



# Minimi quadrati ordinari Interpretazione geometrica

Eduardo Rossi



Il *modello di regressione lineare multipla* è usato per studiare le relazioni tra la variabile dipendente e diverse variabili indipendenti (esplicative).

$$y_t = \beta_1 x_{t1} + \dots + \beta_K x_{tK} + \epsilon_t \quad t = 1, 2, \dots, N \quad (1)$$

$\beta_1, \dots, \beta_K$  parametri fissi ma ignoti,  $\epsilon_t$  ignoto,  $y_t$  regredendo, v.casuale,  $x_{kt}$  regressore, covariata casuale. In genere, uno dei regressori è fissato uguale ad 1, per esempio il primo:  $x_{1t} = 1, \forall t$ ; con  $\beta_1$  intercetta (o costante) dell'equazione.



I caratteri variano simultaneamente tra gli individui. Il **metodo dei minimi quadrati ordinari** è un modo per scomporre le differenze nella variabile dipendente fra diverse caratteristiche osservate (variabili esplicative) per le diverse unità nel campione.

Il metodo dei minimi quadrati ordinari (in inglese **Ordinary Least Squares**, OLS) è usato per stimare il valore di  $\beta_k, k = 1, \dots, K$ . Questi sono scelti in modo tale che siano la soluzione al seguente problema:

$$\min_{\beta_1, \dots, \beta_K} \sum_{t=1}^N [y_t - (\beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_K x_{tK})]^2$$

Il termine “minimi quadrati” si riferisce alla minimizzazione della somma delle differenze al quadrato.  $[y_t - (\beta_1 x_{t1} + \dots + \beta_K x_{tK})]$ , gli *scarti*.



## LA SOMMA DEI QUADRATI

---

La funzione obiettivo

$$f(\beta_1, \dots, \beta_K) = \sum_{t=1}^N [y_t - (\beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_K x_{tK})]^2 \quad (2)$$

è la *sum of squared residuals* (somma dei quadrati dei residui).

Quando i residui sono valutati in  $\hat{\beta}_1, \dots, \hat{\beta}_K$  i residui sono detti *fitted residuals* (residui fittati, o residui della regressione).

Consideriamo il caso in cui l'unica variabile esplicativa è la costante:  $K = 1$  e  $x_{1t} = 1, \forall t$ . OLS trova il valore di  $\beta_1$  che è il più vicino a  $y_t$  nel senso della somma dei quadrati dei residui. OLS è la minimizzazione di una funzione quadratica in  $\beta_1$  e il risultato è la media:

$$\hat{\beta}_1 = \arg \min \sum_{t=1}^N (y_t - \beta_1)^2 = \frac{\sum_{t=1}^N y_t}{N}$$



## NOTAZIONE

---

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_K]' \quad (K \times 1) \quad \mathbf{x}_t = \begin{bmatrix} x_{t1} \\ \vdots \\ x_{tK} \end{bmatrix} \quad (K \times 1) \quad (3)$$

### Notazione matriciale

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad (N \times 1)$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{bmatrix} \quad (N \times K)$$



$$\begin{bmatrix} \mathbf{x}'_1 \boldsymbol{\beta} \\ \vdots \\ \mathbf{x}'_N \boldsymbol{\beta} \end{bmatrix} = \mathbf{X} \boldsymbol{\beta}$$

Il vettore  $\mathbf{y}$  raccoglie tutte le osservazioni della variabile dipendente. La matrice  $\mathbf{X}$  raccoglie le osservazioni sulle variabili esplicative. Ogni colonna di  $\mathbf{X}$  contiene tutte le osservazioni per la singola variabile esplicativa.



## LO STIMATORE DEI MINIMI QUADRATI (OLS)

---

Stimatore = E' una regola per calcolare una stima (un numero) dai dati campionari.

Il metodo dei minimi quadrati risolve il problema

$$\hat{\beta} \equiv \arg \min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

Definiamo

$$S(\beta) \equiv (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$



## LO STIMATORE DEI MINIMI QUADRATI (OLS)

---

$$\begin{aligned}\frac{\partial S(\beta)}{\partial \beta} &= \frac{\partial (y'y - 2\beta'X'y + \beta'X'X\beta)}{\partial \beta} \\ &= \frac{\partial (-2\beta'X'y + \beta'X'X\beta)}{\partial \beta} \\ &= -2\frac{\partial \beta'}{\partial \beta}X'y + \frac{\partial (\beta'X'X\beta)}{\partial \beta} \\ &= -2X'y + 2X'X\beta\end{aligned}$$



## LO STIMATORE DEI MINIMI QUADRATI (OLS)

---

$$\frac{\partial S(\hat{\beta})}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0 \quad (4)$$

Le *equazioni normali*

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\beta} = 0 \quad (5)$$

Lo stimatore OLS è

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (6)$$

Poichè la funzione stimata è lineare nei coefficienti, gli OLS ci danno dei coefficienti stimati che sono somme ponderate delle  $\{y_t\}$ . Le stime OLS sono funzioni lineari della variabile dipendente. Questa linearità in  $\{y_t\}$  semplifica l'analisi statistica degli OLS.



