Modello di regressione lineare multiplo

Struttura del modello: supponiamo che esista una relazione che lega una variabile dipendente Y ad un vettore di $(k-1)$ variabili esplicative o regressori $X_* \equiv (X_2,\dots,X_k)'$. La forma funzionale che lega gli elementi di un campione casuale $\mathbf{Y} = (Y_1,\dots,Y_n)'$ della variabile dipendente alle corrispondenti osservazioni dei regressori è la seguente

$$\begin{cases} Y_1 = \beta_1 + \beta_2 x_{12} + \dots + \beta_j x_{1j} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ \dots \\ Y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik} + \varepsilon_i \\ \dots \\ Y_n = \beta_1 + \beta_2 x_{n2} + \dots + \beta_j x_{nj} + \dots + \beta_k x_{nk} + \varepsilon_n \end{cases},$$

dove x_{ij} indica la i-esima osservazione del j-esimo regressore e $\boldsymbol{\varepsilon} = (\varepsilon_1,\dots,\varepsilon_n)'$ è un vettore di termini casuali non osservabili detti errori.

In <u>forma matriciale</u>:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

dove $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)'$, $\mathbf{X} = (\mathbf{1}_n, \mathbf{X}_*)$, $\mathbf{X}_* = \{x_{ij}; i = 1,...,n, \ j = 2,...,k\})$ e $\mathbf{1}_n = \underbrace{(1,...,1)'}_{n}$

Ipotesi classiche sul modello:

1. $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$;
2. $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n$ (<u>errori</u> <u>incorrelati</u> ed <u>omoschedastici</u>);
3. Rg($\mathbf{X}$) = k < n;
4. la matrice **X** è fissa;
5. $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n)$.

Commenti:

- (1)-(4) $\Rightarrow$ E($\mathbf{Y}$) = $\mathbf{X}\boldsymbol{\beta}$ e Var($\mathbf{Y}$) = $\sigma_\varepsilon^2 \mathbf{I}_n$.
- (1)-(3) + (4′): X_* è una V.A.M. con matrice di varianze-covarianze finita e tale che E($\boldsymbol{\varepsilon}|\mathbf{X}$) = $\mathbf{0}$ $\Rightarrow$ E($\mathbf{Y}|\mathbf{X}$) = $\mathbf{X}\boldsymbol{\beta}$ e Var($\mathbf{Y}|\mathbf{X}$) = $\sigma_\varepsilon^2 \mathbf{I}_n$.

Stimatore dei minimi quadrati ordinari (OLS) dei parametri β:

$$\hat{\boldsymbol{\beta}} = \arg\min\{S(\boldsymbol{\beta})\}, \text{ dove } S(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

La soluzione è data da:

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

(*prova per esercizio*. *Traccia*: prova che $S(\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ e verifica le condizioni del primo e del secondo ordine).

La stima OLS dei parametri $\boldsymbol{\beta}$ è quindi data $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, dove $\mathbf{y} = (y_1,\ldots, y_n)'$ è il campione *ex-post* della variabile dipendente *Y*.

Commento. Date le seguenti definizioni:

- errori: $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$;
- errori di predizione: $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$;
- residui: $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$,

notiamo che i vettori $\boldsymbol{\varepsilon}$ ed $\hat{\boldsymbol{\varepsilon}}$ sono aleatori mentre il vettore $\mathbf{e}$ è una realizzazione di $\hat{\boldsymbol{\varepsilon}}$. La relazione tra $\boldsymbol{\varepsilon}$ ed $\hat{\boldsymbol{\varepsilon}}$ verrà chiarita più tardi.

Le matrici P e M: il modello stimato $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ può riscriversi come

$$\mathbf{y} = \mathbf{Py} + \mathbf{My}, \text{ dove } \mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \text{ e } \mathbf{M} = \mathbf{I}_n - \mathbf{P}.$$

Le matrici **P** e **M** godono delle seguenti proprietà:

1. (simmetria) $\mathbf{P}' = \mathbf{P}$ e $\mathbf{M}' = \mathbf{M}$;
2. (idempotenza) $\mathbf{PP} = \mathbf{P}$ e $\mathbf{MM} = \mathbf{M} \Rightarrow$ (ortogonalità) $\mathbf{PM} = \mathbf{MP} = \mathbf{0}$;
3. $\mathbf{PX} = \mathbf{X}$ e $\mathbf{MX} = \mathbf{0} \Rightarrow \mathrm{Rg}(\mathbf{P}) = k$ e $\mathrm{Rg}(\mathbf{M}) = (n - k)$
4. **P** e **M** sono semi-definite positive (*verifica per esercizio*. *Traccia*: usa le precedenti proprietà)

Proprietà descrittive della stima OLS:

a) (ortogonalità tra **X** ed **e**) $\mathbf{X}'\mathbf{e} = \mathbf{X}'\mathbf{My} = \mathbf{0} \Rightarrow$ (la media dei residui è nulla) $\mathbf{1}_n'\mathbf{e} = \mathbf{0}$;

b) (scomposizione della devianza totale in devianza spiegata e devianza residua) $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2$ dove $\hat{\mathbf{y}} = \mathbf{Xb}$ (*prova per es*. *Traccia*: usa la matrice $\mathbf{L} = \mathbf{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n' = \mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'$ per scrivere le devianze in forma matriciale e la proprietà (a)).

