Note di econometria III

(Variabili risposta limitate - introduzione)

Valerio Potì

5 giugno 2020

QUESTE DISPENSE SONO ALLO STATO DI BOZZA

L'AUTORE SAREBBE GRATO CON IL LETTORE PER SUGGERIMENTI O COMMENTI

(Work-in-progress, comments and feed-back very welcome)

Indice

1. Variabili risposta limitate

1.1 Modelli di regressione Probit e Logit per variabili dicotomiche

Finora, pur senza assumerlo direttamenete, abbiamo lavorato con variabili dipndenti essenzialmente continue, ad esempio il PIL, il tasso di inflazione, etc. Ma esistono situazioni, e sono piuttosto comuni, in cui la variabile dipendente assume un numero limitato di valori discreti. Nel seguito, ci occuperemo principalmente del caso in cui la variabile dipendente sia dicotomica (binaria). Inoltre, la variabile dipendente può essere anche di categoria nominale o ordinale, ma non ci occuperemo di questi modelli.

Esempio: Si supponga che si sia interessati a spiegare la decisione che a una domanda di mutuo venga opposto un rifiuto da parte della banca.

Si supponga che il modello utilizzato per descrivere la relazione tra il rifiuto di concedere il mutuo e il rapporto rata/reddito sia il seguente:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

In questo modello,

$y = 1$ se la decisione è di negare il mutuo;

$y = 0$ se la decisione è di *non* negare (concedere);

Si vuole valutare se ci sia una relazione tra il rifiuto di concedere il mutuo ($y = 1$) e la variabile x (rapporto rata/reddito) che si ritiene possa "spiegare" tale decisione. Data la definizione della variabile y, si avrà che

$$E(y_i|x_i) = \pi_i \times 1 + (1 - \pi_i) \times 0 = \pi_i$$

$$Var(y_i|x_i) = \pi_i (1 - \pi_i)$$

dove $\pi_i \equiv P(y_i = 1|x_i) \in [0,1]$ rappresenta, per ciascun individuo i, la probabilità che, dato x_i, $y_i = 1$. Si noti che, come nel caso più generale della distribuzione di una variabile binomiale, la varianza è interamente determinata dal valore atteso.

Per via della condizione di identificazione del modello di regressione, si ha che $E(u_i|x_i) = 0$ e dunque, come al solito,

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$$

Inoltre, come visto prima,

$$E(y_i \mid x_i) = \pi_i$$

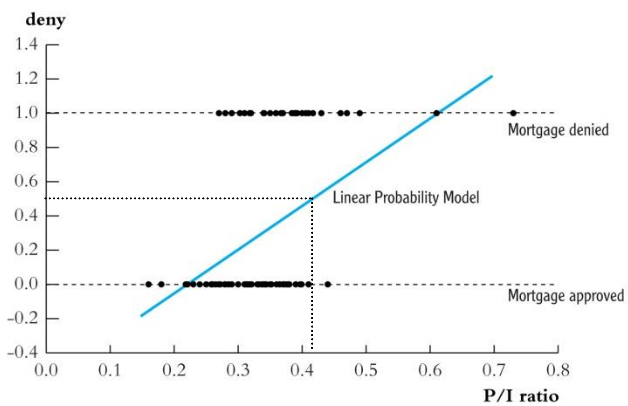
Segue dunque che

$$\pi_i = \beta_0 + \beta_1 x_i$$

e quindi che la probabilità dipende linearmente dai regressori, come mostrato nella figura più in basso. Il grafico in Figura riporta la relazione tra la probabilità che il mutuo venga negato (deny) e il rapporto tra rata del mutuo e reddito (P/I ratio, ovvero un'acronimo per 'payment/income ratio') stimata utilizzando il modello lineare della probabilità appena visto e utilizzando 2380 osservazioni facenti parte del dataset compilato con i dato forniti dalle banche alla Amministrazione Americana in ottemperanza all'Home Mortgage Disclosure Act (HMDA).

