RENATO LEONI

Regressione lineare

UNIVERSITÀ DI FIRENZE
DIPARTIMENTO DI STATISTICA "G. PARENTI"
FIRENZE, 2007

1 IL MODELLO DI REGRESSIONE LINEARE

Dati 1+p caratteri quantitativi, supponiamo che questi siano distinti in due gruppi il primo dei quali è formato da un solo carattere, che indichiamo con Y, il secondo dai rimanenti p caratteri, che indichiamo con $X_1, \ldots, X_p$.

Chiamiamo il carattere che entra nel primo gruppo *variabile dipendente* e i caratteri del secondo gruppo *variabili indipendenti* o *variabili esplicative* o *regressori*.

Indichiamo inoltre con y_i la i-esima determinazione della variabile dipendente Y e con x_{ij} la i-esima determinazione della variabile indipendente X_j $(i = 1,\ldots,n; j = 1,\ldots,p)$.

Considerata una famiglia di funzioni del tipo

$$Y = b_0 + b_1 X_1 + \ldots + b_p X_p \;,$$

si fissi per il momento l'attenzione su un certo insieme di valori (qualsiasi) dei coefficienti $b_0, b_1, \ldots, b_p$ (*coefficienti di regressione lineare*).

Ovviamente, in corrispondenza delle determinazioni $x_{i1}, \ldots, x_{ip}$ assunte dalle variabili indipendenti $X_1, \ldots, X_p$, si ottiene, in generale, un *valore teorico* $b_0 + b_1 x_{i1} + \ldots + b_p x_{ip}$ che si discosta, per difetto o per eccesso, dal valore osservato y_i.

La differenza $y_i - b_0 - b_1 x_{i1} - \ldots - b_p x_{ip}$ riceve la denominazione di *residuo* (*scostamento*, *errore*, *scarto*) e sarà indicata con e_i.

Si ha quindi che (*modello di regressione lineare*)

$$\boxed{y_i = b_0 + b_1 x_{i1} + \ldots + b_p x_{ip} + e_i} \qquad (1)$$

OSSERVAZIONE 1. Come si è ora detto, la (1) assume genericamente la denominazione di modello di regressione lineare.

Più specificatamente, nel caso in cui si abbia una sola variabile indipendente (p = 1) si parla di modello di regressione lineare *semplice*, mentre qualora si abbiano più variabili indipendenti $(p > 1)$ si parla di modello di regressione lineare *multipla*.

OSSERVAZIONE 2. Si consideri un modello del tipo (*modello di regressione funzionale*)

$$g(y_i) = b_0 + b_1\, g_1(x_{i1}) + \ldots + b_p\, g_p(x_{ip}) + e_i$$

dove $g, g_1, \ldots, g_p$ sono funzioni *note*, rispettivamente, di $Y, X_1, \ldots, X_p$.

Chiaramente, posto

$$y_i^* = g(y_i) \ ,\ x_{i1}^* = g_1(x_{i1}) \ ,\ \ldots \ ,\ x_{ip}^* = g_p(x_{ip}) \ ,$$

ci si riconduce a un modello di regressione del tipo considerato in precedenza in cui tuttavia – in luogo delle determinazioni delle variabili originarie $Y, X_1, \ldots, X_p$. – compaiono le determinazioni delle variabili trasformate $Y^*, X_1^*, \ldots, X_p^*$.

OSSERVAZIONE 3. Supposto che si abbia una sola variabile indipendente X, indichiamo le determinazioni da essa assunte con $x_1, \ldots, x_n$.

Considerato un modello del tipo (*modello di regressione funzionale semplice*)

$$y_i = b_0 + b_1\, g_1(x_i) + \ldots + b_p\, g_p(x_i) + e_i$$

dove $g_1, \ldots, g_p$ sono funzioni *note* di X, ci si riconduce al modello di regressione descritto in precedenza ponendo

$$x_{i1}^* = g_1(x_i) \ ,\ \ldots \ ,\ x_{ip}^* = g_p(x_i) \ .$$

2 LA DETERMINAZIONE DEI COEFFICIENTI DI REGRESSIONE LINEARE

Il problema che ci proponiamo di risolvere per primo riguarda la determinazione dei coefficienti di regressione lineare $b_0, b_1, \dots, b_p$ che compaiono nella (1) attraverso il metodo dei *minimi quadrati* (*m.q.*).

In termini formali, si tratta di determinare i coefficienti $b_0, b_1, \dots, b_p$ che rendono minima la quantità

$$\boxed{S = \Sigma_i\, e_i^2 = \Sigma_i\, (y_i - b_0 - b_1 x_{i1} - \dots - b_p x_{ip})^2}$$

Allo scopo di procedere il più rapidamente possibile alla soluzione di tale problema è conveniente introdurre una notazione di tipo matriciale.

Siano

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad , \quad \mathbf{Z} = \begin{bmatrix} 1 & x_{11} \cdots x_{1p} \\ \cdots & \cdots\cdots\cdots \\ 1 & x_{n1} \cdots x_{np} \end{bmatrix} \quad , \quad \mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_p \end{bmatrix} \quad , \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

cosicché la (1) assume la forma

$$\text{(1')} \qquad \boxed{\mathbf{y} = \mathbf{Z}\mathbf{b} + \mathbf{e}}$$

Posto (1)

$$\begin{aligned} S = \Sigma_i\, e_i^2 \qquad & = \mathbf{e}'\mathbf{e} \\ = (\mathbf{y} - \mathbf{Z}\mathbf{b})'(\mathbf{y} - \mathbf{Z}\mathbf{b}) \qquad & = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}\mathbf{b} - \mathbf{b}'\mathbf{Z}'\mathbf{y} + \mathbf{b}'\mathbf{Z}'\mathbf{Z}\mathbf{b} \\ = \mathbf{y}'\mathbf{y} - 2\,\mathbf{y}'\mathbf{Z}\mathbf{b} + \mathbf{b}'\mathbf{Z}'\mathbf{Z}\mathbf{b}\,, & \end{aligned}$$

il vettore $\widehat{\mathbf{b}}$ dei coefficienti che rende minima la quantità S deve essere tale che la derivata di S rispetto a **b**, calcolata in $\widehat{\mathbf{b}}$, risulti nulla.

Ma,

(1) Si osservi che nella espressione che segue è $\mathbf{y}'\mathbf{Z}\mathbf{b} = \mathbf{b}'\mathbf{Z}'\mathbf{y}$ essendo entrambi i membri di questa eguaglianza degli scalari.

$$\frac{\partial S}{\partial \mathbf{b}} = -\,2\mathbf{Z}'\mathbf{y} + 2\mathbf{Z}'\mathbf{Z}\mathbf{b}$$

e quindi, eguagliando a zero tale espressione, si ha (*sistema di equazioni normali*)

$$\boxed{\mathbf{Z}'\mathbf{Z}\mathbf{b} = \mathbf{Z}'\mathbf{y}}$$

D'altro canto, nell'ipotesi a cui ci atterremo anche nel seguito che $\mathbf{Z}$ sia di pieno rango per colonne [2], $\mathbf{Z}'\mathbf{Z}$ risulta invertibile; pertanto, il sistema di cui sopra ammette un'*unica* soluzione rispetto al vettore dei coefficienti incogniti $\mathbf{b}$ data da

$$\boxed{\widehat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}} \tag{2}$$

Inoltre, poiché la matrice $\partial^2 S / \partial \mathbf{b} \partial \mathbf{b}' = 2\mathbf{Z}'\mathbf{Z}$ è definita positiva, la soluzione ottenuta rappresenta effettivamente un minimo (globale) di S.