Coefficiente di determinazione:

$$R^2 = \hat{\mathbf{y}}'\mathbf{L}\hat{\mathbf{y}} \big/ \mathbf{y}'\mathbf{L}\mathbf{y} = (1 - \mathbf{e}'\mathbf{e} \big/ \mathbf{y}'\mathbf{L}\mathbf{y}) = (1 - \mathbf{y}'\mathbf{M}\mathbf{y} \big/ \mathbf{y}'\mathbf{L}\mathbf{y})$$

misura la frazione delle variabilità campionaria della variabile dipendente che è "spiegata" dai regressori.

Commento. Il coefficiente R^2 non decresce al crescere del numero dei regressori k (*verifica per esercizio*. *Traccia*: scrivi $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$ e mostra che $\mathbf{e}_1'\mathbf{e}_1 - \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{M}_1\mathbf{y} - \mathbf{y}'\mathbf{M}\mathbf{y} \geq 0$ dove $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1') \Rightarrow$ <u>Coefficiente di determinazione "corretto"</u>:

$$R_c^2 = 1 - \frac{\mathbf{y}'\mathbf{M}\mathbf{y} \big/ (n-k)}{\mathbf{y}'\mathbf{L}\mathbf{y} \big/ (n-1)}$$

La ragione dell'aggettivo "corretto" è che $s_Y = \mathbf{Y}'\mathbf{L}\mathbf{Y} \big/ (n-1)$ è uno stimatore corretto di σ_Y^2 e che $s_\varepsilon^2 = \mathbf{Y}'\mathbf{M}\mathbf{Y} \big/ (n-k)$ è, come mostreremo più avanti, una stimatore corretto di σ_ε^2.

c) (Relazione tra il modello stimato e la V.A. normale multivariata). Sia $\mathbf{b} = (b_1, \mathbf{b}_*)'$, possiamo allora scrivere la seguente relazione tra le variabili scarto:

$$\mathbf{Ly} = \mathbf{LX}_*\mathbf{b}_* + \mathbf{e},$$

(*verifica per esercizio*. *Traccia*: usa l'ortogonalità tra la matrice $\mathbf{L}$ e il vettore $\mathbf{1}_n$) dove $\mathbf{b}_* = (n^{-1}\mathbf{X}_*'\mathbf{LX}_*)^{-1}(n^{-1}\mathbf{X}_*'\mathbf{Ly}) \equiv \overline{\mathbf{\Sigma}}_{xx}^{-1}\overline{\mathbf{\Sigma}}_{xy}$. Ne segue che:

$$\hat{y}_i = \overline{y} + (\mathbf{X}_{i\bullet} - \overline{\boldsymbol{\mu}}_x)'\overline{\mathbf{\Sigma}}_{xx}^{-1}\overline{\mathbf{\Sigma}}_{xy},$$

$$n^{-1}\mathbf{e}'\mathbf{e} = n^{-1}(\mathbf{y}'\mathbf{Ly}) - \overline{\mathbf{\Sigma}}_{yx}\overline{\mathbf{\Sigma}}_{xx}^{-1}\overline{\mathbf{\Sigma}}_{xy} \equiv \overline{\sigma}_y^2 - \overline{\mathbf{\Sigma}}_{yx}\overline{\mathbf{\Sigma}}_{xx}^{-1}\overline{\mathbf{\Sigma}}_{xy}$$

Se $\begin{pmatrix} y \\ X_* \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \mathbf{\Sigma}_{yx} \\ \mathbf{\Sigma}_{xy} & \mathbf{\Sigma}_{xx} \end{pmatrix}\right]$, dalla proprietà #5 della normale multipla si ha: $y \mid X_* \sim N\left((\mu_y - \mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}(X_* - \boldsymbol{\mu}_x)), (\sigma_y^2 - \mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{yx})\right)$

Il metodo OLS stima quindi i parametri della distribuzione condizionata con i loro analoghi campionari $(\hat{y}_i, n^{-1}\mathbf{e}'\mathbf{e})$.

