STATISTICA I e II

Renato Guseo

Seminario:
ASPETTI ALGEBRICO–GEOMETRICI NEI MODELLI LINEARI

Padova, 2 ottobre 2009

Estratto da: Guseo R. (2006). *Statistica*, Cedam, Padova, pp. 252–266.

1 Interpretazione geometrica del modello lineare

Il modello di regressione multipla può essere rappresentato nella usuale forma esplicita,

$$ {}_Kx_i = \alpha_0 + \alpha_1\, {}_1x_i + \alpha_2\, {}_2x_i + \cdots + \alpha_{K-1}\, {}_{K-1}x_i + \varepsilon_i \qquad i = 1, 2, \cdots, N, \tag{1} $$

ove la diseguaglianza, $N \geq K$, assicura una condizione necessaria per l'esistenza di un'unica soluzione nell'ambito dei minimi quadrati.

Le osservazioni congiunte di ciascuna variabile ${}_\ell X$, $\ell = 1, 2, \cdots, K$, – considerate come successione ordinata – costituiscono un vettore, ${}_\ell\boldsymbol{x}$, nello spazio lineare R^N. In questo spazio, il modello regressivo lineare assume la forma estesa che segue,

$$ \begin{pmatrix} {}_Kx_1 \\ {}_Kx_2 \\ \vdots \\ {}_Kx_N \end{pmatrix} = \begin{pmatrix} 1 & {}_1x_1 & {}_2x_1 & \cdots & {}_{K-1}x_1 \\ 1 & {}_1x_2 & {}_2x_2 & \cdots & {}_{K-1}x_2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & {}_1x_N & {}_2x_N & \cdots & {}_{K-1}x_N \end{pmatrix} \cdot \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{K-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix} \tag{2} $$

ove il prodotto, "·", è l'usuale prodotto *righe per colonne.*

Più sinteticamente, seguendo la notazione vettoriale si ha,

$$ {}_K\boldsymbol{x} = (\mathbf{1}, {}_1\boldsymbol{x}, {}_2\boldsymbol{x}, \cdots, {}_{K-1}\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3} $$

ovvero, usando la matrice $\boldsymbol{X}$ per accogliere i vettori $(\mathbf{1}, {}_1\boldsymbol{x}, {}_2\boldsymbol{x}, \cdots, {}_{K-1}\boldsymbol{x})$ e il vettore $\boldsymbol{y}$ per indicare ${}_K\boldsymbol{x}$, si ottiene la forma più sintetica,

$$ \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{4} $$

Le colonne di $\boldsymbol{X}$, combinate linearmente con il vettore $\boldsymbol{\beta} \in R^K$, definiscono – al variare di $\boldsymbol{\beta}$ in R^K – un vettore "medio" $\boldsymbol{\eta}$ che appartiene ad un *sottospazio* di R^N di dimensione al più K detto *varietà lineare* $\mathcal{V}_K$ o *spazio lineare generato dalle colonne di* $\boldsymbol{X}$. La dimensione della varietà lineare $\mathcal{V}_K$ dipende dal rango di $\boldsymbol{X}$. Se $\boldsymbol{X}$ è di rango pieno, le sue colonne costituiscono una *base* per $\mathcal{V}_K$ (non necessariamente ortonormale).

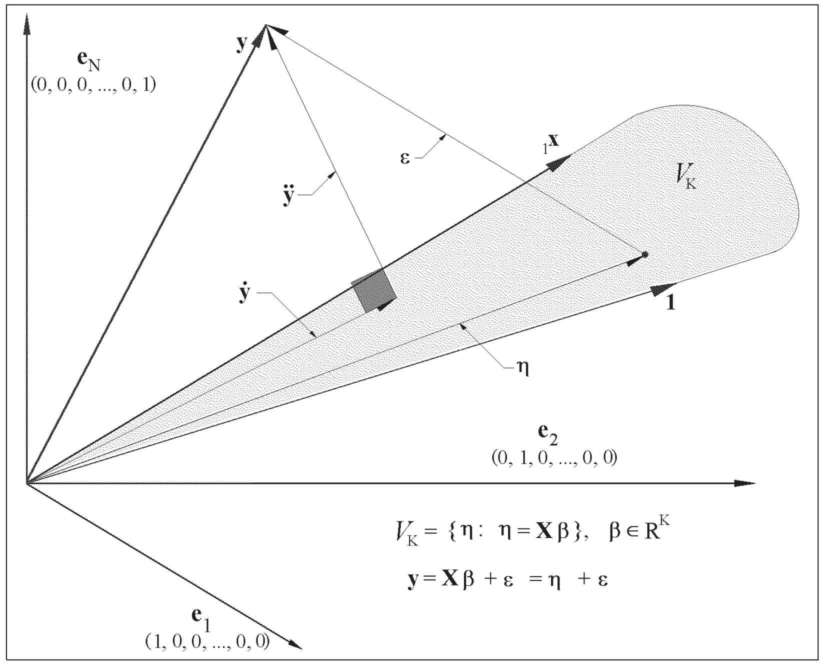

Figura 1: Varietà lineare $\mathcal{V}_K$, proiezione ortogonale e residuo.

Per rendere più immediata l'intuizione, si veda il grafico schematico della Figura 1 in cui il vettore $\boldsymbol{y} \in R^N$ non appartiene alla varietà lineare $\mathcal{V}_K$ e il vettore $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{\eta}$ esprime l'*errore residuo* che separa le osservazioni della variabile dipendente $_KX$, qui rappresentate da $\boldsymbol{y}$, da un generico modello medio $\boldsymbol{\eta} \in \mathcal{V}_K$.

La *devianza* è pari al quadrato del modulo del vettore errore residuo $\boldsymbol{\epsilon}$,

$$\delta^2(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}\ . \tag{5}$$

Criterio dei minimi quadrati.
La selezione del modello che soddisfa il criterio dei minimi quadrati si basa sulla minimizzazione della devianza, $\delta^2(\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$, rispetto al vettore dei parametri $\boldsymbol{\beta} \in R^K$, ovvero,

$$\min_{\boldsymbol{\beta}} \delta^2(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}). \tag{6}$$

Sviluppando la devianza si ottiene,

$$\delta^2(\boldsymbol{\beta}) = (\boldsymbol{y}' - \boldsymbol{\beta}'\boldsymbol{X}')(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}. \tag{7}$$

Derivando rispetto a $\boldsymbol{\beta}$ e uguagliando a zero si ha,

$$-\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{0}, \tag{8}$$

e, semplificando, si ottengono le seguenti *equazioni normali*

$$\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{X}'\boldsymbol{y} \tag{9}$$

Assumendo ora il rango pieno di $\boldsymbol{X}$, $r(\boldsymbol{X}) = K$ (e quindi necessariamente $N \geq K$), si ha che la matrice $\boldsymbol{X}'\boldsymbol{X}$ è di pari rango, $r(\boldsymbol{X}'\boldsymbol{X}) = K$, ed esiste l'inversa $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, per cui il sistema (9) possiede un'unica soluzione, precisamente,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{10}$$

Se $\boldsymbol{X}'\boldsymbol{X}$ non è invertibile, $r(\boldsymbol{X}'\boldsymbol{X}) < K$, esistono infinite soluzioni.

