

RENATO LEONI

**Alcune alternative al metodo classico
nello studio della regressione lineare bivariata**

Università di Firenze
Dipartimento di Statistica, Informatica, Applicazioni "Giuseppe Parenti"
Firenze, novembre 2015

Eventuali commenti e osservazioni critiche possono essere inviati al seguente indirizzo: leoni@disia.unifi.it

Cap. 1 INTRODUZIONE

Ponendoci da un punto di vista che possiamo chiamare «geometrico-descrittivo», il procedimento usualmente impiegato nello studio della regressione lineare tra due variabili X e Y – procedimento noto come *metodo classico* o *dei minimi quadrati ordinari* – può essere riassunto nei seguenti termini ⁽¹⁾.

Supponiamo che \mathbb{R}^2 sia munito della metrica euclidea standard.

Conveniamo inoltre di rappresentare ciascun vettore di \mathbb{R}^2 (scritto come vettore colonna) mediante le sue coordinate rispetto alla base naturale formata dai vettori canonici

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Indicando poi con x_1, \dots, x_n e y_1, \dots, y_n i valori osservati delle variabili X e Y in n individui, sia ($i = 1, \dots, n$)

$$\mathbf{z}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}.$$

Consideriamo adesso in \mathbb{R}^2 la retta $r : ax + by = c$.

Qualora sia $b \neq 0$, possiamo riscrivere r nella forma $r_{Y/X} : y = c/b - a/bx = a_0 + a_1x$, e la *proiezione* di \mathbf{z}_i su r lungo il sottospazio U_2 generato dal vettore \mathbf{u}_2 (asse delle ordinate) è data dal vettore (Cfr. la Fig. 1)

$$\widehat{\mathbf{z}}_i = \begin{bmatrix} x_i \\ \widehat{y}_i \end{bmatrix}$$

dove $\widehat{y}_i = a_0 + a_1x_i$.

A sua volta, qualora sia $a \neq 0$, possiamo riscrivere r nella forma $r_{X/Y} : x = c/a - b/a y = b_0 + b_1y$, e la *proiezione* di \mathbf{z}_i su r lungo il sotto-

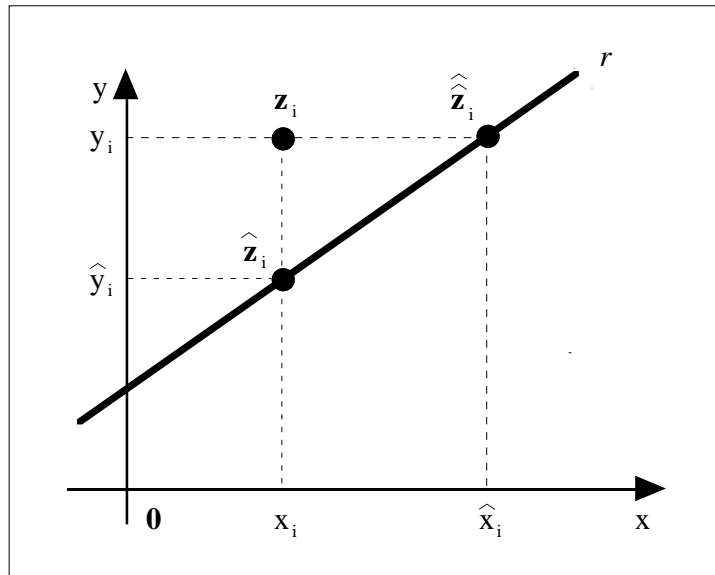
(1) La locuzione «regressione lineare» va intesa, secondo l'uso comune, come sinonimo di «relazione lineare». Nel seguito, ai fini di una maggiore immediatezza espressiva, ometteremo talvolta l'aggettivo «lineare».

spazio U_1 generato dal vettore \mathbf{u}_1 (asse delle ascisse) è data dal vettore (Cfr. la Fig. 1)

$$\hat{\mathbf{z}}_i = \begin{bmatrix} \hat{x}_i \\ y_i \end{bmatrix}$$

dove $\hat{x}_i = b_0 + b_1 y_i$.

Fig. 1



A tali proiezioni sono poi associati i vettori, detti *residui*, definiti da

$$\delta_i = \mathbf{z}_i - \hat{\mathbf{z}}_i = \begin{bmatrix} 0 \\ y_i - a_0 - a_1 x_i \end{bmatrix}$$

e

$$\varepsilon_i = \mathbf{z}_i - \hat{\mathbf{z}}_i = \begin{bmatrix} x_i - b_0 - b_1 y_i \\ 0 \end{bmatrix}.$$

Ciò premesso, designando con $d_{\delta,i}$ e $d_{\varepsilon,i}$ le lunghezze di δ_i e ε_i , il valore dei parametri a_0, a_1 e b_0, b_1 è determinato rendendo minime le espressioni

$$(1) \quad \frac{1}{n} \sum_i d_{\delta,i}^2 = \frac{1}{n} \sum_i (y_i - a_0 - a_1 x_i)^2$$

e

$$(2) \quad \frac{1}{n} \sum_i d_{\varepsilon,i}^2 = \frac{1}{n} \sum_i (x_i - b_0 - b_1 y_i)^2 .$$

Pertanto, qualora \hat{a}_0, \hat{a}_1 e \hat{b}_0, \hat{b}_1 denotino il valore dei parametri determinato minimizzando le espressioni scritte in (1) e (2), risultano definite due *rette di regressione classica* che – posto, con ovvio significato dei simboli,

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_i x_i & , & \quad \bar{y} = \frac{1}{n} \sum_i y_i \\ \sigma_x &= \sqrt{\frac{1}{n} \sum_i (x_i - \bar{x})^2} & , & \quad \sigma_y = \sqrt{\frac{1}{n} \sum_i (y_i - \bar{y})^2} \\ \sigma_{xy} &= \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y}) & , & \quad \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \end{aligned}$$

– indichiamo con ⁽²⁾

$$(3) \quad \hat{r}_{Y/X} : y = \hat{a}_0 + \hat{a}_1 x \quad (\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}, \hat{a}_1 = \rho_{xy} \frac{\sigma_y}{\sigma_x})$$

e

$$(4) \quad \hat{r}_{X/Y} : x = \hat{b}_0 + \hat{b}_1 y \quad (\hat{b}_0 = \bar{x} - \hat{b}_1 \bar{y}, \hat{b}_1 = \rho_{xy} \frac{\sigma_x}{\sigma_y}) .$$

Ciascuna di tali rette descrive sinteticamente la relazione che sussiste tra X e Y.

Tuttavia, $\hat{r}_{Y/X}$ e $\hat{r}_{X/Y}$ – escludendo il caso in cui si verifichi perfetta correlazione lineare tra X e Y ($|\rho_{xy}| = 1$) – non sono coincidenti, tranne che per $x = \bar{x}$ e $y = \bar{y}$ (*baricentro* o *centroide*), e rappresentano dunque modi *diversi* per descrivere la relazione che intercorre tra X e Y.

Diviene quindi importante stabilire in quali circostanze sia lecito usare la (3) e in quali la (4).

