IL METODO ECONOMETRICO

8 maggio 2017

L'obiettivo di questa lezione è quello di fornire alcuni strumenti necessari per l'analisi empirica. In particolare, approfondiremo il metodo econometrico come strumento cardine dell'analisi empirica. Il metodo econometrico cerca di attuare una fusione tra diverse fonti di informazioni: a) *dati*, ossia vettori di osservazioni che riguardano le famiglie, le imprese, i comuni, il debito, ecc...; b) *la teoria economica* (ad esempio le teorie Keynesiane in cui si dice che il conusmo dipende dal reddito) e c) *gli strumenti matematici* che formalizzano, attraverso la formulazione matematica, i concetti delle teorie economiche.

1 Alcuni concetti di base

Ogni variabile che osserviamo ha delle proprie caratteristiche statistiche che permettono di trarre alcune considerazioni. Supponiamo di avere a disposizione la varibile "spesa corrente" di tutti i comuni d'Italia. Tale variabile ha una propria media (o valore atteso) e una determinata varianza (deviazione standard o scarto quadratico medio). Immaginiamo di avere un ulteriore variabile: la popoalzione. Ora, potremmo vedere se esite una relazione tra l'andamento della spesa e la popolazione attraverso lo studio della covarianza delle due variabili oppure del coefficiente di correlazione e potremmo trovare che la spesa dei comuni italiani è collegata alla popolazione attarverso qualche relazione statistica.

Appare evidente, quindi, che per sintetizzare alcune relazioni economiche è necessario conoscere alcuni strumenti matematici e statistici.

1.1 Elementi di algebra lineare

Quando si vogliono analizzare i dati è molto utile rappresentarli sotto forma di vettori o di matrici. In particolare, una matrice è un insieme rettangolare di elementi disposti in righe e colonne. Una matrice A di ordine $N \times K$ è una matrice di $N \times K$ elementi, disposti su N righe e K colonne:

$$\underset{(N \times K)}{A} = \begin{pmatrix} a_{11} & \cdots & a_{1K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix}$$

dove a_{ij} è il generico elemento della matrice (i-esima riga e j-esima colonna). Da una matrice derivano quindi due vettori.

- Vettore colonna (di ordine N) è una matrice di ordine $N \times 1$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$$

- Vettore riga (di ordine K) è una matrice di ordine $1 \times K$

$$y = (y_1\, y_2\, \dots\, y_K)$$

1.1.1 Operazioni tra matrici

Di seguito sono definite le operazioni tra matrici ed alcune proprietà rilevanti di tali operazioni.

a) Addizione

La somma di due matrici dello stesso ordine genera una matrice C definita come:$C = A + B$ dove C è dello stesso ordine di A e B e un qualisiasi elemento generico $c_{ij} = a_{ij} + b_{ij}$.

b) Sottrazione

La differenza di due matrici dello stesso ordine genera una matrice C definita come $C = A - B$ dove C è dello stesso ordine di A e B e un qualisiasi elemento generico $c_{ij} = a_{ij} - b_{ij}$.

c) Moltiplicazione per uno scalare

Se moltiplichiamo una matrice A per uno scalareλ il loro prodotto è una matrice C dello stesso ordine di A dove $c_{ij} = \lambda a_{ij}$.

d) Prodotto fra matrici

Date una matrice A di ordine $N \times K$ e una matrice B di ordine $K \times M$, il loro prodotto AB (A è postmoltiplicata da B) è costituito da una matrice C di ordine $N \times M$. Il prodotto AB esiste perchè il numero delle colonne di A è uguale al numero di righe di B (le due matrici sono conformabili per la moltiplicazione). Viceversa, il prodotto BA non esiste poichè il numero di colonne di B non è uguale al numero di righe di A.

Date le seguenti due matrici di ordine 2×2 confromabili per la moltiplicazione,

$$\underset{(2\times2)}{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \; e \; \underset{(2\times2)}{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

il prodotto è uguale a C:

$$\underset{(2\times2)}{C} = \begin{pmatrix} (a_{11} \times b_{11} + a_{12} \times b_{21}) & (a_{11} \times b_{12} + a_{12} \times b_{22}) \\ (a_{21} \times b_{11} + a_{22} \times b_{21}) & (a_{21} \times b_{12} + a_{22} \times b_{22}) \end{pmatrix}$$

ciascun elemento della matrice C è ottenuto come prodotto interno di un vettore riga A e di un vettore colonna B

PROPRIETA' DEL PRODOTTO FRA MATRICI

- La moltiplicazione fra matrici non è necessariamente commutativa; generalmete, infatti, $AB \neq BA$.
- Il prodotto fra un vettore colonna $(N \times 1)$ e un vettore riga $(1 \times N)$ è una matrice $(N \times N)$.
- Il prodotto fra una matrice $(N \times N)$ e un vettore colonna $(N \times 1)$ è un vettore colonna.
- Il prodotto fra un vettore riga $(1 \times N)$ e una matrice $(N \times N)$ è un vettore riga $(1 \times N)$.