ESEMPIO 1. Dati

$$\mathbf{y} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 2 \end{bmatrix} \quad , \quad \mathbf{Z} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -2 \\ 1 & 1 & 0 \end{bmatrix}$$

risulta

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 4 & 0 & -4 \\ 0 & 4 & 0 \\ -4 & 0 & 6 \end{bmatrix} \quad , \quad (\mathbf{Z}'\mathbf{Z})^{-1} = \frac{1}{4}\begin{bmatrix} 3 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

e, quindi,

$$\widehat{\mathbf{b}} = \frac{1}{4}\begin{bmatrix} 3 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -2 & 0 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}.$$

(2) Si noti che una condizione necessaria affinché ciò si verifichi è che sia $n \geq 1 + p$.

3 ALCUNE PROPRIETÀ ELEMENTARI DELLA SOLUZIONE DEI MINIMI QUADRATI

Allo scopo di evidenziare alcune semplici proprietà della soluzione ottenuta mediante il metodo dei m.q., si osservi anzitutto che, indicando con

$$\boxed{\widehat{\mathbf{y}} = \mathbf{Z}\widehat{\mathbf{b}}}$$

il vettore dei valori teorici e con

$$\boxed{\widehat{\mathbf{e}} = \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{Z}\widehat{\mathbf{b}}}$$

il vettore dei residui, risulta

$$\boxed{\begin{aligned} \mathbf{Z}'\widehat{\mathbf{e}} &= \mathbf{Z}'(\mathbf{y} - \widehat{\mathbf{y}}) = \mathbf{Z}'\mathbf{y} - \mathbf{Z}'\widehat{\mathbf{y}} = \mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{Z}\widehat{\mathbf{b}} \\ &= \mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{y} = \mathbf{0} \end{aligned}}$$

In particolare, tenendo presente che la prima riga della matrice $\mathbf{Z}'$ è costituita dal vettore riga (di ordine n) $\mathbf{u}' = \left[1 \cdots 1\right]$, si ha che

$$\mathbf{u}'\widehat{\mathbf{e}} = \mathbf{u}'(\mathbf{y} - \widehat{\mathbf{y}}) = 0$$

da cui

$$\boxed{\mathbf{u}'\mathbf{y} = \mathbf{u}'\widehat{\mathbf{y}}}$$

ovvero che *la somma dei valori osservati eguaglia la somma dei valori teorici.*

Inoltre, essendo

$$\widehat{\mathbf{e}}'\widehat{\mathbf{y}} = \widehat{\mathbf{y}}'\widehat{\mathbf{e}} = \widehat{\mathbf{b}}'\mathbf{Z}'\widehat{\mathbf{e}} = \widehat{\mathbf{b}}'\mathbf{0} = 0\ ,$$

risulta che

$$\boxed{\begin{aligned}\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}}+\hat{\mathbf{e}})'(\hat{\mathbf{y}}+\hat{\mathbf{e}}) = \hat{\mathbf{y}}'\hat{\mathbf{y}}+\hat{\mathbf{y}}'\hat{\mathbf{e}}+\hat{\mathbf{e}}'\hat{\mathbf{y}}+\hat{\mathbf{e}}'\hat{\mathbf{e}}\\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}}+\hat{\mathbf{e}}'\hat{\mathbf{e}}\end{aligned}}$$

ovvero che *la somma dei quadrati dei valori osservati può essere scomposta in due parti di cui la prima rappresenta la somma dei quadrati dei valori teorici*, *la seconda rappresenta la somma dei quadrati dei residui*.

ESEMPIO 2. Proseguendo nell'Esempio 1, si ha che

$$\hat{\mathbf{y}} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -2 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 0 \\ 2 \end{bmatrix} \quad , \quad \hat{\mathbf{e}} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} .$$

ed, essendo $\mathbf{y}'\mathbf{y} = 24$, $\hat{\mathbf{y}}'\hat{\mathbf{y}} = 22$, $\hat{\mathbf{e}}'\hat{\mathbf{e}} = 2$, la proprietà di scomponibilità della somma dei quadrati dei valori osservati risulta verificata.

OSSERVAZIONE 4. La relazione esistente tra le determinazioni assunte dalla variabile dipendente e le determinazioni assunte dalle variabili indipendenti è, talvolta, descritta da un modello del tipo

$$y_i = b_1 x_{i1} + \ldots + b_p x_{ip} + e_i$$

in cui cioè manca il coefficiente b_0.

Qualora si proceda al calcolo dei coefficienti incogniti $b_1, \ldots, b_p$ mediante il metodo dei m.q. (3), si ottiene, com'è facile verificare, una espressione del tutto analoga alla (2).

In tal caso, tuttavia, pur continuando a valere la proprietà di scomponibilità della somma dei quadrati dei valori osservati, *non* sussiste più, in generale, l'eguaglianza tra la somma dei valori osservati e la somma dei valori teorici.

(3) Si noti che, in questo caso, una condizione necessaria affinché la matrice delle determinazioni assunte dalle variabili indipendenti sia di pieno rango per colonne è che si abbia $n \geq p$.

4 UNA INTERPRETAZIONE GEOMETRICA DELLA SOLUZIONE DEI MINIMI QUADRATI

Vogliamo, adesso, mostrare come i risultati conseguiti in precedenza attraverso il metodo dei m.q. siano suscettibili di ricevere una semplice, ma significativa, interpretazione geometrica.

A questo fine, definiamo anzitutto il *prodotto scalare* $G(\mathbf{v}, \mathbf{w})$ di due generici vettori $\mathbf{v}, \mathbf{w} \in R^n$ ponendo (*metrica standard*)

$$G(\mathbf{v}, \mathbf{w}) = \mathbf{v}' \mathbf{w} .$$

Come è noto, risulta allora possibile esprimere la *lunghezza* $\|\mathbf{v}\|$ di ogni vettore $\mathbf{v} \in R^n$ mediante l'espressione

$$\|\mathbf{v}\| = \sqrt{G(\mathbf{v}, \mathbf{w})} = \sqrt{\mathbf{v}' \mathbf{v}}$$

e la *distanza* tra due vettori $\mathbf{v}, \mathbf{w} \in R^n$ attraverso la relazione

$$d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\| = \sqrt{G(\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w})} = \sqrt{(\mathbf{v} - \mathbf{w})'(\mathbf{v} - \mathbf{w})} .$$

Assunta quale base di R^n quella costituita dagli n vettori canonici (di ordine n) $\mathbf{u}_1, \dots, \mathbf{u}_n$ e posto

$$\mathbf{x}_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix},$$

indichiamo con S_1 il sottospazio di R^n generato da $\mathbf{u}, \mathbf{x}_1, \dots, \mathbf{x}_p$.

Dato il vettore $\mathbf{y}$ definito in precedenza, è immediato riconoscere che il vettore

$$\boxed{\hat{\mathbf{y}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{Z}\hat{\mathbf{b}}}$$

– tale che la distanza al quadrato

$$d^2(\mathbf{y}, \hat{\mathbf{y}}) = (\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}}) = S$$

di **y** da $\widehat{\mathbf{y}}$ sia minima rispetto a ogni altro vettore di S_1 – non è altro che la *proiezione ortogonale* di **y** su S_1.