(2) Utilizzando una metrica euclidea diversa da quella standard, otterremmo ovviamente risultati diversi, ma lo schema di ragionamento rimarrebbe immutato.

A questo fine, occorre anzitutto osservare che le variabili X e Y non costituiscono un sistema *chiuso*, ma un sistema *aperto* all'influenza di una molteplicità di altri fattori che possiamo chiamare *fattori di disturbo*.

Se si suppone che l'influenza dei fattori di disturbo si eserciti *solamente* sulla Y ma *non* sulla X , ciascun y_i si può interpretare come la risultante di due componenti:

- la prima, \hat{y}_i , che esprime il valore di Y che osserveremmo qualora i fattori di disturbo non esercitassero la loro influenza;
- la seconda, $y_i - a_0 - a_1 x_i$, dovuta all'influenza dei fattori di disturbo.

Ne consegue che risulta logico minimizzare la (1), che rappresenta una misura dell'impatto complessivo dei fattori di disturbo su Y , e quindi descrivere la relazione lineare tra X e Y in termini della (3).

Viceversa, se si suppone che l'influenza dei fattori di disturbo si eserciti *solamente* sulla X ma *non* sulla Y , con un ragionamento speculare a quello ora svolto, se ne deduce che risulta logico minimizzare la (2) e dunque esprimere la relazione lineare tra X e Y in termini della (4).

Ci sono, tuttavia, situazioni in cui l'ipotesi che l'influenza dei fattori di disturbo si riversi esclusivamente sulla Y oppure sulla X è scarsamente attendibile o addirittura insostenibile, e appare più giustificato ipotizzare che tale influenza si eserciti *sia* sulla Y *che* sulla X , facendo così cadere il presupposto logico che sta alla base della minimizzazione della (1) o della (2).

Le alternative al metodo classico che esamineremo nei capitoli successivi, sempre ponendoci da un punto di vista «geometrico-descrittivo» e nell'assunto che \mathbb{R}^2 sia dotato della metrica euclidea standard, si propongono appunto di descrivere la relazione lineare che intercorre tra X e Y in questa seconda ipotesi ⁽³⁾.

OSSERVAZIONE 1. Si noti che, benché \hat{a}_0, \hat{a}_1 e \hat{b}_0, \hat{b}_1 non siano invarianti

(3) Lo studio della relazione lineare tra due variabili, nel caso particolare in cui la metrica adottata sia la metrica L_1 , sarà oggetto di un prossimo lavoro.

rispetto a cambiamenti delle unità di misura con cui sono espressi i valori osservati di X e Y, gli effetti di tale cambiamento sono del tutto prevedibili.

Infatti, se ($s_1, s_2 > 0$)

$$x_i \rightarrow x_i s_1 \quad , \quad y_i \rightarrow y_i s_2$$

allora

$$\hat{a}_0 \rightarrow \hat{a}_0 s_2 \quad , \quad \hat{a}_1 \rightarrow \hat{a}_1 \frac{s_2}{s_1}$$

e

$$\hat{b}_0 \rightarrow \hat{b}_0 s_1 \quad , \quad \hat{b}_1 \rightarrow \hat{b}_1 \frac{s_1}{s_2} .$$

OSSERVAZIONE 2. Posto che nella (4) sia $\hat{b}_1 \neq 0$, esplicitando formalmente rispetto a y, possiamo riscrivere $\hat{r}_{X/Y}$ nella forma

$$\hat{r}_{Y/X} : y = -\frac{\hat{b}_0}{\hat{b}_1} + \frac{1}{\hat{b}_1} x = \hat{\hat{b}}_0 + \hat{\hat{b}}_1 x .$$

In questo modo, sia in $\hat{r}_{Y/X}$ che in $\hat{\hat{r}}_{Y/X}$ la y figura come variabile dipendente e, a sua volta, la x assume il ruolo di variabile indipendente.

OSSERVAZIONE 3. Si noti che

$$0 < \hat{a}_1 = \rho_{XY} \frac{\sigma_Y}{\sigma_X} < \frac{1}{\rho_{XY}} \frac{\sigma_Y}{\sigma_X} = \hat{\hat{b}}_1 \quad \text{per } 0 < \rho_{XY} < 1$$

e

$$0 > \hat{a}_1 = \rho_{XY} \frac{\sigma_Y}{\sigma_X} > \frac{1}{\rho_{XY}} \frac{\sigma_Y}{\sigma_X} = \hat{\hat{b}}_1 \quad \text{per } -1 < \rho_{XY} < 0 .$$

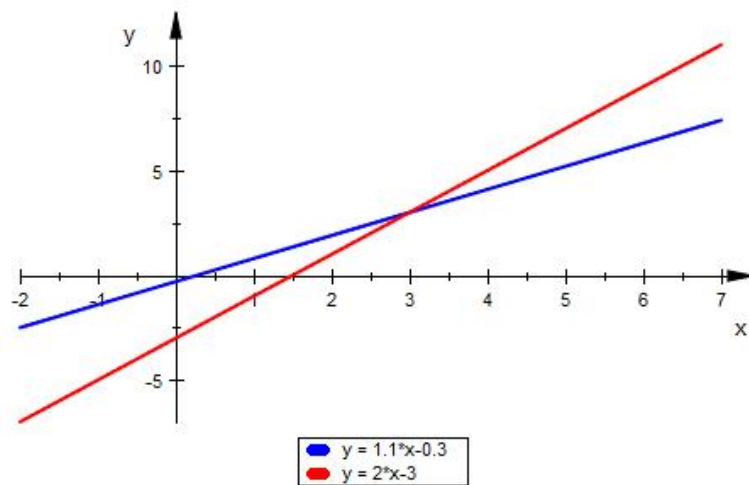
ESEMPIO 1. Qualora i valori osservati delle variabili X e Y siano

X	1	2	3	4	5
Y	2	1	3	2	7

si verifica facilmente che

$$\hat{a}_0 = -0.3, \hat{a}_1 = 1.1, \hat{b}_0 = 1.5, \hat{b}_1 = 0.5, \hat{b}_0 = -3, \hat{b}_1 = 2, \rho_{YX} = 0.74.$$

Le rette $\hat{r}_{Y|X}$ e $\hat{r}_{X|Y}$ che ne derivano sono riportate nel grafico seguente.



Cap. 2 REGRESSIONE ORTOGONALE

Consideriamo nuovamente in \mathbb{R}^2 la retta $r : ax + by = c$.

