Modello di Regressione Lineare Multiplo

Analisi dei Dati

Corso di Laurea in Scienze Statistiche e Attuariali
Dipartimento di Diritto, Economia, Management e Metodi Quantitativi (DEMM)
Università degli Studi del Sannio

Prof. Pietro Amenta

Fonte: *Lebart, Morineau, Fénelon, Traitement des données statistiques. Dunod*

Stima OLS. Definizione di un modello lineare.

Problema: Un imprenditore compra un negozio di alimentari che ha una superficie S in un quartiere la cui popolazione residente circostante è P. Si ipotizza che l'incasso giornaliero per questo tipo di negozio dipenda linearmente dalla superficie e dal numero di residenti. Vengono raccolti gli incassi di 13 negozi che insistono nel quartiere. Quale incasso può sperare l'acquirente nell'avviare l'attività ?

Popolazione	**Superficie**	**Incasso**
70	21	198
35	26	209
55	14	197
25	10	156
28	12	85
43	20	187
15	5	43
33	28	211
23	9	120
4	6	62
45	10	176
20	8	117
56	36	273

Stima OLS. Definizione di un modello lineare.

Possiamo scrivere

$$198 = \alpha_1 70 + \alpha_2 21 + \alpha_3 + \textit{residuo}_1$$
$$\vdots$$
$$273 = \alpha_1 56 + \alpha_2 36 + \alpha_3 + \textit{residuo}_{13}$$

quindi, con $i = 1, \ldots, 13$

$$y_i = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 + \epsilon_i$$

dove

- y_i è il valore della variabile endogena o da spiegare sull'i-esima u.s.;
- x_{i1} e x_{i2} sono i valori delle variabili esogene o esplicative (predittrici) sull'i-esima u.s.;
- α_1 e α_2 sono i coefficienti incogniti della combinazione lineare delle variabili esogene o esplicative (predittrici);
- α_3 è il termine costante del modello.

Stima OLS. Definizione di un modello lineare.

Vi sono quindi 13 equazioni per stimare 3 coefficienti α_i incogniti più altri 13 valori (incogniti) dati dai residui ϵ_i che rappresentano lo scarto fra il valore y_i e la parte lineare $(\alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3)$ dell'osservazione i-esima.

Abbiamo un sistema di 13 equazioni in 16 incognite: il sistema ammette quindi una infinità di soluzioni e siano a_1, a_2 e a_3 una di queste.

Si cercherà allora una soluzione a_1, a_2 e a_3 che minimizza "globalmente" l'insieme degli scarti dalla linearità e_i

$$e_i = y_i - (a_1 x_{i1} + a_2 x_{i2} + a_3)$$

Abbiamo diversi criteri di minimizzazione: $\min\{\sum_i^n e_i^2\}$ (**minimi quadrati**), $\min\{\sum_i^n |e_i|\}$, $\min\{\max e_i\}$, ecc..

Stima OLS. Definizione di un modello lineare.

Generalizziamo a p variabili e utilizziamo il criterio dei minimi quadrati per la stima dei coefficienti $a_1, a_2, ..., a_p$. Il modello per le n u.s. può essere scritto

$$\underset{(n,1)}{\mathbf{y}} = \alpha_1 \underset{(n,1)}{\mathbf{x}_1} + \alpha_2 \underset{(n,1)}{\mathbf{x}_2} + \ldots + \alpha_p \underset{(n,1)}{\mathbf{x}_p} + \underset{(n,1)}{\boldsymbol{\epsilon}}$$

o ancora, in termini matriciali

$$\underset{(n,1)}{\mathbf{y}} = \underset{(n,p)}{\mathbf{X}} \underset{(p,1)}{\boldsymbol{\alpha}} + \underset{(n,1)}{\boldsymbol{\epsilon}}$$

dove $\underset{(n,p)}{\mathbf{X}} = \left[\underset{(n,1)}{\mathbf{x}_1} \Big| \underset{(n,1)}{\mathbf{x}_2} \Big| \ldots \Big| \underset{(n,1)}{\mathbf{x}_p} \right]$ e $\underset{(p,1)}{\boldsymbol{\alpha}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$.

Stima OLS. Definizione di un modello lineare.