1.1.2 Trasposizione di una matrice

La matrice trasposta $A^{'}$ di una matrice A di ordine $N \times K$ è una matrice di ordine $K \times N$ ottenuta sostituendo le colonne (o le righe) con le righe (o le colonne) di A. Definiamo $A = (a_1\, a_2\, \ldots a_K)$ la matrice trasposta di A è

$$A^{'} = \begin{pmatrix} a_1^{'} \\ a_2^{'} \\ \vdots \\ a_K^{'} \end{pmatrix}$$

Ne consegue che la matrice trasposta di un vettore colonna x è un vettore riga: $x^{'} = (x_1\, x_2\, \cdots\, x_N)$.

1.1.3 Determinante

Ad ogni matrice quadrata è associato uno scalare, noto come determinante della matrice, det A. Se la matrice quadrata è di ordine 2×2 il determinante è:

$$det\, A = a_{11}a_{22} - a_{12}a_{21}$$

Se la matrice quadrata è di ordine 3×3 il determinante è:

$$det\, A = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

1.1.4 Inversione di una matrice

L'inversa di una matrice quadrata A è definita A^{-1}. Se esiste, è una matrice quadrata tale che

$$AA^{-1} = A^{-1}A = I$$

dove I è una matrice identità[1] dello stesso ordine di A.

Data la seguente matrice A di ordine 2×2

$$\underset{(2\times 2)}{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

occorre individuare il determinante della matrice:

$$det\, A = a_{11}a_{22} - a_{12}a_{21}$$

poi è necessario individuare la matrice dei cofattori o dei complementi algebrici. Il complemento algebrico in posizione i, j è definito come $cof(A, x_{i,j}) = (-1)^{i+j} \times minore(A, i, j)$ dove $minore(A, i, j)$ rappresenta il minore di A che si ottiene cancellando la riga i-esima e la colonna j-esima. Il segno $(-1)^{i+j}$ va nel seguente modo:

$$\begin{pmatrix} + & - & + & \ldots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

[1] La matrice identità, anche detta matrice identica o matrice unità, è una matrice quadrata in cui tutti gli elementi della diagonale principale sono costituiti dal numero 1, mentre i restanti elementi sono 0.

Nel nostro caso, il complemento algebrico di $a_{11} = a_{22}$, $a_{12} = -a_{21}$, $a_{21} = -a_{12}$e $a_{22} = a_{11}$.

Infine, la matrice dei complementi algebrici deve essere trasposta. Di conseguenza:

$$A^{-1} = \frac{1}{(a_{11}a_{22} - a_{12}a_{21})} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}$$

2 La stima econometrica

Supponiamo di volere indagare la relazione che esiste tra la spesa di un comune e determinate variabili, come ad esempio la popolazione, la ricchezza del comune, la superficie, ecc.. come facciamo?

Il metodo econometrico è utile quando si vuole indagare una relazione di causalità tra diverse variabili e può essere classificato in tre fasi: specificazione del modello, stima dei parametri e test.

2.1 Specificazione del modello

In questa prima fase di specificazione del modello un ruolo importante lo svolgono le ipotesi che si fanno su come è fatto il processo statistico che ha generato i nostri dati. La teoria economica suggerisce l'elenco delle variabili di interesse del problema che si intende affrontare e la direzione di causalità. Ad esempio $Y = f(X)$ cioè la variabile dipendente Y è spiegata dalla variabile esplicativa X: il nesso di causalità tra reddito e consumo non è spiegato nè dalla statistica nè dalla matematica ma dalla teoria economica.

D'altro canto la teoria economica da sola non basta per definire compiutamente tutti gli elementi in cui si compone un modello econometrico stimabile: per questo sono necessarie ipotesi di specificazione quali la scelta della forma funzionale. La matematica trasforma la semplice relazione $Y = f(X)$, in una relazione specifica evidenziando così il punto di vista quantitativo.

