Lezione n. 9 (a cura di Simona Cretaro)

Modello di regressione lineare multipla

È l'estensione del modello di regressione lineare semplice al caso di più variabili esplicative. La singola variabile esplicativa è chiamata covariata X.

Il modello è espresso nella forma:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots\dots\dots + \beta_k x_{ik} + \varepsilon_i$$

$\{ \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots\dots\dots + \beta_k x_{ik} \}$ $\Longrightarrow$ componente deterministica

ε_i $\Longrightarrow$ componente stocastica

Esistono k variabili x che determinano la y e k+1 parametri; xi1 è il valore della prima covariata in corrispondenza del soggetto i; in generale xij è il valore della covariata j sempre in corrispondenza del soggetto i. Naturalmente quando k=1 la relazione tra X e Y è descritta da un modello di regressione lineare semplice. Adattando l'esempio relativo a quest'ultimo, adesso si ha Y=reddito, x1=età, x2=anni di istruzione.

Nella popolazione di riferimento, la media della Y per chi ha una certa configurazione delle x, è uguale a:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta 2 x_{i2} + \dots\dots\dots + \beta_k x_{ik}$$

L'intercetta β_0, analogamente al caso di regressione lineare semplice, è il valore atteso della Y quando tutte le k covariate sono uguali a 0; i valori delle singole persone oscillano intorno alla media e ciò dipende dalla componente stocastica, mentre β_j=(j:1….k), il coefficiente angolare per la j-esima covariata, indica l'incremento del valore atteso $E(Y_i)$ se xj aumenta di 1 a parità delle altre covariate.

Nel nostro esempio;

β_0= 12= reddito medio che ho quando le 2 covariate sono uguali a 0 o comunque ad un valore fissato (es.l'età).

β_1= 1= se l'età aumenta di un anno, a parità di anni di istruzione, il reddito medio aumenta di 1000 euro.

β_2=2= se si aggiunge un anno di istruzione, a parità di età, il reddito medio aumenta di 2000 euro.

Assunzioni riguardo ε_i:

$\varepsilon_i = Yi - E(Y_i)$

$E(\varepsilon_i)=0$

$V(\varepsilon_i)=\sigma^2$ $i=1,\dots,n$ omoschedasticità

$\varepsilon i \perp\!\!\!\perp \varepsilon j$

La stima dei parametri presuppone la scrittura in forma matriciale.

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

Matrice del disegno X (n*k+1)=

$$\begin{pmatrix} 1 & X_{11} & X_{12} \dots\dots\dots & X_{1k} \\ 1 & X_{21} & X_{22} \dots\dots\dots & X_{2k} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ 1 & X_{n1} & X_{n2} \dots\dots\dots & X_{nk} \end{pmatrix} * \beta(k+1)*1 \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots\dots \\ \beta_n \end{pmatrix}$$

Il vettore colonna composto da tutti 1 serve per introdurre l'intercetta.

Isolo la riga della matrice x;

$\underline{\mathbf{X}}_i = (1\ x_{i1}\ x_{i2} \dots\dots x_{ik})$

$$X\underline{\boldsymbol{\beta}} = \begin{pmatrix} X'_1\, \underline{\boldsymbol{\beta}} \\ X'_2\, \underline{\boldsymbol{\beta}} \\ \dots\dots \\ X'_n\, \underline{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \beta_0 + x_{11}\, \beta_1 + \dots\dots\dots\dots x_{1k}\, \beta_k \\ \beta_0 + x_{21}\, \beta_1 + \dots\dots\dots\dots x_{2k}\, \beta_k \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \beta_0 + x_{n1}\, \beta_1 + \dots\dots\dots\dots x_{nk}\, \beta_k \end{pmatrix}$$

$$\underline{Y} = \begin{pmatrix} Y1 \\ Y2 \\ \ldots \\ Yn \end{pmatrix} \qquad \underline{\varepsilon} = \begin{pmatrix} \varepsilon 1 \\ \varepsilon 2 \\ \ldots \\ \varepsilon n \end{pmatrix}$$