L'adattamento del modello sarà quindi

$$\underset{(n,1)}{\mathbf{y}} = a_1 \underset{(n,1)}{\mathbf{x}_1} + a_2 \underset{(n,1)}{\mathbf{x}_2} + \ldots + a_p \underset{(n,1)}{\mathbf{x}_p} + \underset{(n,1)}{\mathbf{e}}$$

oppure

$$\underset{(n,1)}{\mathbf{y}} = \underset{(n,p)}{\mathbf{X}} \underset{(p,1)}{\mathbf{a}} + \underset{(n,1)}{\mathbf{e}}$$

dove $\underset{(n,p)}{\mathbf{X}} = \left[\underset{(n,1)}{\mathbf{x}_1} \,\middle|\, \underset{(n,1)}{\mathbf{x}_2} \,\middle|\, \ldots \,\middle|\, \underset{(n,1)}{\mathbf{x}_p} \right]$, $\underset{(p,1)}{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$, $\underset{(n,1)}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$.

Stima OLS. Definizione di un modello lineare.

L'adattamento del modello ai minimi quadrati quindi corrisponde alla soluzione che minimizza il prodotto scalare

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{a})^T(\mathbf{y} - \mathbf{X}\mathbf{a}) = \mathbf{y}^T\mathbf{y} - 2\mathbf{a}^T\mathbf{X}^T\mathbf{y} + \mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{a}$$

e risulta essere funzione del vettore incognito **a**. Una condizione necessaria per il minimo è che sia nulla la derivata prima parziale

$$\frac{\partial(\mathbf{e}^T\mathbf{e})}{\partial\mathbf{a}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\mathbf{a} = 0$$

da cui si ottiene la condizione per l'estremo

$$\mathbf{X}^T\mathbf{X}\mathbf{a} = \mathbf{X}^T\mathbf{y}$$

che risulta essere un sistema di p equazioni in p incognite.

Stima OLS. Definizione di un modello lineare.

Se $n > p$ e se $\mathbf{X}$ è di rango massimo (cioè p), allora $\mathbf{X}^T\mathbf{X}$ è una matrice di ordine p e di rango p, dunque è invertibile, e quindi avremo

$$\mathbf{a} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Verifichiamo se la soluzione è anche un minimo. Sia $\tilde{\mathbf{a}}$ un'altra soluzione ed $\tilde{\mathbf{e}}$ il vettore degli scarti corrispondente

$$\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{X}\tilde{\mathbf{a}} = (\mathbf{y} - \mathbf{X}\mathbf{a}) + (\mathbf{X}\mathbf{a} - \mathbf{X}\tilde{\mathbf{a}}) = \mathbf{e} + \mathbf{X}(\mathbf{a} - \tilde{\mathbf{a}})$$

$$\begin{aligned}\tilde{\mathbf{e}}^T\tilde{\mathbf{e}} &= \mathbf{e}^T\mathbf{e} + 2(\mathbf{a} - \tilde{\mathbf{a}})^T\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{a}) + (\mathbf{a} - \tilde{\mathbf{a}})^T\mathbf{X}^T\mathbf{X}(\mathbf{a} - \tilde{\mathbf{a}}) \\ &= \mathbf{e}^T\mathbf{e} + (\mathbf{X}(\mathbf{a} - \tilde{\mathbf{a}}))^T(\mathbf{X}(\mathbf{a} - \tilde{\mathbf{a}}))\end{aligned}$$

dalla quale si evince che $\mathbf{e}^T\mathbf{e}$ è la somma di quadrati più piccola ottenibile (minimo assoluto), essendo tutte le quantità maggiori o uguali a zero.

Stima OLS. Definizione di un modello lineare.

