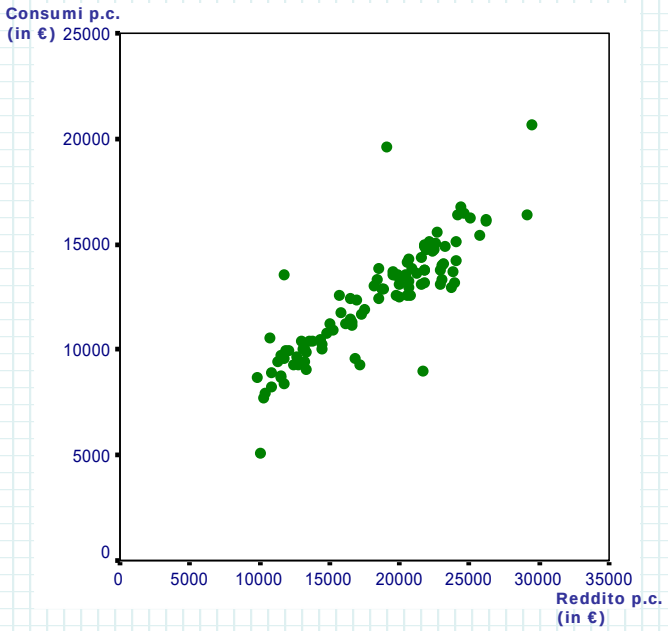
Avendo tanti punti quanti sono i parametri, è possibile scrivere un sistema di *n* equazioni con *n* incognite (i parametri, appunto) che, in generale, ammette ***una e una sola soluzione*** che individua in modo univoco l'interpolante.

L'interpolazione

Interpolazione statistica

Quando l'insieme dei punti a disposizione è numeroso, è molto probabile che questi si dispongano, sul piano, definendo una ***nube di punti***.

In questi casi, all'interpolante matematica si sostituisce l'***interpolante statistica***, che abbandona il vincolo di passare ***per*** i punti, a favore di una condizione più realistica, di passare, cioè, ***fra*** i punti dati.

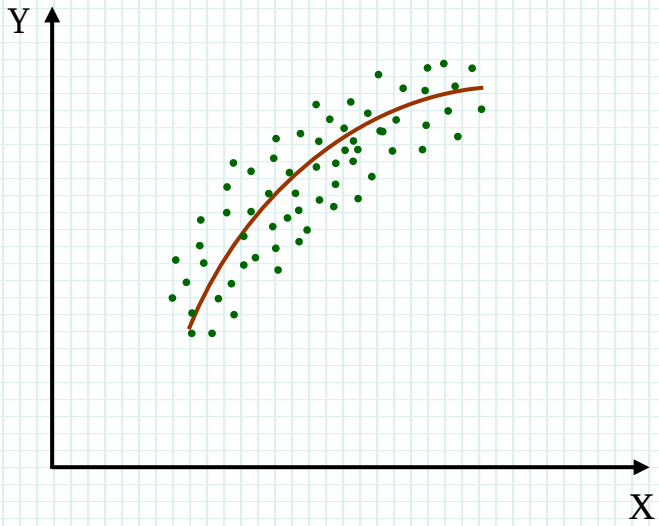


L'interpolazione

Interpolazione statistica

Quando l'insieme dei punti a disposizione è numeroso, è molto probabile che questi si dispongano, sul piano, definendo una ***nube di punti***.

In questi casi, all'interpolante matematica si sostituisce l'***interpolante statistica***, che abbandona il vincolo di passare ***per*** i punti, a favore di una condizione più realistica, di passare, cioè, ***fra*** i punti dati.



Mentre, però, nell'interpolazione matematica avevamo un'unica soluzione individuata dalla soluzione del sistema di equazioni, nell'interpolazione statistica esistono ***infinite curve*** che possono passare fra i punti.

E' quindi necessario stabilire delle ***condizioni*** cui la funzione interpolante deve soddisfare per far sì che il problema sia definito in modo univoco.

Ancora, nell'interpolazione statistica non c'è una relazione fissa tra il numero dei parametri dell'interpolante e il numero dei punti, risultando sufficiente che i secondi superino i primi.

L'interpolazione

Interpolazione statistica

Poiché valori osservati e valori interpolati sono diversi, è necessario distinguerli, indicando con:

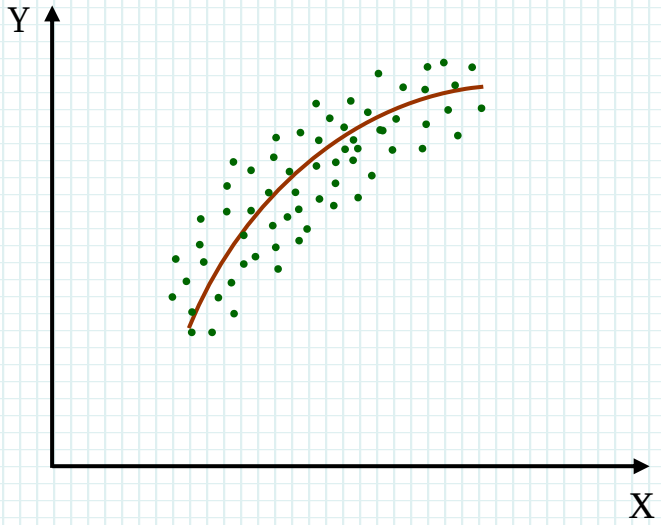
y_i le ordinate osservate corrispondenti a un certo valore x_i;

$\hat{y}_i$ le ordinate *calcolate* corrispondenti a un certo valore x_i;

La teoria dell'interpolazione statistica è relativamente semplice quando la curva interpolante è un *polinomio completo di grado k*, rappresentata, cioè, dall'espressione:

$$\hat{y} = a_0 + a_1 x + a_2 x^2 + \square + a_k x^k$$

Dove a_0, a_1, a_2, ..., a_k sono delle costanti e rappresentano i parametri della curva.



Le funzioni che contengono tutti i parametri alla prima potenza vengono definite *lineari nei parametri*.

L'interpolazione

Metodo dei minimi quadrati

(Gauss, 1795; Legendre, 1805)

Funzione interpolante: $\hat{y}_i = \varphi(x_i, a_0, a_1, \ldots, a_k)$

La condizione dei minimi quadrati determina i parametri incogniti in modo da rendere minima la somma dei quadrati degli scarti fra valori interpolati e valori osservati:

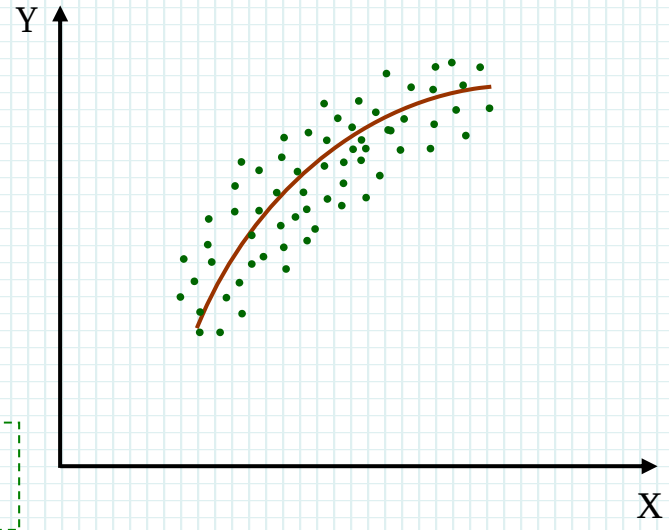
$$S = \sum_i (\hat{y}_i - y_i)^2 = min$$

$$S = \sum_i \left[\varphi(x_i, a_0, a_1, \ldots, a_k) - y_i\right]^2 = min$$

La quantità da minimizzare, ***S***, è funzione dei $k+1$ parametri incogniti. Condizione necessaria perché ciò si verifichi è che le $k+1$ derivate parziali di ***S*** rispetto ai parametri siano nulle. Si ha allora il sistema:

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 2\sum_i \left[\varphi(x_i, a_0, a_1, \ldots, a_k) - y_i\right]\dfrac{\partial \varphi}{\partial a_0} = 0 \\ \dfrac{\partial S}{\partial a_1} = 2\sum_i \left[\varphi(x_i, a_0, a_1, \ldots, a_k) - y_i\right]\dfrac{\partial \varphi}{\partial a_1} = 0 \\ \vdots \\ \dfrac{\partial S}{\partial a_k} = 2\sum_i \left[\varphi(x_i, a_0, a_1, \ldots, a_k) - y_i\right]\dfrac{\partial \varphi}{\partial a_k} = 0 \end{cases}$$

Sistema di equazioni normali a minimi quadrati



L'interpolazione

Metodo dei minimi quadrati *(Gauss, 1795; Legendre, 1805)*

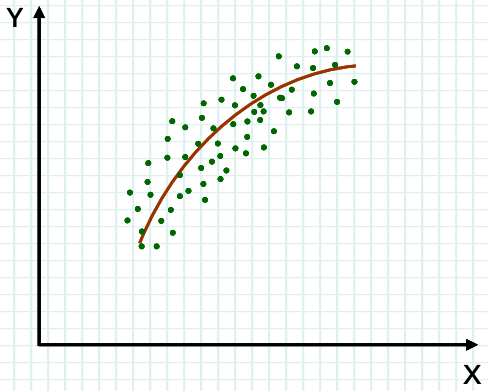
Funzione interpolante: $\hat{y}_i = \varphi\left(x_i, a_0, a_1, \ldots, a_k\right)$

La condizione dei minimi quadrati determina i parametri incogniti in modo da rendere minima la somma dei quadrati degli scarti fra valori interpolati e valori osservati:

$$S = \sum_i \left(\hat{y}_i - y_i\right)^2 = min$$

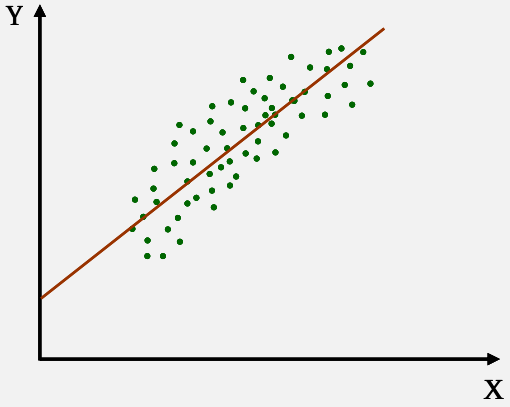
$$S = \sum_i \left[\varphi\left(x_i, a_0, a_1, \ldots, a_k\right) - y_i\right]^2 = min$$

La soluzione individuata rende minima la somma dei quadrati degli scarti rispetto a qualunque altra curva *dello stesso tipo di quella scelta*.