## L'INTERPRETAZIONE GEOMETRICA DEGLI OLS

---

Lo spazio delle colonne di  $\mathbf{X}$ ,  $Col(\mathbf{X})$ , è il sottospazio lineare di  $\mathbb{R}^N$  coperto dalle combinazioni lineari dei vettori colonna di  $\mathbf{X}$ :

$$Col(\mathbf{X}) \equiv \{\mathbf{z} \in \mathbb{R}^N \mid \mathbf{z} = \mathbf{X}\alpha, \alpha \in \mathbb{R}^k\}$$

La procedura di stima OLS trova il vettore in  $Col(\mathbf{X})$ ,  $\hat{\boldsymbol{\mu}}$ , che è più vicino a  $\mathbf{y}$ .

$\hat{\boldsymbol{\mu}}$  è detta *proiezione* di  $\mathbf{y}$  sul  $Col(\mathbf{X})$ .



Il metodo OLS risolve:

$$\hat{\beta} \equiv \arg \min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \quad (7)$$

La somma delle deviazioni al quadrato tra gli elementi di  $\mathbf{y}$  e  $\mathbf{X}\beta$  è il quadrato della distanza Euclidea fra  $\mathbf{y}$  e  $\mathbf{X}\beta$ :

$$(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) = \sum_{t=1}^N (y_t - \mathbf{x}'_t\beta)^2 = \|\mathbf{y} - \mathbf{X}\beta\|^2$$



Procedura in due passi:

1. Trovare il punto in un sottospazio che è il più vicino ad un punto che non si trova in quel sottospazio. Il sottospazio è l'insieme dei possibili vettori reali  $N$  dimensionali  $\mathbf{X}\boldsymbol{\beta}$  che può essere creato cambiando  $\boldsymbol{\beta}$  e questo sottospazio è lo spazio delle colonne di  $\mathbf{X}$ .

$$\hat{\boldsymbol{\mu}} \equiv \arg \min_{\boldsymbol{\mu} \in \text{Col}(\mathbf{X})} \|\mathbf{y} - \boldsymbol{\mu}\|^2$$

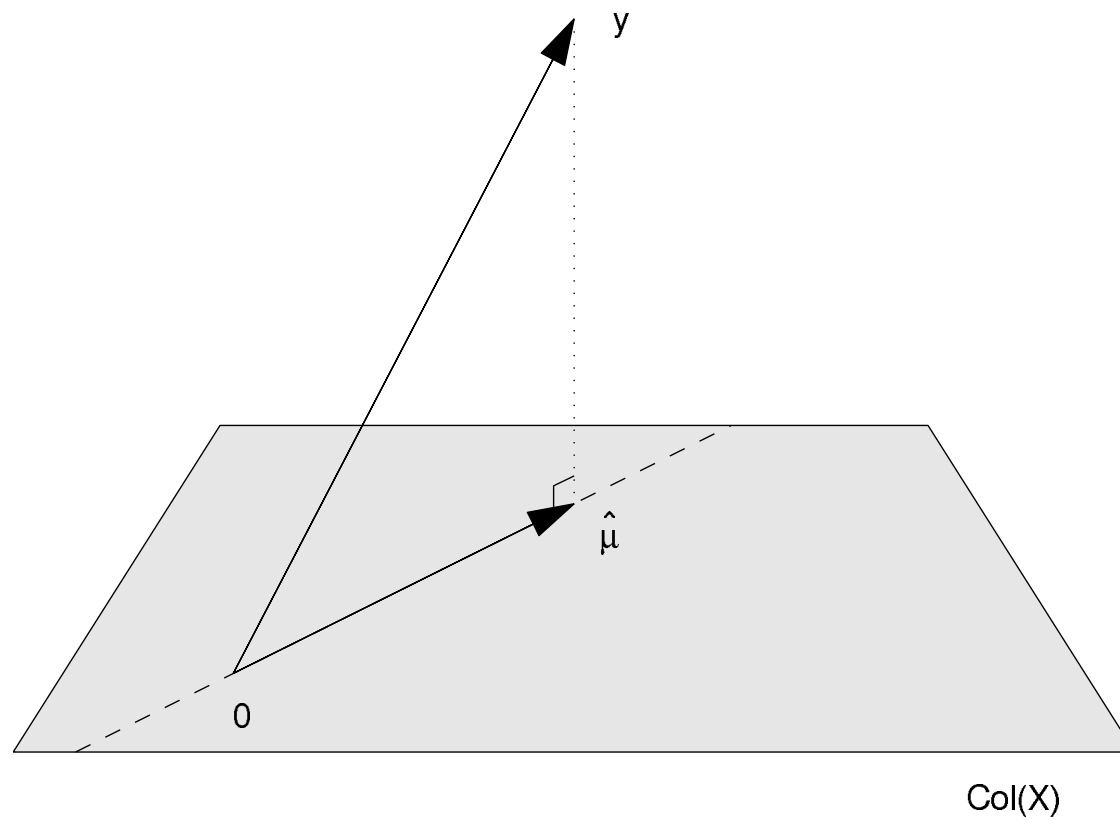
2. Trovare un  $\hat{\boldsymbol{\beta}}$  che sia soluzione a:

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$



# L'INTERPRETAZIONE GEOMETRICA DEGLI OLS

---





La soluzione al primo passo è unica mentre ci possono essere molte soluzioni al secondo problema. Sia  $\hat{\beta}$  una soluzione di (7) e sia  $\hat{\mu} = \mathbf{X}\hat{\beta}$ .