Il vettore $\hat{\boldsymbol{\beta}}$ ha come componenti le coordinate della proiezione ortogonale $\dot{\boldsymbol{y}}$ di $\boldsymbol{y}$ in $\mathcal{V}_K$ in relazione alla base $\boldsymbol{X}$.
La proiezione $\dot{\boldsymbol{y}}$ è definita infatti da una trasformazione lineare, precisamente,

$$\dot{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = {}_X\boldsymbol{P}\boldsymbol{y}, \tag{11}$$

ove ${}_X\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ è una matrice simmetrica ed *idempotente*, ${}_X\boldsymbol{P}\,{}_X\boldsymbol{P} = {}_X\boldsymbol{P}$, detta *proiettore* sullo spazio lineare $\mathcal{V}_K$ *generato* dalle colonne di $\boldsymbol{X}$.
Il vettore *residuo*, distinto in generale dal vettore *errore residuo* $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{\eta}$, si ottiene per differenza, $\ddot{\boldsymbol{y}} = \boldsymbol{y} - \dot{\boldsymbol{y}}$, ovvero,

$$\ddot{\boldsymbol{y}} = \boldsymbol{y} - \dot{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y} = (\boldsymbol{I} - {}_X\boldsymbol{P})\boldsymbol{y}, \tag{12}$$

ed è anch'esso funzione lineare di una matrice simmetrica e idempotente, precisamente, $\boldsymbol{I} - {}_X\boldsymbol{P}$.

Teorema (Ortogonalità del vettore residuo). Sia $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ un modello di regressione multipla con $\boldsymbol{y},\ \boldsymbol{\epsilon} \in R^N,\ \boldsymbol{\beta} \in R^K$ e $\boldsymbol{X}$ una matrice $N \times K$ di rango pieno, $r(\boldsymbol{X}) = K$. Il vettore residuo $\ddot{\boldsymbol{y}} = \boldsymbol{y} - \dot{\boldsymbol{y}}$, ove $\dot{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = {}_X\boldsymbol{P}\boldsymbol{y}$ con ${}_X\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, è *ortogonale* alla varietà lineare $\mathcal{V}_K \subset R^N$ generata dalle colonne di $\boldsymbol{X}$,

$$\ddot{\boldsymbol{y}} \perp \boldsymbol{X}\boldsymbol{\alpha}, \qquad \boldsymbol{\alpha} \in R^K. \tag{13}$$

Prova. È sufficiente controllare l'annullamento del prodotto scalare tra $\ddot{\boldsymbol{y}}$ e il generico vettore $\boldsymbol{X}\boldsymbol{\alpha}$ di $\mathcal{V}_K$, precisamente,

$$\ddot{\boldsymbol{y}}'\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{X}\boldsymbol{\alpha} = \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\alpha}) = 0.$$

$\square$

Si veda a scopo illustrativo la Figura 1.

1.1 Teorema di Pitagora

Conseguenza immediata della ortogonalità del vettore residuo $\ddot{\boldsymbol{y}}$ rispetto alla varietà lineare $\mathcal{V}_K$ è la proposizione che segue.

Teorema (Relazione pitagorica). Sotto le ipotesi del teorema precedente, le norme al quadrato dei vettori $\boldsymbol{y}$, $\dot{\boldsymbol{y}}$ e $\ddot{\boldsymbol{y}}$ soddisfano la relazione pitagorica,

$$\boldsymbol{y}'\boldsymbol{y} = \dot{\boldsymbol{y}}'\dot{\boldsymbol{y}} + \ddot{\boldsymbol{y}}'\ddot{\boldsymbol{y}}. \tag{14}$$

Prova. Per la idempotenza della matrice ${}_X\boldsymbol{P}$ si ha,

$$\dot{\boldsymbol{y}}'\dot{\boldsymbol{y}} = \boldsymbol{y}'\,{}_X\boldsymbol{P}'\,{}_X\boldsymbol{P}\boldsymbol{y} = \boldsymbol{y}'\,{}_X\boldsymbol{P}\boldsymbol{y},$$

e, analogamente, per la idempotenza della matrice $(\boldsymbol{I} - {}_X\boldsymbol{P})$ si ha,

$$\ddot{\boldsymbol{y}}'\ddot{\boldsymbol{y}} = \boldsymbol{y}'(\boldsymbol{I} - {}_X\boldsymbol{P})'(\boldsymbol{I} - {}_X\boldsymbol{P})\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - {}_X\boldsymbol{P})\boldsymbol{y},$$

per cui,

$$\dot{\boldsymbol{y}}'\dot{\boldsymbol{y}} + \ddot{\boldsymbol{y}}'\ddot{\boldsymbol{y}} = \boldsymbol{y}'\,{}_X\boldsymbol{P}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\,{}_X\boldsymbol{P}\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{y}.$$

□

1.2 Proiezione ortogonale in sottospazi nidificati.

Talvolta i modelli regressivi utilizzati differiscono per la soppressione di alcune componenti esplicative. È del tutto naturale chiedersi, in questi contesti, se il modello nidificato più semplice, inteso come contrazione di un modello più generale, non riduca eccessivamente l'informazione congiunta utilizzata nella direzione della previsione degli ordini di grandezza medi di una variabile $\boldsymbol{y}$ a partire da configurazioni esplicative assegnate $\boldsymbol{W}$.

Sia $\mathcal{V}_r$ un sottospazio lineare del sottospazio $\mathcal{V}_K$, $\mathcal{V}_r \subset \mathcal{V}_K \subset R^N$. Siano

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e}, \qquad \boldsymbol{\beta} \in R^K,\ \boldsymbol{\alpha} \in R^r,\ \boldsymbol{y}, \boldsymbol{\epsilon}, \boldsymbol{e} \in R^N, \tag{15}$$

due modelli regressivi lineari nidificati ove, in particolare,

$$\boldsymbol{W} = (\boldsymbol{X}, \boldsymbol{D}). \tag{16}$$

Sia $\dot{\boldsymbol{y}} = \boldsymbol{W}\hat{\boldsymbol{\beta}}$ la proiezione ortogonale di $\boldsymbol{y}$ su $\mathcal{V}_K$, sottospazio di R^N generato dalle colonne di $\boldsymbol{W}$. Analogamente, $\tilde{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\alpha}}$ è la proiezione di $\boldsymbol{y}$ su $\mathcal{V}_r$, sottospazio generato dal sottoinsieme di colonne $\boldsymbol{X}$ di $\boldsymbol{W}$.