Quando è k=1, la funzione interpolante è la retta e l'espressione si riduce a: $\hat{y} = a_0 + a_1 x$

Che può anche essere trovata nelle forme: $\hat{y} = a + bx$ $\hat{y} = b_0 + b_1 x$



L'interpolazione

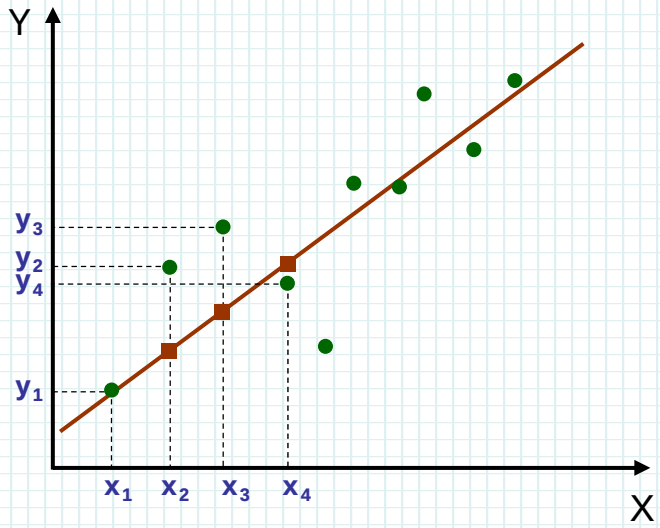
$$S = \sum_i \left(y_i - \hat{y}_i\right)^2 = min$$

Interpolante lineare ⇒ $\hat{y} = b_0 + b_1 x$

Metodo dei minimi quadrati *(Gauss, 1795; Legendre, 1805)*

$$S(b_0, b_1) = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2 = \min$$

$$\frac{\partial\left[S(b_0, b_1)\right]}{\partial b_0} = 0 \quad ; \quad \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_1} = 0$$



L'interpolazione

$$S = \sum_i \left(y_i - \hat{y}_i\right)^2 = min$$

Interpolante lineare ⇒ $\hat{y} = b_0 + b_1 x$

Metodo dei minimi quadrati *(Gauss, 1795; Legendre, 1805)*

$$S(b_0, b_1) = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2 = \min \quad \Rightarrow \quad \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_0} = 0 \quad ; \quad \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_1} = 0$$

NOTA:

La derivata di una funzione quadratica è uguale a due volte la funzione non derivata moltiplicato la derivata della funzione:

$$D\left(f^2(x)\right) = 2 f(x) \cdot f'(x)$$

Quindi, le derivate, rispetto a b_0 e a b_1 di $\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2$ sono:

$$\frac{\partial}{\partial b_0}\left[\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2\right] = 2\sum_i \left(y_i - b_0 - b_1 x_i\right)\cdot(-1) = -2\sum_i \left(y_i - b_0 - b_1 x_i\right)$$

$$\frac{\partial}{\partial b_1}\left[\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2\right] = 2\sum_i \left(y_i - b_0 - b_1 x_i\right)\cdot(-x_i) = -2x_i\sum_i \left(y_i - b_0 - b_1 x_i\right)$$

L'interpolazione

$$S = \sum_i \left(y_i - \hat{y}_i\right)^2 = min$$

Interpolante lineare ⇒ $\hat{y} = b_0 + b_1 x$

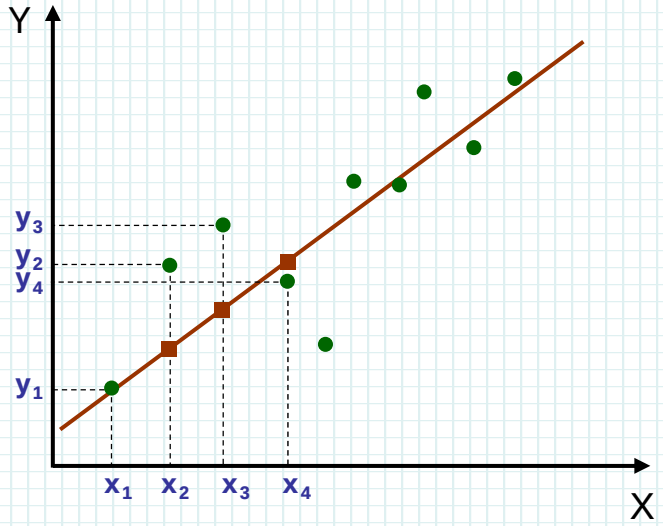
Metodo dei minimi quadrati *(Gauss, 1795; Legendre, 1805)*

$$S(b_0, b_1) = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2 = \min$$

$$\frac{\partial\left[S(b_0, b_1)\right]}{\partial b_0} = 0 \quad ; \quad \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_1} = 0$$

$$\begin{cases} -2\left[\sum_i \left(y_i - b_0 - b_1 x_i\right)\right] = 0 \\ -2x_i\left[\sum_i \left(y_i - b_0 - b_1 x_i\right)\right] = 0 \end{cases}$$

$$\begin{cases} -2\left[\sum_i y_i - n b_0 - b_1 \sum_i x_i\right] = 0 \\ -2\left[\sum_i x_i y_i - b_0 \sum_i x_i - b_1 \sum_i x_i^2\right] = 0 \end{cases}$$



L'interpolazione

$$S = \sum_i \left(\hat{y}_i - y_i\right)^2 = min$$

Interpolante lineare $\hat{y} = b_0 + b_1 x$

Metodo dei minimi quadrati *(Gauss, 1795; Legendre, 1805)*

$$S(b_0, b_1) = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2 = \min \quad \Rightarrow \quad \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_0} = \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_1} = 0$$

dividiamo per n

① $\begin{cases} \sum_i y_i - n b_0 - b_1 \sum_i x_i = 0 \\ \sum_i x_i y_i - b_0 \sum_i x_i - b_1 \sum_i x_i^2 = 0 \end{cases}$ ⇒ ② $\begin{cases} n\bar{y} - n b_0 - b_1 n\bar{x} = 0 \\ \sum_i x_i y_i - b_0 n\bar{x} - b_1 \sum_i x_i^2 = 0 \end{cases}$ → $\bar{y} - b_0 - b_1\bar{x} = 0$

$$b_0 = \bar{y} - b_1\bar{x}$$

③ $\begin{cases} b_0 = \bar{y} - b_1\bar{x} \\ \sum_i x_i y_i - \left(\bar{y} - b_1\bar{x}\right) n\bar{x} - b_1 \sum_i x_i^2 = 0 \end{cases}$ ⇒ ④ $\begin{cases} b_0 = \bar{y} - b_1\bar{x} \\ \sum_i x_i y_i - n\bar{x}\bar{y} + b_1 n\left(\bar{x}\right)^2 - b_1 \sum_i x_i^2 = 0 \end{cases}$

dividiamo per n

⑤ $\begin{cases} b_0 = \bar{y} - b_1\bar{x} \\ \frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y} - b_1\left[\frac{\sum_i x_i^2}{n} - \left(\bar{x}\right)^2\right] = 0 \end{cases}$

L'interpolazione

$$S = \sum_i \left(\hat{y}_i - y_i\right)^2 = min$$

Interpolante lineare $\hat{y} = b_0 + b_1 x$

Metodo dei minimi quadrati *(Gauss, 1795; Legendre, 1805)*

$$S(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = \min \quad \Rightarrow \quad \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_0} = \frac{\partial\left[S(b_0, b_1)\right]}{\partial b_1} = 0$$

$$\begin{cases} b_0 = \bar{y} - b_1 \bar{x} \\ \underbrace{\frac{1}{n}\sum_i x_i y_i - \bar{x}\bar{y}}_{Cov(XY)} - b_1 \left[\underbrace{\frac{\sum_i x_i^2}{n} - (\bar{x})^2}_{Var(X)}\right] = 0 \end{cases} \Rightarrow Cov(XY) - b_1 \cdot Var(X) = 0 \Rightarrow b_1 = \frac{Cov(XY)}{Var(X)}$$

Cov(XY) **Var(X)**

Parametri della interpolante lineare ricavati con il metodo dei minimi quadrati: $b_0 = \bar{y} - b_1 \bar{x}$ $b_1 = \frac{Cov(XY)}{Var(X)}$

La retta costruita con i valori di b_0 e b_1 ottenuti dalla risoluzione del sistema, sarà dunque quella più "vicina" ai punti, ossia quella che rende minima la somma dei quadrati delle distanze tra valori osservati e valori teorici della variabile dipendente y.

La Regressione

La Regressione sul testo *Statistica per le decisioni*

18. Il modello di Regressione lineare

I paragrafi 18.4, 18.6 e 18.7 verranno trattati nella **Lezione 23**.

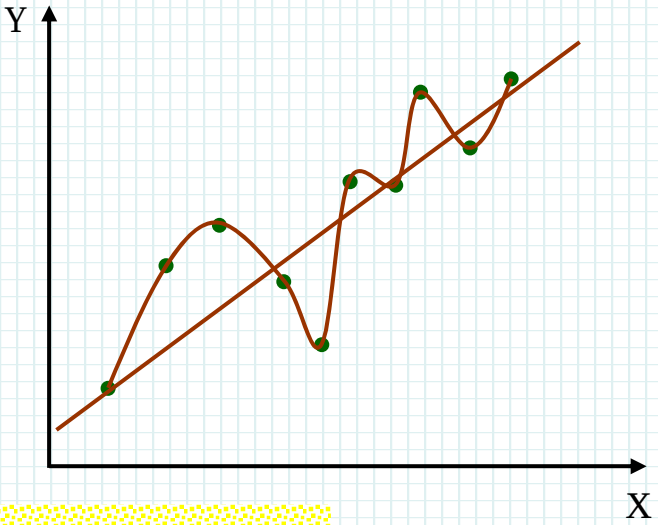
La Regressione

X ⟶ Variabile indipendente (data)

Y ⟶ Variabile dipendente

Dipendenza funzionale (o deterministica): $Y = f(X;\theta)$

- Da un ***punto di vista analitico***, i valori della **Y** possono essere determinati ***senza errore*** a partire dai soli valori della **X**;
- Da un ***punto di vista grafico***, la dipendenza funzionale implica la definizione di una funzione che passi ***per*** tutti i punti, e che quindi richiede la determinazione di tanti parametri quanti sono i punti.