La relazione è inoltre *stocastica* per la presenza di un termine di errore, spesso additivo, cioè di una variabile casaule che serve a cogliere qualsiasi altro effetto non esplicitato (spesso non osservabile direttamente e non misurabile) e che dipende dalla qualità del modello. Tale variabile casuale è rappresentata da ε.

In maniera formale, la relazione lineare che lega una variabile ad un'altra in un rapporto di causalità si esplicita in questo modo:

$$Y = X\beta + \varepsilon$$

dove Y rappresenta la varibile dipendente del modello (ad esempio la spesa dei comuni) e X rappresenta l'insieme delle variabili esplicative (popolazione, reddito, superficie, ecc.) che cercano di "spiegare" la variabile dipendente.

Oltre all'ipotesi (1) di linearità della funzione si fanno altre quattro ipotesi:

$ipotesi\,(2):\ cov(X, \varepsilon) = 0$
$ipotesi\,(3):\ E(\varepsilon) = 0$
$ipotesi\,(4):\ E(\varepsilon^2) = \sigma^2$
$ipotesi\,(5):\ cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0\, \forall i \neq j$

2.2 Stima dei parametri e il modello OLS

Il parametro β è incognito e costituisce l'oggetto dell'inferenza statistica che viene condotta nell'ambito del modello di regressione lineare. Poter disporre delle stime di tale parametro ci permette di quantificare la relazione di causalità fra le variabili esplicative e la variabile dipendente.

Un metodo largamente utilizzato per la stima del modello parametrico è quello dei minimi quadrati ordinari (OLS ordinary least squares), che attribuisce ai parametri della relazione quei valori che minimizzano il quadrato delle distanze fra le osservazioni disponibili e la corrispondente retta di regressione: tali distanze sono anche detti residui ε. La stima di β viene scelta in modo da rendere più piccoli possibili i residui, per cui la miglior retta di regressione è ottenuta minimizzando la somma dei quadrati dei residui.

Definiamo il vettore colonna contenente le n osservazioni della variabile dipendente y e la matrice X contenente le n osservazioni campioanrie delle K variabili esplicative:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

$$\underset{n \times K}{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{pmatrix}$$

Infine, εè il vettore colonna contenente gli n termini di errore:

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

Il modello lineare può essere scritto nel modo seguente:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} \beta_1 + \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix} \beta_2 + \cdots + \begin{pmatrix} x_{1K} \\ x_{2K} \\ \vdots \\ x_{nK} \end{pmatrix} \beta_K + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

o in maniera pià compatta:

$$y = X\beta + \varepsilon$$

dove β è il vettore colonna dei K parametri da stimare.

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}$$

Con il metodo OLS si definiscono gli stimatori $\beta_1\, \beta_2 \ldots \beta_K$ tali da minimizzare la somma dei quadrati dei reisdui (RSS).

Definiamo il vettore dei residui come:

$$\varepsilon = y - X\beta$$

e quindi la somma dei quadrati dei reisdui è:

$$RSS = \sum_{i=1}^{N} \varepsilon_i^2 = \varepsilon^{'}\varepsilon = (y - X\beta)^{'}(y - X\beta)$$

Svuluppando tale formula si ottiene

$$RSS = (y - X\beta)^{'}(y - X\beta)$$

minimizzando RSS rispetto a β si ottiene:

$$\beta^{OLS} = (X^{'}X)^{-1}X^{'}y$$

3 Un caso applicato

Consideriamo di avere un campione di 15 studenti dell'Università di Ferrara, di cui il 33% sono femmine, e siamo interessati a misurare l'impatto che può avere l'essere maschio/femmina sull'atezza. Al fine di misurare tale impatto, decidiamo di utilizzare lo stimatore OLS.

Come procediamo?

In primo luogo definiamo la funzione che vogliamo stimare:

$$h_i = \beta_1 + \beta_2 g_i + u_i$$

dove h_i è la nostra variabile dipendente ossia l'altezza degli studenti, g_i è una variabile dicotomica (o dummy) che è uguale a 0 quando lo studente è un maschio e uguale a 1 quando lo studente e femmina e u_i è il termine di errore. L'equazione sopra prevede la stima di due parametri: β_1 β_2 e, in particolare, β_1 misura l'altezza media degli studenti maschi, mentre $\beta_2$cattura la differenza tra l'altezza media delle femmine rispetto ai maschi.