Siano $\ddot{\boldsymbol{y}} = \boldsymbol{y} - \dot{\boldsymbol{y}}$ il vettore residuo del modello più ampio e $\tilde{\tilde{\boldsymbol{y}}} = \boldsymbol{y} - \tilde{\boldsymbol{y}}$ il vettore residuo del modello ridotto. La Figura 2 descrive in termini intuitivi le relazioni tra le quantità vettoriali coinvolte dai minimi quadrati.

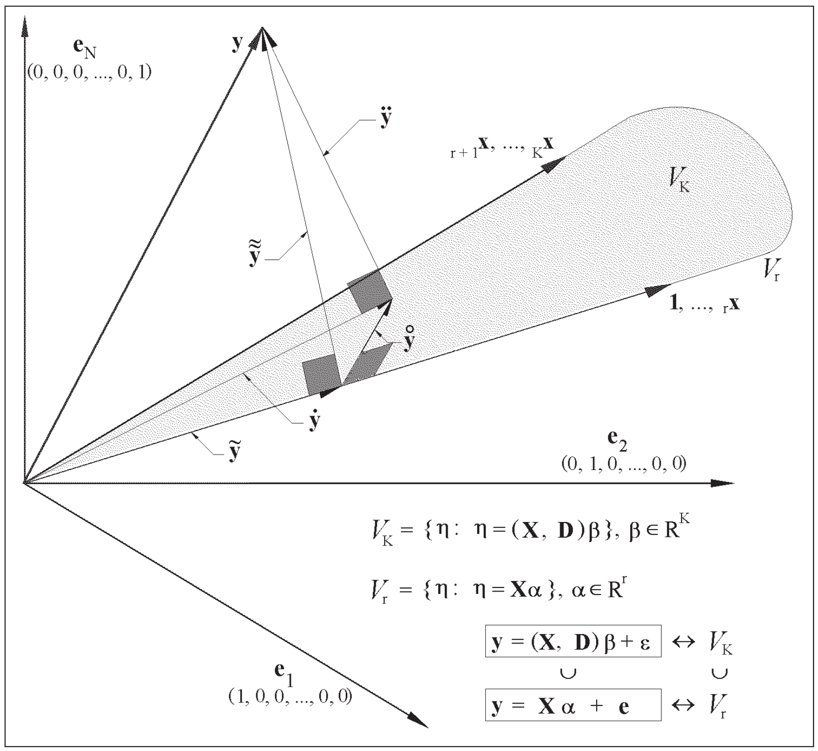

Figura 2: Proiezioni in varietà nidificate: $\mathcal{V}_K \supset \mathcal{V}_r,\ r \leq K$.

Il teorema generale più volte menzionato in relazione alla monotonia delle devianze residue per modelli appartenenti a famiglie monotone, oppure, con altra terminologia, a famiglie nidificate, trova una ovvia conferma nel lemma che segue.

Lemma (Monotonia). Sotto le assunzioni espresse dalla (15) e dalla (16), siano $\tilde{\tilde{\boldsymbol{y}}}$ e $\ddot{\boldsymbol{y}}$ i vettori dei residui, secondo i minimi quadrati, dei modelli ridotto e completo. La devianza residua del modello ridotto è *non inferiore* alla corrispondente del modello completo,

$$\delta^{*2}_{\boldsymbol{X}} = \delta^2(\hat{\boldsymbol{\alpha}}) = \tilde{\tilde{\boldsymbol{y}}}'\tilde{\tilde{\boldsymbol{y}}} \geq \ddot{\boldsymbol{y}}'\ddot{\boldsymbol{y}} = \delta^2(\hat{\boldsymbol{\beta}}) = \delta^{*2}_{\boldsymbol{W}} \ . \tag{17}$$

Prova. Per costruzione valgono le seguenti proprietà, $\mathring{\boldsymbol{y}} = \dot{\boldsymbol{y}} - \tilde{\boldsymbol{y}} \in \mathcal{V}_K$ e $\ddot{\boldsymbol{y}} \perp \mathcal{V}_K$. Per il teorema di Pitagora, (14), vale inoltre l'equazione

$$\tilde{\tilde{\boldsymbol{y}}}'\tilde{\tilde{\boldsymbol{y}}} = \ddot{\boldsymbol{y}}'\ddot{\boldsymbol{y}} + \mathring{\boldsymbol{y}}'\mathring{\boldsymbol{y}} \ .$$

Poiché $\mathring{\boldsymbol{y}}'\mathring{\boldsymbol{y}} \geq 0$, segue immediatamente la tesi. □

Con riferimento alle stesse ipotesi del lemma precedente è immediato dimostrare il seguente teorema che va sotto il nome di *proiezione in due stadi.*

Teorema (Proiezione in due stadi). Il vettore $\mathring{\boldsymbol{y}}$, differenza di proiezioni o differenza dei vettori residuo di due modelli regressivi lineari nidificati, detto talvolta *vettore guadagno*, $\mathring{\boldsymbol{y}} = \dot{\boldsymbol{y}} - \tilde{\boldsymbol{y}} = \ddot{\boldsymbol{y}} - \tilde{\tilde{\boldsymbol{y}}}$, è *ortogonale* al sottospazio ridotto $\mathcal{V}_r$,

$$\mathring{\boldsymbol{y}} \perp \mathcal{V}_r. \tag{18}$$

Prova. È sufficiente provare che $\mathring{\boldsymbol{y}}' \boldsymbol{X}\boldsymbol{\gamma} = 0$ ove $\boldsymbol{\gamma}$ è un generico vettore di R^r. Infatti, passando alle definizioni si ha,

$$\begin{aligned}
\mathring{\boldsymbol{y}}' \boldsymbol{X}\boldsymbol{\gamma} &= \boldsymbol{y}'[\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}' - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{X}\boldsymbol{\gamma} \\
&= \boldsymbol{y}'[\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{X} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}]\gamma \\
&= \boldsymbol{y}'[\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{X} - \boldsymbol{X}]\gamma.
\end{aligned}$$

Il proiettore $\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'$ non muta (talvolta si dice che lascia in sé) i vettori della base identificata dalle colonne di $\boldsymbol{W}$ e pertanto, poiché $\boldsymbol{X}$ rappresenta le prime r colonne di $\boldsymbol{W}$, si ha,

$$\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{X} = \boldsymbol{X},$$

e quindi,

$$\mathring{\boldsymbol{y}}' \boldsymbol{X}\boldsymbol{\gamma} = \boldsymbol{y}'[\boldsymbol{X} - \boldsymbol{X}]\gamma = 0.$$

$\square$

Adesso può risultare più evidente il significato della terminologia *proiezione in due stadi* utilizzata in apertura. Il vettore *guadagno* $\mathring{\boldsymbol{y}}$ funge da residuo ortogonale nella proiezione, secondo i minimi quadrati, di $\dot{\boldsymbol{y}}$ su $\mathcal{V}_r \subset \mathcal{V}_K$ ove, a sua volta, $\dot{\boldsymbol{y}}$ stesso è proiezione ortogonale di $\boldsymbol{y}$ su $\mathcal{V}_K$. Si esamini a questo proposito la Figura 2.