Dipendenza statistica: $Y = f(X;\theta) + e$

- Il valore della variabile dipendente ***non è univocamente determinato a partire dal solo valore della variabile esplicativa***, potendosi osservare, per ciascun di **X**, più valori di **Y**;
- Da un ***punto di vista grafico***, la dipendenza statistica implica una funzione che passi ***fra*** i punti osservati. Il numero di parametri da determinare dipende, in questo caso, dal tipo di funzione scelta e ***non*** dal numero di punti osservati.

La Regressione

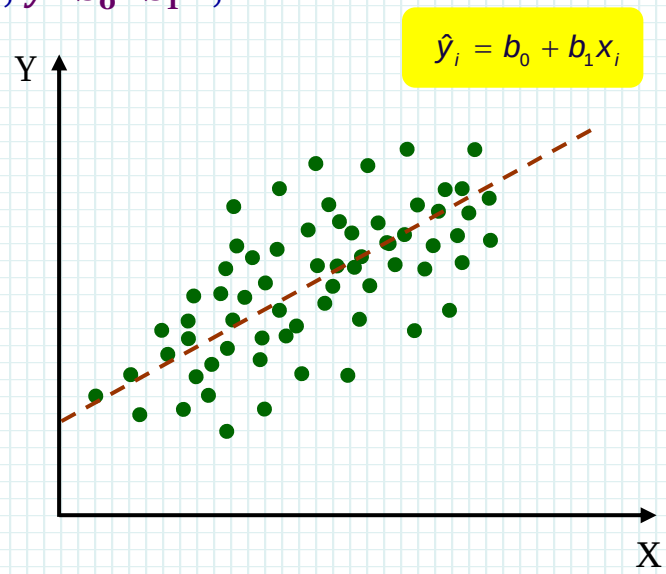
$$Y = f(X;\theta) + e$$

Riepilogo

X ⟶ Variabile indipendente

Y ⟶ Variabile dipendente

- **Decidiamo di rappresentare la nube di punti con una *funzione* che passi *tra* i punti stessi;**
- **Tra tutte le possibili funzioni, scegliamo la funzione lineare, $y=b_0+b_1x$;**
- **Tra tutte le infinite possibili rette, scegliamo quella che ottimizza un criterio che definiamo arbitrariamente, per esempio *quella che minimizza la somma dei quadrati degli scarti tra valori osservati e valori teorici*:**



$$S = \sum_i \left(y_i - \hat{y}_i\right)^2 = \sum_i \left(y_i - b_0 - b_1 x\right)^2 = min$$

- **Il *metodo dei minimi quadrati* ci consente di ottenere le soluzioni di questo problema, soluzioni che rappresentano i parametri della retta:**

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

- **Sostituendo questi valori nell'equazione $\hat{y}_i = b_0 + b_1 x_i$, per ogni valore dato di *X* otterremo il corrispondente valore teorico di *Y*.**

La Regressione $Y = f(X;\theta) + e$

X ⟶ Variabile indipendente

Y ⟶ Variabile dipendente

La scelta del tipo di funzione: ⟶ *Lineare* $\hat{y}_i = b_0 + b_1 x_i$

$b_0 = \bar{y} - b_1 \bar{x}$

$\bar{y} = b_0 + b_1 \bar{x}$

E' l'*intercetta* sull'asse delle ordinate. Può essere interpretato come il valore di Y per X=0 (quando ciò ha senso).

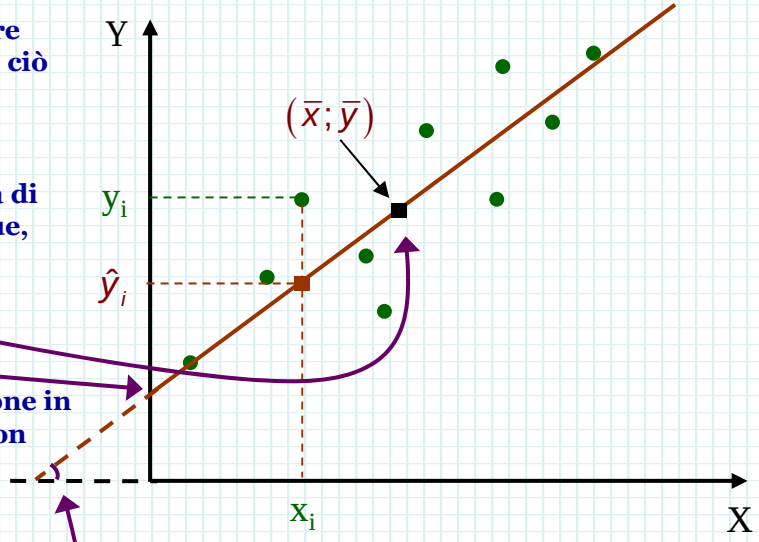
Il punto di coordinate $(\bar{x};\bar{y})$ è un punto della retta di regressione. La retta di regressione passa, dunque, sempre per il baricentro della nube.

$b_1 = \dfrac{Cov(XY)}{Var(X)}$

E' il *coefficiente angolare* della retta di regressione in quanto funzione dell'angolo che la retta forma con l'asse delle ascisse.

Esprime dunque la *pendenza* (positiva, negativa o nulla) della retta.

Esprime anche quanto varia la variabile Y al variare unitario della variabile X.

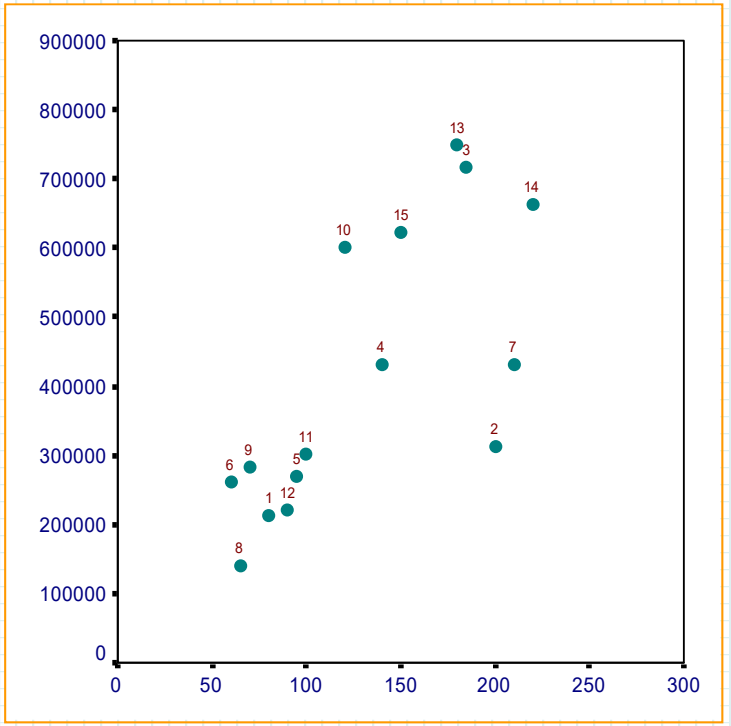


La Regressione

$$Y = f(X;\theta) + e$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

App.	mq (X)	Prezzo in € (Y)
1	80	212000
2	200	313000
3	185	717000
4	140	431000
5	95	270000
6	60	261000
7	210	431000
8	65	140000
9	70	282000
10	120	600000
11	100	303000
12	90	220000
13	180	749000
14	220	663000
15	150	623000
	1.965	*6.215.000*



μ_X=131,0 μ_Y=414.333,3

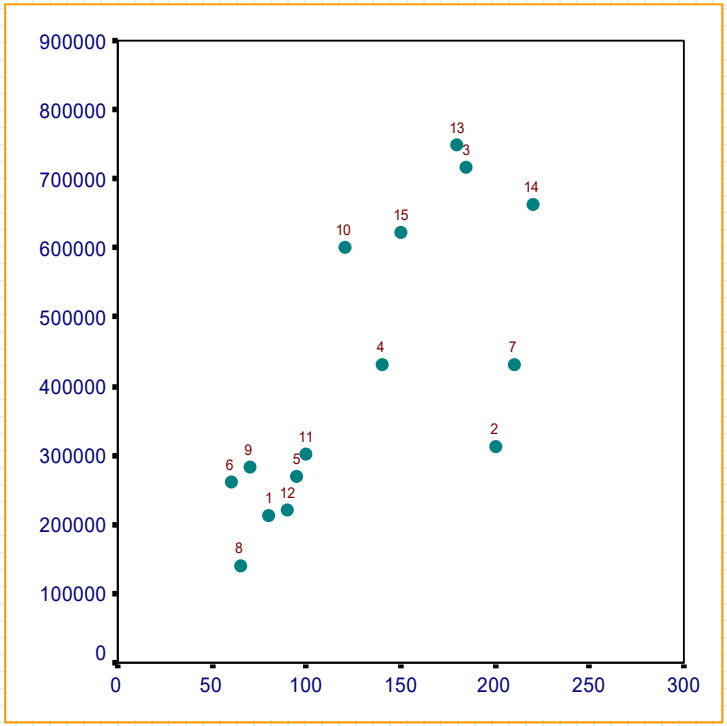
σ_X=54,44 σ_Y=197.060,96

La Regressione

$$Y = f(X;\theta) + e$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

App.	mq (X)	Prezzo in € (Y)	X-M(X)	Y-M(Y)	[X-M(X)]*[Y-M(Y)]
1	80	212000	-51,0	-202333,3	10.319.000
2	200	313000	69,0	-101333,3	-6.992.000
3	185	717000	54,0	302666,7	16.344.000
4	140	431000	9,0	16666,7	150.000
5	95	270000	-36,0	-144333,3	5.196.000
6	60	261000	-71,0	-153333,3	10.886.667
7	210	431000	79,0	16666,7	1.316.667
8	65	140000	-66,0	-274333,3	18.106.000
9	70	282000	-61,0	-132333,3	8.072.333
10	120	600000	-11,0	185666,7	-2.042.333
11	100	303000	-31,0	-111333,3	3.451.333
12	90	220000	-41,0	-194333,3	7.967.667
13	180	749000	49,0	334666,7	16.398.667
14	220	663000	89,0	248666,7	22.131.333
15	150	623000	19,0	208666,7	3.964.667
	1.965	*6.215.000*	*0,0*	*0,0*	*115.270.000*