Per procedere alla stima OLS, innanzitutto riscriviamo la nostra equazione da stimare in maniera compatta:

$$h = X\beta + u$$

dove β è un vettore colonna: $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$.

Definiamo il vettore colonna della variabile dipendente:

$$\underset{(15\times 1)}{h} = \begin{pmatrix} 183 \\ 151 \\ 154 \\ 185 \\ 192 \\ 175 \\ 159 \\ 191 \\ 146 \\ 168 \\ 166 \\ 175 \\ 157 \\ 158 \\ 175 \end{pmatrix}$$

e la matrice X e di conseguenza $X^{'}$, delle variabili esplicative dove la prima colonna è composta da tutti 1 poichè rappresenta la costante del nostro modello e la seconda colonna è uguale a 1 quando lo studente osservato è femmina e 0 quando è maschio.

$$\underset{(15\times 2)}{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$$

e

$$\underset{(2\times 15)}{X^{'}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Abbiamo visto che lo stimatore OLS è uguale a : $\beta^{OLS} = (X^{'}X)^{-1}X^{'}y$. Calcoliamo prima $(X^{'}X)$ che ci darà una matrice di ordine 2×2, infatti stiamo moltiplicando una matrice $(X^{'})$ di dimensioni 2×15 con una matrice X di dimensioni 15×2

$$\underset{(2\times 15)}{X'}\underset{(15\times 2)}{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 15 & 5 \\ 5 & 5 \end{pmatrix}$$

Ora invertiamo la matrice $(X'X)$:

$$(X'X)^{-1} = \frac{1}{50}\begin{pmatrix} 5 & -5 \\ -5 & 15 \end{pmatrix} = \begin{pmatrix} 0,1 & -0,1 \\ -0,1 & 0,3 \end{pmatrix}$$

Infine, calcoliamo $X'y$ che ci darà un vettore colonna di ordine 2×1. Stiamo infatti moltipicando una matrice di dimensioni 2×15 con un vettore colonna di dimensioni 15×1

$$X'y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 183 \\ 151 \\ 154 \\ 185 \\ 192 \\ 175 \\ 159 \\ 191 \\ 146 \\ 168 \\ 166 \\ 175 \\ 157 \\ 158 \\ 175 \end{pmatrix} = \begin{pmatrix} 2535 \\ 805 \end{pmatrix}$$

Mettendo insieme i pezzi, otteniamo la stima dei parametri β_1 e β_2 come:

$$\beta^{OLS} = \underbrace{\begin{pmatrix} 0,1 & -0,1 \\ -0,1 & 0,3 \end{pmatrix}}_{(X'X)^{-1}} \underbrace{\begin{pmatrix} 2535 \\ 805 \end{pmatrix}}_{X'y} = \begin{pmatrix} 173 \\ -12 \end{pmatrix}$$

Per cui, il parametro stimato $\hat{\beta}_1$ è uguale a 173,4 mentre $\hat{\beta}_2$=-12,4. Ciò implica che l'altezza media dei maschi del nostro campione è pari a 173,4 cm, mentre le femmine sono, in media, 12,4 cm più basse.

Nella Fig.1 è mostrato l'esempio di cui sopra utilizzando il sofwtare econometrico Stata.

Figura 1: Tabella di regressione

```
      Source |       SS       df       MS              Number of obs =      15
-------------+------------------------------           F(  1,    13) =    2.45
       Model |         480     1         480           Prob > F      =  0.1415
    Residual |        2546    13  195.846154           R-squared     =  0.1586
-------------+------------------------------           Adj R-squared =  0.0939
       Total |        3026    14  216.142857           Root MSE      =  13.995

------------------------------------------------------------------------------
           h |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |      -12.4   7.665106    -1.57   0.141    -28.55945    4.559454
       _cons |      173.4   4.425451    39.09   0.000     163.4394    182.5606
------------------------------------------------------------------------------
```

Quindi la stima dell'altezza si ottiene utlizzandola seguente equazione:

$$h_i = 173,4 - 12,4g_i$$

Il valore dell'altezza standard (media) per gli uomini è 173,4, mentre per le donne è pari a 173,4-12,4=161. Questi ultmi valori sonno i cosidetti valori fittati della regressione rispettivamento quando $g_i = 0$, ovvero considero il caso dei maschi e $g_i = 1$, ovvero considero il caso delle femmine.