1. Il vettore dei valori fittati  $\hat{\mu}$  è l'unica proiezione ortogonale di  $\mathbf{y}$  su  $\text{Col}(\mathbf{X})$ .
2. Il vettore dei residui fittati  $\mathbf{y} - \hat{\mu}$  è ortogonale a  $\text{Col}(\mathbf{X})$ .
3. Se  $\dim[\text{Col}(\mathbf{X})] = K$ , allora (7) ha una soluzione unica:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mu}'$$



Tre idee base:

1. La regressione OLS significa **minimizzare la distanza al quadrato** tra il vettore osservato  $\mathbf{y}$  e un vettore di regressione  $\mathbf{X}\boldsymbol{\beta}$  che appartiene a  $\text{Col}(\mathbf{X})$ .
2. Il vettore dei valori fittati  $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$  è la **proiezione ortogonale** su  $\text{Col}(\mathbf{X})$ . Il vettore dei residui  $(\mathbf{y} - \hat{\boldsymbol{\mu}})$  è **perpendicolare** a  $\hat{\boldsymbol{\mu}}$  e ad ogni altro vettore in  $\text{Col}(\mathbf{X})$ .
3. Se  $\dim[\text{Col}(\mathbf{X})] = K$  allora  $\hat{\boldsymbol{\beta}}$  è **unico**.



Due osservazioni ed una sola variabile esplicativa ( $N = 1, K = 1$ )

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \equiv \boldsymbol{\iota}$$

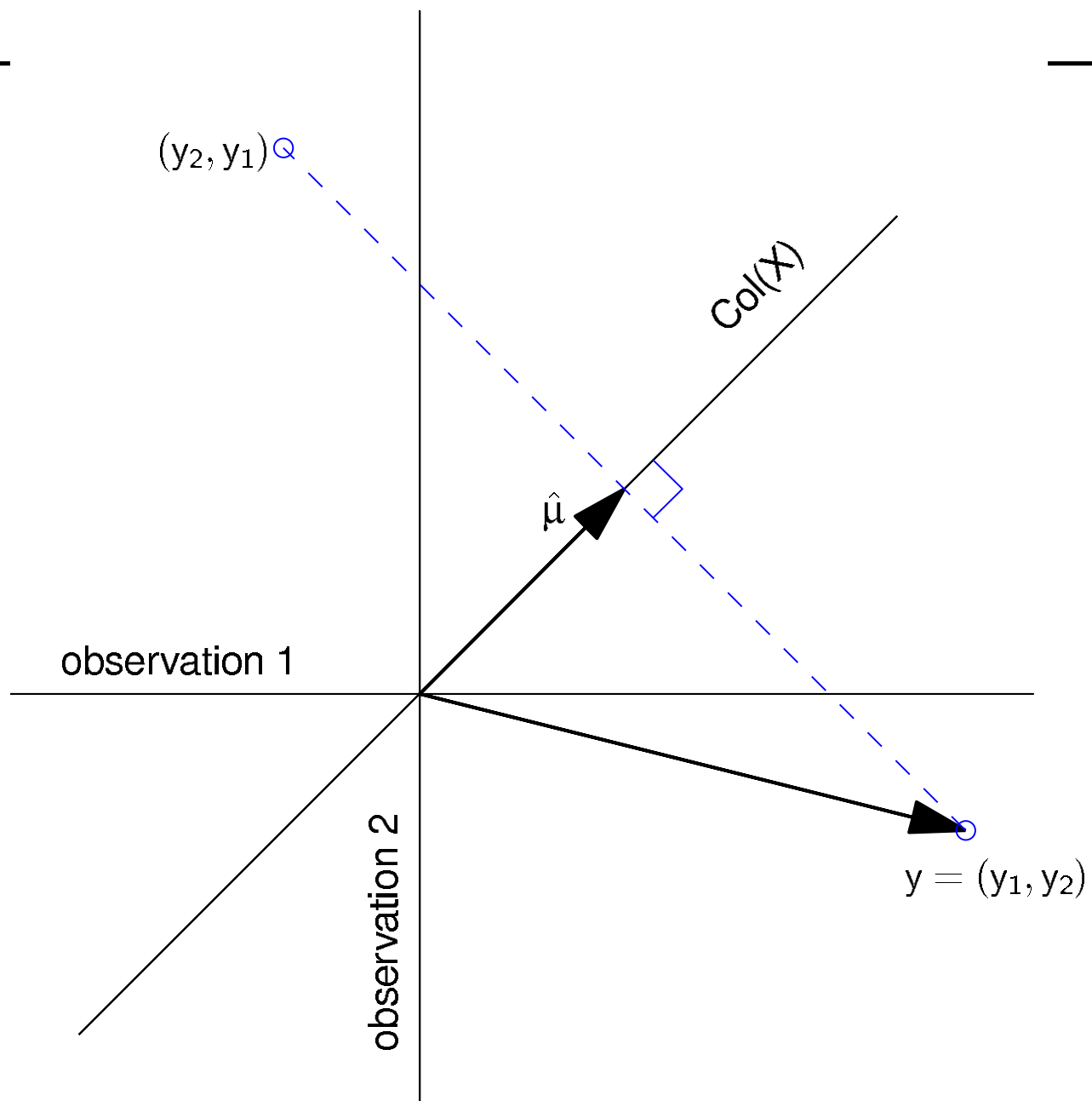
$$\text{Col}(\mathbf{X}) = \{\mathbf{z} \in \mathbb{R}^2 \mid z_1 = z_2\}, \text{ e } \hat{\boldsymbol{\beta}} = \bar{y}$$