μ_X=131,0 μ_Y=414.333,3

σ_X=54,44 σ_Y=197.060,96

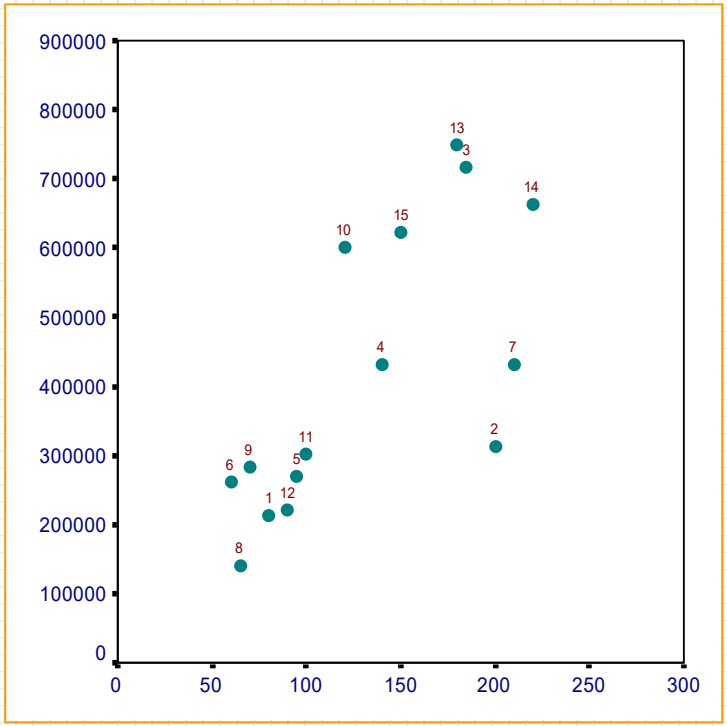
$$Cov(XY) = \frac{\sum_i (x_i - \mu_X)\cdot(y_i - \mu_Y)}{n} = \frac{115.270.000}{15} = 7.684.666,7$$

La Regressione

$$Y = f(X;\theta) + e$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

App.	mq (X)	Prezzo in € (Y)	X-M(X)	Y-M(Y)	[X-M(X)]*[Y-M(Y)]
1	80	212000	-51,0	-202333,3	10.319.000
2	200	313000	69,0	-101333,3	-6.992.000
3	185	717000	54,0	302666,7	16.344.000
4	140	431000	9,0	16666,7	150.000
5	95	270000	-36,0	-144333,3	5.196.000
6	60	261000	-71,0	-153333,3	10.886.667
7	210	431000	79,0	16666,7	1.316.667
8	65	140000	-66,0	-274333,3	18.106.000
9	70	282000	-61,0	-132333,3	8.072.333
10	120	600000	-11,0	185666,7	-2.042.333
11	100	303000	-31,0	-111333,3	3.451.333
12	90	220000	-41,0	-194333,3	7.967.667
13	180	749000	49,0	334666,7	16.398.667
14	220	663000	89,0	248666,7	22.131.333
15	150	623000	19,0	208666,7	3.964.667
	1.965	*6.215.000*	*0,0*	*0,0*	*115.270.000*



μ_X=131,0 μ_Y=414.333,3

σ_X=54,44 σ_Y=197.060,96

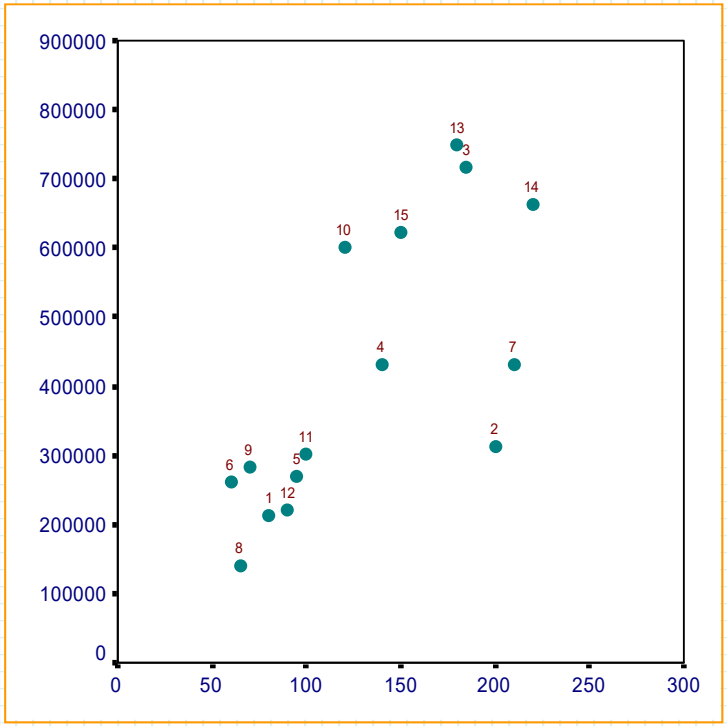
$$b_1 = \frac{\text{cov}(XY)}{Var(X)} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{7.684.666,7}{2964,00} = 2.592,67$$

La Regressione

$$Y = f(X;\theta) + e$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

App.	mq (X)	Prezzo in € (Y)	X-M(X)	Y-M(Y)	[X-M(X)]*[Y-M(Y)]
1	80	212000	-51,0	-202333,3	10.319.000
2	200	313000	69,0	-101333,3	-6.992.000
3	185	717000	54,0	302666,7	16.344.000
4	140	431000	9,0	16666,7	150.000
5	95	270000	-36,0	-144333,3	5.196.000
6	60	261000	-71,0	-153333,3	10.886.667
7	210	431000	79,0	16666,7	1.316.667
8	65	140000	-66,0	-274333,3	18.106.000
9	70	282000	-61,0	-132333,3	8.072.333
10	120	600000	-11,0	185666,7	-2.042.333
11	100	303000	-31,0	-111333,3	3.451.333
12	90	220000	-41,0	-194333,3	7.967.667
13	180	749000	49,0	334666,7	16.398.667
14	220	663000	89,0	248666,7	22.131.333
15	150	623000	19,0	208666,7	3.964.667
	1.965	*6.215.000*	*0,0*	*0,0*	*115.270.000*



μ_X=131,0 μ_Y=414.333,3

σ_X=54,44 σ_Y=197.060,96

$$b_0 = \bar{y} - b_1\bar{x} = 414.333,3 - 2.592,67 \times 131 = 74.693,5$$

La Regressione

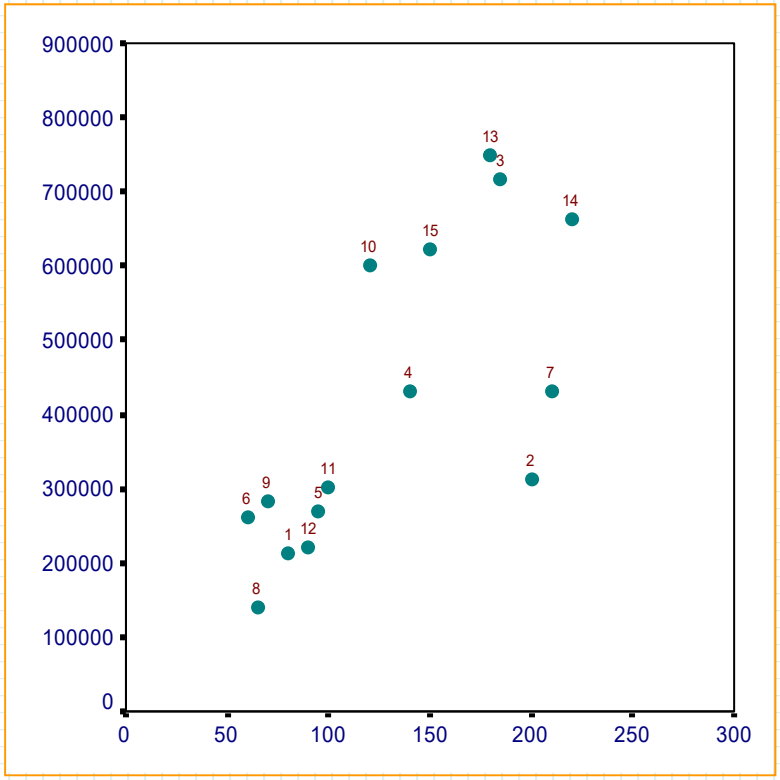
$$Y = f(X;\theta) + e$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

$$\hat{y} = b_0 + b_1 x$$

$$\hat{y} = 74.693,88 + 2.592,67x$$

App.	mq (X)	Prezzo in € (Y)	$\hat{Y}$	$Y - \hat{Y}$
1	80	212.000	282.107	-70.107
2	200	313.000	593.227	-280.227
3	185	717.000	554.337	162.663
4	140	431.000	437.667	-6.667
5	95	270.000	320.997	-50.997
6	60	261.000	230.254	30.746
7	210	431.000	619.154	-188.154
8	65	140.000	243.217	-103.217
9	70	282.000	256.181	25.819
10	120	600.000	385.814	214.186
11	100	303.000	333.961	-30.961
12	90	220.000	308.034	-88.034
13	180	749.000	541.374	207.626
14	220	663.000	645.081	17.919
15	150	623.000	463.594	159.406
	1.965	6.215.000	6.215.000	0