1.3 Proiezioni ortogonali in sottospazi comuni

In altri contesti si desidera studiare la relazione tra due variabili, ad esempio tra ${}_KX$ e ${}_sX$, tenendo sotto controllo la dipendenza di ciascuna di queste da un insieme comune di variabili $1, {}_1X, {}_2X, \cdots, {}_{r-1}X$, ove la prima è l'usuale variabile degenere adottata a partire dal sottomodello costante. Passando alla notazione matriciale si indica con la matrice $\boldsymbol{X} = (\mathbf{1}, {}_1\boldsymbol{x}, {}_2\boldsymbol{x}, \cdots, {}_{r-1}\boldsymbol{x})$ l'insieme dei vettori corrispondenti.

Sia $\mathcal{V}_r$ un sottospazio lineare di R^N generato dalle prime r colonne, $\boldsymbol{X}$, di una matrice $\boldsymbol{Z} = (\boldsymbol{X}, {}_s\boldsymbol{x}, {}_K\boldsymbol{x})$ ove ${}_s\boldsymbol{x}$ e ${}_K\boldsymbol{x}$ sono vettori colonna corrispondenti alle variabili statistiche ${}_sX$ e ${}_KX$, $N \geq r + 2$.

Caso elementare, $\mathcal{V}_r \equiv \mathbf{1}$.

Il caso più semplice è il caso in cui $\boldsymbol{X}$ è costituita dal solo vettore $\mathbf{1}$.

Siano ${}_s\dot{\boldsymbol{x}}$ e ${}_K\dot{\boldsymbol{x}}$ le proiezioni ortogonali di ${}_s\boldsymbol{x}$ e ${}_K\boldsymbol{x}$ su $\mathcal{V}_1$ e ${}_s\ddot{\boldsymbol{x}}$ e ${}_K\ddot{\boldsymbol{x}}$ i vettori residui corrispondenti, Figura 3. Si noti che in questo caso l'ortogonalità di ${}_s\ddot{\boldsymbol{x}}$ rispetto al vettore unitario $\mathbf{1}$, coincide con la ben nota proprietà della somma degli scarti lineari della media, ${}_s\ddot{\boldsymbol{x}}'\mathbf{1} = 0$. Analoga considerazione vale per ${}_K\ddot{\boldsymbol{x}}$ per cui ${}_K\ddot{\boldsymbol{x}}'\mathbf{1} = 0$.

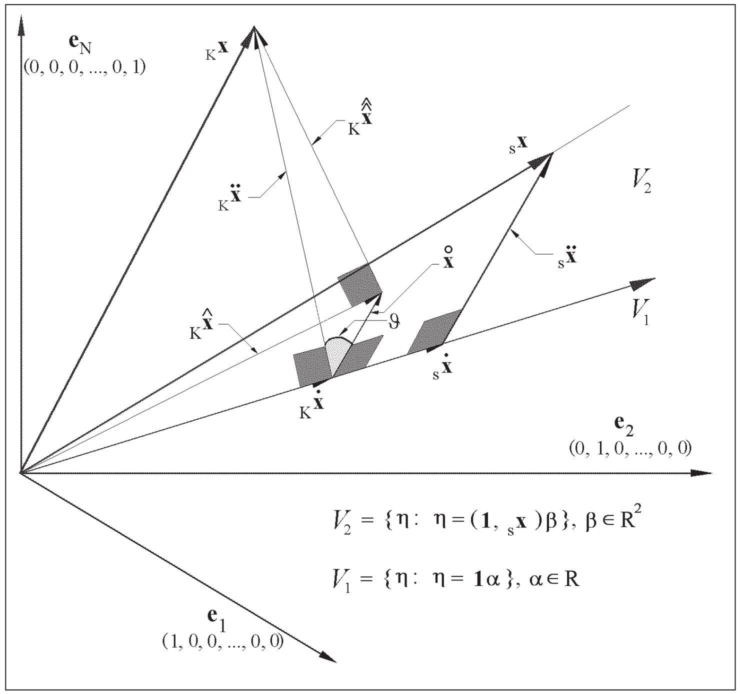

Figura 3: Proiezioni in sottospazi comuni: caso elementare $\mathcal{V}_r \equiv \mathbf{1}$.

Il quadrato del prodotto scalare tra i vettori ${}_s\ddot{\boldsymbol{x}}$ e ${}_K\ddot{\boldsymbol{x}}$ dà luogo a,

$$({}_s\ddot{\boldsymbol{x}}'{}_K\ddot{\boldsymbol{x}})^2 = \|{}_s\ddot{\boldsymbol{x}}\|^2\|{}_K\ddot{\boldsymbol{x}}\|^2\cos^2(\vartheta) \tag{19}$$

e quindi, portando a primo membro nella (19) i quadrati dei moduli dei vettori ${}_s\ddot{\boldsymbol{x}}$ e ${}_K\ddot{\boldsymbol{x}}$ si ha, dividendo numeratore e denominatore per N^2,

$$\frac{({}_s\ddot{\boldsymbol{x}}'\,{}_K\ddot{\boldsymbol{x}})^2}{\|{}_s\ddot{\boldsymbol{x}}\|^2\|{}_K\ddot{\boldsymbol{x}}\|^2} = \frac{\mathrm{Cov}^2({}_sX,{}_KX)}{\mathrm{Var}({}_sX)\mathrm{Var}({}_KX)} = \rho^2_{{}_sX{}_KX} = \cos^2(\vartheta), \tag{20}$$

ove, ϑ è l'angolo compreso tra ${}_s\ddot{\boldsymbol{x}}$ e ${}_K\ddot{\boldsymbol{x}}$. Si ricorda, a questo proposito, che i vettori sono definiti a meno di una congruenza (spostamento rigido che conserva la direzione).

Si considera ora la matrice $\boldsymbol{W} = (\mathbf{1}, {}_s\boldsymbol{x})$ e il modello lineare esteso

$${}_K\boldsymbol{x} = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

che naturalmente include il modello ridotto ${}_KX = \tau + \varepsilon$.

