IL MODELLO DI REGRESSIONE MULTIPLA

SOMMARIO: 1. Introduzione. - 2. Specificazione di un modello di regressione multipla. - 3. Stima con il metodo dei minimi quadrati ordinari (OLS). - 4. Proprietà degli stimatori dei minimi quadrati. - 5. Indice di determinazione multipla. - 6. Test e intervalli di confidenza per il modello di regressione multipla. - 7. Previsione nel modello di regressione mutipla. - 8. Variabili dummy. - Questionario.

1. INTRODUZIONE

Nella realtà è opportuno analizzare modelli in cui le variabili indipendenti, che *spiegano* una variabile dipendente, sono più di una. Quando si analizza un fenomeno, esso presenta una serie di caratteristiche che obbligano a sottolineare certe variabili influenti sullo stesso e a trascurarne altre; la scelta dell'una o dell'altra variabile esplicativa da inserire nel modello è rimessa alla competenza dell'elaboratore.

Supposta giusta la scelta delle variabili inserite nel modello, in questo capitolo tratteremo, in generale, un modello in cui le variabili esplicative sono più di una. Dopo aver trattato la fase di specificazione del modello, ci occuperemo della procedura di stima dei parametri dello stesso e delle proprietà statistiche degli stimatori. Tratteremo, infine, i problemi di inferenza relativi al modello, quali la costruzione dei test d'ipotesi e degli intervalli di confidenza per i coefficienti.

Per illustrare il modello di regressione multipla ci serviremo dell'algebra lineare e indicheremo con lettere in **grassetto** i vettori e le matrici; più precisamente:

— un **vettore** sarà indicato con una lettera in grassetto **minuscolo** o con un numero in grassetto, ad esempio, il vettore nullo sarà indicato con **0**;
— una **matrice** sarà indicata con una lettera in grassetto **maiuscolo**.

La stessa convenzione è stata adoperata in **Appendice A** per illustrare le matrici e le loro proprietà.

2. SPECIFICAZIONE DI UN MODELLO DI REGRESSIONE MULTIPLA

La rappresentazione formale del modello considera la variabile dipendente Y come una funzione lineare di $k - 1$ variabili esplicative $X_2, X_3, ..., X_k$.

Il modello di regressione multipla sarà:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_k X_k + \varepsilon \tag{2.1}$$

in cui la variabile Y è funzione di una combinazione lineare di valori della variabile X. Geometricamente l'equazione data definisce l'**iperpiano di regressione** nello spazio a k dimensioni.

La (2.1), in presenza di n osservazioni campionarie, dà luogo al seguente sistema di equazioni:

$$\begin{array}{l} y_1 = \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + ... + \beta_k x_{1k} + \varepsilon_1 \\ y_2 = \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + ... + \beta_k x_{2k} + \varepsilon_2 \\ \dots \; \dots \; \dots \; \dots \; \dots \; \dots \; \dots \; \dots \; \dots \\ y_n = \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + ... + \beta_k x_{nk} + \varepsilon_n \end{array}$$

Il modello può essere rappresentato con notazione matriciale nel seguente modo:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

dove:

— $\boldsymbol{y}$ è il vettore $n \times 1$ costituito dall'insieme ordinato delle n osservazioni sulla variabile dipendente, realizzazioni della v.c. Y, ossia:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{2.3}$$

— $\boldsymbol{X}$ è la matrice $n \times k$ ottenuta dall'insieme ordinato delle n osservazioni sulle $k-1$ variabili esplicative, ossia:

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{12} & x_{13} & ... & x_{1k} \\ 1 & x_{22} & x_{23} & ... & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} & ... & x_{nk} \end{bmatrix} \tag{2.4}$$

in cui il vattore unitario nella prima colonna ha la funzione di inserire la costante β_1 e rappresenta, quindi, una varibile esplicativa artificiosa, pari a 1 per ogni unità statistica.