μ_X=131,0 μ_Y=414.333,3

σ_X=54,44 σ_Y=197.060,96

La Regressione

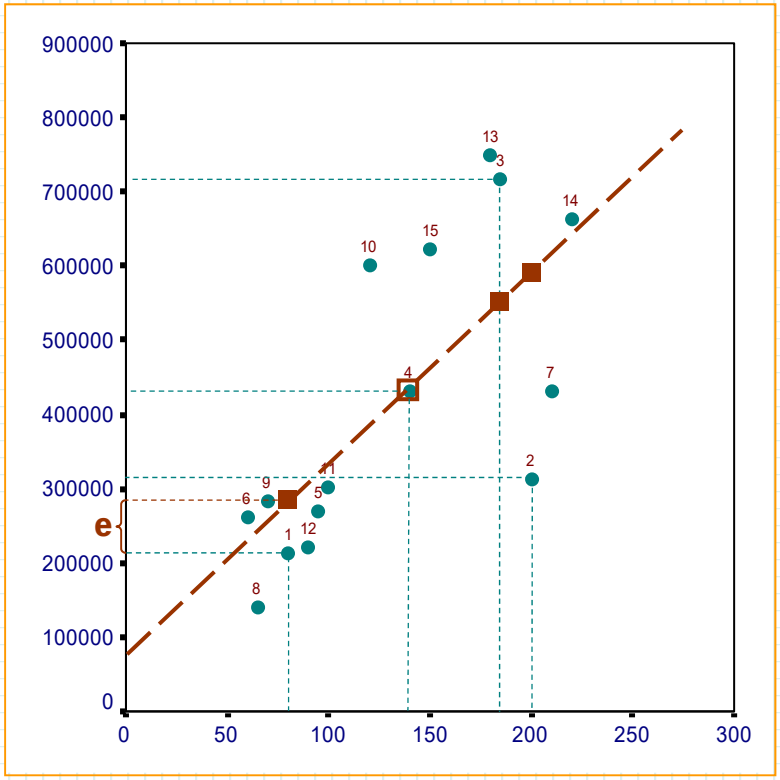
$$Y = f(X;\theta) + e$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

$$\hat{y} = b_0 + b_1 x$$

$$\hat{y} = 74.693,88 + 2.592,67x$$

App.	mq (X)	Prezzo in € (Y)	$\hat{Y}$	$Y - \hat{Y}$
1	80	212.000	282.107	-70.107
2	200	313.000	593.227	-280.227
3	185	717.000	554.337	162.663
4	140	431.000	437.667	-6.667
5	95	270.000	320.997	-50.997
6	60	261.000	230.254	30.746
7	210	431.000	619.154	-188.154
8	65	140.000	243.217	-103.217
9	70	282.000	256.181	25.819
10	120	600.000	385.814	214.186
11	100	303.000	333.961	-30.961
12	90	220.000	308.034	-88.034
13	180	749.000	541.374	207.626
14	220	663.000	645.081	17.919
15	150	623.000	463.594	159.406
	1.965	6.215.000	6.215.000	0



μ_X=131,0 μ_Y=414.333,3

σ_X=54,44 σ_Y=197.060,96

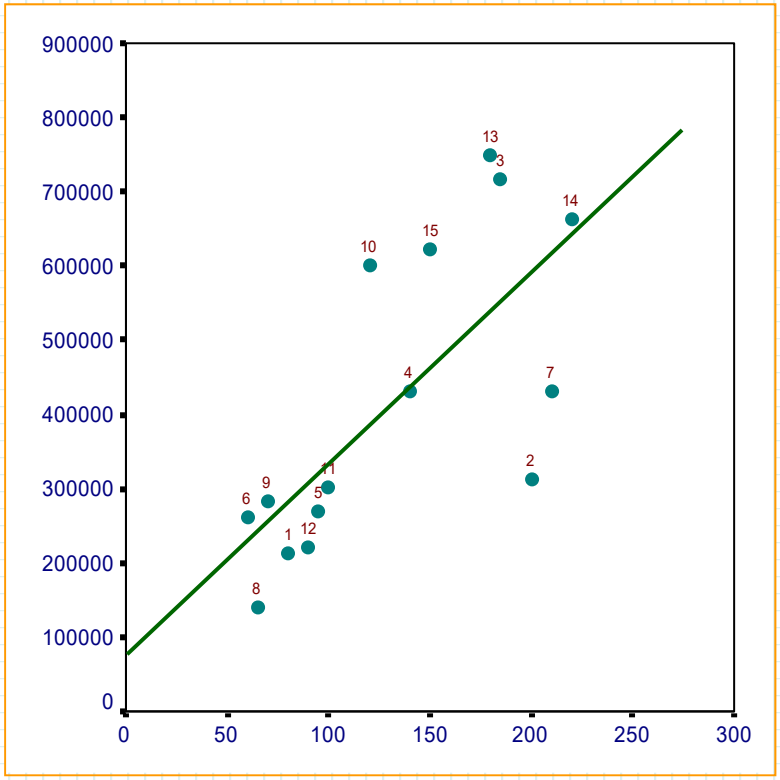
La Regressione: Interpolazione ed Estrapolazione

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

$$\hat{y} = b_0 + b_1 x$$

$$\hat{y} = 74.693,88 + 2.592,67x$$

App.	mq (X)	Prezzo in € (Y)	$\hat{Y}$	$Y - \hat{Y}$
1	80	212.000	282.107	-70.107
2	200	313.000	593.227	-280.227
3	185	717.000	554.337	162.663
4	140	431.000	437.667	-6.667
5	95	270.000	320.997	-50.997
6	60	261.000	230.254	30.746
7	210	431.000	619.154	-188.154
8	65	140.000	243.217	-103.217
9	70	282.000	256.181	25.819
10	120	600.000	385.814	214.186
11	100	303.000	333.961	-30.961
12	90	220.000	308.034	-88.034
13	180	749.000	541.374	207.626
14	220	663.000	645.081	17.919
15	150	623.000	463.594	159.406
	1.965	6.215.000	6.215.000	0



Qual è il prezzo previsto
di un appartamento di 160 mq?

$$\hat{y} = 74.693,88 + 2.592,67 \cdot 160$$
$$= 489.520,7$$

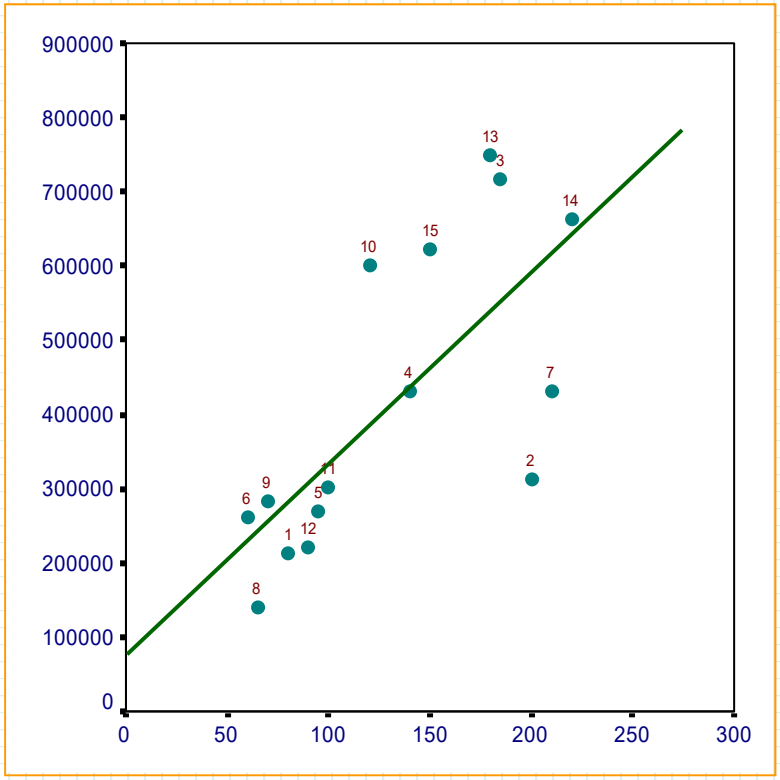
La Regressione: Interpolazione ed Estrapolazione

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = \frac{Cov(XY)}{Var(X)}$$

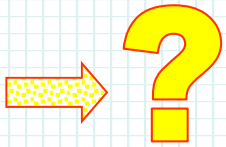
$$\hat{y} = b_0 + b_1 x$$

$$\hat{y} = 74.693,88 + 2.592,67x$$

App.	mq (X)	Prezzo in € (Y)	$\hat{Y}$	$Y - \hat{Y}$
1	80	212.000	282.107	-70.107
2	200	313.000	593.227	-280.227
3	185	717.000	554.337	162.663
4	140	431.000	437.667	-6.667
5	95	270.000	320.997	-50.997
6	60	261.000	230.254	30.746
7	210	431.000	619.154	-188.154
8	65	140.000	243.217	-103.217
9	70	282.000	256.181	25.819
10	120	600.000	385.814	214.186
11	100	303.000	333.961	-30.961
12	90	220.000	308.034	-88.034
13	180	749.000	541.374	207.626
14	220	663.000	645.081	17.919
15	150	623.000	463.594	159.406
	1.965	6.215.000	6.215.000	0

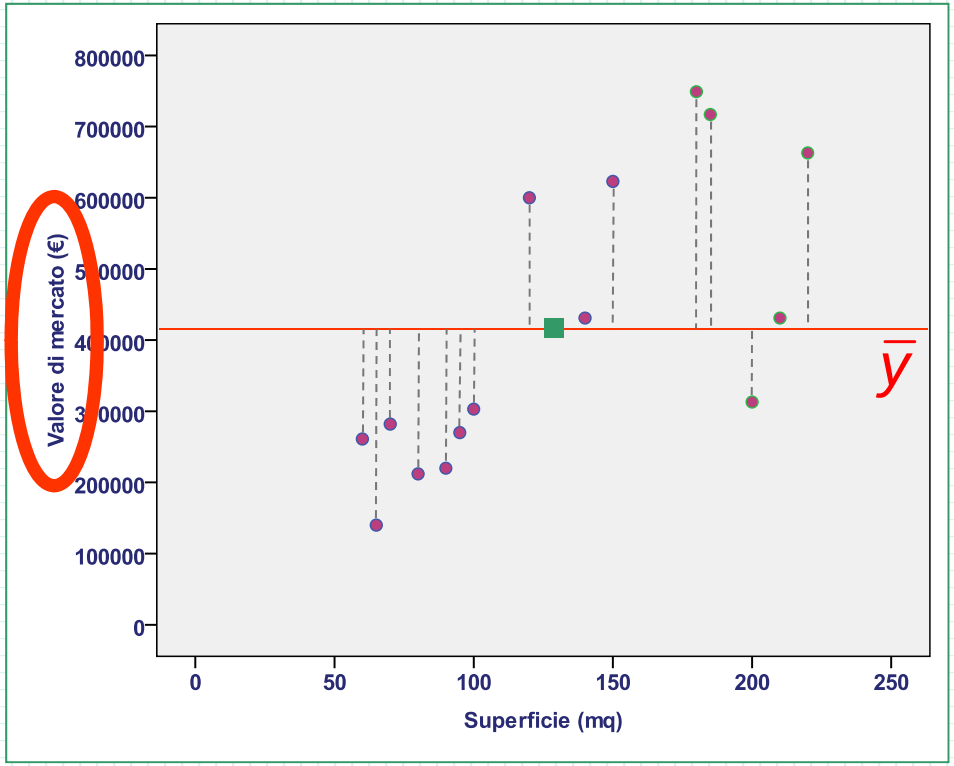


Qual è il prezzo previsto
di un appartamento di 260 mq?