$$\hat{\boldsymbol{\beta}} = \bar{y} = \arg \min_{\beta} [(y_1 - \beta)^2 + (y_2 - \beta)^2]$$

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\iota}y$$



# ESEMPIO





## ESEMPIO

---

Tre osservazioni e due variabili esplicative ( $N = 3, K = 2$ )

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ 1 & x_{32} \end{bmatrix} \equiv \boldsymbol{\iota}$$

$\text{Col}(\mathbf{X})$  è un piano che contiene tre punti:

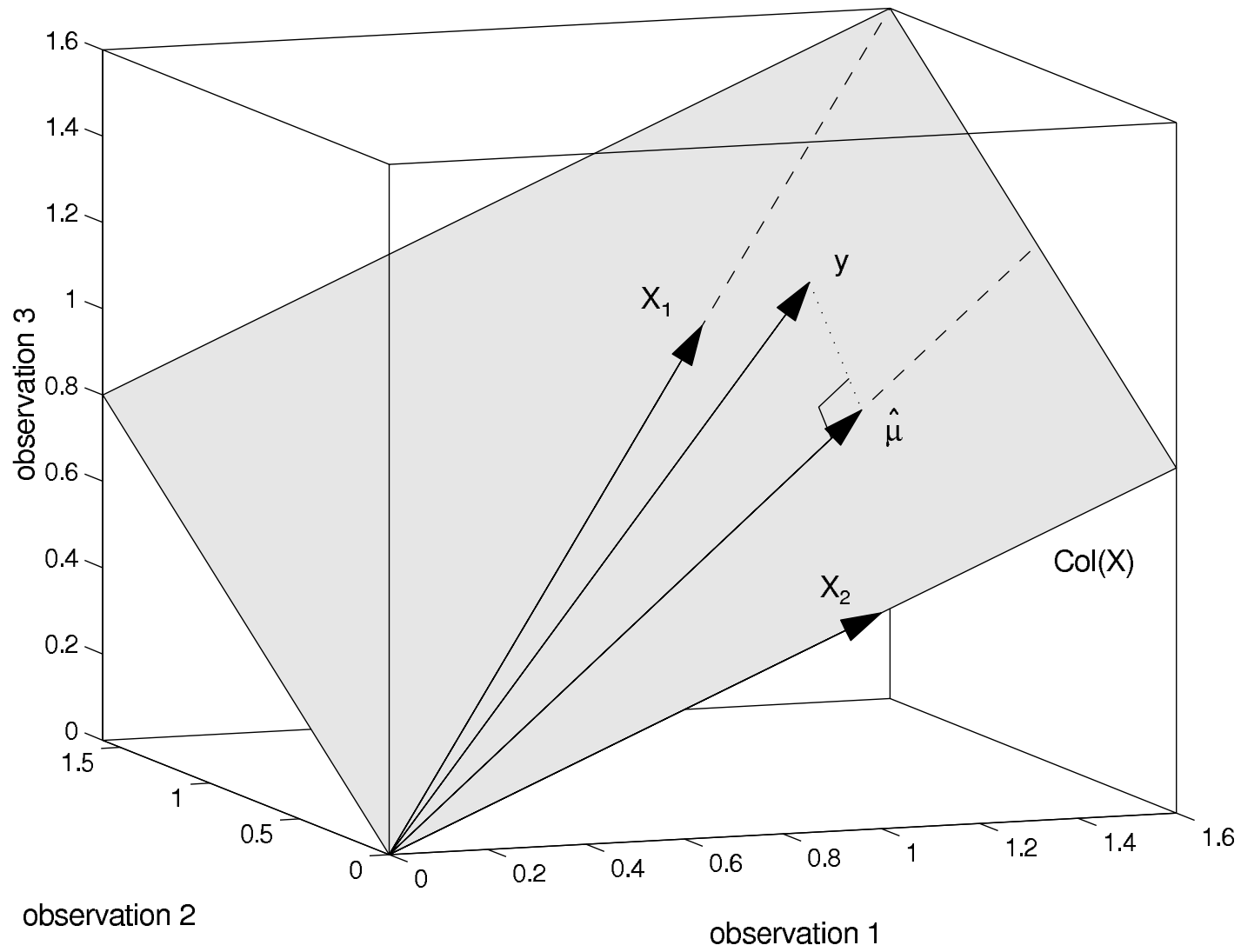
$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix}$$