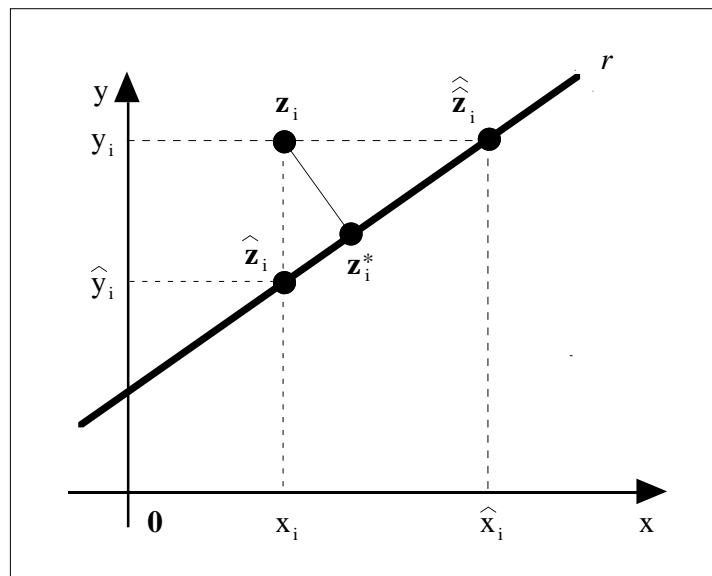
Come si è visto nel capitolo precedente – riscritta r nelle forme $(a, b \neq 0)$

$$r_{Y/X} : y = c/b - a/bx = a_0 + a_1x \quad , \quad r_{X/Y} : x = c/a - b/ay = b_0 + b_1y$$

– le proiezioni di \mathbf{z}_i su r lungo ciascuno degli assi danno luogo a due vettori che abbiamo indicato con $\widehat{\mathbf{z}}_i$ e $\widehat{\widehat{\mathbf{z}}}_i$.

Supponiamo adesso di proiettare *ortogonalmente* \mathbf{z}_i su r e sia \mathbf{z}_i^* il risultato di questa operazione (Cfr. la Fig. 2).

Fig. 2



Associamo poi a \mathbf{z}_i^* il vettore $\omega_i = \mathbf{z}_i - \mathbf{z}_i^*$, denominato *residuo ortogonale*, di lunghezza pari a $d_{\omega,i}$.

Dato che i residui $\delta_i = \mathbf{z}_i - \widehat{\mathbf{z}}_i$ e $\varepsilon_i = \mathbf{z}_i - \widehat{\widehat{\mathbf{z}}}_i$ – di lunghezze pari a $d_{\delta,i}$ e $d_{\varepsilon,i}$ – sono ortogonali, per un noto teorema di Geometria analitica risulta

$$(5) \quad d_{\omega,i}^2 = \frac{(y_i - a_0 - a_1x_i)^2}{1 + a_1^2}$$

oppure

$$(5') \quad d_{\omega,i}^2 = \frac{(x_i - b_0 - b_1 y_i)^2}{1 + b_1^2}.$$

Ciò premesso, facendo ricorso per concretezza alla (5) che sottintende il riferimento alla retta r scritta nella forma $r_{Y/X} : y = a_0 + a_1 x$, il valore dei parametri a_0, a_1 è determinato mediante un procedimento che possiamo chiamare *metodo ortogonale* e che consiste nel rendere minima l'espressione

$$(6) \quad \frac{1}{n} \sum_i d_{\omega,i}^2 = \frac{1}{n} \sum_i \frac{(y_i - a_0 - a_1 x_i)^2}{1 + a_1^2}$$

la quale rappresenta una misura della dispersione (*inerzia*) di $\mathbf{z}_1, \dots, \mathbf{z}_n$ attorno a r .

Siano adesso a_0^* e a_1^* il valore dei parametri che minimizzano la (6).

Poiché, nell'ipotesi che sia $\sigma_{XY} \neq 0$, si dimostra – seguendo per esempio Casella&Berger (2002) – che

$$a_0^* = \bar{y} - a_1^* \bar{x}, \quad a_1^* = \frac{\sigma_Y^2 - \sigma_X^2 + \sqrt{(\sigma_Y^2 - \sigma_X^2)^2 + 4\sigma_{XY}^2}}{2\sigma_{XY}}$$

risulta definita la cosiddetta *retta di regressione ortogonale*, ovviamente passante per il baricentro, che indichiamo con

$$(7) \quad r_{Y,X}^* : y = a_0^* + a_1^* x.$$

OSSERVAZIONE 4. Come si riconosce facilmente, a_0^* e a_1^* *non* sono invarianti rispetto a cambiamenti delle unità di misura con cui sono espressi i valori osservati di X e Y .

Ed è bensì vero che, in linea teorica, sarebbe possibile individuare il legame che sussiste tra «vecchi» e «nuovi» valori dei parametri in questione, ma tale legame risulta relativamente complesso e non molto utile.

OSSERVAZIONE 5. Consideriamo nuovamente le lunghezze al quadrato $d_{\delta,i}^2$ e

$d_{\varepsilon,i}^2$ dei residui δ_i e ε_i .

La media armonica di $d_{\delta,i}^2$ e $d_{\varepsilon,i}^2$ è

$$M_i^{(h)} = \begin{cases} 2 \frac{d_{\delta,i}^2 d_{\varepsilon,i}^2}{d_{\delta,i}^2 + d_{\varepsilon,i}^2} & \text{per } d_{\delta,i}^2, d_{\varepsilon,i}^2 \neq 0 \ (\mathbf{z}_i \notin r) \\ 0 & \text{per } d_{\delta,i}^2, d_{\varepsilon,i}^2 = 0 \ (\mathbf{z}_i \in r). \end{cases}$$

Ma, come è subito visto ($a, b \neq 0$)

$$b_0 = -\frac{a_0}{a_1}, \quad b_1 = \frac{1}{a_1}$$

da cui

$$x_i - b_0 - b_1 y_i = x_i + \frac{a_0}{a_1} - \frac{1}{a_1} y_i.$$

Pertanto,

$$\frac{1}{2} M_i^{(h)} = \frac{(y_i - a_0 - a_1 x_i)^2 (x_i + \frac{a_0}{a_1} - \frac{1}{a_1} y_i)^2}{(y_i - a_0 - a_1 x_i)^2 + (x_i + \frac{a_0}{a_1} - \frac{1}{a_1} y_i)^2} = \frac{(y_i - a_0 - a_1 x_i)^2}{1 + a_1^2} = d_{\omega,i}^2.$$

Ne consegue che minimizzare la (6) produce gli stessi risultati che otterremmo minimizzando la media (aritmetica semplice) delle medie armoniche delle lunghezze al quadrato $d_{\delta,i}^2$ e $d_{\varepsilon,i}^2$ dei residui δ_i e ε_i .

OSSERVAZIONE 6. Consideriamo la matrice delle covarianze

$$\mathbf{V} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

e l'equazione

$$\mathbf{Vc} = \lambda \mathbf{c}$$

nelle incognite λ e $\mathbf{c} \neq \mathbf{0}$ associata a \mathbf{V} .

Soluzioni di tale equazione sono date, come è ben noto, dagli autovalori

$\tilde{\lambda}_1$ e $\tilde{\lambda}_2$ i quali, escluso il caso banale in cui \mathbf{V} sia una matrice scalare, risultano distinti ($\tilde{\lambda}_1 \neq \tilde{\lambda}_2$) e dagli autovettori ortonormali (*vettori principali*) $\tilde{\mathbf{c}}_1$ e $\tilde{\mathbf{c}}_2$ corrispondenti a $\tilde{\lambda}_1$ e $\tilde{\lambda}_2$.