Rigetto della domanda di mutuo vs. rapport tra rata del mutuo e reddito (P/I ratio)

Porzione del data-set HMDA



Il modello stimato è il seguente:

$$\widehat{deny}_i = -0.080 + 0.604\ P/I\ ratio$$
$$(0.032)\ (0.098)$$

Per un valore di *P/I ratio* = 0.3, abbiamo che $\widehat{deny}_i$ = -0.080 + 0.604 × 0.3 = 0.151 e per *P/I ratio* = 0.4, abbiamo che $\widehat{deny}_i$ = -0.080 + 0.604 × 0.4 = 0.212. Dunque, un incremento del rapporto rata/reddito del 10% determina un incremento lineare stimato della probabilità di rifiuto di $\frac{0.212 - 0.151}{0.151} = 0.061$, ovvero di circa il 6%. Il problema è che, proprio per la linearità del modello, la variabile dipendente può assumere valori negativi o superiori a uno. Nell'esempio, ciò accade per valori del rapporto rata/reddito inferiori al 20% e superiori a poco più del 60%. Ma che senso hanno probailità come queste?

Più in generale, supponendo che si intenda utilizzare il seguente modello di regressione

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

si avrebbe che

$$E(y_i \mid \boldsymbol{x}_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} = \pi_i$$

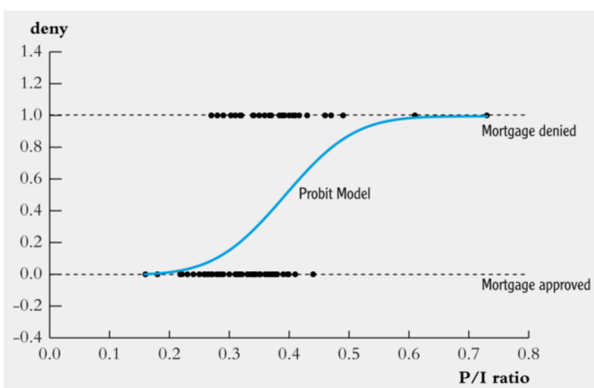
In questo modello, ciascun coefficiente quantifica di quanto cambi la probabilità che y = 1 per una variazione unitaria della corrispondente variabile indipendente. Come nel caso in cui era presente una sola variabile indipendente, la linearità del modello da luogo al problema che, a seconda dei valori assunti dalle variabili indipendenti, non necessariamente

$$\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$$

assume valori compresi tra 0 e 1. Quindi il problema è che la linearità del modello non garantisce che le probabilità previste siano comprese nell'intervallo [0, 1]. Ciò di cui avremmo bisogno è un modello in cui

1. $0 \le \Pr(y_i = 1 \mid x_{k,i}) \le 1$
2. $\Pr(y_i = 1 \mid x_{k,i})$ cresca al crescere di $x_{k,i}$ quando β_k sia positivo (e viceversa)

Ciò richiede una forma funzionale *non-lineare* della probabilità. In particolare, una forma ad '*S*' appare particolarmente adatta. Ad esempio,



Questa è la forma funzionale che caratterizza il **modello Probit**. In tale modello di regressione, la probabilità che y_i = 1 viene specificata utilizzando la funzione cumulativa della distribuzione normale standardizzata valutata a $z_i = \boldsymbol{\beta}'\boldsymbol{x_i} = \beta_0 + \beta_1 x_{1,i} + \cdots + \ + \beta_k x_{k,1}$

$$\Pr(y_i = 1 \mid \boldsymbol{x}_i) = \Phi(z_i) = \Phi(\boldsymbol{\beta}'\boldsymbol{x}_i)$$