Applicando il criterio dei minimi quadrati si consegue la proiezione

$${}_K\hat{\boldsymbol{x}} = \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}\ {}_K\boldsymbol{x}$$

e il residuo ${}_K\hat{\ddot{\boldsymbol{x}}}$ il cui modulo $\|{}_K\hat{\ddot{\boldsymbol{x}}}\|$ è inferiore al modulo $\|{}_K\ddot{\boldsymbol{x}}\|$, in forza del teorema di Pitagora. Per il teorema sulla proiezione in due stadi, ove vale la (18), si ha che $\mathring{\boldsymbol{x}} = {}_K\dot{\boldsymbol{x}} - {}_K\hat{\boldsymbol{x}}$ è ortogonale al vettore $\mathbf{1}$ e coincide quindi con il vettore ${}_s\ddot{\boldsymbol{x}}$ "applicato" al piede della proiezione ${}_K\dot{\boldsymbol{x}}$ di ${}_K\boldsymbol{x}$ a meno di una costante moltiplicativa, $\mathring{\boldsymbol{x}} = \gamma\ {}_s\ddot{\boldsymbol{x}}$. L'angolo compreso tra il vettore

$\mathring{\boldsymbol{x}}$ e ${}_K\ddot{\boldsymbol{x}}$ è il medesimo, ϑ, per cui, essendo $\mathring{\boldsymbol{x}} \perp {}_K\hat{\ddot{\boldsymbol{x}}}$ si ha,

$$\cos^2(\vartheta) = \frac{\mathring{\boldsymbol{x}}' \; \mathring{\boldsymbol{x}}}{{}_K\ddot{\boldsymbol{x}}' \; {}_K\ddot{\boldsymbol{x}}} = \rho^2_{{}_sX\,{}_KX}. \tag{21}$$

Tale quantità coincide quindi con l'indice di miglioramento relativo nell'interpretazione delle modalità di ${}_KX$ passando da un modello costante

$${}_KX = \tau + e,$$

al modello che aggiunge una nuova variabile ${}_sX$,

$${}_KX = \alpha + \beta \; {}_sX + \varepsilon.$$

L'estensione al caso $\mathcal{V}_r, \; r > 1$ sarà introdotta nell'ambito della correlazione parziale.

1.4 Rapporto di correlazione multipla

Si ricorda che l'indice di determinazione multipla o rapporto di correlazione multipla assume la forma,

$$ {}_K\eta^2_{1,2,\cdots,K-1} = \frac{{}_K\delta_0^{*2} - {}_K\delta^{*2}_{1,2,\cdots,K-1}}{{}_K\delta_0^{*2}}, \tag{22} $$

ove implicitamente si fa riferimento a due modelli regressivi lineari, tipicamente,

$$ {}_KX = \tau + e $$

e

$$ {}_KX = \alpha_0 + \alpha_1\, {}_1X + \alpha_2\, {}_2X + \cdots \alpha_{K-1}\, {}_{K-1}X + \varepsilon, $$

in cui il primo è nidificato nel secondo.

Passando all'interpretazione geometrica, il primo modello assume l'aspetto

$$ {}_K\boldsymbol{x} = \mathbf{1}\tau + \boldsymbol{e} $$

cui corrispondono, secondo il criterio dei minimi quadrati, la stima $\hat{\tau} = (\mathbf{1}'\mathbf{1})^{-1}\ \mathbf{1}'{}_K\boldsymbol{x} = \frac{\mathbf{1}'{}_K\boldsymbol{x}}{N} = \mu_{{}_KX}$, la proiezione ${}_K\dot{\boldsymbol{x}} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'{}_K\boldsymbol{x} = \boldsymbol{U}_N\, {}_K\boldsymbol{x}$, ove $\boldsymbol{U}_N = (\mathbf{1}\,\mathbf{1}')\frac{1}{N}$ e il residuo ${}_K\ddot{\boldsymbol{x}} = {}_K\boldsymbol{x} - {}_K\dot{\boldsymbol{x}} = (\boldsymbol{I} - \boldsymbol{U}_N){}_K\boldsymbol{x}$.

Analogamente, per il caso completo si avrà la stima del modello lineare di tipo

$$_K\boldsymbol{x} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad , \qquad \boldsymbol{X} = (\mathbf{1}, {}_1\boldsymbol{x}, \cdots, {}_{K-1}\boldsymbol{x})$$

cui corrispondono, per $\boldsymbol{X}$ di rango pieno e secondo il criterio dei minimi quadrati, le stime $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'{}_K\boldsymbol{x}$, la proiezione ${}_K\tilde{\boldsymbol{x}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'{}_K\boldsymbol{x} = {}_X\boldsymbol{P}{}_K\boldsymbol{x}$ con ${}_X\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ e il residuo ${}_K\tilde{\tilde{\boldsymbol{x}}} = {}_K\boldsymbol{x} - {}_K\tilde{\boldsymbol{x}} = (\boldsymbol{I} - {}_X\boldsymbol{P}){}_K\boldsymbol{x}$. Il numeratore della (22) può essere espresso come differenza tra le norme al quadrato delle proiezioni, oppure come differenza delle norme al quadrato dei residui, per cui si ha, tenendo conto che ${}_X\boldsymbol{P}$ e $\boldsymbol{U}_N$ sono proiettori ortogonali,

$$\begin{aligned} {}_X\eta^2_{1,2,\cdots,K-1} &= \frac{{}_K\boldsymbol{x}'{}_X\boldsymbol{P}{}_K\boldsymbol{x} - {}_K\boldsymbol{x}'\boldsymbol{U}_N{}_K\boldsymbol{x}}{{}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N){}_K\boldsymbol{x}} \\ &= \frac{{}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N){}_K\boldsymbol{x} - {}_K\boldsymbol{x}'(\boldsymbol{I} - {}_X\boldsymbol{P}){}_K\boldsymbol{x}}{{}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N){}_K\boldsymbol{x}} = \frac{{}_K\boldsymbol{x}'({}_X\boldsymbol{P} - \boldsymbol{U}_N){}_K\boldsymbol{x}}{{}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N){}_K\boldsymbol{x}} \end{aligned} \tag{23}$$

Seguendo un altro punto di vista si può studiare il coefficiente di correlazione al quadrato tra ${}_KX$ e ${}_K\tilde{X}$. Si dimostra il teorema sorprendente che segue.

Teorema (Correlazione *predicted–observed*) Sotto le assunzioni che precedono, il rapporto di correlazione multipla coincide con il quadrato del coefficiente di correlazione di Bravais–Pearson tra i valori osservati della variabile dipendente ${}_K\boldsymbol{x}$ e i valori calcolati della medesima secondo i minimi quadrati, ${}_K\tilde{\boldsymbol{x}}$,

$$\rho^2_{{}_KX\,{}_K\tilde{X}} = {}_K\eta^2_{1,2,\cdots,K-1} \;. \tag{24}$$

Prova. Si considerino gli scostamenti dalla media di ${}_X\boldsymbol{x}$, ${}_K\ddot{\boldsymbol{x}} = {}_X\boldsymbol{x} - {}_X\dot{\boldsymbol{x}} = (\boldsymbol{I} - \boldsymbol{U}_N)_K\boldsymbol{x}$ e gli analoghi scostamenti di ${}_K\tilde{\boldsymbol{x}}$, ${}_K\ddot{\tilde{\boldsymbol{x}}} = (\boldsymbol{I} - \boldsymbol{U}_N)_K\tilde{\boldsymbol{x}}$. Si ricavano immediatamente le seguenti devianze,

$$N\,\mathrm{Var}({}_KX) \;=\; {}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N)_K\boldsymbol{x};$$