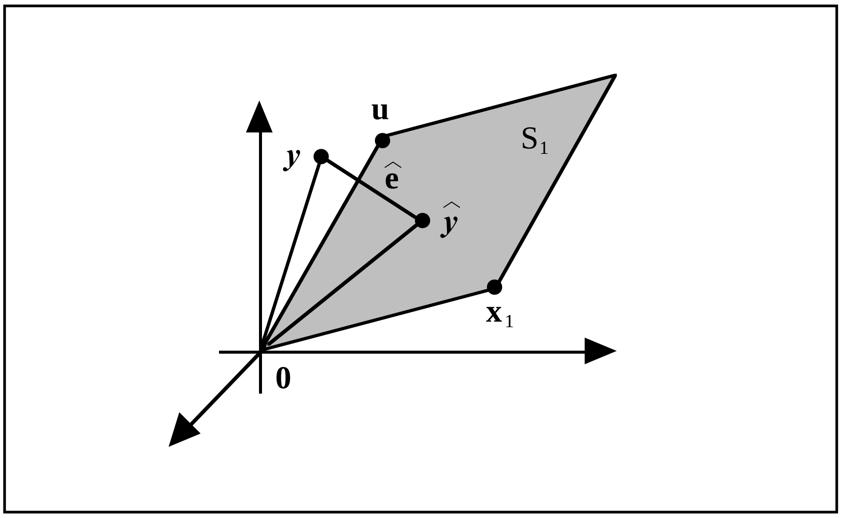
Inoltre, il teorema di Pitagora, applicato ai vettori (ortogonali) $\mathbf{y} - \widehat{\mathbf{y}} = \widehat{\mathbf{e}}$ e $\widehat{\mathbf{y}}$, fornisce immediatamente la proprietà di scomponibilità della somma dei quadrati dei valori osservati.

La Fig. 1 che segue, in cui p = 1, illustra quanto ora detto.

In essa il vettore **y** è rappresentato come un vettore non appartenente al sottospazio S_1 di R^3 generato dai vettori **u** e $\mathbf{x}_1$ e il vettore $\widehat{\mathbf{y}}$ è ottenuto come proiezione ortogonale di **y** su S_1.

Le coordinate di $\widehat{\mathbf{y}}$ rispetto a **u** e a $\mathbf{x}_1$ (non disegnate in figura) rappresentano i coefficienti di regressione $\widehat{b}_0$ e $\widehat{b}_1$.

Fig. 1



5 LA DETERMINAZIONE DEI COEFFICIENTI DI REGRESSIONE LINEARE: CONSIDERAZIONI ULTERIORI

Come si è visto, il calcolo dei coefficienti di regressione può essere eseguito direttamente sulla base della (2).

Tuttavia, sia allo scopo di rendere più agevole l'esecuzione dei calcoli, sia perché i coefficienti che realmente interessano sono quelli che si accompagnano a $\mathbf{x}_1, \dots, \mathbf{x}_p$, può essere conveniente impostare il problema in maniera leggermente differente da quella che abbiamo esposto in precedenza.

A questo fine, consideriamo i vettori

$$\overline{\mathbf{y}} = \mathbf{u}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{u}'\mathbf{y} = \mathbf{u}\frac{\mathbf{u}'\mathbf{y}}{n} = \mathbf{u}\,\overline{y} \quad , \quad \overline{\mathbf{x}}_j = \mathbf{u}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{u}'\mathbf{x}_j = \mathbf{u}\frac{\mathbf{u}'\mathbf{x}_j}{n} = \mathbf{u}\,\overline{x}_j$$

dove $\overline{y}$ e $\overline{x}_j$ rappresentano le medie degli elementi contenuti, rispettivamente, in $\mathbf{y}$ e $\mathbf{x}_j$ (4).

Definiti i vettori (scarti dalle rispettive medie)

$$\widetilde{\mathbf{y}} = \mathbf{y} - \overline{\mathbf{y}} = \begin{bmatrix} y_1 - \overline{y} \\ \vdots \\ y_n - \overline{y} \end{bmatrix} \quad , \quad \widetilde{\mathbf{x}}_j = \mathbf{x}_j - \overline{\mathbf{x}}_j = \begin{bmatrix} x_{1j} - \overline{x}_j \\ \vdots \\ x_{nj} - \overline{x}_j \end{bmatrix}$$

e posto

$$\widetilde{\mathbf{X}} = [\,\widetilde{\mathbf{x}}_1 \cdots \widetilde{\mathbf{x}}_p\,] \,,$$

vogliamo anzitutto mostrare che gli elementi del vettore

(3) $$\boxed{\hat{\widetilde{\mathbf{b}}} = (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{y}}}$$

– ottenuto minimizzando, rispetto a $\widetilde{\mathbf{b}}$, l'espressione

$$(\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\widetilde{\mathbf{b}})'(\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\widetilde{\mathbf{b}})$$

---

(4) Si noti che $\overline{\mathbf{y}}$ e $\overline{\mathbf{x}}_j$ si possono interpretare come le proiezioni di $\mathbf{y}$ e $\mathbf{x}_j$ nel sottospazio generato dal vettore $\mathbf{u}$.

– sono gli stessi, nell'ordine, di quelli che compaiono nel vettore $\widehat{\mathbf{b}}$ (Cfr. la (2)) in corrispondenza di $\mathbf{x}_1, \dots, \mathbf{x}_p$ e che $\widehat{b}_0$ può essere facilmente calcolato, qualora interessi, sulla base di $\widehat{\widetilde{\mathbf{b}}}$.

A questo scopo, supponiamo di ripartire la matrice $\mathbf{Z}$ e il vettore $\widehat{\mathbf{b}}$ nel seguente modo

$$\mathbf{Z} = \left[\mathbf{u} \; \vdots \; \mathbf{x}_1 \; \cdots \; \mathbf{x}_p\right] = \left[\mathbf{u} \;\; \mathbf{X}\right] \quad , \quad \widehat{\mathbf{b}} = \begin{bmatrix} \widehat{b}_0 \\ \cdots \\ \widehat{b}_1 \\ \vdots \\ \widehat{b}_p \end{bmatrix} = \begin{bmatrix} \widehat{b}_0 \\ \widehat{\mathbf{b}}_1 \end{bmatrix} .$$

Le posizioni fatte consentono di scrivere la (2) nella forma

$$\widehat{\mathbf{b}} = \begin{bmatrix} \widehat{b}_0 \\ \widehat{\mathbf{b}}_1 \end{bmatrix} = \left(\begin{bmatrix} \mathbf{u}' \\ \mathbf{X}' \end{bmatrix} \left[\mathbf{u} \;\; \mathbf{X}\right] \right)^{-1} \begin{bmatrix} \mathbf{u}' \\ \mathbf{X}' \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{u}'\mathbf{u} & \mathbf{u}'\mathbf{X} \\ \mathbf{X}'\mathbf{u} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u}'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \end{bmatrix}$$

da cui, tenuto conto della formula di inversione di una matrice a blocchi e posto

$$\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{u}\mathbf{u}' ,$$

dopo qualche passaggio si ottiene

$$\begin{aligned} \widehat{b}_0 &= \frac{1}{n}\left\{1 + \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{H}\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\frac{1}{n}\right\}\mathbf{u}'\mathbf{y} - \frac{1}{n}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{H}\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \frac{1}{n}\left\{\mathbf{u}'\mathbf{y} - \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{H}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}\mathbf{y}\right\} , \\ \widehat{\mathbf{b}}_1 &= -(\mathbf{X}'\mathbf{H}\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\frac{1}{n}\mathbf{u}'\mathbf{y} + (\mathbf{X}'\mathbf{H}\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \;\; (\mathbf{X}'\mathbf{H}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}\mathbf{y} . \end{aligned}$$

Si osservi che la matrice $\mathbf{H}$ qui sopra definita è simmetrica ($\mathbf{H} = \mathbf{H}'$) e idempotente ($\mathbf{H} = \mathbf{H}^2$) per cui possiamo anche scrivere (5)