Adesso possiamo applicare i metodi di stima:

<u>OLS= metodo dei minimi quadrati</u>

Devo minimizzare $S(\underline{\boldsymbol{\beta}}) = \sum_i (Y_i - \hat{Y}_i)^2 = \sum_i (Y_i - x'_i \underline{\boldsymbol{\beta}})^2$

$\hat{Y}_i = x'_i \underline{\boldsymbol{\beta}} = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$

$\hat{Y}_i$ è il valore previsto per un certo $\underline{\boldsymbol{\beta}}$

Esempio: $\hat{Y}_i = 10 + 1*18 + 2*8 = 44000$ è la previsione di reddito per un soggetto che ha 18 anni e ha studiato 8 anni dopo la scuola dell'obbligo

1) $\sum_i (Y_i - \underline{\boldsymbol{x}}'_i \underline{\boldsymbol{\beta}})^2 = (\underline{\boldsymbol{Y}} - x\,\underline{\boldsymbol{\beta}})'(\underline{\boldsymbol{Y}} - x\,\underline{\boldsymbol{\beta}})$

La 1) equivale alla somma dei quadrati. **Y** corrisponde ai valori osservati mentre x $\underline{\boldsymbol{\beta}}$ sono i valori previsti. Si fa il prodotto riga per colonna e otteniamo uno scalare.

$$V = \begin{pmatrix} V1 \\ \ldots \\ \ldots \\ Vn \end{pmatrix} \qquad = \qquad \underline{\boldsymbol{V'V}} = v1^2 + v2^2 + \ldots + vn^2$$

Pongo la derivata prima =0

$\partial S(\underline{\boldsymbol{B}}) / \partial \beta = 0$

$\hat{\underline{\boldsymbol{\beta}}}_{(k+1)*1} = (X'X)^{-1}_{(k+1)(k+1)} X'\underline{\boldsymbol{Y}}(k+1)*1$

$$\underline{\hat{\boldsymbol{\beta}}}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \dots\\ \hat{\beta}_k\end{pmatrix}$$

<u>ML= Metodo della massima verosimiglianza</u>

Le assunzioni che ho finora mi permettono di dire qual è il valore atteso e la varianza di Yi, ma non di conoscere la distribuzione di Y.

$E(Y_i)= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} +.........+\beta_k x_{ik}$

$V (Y_i)= \sigma^2$

Devo fare assunzioni su ε_i

$\varepsilon_i \sim N(0, \sigma^2)$

$Y_i \sim N(x_i'\underline{\boldsymbol{\beta}}, \sigma^2)$

$L(\underline{\boldsymbol{\beta}}, \sigma^2)=\prod_i f(Y_i)= \prod_i (1/\sqrt{2\Pi \sigma^2})*e-1/2(Y_i-x'_i\underline{\boldsymbol{\beta}}/ \sigma^2)^2$

$= (1/\sqrt{2\Pi \sigma^2})^n* e-1/2\sum_i(Y_i-x'_i\underline{\boldsymbol{\beta}}/ \sigma^2)^2$

Devo massimizzare rispetto a β

$\ell(\underline{\boldsymbol{\beta}}, \sigma^2)=-n/2 \log(2\Pi \sigma^2)-1/2\sum_i(Y_i-x'_i\underline{\boldsymbol{\beta}}/ \sigma^2)^2$

Solo la quantità $\sum_i(Y_i-x'_i\underline{\boldsymbol{\beta}}/)^2$ coinvolge β e massimizzare $\ell(\underline{\boldsymbol{\beta}}, \sigma^2)$ equivale a minimizzare $\sum_i(Y_i-x'_i\underline{\boldsymbol{\beta}})^2$, ovvero al metodo dei minimi quadrati. Quindi esclusivamente per $\underline{\boldsymbol{\beta}}$, lo stimatore di massima verosimiglianza è uguale a quello calcolato con il metodo dei minimi quadrati.

Per la varianza invece si hanno dei risultati diversi.