— $\boldsymbol{\beta}$ è il vettore costituito dai k coefficienti delle variabili esplicative X_i, ossia:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \tag{2.5}$$

— $\boldsymbol{\varepsilon}$ è il vettore $n \times 1$ delle manifestazioni non osservabili della componente stocastica di disturbo negli n tempi cui si riferiscono le rilevazioni campionarie, ossia:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.6}$$

Ipotesi classiche del modello di regressione multipla

Le ipotesi su cui si basa il modello di regressione multipla sono:

1. Ipotesi di linearità

L'ipotesi specifica il vettore delle osservazioni campionarie sulla variabile *Y* osservata negli *n* tempi come una combinazione lineare tramite i coefficienti β delle osservazioni sulle variabili esplicative X_i, $i = 1, 2, \ldots, k$, cui si aggiunge, in ogni tempo, un vettore di disturbo $\boldsymbol{\varepsilon}$; in simboli:

$$\mathbf{y} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_3 + \ldots + \beta_k \mathbf{x}_k + \boldsymbol{\varepsilon}$$

in cui $\mathbf{x}_i$ (i = 1, 2, …, k) sono i vettori colonna di osservazioni nella matrice $\mathbf{X}$ (2.4), dove:

$$\mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix} \qquad i = 1, 2, \ldots k$$

2. Ipotesi di non sistematicità degli errori

L'influenza dell'errore su *Y* è, in media, nulla, o ciò che è lo stesso, il valore atteso:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

o, in forma estesa, per ciascuna delle *n* osservazioni:

$$E(\boldsymbol{\varepsilon}) = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Tale ipotesi, formalmente, implica la seguente:

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

3. Ipotesi di sfericità degli errori

Si tratta di una duplice ipotesi distinguibile a sua volta, in:

3.1 Ipotesi di omoschedasticità per cui la variabilità dell'errore è costante:

$$Var(\varepsilon_1) = Var(\varepsilon_2) = \ldots = Var(\varepsilon_n) = \sigma^2$$

3.2 Ipotesi di covarianza nulla di errori relativi a osservazioni diverse:

$$Cov(\varepsilon_1, \varepsilon_2) = Cov(\varepsilon_1, \varepsilon_3) = \ldots = Cov(\varepsilon_n, \varepsilon_{n-1}) = 0$$

Sia $E(\boldsymbol{\varepsilon\varepsilon}')$ la matrice di varianze e covarianze di $\boldsymbol{\varepsilon}$, questa duplice ipotesi in simboli si traduce nella seguente espressione:

$$E(\boldsymbol{\varepsilon\varepsilon}') = \sigma^2 \boldsymbol{I}$$

dove $\boldsymbol{I}$ è la matrice identità; in altri termini, la matrice di varianze e covarianze dei termini d'errore è:

$$\sum = \begin{bmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1,\varepsilon_2) & \ldots & Cov(\varepsilon_1,\varepsilon_n) \\ Cov(\varepsilon_2,\varepsilon_1) & Var(\varepsilon_2) & \ldots & Cov(\varepsilon_2,\varepsilon_n) \\ \ldots & \ldots & \ldots & \ldots \\ Cov(\varepsilon_n,\varepsilon_1) & Cov(\varepsilon_n,\varepsilon_2) & \ldots & Var(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix}$$

4. Ipotesi di non collinearità delle variabili esplicative

Fra due o più delle variabili esplicative non esiste una perfetta dipendenza lineare; in altri termini, nessuna variabile esplicativa può essere espressa come combinazione lineare delle altre variabili esplicative.

Il rango della matrice $\boldsymbol{X}$, ossia il numero massimo di righe e colonne linearmente indipendenti, è:

$$\rho(\boldsymbol{X}) = k$$

5. Ipotesi di non stocasticità delle variabili esplicative

Formalmente, la matrice $\boldsymbol{X}$ non è stocastica ma deterministica.