$$\begin{aligned} N\,\mathrm{Var}({}_K\tilde{X}) \;&=\; {}_K\tilde{\boldsymbol{x}}'(\boldsymbol{I} - \boldsymbol{U}_N)_K\tilde{\boldsymbol{x}} = {}_K\boldsymbol{x}'\,{}_X\boldsymbol{P}(\boldsymbol{I} - \boldsymbol{U}_N)_X\boldsymbol{P}_K\boldsymbol{x} \\ &=\; {}_K\boldsymbol{x}'\,{}_X\boldsymbol{P}_K\boldsymbol{x} - {}_K\boldsymbol{x}'\boldsymbol{U}_{N\,K}\boldsymbol{x} = {}_K\boldsymbol{x}'({}_X\boldsymbol{P} - \boldsymbol{U}_N)_K\boldsymbol{x} \end{aligned}$$

poiché $\boldsymbol{U}_N\,{}_X\boldsymbol{P} = \boldsymbol{U}_N$ essendo $\mathbf{1}$ una colonna di $\boldsymbol{X}$.
Per il calcolo della covarianza tra ${}_KX$ e ${}_K\tilde{X}$ si ha che $N\,\mathrm{Cov}({}_KX, {}_K\tilde{X}) = {}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N)_K\tilde{\boldsymbol{x}} = {}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N)_X\boldsymbol{P}_K\boldsymbol{x} = {}_K\boldsymbol{x}'\,{}_X\boldsymbol{P}_K\boldsymbol{x} - {}_K\boldsymbol{x}'\boldsymbol{U}_{N\,K}\boldsymbol{x} = {}_K\boldsymbol{x}'({}_X\boldsymbol{P} - \boldsymbol{U}_N)_K\boldsymbol{x}$.
Ora, con semplici sostituzioni, si ottiene,

$$\rho^2_{{}_KX\,{}_K\tilde{X}} = \frac{\mathrm{Cov}^2({}_KX, {}_K\tilde{X})}{\mathrm{Var}({}_KX)\mathrm{Var}({}_K\tilde{X})} = \frac{({}_K\boldsymbol{x}'({}_X\boldsymbol{P} - \boldsymbol{U}_N)_K\boldsymbol{x})^2}{{}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N)_K\boldsymbol{x}\,{}_K\boldsymbol{x}'({}_X\boldsymbol{P} - \boldsymbol{U}_N)_K\boldsymbol{x}} = \frac{{}_K\boldsymbol{x}'({}_X\boldsymbol{P} - \boldsymbol{U}_N)_K\boldsymbol{x}}{{}_K\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{U}_N)_K\boldsymbol{x}} = {}_K\eta^2_{1,2,\cdots,K-1} \;. \tag{25}$$

□

1.5 Trasformazioni e riparametrizzazioni

Si consideri il modello regressivo lineare

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{y}, \boldsymbol{\epsilon} \in R^N,\ \boldsymbol{\beta} \in R^K, \tag{26}$$

ove, $\boldsymbol{X} = (\mathbf{1}, {}_1\boldsymbol{x}, {}_2\boldsymbol{x}, \cdots, {}_{K-1}\boldsymbol{x})$ ha rango pieno, $r(\boldsymbol{X}) = K$.
Talvolta, in luogo delle variabili descritte nella matrice $\boldsymbol{X}_{N\times K}$ si preferisce utilizzare un nuovo insieme descritto dalla matrice

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A}, \tag{27}$$

ove $\boldsymbol{A}$ è una matrice quadrata di ordine K e di rango pieno, $r(\boldsymbol{A}) = K$ per cui $r(\boldsymbol{Z}) = r(\boldsymbol{X})$.
La trasformazione lineare (27) è abbastanza generale e include, come caso particolare, i cambiamenti di scala di ciascuna variabile ${}_\ell X,\ \ell = 1, 2, \cdots, K-1$. Un esempio elementare è il seguente

$$\boldsymbol{Z} = \begin{pmatrix} 1 & {}_1z_1 & {}_2z_1 \\ 1 & {}_1z_2 & {}_2z_2 \\ 1 & {}_1z_3 & {}_2z_3 \\ 1 & {}_1z_4 & {}_2z_4 \\ 1 & {}_1z_5 & {}_2z_5 \\ 1 & {}_1z_6 & {}_2z_6 \end{pmatrix} = \begin{pmatrix} 1 & {}_1x_1 & {}_2x_1 \\ 1 & {}_1x_2 & {}_2x_2 \\ 1 & {}_1x_3 & {}_2x_3 \\ 1 & {}_1x_4 & {}_2x_4 \\ 1 & {}_1x_5 & {}_2x_5 \\ 1 & {}_1x_6 & {}_2x_6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & a_1 & 0 \\ 0 & 0 & a_2 \end{pmatrix} = \boldsymbol{X}\boldsymbol{A}\ .$$

La trasformazione di variabili (27) ha alcune conseguenze con riferimento al modello regressivo lineare

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \qquad \boldsymbol{y}, \boldsymbol{\epsilon} \in R^N,\ \boldsymbol{\gamma} \in R^K, \tag{28}$$

ove, $\boldsymbol{Z} = (\mathbf{1}, {}_1\boldsymbol{z}, {}_2\boldsymbol{z}, \cdots, {}_{K-1}\boldsymbol{z})$.
Si studi dapprima il proiettore ${}_Z\boldsymbol{P}$,

$$\begin{aligned} {}_Z\boldsymbol{P} &= \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}' = \boldsymbol{X}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}' \\ &= \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{A}')^{-1}\boldsymbol{A}'\boldsymbol{X}' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = {}_X\boldsymbol{P}\ . \end{aligned} \tag{29}$$

Questa identità assicura la medesima proiezione $\dot{\boldsymbol{y}}$ ed il medesimo residuo $\ddot{\boldsymbol{y}}$ sotto entrambi i modelli (26) e (28).
Per quanto attiene alle stime dei parametri si ha naturalmente

$$\hat{\gamma} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} = (\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{A}')^{-1}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}^{-1}\hat{\boldsymbol{\beta}}\ . \tag{30}$$

Quest'ultima relazione non è del tutto inaspettata essendo $\boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\gamma}$ una riparametrizzazione[1].
Si noti, in conclusione, che la trasformazione (27) cambia solo la *base* della varietà lineare $\mathcal{V}_K$ che accoglie le "medie" dei modelli regressivi e non già la varietà medesima.

[1] Una riparametrizzazione è una funzione biunivoca da R^K in R^K.

