

## Lezione n. 16 (a cura di Peluso Filomena Francesca)

Oltre alle normali variabili risposta che presentano una continuità almeno all'interno di un certo intervallo di valori, esistono variabili risposta di tipo binario, che cioè possono assumere unicamente valori (0, 1). Per esempio:

$$y = \begin{cases} 0 & \text{altimenti} \\ 1 & \text{soggetto trova lavoro entro 6 mesi} \end{cases}$$

In tali casi il modello normalmente utilizzato, cioè

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$$

non è più valido in quanto ha come assunzione che la variabile risposta sia, almeno in un determinato intervallo, continua.

Nei casi in cui la  $y$  sia una variabile risposta di tipo binario, si pongono due differenti assunzioni:

–  $y_i$  si assume abbia una distribuzione Bernoulliana con parametro  $P_i$ , cioè la probabilità di successo riferita quindi all'eventualità la  $y$  assuma valore 1.

$$y_i \sim (P_i)$$

$P_i = P(y_i = 1) \rightarrow$  la  $y$  deve essere posta pari a uno

– una eventuale trasformazione di  $P_i$  si assume sia uguale al modello di regressione multipla.

$$g(P_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

Normalmente, a differenza di quanto qui mostrato nella formula, la regressione è imposta direttamente al valore atteso.

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

Relativamente alla precedente equazione, si potrebbe avere un qualunque valore, ma nel caso binario che invece si sta esaminando, si deve impostare la funzione chiamata *funzione link* ( $= g(P_i)$ ), cioè si impone una trasformazione lineare sul valore atteso.

$$P_i = g^{-1}(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik})$$

Un esempio di funzione link è la funzione logit:

$$g(P_i) = \log\left(\frac{P_i}{1 - P_i}\right)$$

$$\log\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

La funzione è una probabilità che inoltre proietta l'intera relazione nell'intervallo più ampio che ha come limiti gli infiniti (  $-\infty$  ,  $+\infty$  ).

Prendendo un numero molto vicino allo zero, il rapporto  $P_i / (1 - P_i)$  sarà a sua volta molto vicino allo zero e quindi la funzione logaritmo assumerà valori negativi spostati verso il basso.

Nel caso invece  $P_i$  fosse un numero grande, la funzione logaritmo sarà un numero positivo e molto elevato.

Quando, per esempio,  $P_i < 0.5$  , si ha  $g ( P_i )$  che tende a  $-\infty$ .

Quando invece  $P_i > 0.5$  ,  $g ( P_i )$  tende a  $+\infty$ .

La funzione, quando rappresentata graficamente, è sempre crescente e presenta due asintoti verticali in ascissa zero e in ascissa uno.

La funzione logistica è la funzione inversa della funzione logit

$$g^{-1}(\eta) = \frac{e^\eta}{1 + e^\eta}$$

È una trasformazione logistica e di conseguenza è sempre compresa fra zero e uno come valori. In particolare:

→ con valori di  $\eta$  molto piccoli, si ha un numero molto vicino allo zero;

→ se  $\eta$  è negativo, essendo la funzione basata su un esponenziale, si ottiene un numero positivo;

→ se  $\eta$  è invece un numero molto grande e positivo, si ha un numero che si avvicina molto al valore unitario, ma non uguale;

→ con  $\eta = 0$  la funzione assume valore 0.5.

La trasformazione inversa proietta un qualsiasi numero reale in un intervallo compreso fra zero e uno, cioè la trasformazione rende il numero una probabilità.

### ***Modello di regressione logistica***

$$g ( P_i ) = \text{logit}$$

Si assume in questo caso che

$$E ( y_i ) = P_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}} \in (0,1)$$

Si fa cioè un'assunzione su una sua trasformazione, non si fa direttamente sul valore atteso che in questo caso è sempre compreso fra zero e uno.

Il modello in questo modo è più difficile da stimare rispetto a quello della regressione multipla. Soprattutto per l'interpretazione in quanto si è compiuto un cambiamento di scala.