Il vettore dei coefficienti stimati  $\hat{\boldsymbol{\beta}}$  è l'unica combinazione lineare di  $\boldsymbol{\iota}$  e  $\mathbf{X}_2$  che eguaglia  $\hat{\boldsymbol{\mu}}$ .

$$\hat{\boldsymbol{\mu}}_1 = \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 \quad \hat{\boldsymbol{\mu}}_2 = \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 \quad \hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2$$



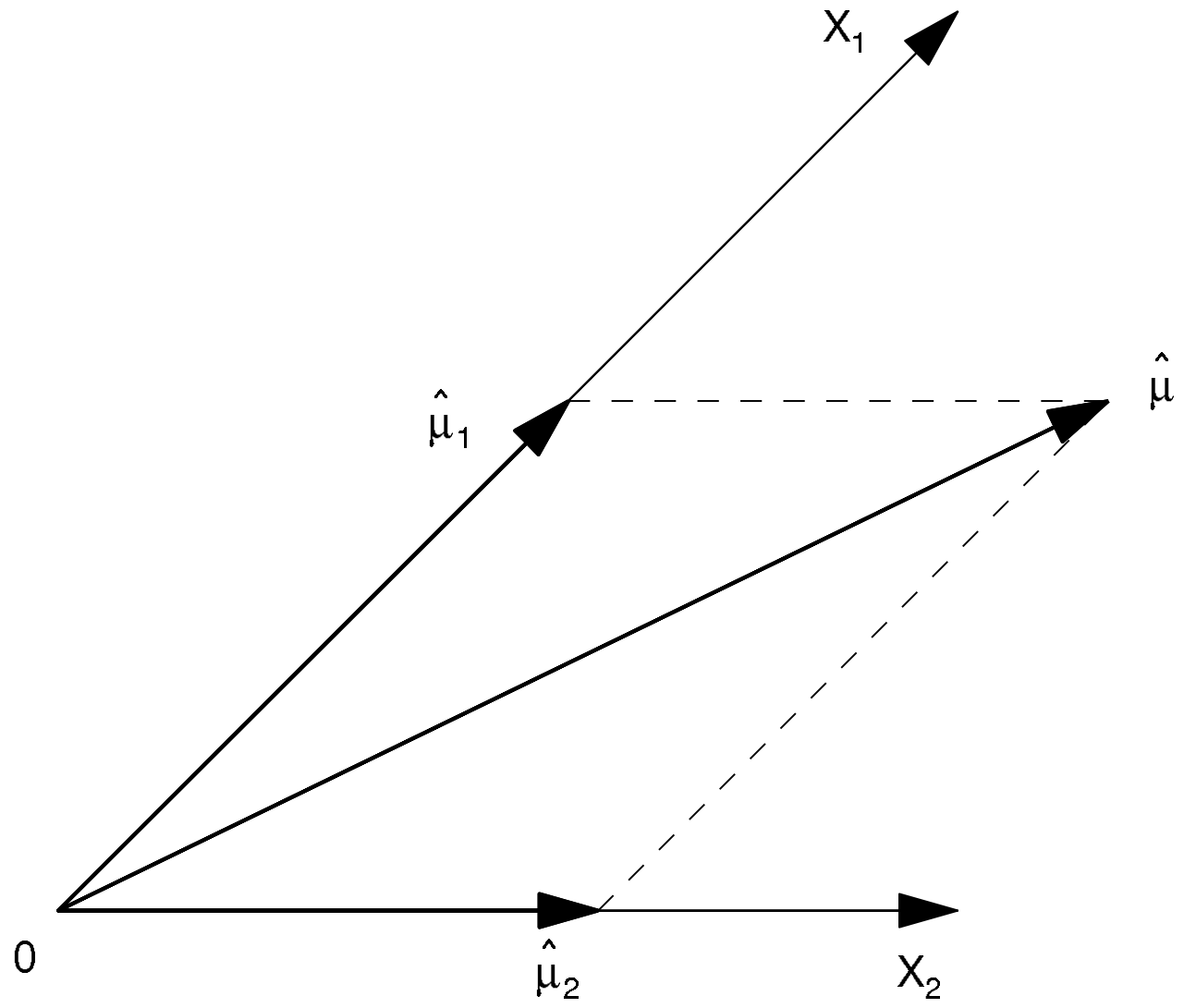
# ESEMPIO





# ESEMPIO

---





## $\hat{\mu}$ COME PROIEZIONE ORTOGONALE

---

La dipendenza lineare fra le variabile esplicative **non ha un ruolo fondamentale** su quanto bene una regressione lineare spiega  $\mathbf{y}$ . La distanza dipende solo da  $\hat{\mu}$ .