Posto $\tilde{\lambda}_1 > \tilde{\lambda}_2$ – essendo

$$\tilde{\lambda}_1 = \frac{\sigma_X^2 + \sigma_Y^2 + \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2}}{2}$$

e

$$\tilde{\mathbf{c}}_1 = \left[\frac{1}{\sqrt{1 + a_1^{*2}}} \quad \frac{a_1^*}{\sqrt{1 + a_1^{*2}}} \right]' = [\tilde{c}_{11} \quad \tilde{c}_{21}]'$$

– possiamo definire in \mathbb{R}^2 la varietà lineare (retta)

$$\tilde{r}_{Y,X} : \begin{cases} x = \bar{x} + \tilde{c}_{11}t \\ y = \bar{y} + \tilde{c}_{21}t \end{cases}$$

che, eliminando il parametro t , riscriviamo nella forma

$$\tilde{r}_{Y,X} : y = \bar{y} - \frac{\tilde{c}_{21}}{\tilde{c}_{11}}\bar{x} + \frac{\tilde{c}_{21}}{\tilde{c}_{11}}x.$$

Ma, poiché $\tilde{c}_{21}/\tilde{c}_{11} = a_1^*$, se ne deduce che $\tilde{r}_{Y,X} = r_{Y,X}^*$.

In sostanza, il procedimento che ha condotto a definire la retta di regressione ortogonale e quello ora delineato, tipico della analisi delle componenti principali basata sulla matrice delle covarianze, portano allo stesso risultato.

OSSERVAZIONE 7. Il metodo ortogonale, nonostante posseda sul piano concettuale una indubbia attrattiva, presenta alcuni inconvenienti, che sono gli stessi di una analisi delle componenti principali basata sulla matrice delle covarianze.

In particolare, supposto che i valori osservati di X e Y siano espressi in unità di misura diverse, il procedimento cade in difetto.

In tal caso, l'espedito, perché di questo si tratta, a cui usualmente si

ricorre consiste nello standardizzare i valori osservati di X e Y , e ciò equivale a lavorare sulla matrice di correlazione piuttosto che sulla matrice delle covarianze.

OSSERVAZIONE 8. Vogliamo dimostrare che

$$0 < \hat{a}_1 < a_1^* < \hat{b}_1 \quad \text{per } 0 < \rho_{XY} < 1$$

e

$$0 > \hat{a}_1 > a_1^* > \hat{b}_1 \quad \text{per } -1 < \rho_{XY} < 0.$$

Possiamo limitarci a provare la prima di queste relazioni, dato che la seconda si dimostra in modo del tutto analogo.

Supposto dunque $0 < \rho_{XY} < 1$, si consideri il polinomio di secondo grado

$$P(a_1) = \sigma_{XY} a_1^2 + (\sigma_X^2 - \sigma_Y^2) a_1 - \sigma_{XY}$$

di cui a_1^* costituisce, come è subito visto, una radice.

$P(a_1)$ possiede le proprietà seguenti: ⁽⁴⁾

(a) $P(\hat{a}_1) < 0$.

(b) $P(\hat{b}_1) > 0$.

(c) La derivata seconda di $P(a_1)$ è positiva per ogni a_1 appartenente all'intervallo (\hat{a}_1, \hat{b}_1) .

Esiste quindi un unico numero \bar{c} tale che $P(\bar{c}) = 0$ nell'intervallo (\hat{a}_1, \hat{b}_1) (secondo teorema sulla unicità della radice di una funzione) e questo numero identifica appunto a_1^* .

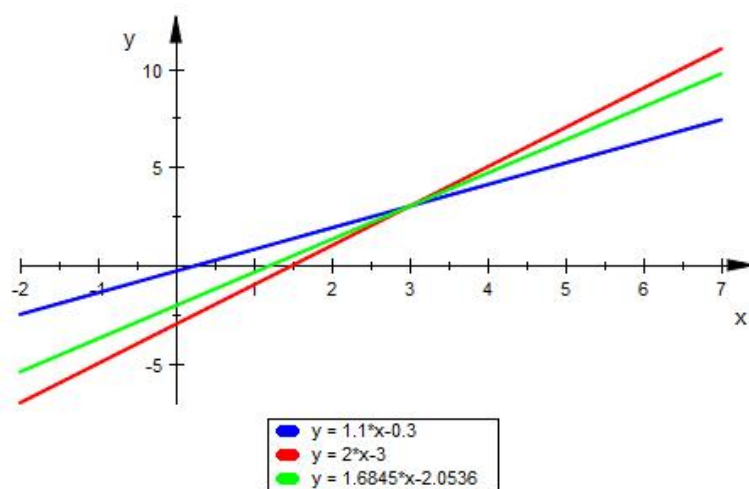
OSSERVAZIONE 9. La regressione ortogonale trae origine dai lavori di Adcock (1878) e Pearson (1901), ma ha assunto nel tempo denominazioni assai diverse tra cui: *major axis regression*, *total least squares regression*.

(4) Per i dettagli si rinvia all'Appendice.

ESEMPIO 2. Riprendendo i dati di cui all' Esempio 1, risulta

$$a_0^* = -2.0536, a_1^* = 1.6845.$$

La retta di regressione ortogonale che ne deriva è riportata nel grafico seguente.



Cap. 3 REGRESSIONE AREALE

Consideriamo ancora una volta in \mathbb{R}^2 la retta $r : ax + by = c$.

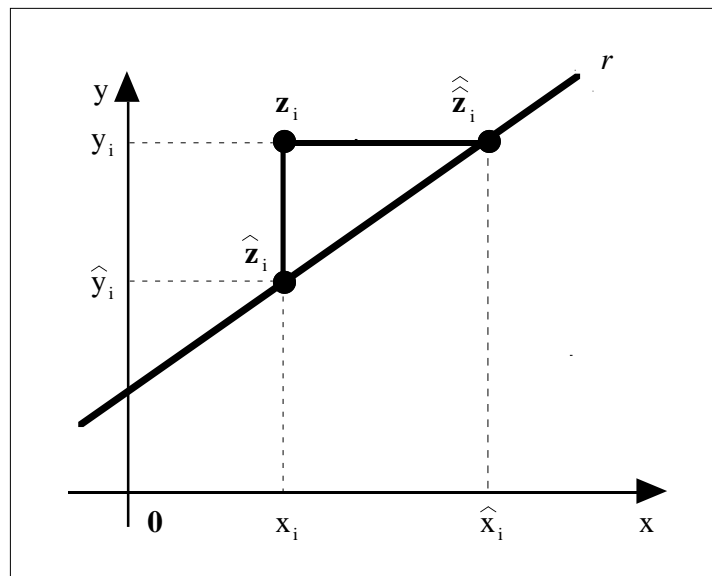
Come si è visto – riscritta r nelle forme ($a, b \neq 0$)

$$r_{Y/X} : y = c/b - a/b x = a_0 + a_1 x \quad , \quad r_{X/Y} : x = c/a - b/a y = b_0 + b_1 y$$

– le proiezioni di z_i su r lungo ciascuno degli assi danno luogo a due vettori che abbiamo indicato con \hat{z}_i e $\hat{\hat{z}}_i$.