1.6 Forme equivalenti delle equazioni normali

Come è noto, il sistema delle equazioni normali basato sulla matrice di varianze e covarianze, $\boldsymbol{\Sigma}_K$, e sul vettore delle medie, $\boldsymbol{\mu}$, è del tipo,

$$\begin{cases} \alpha_0^* = {}_K\mu - \sum_{\ell=1}^{K-1} \alpha_\ell^* \ {}_\ell\mu \\ \\ \sum_{\ell=1}^{K-1} \alpha_\ell^* \sigma_{\ell j} = \sigma_{jK}, \qquad\qquad j = 1, 2, \cdots, K-1. \end{cases} \tag{31}$$

se $\det \boldsymbol{\Sigma}_{K-1} \neq 0$.
Si consideri la divisione per $\sigma_{jj}, \ \ j = 1, 2, \cdots, K-1$ di ciascuna delle ultime $K-1$ equazioni. Si ottiene il seguente sistema equivalente alla (31),

$$\begin{cases} \alpha_0^* = {}_K\mu - \sum_{\ell=1}^{K-1} \alpha_\ell^* \ {}_\ell\mu \\ \\ \sum_{\ell=1}^{K-1} \alpha_\ell^* b_{\ell j} = b_{jK}, \qquad\qquad j = 1, 2, \cdots, K-1, \end{cases} \tag{32}$$

ove $b_{\ell j} = \frac{\sigma_{\ell j}}{\sigma_{jj}}$ è il coefficiente di regressione lineare, secondo i minimi quadrati, nel modello ${}_\ell X = a + b_{\ell j} \ {}_j X + \varepsilon$.
In modo del tutto analogo, posto

$$\rho_{\ell j} = \frac{\sigma_{\ell j}}{\sqrt{\sigma_{\ell\ell}}\sqrt{\sigma_{jj}}} = \frac{\sigma_{\ell j}}{\sigma_\ell \sigma_j}$$

si ottiene un altro sistema equivalente che dipende, in parte, dalla matrice di correlazione $(\rho_{\ell j})$,

$$\begin{cases} \alpha_0^* = {}_K\mu - \sum_{\ell=1}^{K-1} \alpha_\ell^* \ {}_\ell\mu \\ \\ \sum_{\ell=1}^{K-1} \alpha_\ell^* \rho_{\ell j} \sigma_\ell = \rho_{Kj}\sigma_K, \qquad\qquad j = 1, 2, \cdots, K-1. \end{cases} \tag{33}$$

"Peso" delle componenti esplicative.
Si ipotizzi un insieme di K variabili statistiche ${}_1X, {}_2X, \cdots, {}_KX$. Si consideri la variabile ${}_KX$ "dipendente", secondo un modello regressivo lineare, dalle altre variabili esplicative,

$$ {}_KX = \alpha_0 + \alpha_1\ {}_1X + \cdots + \alpha_{K-1}\ {}_{K-1}X + \varepsilon. \tag{34} $$

Nelle applicazioni pratiche è rilevante individuare quali variabili "pesano" maggiormente nella determinazione delle variazioni medie di ${}_KX$.

Il confronto diretto delle soluzioni $\alpha_j^*,\ j = 1, 2, \ldots, K-1$ è fuorviante poiché ciascuna componente α_j^* reca traccia dello scarto quadratico medio della variabile corrispondente. Un modo per risolvere il problema si basa sulla *standardizzazione* delle variabili esplicative.

Variabile standardizzata.
Sia Y una variabile statistica di media $M(Y) = \mu$ e varianza $\mathrm{Var}(Y) = \sigma^2$. Dicesi variabile standardizzata Z la variabile,

$$ Z = \frac{Y - \mu}{\sigma}. \tag{35} $$

È immediato controllare che la media e la varianza della variabile Z sono pari a,

$$ M(Z) = 0 \qquad \text{e} \qquad \mathrm{Var}(Z) = 1. \tag{36} $$

Naturalmente, due variabili standardizzate hanno in comune i primi due momenti e differiscono, in generale, per i successivi. Si può provare tuttavia che se Z e W sono variabili standardizzate delle corrispondenti X e, rispettivamente, Y, allora $\rho^2_{ZW} = \rho^2_{XY}$.

Regressione con variabili indipendenti standardizzate.
Si osservi che una variabile standardizzata è dotata, in particolare, di una varianza unitaria che non dipende piú dalla scala originaria del fenomeno rappresentato.

Siano Z_j le variabili esplicative standardizzate,

$$Z_j = \frac{{}_jX - {}_j\mu}{\sigma_j}, \qquad j = 1, 2, \cdots, K-1\ . \tag{37}$$

Il modello di regressione (34) può essere ripensato con riferimento alle nuove variabili,

$${}_KX = \beta_0 + \beta_1\ {}_1Z + \beta_2\ {}_2Z + \beta_3\ {}_3Z + \cdots + \beta_{K-1}\ {}_{K-1}Z + \varepsilon. \tag{38}$$

Le costanti β^*_j si ottengono con i minimi quadrati usuali e sono tra di loro confrontabili: rappresentano, in termini assoluti, il contributo alle variazioni medie di ${}_KX$ dovuto alla variabile ${}_jZ,\ j = 1, 2, \cdots, K-1$. Si osservi che $\beta^*_0 = {}_K\mu$.

Ci si chiede se si possano ottenere i coefficienti β^*_j utilizzando i coefficienti ordinari α^*_j. La risposta è affermativa. Con semplici sostituzioni è immediato studiare la relazione tra α^*_j e

$\beta_j^*,\ j = 1, 2, \cdots, K-1,$

$$_{K}X = \beta_0 + \beta_1 \frac{{}_{1}X - {}_{1}\mu}{\sigma_1} + \beta_2 \frac{{}_{2}X - {}_{2}\mu}{\sigma_2} + \cdots + \beta_{K-1} \frac{{}_{K-1}X - {}_{K-1}\mu}{\sigma_{K-1}} + \varepsilon,$$

ovvero

$$_{K}X = \beta_0 - \frac{\beta_1\ {}_{1}\mu}{\sigma_1} - \frac{\beta_2\ {}_{2}\mu}{\sigma_2} + \cdots + \frac{\beta_1}{\sigma_1}\ {}_{1}X + \cdots + \frac{\beta_{K-1}}{\sigma_{K-1}}\ {}_{K-1}X + \varepsilon$$

da cui discende immediatamente

$$\alpha_j^* = \frac{\beta_j^*}{\sigma_j}, \qquad j = 1, 2, \cdots, K-1, \tag{39}$$

ovvero, $\beta_j^* = \sigma_j \alpha_j^*$.