---

(5) Si noti che $\mathbf{H}$ è la matrice di proiezione ortogonale nel complemento ortogonale del sottospazio generato da $\mathbf{u}$.

$$\hat{\mathbf{b}}_1 = \left((\mathbf{H}\mathbf{X})'(\mathbf{H}\mathbf{X})\right)^{-1}(\mathbf{H}\mathbf{X})'(\mathbf{H}\mathbf{y}) \, .$$

D'altro canto, si riconosce facilmente che

$$\mathbf{H}\mathbf{X} = \tilde{\mathbf{X}} \quad , \quad \mathbf{H}\mathbf{y} = \tilde{\mathbf{y}}$$

e, quindi, risulta intanto che

$$\boxed{\hat{\mathbf{b}}_1 = \hat{\tilde{\mathbf{b}}}}$$

Infine, noto $\hat{\tilde{\mathbf{b}}}$, si ha che

(4) $$\boxed{\hat{b}_0 = \frac{1}{n}\{\mathbf{u}'\mathbf{y} - \mathbf{u}'\mathbf{X}\hat{\tilde{\mathbf{b}}}\}}$$

OSSERVAZIONE 5. Vogliamo mostrare che la matrice $\tilde{\mathbf{X}}$ di ordine (n,p), ha rango pari a p, vale a dire è una matrice di pieno rango per colonne, e che pertanto $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ è invertibile, come si è implicitamente supposto nella (3).

A questo fine, osserviamo intanto che, essendo $\mathbf{Z} = [\mathbf{u} \;\; \mathbf{X}]$ e $r(\mathbf{Z}) = 1+p$ risulta $r(\mathbf{X}) = p$.

Allora, tenuto conto della formula sullo sviluppo del determinante di una matrice a blocchi, si ha

$$\begin{aligned}\det(\mathbf{Z}'\mathbf{Z}) &= \det\begin{bmatrix} \mathbf{u}'\mathbf{u} & \mathbf{u}'\mathbf{X} \\ \mathbf{X}'\mathbf{u} & \mathbf{X}'\mathbf{X} \end{bmatrix} \\ &= \det\left(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{u}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{u}'\mathbf{X}\right)\det(\mathbf{u}'\mathbf{u}) \\ &\neq 0\end{aligned}$$

da cui

$$\begin{aligned}\det\left(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{u}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{u}'\mathbf{X}\right) &= \det\left(\mathbf{X}'(\mathbf{I} - \frac{1}{n}\mathbf{u}\mathbf{u}')\mathbf{X}\right) \\ &= \det\left(\mathbf{X}'\mathbf{H}\mathbf{X}\right) \\ &\neq 0\end{aligned}$$

e, quindi,

$$r(\mathbf{X}'\mathbf{H}'\mathbf{H}\mathbf{X}) = p = r(\mathbf{H}\mathbf{X}) = (\tilde{\mathbf{X}}) \, .$$

OSSERVAZIONE 6. Si noti che il vettore dei residui $\hat{\tilde{\mathbf{e}}}$ che si ottiene come differenza tra $\tilde{\mathbf{y}}$ e la proiezione ortogonale

$$\boxed{\hat{\tilde{\mathbf{y}}} = \tilde{\mathbf{X}}\hat{\mathbf{b}}_1 = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}}$$

di $\tilde{\mathbf{y}}$ nel sottospazio generato dai vettori colonna di $\tilde{\mathbf{X}}$ è eguale al vettore dei residui $\hat{\mathbf{e}}$ che si ottiene come differenza tra $\mathbf{y}$ e la proiezione ortogonale $\hat{\mathbf{y}}$ di $\mathbf{y}$ nel sottospazio generato dai vettori colonna di $\mathbf{Z}$.

Infatti, posto

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{Z}\hat{\mathbf{b}} = \mathbf{y} - [\mathbf{u} \;\; \mathbf{X}]\begin{bmatrix} \hat{b}_0 \\ \hat{\mathbf{b}}_1 \end{bmatrix} = \mathbf{y} - \mathbf{u}\hat{b}_0 - \mathbf{X}\hat{\mathbf{b}}_1 \, ,$$

premoltiplicando il primo e l'ultimo membro per $\mathbf{H}$ e osservando che

$$\mathbf{H}\hat{\mathbf{e}} = (\mathbf{I} - \frac{\mathbf{u}\mathbf{u}'}{n})\hat{\mathbf{e}} = \hat{\mathbf{e}} \quad , \quad \mathbf{H}\mathbf{u} = (\mathbf{I} - \frac{\mathbf{u}\mathbf{u}'}{n})\mathbf{u} = \mathbf{0}$$

si ottiene

$$\boxed{\hat{\mathbf{e}} = \tilde{\mathbf{y}} - \hat{\tilde{\mathbf{y}}} = \hat{\tilde{\mathbf{e}}}}$$

OSSERVAZIONE 7. Si noti che risulta

$$\boxed{\hat{\tilde{\mathbf{y}}} = \tilde{\mathbf{X}}\hat{\mathbf{b}}_1 = \mathbf{H}\mathbf{X}\hat{\mathbf{b}}_1 = \mathbf{H}\mathbf{u}\hat{b}_0 + \mathbf{H}\mathbf{X}\hat{\mathbf{b}}_1 = \mathbf{H}\hat{\mathbf{y}} = (\mathbf{I} - \frac{\mathbf{u}\mathbf{u}'}{n})\hat{\mathbf{y}} = \hat{\mathbf{y}} - \bar{\mathbf{y}}}$$

6 SCOMPONIBILITÀ DELLA DEVIANZA E INDICE DI DETERMINAZIONE LINEARE

Applicando il teorema di Pitagora ai vettori (ortogonali) $\hat{\tilde{\mathbf{y}}}$ ed $\hat{\mathbf{e}}$, si ottiene (Fig. 2)

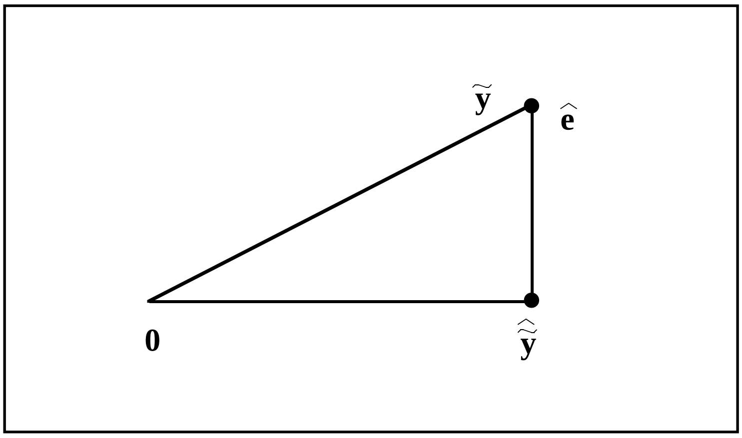
$$\boxed{\tilde{\mathbf{y}}'\,\tilde{\mathbf{y}} = \hat{\tilde{\mathbf{y}}}'\,\hat{\tilde{\mathbf{y}}} + \hat{\mathbf{e}}'\,\hat{\mathbf{e}}}$$

Tenuto conto dell'Osservazione 7, si ha poi che

$$\boxed{\tilde{\mathbf{y}}'\,\tilde{\mathbf{y}} = \Sigma_i\,(y_i - \bar{y})^2 \quad , \quad \hat{\tilde{\mathbf{y}}}'\,\hat{\tilde{\mathbf{y}}} = \Sigma_i\,(\hat{y}_i - \bar{y})^2 \quad , \quad \hat{\mathbf{e}}'\,\hat{\mathbf{e}} = \Sigma_i\,\hat{e}_i^2}$$

I termini di qui sopra ricevono la denominazione, rispettivamente, di *devianza totale*, *devianza di regressione*, *devianza residua* e il loro legame esprime la cosiddetta *scomponibilità della devianza*.