e dunque il modello di regressione è

$$y_i = \Phi(\boldsymbol{\beta}'\boldsymbol{x}_i) + u_i$$

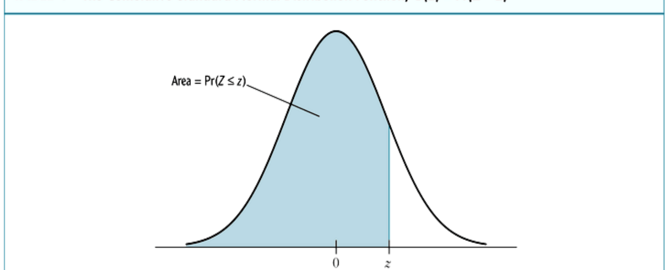
Qui sopra, $\Phi(z_i) \equiv \Pr(Z_i \le z_i)$ è la funzione cumulativa della distribuzione normale standardizzata e $z_i \equiv \boldsymbol{\beta}'\boldsymbol{x}_i$ è il “valore z” or “indice *z*” del modello probit. Tale indice è una sorta di variabile latente che può assumere in linea di principio qualunque valore (reale), la cui probabilità cumulata di assumere un valore pari o inferiore a una certa soglia (*e qui il tutto diventa un pò intricato!*), a sua volta, rappresenta la probabilità che *y* assuma valore pari a 1. Intuitivamente, possiamo trattare z_i come una sorta di indice di propensità che la y_i assuma valore pari a uno.

Esempio: Si supponga che $\beta_0 = -2$, $\beta_1 = 3$, $x_i = 0.4$, cosicchè

$$\Pr(y_i = 1 | x_i = 0.4) = \Phi(-2 + 3\times 0.4) = \Phi(-0.8)$$

$\Pr(y_i = 1|x_i = 0.4)$ = area al di sotto della funzione di densità della distribuzione normale standardizzata alla sinistra di $z = -.8$, ovvero…

TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z)$ = Pr(Z ″ z)



[illegible]	Second Decimal Value of z									
	0	1	2	3	4	5	6	7	8	9
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	[illegible]	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121

…..$\Pr(z \le -0.8) = .2119$

Nel modello Probit, il ‘trucco’ per far si che le probabilità assumano valori compresi nell’intervallo [0, 1] è quello di specificare la probabilità come una determinata trasformazione (non lineare) del modello lineare considerato inizialmente, che diventa il modello di una variabile latente che rappresenta un indice da cui dipende la probabilità che la variabile dipendente assuma valore unitario. La trasformazione ha la proprietà peculiare di assumere valori compresi nell'intervallo desiderato e di crescere al crescere dell'argomento, così da risultare crescente nelle variabili esplicative dell'indice quando quest'ultimo dipenda positivamente dalle variabili stesse. Nell'esempio dei mutui considerato prima, la variabile latente da csui dipende la probabilità secondo la trasformazione non lineare utilizzata può essere vista come una sorta di 'indice di rifiutabilità' del mutuo a fronte della richiesta da parte dell'individuo *i*-esimo.

Tutti i modelli per variabili risposta limitate utilizzano grosso modo un 'trucco' di questo tipo. Un'altro esempio è il modello Logit, che utilizza una trasformazione logistica dell'indice $z_i = \beta ' \boldsymbol{x}_i$ ovvero

$$\Pr(y_i = 1 | \boldsymbol{x}_i) = F(z_i) = F(\beta ' \boldsymbol{x}_i)$$

dove F è la funzione cumulativa della distribuzione logistica

$$F(\beta ' \boldsymbol{x}_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Esempio: Si supponga che β_0 = -3, β_1= 2, x_i = 0.4, cosicchè

$$\Pr(y_i = 1|x_i = 0.4) = F(-3 + 2\times 0.4) = F(-2.2) = 1/(1+e^{-(-2.2)}) = 0.0998$$

Dunque, cosi si vede, Logit e Probit sono molto simili.

1.2 Interpretazione dei parametri

In un modello lineare, i parametri quantificano l'effetto atteso sulla variabile dipendente di una variazione marginale della variabile esplicativa cui si riferiscono. Ad esempio, nel modello lineare, β_1 quantifica l'effetto marginale di $x_{1,i}$ su $E(y_i|x_{1,i}, \dots, x_{k,i})$. Ovvero,

$$\beta_j = \frac{\partial E(y_i|x_{1,i}, \dots, x_{k,i})}{\partial x_{ij}} = \frac{\partial Pr(y_i = 1|x_{1,i}, \dots, x_{k,i})}{\partial x_{ij}}$$