Mostriamo che

$$\hat{\mu} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

quando le colonne di  $\mathbf{X}$  sono linearmente indipendenti. Per due vettori  $\mu$  e  $\hat{\mu}$ :

$$\begin{aligned}\|\mathbf{y} - \mu\|^2 &= \|\mathbf{y} - \hat{\mu} + \hat{\mu} - \mu\|^2 \\ &= \|\mathbf{y} - \hat{\mu}\|^2 + \|\hat{\mu} - \mu\|^2 + 2(\mathbf{y} - \hat{\mu})'(\hat{\mu} - \mu)\end{aligned}$$

ma

$$(\mathbf{y} - \hat{\mu}) \perp (\hat{\mu} - \mu) \Leftrightarrow (\mathbf{y} - \hat{\mu})'(\hat{\mu} - \mu) = 0$$



## $\hat{\boldsymbol{\mu}}$ COME PROIEZIONE ORTOGONALE

---

**Teorema di Pitagora:** Se  $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^N$  e  $\mathbf{z}_1 \perp \mathbf{z}_2$  allora

$$\|\mathbf{z}_1 + \mathbf{z}_2\|^2 = \|\mathbf{z}_1\|^2 + \|\mathbf{z}_2\|^2.$$

in questo caso:

$$\|\mathbf{y} - \boldsymbol{\mu}\|^2 = \|\mathbf{y} - \hat{\boldsymbol{\mu}}\|^2 + \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2.$$

Se c'è un  $\hat{\boldsymbol{\mu}} \in Col(\mathbf{X})$  tale che

$$\mathbf{X}'(\mathbf{y} - \hat{\boldsymbol{\mu}}) = \mathbf{0}$$

allora per tutti gli altri  $\boldsymbol{\mu} \in Col(\mathbf{x})$

$$\boldsymbol{\mu}'(\mathbf{y} - \hat{\boldsymbol{\mu}}) = 0$$

$$(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'(\mathbf{y} - \hat{\boldsymbol{\mu}}) = 0$$

$$\begin{aligned} \|\mathbf{y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{y} - \hat{\boldsymbol{\mu}}\|^2 + \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 \\ &\geq \|\mathbf{y} - \hat{\boldsymbol{\mu}}\|^2 \end{aligned}$$



## $\hat{\mu}$ COME PROIEZIONE ORTOGONALE

---

Poichè  $\mathbf{y} - \hat{\mu}$  è ortogonale a  $Col(\mathbf{X})$ ,  $\hat{\mu}$  è vicino a  $\mathbf{y}$  almeno quanto un qualunque  $\mu$  in  $Col(\mathbf{X})$

Quindi  $\hat{\mu}$  è una soluzione al problema della distanza minima dei OLS

$$\hat{\mu} = \arg \min_{\mu \in Col(\mathbf{X})} \|\mathbf{y} - \mu\|^2$$

Ma  $\hat{\mu}$  è la **soluzione unica!**

La soluzione è unica perchè per ogni altra possibile soluzione  $\tilde{\mu}$  deve essere che

$$\|\mathbf{y} - \tilde{\mu}\|^2 = \|\mathbf{y} - \hat{\mu}\|^2$$

poichè nessun altro  $\mu$  è più vicino a  $\hat{\mu}$ .



## $\hat{\boldsymbol{\mu}}$ COME PROIEZIONE ORTOGONALE

---

Infatti, il teorema di Pitagora implica che

$$\|\mathbf{y} - \tilde{\boldsymbol{\mu}}\|^2 = \|(\mathbf{y} - \hat{\boldsymbol{\mu}}) + (\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}})\|^2 = \|\mathbf{y} - \hat{\boldsymbol{\mu}}\|^2 + \|\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}\|^2.$$

perchè  $(\mathbf{y} - \hat{\boldsymbol{\mu}}) \perp (\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}})$  quindi

$$\|\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}\|^2 = 0 \Rightarrow \hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}$$

La condizione di ortogonalità caratterizza completamente il vettore OLS dei valori fittati  $\hat{\boldsymbol{\mu}}$ .

Costruiamo  $\hat{\boldsymbol{\mu}}$  per una caso particolare e mostriamo che una soluzione unica esiste.

Le *equazioni normali* stabiliscono che

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}$$



## $\hat{\mu}$ COME PROIEZIONE ORTOGONALE

---

risolvendo per

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}$$

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}'\mathbf{y} = \mathbf{0}$$

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

dato che  $\mathbf{X}'\mathbf{X}$  è nonsingolare.