Fig. 2



OSSERVAZIONE 8. In generale, la proprietà di scomponibilità della devianza *non* risulta verificata in un modello in cui manchi il coefficiente b_0 [6]. ▮

---

(6) Per una discussione di questo punto si veda: Leoni, R., *Una osservazione sulla scomponibilità della devianza nel modello di regressione lineare multipla "senza intercetta"*, Rivista di Statistica Applicata, N. 4, 1985.

Il rapporto

$$\boxed{\rho = \frac{\widehat{\widetilde{\mathbf{y}}}'\widehat{\widetilde{\mathbf{y}}}}{\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}}} = \frac{\Sigma_i\,(\widehat{y}_i - \overline{y})^2}{\Sigma_i\,(y_i - \overline{y})^2} = 1 - \frac{\widehat{\mathbf{e}}'\widehat{\mathbf{e}}}{\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}}} = 1 - \frac{\Sigma_i\,\widehat{e}_i^2}{\Sigma_i\,(y_i - \overline{y})^2}}$$

riceve la denominazione di *indice di determinazione lineare*.

Da un punto di vista geometrico, ρ è interpretabile come il quadrato del coseno dell'angolo formato dai vettori $\widetilde{\mathbf{y}}$ e $\widehat{\widetilde{\mathbf{y}}}$, cioè $\rho = \cos^2(\widetilde{\mathbf{y}}, \widehat{\widetilde{\mathbf{y}}})$.

Tale indice assume il valore 1 se e soltanto se, per ogni $i = 1, \ldots, n$, risulta $y_i = \widehat{y}_i$, ovvero quando il vettore $\widetilde{\mathbf{y}}$ è eguale al vettore $\widehat{\widetilde{\mathbf{y}}}$; assume, invece, il valore 0 se e soltanto se, per ogni $i = 1, \ldots, n$, risulta $\widehat{y}_i = \overline{y}$, vale a dire quando $\widehat{\widetilde{\mathbf{y}}} = \mathbf{0}$.

L'indice di determinazione lineare rappresenta quindi una misura del grado di accostamento tra i valori osservati e i valori teorici.

OSSERVAZIONE 9. Nel caso di un modello di regressione semplice, la radice quadrata di ρ – a cui è attribuito un segno (sgn), quello stesso di $\widehat{b}_1$ (*concordanza positiva* o *negativa*) – è solitamente indicata con r e riceve la denominazione di *coefficiente di correlazione lineare semplice*; quindi, si ha che $-1 \le r \le +1$.

Come si verifica facilmente, r può essere espresso nelle forme

$$\begin{aligned}
r &= \operatorname{sgn}(\widehat{b}_1)\sqrt{\frac{\widehat{\widetilde{\mathbf{y}}}'\widehat{\widetilde{\mathbf{y}}}}{\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}}}} &&= \operatorname{sgn}(\widehat{b}_1)\sqrt{\frac{(\widetilde{\mathbf{y}} - \widehat{\mathbf{e}})'\widetilde{\mathbf{x}}_1\widehat{b}_1}{\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}}}} \\
&= \operatorname{sgn}(\widehat{b}_1)\sqrt{\frac{\widetilde{\mathbf{y}}'\widetilde{\mathbf{x}}_1\widehat{b}_1 - \widehat{\mathbf{e}}'\widetilde{\mathbf{x}}_1\widehat{b}_1}{\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}}}} &&= \operatorname{sgn}(\widehat{b}_1)\sqrt{\frac{\widetilde{\mathbf{y}}'\widetilde{\mathbf{x}}_1\widehat{b}_1}{\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}}}} \\
&= \operatorname{sgn}(\widehat{b}_1)\sqrt{\frac{(\widetilde{\mathbf{y}}'\widetilde{\mathbf{x}}_1)^2}{(\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}})\,(\widetilde{\mathbf{x}}_1'\widetilde{\mathbf{x}}_1)}} &&= \operatorname{sgn}(\widehat{b}_1)\frac{\left|\widetilde{\mathbf{y}}'\widetilde{\mathbf{x}}_1\right|}{\sqrt{(\widetilde{\mathbf{y}}'\widetilde{\mathbf{y}})\,(\widetilde{\mathbf{x}}_1'\widetilde{\mathbf{x}}_1)}}\,.
\end{aligned}$$

ESEMPIO 3. Riprendendo l'Esempio 1, risulta (p = 2)

$$\tilde{\mathbf{y}} = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 0 \end{bmatrix} \quad , \quad \tilde{\mathbf{X}} = \begin{bmatrix} -1 & 0 \\ -1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} .$$

Pertanto,

$$\hat{\mathbf{b}}_1 = \left(\begin{bmatrix} -1 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ -1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 & -1 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} ,$$

$$\hat{\mathrm{b}}_0 = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ -1 & -1 \\ 1 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 3$$

e, inoltre,

$$\hat{\tilde{\mathbf{y}}} = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix} \quad , \quad \hat{\mathbf{e}} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad , \quad \rho = \frac{6}{8} = \frac{3}{4} .$$

8 CAMBIAMENTI DI UNITÀ DI MISURA

Ci proponiamo di mostrare come muta $\hat{\mathbf{b}}_1 = \hat{\tilde{\mathbf{b}}}$ quando cambiano le unità di misura in cui sono espresse le determinazioni delle variabili considerate.

A questo fine, posto ($k_j > 0$, per $j = 1, \ldots, p$)

$$\mathbf{K} = \mathrm{diag}(k_1 , \ldots , k_p) \ ,$$

siano $(k > 0)$

$$\tilde{\mathbf{y}}^* = \tilde{\mathbf{y}}k \quad , \quad \tilde{\mathbf{X}}^* = \tilde{\mathbf{X}}\mathbf{K} .$$

Allora ($(\mathbf{K}^{-1})' = \mathbf{K}^{-1}$),

$$\begin{aligned}
(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} &= ((\tilde{\mathbf{X}}^*\mathbf{K}^{-1})'(\tilde{\mathbf{X}}^*\mathbf{K}^{-1}))^{-1} = \mathbf{K}((\tilde{\mathbf{X}}^*)'(\tilde{\mathbf{X}}^*))^{-1}\mathbf{K} , \\
\tilde{\mathbf{X}}'\tilde{\mathbf{y}} &= \mathbf{K}^{-1}(\tilde{\mathbf{X}}^*)'\tilde{\mathbf{y}}^* k^{-1}
\end{aligned}$$

e, quindi,

$$\begin{aligned}
\hat{\mathbf{b}}_1 &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \mathbf{K}((\tilde{\mathbf{X}}^*)'(\tilde{\mathbf{X}}^*))^{-1}\mathbf{K}\mathbf{K}^{-1}(\tilde{\mathbf{X}}^*)'\tilde{\mathbf{y}}^* k^{-1} \\
&= k^{-1}\mathbf{K}((\tilde{\mathbf{X}}^*)'(\tilde{\mathbf{X}}^*))^{-1}(\tilde{\mathbf{X}}^*)'\tilde{\mathbf{y}}^* = k^{-1}\mathbf{K}\hat{\mathbf{b}}_1^*
\end{aligned}$$

ovvero

$$\boxed{\hat{\mathbf{b}}_1^* = k\mathbf{K}^{-1}\hat{\mathbf{b}}_1}$$

A sua volta, si verifica facilmente che ($\mathbf{K}' = \mathbf{K}$)

$$\boxed{\rho = \frac{\hat{\tilde{\mathbf{y}}}'\hat{\tilde{\mathbf{y}}}}{\tilde{\mathbf{y}}'\tilde{\mathbf{y}}} = \frac{(\hat{\tilde{\mathbf{y}}}^*)'(\hat{\tilde{\mathbf{y}}}^*)}{(\tilde{\mathbf{y}}^*)'(\tilde{\mathbf{y}}^*)} = \cos^2(\tilde{\mathbf{y}}^*, \hat{\tilde{\mathbf{y}}}^*)}$$

In conclusione, cambiando le unità di misura, il vettore dei coefficienti di regressione $\hat{\mathbf{b}}_1$ varia nel modo indicato qui sopra; invece, l'indice di determinazione lineare, come è facilmente intuibile quando si rifletta sul significato geometrico di ρ, rimane invariato.