Nei modelli di probabilità non lineare, invece, abbiamo che

$$\frac{\partial E(y_i|x_{1,i}, \dots, x_{k,i})}{\partial x_{ji}} = \frac{\partial Pr(y_i = 1|x_{1,i}, \dots, x_{k,i})}{\partial x_{ji}} = \frac{\partial \Gamma(z_i)}{\partial x_{ji}} = \frac{\partial \Gamma(z_i)}{\partial z_i} \times \frac{\partial z_i}{\partial x_{ji}}$$

$$= \frac{\partial \Gamma(z_i)}{\partial z_i} \times \frac{\partial(\beta_0 + \beta_1 x_{1,i} + \dots + +\beta_k x_{k,i})}{\partial x_{ji}} = \frac{\partial \Gamma(z_i)}{\partial z_i} \times \beta_j$$

Nell'espressione di cui sopra, $\frac{\partial \Gamma(z_i)}{\partial z_i}$ non è altro che la funzione di densità corrispondente alla funzione di densità cumulata $\Gamma(Z)$, valuata a $Z = z_i$.

Ad esempio, nel caso molto semplice esampinato sopra in cui $Pr(y_i = 1|x_i) = \Phi(-2 + 3x_i)$, abbiamo che

$$\frac{\partial E(y_i|x_{1i}, \dots, x_{ki})}{\partial x_{ji}} = \frac{\partial \Phi(z_i)}{\partial z_i}\frac{\partial z_i}{\partial x_{1i}} = \frac{\partial \Phi(-2 + 3x_i)}{\partial z_i}\beta_1 = \phi(-2 + 3x_i)\beta_1 = 3\phi(-2 + 3x_i)$$

Qui sopra, ϕ denota la derivata prima di Φ, e dunque la funzione di densità di una variabile normale standardizzata. Si nota subito che tale derivata, come possiamo aspettarci, non è costante poiché dipende da $\phi(-2 + 3x_i)$, che dipende dal valore assunto da x_i per l'individuo i-esimo. Dunque, per un individuo il cui rapporto rata reddito sia pari a 0,3, si ha che $x_i = 0{,}4$ e quindi

$$\frac{\partial E(y_i|x_{1i})}{\partial x_{1i}} = 3\phi(-2 + 3 \times 0{,}4) = 3\phi(-2 + 3 \times 0{,}4) = 3\phi(-0{,}8) = 3 \times 0{,}29 = 0{,}87$$

poiché la funzione di densità di una variabile normale standardizzata $\phi(Z)$ valorizzata a $Z = 0{,}4$ è uguale (approssimativamante) a $0{,}29$.

Invece, per un individuo il rapporto rata reddito sia pari a 0,3, si ha che $x_i = 0{,}3$ e quindi

$$\frac{\partial E(y_i|x_{1i})}{\partial x_{1i}} = 3\phi(-2 + 3 \times 0{,}3) = 3\phi(-2 + 3 \times 0{,}3) = 3\phi(-1{,}1) = 3 \times 0{,}22 = 0{,}65$$

Dunque, per il primo e il secondo individuo un piccolo aumento del rapporto tra rata e reddito comporta un aumento proporzionale della probabilità di rifiuto secondo un fattore di proporzionalità rispettivamente dello 0,87 e dello 0,65. Ad esempio, un aumento del rapporto tra rata e reddito dell'1% comporta un aumento dello 0,87% della probabilità di rifiuto per il primo individuo e dello 0,65% per il secondo individuo.

1.3 Stima dei parametri e inferenza (si veda il libro di testo per una trattazione più approfondita)

Ma come vengono stimati i parametri? In campioni ampi, possiamo usare sia una versione dello stimatore OLS che si applica al caso di modelli di regressione non lineari, quali sono i nostri Logit e probit, che (più agevolmente) utililizzare lo stimatore di massima verosimiglianza (Maxiumm Likelihood, ovvero in breve ML). Quest'ultimo stimatore è anche, almeno in teoria, più efficente e comunque entrambi sono consistenti (sempre in campioni ampi). Per dettagli sulle procedure di stima si rimanda all'appendice e al libro di testo. Questi stimatori, ovviamente, forniscono anche errori standard e altre stime necessarie per costruire le statistiche di numerosi test, sia per verificare la significatività di singoli parametri che per sottoporre a verifica ipotesi multiple.