6. Ipotesi di normalità degli errori

Un'ulteriore ipotesi del modello di regrssione concerne la normalità dei termini d'errore ed, essendo, dalle ipotesi 2 e 3, rispettivamente, $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ e $E(\boldsymbol{\varepsilon\varepsilon}') = \sigma^2\boldsymbol{I}$, tale ipotesi si traduce nella seguente espressione analitica:

$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$$

Talvolta, i dati empirici violano una o più ipotesi del modello lineare di regressione. Nei capitoli seguenti ci occuperemo della violazione di alcune di tali assunzioni.

Altra ipotesi fondamentale, che non abbiamo esplicitato, è quella di correlazione nulla tra variabili esplicative e termine di errore. Vedremo nel capitolo sesto le conseguenze sul modello della violazione di tale ipotesi.

3. STIMA CON IL METODO DEI MINIMI QUADRATI ORDINARI (OLS)

Il procedimento di **stima** in un modello di regressione multipla è volto alla determinazione del vettore $\boldsymbol{\beta}$ costituito dagli elementi $(\beta_1, \beta_2, \ldots, \beta_k)$, e alla varianza dei termini di errore σ^2. Occupiamoci dapprima della stima del vettore dei coefficienti di regressione.

3.1 Stima di $\boldsymbol{\beta}$

Per applicare il **metodo dei minimi quadrati ordinari** *OLS*, consideriamo il modello espresso in forma compatta $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$; esso può essere scritto nel modo seguente:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \ldots & \boldsymbol{x_k} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \boldsymbol{\varepsilon} \tag{3.1}$$

in cui $\boldsymbol{x}_i$ (i = 1, 2, ..., k) sono i vettori colonna delle osservazioni nella matrice $\boldsymbol{X}$ (2.4).

La (3.1), eseguendo i prodotti, diviene:

$$\boldsymbol{y} = \beta_1 \boldsymbol{x_1} + \beta_2 \boldsymbol{x_2} + \ldots + \beta_k \boldsymbol{x_k} + \boldsymbol{\varepsilon} \tag{3.2}$$

Se nella (3.2) al vettore (2.5) di parametri incogniti β_i, i = 1, 2, ..., k, si sostituisce il vettore dei valori arbitrari seguente:

$$\hat{\boldsymbol{\beta}}^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_k^* \end{bmatrix}$$

allora a ciascuna delle t osservazioni y_t corrisponde un valore calcolato:

$$\hat{y}_t = \hat{\beta}_1^* x_{t1} + \hat{\beta}_2^* x_{t2} + \ldots + \hat{\beta}_k^* x_{tk} \qquad t = 1, 2, \ldots, n$$

e la differenza:

$$e_t = y_t - \hat{\beta}_1^* x_{t1} - \hat{\beta}_2^* x_{t2} - \ldots - \hat{\beta}_k^* x_{tk} \qquad t = 1, 2, \ldots, n$$

costituisce un residuo che, espresso in termini matriciali, assume la seguente notazione:

$$\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{3.3}$$

Pertanto, ripetendo per tutte le osservazioni campionarie, il modello diviene:

$$\boldsymbol{y} = \hat{\beta}_1^* \boldsymbol{x_1} + \hat{\beta}_2^* \boldsymbol{x_2} + \ldots + \hat{\beta}_k^* \boldsymbol{x_k} + \boldsymbol{e} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^* + \boldsymbol{e}$$

ossia:

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^* + \boldsymbol{e} \tag{3.4}$$

Per la determinazione dei parametri incogniti, il metodo dei minimi quadrati consiste nel minimizzare la somma dei quadrati dei residui espressa da:

$$\boldsymbol{e}'\boldsymbol{e} = \begin{bmatrix} e_1 & e_2 & \ldots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = e_1^2 + e_2^2 + \ldots + e_n^2$$