9 REGRESSIONE PARZIALE

Circa il significato da attribuire ai coefficienti di regressione $\widehat{b}_1, \ldots, \widehat{b}_p$, si può intanto osservare che, considerata la *funzione di regressione*

$$\widehat{Y} = \widehat{b}_0 + \widehat{b}_1 X_1 + \ldots + \widehat{b}_p X_p ,$$

chiaramente $\widehat{b}_j$ $(j = 1, \ldots, p)$ esprime la variazione (incremento o decremento) che subisce la variabile dipendente $\widehat{Y}$ allorché la variabile indipendente X_j subisce un incremento unitario, *fermo restando il valore assunto dalle altre variabili indipendenti.*

Oltre a quello ora indicato, il coefficiente di regressione $\widehat{b}_j$ presenta un altro significato che vogliamo porre in evidenza.

A questo fine, riprendiamo in considerazione la matrice $\mathbf{Z}$ definita in precedenza e supponiamo di ripartirla nel seguente modo

$$\mathbf{Z} = \left[\mathbf{u} \;\; \mathbf{x}_1 \;\; \cdots \;\; \mathbf{x}_{p-1} \;\vdots\; \mathbf{x}_p\right] = \left[\mathbf{Z}_{(p-1)} \;\; \mathbf{x}_p\right] .$$

Siano

$$\widehat{\mathbf{y}}_{\mathbf{z}(p-1)} = \mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{y} \;\; , \;\; \widehat{\mathbf{x}}_{p,\mathbf{z}(p-1)} = \mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{x}_p$$

i vettori che si ottengono dalla regressione di $\mathbf{y}$ e $\mathbf{x}_p$ rispetto ai vettori colonna che compongono $\mathbf{Z}_{(p-1)}$, ovverosia le proiezioni ortogonali di $\mathbf{y}$ e $\mathbf{x}_p$ nel sottospazio generato dai vettori colonna di $\mathbf{Z}_{(p-1)}$.

Con un linguaggio assai espressivo – al quale, tuttavia, non sempre corrisponde un significato sostanziale – si suole dire che tali vettori *rappresentano l'influenza determinata dalle variabili* $X_1, \ldots, X_{p-1}$ *sulle variabili* Y e X_p.

Ne consegue che i vettori

$$\widetilde{\mathbf{y}}_{\mathbf{z}(p-1)} = \mathbf{y} - \widehat{\mathbf{y}}_{\mathbf{z}(p-1)} \;\; , \;\; \widetilde{\mathbf{x}}_{p,\mathbf{z}(p-1)} = \mathbf{x}_p - \widehat{\mathbf{x}}_{p,\mathbf{z}(p-1)}$$

rappresentano ciò che rimane dopo che tale influenza è stata eliminata.

Eseguendo la regressione di $\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}$ rispetto a $\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}$ – ovvero, proiettando ortogonalmente $\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}$ nel sottospazio generato da $\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}$ – si ottiene

$$\boxed{\hat{\tilde{\mathbf{y}}}_{\mathbf{z}(p-1)} = \tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}(\tilde{\mathbf{x}}'_{p,\mathbf{z}(p-1)}\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)})^{-1}\tilde{\mathbf{x}}'_{p,\mathbf{z}(p-1)}\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}}$$

Il coefficiente

$$\boxed{\hat{b}_{p,\mathbf{z}(p-1)} = (\tilde{\mathbf{x}}'_{p,\mathbf{z}(p-1)}\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)})^{-1}\tilde{\mathbf{x}}'_{p,\mathbf{z}(p-1)}\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}}$$

è detto *coefficiente di regressione lineare parziale* (o *netta*) di $\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}$ rispetto a $\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}$.

A sua volta, il coefficiente

$$\boxed{\begin{aligned} r_{\mathbf{z}(p-1)} &= \operatorname{sgn}(\hat{b}_{p,\mathbf{z}(p-1)})\sqrt{\frac{\hat{\tilde{\mathbf{y}}}'_{\mathbf{z}(p-1)}\hat{\tilde{\mathbf{y}}}_{\mathbf{z}(p-1)}}{\tilde{\mathbf{y}}'_{\mathbf{z}(p-1)}\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}}} \\ &= \operatorname{sgn}(\hat{b}_{p,\mathbf{z}(p-1)})\sqrt{\cos^2(\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}, \hat{\tilde{\mathbf{y}}}_{\mathbf{z}(p-1)})} \end{aligned}}$$

è denominato *coefficiente di correlazione lineare parziale* (o *netta*) tra $\tilde{\mathbf{y}}_{\mathbf{z}(p-1)}$ e $\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}$.

Si tratta, in entrambi i casi, di coefficienti calcolati dopo che l'influenza di $X_1, \dots, X_{p-1}$ su Y e X_p è stata eliminata o, come anche si dice, *al netto* dell'influenza di $X_1, \dots, X_{p-1}$.

OSSERVAZIONE 10. Con un procedimento analogo a quello utilizzato nella Osservazione 5 si può facilmente dimostrare mostrare che il vettore $\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}$ è diverso dal vettore zero e che pertanto $\tilde{\mathbf{x}}'_{p,\mathbf{z}(p-1)}\tilde{\mathbf{x}}_{p,\mathbf{z}(p-1)}$ è un numero reale non nullo, come si è implicitamente supposto qui sopra. ▮

Vogliamo mettere in evidenza, adesso, che $\hat{b}_{p,\mathbf{z}(p-1)}$ è uguale a $\hat{b}_p$, cosicché quest'ultimo ha anche il significato di coefficiente di regressione netta.

Eseguiamo, a tal fine, una ripartizione del vettore $\widehat{\mathbf{b}}$ nel seguente modo

$$\widehat{\mathbf{b}} = \begin{bmatrix} \widehat{b}_0 \\ \vdots \\ \widehat{b}_{p-1} \\ \cdots\cdots \\ \widehat{b}_p \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{b}}_{(p-1)} \\ \widehat{b}_p \end{bmatrix}.$$

Le posizioni fatte consentono di scrivere la (2) nella forma

$$\widehat{\mathbf{b}} = \begin{bmatrix} \widehat{\mathbf{b}}_{(p-1)} \\ \widehat{b}_p \end{bmatrix} = \left(\begin{bmatrix} \mathbf{Z}'_{(p-1)} \\ \mathbf{x}'_p \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{(p-1)} & \mathbf{x}_p \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{Z}'_{(p-1)} \\ \mathbf{x}'_p \end{bmatrix} \mathbf{y}$$
$$= \begin{bmatrix} \mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)} & \mathbf{Z}'_{(p-1)}\mathbf{x}_p \\ \mathbf{x}'_p\mathbf{Z}_{(p-1)} & \mathbf{x}'_p\mathbf{x}_p \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Z}'_{(p-1)}\mathbf{y} \\ \mathbf{x}'_p\mathbf{y} \end{bmatrix}$$

da cui, tenuto conto della formula di inversione di una matrice a blocchi, dopo qualche passaggio, si ottiene

$$\widehat{b}_p = \left(\mathbf{x}'_p\mathbf{x}_p - \mathbf{x}'_p\mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{x}_p\right)^{-1}\left(-\mathbf{x}'_p\mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{y} + \mathbf{x}'_p\mathbf{y}\right).$$