$OLS= S^2=DR/n-(k+1)=DR/gdl$

$ML=\hat{\sigma}^2=DR/n$

$DR= \sum_i(Y_i-\hat{Y}_i)^2 \quad \hat{Y}_i=\underline{\mathbf{x}}_i'\underline{\hat{\boldsymbol{\beta}}}$

S^2 è uno stimatore non distorto mentre $\hat{\sigma}^2$ non tiene conto del numero dei parametri.

$\varepsilon_i \sim N(0, \sigma^2)$ $i=1,\dots,n$

Si può esprimere sul vettore degli errori:

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

Il vettore degli errori si distribuisce come una normale multivariata e $\sigma^2 I$ è la matrice di varianza covarianza.

$$\sigma^2 I(n*n)=\begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

PROPRIETÀ DEGLI STIMATORI

$\hat{\underline{\mathbf{B}}}$ è lo stimatore di $\underline{\mathbf{B}}$

$$E(\hat{\underline{\mathbf{B}}})=\begin{pmatrix} E(\hat{\beta}_0) \\ E(\hat{\beta}_1) \\ \dots \\ E(\hat{\beta}_k) \end{pmatrix} = E((X'X)^{-1}X'\underline{\mathbf{Y}})$$

$=(X'X)^{-1}X'E(\underline{\mathbf{Y}})$

$=(X'X)^{-1}X'E(x\underline{\boldsymbol{\beta}}+\underline{\boldsymbol{\varepsilon}})$ $E(\underline{\boldsymbol{\varepsilon}})=0$

$=(X'X)^{-1}X'X\underline{\boldsymbol{\beta}}=\underline{\boldsymbol{\beta}}$

$E(\hat{\underline{\boldsymbol{\beta}}})=\underline{\boldsymbol{\beta}}$

OLS e ML danno stimatori non distorti

Varianza $\underline{\hat{\boldsymbol{\beta}}}$

Dalla regola $V(mz)= m^2V(Z)$ si può scrivere:

$V(M\underline{\mathbf{Z}})_{(r*r)}= MV(\underline{\mathbf{Z}})M'$

M ha dimensione r*c, $\underline{\mathbf{Z}}$ (c*1), $M(r*c)V(c*c)M'(c*r)$

$V(\underline{\hat{\boldsymbol{\beta}}})= V((X'X)^{-1}X'\underline{\mathbf{Y}})$ M= $(X'X)^{-1} X'$ $\underline{\mathbf{z}}=\underline{\mathbf{y}}$

$= (X'X)^{-1} X' V(\underline{\mathbf{Y}}) X(X'X)^{-1}$

$= (X'X)^{-1} X'[\sigma^2I] X (X'X)^{-1}$

$= \sigma^2(X'X)^{-1} X'X(X'X)^{-1}= \sigma^2(X'X)^{-1} \Longrightarrow$ è la matrice di varianza covarianza dello stimatore $\underline{\hat{\boldsymbol{\beta}}}$

$$V(\underline{\hat{\boldsymbol{\beta}}})= \sigma^2(X'X)^{-1}= \begin{pmatrix} V(\hat{\beta}0) & Cov(\hat{\beta}0,\hat{\beta}1) & \ldots & \ldots & Cov(\hat{\beta}0,\hat{\beta}k) \\ Cov(\hat{\beta}0,\hat{\beta}1) & V(\hat{\beta}1) & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ Cov(\hat{\beta}0,\hat{\beta}k) & \ldots & \ldots & \ldots & Var(\hat{\beta}k) \end{pmatrix}$$

$\hat{V}(\underline{\hat{\boldsymbol{\beta}}})$ è la varianza teorica se conosco σ^2

$\hat{V}(\underline{\hat{\boldsymbol{\beta}}})= S^2(X'X)^{-1}$ è la varianza nella pratica perché σ^2 non lo conosco

$\hat{V}(\hat{\beta}j)= diagj\ (S^2(X'X)^{-1})$ j=0,….,k è la varianza di uno specifico elemento

s.e. $(\hat{\beta}j)= \sqrt{V}(\hat{\beta}j)$ è la misura di errore dello stimatore