Si tenga presente, comunque, che (ovviamente) sia l'R^2 che la sua versione aggiustata non hanno molto senso in questo contesto. Vengono usate, al loro posto, due altre misure:

1. la frazione predetta correttamente, ovvero la frazione delle $y_i = 1$ ($y_i = 0$) per cui la probabilità predetta è maggiore di 0.5 (minore di 0.5).
2. Lo pseudo-R^2 dato dal miglioramento nel valore della (log-)verosimiglianza relativamente al modello che non contiene alcuna delle variabili esplicative eccetto l'intercetta, e dunque rappresenta una misura di bontà del modello basata sulla funzione di verosimiglianza (e coincide con l'R^2 tradizionale in un modello lineare con errori normalmente distribuiti).

Appendice - Maximum Likelihood (Massima Verosimiglianza)

Abbiamo, con *n* osservazioni e per *i* = {1, 2 ..., *n*},

$$y_i = \begin{cases} 1 \; con \; probabilit\grave{a} \; P(y_i = 1|\boldsymbol{x_i}) \\ 0 \; con \; probabilit\grave{a} \; 1 - P(y_i = 1|\boldsymbol{x_i}) \end{cases}$$

dove

$$P(y_i = 1|\boldsymbol{x_i}) := \varphi(z_i) = \varphi(\boldsymbol{\beta}'\boldsymbol{x_i})$$

cosicchè

$$\begin{aligned} E(y_i|\boldsymbol{x_i}) &= P(y_i = 1|\boldsymbol{x_i}) \times 1 + P(y_i = 0|\boldsymbol{x_i}) \times 0 \\ &= P(y_i = 1|\boldsymbol{x_i}) \\ &= \varphi(z_i) \\ &= \varphi(\boldsymbol{\beta}'\boldsymbol{x_i}) \end{aligned}$$

poichè, come visto prima, $z_i = \boldsymbol{\beta}'\boldsymbol{x_i}$.

Risulta inoltre che

$$\begin{aligned} y_i &= E(y_i|\boldsymbol{x_i}) + u_i \\ &= P(y_i = 1|\boldsymbol{x_i}) \times 1 + u_i \end{aligned}$$

ove gli u_i sono indipendentemente distribuiti e $E(u_i|\boldsymbol{x_i}) = 0$. Quindi, per costruzione, abbiamo

$$\begin{aligned} u_i &= y_i - E(y_i|\boldsymbol{x_i}) \\ &= y_i - P(y_i = 1|\boldsymbol{x_i}) \end{aligned}$$

$$= \begin{cases} 1 - P(y_i = 1|\boldsymbol{x_i}) \; con \; probabilit\grave{a} \; P(y_i = 1|\boldsymbol{x_i}) \\ 0 - P(y_i = 1|\boldsymbol{x_i}) \; con \; probabilit\grave{a} \; P(y_i = 0|\boldsymbol{x_i}) = 1 - P(y_i = 1|\boldsymbol{x_i}) \end{cases}$$

Dunque, la verosimiglianza, ovvero la likelihood, di $\boldsymbol{\beta}$ è

$$\begin{aligned} L(\boldsymbol{\beta}) &= P\{U_1 = u_1, U_2 = u_2, \dots, U_n = u_n\} \\ &= \prod_{i=1}^{n-s} P(y_i = 1|\boldsymbol{x_i}) \prod_{i=1}^{s} P(y_i = 0|\boldsymbol{x_i}) \\ &= \prod_{i=1}^{n-s} P(y_i = 1|\boldsymbol{x_i}) \prod_{i=1}^{s} [1 - P(y_i = 1|\boldsymbol{x_i})] \\ &= \prod_{i=1}^{n-s} \varphi(\boldsymbol{\beta}'\boldsymbol{x_i}) \prod_{i=1}^{s} [1 - \varphi(\boldsymbol{\beta}'\boldsymbol{x_i})] \end{aligned}$$

dove *s* denota il numero di osservazioni per cui $y_i = 1$.