Ma – tenuto conto delle espressioni di $\widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)}$, $\widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)}$, $\widehat{\mathbf{y}}_{\mathbf{Z}(p-1)}$, $\widetilde{\mathbf{y}}_{\mathbf{Z}(p-1)}$ date in precedenza e del fatto che

$$\widehat{\mathbf{x}}'_{p,\mathbf{Z}(p-1)}\widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)} = \widehat{\mathbf{x}}'_{p,\mathbf{Z}(p-1)}(\mathbf{x}_p - \widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)})$$
$$= (\mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{x}_p)'(\mathbf{x}_p - \mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{x}_p) = 0$$

e

$$\widetilde{\mathbf{x}}'_{p,\mathbf{Z}(p-1)}\widehat{\mathbf{y}}_{\mathbf{Z}(p-1)} = (\mathbf{x}_p - \widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)})'\widehat{\mathbf{y}}_{\mathbf{Z}(p-1)}$$
$$= (\mathbf{x}_p - \mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{x}_p)'\mathbf{Z}_{(p-1)}(\mathbf{Z}'_{(p-1)}\mathbf{Z}_{(p-1)})^{-1}\mathbf{Z}'_{(p-1)}\mathbf{y} = 0$$

– si ha

$$\widehat{b}_p = (\mathbf{x}'_p\mathbf{x}_p - \mathbf{x}'_p\widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)})^{-1}(-\widehat{\mathbf{x}}'_{p,\mathbf{Z}(p-1)}\mathbf{y} + \mathbf{x}'_p\mathbf{y})$$
$$= (\mathbf{x}'_p(\mathbf{x}_p - \widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)}))^{-1}(\mathbf{x}_p - \widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)})'\mathbf{y}$$

$$
\begin{aligned}
&= (\mathbf{x}'_p \widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)})^{-1} (\widetilde{\mathbf{x}}'_{p,\mathbf{Z}(p-1)} \mathbf{y}) \\
&= (\mathbf{x}'_p \widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)} - \widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)} \widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)})^{-1} (\widetilde{\mathbf{x}}'_{p,\mathbf{Z}(p-1)} \mathbf{y} - \widetilde{\mathbf{x}}'_{p,\mathbf{Z}(p-1)} \widehat{\mathbf{y}}_{p,\mathbf{Z}(p-1)}) \\
&= (\widetilde{\mathbf{x}}'_{p,\mathbf{Z}(p-1)} \widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)})^{-1} (\widetilde{\mathbf{x}}'_{p,\mathbf{Z}(p-1)} \widetilde{\mathbf{y}}_{\mathbf{Z}(p-1)})
\end{aligned}
$$

ovvero

$$
\boxed{\widehat{b}_p = \widehat{b}_{p,\mathbf{Z}(p-1)}}
$$

Ovviamente, il ragionamento ora svolto può essere ripetuto per ciascuno degli altri coefficienti di regressione e, quindi, a ciascuno di essi può essere attribuito il significato di coefficiente di regressione netta.

ESEMPIO 4. Riprendendo l'Esempio 1 e con le notazioni ora introdotte, si ha che (p = 2)

$$
\mathbf{Z}_{(p-1)} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \mathbf{x}_p = \begin{bmatrix} -1 \\ -1 \\ -2 \\ 0 \end{bmatrix}
$$

$$
\widehat{\mathbf{y}}_{\mathbf{Z}(p-1)} = \begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix} \quad , \quad \widehat{\mathbf{x}}_{p,\mathbf{Z}(p-1)} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \quad , \quad \widetilde{\mathbf{y}}_{\mathbf{Z}(p-1)} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \quad , \quad \widetilde{\mathbf{x}}_{p,\mathbf{Z}(p-1)} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}
$$

e, quindi,

$$
\widehat{b}_{p,\mathbf{Z}(p-1)} = 1 = \widehat{b}_p \quad , \quad \widehat{\widetilde{\mathbf{y}}}_{\mathbf{Z}(p-1)} = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \quad , \quad r_{\mathbf{Z}(p-1)} = \frac{1}{2\sqrt{2}} . \qquad \blacksquare
$$

Supponiamo adesso di eseguire una ripartizione della matrice **Z** nel seguente modo

$$
\mathbf{Z} = [\mathbf{u} \;\; \mathbf{x}_1 \; \cdots \; \mathbf{x}_h \; \vdots \; \mathbf{x}_{h+1} \; \cdots \; \mathbf{x}_p] = [\mathbf{Z}_{(h)} \;\; \mathbf{X}_{(p-h)}] .
$$

Siano (j = h+1, ... , p)

$$\hat{\mathbf{y}}_{\mathbf{Z}(h)} = \mathbf{Z}_{(h)}(\mathbf{Z}'_{(h)}\mathbf{Z}_{(h)})^{-1}\mathbf{Z}'_{(h)}\mathbf{y} \quad , \quad \hat{\mathbf{x}}_{j,\mathbf{Z}(h)} = \mathbf{Z}_{(h)}(\mathbf{Z}'_{(h)}\mathbf{Z}_{(h)})^{-1}\mathbf{Z}'_{(h)}\mathbf{x}_j$$

i vettori che si ottengono dalla regressione di $\mathbf{y}$ e $\mathbf{x}_j$ rispetto ai vettori colonna che compongono $\mathbf{Z}_{(h)}$, ovverosia le proiezioni ortogonali di $\mathbf{y}$ e $\mathbf{x}_j$ nel sottospazio generato dai vettori colonna di $\mathbf{Z}_{(h)}$.

Chiaramente, tali vettori *rappresentano l'influenza determinata dalle variabili* $X_1, \dots, X_h$ *sulle variabili* Y e X_j.

Ne consegue che i vettori

$$\tilde{\mathbf{y}}_{\mathbf{Z}(h)} = \mathbf{y} - \hat{\mathbf{y}}_{\mathbf{Z}(h)} \quad , \quad \tilde{\mathbf{x}}_{j,\mathbf{Z}(h)} = \mathbf{x}_j - \hat{\mathbf{x}}_{j,\mathbf{Z}(h)}$$

rappresentano ciò che rimane dopo che tale influenza è stata eliminata.

Ciò premesso, operando la regressione di $\tilde{\mathbf{y}}_{\mathbf{Z}(h)}$ rispetto ai vettori colonna della matrice (7)

$$\tilde{\mathbf{X}}_{p,\mathbf{Z}(h)} = \left[\tilde{\mathbf{x}}_{h+1,\mathbf{Z}(h)} \ \cdots \ \tilde{\mathbf{x}}_{p,\mathbf{Z}(h)}\right],$$

si ottiene

$$\boxed{\hat{\tilde{\mathbf{y}}}_{\mathbf{Z}(h)} = \tilde{\mathbf{X}}_{p,\mathbf{Z}(h)}\left(\tilde{\mathbf{X}}'_{p,\mathbf{Z}(h)}\tilde{\mathbf{X}}_{p,\mathbf{Z}(h)}\right)^{-1}\tilde{\mathbf{X}}'_{p,\mathbf{Z}(h)}\tilde{\mathbf{y}}_{\mathbf{Z}(h)}}$$

Il vettore

$$\boxed{\hat{\mathbf{b}}_{p,\mathbf{Z}(h)} = \left(\tilde{\mathbf{X}}'_{p,\mathbf{Z}(h)}\tilde{\mathbf{X}}_{p,\mathbf{Z}(h)}\right)^{-1}\tilde{\mathbf{X}}'_{p,\mathbf{Z}(h)}\tilde{\mathbf{y}}_{\mathbf{Z}(h)}}$$

è anche detto *vettore dei coefficienti di regressione multipla-parziale* di $\tilde{\mathbf{y}}_{\mathbf{Z}(h)}$ rispetto a $\tilde{\mathbf{X}}_{p,\mathbf{Z}(h)}$.