## $\hat{\boldsymbol{\mu}}$ COME PROIEZIONE ORTOGONALE

---

La soluzione per  $\hat{\boldsymbol{\mu}}$  segue

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$\hat{\boldsymbol{\beta}}$  e  $\hat{\boldsymbol{\mu}}$  hanno una relazione 1-a-1. Possiamo anche ottenere  $\hat{\boldsymbol{\beta}}$  da  $\hat{\boldsymbol{\mu}}$ : premoltiplicando per  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\mu}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$$



## **Teorema** *Proiezione*

Sia  $y \in \mathbb{R}^N$  e  $S \subseteq \mathbb{R}^N$  un sottospazio lineare. Allora  $\hat{\mu} \in S$  è una soluzione al problema

$$\min_{\mu \in S} \|y - \mu\|^2$$

**se e solo se**  $(y - \hat{\mu}) \perp S$ . Inoltre,  $\hat{\mu}$  esiste ed è unico.



Il teorema identifica il meccanismo di minimizzazione che significa trovare un  $\hat{\boldsymbol{\mu}} \in \text{Col}(\mathbf{X})$  tale che

$$\mathbf{y} - \hat{\boldsymbol{\mu}} \perp \text{Col}(\mathbf{X})$$

Secondo, il teorema chiarisce che  $\text{Col}(\mathbf{X})$  determina l'ottimale  $\hat{\boldsymbol{\mu}}$ .



## PROIETTORI ORTOGONALI

---

Per ogni  $\mathbf{y}$ , c'è un'unica  $\hat{\boldsymbol{\mu}}$ ,

$$\hat{\boldsymbol{\mu}} = \arg \min_{\boldsymbol{\mu} \in S} \|\mathbf{y} - \boldsymbol{\mu}\|^2$$

chiamata *proiezione* di  $\mathbf{y}$ . La proiezione ortogonale di  $\mathbf{y}$  è sempre una trasformazione lineare di  $\mathbf{y}$ :

$$\hat{\boldsymbol{\mu}} = \mathbf{P}\mathbf{y}$$

$\mathbf{P}$  proiettore ortogonale.

Nel caso generale che  $S = \text{Col}(\mathbf{X})$  e  $\mathbf{X}$  sia di rango-colonna pieno, la matrice

$$\mathbf{P}_X \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\hat{\boldsymbol{\mu}} = \mathbf{P}_X\mathbf{y}$$

è la trasformazione lineare di  $\mathbf{y}$  su  $\text{Col}(\mathbf{X})$  che produce  $\hat{\boldsymbol{\mu}}$ .



$\mathbf{P}_X$  ha due proprietà:

- non modifica i vettori in  $Col(\mathbf{X})$

$$\mathbf{z} \in Col(\mathbf{X}) \Rightarrow \mathbf{P}_X \mathbf{z} = \mathbf{z}$$

- trasforma i vettori ortogonali a  $Col(\mathbf{X})$  nel vettore zero.

$$\mathbf{z} \perp Col(\mathbf{X}) \Rightarrow \mathbf{P}_X \mathbf{z} = \mathbf{0}$$



## Prova

$\forall \mathbf{z} \in Col(\mathbf{X})$  esiste un  $\alpha : \mathbf{z} = \mathbf{X}\alpha$

$$\mathbf{P}_X \mathbf{z} = \mathbf{P}_X \mathbf{X}\alpha = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\alpha = \mathbf{X}\alpha = \mathbf{z}$$

Se  $\mathbf{z} \perp Col(\mathbf{X}) : \mathbf{z}'\mathbf{X} = \mathbf{0}, \forall \mathbf{X} \in Col(\mathbf{X})$  cosicchè  $\mathbf{X}'\mathbf{z} = \mathbf{0}$  e

$$\mathbf{P}_X \mathbf{z} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z} = \mathbf{0}$$

□



$\forall \mathbf{z} \in \mathbb{R}^N$ , possiamo scomporre  $\mathbf{z}$  *univocamente* nel vettore somma  $\mathbf{z}_1 + \mathbf{z}_2$  dove  $\mathbf{z}_1 \in Col(\mathbf{X})$  e  $\mathbf{z}_2 \in Col^\perp(\mathbf{X}) \equiv \{\mathbf{z} \in \mathbb{R}^N | \mathbf{X}'\mathbf{z} = 0\}$ .  
Dove  $Col^\perp(\mathbf{X})$  è il *complemento ortogonale*.

### Complemento ortogonale

Il sottospazio lineare di vettori  $S^\perp$ , ortogonale al sottospazio  $S \subseteq \mathbb{V}$ :

$$S^\perp = \{\mathbf{v} \in \mathbb{V} | \mathbf{u}'\mathbf{v} = 0, \forall \mathbf{u} \in S\}$$

è chiamato *complemento ortogonale* di  $S$ . E' equivalente a scrivere  $\mathbf{v} \in S^\perp$  come  $\mathbf{v} \perp S$ . Notiamo che se  $\mathbf{v} \in S \cap S^\perp$  allora  $\mathbf{v}'\mathbf{v} = 0$  tale che  $\mathbf{v}$  deve essere il vettore zero. In altre parole  $S \cap S^\perp = \{\mathbf{0}\}$



## PROIEZIONE ORTOGONALE

---

Sia  $\mathbb{S} \subseteq \mathbb{R}^N$  (sottospazio lineare) tale che per ogni  $\mathbf{z} \in \mathbb{R}^N$  c'è un unico  $\mathbf{z}_1 \in \mathbb{S}$  ed un unico  $\mathbf{z}_2 \in \mathbb{S}^\perp$  tale che  $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$ .