Per capire l'interpretazione dei parametri si dovrebbe fare la derivata del valore atteso rispetto a una singola covariata che ha il segno uguale al parametro riferito alla stessa.

$$\frac{\partial E(y_i)}{\partial x_{ij}}$$

Visto che la formula riporta lo stesso segno di  $\beta_j$ , si ha che:

→ se  $\beta_j < 0$ , la funzione decresce con tale covariata;

→ se  $\beta_j > 0$ , la funzione cresce con tale covariata.

In tal modo si vede l'influenza positiva o negativa sul valore atteso della covariata.

### Esempio

y = soggetto entro 6 mesi trova un'occupazione lavorativa

$x_1$  = voto di laurea       $\beta_1 = 0.5$  → la probabilità di trovare lavoro dipendente dal voto ottenuto alla laurea fa aumentare positivamente anche la probabilità della y, cioè di trovare lavoro.

$x_2$  = anni impiegati per laurearsi       $\beta_2 = - 0.2$  → minore sarà il tempo che si impiega per laurearsi, minore sarà la probabilità di trovare lavoro entro determinati termini.

### *Massima verosimiglianza*

In questo caso si imposta per andare poi a massimizzare i parametri  $\beta$  contenuti nel vettore

$$\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$$

$$L(\underline{\beta}) = \prod f(y) = \prod P_i^{y_i} (1 - P_i)^{1 - y_i}$$

$$l(\underline{\beta}) = \sum y_i \log(P_i) + (1 - y_i) \log(1 - P_i)$$

$$l(\underline{\beta}) = \sum y_i [\log(P_i) - \log(1 - P_i)] + \log(1 - P_i) = \sum y_i \log\left(\frac{P_i}{1 - P_i}\right) + \log(1 - P_i)$$

$$l(\underline{\beta}) = \sum y_i \underline{X}_i' \underline{\beta} + \log\left(\frac{1}{1 + e^{\underline{X}_i' \underline{\beta}}}\right)$$

In questo modo si ha:

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

$$P_i = \frac{e^{\underline{X}_i' \underline{\beta}}}{1 + e^{\underline{X}_i' \underline{\beta}}}$$

ottenendo quindi

$$1 - P_i = \frac{1 + e^{\underline{X}_i' \underline{\beta}} - e^{\underline{X}_i' \underline{\beta}}}{1 + e^{\underline{X}_i' \underline{\beta}}} = \frac{1}{1 + e^{\underline{X}_i' \underline{\beta}}}$$

$$l(\underline{\beta}) = \sum_i y_i x_i^T \underline{\beta} - \log(1 + e^{x_i^T \underline{\beta}})$$

Per stimare ora  $\underline{\beta}$ , si deve trovare il massimo della funzione

$$\max_{\underline{\beta}} l(\underline{\beta})$$

Si usa quindi l'algoritmo di *Newton-Raphson* con cui a ogni passo si propone una nuova soluzione fino ad arrivare al valore massimo.

$$\underline{\beta}^{(0)} = \underline{0}$$

$$\underline{\beta}^{(1)} = \underline{\beta}^{(0)} -^{-1} S(\underline{\beta}^{(0)})$$

Si vuole trovare una soluzione migliore e quindi un valore più alto della verosimiglianza

$$J_{(k+1) \times (k+1)}(\underline{\beta}^{(0)}) = \frac{\partial^2 l(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}^T}$$

La matrice J così indicata è chiamata *matrice d'informazione* ed è la derivata seconda della verosimiglianza.

$$J(\underline{\beta}) = \begin{pmatrix} \frac{\partial^2 l(\underline{\beta})}{\partial \beta_0} & \frac{\partial^2 l(\underline{\beta})}{\partial \beta_0 \partial \beta_1} & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix}$$

$$S_{(k+1) \times 1}(\underline{\beta}) = \frac{\partial l(\underline{\beta})}{\partial \underline{\beta}} = \begin{pmatrix} \frac{\partial l(\underline{\beta})}{\partial \beta_0} \\ \frac{\partial l(\underline{\beta})}{\partial \beta_1} \\ \dots \end{pmatrix}$$

La funzione così scritta è la derivata prima della precedente. Da qui si ottiene:

$$\underline{\beta}^{(2)} = \underline{\beta}^{(1)} - J(\underline{\beta}^{(1)})^{-1} S(\underline{\beta}^{(1)})$$

$$\underline{\beta}^{t+1} = \underline{\beta}^{(t)} - J(\underline{\beta}^{(t)})^{-1} S(\underline{\beta}^{(t)})$$