Consideriamo adesso il triangolo (rettangolo) T_i di vertici z_i, \hat{z}_i e $\hat{\hat{z}}_i$ i cui lati delimitano la superficie S_i che possiamo chiamare *residuo superficiale* (Cfr. la Fig. 3).

Fig. 3



L'area A_i di S_i vale $\frac{1}{2} d_{\delta,i} d_{\varepsilon,i}$ e quindi ($a_1, b_1 \neq 0$)

$$\begin{aligned}
 A_i &= \frac{1}{2} |y_i - a_0 - a_1 x_i| |x_i - b_0 - b_1 y_i| &= \frac{1}{2} |y_i - a_0 - a_1 x_i| \left| x_i + \frac{a_0}{a_1} - \frac{1}{a_1} y_i \right| \\
 (8) \quad &= \frac{1}{2} |y_i - a_0 - a_1 x_i| \left| -x_i - \frac{a_0}{a_1} + \frac{1}{a_1} y_i \right| &= \frac{1}{2} |y_i - a_0 - a_1 x_i| |y_i - a_0 - a_1 x_i| \frac{1}{|a_1|} \\
 &= \frac{1}{2} \frac{1}{|a_1|} (y_i - a_0 - a_1 x_i)^2
 \end{aligned}$$

oppure

$$\begin{aligned}
 A_i &= \frac{1}{2} |y_i - a_0 - a_1 x_i| |x_i - b_0 - b_1 y_i| = \frac{1}{2} \left| y_i + \frac{b_0}{b_1} - \frac{1}{b_1} x_i \right| |x_i - b_0 - b_1 y_i| \\
 (8') \quad &= \frac{1}{2} \left| -y_i - \frac{b_0}{b_1} + \frac{1}{b_1} x_i \right| |x_i - b_0 - b_1 y_i| = \frac{1}{2} \frac{1}{|b_1|} |x_i - b_0 - b_1 y_i| |x_i - b_0 - b_1 y_i| \\
 &= \frac{1}{2} \frac{1}{|b_1|} (x_i - b_0 - b_1 y_i)^2.
 \end{aligned}$$

Ciò premesso, facendo ricorso per concretezza alla (8) che sottintende il riferimento alla retta r scritta nella forma $r_{Y/X} : y = a_0 + a_1 x$, il valore dei parametri a_0, a_1 è determinato mediante un procedimento che possiamo chiamare *metodo areale* e che consiste nel rendere minima l'espressione

$$(9) \quad \frac{1}{n} \sum_i A_i = \frac{1}{2} \frac{1}{n} \sum_i \frac{(y_i - a_0 - a_1 x_i)^2}{|a_1|}$$

la quale rappresenta una misura della dispersione di $\mathbf{z}_1, \dots, \mathbf{z}_n$ attorno a r .

Siano adesso $a_0^\#$ e $a_1^\#$ il valore dei parametri che minimizzano la (9).

Poiché, nell'ipotesi che sia $\sigma_{XY} \neq 0$, si dimostra – seguendo per esempio Greene (2013) – che

$$a_0^\# = \bar{y} - a_1^\# \bar{x}, \quad a_1^\# = \text{sgn}(\rho_{XY}) \sqrt{\widehat{a}_1 \widehat{b}_1} = \text{sgn}(\rho_{XY}) \frac{\sigma_Y}{\sigma_X}$$

risulta definita la cosiddetta *retta di regressione areale* ⁽⁵⁾, ovviamente passante per il baricentro, che indichiamo con

$$(10) \quad r_{Y,X}^\# : y = a_0^\# + a_1^\# x.$$

OSSERVAZIONE 10. La pendenza $a_1^\#$ della retta di regressione areale, come subito si verifica, è invariante rispetto a cambiamenti delle unità di misura con cui sono espressi i valori osservati di X e Y .

Questo fatto è spesso ritenuto in letteratura quale motivo di preferenza del metodo areale nei confronti del metodo ortogonale.

(5) Questa dizione è tutt'altro che comune.

OSSERVAZIONE 11. Consideriamo le lunghezze al quadrato $d_{\delta,i}^2$ e $d_{\varepsilon,i}^2$ dei residui δ_i e ε_i .

La media geometrica di $d_{\delta,i}^2$ e $d_{\varepsilon,i}^2$ è

$$M_i^{(g)} = \sqrt{d_{\delta,i}^2 d_{\varepsilon,i}^2} = |y_i - a_0 - a_1 x_i| |x_i - b_0 - b_1 y_i| = 2 A_i.$$

Quindi, minimizzare la (9) produce gli stessi risultati che otterremmo minimizzando la media (aritmetica semplice) delle medie geometriche delle lunghezze al quadrato $d_{\delta,i}^2$ e $d_{\varepsilon,i}^2$ dei residui δ_i e ε_i .

OSSERVAZIONE 12. Consideriamo la matrice di correlazione

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{bmatrix}$$

e l'equazione

$$\mathbf{R}\mathbf{c} = \lambda \mathbf{c}$$

nelle incognite λ e $\mathbf{c} \neq \mathbf{0}$ associata a \mathbf{R} .

Soluzioni di tale equazione sono date dagli autovalori $\tilde{\lambda}_1$ e $\tilde{\lambda}_2$ i quali, escluso il caso banale in cui sia $\rho_{XY} = 0$, risultano distinti ($\tilde{\lambda}_1 \neq \tilde{\lambda}_2$) e dagli autovettori ortonormali (*vettori principali*) $\tilde{\mathbf{c}}_1$ e $\tilde{\mathbf{c}}_2$ corrispondenti ai suddetti autovalori.

Limitiamoci per semplicità a esaminare il caso in cui sia $0 < \rho_{XY} < 1$.

Posto $\tilde{\lambda}_1 > \tilde{\lambda}_2$ – essendo

$$\tilde{\lambda}_1 = 1 + \rho_{XY}$$

e

$$\tilde{\mathbf{c}}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}' = [\tilde{\mathbf{c}}_{11} \ \tilde{\mathbf{c}}_{21}]'$$

– possiamo definire in \mathbb{R}^2 la varietà lineare (retta)

$$\tilde{r}_{Y,X} : \begin{cases} x = \bar{x} + \sigma_X \tilde{c}_{11} t \\ y = \bar{y} + \sigma_Y \tilde{c}_{21} t \end{cases}$$

che, eliminando il parametro t , riscriviamo nella forma

$$\tilde{r}_{Y,X} : y = \bar{y} - \frac{\sigma_{Y\bar{X}}}{\sigma_X} + \frac{\sigma_{YX}}{\sigma_X} .$$

Ma, poiché $\sigma_Y/\sigma_X = a_1^\#$, se ne deduce che $\tilde{r}_{Y,X} = r_{Y,X}^\#$.

In sostanza, il procedimento che ha condotto a definire la retta di regressione areale e quello ora delineato, tipico della analisi delle componenti principali basata sulla matrice di correlazione, portano allo stesso risultato.