Proprietà della stimatore OLS: (è utile scrivere $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$)

1. (<u>correttezza</u>) $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ (*prova*: immediata dall'equazione di sopra);
2. $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$ (*prova*: *idem*)
3. (Teorema di Gauss-Markov): Lo stimatore OLS è BLUE.
   Dimostrazione: sia $\tilde{\boldsymbol{\beta}} = \mathbf{A}'\mathbf{Y} = \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathbf{A}'\boldsymbol{\varepsilon}$ un generico stimatore lineare e corretto. Poiché $\mathrm{E}(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ ne segue

$$(i)\ \mathbf{A}'\mathbf{X} = \mathbf{I}_k \Rightarrow \tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{A}'\boldsymbol{\varepsilon} \Rightarrow \mathrm{Var}(\tilde{\boldsymbol{\beta}}) = \sigma_\varepsilon^2\mathbf{A}'\mathbf{A}.$$

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2[\mathbf{A}'\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}] = \sigma_\varepsilon^2[\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}]$$

   $= \sigma_\varepsilon^2\mathbf{A}'\mathbf{M}\mathbf{A} = \sigma_\varepsilon^2\mathbf{A}'\mathbf{M}'\mathbf{M}\mathbf{A}$, che è pertanto una matrice s.d.p. la quale si annulla se e solo se $\mathbf{M}\mathbf{A} = \mathbf{0} \Rightarrow \mathbf{A}' = \boldsymbol{\alpha}'\mathbf{X}' \overset{(i)}{\Rightarrow} \boldsymbol{\alpha}' = (\mathbf{X}'\mathbf{X})^{-1}$. ■

Stimatore della varianza degli errori σ_ε^2: $s_\varepsilon^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/(n-k)$. *Dimostra per esercizio che* $\mathrm{E}(s_\varepsilon^2) = \sigma_\varepsilon^2$ (*suggerimento*: prova che $s_\varepsilon^2 = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}/(n-k)$ e usa le proprietà dell'operatore traccia).

Inferenza nel modello di regressione: se $\varepsilon \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n)$, allora

1. $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1}) \Rightarrow \dfrac{\hat{\beta}_i - \beta_i}{\sigma_\varepsilon \sqrt{a_{ii}}} \sim N(0,1)$, dove $\{a_{ij}\} = (\mathbf{X}'\mathbf{X})^{-1}$

2. $\dfrac{(n-k)s_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(n-k)$ (*prova per esercizio*. *Traccia*: usa la relazione $s_\varepsilon^2 = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} \big/ (n-k)$ e la diagonalizzazione della matrice **M**).

3. Poiché $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ e $\mathbf{M}\boldsymbol{\varepsilon}$ sono indipendenti (*prova per esercizio*) $\overset{(1) \text{e} (2)}{\Rightarrow}$
$\dfrac{\hat{\beta}_i - \beta_i}{s_\varepsilon \sqrt{a_{ii}}} \sim T(n-k)$

4. Sia $\underbrace{\mathbf{R}}_{k \times p}$ una matrice con Rg($\mathbf{R}$) = $p \Rightarrow \mathbf{R}'\hat{\boldsymbol{\beta}} \sim N(\mathbf{R}'\boldsymbol{\beta}, \sigma_\varepsilon^2 \mathbf{R}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R})$
$\Rightarrow (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{R}(\mathbf{R}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R})^{-1}\mathbf{R}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \big/ \sigma_\varepsilon^2 \sim \chi^2(q)$

5. Poiché $\mathbf{R}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ e $\mathbf{M}\boldsymbol{\varepsilon}$ sono indipendenti $\overset{(2) \text{e} (4)}{\Rightarrow}$
$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{R}(\mathbf{R}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R})^{-1}\mathbf{R}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \big/ s_\varepsilon^2 q \sim F(q, n-k)$

Test per restrizioni lineari sui coefficienti β:

- $H_0 : \beta_i = {}_0\beta_i$ vs. $H_1 : \beta_i \neq {}_0\beta_i$
- Statistica test: $T = \dfrac{\hat{\beta}_i - {}_0\beta_i}{s_\varepsilon \sqrt{a_{ii}}} \sim T(n-k)$
- Valore critico $t_\alpha : \Pr\{T(n-k) > t_\alpha\} = \alpha$.
- Regione di rifiuto di ampiezza α: $|T| > t_{\alpha/2}$

- $H_0 : \mathbf{R}'\boldsymbol{\beta} = \mathbf{r}$ vs. $H_1 : \mathbf{R}'\boldsymbol{\beta} \neq \mathbf{r}$
- Alcuni esempi: $\begin{cases} \beta_{k-s+1} = \beta_{k-s+2} = ... = \beta_k = 0 \Leftrightarrow \mathbf{R}' = [\underbrace{\mathbf{0}}_{s\times(k-s)}, \mathbf{I}_s], \mathbf{r} = \underbrace{\mathbf{0}}_{s\times 1} \\ \beta_2 + \beta_3 = 1 \Leftrightarrow \mathbf{R}' = [0,1,1,0,...,0], \mathbf{r} = 1 \end{cases}$
- Statistica test: $F = (\mathbf{R}'\hat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R})^{-1}(\mathbf{R}'\hat{\boldsymbol{\beta}} - \mathbf{r}) \big/ s_\varepsilon^2 q \sim F(q, n-k)$
- Valore critico $f_\alpha : \Pr\{F(q, n-k) > f_\alpha\} = \alpha$;
- Regione di rifiuto di ampiezza α: $F > f_\alpha$