Ci si ferma nella ricerca di una migliore soluzione quando la differenza fra la log-verosimiglianza al passo ( t + 1 ) e la log-verosimiglianza al passo precedente t, è minore di un certo valore  $\epsilon$  pari normalmente a  $10^{-6} = 0.000001$ .

$$l(\underline{\beta}^{(t+1)}) - l(\underline{\beta}^{(t)}) < \varepsilon = 10^{-6}$$

I risultato trovato quindi dalla precedente funzione sarà la stima di massima verosimiglianza

$\hat{\beta} = \underline{\beta}^{(t+1)}$  → è il primo numero del vettore per cui non aumenta più il valore della funzione, è cioè la stima di massima verosimiglianza ML.

Una volta ottenuta la stima del  $\beta$ , si ha la possibilità di fare inferenza

$$s.e. = \underset{(k+1) \times (k+1)}{\widehat{Var}}(\hat{\underline{\beta}}) = [-J(\hat{\underline{\beta}})]^{-1}$$

di conseguenza risulta che:

$$s.e.(\hat{\beta}_j) = \sqrt{\text{diag}_j(\widehat{Var}(\hat{\underline{\beta}}))}$$

Nel caso k sia uguale a 2, la matrice di varianza e covarianze stimate risulta essere

$$\underset{3 \times 3}{\widehat{Var}}(\hat{\underline{\beta}}) = \begin{pmatrix} \widehat{Var}(\hat{\beta}_0) & \widehat{cov}(\hat{\beta}_0, \hat{\beta}_1) & \widehat{cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \widehat{cov}(\hat{\beta}_1, \hat{\beta}_0) & \widehat{Var}(\hat{\beta}_1) & \widehat{cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \widehat{cov}(\hat{\beta}_2, \hat{\beta}_0) & \widehat{cov}(\hat{\beta}_2, \hat{\beta}_1) & \widehat{Var}(\hat{\beta}_2) \end{pmatrix}$$

Come intervallo di confidenza si avrà: 
$$\left( \begin{array}{c} \hat{\beta}_j - Z_{\frac{\alpha}{2}} s.e.(\hat{\beta}_j) \\ \hat{\beta}_j + Z_{\frac{\alpha}{2}} s.e.(\hat{\beta}_j) \end{array} \right)$$

La distribuzione usata, come si vede dalla formula, è la normale standard in quanto in questi casi l'uso di una distribuzione t di Student non risulterebbe più efficace.

### ***Verifica delle ipotesi***

In questi casi, per la verifica delle ipotesi, si usa la statistica Z, cioè:

$$Z = \frac{\hat{\beta}_j - \beta_{j0}}{s.e. \hat{\beta}_j}$$

in cui  $\beta_{j0}$  è il valore del parametro sotto ipotesi nulla  $H_0$ .

Nel caso  $H_1: \beta_j \neq \beta_{j0}$ , si rifiuterebbe  $H_0$  nel caso in cui  $|Z| \geq Z_{\frac{\alpha}{2}}$

Si pongono di volta in volta i singoli parametri pari a zero per vedere se hanno una reale influenza sulla probabilità di successo .

Il p-value si usa, come di norma, ponendo come valore limite 0.05:

→ se p-value < 0.05 , si Rifiuta l'ipotesi nulla  $H_0$ ;

→ se il p-value > 0.05 , si Accetta  $H_0$ , la relativa covariata  $x_j$  non ha quindi una influenza significativa ed è quindi eliminabile.

In R si ha come output la seguente tabella di valori:

	Stime	SE	Z-test	Stime	P-value
intercetta	$\beta_0$	$\beta_0$	$\beta_0\beta_0$	$\hat{\beta}_0 = \dots$	...
$x_1$	$\beta_1$	$\beta_1$	$\beta_1\beta_1$	$\hat{\beta}_1 = \dots$	...
$x_2$	$\beta_2$	$\beta_2$	$\beta_2\beta_2$	$\hat{\beta}_2 = \dots$	...

Per vedere la significatività si vedono i valori contenuti nella colonna riferita al p-value tenendo sempre conto del valore limite posto pari a 0.05.