OSSERVAZIONE 13. Dato che $a_1^\#$ è media geometrica con segno di \hat{a}_1 e \hat{b}_1 , si ha

$$\hat{a}_1 < a_1^\# < \hat{b}_1 \quad \text{per } 0 < \rho_{XY} < 1$$

e

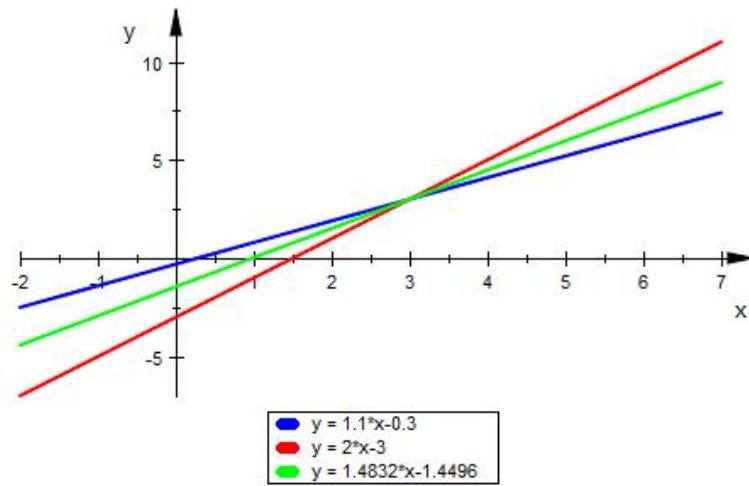
$$\hat{a}_1 > a_1^\# > \hat{b}_1 \quad \text{per } -1 < \rho_{XY} < 0 .$$

OSSERVAZIONE 14. La regressione areale, sotto la denominazione di *diagonal regression*, trae origine da un lavoro di Frisch (1928), ma attualmente le denominazioni più comuni sono: *reduced major axis regression*, *geometric mean regression*.

ESEMPIO 3. Riprendendo i dati di cui all' Esempio 1, risulta

$$a_0^\# = -1.4496, a_1^\# = 1.4832 .$$

La retta di regressione areale che ne deriva è riportata nel grafico alla pagina seguente.



Cap. 4 REGRESSIONE PITAGORICA

Considerato nuovamente il triangolo (rettangolo) T_i di vertici \mathbf{z}_i , $\widehat{\mathbf{z}}_i$ e $\widehat{\widehat{\mathbf{z}}}_i$ introdotto nel capitolo precedente, concentriamo l'attenzione sul vettore $\mathbf{h}_i = \widehat{\widehat{\mathbf{z}}}_i - \widehat{\mathbf{z}}_i$ (Cfr. la Fig. 3) ⁽⁶⁾.

Il quadrato della lunghezza di \mathbf{h}_i , che denotiamo con $d_{h,i}^2$, vale ovviamente (teorema di Pitagora) $d_{\delta,i}^2 + d_{\varepsilon,i}^2$.

Pertanto – supposto $a_1, b_1 \neq 0$ – risulta

$$(11) \quad d_{h,i}^2 = (y_i - a_0 - a_1 x_i)^2 + \left(x_i + \frac{a_0}{a_1} - \frac{1}{a_1} y_i\right)^2$$

oppure

$$(11') \quad d_{h,i}^2 = \left(y_i + \frac{b_0}{b_1} - \frac{1}{b_1} x_i\right)^2 + (x_i - b_0 - b_1 y_i)^2.$$

Ciò premesso, facendo ricorso per concretezza alla (11) che sottintende il riferimento alla retta r scritta nella forma $r_{Y/X} : y = a_0 + a_1 x$, il valore dei parametri a_0, a_1 è determinato mediante un procedimento che possiamo chiamare *metodo pitagorico* e che consiste nel rendere minima l'espressione

$$(12) \quad \frac{1}{n} \sum_i d_{h,i}^2 = \frac{1}{n} \sum_i \left[(y_i - a_0 - a_1 x_i)^2 + \left(x_i + \frac{a_0}{a_1} - \frac{1}{a_1} y_i\right)^2 \right].$$

Siano a_0^{**} e a_1^{**} il valore dei parametri che minimizzano la (12).

Qualora sia $\sigma_{XY} \neq 0$, si dimostra – seguendo per esempio Greene (2013) – che

$$a_0^{**} = \bar{y} - a_1^{**} \bar{x}$$

e a_1^{**} è ottenuto come radice del polinomio di quarto grado

$$P(a_1) = \sigma_X^2 a_1^4 - \sigma_{XY} a_1^3 + \sigma_{XY} a_1 - \sigma_Y^2.$$

(6) Si noti esplicitamente che, dal punto di vista geometrico, \mathbf{h}_i rappresenta l'ipotenusa di T_i e, a sua volta, ω_i (residuo ortogonale) costituisce l'altezza relativa all'ipotenusa.

Quest'ultimo polinomio ammette due radici complesse coniugate, che non interessano nel presente contesto, e due radici (reali) di segno opposto delle quali occorre scegliere quella di segno concorde con ρ_{XY} ⁽⁷⁾.

In definitiva, attraverso il procedimento ora descritto risulta definita la cosiddetta *retta di regressione pitagorica*, ovviamente passante per il baricentro, che indichiamo con

$$(13) \quad r_{Y,X}^{**} : y = a_0^{**} + a_1^{**} x .$$

OSSERVAZIONE 14. Chiaramente, a_0^{**} e a_1^{**} *non* sono invarianti rispetto a cambiamenti delle unità di misura con cui sono espressi i valori osservati di X e Y.

OSSERVAZIONE 15. Vogliamo dimostrare che

$$0 < \hat{a}_1 < a_1^{**} < \hat{b}_1 \quad \text{per } 0 < \rho_{XY} < 1$$

e

$$0 > \hat{a}_1 > a_1^{**} > \hat{b}_1 \quad \text{per } -1 < \rho_{XY} < 0 .$$

Possiamo limitarci a provare la prima delle relazioni di cui sopra, dato che la seconda si dimostra con un ragionamento del tutto simile.

Supponiamo dunque che sia $0 < \rho_{XY} < 1$.

In tal caso, risulta ⁽⁸⁾

$$(a) \quad P(\hat{a}_1) < 0 .$$

$$(b) \quad P(\hat{b}_1) > 0 .$$

(c) La derivata seconda di $P(a_1)$ è positiva per ogni a_1 appartenente all'intervallo (\hat{a}_1, \hat{b}_1) .

(7) Di fatto, le radici del polinomio in questione vengono calcolate per via numerica, in modo approssimato. In Matlab, la funzione ROOTS serve egregiamente allo scopo.

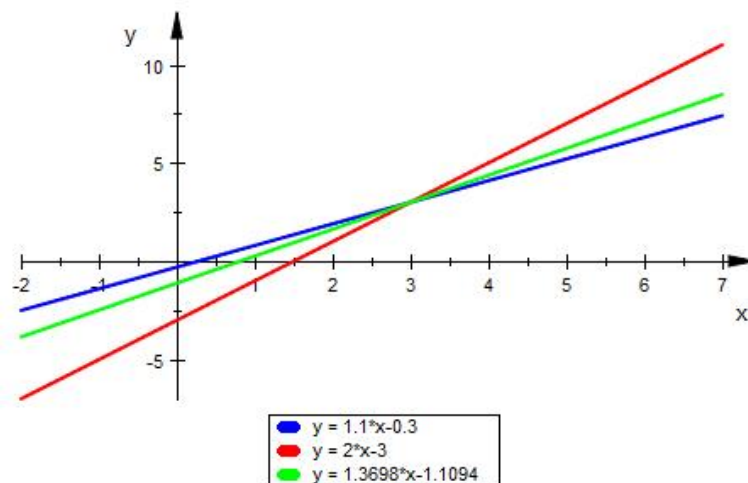
(8) Per i dettagli si rinvia all'Appendice.