Ora, con un procedimento analogo a quello esposto in precedenza, si dimostra che gli elementi $\hat{b}_{h+1,\mathbf{Z}(h)}, \dots, \hat{b}_{p,\mathbf{Z}(h)}$ di $\hat{\mathbf{b}}_{p,\mathbf{Z}(h)}$ sono gli stessi, nell'ordine, dei coefficienti di regressione $\hat{b}_{h+1}, \dots, \hat{b}_p$ che si accompagnano alle variabili $X_{h+1}, \dots, X_p$ e che sono parte del vettore $\hat{\mathbf{b}}$.

(7) Si può facilmente dimostrare, seguendo una linea di ragionamento simile a quella seguita nella Osservazione 5, che tale matrice è di pieno rango per colonne.

Ne consegue che questi ultimi, considerati congiuntamente, assumono il significato di coefficienti di regressione multipla parziale, mentre, presi ad uno ad uno, assumono quello di coefficienti di regressione parziale o netta.

10 IL MODELLO MULTIVARIATO DI REGRESSIONE LINEARE

Finora ci siamo occupati del caso in cui il gruppo delle variabili dipendenti fosse costituito da una sola variabile Y.

Supponiamo adesso che tale gruppo sia composto da $t \geq 2$ variabili $Y_1, \dots, Y_t$.

Considerati i t modelli di regressione

$$(5) \qquad \boxed{\begin{array}{l} y_{i1} = a_{01} + a_{11} x_{i1} + \cdots + a_{p1} x_{ip} + e_{i1} \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ y_{it} = a_{0t} + a_{1t} x_{i1} + \cdots + a_{pt} x_{ip} + e_{it} \end{array}}$$

e posto ($s = 1,\dots,t$)

$$\mathbf{y}_s = \begin{bmatrix} y_{1s} \\ \vdots \\ y_{ns} \end{bmatrix} \quad , \quad \mathbf{a}_s = \begin{bmatrix} a_{0s} \\ \vdots \\ a_{ps} \end{bmatrix} \quad , \quad \mathbf{e}_s = \begin{bmatrix} e_{1s} \\ \vdots \\ e_{ns} \end{bmatrix}$$

$$\mathbf{Y} = [\mathbf{y}_1 \cdots \mathbf{y}_t] \quad , \quad \mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_t] \quad , \quad \mathbf{E} = [\mathbf{e}_1 \cdots \mathbf{e}_t]$$

è intanto ovvio che possiamo scrivere, più compattamente, la (5) nella forma (*modello multivariato di regressione lineare*)

$$(5') \qquad \boxed{\mathbf{Y} = \mathbf{Z}\mathbf{A} + \mathbf{E}}$$

Inoltre, poiché il metodo dei m.q., applicato a *ciascuno* dei t modelli di regressione di cui sopra, fornisce

$$\widehat{\mathbf{a}}_s = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_s \ ,$$

si ha che

$$(6) \qquad \boxed{\widehat{\mathbf{A}} = [\widehat{\mathbf{a}}_1 \cdots \widehat{\mathbf{a}}_t] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}}$$

Vogliamo adesso mostrare che, applicando il metodo dei m.q. *simultaneamente* ai t modelli di regressione scritti nella (5), si perviene allo stesso risultato.

In effetti, considerata la quantità

$$\boxed{\begin{aligned} S &= \Sigma_s \, \Sigma_i \, e_{is}^2 = \mathrm{tr}\{\mathbf{E}'\mathbf{E}\} = \mathrm{tr}\{(\mathbf{Y}-\mathbf{ZA})'(\mathbf{Y}-\mathbf{ZA})\} \\ &= \mathrm{tr}\{\mathbf{Y}'\mathbf{Y}\} - \mathrm{tr}\{\mathbf{Y}'\mathbf{ZA}\} - \mathrm{tr}\{\mathbf{A}'\mathbf{Z}'\mathbf{Y}\} + \mathrm{tr}\{\mathbf{A}'\mathbf{Z}'\mathbf{ZA}\} \\ &= \mathrm{tr}\{\mathbf{Y}'\mathbf{Y}\} - 2\,\mathrm{tr}\{\mathbf{Y}'\mathbf{ZA}\} + \mathrm{tr}\{\mathbf{A}'\mathbf{Z}'\mathbf{ZA}\} \end{aligned}}$$

la matrice $\widehat{\mathbf{A}}$ che minimizza S deve esser tale che la derivata di S rispetto ad $\mathbf{A}$, calcolata in $\widehat{\mathbf{A}}$, risulti nulla.

Ma,

$$\frac{\partial S}{\partial \mathbf{A}} = -\,2\mathbf{Z}'\mathbf{Y} + 2\mathbf{Z}'\mathbf{ZA}$$

ed eguagliando a zero tale espressione, si ottiene immediatamente la (6).

OSSERVAZIONE 11. Si noti che applicando l'operatore *vec* a entrambi i membri della (5'), si ottiene

$$\mathrm{vec}(\mathbf{Y}) = \mathrm{vec}(\mathbf{ZA}) + \mathrm{vec}(\mathbf{E}) = (\mathbf{I}\otimes\mathbf{Z})\mathrm{vec}(\mathbf{A}) + \mathrm{vec}(\mathbf{E})$$

da cui

$$\begin{aligned} \mathrm{vec}(\widehat{\mathbf{A}}) &= \big((\mathbf{I}\otimes\mathbf{Z})'(\mathbf{I}\otimes\mathbf{Z})\big)^{-1}(\mathbf{I}\otimes\mathbf{Z})'\mathrm{vec}(\mathbf{Y}) \\ &= \big((\mathbf{I}\otimes\mathbf{Z}')(\mathbf{I}\otimes\mathbf{Z})\big)^{-1}(\mathbf{I}\otimes\mathbf{Z}')\mathrm{vec}(\mathbf{Y}) \\ &= \big(\mathbf{I}\otimes(\mathbf{Z}'\mathbf{Z})\big)^{-1}(\mathbf{I}\otimes\mathbf{Z}')\mathrm{vec}(\mathbf{Y}) \\ &= \big(\mathbf{I}\otimes(\mathbf{Z}'\mathbf{Z})^{-1}\big)(\mathbf{I}\otimes\mathbf{Z}')\mathrm{vec}(\mathbf{Y}) \\ &= \big(\mathbf{I}\otimes[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\big)\mathrm{vec}(\mathbf{Y}) \end{aligned}$$

la quale non è altro che un modo diverso di scrivere la (6).

OSSERVAZIONE 12. Si potrebbe dimostrare che la matrice dei coefficienti di regressione $\widehat{\mathbf{A}}$ che minimizza S, ovvero $\mathrm{tr}(\mathbf{E}'\mathbf{E})$, minimizza $\det(\mathbf{E}'\mathbf{E})$ e, quindi, anche la varianza generalizzata dei residui, vale a dire $\det(\frac{1}{n}\mathbf{E}'\mathbf{E})$.