Essendo, dalla (3.4), $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*$, si ha che la quantità da minimizzare è la seguente:

$$\boldsymbol{e}'\boldsymbol{e} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right)' \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right) \tag{3.5}$$

ossia deve essere:

$$\boldsymbol{f}\left(\hat{\boldsymbol{\beta}}^*\right) = \boldsymbol{e}'\boldsymbol{e} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right)' \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right) = \min$$

Effettuando i prodotti, la (3.5) diviene:

$$\begin{aligned} \boldsymbol{e}'\boldsymbol{e} &= \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right)' \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right) \\ &= \left(\boldsymbol{y}' - \hat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\right)\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^*\right) \\ &= \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}^* + \hat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}^* \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\hat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\boldsymbol{y} + \hat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}^* \end{aligned}$$

in cui $\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}^*$, essendo uno scalare, è uguale al trasposto $\hat{\boldsymbol{\beta}}^{*\prime}\boldsymbol{X}'\boldsymbol{y}$.

Per minimizzare $\boldsymbol{f}\left(\hat{\boldsymbol{\beta}}^*\right) = \boldsymbol{e}'\boldsymbol{e}$, per la **condizione del primo ordine** per l'esistenza di un punto di minimo, occorre uguagliare a zero la derivata parziale rispetto a $\hat{\boldsymbol{\beta}}^*$:

$$\frac{\partial f\left(\hat{\boldsymbol{\beta}}^*\right)}{\partial \hat{\boldsymbol{\beta}}^*} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}^* = \boldsymbol{0} \tag{3.6}$$

da cui si definiscono **equazioni normali** le equazioni risultanti dalla relazione:

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}^* = \boldsymbol{X}'\boldsymbol{y} \tag{3.7}$$

Moltiplicando ambo i membri della (3.7) per $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ si ha:

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

da cui, lo stimatore OLS è dato dal vettore dei parametri $\hat{\boldsymbol{\beta}}$ espresso da:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{3.8}$$

È importante rilevare che la determinazione di $\hat{\boldsymbol{\beta}}$ è possibile solo se la matrice prodotto $\boldsymbol{X'X}$ è invertibile. Essendo le matrici $\boldsymbol{X'}$ e $\boldsymbol{X}$, di ordine, rispettivamente, $k \times n$ e $n \times k$, la matrice $\boldsymbol{X'X}$ è una matrice quadrata $k \times k$; essa ha rango:

$$\rho(\boldsymbol{X'X}) = k$$

pertanto, è non singolare ed esiste la sua inversa $(\boldsymbol{X'X})^{-1}$.

Affinché si tratti di un minimo occorre differenziare nuovamente l'equazione $\dfrac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}}$;

ottenendo: $\dfrac{\partial^2 f(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}^2} = 2(\boldsymbol{X'X})$.

Essendo la matrice $\boldsymbol{X'X}$ definita positiva, la **condizione del secondo ordine** è soddisfatta.

Esempio

La tabella seguente riporta le osservazioni campionarie su una variabile Y e su due corrispondenti variabili esplicative:

Y	X_2	X_3
4	4	9
4	4	11
6	9	21
6	5	17
7	3	16
7	4	17
6	2	12
5	7	16
4	5	12
5	4	13

Tabella 1

Stimare i parametri del modello di regressione multipla:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

In realtà si tratta di un modello di regressione trivariata già visto nel capitolo precedente e per il quale non è strettamente necessaria la notazione matriciale. Abbiamo presentato qui questo modello, in quanto l'estensione per stimare il vettore dei parametri del modello applicando la (3.8) occorre specificare la matrice $\boldsymbol{X'X}$ e il vettore $\boldsymbol{X'y}$. Praticamente, il calcolo delle stime OLS si ottiene utilizzando uno dei numerosi pacchetti statistici presenti sul mercato; non è agevole calcolare, infatti, la matrice inversa di $\boldsymbol{X'X}$.