La trasformazione che porta il modello regressivo dalla versione ordinaria a quella con variabili esplicative standardizzate rientra, come caso particolare, in quelle trasformazioni lineari descritte dalla (27) ove,

$$\boldsymbol{A} = \begin{pmatrix} 1 & -\frac{\mu_1}{\sigma_1} & -\frac{\mu_2}{\sigma_2} & \cdots & -\frac{\mu_{K-1}}{\sigma_{K-1}} \\ 0 & \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_{K-1}} \end{pmatrix} \tag{40}$$

e

$$\boldsymbol{A}^{-1} = \begin{pmatrix} 1 & \mu_1 & \mu_2 & \cdots & \mu_{K-1} \\ 0 & \sigma_1 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{K-1} \end{pmatrix} . \tag{41}$$

Le conseguenze ovvie riguardano i vettori delle previsioni e i vettori residui che restano invarianti rispetto alla trasformazione

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A} .$$

I coefficienti $\hat{\boldsymbol{\beta}}$ standardizzati si ottengono dai coefficienti ordinari, associati alle variabili originarie, mediante la forma

$$\hat{\boldsymbol{\beta}} = \boldsymbol{A}^{-1}\hat{\boldsymbol{\alpha}} . \tag{42}$$

Si osservi che il sistema delle equazioni normali è caratterizzato, ad esempio, secondo la (33) da

$$\begin{cases} \alpha_0^* = {}_K\mu - \sum_{\ell=1}^{K-1} \alpha_\ell^* \ {}_\ell\mu \\ \\ \sum_{\ell=1}^{K-1} \alpha_\ell^* \rho_{\ell j} \sigma_\ell = \rho_{Kj}\sigma_K \qquad j = 1, 2, \cdots, K-1 \end{cases}$$

Nell'ipotesi di standardizzazione delle variabili esplicative, il sistema si contrae, in senso lato, a $K-1$ equazioni, precisamente,

$$\sum_{\ell=1}^{K-1} \beta_\ell^* \rho_{\ell j} = \rho_{Kj}\sigma_K \qquad j = 1, 2, \cdots, K-1 \; .$$

Quest'ultima versione chiarisce in quale modo intervengono le variabili originarie ${}_1X$, ${}_2X$, $\cdots$, ${}_{K-1}X$ nella spiegazione delle variazioni dei valori medi della variabile ${}_KX$: solamente attraverso le costanti adimensionali $\rho_{\ell j}$.

APPENDICE A

Per il prosieguo si richiamano qui alcune regole di derivazione vettoriale.

Derivazione di funzioni reali e vettoriali rispetto ad un vettore.
In letteratura esistono vari modi di disporre le derivate di una funzione reale o vettoriale, $f(\boldsymbol{\beta})$, $\boldsymbol{f}(\boldsymbol{\beta})$, rispetto ad un vettore *colonna* $\boldsymbol{\beta} \in R^K$. Qui si seguirà un approccio largamente diffuso.

Le idee di fondo sono le seguenti. Se si deriva una funzione reale $f = f(\boldsymbol{\beta})$ rispetto a $\boldsymbol{\beta}$, le derivate si dispongono secondo una colonna, $K \times 1$,

$$\frac{\partial f}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial f}{\partial \beta_1} \\ \frac{\partial f}{\partial \beta_2} \\ \vdots \\ \frac{\partial f}{\partial \beta_K} \end{pmatrix}, \qquad f \in R. \tag{43}$$

Se $\boldsymbol{g} = \boldsymbol{g}(\boldsymbol{\beta}) \in R^N$ è un vettore *colonna* funzione di $\boldsymbol{\beta} \in R^K$, allora le derivate di $\boldsymbol{g}$ rispetto a $\boldsymbol{\beta}$ si accolgono in una matrice $N \times K$ e si usa la notazione con $\partial \boldsymbol{\beta}'$ a denominatore,

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\beta}'} = \begin{pmatrix} \frac{\partial g_1}{\partial \beta_1} & \frac{\partial g_1}{\partial \beta_2} & \cdots & \frac{\partial g_1}{\partial \beta_K} \\ \frac{\partial g_2}{\partial \beta_1} & \frac{\partial g_2}{\partial \beta_2} & \cdots & \frac{\partial g_2}{\partial \beta_K} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_N}{\partial \beta_1} & \frac{\partial g_N}{\partial \beta_2} & \cdots & \frac{\partial g_N}{\partial \beta_K} \end{pmatrix}, \qquad \boldsymbol{g} \in R^N. \tag{44}$$

Se $\boldsymbol{h} = \boldsymbol{h}(\boldsymbol{\beta}) \in R^N$ è un vettore *riga* funzione di $\boldsymbol{\beta} \in R^K$, allora le derivate di $\boldsymbol{h}$ rispetto a $\boldsymbol{\beta}$ si accolgono in una matrice $K \times N$ e si usa la notazione con $\partial \boldsymbol{\beta}$ a denominatore,

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial h_1}{\partial \beta_1} & \frac{\partial h_2}{\partial \beta_1} & \cdots & \frac{\partial h_N}{\partial \beta_1} \\ \frac{\partial h_1}{\partial \beta_2} & \frac{\partial h_2}{\partial \beta_2} & \cdots & \frac{\partial h_N}{\partial \beta_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_1}{\partial \beta_K} & \frac{\partial h_2}{\partial \beta_K} & \cdots & \frac{\partial h_N}{\partial \beta_K} \end{pmatrix}, \qquad \boldsymbol{h} \in R^N. \tag{45}$$

La derivazione di un prodotto *riga per colonna* di due vettori, $\boldsymbol{hg}$, rispetto al vettore colonna $\boldsymbol{\beta} \in R^K$ deve porgere, per quanto si è assunto con la (43), un vettore colonna $K \times 1$.

Sfruttando, in modo iterato, la regola elementare della derivazione del prodotto di funzioni reali di variabile reale, si ha

$$\frac{\partial \boldsymbol{hg}}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\beta}} \boldsymbol{g} + \left(\boldsymbol{h} \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\beta}'} \right)' \; . \tag{46}$$

Si considerino ora alcuni esempi.

Sia $\boldsymbol{A}$ una matrice $N \times K$. La derivata del vettore colonna $\boldsymbol{A\beta}$ rispetto a $\boldsymbol{\beta} \in R^K$ si indicherà, per la (44), come segue,

$$\frac{\partial \boldsymbol{A\beta}}{\partial \boldsymbol{\beta}'} = \boldsymbol{A} \; . \tag{47}$$

Viceversa, se si considera il vettore riga $\boldsymbol{\beta}'\boldsymbol{A}'$, allora la sua derivata rispetto a $\boldsymbol{\beta} \in R^K$ si indica, per la (45), nel modo che segue,

$$\frac{\partial \boldsymbol{\beta}'\boldsymbol{A}'}{\partial \boldsymbol{\beta}} = \boldsymbol{A}' \; . \tag{48}$$

Si consideri, infine, la derivata rispetto a $\boldsymbol{\beta} \in R^K$ del prodotto $\boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{A\beta}$. Per la regola (46) si ha subito,

$$\begin{aligned} \frac{\partial \boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{A\beta}}{\partial \boldsymbol{\beta}} &= \frac{\partial \boldsymbol{\beta}'\boldsymbol{A}'}{\partial \boldsymbol{\beta}} \boldsymbol{A\beta} + \left(\boldsymbol{\beta}'\boldsymbol{A}' \frac{\partial \boldsymbol{A\beta}}{\partial \boldsymbol{\beta}'} \right)' \\ &= \boldsymbol{A}'\boldsymbol{A\beta} + (\boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{A})' = 2\boldsymbol{A}'\boldsymbol{A\beta} \; . \end{aligned} \tag{49}$$