Se il valore della X è ***esterno*** all'intervallo dei valori considerati, il valore della Y ***non può essere previsto*** applicando la funzione di regressione. In questo caso si tratterebbe di "*estrapolazione*" e non di "interpolazione".

La Regressione: Valutazione dell' adattamento



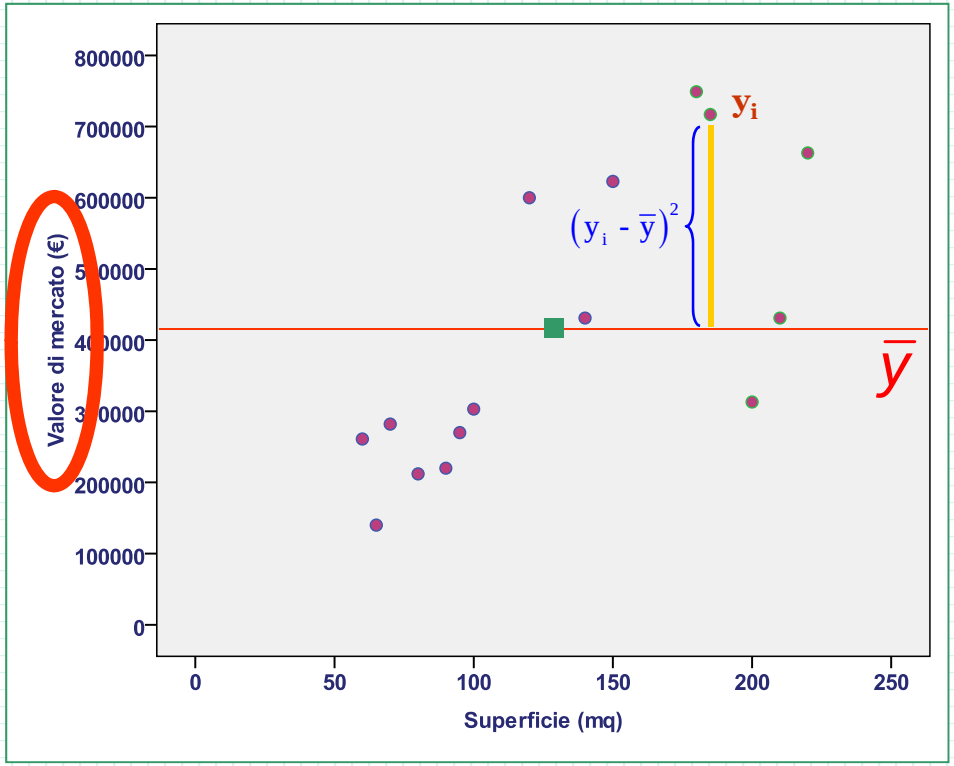
• Se consideriamo **la sola variabile y**, la previsione più attendibile per y è data dalla media;

• l' **errore complessivo di previsione** sarà dato dalla somma delle distanze tra valori y osservati e valori y teorici (che coincidono, in questo caso, con il valore medio);

• tale errore è dunque pari alla **devianza di y**:

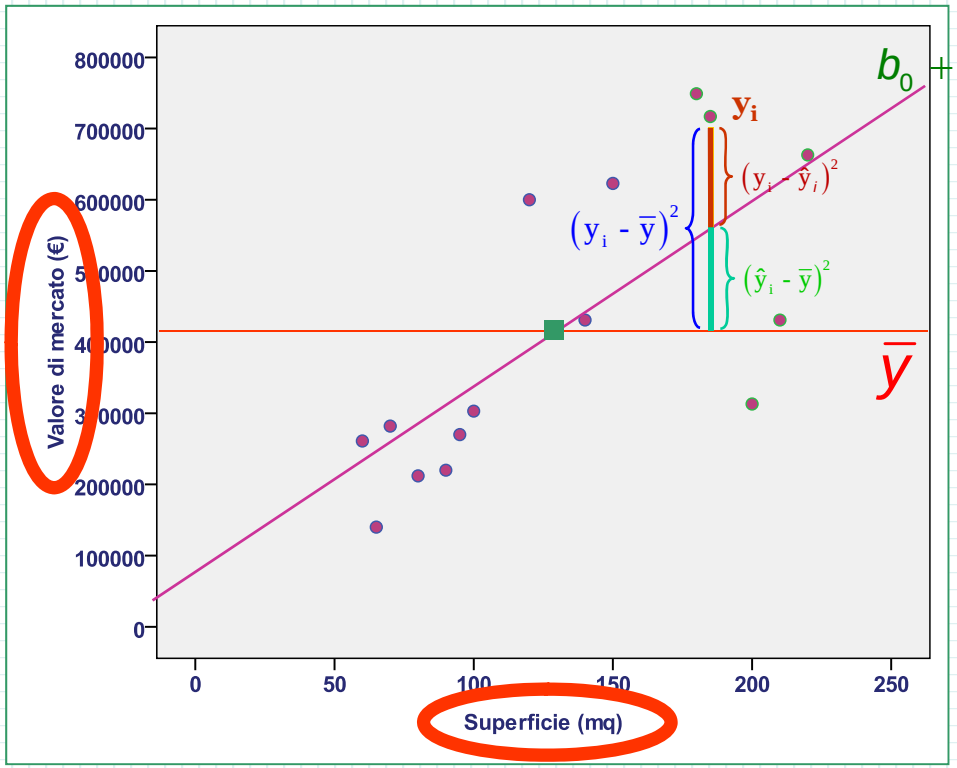
$$Dev(Y) = \sum_i (y_i - \bar{y})^2$$

La Regressione: Valutazione dell' adattamento



• Il generico punto y_i partecipa all' **errore complessivo** (devianza di y) con la quantità $(y_i - \bar{y})^2$

La Regressione: Valutazione dell'adattamento



• Il generico punto y_i partecipa all'**errore complessivo** (devianza di y) con la quantità $(y_i - \bar{y})^2$

• Quando consideriamo anche **la variabile X,** ci aspettiamo che l'errore di previsione si riduca;

• La **retta interpolante** si dispone nella direzione di "allungamento" della nube di punti, facendo "perno" sul baricentro della nube, il punto medio, per il quale le retta di regressione deve passare;

• Considerando ancora il generico punto y_i, una parte dell'errore $(y_i - \bar{y})^2$ viene ora eliminata, "spiegata" dalla retta di regressione.

• La **parte "spiegata"** è pari alla quantità:

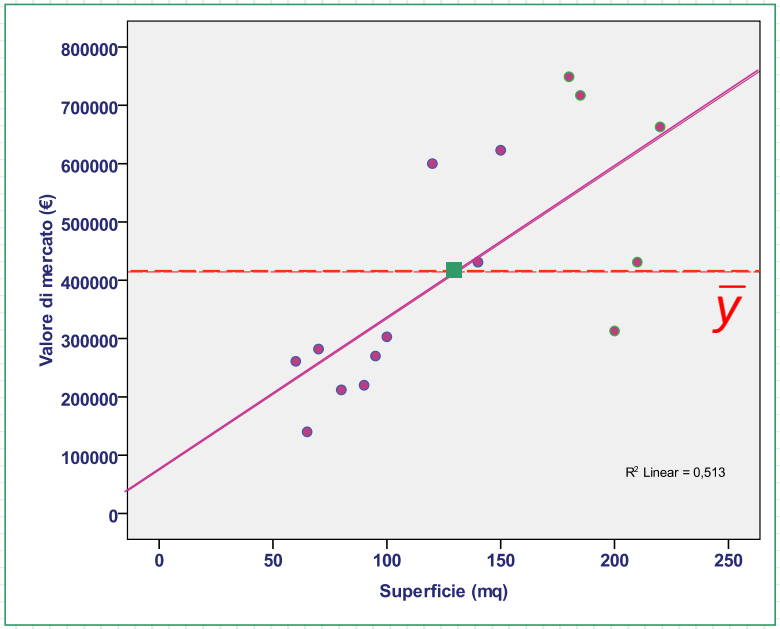
$$(\hat{y}_i - \bar{y})^2$$

• La **parte di errore "residua"**, che rimane anche dopo la costruzione della retta di regressione, è la parte di errore che non viene spiegata nemmeno dopo l'osservazione della variabile X;

• La **parte "residua"** risulta pari a:

$$(y_i - \hat{y}_i)^2$$

La Regressione: Valutazione dell'adattamento



Una misura della **bontà dell'adattamento della retta di regressione ai dati** può quindi essere data dal **rapporto tra la devianza spiegata e la devianza totale**.

R^2 → **Indice di determinazione**

$$R^2 = \frac{Dev(\hat{y})}{Dev(y)} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \qquad 0 \le R^2 \le 1$$

• Quando è **$R^2=0$**, **la devianza spiegata è pari a zero**. Questo vuol dire che **l'osservazione della variabile X non ha aggiunto nulla a quanto già si sapeva dalla sola osservazione della Y**. Dal punto di vista geometrico, la retta di regressione coincide con la retta M(Y); dal punto di vista interpretativo, le variabili X e Y sono incorrelate;

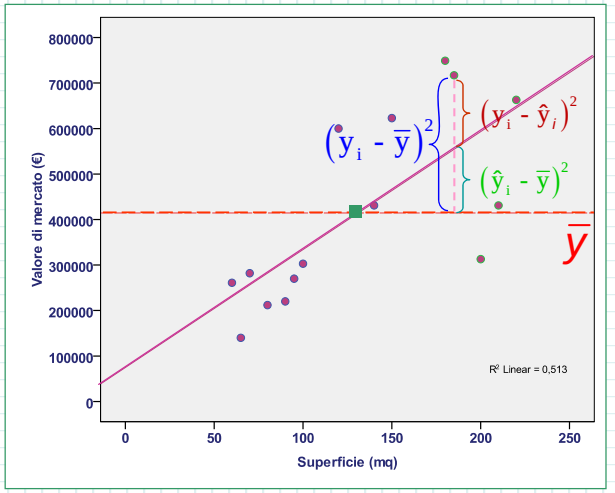
• Quando è **$R^2=1$**, **la devianza spiegata è uguale alla devianza totale**. Questo vuol dire che **l'osservazione della variabile X spiega perfettamente la variabile Y**, e ne rende possibile la previsione senza possibilità di errore. Dal punto di vista geometrico, tutti i punti sono allineati e la retta di regressione passa per tutti i punti (siamo quindi nel caso di una dipendenza funzionale, deterministica, esatta); dal punto di vista interpretativo, le variabili X e Y sono massimamente correlate.