Esiste quindi un unico numero \bar{c} tale che $P(\bar{c}) = 0$ nell'intervallo sopra menzionato (secondo teorema sulla unicità della radice di una funzione) e questo numero identifica appunto a_1^{**} .

ESEMPIO 4. Riprendendo i dati di cui all' Esempio 1, risulta

$$a_0^{**} = -1.1094, a_1^{**} = 1.3698.$$

La retta di regressione areale che ne deriva è riportata nel grafico alla seguente.



Cap. 5 ULTERIORI ALTERNATIVE

In questo capitolo ci proponiamo di suggerire un procedimento che consenta di allargare lo spettro delle possibili alternative al metodo classico nello studio della regressione lineare bivariata.

A questo fine, osserviamo anzitutto che le rette di regressione classica $\hat{r}_{Y/X}$ e $\hat{r}_{X/Y}$ – entrambe passanti per il baricentro e con pendenze, \hat{a}_1 e \hat{b}_1 , di segno concorde – possono essere assunte quali rette generatrici di un fascio proprio di rette ⁽⁹⁾.

In questa ottica, è abbastanza naturale assumere come membri del fascio quelli risultanti da una media delle suddette pendenze, media che indichiamo genericamente con

$$M(\hat{a}_1, \hat{b}_1).$$

La questione diviene dunque quale tipo di media utilizzare.

Qui di seguito sono indicate alcune medie, scelte tra quelle di uso più comune ($\hat{a}_1, \hat{b}_1 \neq 0$):

$$\textit{Media aritmetica} \quad M(\hat{a}_1, \hat{b}_1) = \text{sgn}(\rho_{XY}) \frac{|\hat{a}_1| + |\hat{b}_1|}{2}$$

$$\textit{Media armonica} \quad M(\hat{a}_1, \hat{b}_1) = \text{sgn}(\rho_{XY}) \frac{2|\hat{a}_1||\hat{b}_1|}{(|\hat{a}_1| + |\hat{b}_1|)}$$

$$\textit{Media antiarmonica} \quad M(\hat{a}_1, \hat{b}_1) = \text{sgn}(\rho_{XY}) \frac{\hat{a}_1^2 + \hat{b}_1^2}{(|\hat{a}_1| + |\hat{b}_1|)}$$

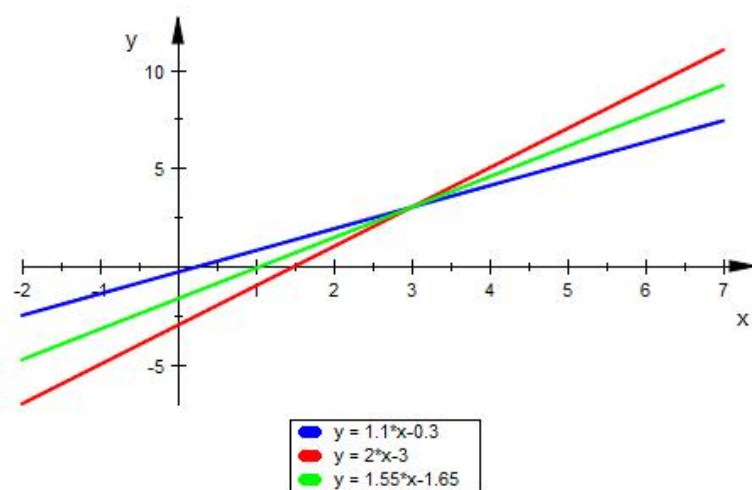
$$\textit{Media quadratica} \quad M(\hat{a}_1, \hat{b}_1) = \text{sgn}(\rho_{XY}) \left(\frac{\hat{a}_1^2 + \hat{b}_1^2}{2} \right)^{1/2}$$

(9) In realtà, avremmo potuto prendere come rette generatrici del fascio qualsiasi coppia tra quelle di cui abbiamo discusso in questo lavoro: per esempio, la coppia retta di regressione ortogonale/retta di regressione areale.

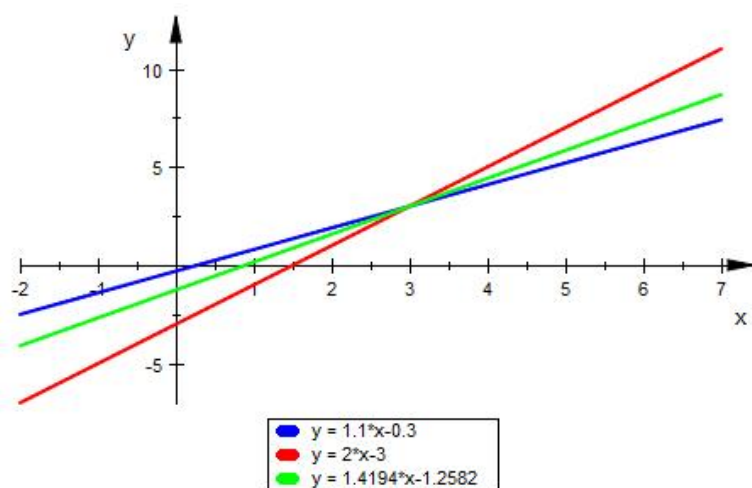
Si noti che le medie di cui sopra godono delle fondamentali proprietà, tra le altre, di *internalità* e *simmetria*.

ESEMPIO 5. Riprendendo i dati dell' Esempio 1, le rette di regressione che risultano in applicazione di quanto detto sono riportate nei grafici che seguono.