Allora la funzione da  $\mathbb{R}^N$  a  $\mathbb{S}^\perp$  che associa ogni  $\mathbf{z}$  con il suo corrispondente  $\mathbf{z}_1$  è una *proiezione ortogonale*.

Quando  $\mathbb{S} = \text{Col}(\mathbf{X})$  allora  $\mathbf{P}_X \mathbf{z} = \mathbf{z}_1$  è la proiezione ortogonale di  $\mathbf{z}$  su  $\text{Col}(\mathbf{X})$ . Solo la componente di  $\mathbf{z}$  in  $\text{Col}(X)$  sopravvive alla premoltiplicazione per  $\mathbf{P}_X$ .

La proiezione ortogonale da  $\mathbb{R}^N$  su un sottospazio  $\mathbb{S}$  è una **trasformazione lineare**. (La proiezione ortogonale di una combinazione lineare di vettori uguaglia la combinazione lineare delle proiezioni ortogonali dei singoli vettori).



## PROIETTORE ORTOGONALE

---

Ogni proiezione ortogonale da  $\mathbb{R}^N$  in un sottospazio  $\mathbb{S}$  può essere rappresentata da una matrice  $\mathbf{P}$ , chiamata **proiettore ortogonale**.

Sia  $\mathbb{S} \subseteq \mathbb{R}^N$ ,  $\forall \mathbf{z} \in \mathbb{R}^N$  c'è un unico  $\mathbf{z}_1 \in \mathbb{S}$  ed un unico  $\mathbf{z}_2 \in \mathbb{S}^\perp$  tale che  $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$ . Allora una matrice  $(N \times N)$   $\mathbf{P}$  tale che  $\mathbf{P}\mathbf{z} = \mathbf{z}_1$  è un *proiettore ortogonale* su  $\mathbb{S}$ .

Un proiettore ortogonale preserva la componente di un vettore in un sottospazio  $\mathbb{S}$  e annulla la componente nel sottospazio complementare ortogonale  $\mathbb{S}^\perp$ .

Se  $\mathbf{P}$  è un proiettore ortogonale su un sottospazio di  $\mathbb{R}^N$ , allora  $\mathbf{P}$  è unica.



## PROPRIETÀ DEI PROIETTORI ORTOGONALI

---

### 1. Simmetria

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}' = [\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}']' = \mathbf{P}'_X$$

### 2. Idempotenza

$$\mathbf{P}_X\mathbf{P}_X = [\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'][\mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'] = \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}' = \mathbf{P}_X$$

### 3. Semidefinitezza positiva

Per ogni  $\mathbf{w} \in \mathbb{R}^N$

$$\mathbf{w}'\mathbf{P}_X\mathbf{w} = \mathbf{w}'\mathbf{P}_X\mathbf{P}_X\mathbf{w} = \mathbf{w}'\mathbf{P}'_X\mathbf{P}_X\mathbf{w} = (\mathbf{P}_X\mathbf{w})'(\mathbf{P}_X\mathbf{w}) = \|\mathbf{P}_X\mathbf{w}\|^2 \geq 0$$



## PROPRIETÀ DEI PROIETTORI ORTOGONALI

---

Osserviamo che

$$\mathbf{z} \in Col^\perp(\mathbf{X}) \Rightarrow (\mathbf{I} - \mathbf{P}_X)\mathbf{z} = \mathbf{z}$$

$$\mathbf{z} \in Col(\mathbf{X}) \Rightarrow (\mathbf{I} - \mathbf{P}_X)\mathbf{z} = \mathbf{0}$$

cioè  $\mathbf{M}_X = (\mathbf{I} - \mathbf{P}_X)$  è un proiettore ortogonale su  $Col^\perp(\mathbf{X})$ , il complemento ortogonale di  $Col(\mathbf{X})$ .



Se esiste un vettore  $\alpha \in \mathbb{R}^K$  tale che  $\mathbf{X}\alpha = 0$  allora le colonne di  $\mathbf{X}$  sono linearmente indipendenti. Questa situazione è detta *multicollinearità esatta*.

Un unico  $\hat{\mu}$  esiste anche quando  $\mathbf{X}$  è di rango ridotto. Quando  $\mathbf{X}$  e  $(\mathbf{X}'\mathbf{X})$  sono singolari non possiamo usare  $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  per trovare  $\mathbf{P}_X$ .

Quando  $\dim[\text{Col}(\mathbf{X})] < K$ , possiamo trovare  $\mathbf{P}_X$  applicando la formula ad ogni sottoinsieme linearmente indipendente delle colonne di  $\mathbf{X}$  cioè una base per  $\text{Col}(\mathbf{X})$ .



Indichiamo con  $\mathbf{P}_X$  il proiettore ortogonale su  $Col(\mathbf{X})$  e sia  $\mathbf{X}_1$  una matrice composta da un sottoinsieme linearmente di colonne di  $\mathbf{X}$  tale che

$$Col(\mathbf{X}_1) = Col(\mathbf{X})$$

allora

$$\mathbf{P}_X = \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$$