• Quando è **$0 \le R^2 \le 1$**, **la devianza spiegata è pari a una quota della devianza totale**. L'osservazione della variabile X migliora quindi la previsione della variabile Y, con una quota di errore residua dovuta in parte alle variabili non osservate, in parte alla sempre presente quota di imponderabilità dei fenomeni osservati.

La Regressione

$$\hat{y} = 74.693,88 + 2.592,67x$$

App.	mq (X)	Prezzo in € (Y)	$\hat{Y}$	$(Y - \bar{Y})^2$	$(\hat{Y} - \bar{Y})^2$	$(Y - \hat{Y})^2$
1	80	212.000	282.107	40.938.777.778	17.483.727.210	4.915.031.754
2	200	313.000	593.227	10.268.444.444	32.003.085.447	78.527.393.139
3	185	717.000	554.337	91.607.111.111	19.601.133.620	26.459.127.322
4	140	431.000	437.667	277.777.778	544.475.934	44.453.442
5	95	270.000	320.997	20.832.111.111	8.711.614.942	2.600.724.704
6	60	261.000	230.254	23.511.111.111	33.885.224.478	945.320.444
7	210	431.000	619.154	277.777.778	41.951.534.609	35.401.954.462
8	65	140.000	243.217	75.258.777.778	29.280.705.777	10.653.805.643
9	70	282.000	256.181	17.512.111.111	25.012.283.333	666.640.808
10	120	600.000	385.814	34.472.111.111	813.352.938	45.875.646.835
11	100	303.000	333.961	12.395.111.111	6.459.770.030	958.561.153
12	90	220.000	308.034	37.765.444.444	11.299.556.109	7.749.978.661
13	180	749.000	541.374	112.001.777.778	16.139.342.188	43.108.537.570
14	220	663.000	645.081	61.835.111.111	53.244.368.793	321.099.637
15	150	623.000	463.594	43.541.777.778	2.426.614.965	25.410.267.386
	1.965	6.215.000	6.215.000	582.495.333.333	298.856.790.373	283.638.542.960



$$Dev(Y) = \sum_i (y_i - \bar{y})^2 = 582.495.333.333$$

$$Dev(\hat{Y}) = \sum_i (\hat{y}_i - \bar{y})^2 = 298.856.790.373$$

$$Dev(e) = \sum_i (y_i - \hat{y})^2 = 283.638.542.960$$

$$R^2 = \frac{Dev(\hat{Y})}{Dev(Y)} = \frac{298.856.790.373}{582.495.333.333} = 0,513$$

Regressione e Correlazione

a. Con *Y* variabile dipendente: $b_0^{(y)} = \bar{y} - b_1^{(y)}\bar{x}$; $b_1^{(y)} = \frac{Cov(XY)}{Var(X)}$

b. Con *X* variabile dipendente: $b_0^{(x)} = \bar{x} - b_1^{(x)}\bar{y}$; $b_1^{(x)} = \frac{Cov(XY)}{Var(Y)}$

Indice di interdipendenza

$$\sqrt{b_1^{(y)} \times b_1^{(x)}} = \sqrt{\frac{Cov(XY)}{Var(X)} \times \frac{Cov(XY)}{Var(Y)}}$$

$$= \sqrt{\frac{\left[Cov(XY)\right]^2}{Var(X) \cdot Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = r \rightarrow$$ **Coefficiente di correlazione**

Il coefficiente di correlazione *r*, indice simmetrico di interdipendenza, può dunque essere considerato come una media (geometrica) dei due indici asimmetrici di dipendenza rappresentati dai coefficienti di regressione.

Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)
Closer	5.611,4	42.340,1
Birth - Io sono Sean	3.933,8	25.420,2
Saw	5.161,9	34.475,7
The Aviator	5.874,6	40.150,1
Neverland - Un sogno per la vita	5.856,7	45.063,8
Il mercante di Venezia	3.845,5	20.419,9
Million dollar baby	5.643,5	36.129,3
Shark tale	7.655,2	57.472,3
Constantine	5.044,2	25.334,4
Cuore sacro	2.915,4	18.279,8

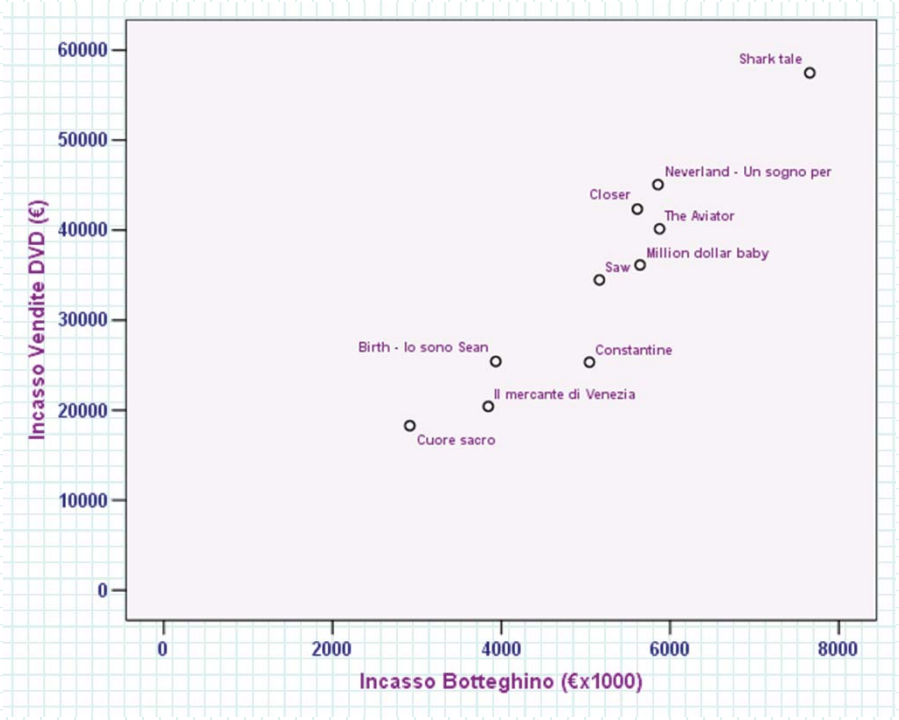
Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)
Closer	5.611,4	42.340,1
Birth - Io sono Sean	3.933,8	25.420,2
Saw	5.161,9	34.475,7
The Aviator	5.874,6	40.150,1
Neverland - Un sogno per la vita	5.856,7	45.063,8
Il mercante di Venezia	3.845,5	20.419,9
Million dollar baby	5.643,5	36.129,3
Shark tale	7.655,2	57.472,3
Constantine	5.044,2	25.334,4
Cuore sacro	2.915,4	18.279,8
	51.542,4	*345.085,6*



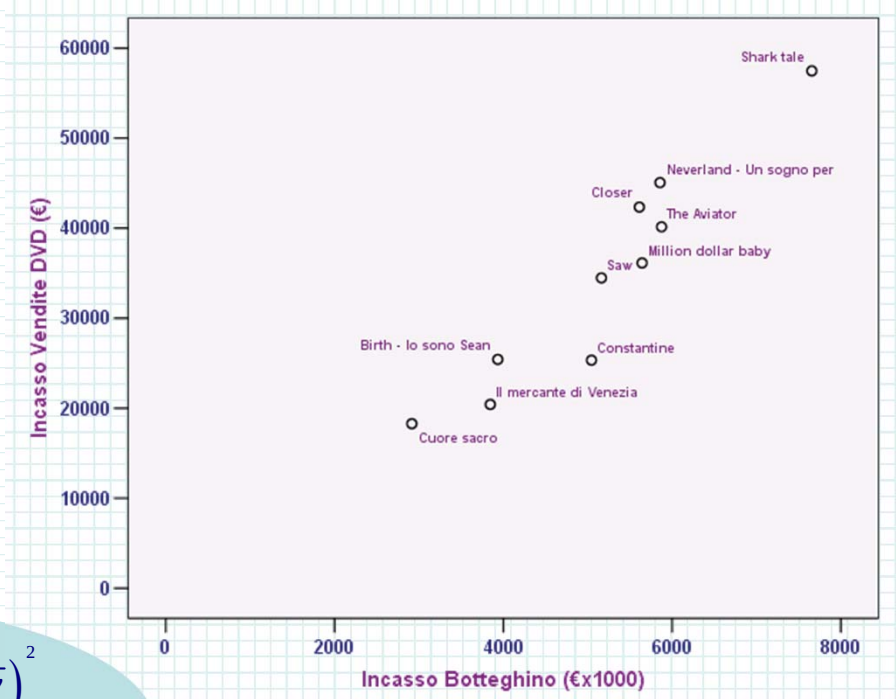
Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)
Closer	5.611,4	42.340,1
Birth - Io sono Sean	3.933,8	25.420,2
Saw	5.161,9	34.475,7
The Aviator	5.874,6	40.150,1
Neverland - Un sogno per la vita	5.856,7	45.063,8
Il mercante di Venezia	3.845,5	20.419,9
Million dollar baby	5.643,5	36.129,3
Shark tale	7.655,2	57.472,3
Constantine	5.044,2	25.334,4
Cuore sacro	2.915,4	18.279,8
	51.542,4	*345.085,6*