$\underset{(13,1)}{\mathbf{y}} =$

198
209
197
156
85
187
43
211
120
62
176
117
273

$\underset{(13,3)}{\mathbf{X}} =$

$\mathbf{x}_1$	$\mathbf{x}_2$	**1**
70	21	1
35	26	1
55	14	1
25	10	1
28	12	1
43	20	1
15	5	1
33	28	1
23	9	1
4	6	1
45	10	1
20	8	1
56	36	1

$\underset{(3,3)}{\mathbf{X}^T\mathbf{X}} =$

19828	8452	452
8452	4343	205
452	205	13

$\underset{(3,3)}{(\mathbf{X}^T\mathbf{X})^{-1}} =$

0.00039	-0.00047	-0.00630
-0.00047	0.00146	-0.00669
-0.00630	-0.00669	0.40147

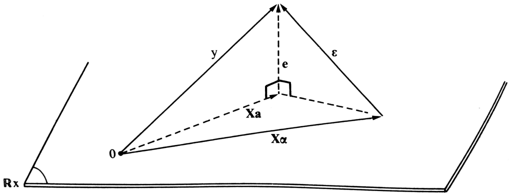
$\underset{(3,1)}{\mathbf{X}^T\mathbf{y}} =$

82495
38769
2034

$$\underset{(3,1)}{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \boxed{\begin{array}{c} 1.50 \\ 4.24 \\ 37.50 \end{array}}$$

$$\boxed{\underset{(13,1)}{\hat{\mathbf{y}}} = 1.50 \underset{(13,1)}{\mathbf{x}_1} + 4.24 \underset{(13,1)}{\mathbf{x}_2} + 37.50}$$

Stima OLS. Definizione di un modello lineare.



- Moltiplicando a sinistra $\mathbf{y} = \mathbf{Xa} + \mathbf{e}$ per $\mathbf{X}^T$ otteniamo $\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{Xa} + \mathbf{X}^T\mathbf{e}$. Ciò implica che $\mathbf{X}^T\mathbf{e} = 0$ e quindi il vettore $\mathbf{e}$ risulta essere ortogonale allo spazio $\Re_\mathbf{X}$.
- Il modello $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ definisce una decomposizione di $\mathbf{y}$ in due parti sconosciute: $\mathbf{X}\boldsymbol{\alpha} \in \Re_\mathbf{X}$ ed $\boldsymbol{\epsilon} \in \Re^n$.
- La tecnica dei minimi quadrati fornisce come soluzione la decomposizione $\mathbf{y} = \mathbf{Xa} + \mathbf{e}$ che minimizza la lunghezza (norma) di $\mathbf{e}$ proiettando ortogonalmente $\mathbf{y}$ su $\Re_\mathbf{X}$: i due vettori $\mathbf{Xa}$ ed $\mathbf{e}$ sono ortogonali.

Stima OLS. Definizione di un modello lineare.

La proiezione ortogonale di **y** su $\Re_{\mathbf{X}}$ è una trasformazione lineare caratterizzata da una matrice $\mathbf{P_X}$ denominata **Operatore di Proiezione Ortogonale**

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{a} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{P_X}\mathbf{y}$$

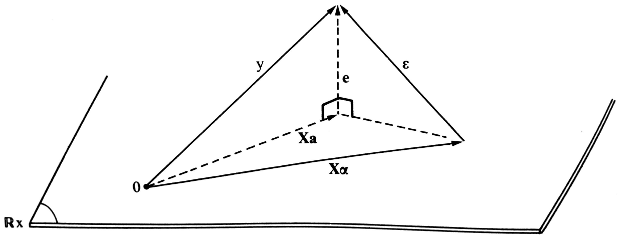
dove $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ e con $\mathbf{Q_X} = \mathbf{I} - \mathbf{P_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ operatore di proiezione orto-complementare rispetto allo spazio $\Re_{\mathbf{X}}$.
Si verifica con facilità che $\mathbf{e} = \mathbf{Q_X}\mathbf{y}$ e $\mathbf{Q_X}\mathbf{X} = \mathbf{0}$, ed è vero inoltre

$$\mathbf{e}^T\mathbf{e} = (\mathbf{Q_X}\mathbf{y})^T\mathbf{Q_X}\mathbf{y} = \mathbf{y}^T\mathbf{Q}_{\mathbf{X}}^2\mathbf{y} = \mathbf{y}^T\mathbf{Q_X}\mathbf{y}$$

Ma dato che $\mathbf{e} = \mathbf{Q_X}(\mathbf{X}\alpha + \epsilon) = \mathbf{Q_X}\epsilon$, allora possiamo scrivere

$$\mathbf{e}^T\mathbf{e} = \epsilon^T\mathbf{Q_X}\epsilon$$

Stima OLS. Definizione di un modello lineare.