Inserendo nella prima colonna della matrice $\boldsymbol{X}$ il vettore unitario, la matrice $\boldsymbol{X}$ e la matrice trasposta $\boldsymbol{X}'$ sono pari, rispettivamente, a:

$$\boldsymbol{X}=\begin{bmatrix}1 & 4 & 9\\ 1 & 4 & 11\\ \dots & \dots & \dots\\ 1 & 5 & 12\\ 1 & 4 & 13\end{bmatrix} \qquad \boldsymbol{X}'=\begin{bmatrix}1 & 1 & \dots & 1 & 1\\ 4 & 4 & \dots & 5 & 4\\ 9 & 11 & \dots & 12 & 13\end{bmatrix}$$

Pertanto, la matrice $\boldsymbol{X}'\boldsymbol{X}$, di ordine 3×3, è pari a:

$$\boldsymbol{X}'\boldsymbol{X}=\begin{bmatrix}10 & 47 & 144\\ 47 & 257 & 718\\ 144 & 718 & 2.190\end{bmatrix}$$

Il vettore $\boldsymbol{X}'\boldsymbol{y}$ è, invece, pari a:

$$\boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}1 & 1 & \dots & 1 & 1\\ 4 & 4 & \dots & 5 & 4\\ 9 & 11 & \dots & 12 & 13\end{bmatrix}\begin{bmatrix}4\\ 4\\ \vdots\\ 4\\ 5\end{bmatrix}=\begin{bmatrix}54\\ 252\\ 804\end{bmatrix}$$

Lo stimatore dei minimi quadrati del vettore $\boldsymbol{\beta}$, applicando la (3.8), è:

$$\hat{\boldsymbol{\beta}}=\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}10 & 47 & 144\\ 47 & 257 & 718\\ 144 & 718 & 2.190\end{bmatrix}^{-1}\begin{bmatrix}54\\ 252\\ 804\end{bmatrix}=\begin{bmatrix}1{,}92845\\ -0{,}51793\\ 0{,}41013\end{bmatrix}$$

Pertanto, l'equazione del modello di regressione è la seguente:

$$\hat{Y}=1{,}92845-0{,}51793X_2+0{,}41013X_3$$

Nel calcolare $\hat{\boldsymbol{\beta}}$ si è dovuta calcolare la matrice inversa di $\boldsymbol{X}'\boldsymbol{X}$; si rimette al lettore la determinazione della stessa con l'ausilio della teoria sulle matrici riportata nell'Appendice A.

3.2 Stima di σ^2

Una stima della varianza dei termini di errore è ottenuta in funzione della devianza residua $\boldsymbol{e}'\boldsymbol{e}$. Ma, prima, occorre introdurre la matrice $\boldsymbol{M}$ desumibile da:

$$\begin{aligned}\boldsymbol{e}&=\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}=\boldsymbol{y}-\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}\\ &=\left[\boldsymbol{I}-\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right]\boldsymbol{y}=\boldsymbol{M}\boldsymbol{y}\end{aligned} \tag{3.9}$$

Nella (3.9) si è introdotta la nuova matrice:

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' \tag{3.10}$$

per la quale valgono le seuenti relazioni:
1. $\boldsymbol{M} = \boldsymbol{M}'$ (la matrice è simmetrica);
2. $\boldsymbol{M}'\boldsymbol{M} = \boldsymbol{M}\boldsymbol{M}' = \boldsymbol{M}$ (la matrice è idempotente);
3. $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$;
4. $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{\varepsilon}$;
5. $\boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$, in cui:

$$\begin{aligned} E\left(\boldsymbol{e}'\boldsymbol{e}\right) &= E\left(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}\right) \\ &= E\left\{tr\left(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}\right)\right\} && \text{essendo } \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon} \text{ uno scalare} \\ &= E\left\{tr\left(\boldsymbol{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\right\} \\ &= \sigma^2 tr\left(\boldsymbol{M}\right) && \text{essendo } E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) = \sigma^2\boldsymbol{I} \end{aligned} \tag{3.11}$$

in cui $tr(\mathbf{M})$ è la traccia della matrice $\boldsymbol{M}$. Dalla (3.10) si ha:

$$\begin{aligned} tr\left(\boldsymbol{M}\right) &= tr\left(\boldsymbol{I}\right) - tr\left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) \\ &= tr\left(\boldsymbol{I}\right) - tr\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) \\ &= n - k \end{aligned}$$

per cui, la (3.11) assume la seguente espressione:

$$E\left(\boldsymbol{e}'\boldsymbol{e}\right) = \sigma^2\left(n-k\right)$$

Pertanto, lo **stimatore** OLS della varianza σ^2 è definito da:

$$s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k} \tag{3.12}$$

La radice quadrata di s^2 rappresenta, come è noto, l'**errore standard della regressione**.
Il valore atteso dello stimatore s^2 è:

$$E\left(s^2\right) = \frac{E\left(\boldsymbol{e}'\boldsymbol{e}\right)}{n-k} = \frac{\sigma^2\left(n-k\right)}{n-k} = \sigma^2$$

e, quindi, lo stimatore è **non distorto**.

Pertanto, la stima della matrice delle varianze e covarianze degli stimatori $\hat{\boldsymbol{\beta}}$, ossia di $Var\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$, è:

$$es^2\left(\hat{\boldsymbol{\beta}}\right) = s^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

In particolare, la stima della varianza dello stimatore $\hat{\beta}_i$, $i = 1, 2, ..., k$, è:

$$es^2\left(\hat{\beta}_i\right) = s^2 a_{ii}$$

in cui a_{ii} è l'elemento i-esimo sulla diagonale principale della matrice $\left(\boldsymbol{X'X}\right)^{-1}$ e la cui espressione analitica sarà data nel capitolo quinto.

Se le v.c. errori $\boldsymbol{\varepsilon}$ sono **normali**, allora anche il vettore $\boldsymbol{e} = \boldsymbol{M\varepsilon}$ è una **v.c. normale multivariata** con:

$$E\left(\boldsymbol{e}\right) = \boldsymbol{0} \qquad \text{e} \qquad Var\left(\boldsymbol{e}\right) = Var\left(\boldsymbol{M\varepsilon}\right) = \boldsymbol{M}Var\left(\boldsymbol{\varepsilon}\right)\boldsymbol{M'} = \sigma^2\boldsymbol{M}$$

Per cui, si scrive:

$$\boldsymbol{\varepsilon} \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right) \Rightarrow \boldsymbol{e} \sim MN\left(\boldsymbol{0}, \sigma^2\boldsymbol{M}\right)$$

4. PROPRIETÀ DEGLI STIMATORI DEI MINIMI QUADRATI

Dalla (3.8) si evince che le stime OLS del vettore dei parametri $\boldsymbol{\beta}$ sono funzioni del vettore dei valori osservati su Y; al variare del campione casuale le stime $\hat{\boldsymbol{\beta}}$ generano **stimatori dei minimi quadrati** definiti da:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'y} \tag{4.1}$$

che, essendo $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, può essere espresso in maniera equivalente:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'y} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}\left(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}\right) \\ &= \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'X\beta} + \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'\varepsilon} = \boldsymbol{\beta} + \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'\varepsilon}\end{aligned} \tag{4.2}$$

Sia assegnata la matrice $\boldsymbol{C} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}$ per cui:

$$\boldsymbol{CC'} = \left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X} = \left(\boldsymbol{X'X}\right)^{-1}$$

lo stimatore $\hat{\boldsymbol{\beta}}$ (4.1) può essere scritto in maniera equivalente:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{Cy} = \boldsymbol{\beta} + \boldsymbol{C\varepsilon} \tag{4.3}$$

Le proprietà dello stimatore $\hat{\boldsymbol{\beta}}$ sono espresse nel seguente teorema di Gauss-Markov già visto nel capitolo secondo per un modello di regressione semplice. Ovviamente, in questo contesto, esso si esprime sotto forma matriciale.