$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{Cov(XY)}{Var(X)}$$

$$R^2 = \frac{Dev(Reg)}{Dev(Y)} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)	X^2	Y^2	XY
Closer	5.611,4	42.340,1	31.488.072,5	1.792.685.568,3	237.588.327,1
Birth - Io sono Sean	3.933,8	25.420,2	15.475.005,1	646.187.034,1	99.998.738,1
Saw	5.161,9	34.475,7	26.645.676,3	1.188.574.428,1	177.961.707,9
The Aviator	5.874,6	40.150,1	34.510.996,2	1.612.032.742,4	235.866.182,2
Neverland - Un sogno per la vita	5.856,7	45.063,8	34.300.392,6	2.030.750.081,2	263.923.331,8
Il mercante di Venezia	3.845,5	20.419,9	14.787.775,5	416.970.686,6	78.524.320,5
Million dollar baby	5.643,5	36.129,3	31.849.411,4	1.305.323.576,6	203.896.511,8
Shark tale	7.655,2	57.472,3	58.602.588,5	3.303.060.113,5	439.963.490,0
Constantine	5.044,2	25.334,4	25.444.193,9	641.830.724,6	127.792.274,6
Cuore sacro	2.915,4	18.279,8	8.499.683,6	334.151.139,3	53.293.329,4
	51.542,4	*345.085,6*	*281.603.795,7*	*13.271.566.094,6*	*1.918.808.213,5*

$$M(X)=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad Var(X)=M(X^2)-[M(X)]^2 \qquad Cov(XY)=M(XY)-M(X)\cdot M(Y) \qquad b_0=\bar{y}-b_1\bar{x}$$

$$M(Y)=\frac{1}{n}\sum_{i=1}^{n}Y_i \qquad Var(Y)=M(Y^2)-[M(Y)]^2 \qquad r(XY)=\frac{Cov(XY)}{\sigma_X\cdot\sigma_Y} \qquad b_1=\frac{Cov(XY)}{Var(X)}$$

Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)	X^2	Y^2	XY
Closer	5.611,4	42.340,1	31.488.072,5	1.792.685.568,3	237.588.327,1
Birth - Io sono Sean	3.933,8	25.420,2	15.475.005,1	646.187.034,1	99.998.738,1
Saw	5.161,9	34.475,7	26.645.676,3	1.188.574.428,1	177.961.707,9
The Aviator	5.874,6	40.150,1	34.510.996,2	1.612.032.742,4	235.866.182,2
Neverland - Un sogno per la vita	5.856,7	45.063,8	34.300.392,6	2.030.750.081,2	263.923.331,8
Il mercante di Venezia	3.845,5	20.419,9	14.787.775,5	416.970.686,6	78.524.320,5
Million dollar baby	5.643,5	36.129,3	31.849.411,4	1.305.323.576,6	203.896.511,8
Shark tale	7.655,2	57.472,3	58.602.588,5	3.303.060.113,5	439.963.490,0
Constantine	5.044,2	25.334,4	25.444.193,9	641.830.724,6	127.792.274,6
Cuore sacro	2.915,4	18.279,8	8.499.683,6	334.151.139,3	53.293.329,4
	51.542,4	*345.085,6*	*281.603.795,7*	*13.271.566.094,6*	*1.918.808.213,5*

M(X) = 5.154,2

M(Y) = 34.508,6

Var(X) = 1.594.602,0

Var(Y) = 136.136.146,2

Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)	X^2	Y^2	XY
Closer	5.611,4	42.340,1	31.488.072,5	1.792.685.568,3	237.588.327,1
Birth - Io sono Sean	3.933,8	25.420,2	15.475.005,1	646.187.034,1	99.998.738,1
Saw	5.161,9	34.475,7	26.645.676,3	1.188.574.428,1	177.961.707,9
The Aviator	5.874,6	40.150,1	34.510.996,2	1.612.032.742,4	235.866.182,2
Neverland - Un sogno per la vita	5.856,7	45.063,8	34.300.392,6	2.030.750.081,2	263.923.331,8
Il mercante di Venezia	3.845,5	20.419,9	14.787.775,5	416.970.686,6	78.524.320,5
Million dollar baby	5.643,5	36.129,3	31.849.411,4	1.305.323.576,6	203.896.511,8
Shark tale	7.655,2	57.472,3	58.602.588,5	3.303.060.113,5	439.963.490,0
Constantine	5.044,2	25.334,4	25.444.193,9	641.830.724,6	127.792.274,6
Cuore sacro	2.915,4	18.279,8	8.499.683,6	334.151.139,3	53.293.329,4
	51.542,4	*345.085,6*	*281.603.795,7*	*13.271.566.094,6*	*1.918.808.213,5*

M(X) = 5.154,2

M(Y) = 34.508,6

Sqm(X) = 1.262.8

Sqm(Y) = 11.667,7

$$Cov(XY) = M(XY) - [M(X) \cdot M(Y)]$$

$$= \frac{1.918.808.213,5}{10} - (5.145,2 \cdot 34.508,6) = 14.016.595,2$$

$$r_{XY} = \frac{Cov(XY)}{\sigma_X \cdot \sigma_Y} = \frac{14.016.595,2}{1.262,8 \cdot 11.667,7} = 0,95$$

Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)	X^2	Y^2	XY
Closer	5.611,4	42.340,1	31.488.072,5	1.792.685.568,3	237.588.327,1
Birth - Io sono Sean	3.933,8	25.420,2	15.475.005,1	646.187.034,1	99.998.738,1
Saw	5.161,9	34.475,7	26.645.676,3	1.188.574.428,1	177.961.707,9
The Aviator	5.874,6	40.150,1	34.510.996,2	1.612.032.742,4	235.866.182,2
Neverland - Un sogno per la vita	5.856,7	45.063,8	34.300.392,6	2.030.750.081,2	263.923.331,8
Il mercante di Venezia	3.845,5	20.419,9	14.787.775,5	416.970.686,6	78.524.320,5
Million dollar baby	5.643,5	36.129,3	31.849.411,4	1.305.323.576,6	203.896.511,8
Shark tale	7.655,2	57.472,3	58.602.588,5	3.303.060.113,5	439.963.490,0
Constantine	5.044,2	25.334,4	25.444.193,9	641.830.724,6	127.792.274,6
Cuore sacro	2.915,4	18.279,8	8.499.683,6	334.151.139,3	53.293.329,4
	51.542,4	*345.085,6*	*281.603.795,7*	*13.271.566.094,6*	*1.918.808.213,5*

M(X) = 5.154,2

M(Y) = 34.508,6

Sqm(X) = 1.262.8

Sqm(Y) = 11.667,7

Cov(XY) = 14.016.595,2

r_{XY} = 0,95

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{Cov(XY)}{Var(X)}$$

Esercizio 1

Il responsabile di un negozio operante nel settore dell'*home entertainment* ipotizza l'esistenza di una relazione tra gli incassi realizzati ai botteghini e quelli derivanti dalla vendita di DVD.

Considerando i seguenti 10 film (stagione 2004-'05), si valuti l'ipotesi del responsabile determinando:

- **Il diagramma di dispersione**
- **la retta di regressione e l'indice di determinazione lineare**

Film	Incasso al botteghino (€ x 1000)	Incasso vendite DVD (€)	X^2	Y^2	XY
Closer	5.611,4	42.340,1	31.488.072,5	1.792.685.568,3	237.588.327,1
Birth - Io sono Sean	3.933,8	25.420,2	15.475.005,1	646.187.034,1	99.998.738,1
Saw	5.161,9	34.475,7	26.645.676,3	1.188.574.428,1	177.961.707,9
The Aviator	5.874,6	40.150,1	34.510.996,2	1.612.032.742,4	235.866.182,2
Neverland - Un sogno per la vita	5.856,7	45.063,8	34.300.392,6	2.030.750.081,2	263.923.331,8
Il mercante di Venezia	3.845,5	20.419,9	14.787.775,5	416.970.686,6	78.524.320,5
Million dollar baby	5.643,5	36.129,3	31.849.411,4	1.305.323.576,6	203.896.511,8
Shark tale	7.655,2	57.472,3	58.602.588,5	3.303.060.113,5	439.963.490,0
Constantine	5.044,2	25.334,4	25.444.193,9	641.830.724,6	127.792.274,6
Cuore sacro	2.915,4	18.279,8	8.499.683,6	334.151.139,3	53.293.329,4
	51.542,4	*345.085,6*	*281.603.795,7*	*13.271.566.094,6*	*1.918.808.213,5*

M(X) = 5.154,2

M(Y) = 34.508,6

Sqm(X) = 1.262.8

Sqm(Y) = 11.667,7

Cov(XY) = 14.016.595,2

r_{XY} = 0,95

b_1 =8,79 **b_o =-10.796,8**

$$\hat{y} = -10.796,8 + 8,79\,X$$

$$R^2 = 0,95^2 = 0,902$$

Esercizio 2

300 studenti Esame di matematica **(X)** Esame di statistica **(Y)**

$\bar{x}=24,2$ $\bar{y}=26,9$

$r_{(XY)}=0,78$

$s_X=2,9$ $s_Y=2,4$

Qual è il voto previsto all'esame di statistica per uno studente che ha avuto 25 all'esame di matematica?

Esercizio 2

300 studenti Esame di matematica **(X)** Esame di statistica **(Y)**

$\bar{x}_X = 24{,}2$ $\bar{x}_Y = 26{,}9$

$r_{(XY)} = 0{,}78$

$s_X = 2{,}9$ $s_Y = 2{,}4$

Qual è il voto previsto all'esame di statistica per uno studente che ha avuto 25 all'esame di matematica?

$$\hat{y} = b_0 + b_1 x \qquad b_1 = \frac{Cov(XY)}{Var(X)} \qquad b_0 = \bar{y} - b_1 \bar{x}$$

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \Rightarrow \sigma_{XY} = r \cdot \sigma_X \cdot \sigma_Y = 0{,}78 \times 2{,}9 \times 2{,}4 = 5{,}4288$$

$$b_1 = \frac{Cov(XY)}{Var(X)} = \frac{5{,}4288}{8{,}41} = 0{,}646 \qquad b_0 = \bar{y} - b_1 \bar{x} = 26{,}9 - 0{,}646 \times 24{,}2 = 11{,}27$$

$$\hat{y} = b_0 + b_1 x = 11{,}27 + 0{,}646 x$$

per X=25 ➡ $\hat{y} = 11{,}27 + 0{,}646 \cdot 25 = 27{,}42$