- La tecnica dei minimi quadrati fornisce come soluzione la decomposizione $\mathbf{y} = \mathbf{Xa} + \mathbf{e}$ che minimizza la lunghezza (norma) di $\mathbf{e}$ proiettando ortogonalmente $\mathbf{y}$ su $\Re_{\mathbf{X}}$: i due vettori $\mathbf{Xa}$ ed $\mathbf{e}$ sono ortogonali.
- $\hat{\mathbf{y}} = \mathbf{Xa} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$
- $\mathbf{e} = \mathbf{Q}_{\mathbf{X}}\mathbf{y} = \mathbf{Q}_{\mathbf{X}}\epsilon$
- $\hat{\epsilon} = \mathbf{P}_{\mathbf{X}}\epsilon$

Stima OLS. Modello senza intercetta.

- Il modello lineare precedente si può anche scrivere

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\alpha}_0 + \mathbf{1}\alpha_p + \boldsymbol{\epsilon}$$

  con α_p termine costante del modello e $\mathbf{X}_0$ matrice delle $p-1$ variabili esogene. Risulta essere allora un caso particolare di $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ dove $\mathbf{X} = [\mathbf{X}_0|\mathbf{1}]$ e $\boldsymbol{\alpha}^T = [\boldsymbol{\alpha}_0|\alpha_p]$.
- $\mathbf{Q_1} = \mathbf{I} - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T = \mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T$ proiettore ortocomplementare allo spazio genenato da **1**.
- Se $\mathbf{z}$ è un punto di $\Re^n$ allora $\mathbf{Q_1}\mathbf{z} = (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{z} = \mathbf{z} - (1/n)\mathbf{1}\mathbf{1}^T\mathbf{z} = \mathbf{z} - \mathbf{1}\bar{z} = \hat{\mathbf{z}}$ (dove $\hat{\mathbf{z}}$ è il vettore $\mathbf{z}$ centrato rispetto alla media $\bar{z}$).
- $\hat{\mathbf{X}} = \mathbf{Q_1}\mathbf{X}$ matrice centrata rispetto alle medie di colonna

Stima OLS. Modello senza intercetta.

Se cerchiamo nuovamente la stima OLS dei parametri $\mathbf{a} = [\mathbf{a}_0 | a_p]$ del modello utilizzando la matrice $\mathbf{X} = [\mathbf{X}_0 | \mathbf{1}]$ otteniamo

$$\begin{cases} \mathbf{a}_0 = (\hat{\mathbf{X}}_0^T \hat{\mathbf{X}}_0)^{-1} \hat{\mathbf{X}}_0^T \hat{\mathbf{y}} & \text{per i } p-1 \text{ coefficienti } a_1, a_2, \ldots, a_{p-1} \\ a_p = \bar{y} - \sum_{k=1}^{p-1} a_k \bar{x}_k & \text{per il termine costante} \end{cases}$$

Se indichiamo con

- $\mathbf{V}_{\mathbf{XX}} = \frac{1}{n} \hat{\mathbf{X}}_0^T \hat{\mathbf{X}}_0$ la matrice delle varianze e covarianze empiriche fra le variabili esogene
- $\mathbf{V}_{\mathbf{Xy}} = \frac{1}{n} \hat{\mathbf{X}}_0^T \hat{\mathbf{y}}$ il vettore delle $p-1$ covarianze fra le $\mathbf{x}_k$ e $\mathbf{y}$

allora possiamo scrivere

$$\begin{cases} \mathbf{a}_0 = \mathbf{V}_{\mathbf{XX}}^{-1} \mathbf{V}_{\mathbf{Xy}} & \text{per i } p-1 \text{ coefficienti } a_1, a_2, \ldots, a_{p-1} \\ a_p = \bar{y} - \mathbf{a}_0^T \bar{\mathbf{x}} & \text{per il termine costante} \end{cases}$$

Stima OLS. Modello senza intercetta.