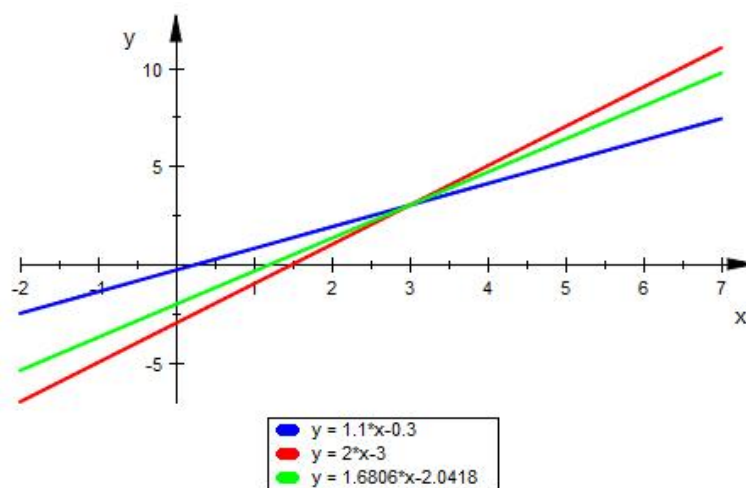
Media aritmetica



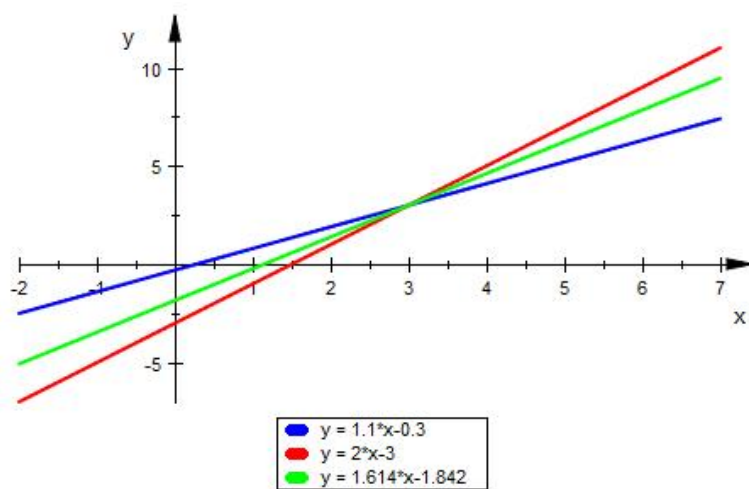
Media armonica



Media antiarmonica



Media quadratica



RIFERIMENTI BIBLIOGRAFICI

Adcock R.J.: *A problem in least squares*. The Analyst, London, 1878.

Casella G. & Berger R.L.: *Statistical inference*. Brooks/Cole, 2002.

Frisch R.: *Correlation and scatter in statistical variables*. Nordic Statistical Journal, 1929.

Greene N.: *Generalized least-squares regression I: efficient derivations*, Proceedings of the 1th International Conference on Computational Science and Engineering, Valencia, Spain, 2013.

Pearson K.: *On lines and planes of closest fit to systems of points in space*. The Philosophical Magazine, 1901.

APPENDICE

OSSERVAZIONE 8.

(a) $P(\hat{a}_1) < 0$.Infatti, poiché $\sigma_{XY}^2 - \sigma_X^2 \sigma_Y^2 < 0$,

$$\begin{aligned}
P(\hat{a}_1) &= \sigma_{XY} \frac{\sigma_{XY}^2}{\sigma_X^4} + (\sigma_X^2 - \sigma_Y^2) \frac{\sigma_{XY}}{\sigma_X^2} - \sigma_{XY} \\
&= \frac{\sigma_{XY}^3}{\sigma_X^4} + \sigma_{XY} - \frac{\sigma_Y^2}{\sigma_X^2} \sigma_{XY} - \sigma_{XY} \\
&= \frac{\sigma_{XY}}{\sigma_X^2} \left(\frac{\sigma_{XY}^2}{\sigma_X^2} - \sigma_Y^2 \right) \\
&= \frac{\sigma_{XY}}{\sigma_X^2} \left(\frac{\sigma_{XY}^2 - \sigma_X^2 \sigma_Y^2}{\sigma_X^2} \right) < 0.
\end{aligned}$$

(b) $P(\hat{b}_1) > 0$.Infatti, poiché $\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2 > 0$,

$$\begin{aligned}
P(\hat{b}_1) &= \sigma_{XY} \frac{\sigma_Y^4}{\sigma_{XY}^2} + (\sigma_X^2 - \sigma_Y^2) \frac{\sigma_Y^2}{\sigma_{XY}} - \sigma_{XY} \\
&= \frac{\sigma_Y^4}{\sigma_{XY}} + \frac{\sigma_X^2 \sigma_Y^2}{\sigma_{XY}} - \frac{\sigma_Y^4}{\sigma_{XY}} - \sigma_{XY} \\
&= \frac{\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2}{\sigma_{XY}} > 0.
\end{aligned}$$

(c) La derivata seconda di $P(a_1)$ è positiva per ogni a_1 appartenente all'intervallo (\hat{a}_1, \hat{b}_1) .

Infatti,

$$D^2(P(a_1)) = \sigma_{XY}$$

ed essendo tale derivata seconda positiva per ogni a_1 appartenente a \mathbb{R} , lo è *a fortiori* per ogni a_1 appartenente all'intervallo (\hat{a}_1, \hat{b}_1) .

OSSERVAZIONE 15.(a) $P(\hat{a}_1) < 0$.Infatti, poiché $\sigma_{XY}^2 - \sigma_X^2 \sigma_Y^2 < 0$,

$$\begin{aligned}
P(\hat{a}_1) &= \sigma_X^2 \frac{\sigma_{XY}^4}{\sigma_X^8} - \sigma_{XY} \frac{\sigma_{XY}^3}{\sigma_X^6} + \frac{\sigma_{XY}^2}{\sigma_X^2} - \sigma_Y^2 \\
&= \frac{\sigma_{XY}^4}{\sigma_X^6} - \frac{\sigma_{XY}^4}{\sigma_X^6} + \frac{\sigma_{XY}^2}{\sigma_X^2} - \sigma_Y^2 \\
&= \frac{\sigma_{XY}^2}{\sigma_X^2} - \sigma_Y^2 \\
&= \frac{\sigma_{XY}^2 - \sigma_X^2 \sigma_Y^2}{\sigma_X^2} < 0.
\end{aligned}$$

(b) $P(\hat{b}_1) > 0$.Infatti, poiché $\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2 > 0$,

$$\begin{aligned}
P(\hat{b}_1) &= \sigma_X^2 \frac{\sigma_Y^8}{\sigma_{XY}^4} - \sigma_{XY} \frac{\sigma_Y^6}{\sigma_{XY}^3} + \sigma_{XY} \frac{\sigma_Y^2}{\sigma_{XY}} - \sigma_Y^2 \\
&= \sigma_X^2 \frac{\sigma_Y^8}{\sigma_{XY}^4} - \frac{\sigma_Y^6}{\sigma_{XY}^2} \\
&= \frac{\sigma_X^2 \sigma_Y^8 - \sigma_{XY}^2 \sigma_Y^6}{\sigma_{XY}^4} \\
&= \frac{\sigma_Y^6}{\sigma_{XY}^4} (\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2) > 0.
\end{aligned}$$

(c) La derivata seconda di $P(a_1)$ è positiva per ogni a_1 appartenente all'intervallo (\hat{a}_1, \hat{b}_1) .

Infatti,

$$D^2(P(a_1)) = 12 \sigma_X^2 a_1^2 - 6 \sigma_{XY} a_1$$

e questo polinomio di secondo grado ammette, come subito si verifica, le radici

$$a_{1(-)} = 0 \quad , \quad a_{1(+)} = \frac{1}{2} \hat{a}_1 .$$

Ne consegue che, essendo la derivata seconda di $P(a_1)$ positiva per ogni a_1 esterno all'intervallo $[0, 1/2 \hat{a}_1]$, lo è *a fortiori* per ogni a_1 appartenente all'intervallo (\hat{a}_1, \hat{b}_1) .