Sia $\tilde{\mathbf{y}} = \mathbf{X}\mathbf{a} = \mathbf{X}_0\mathbf{a}_0 + \mathbf{1}a_p$, allora possiamo scrivere

$$var(\tilde{\mathbf{y}}) = \frac{1}{n}\mathbf{a}_0^T\mathbf{X}_0^T\mathbf{X}_0\mathbf{a}_0 = var(\mathbf{X}_0\mathbf{a}_0) = cov(\mathbf{y}, \mathbf{X}_0\mathbf{a}_0) = cov(\mathbf{y}, \tilde{\mathbf{y}}) = R$$

dove R è il coefficiente di correlazione multiplo. Il suo quadrato si può esprimere in diversi modi

$$R^2 = \frac{cov(\mathbf{y}, \tilde{\mathbf{y}})}{var(\mathbf{y})var(\tilde{\mathbf{y}})} = \frac{var(\tilde{\mathbf{y}})}{var(\mathbf{y})} = \frac{cov(\mathbf{y}, \mathbf{X}_0\mathbf{a}_0)}{var(\mathbf{y})}$$

$$R^2 = \frac{\mathbf{a}_0^T\mathbf{X}_0^T\mathbf{X}_0\mathbf{a}_0}{\hat{\mathbf{y}}^T\hat{\mathbf{y}}} = \frac{\mathbf{a}_0^T\mathbf{X}_0^T\hat{\mathbf{y}}}{\hat{\mathbf{y}}^T\hat{\mathbf{y}}} = \frac{\sum_k a_k cov(\mathbf{y}, \mathbf{x}_k)}{\hat{\mathbf{y}}^T\hat{\mathbf{y}}}$$

Il termine R^2 assume quindi un significato in termini di varianza totale divisa in varianza spiegata e in quella residuale

$$var(\mathbf{y}) = var(\tilde{\mathbf{y}}) + var(\mathbf{e})$$

Stima OLS. Modello senza intercetta.

Possiamo scrivere quindi

$$\begin{cases} \text{varianza spiegata} & R^2 var(\mathbf{y}) = var(\tilde{\mathbf{y}}) \\ \text{varianza residuale} & \underline{(1 - R^2) var(\mathbf{y}) = var(\mathbf{e})} \\ \text{varianza totale} & var(\mathbf{y}) = var(\tilde{\mathbf{y}}) + var(\mathbf{e}) \end{cases}$$

L'R^2 è anche pari a

$$R^2 = 1 - \frac{var(\mathbf{e})}{var(\mathbf{y})} = 1 - \frac{\sum e_i^2}{n \times var(\mathbf{y})}$$

Minimizzare $\sum e_i^2$ equivale allora a massimizzare R^2 e quindi l'adattamento OLS determina la combinazione lineare delle variabili esogene di correlazione massima con la variabile endogena $\mathbf{y}$

Stima OLS. Modello senza intercetta.

- Si evidenzia che l'introduzione nel modello di una nuova variabile esogena fa diminuire la somma degli scarti al quadrato e di conseguenza fa aumentare l'R^2.
- Il queste condizioni, allora, il valore assunto dall'R^2 non può essere un riterio assoluto per apprezzare la qualità dell'adattamento.

Stima OLS. Modello senza intercetta.

Tornando ai dati dell'esempio, abbiamo i seguenti risultati

- $\mathbf{V}_{\mathbf{XX}}^{-1} = \begin{bmatrix} 316.3 & 101.9 \\ 101.9 & 85.4 \end{bmatrix}^{-1} = \begin{bmatrix} 0.00513 & -0.00612 \\ -0.00612 & 0.01901 \end{bmatrix}$
- $\mathbf{a}_0 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \mathbf{V}_{\mathbf{XX}}^{-1}\mathbf{V}_{\mathbf{Xy}} = \mathbf{V}_{\mathbf{XX}}^{-1} \begin{bmatrix} 905.7 \\ 515.0 \end{bmatrix} = \begin{bmatrix} 1.50 \\ 4.24 \end{bmatrix}$
- $a_3 = \bar{y} - (a_1\bar{x}_1 + a_2\bar{x}_2) = 37.50$
- $\sum e_i^2 = 13 \times [4137.6 - (1.50 \times 905.7 + 4.24 \times 515.01)] = 7756$
- $R = 0.93$
- $R^2 = 0.86 = 1 - \frac{7756}{13 \times 4137.6}$