

# LA REGRESSIONE LOGISTICA NELLA RICERCA CLINICA

**Fabio Provenzano, Graziella D'Arrigo, Carmine Zoccali, Giovanni Tripepi**

CNR-IBIM, Unità di Ricerca di Epidemiologia Clinica e Fisiopatologia delle Malattie Renali e dell'Ipertensione Arteriosa, Reggio Calabria

## INTRODUZIONE

La regressione lineare e la regressione logistica sono due metodi statistici utilizzati negli studi eziologici ed in quelli prognostici. In un precedente articolo di questa serie (1), abbiamo descritto la regressione lineare, che viene utilizzata per modellare il rapporto tra una variabile dipendente continua ed una o più variabili indipendenti (sia continue che categoriche). Quando l'indicatore di risultato è dicotomico, per esempio essere affetto/non affetto da una specifica malattia, non è possibile utilizzare la regressione lineare. In questo caso, la regressione logistica permette di analizzare il rapporto tra la variabile dicotomica ed una o più variabili esplicative (sia continue che categoriche). In questo articolo, descriveremo la regressione logistica semplice e multipla, attraverso esempi specifici.

## KEY WORDS:

Odds,  
Odds ratio,  
Logistic  
regression,  
Clinical research

## PAROLE CHIAVE:

Odds,  
Odds ratio,  
Regressione  
logistica,  
Ricerca clinica

## REGRESSIONE LOGISTICA SEMPLICE

In un recente studio che includeva 500 ipertesi essenziali, gli Autori hanno analizzato il rapporto tra disfunzione endoteliale e malattia renale cronica (filtrato glomerulare  $<60$  mL/min/1.73 m<sup>2</sup>) (2). Al momento dell'arruolamento, i pazienti sono stati sottoposti alla misurazione della funzione endoteliale (definita sulla base della risposta vasodilatatoria all'infusione di acetilcolina nell'avambraccio) e della funzione renale. Per analizzare l'associazione disfunzione endoteliale-malattia renale, i pazienti sono stati suddivisi in due gruppi sulla base della risposta vasodilatatoria all'infusione di acetilcolina ( $<400\%$ : disfunzione endoteliale;  $\geq 400\%$ : normale funzione endoteliale) e alla presenza/assenza di malattia renale cronica (filtrato glomerulare  $<60$  mL/min/1.73m<sup>2</sup>). La Tabella I descrive il rapporto tra la proporzione di individui con malattia renale cronica ed alterata funzione endoteliale. La proporzione di individui con malattia renale cronica era circa tre volte più alta (17.0%) nei pazienti con disfunzione endoteliale rispetto a quella dei pazienti con normale funzione endoteliale (6.3%) (Tab. I).

**TABELLA I - RAPPORTO TRA ALTERATA FUNZIONE ENDOTELIALE E PROPORZIONE DI INDIVIDUI CON MALATTIA RENALE CRONICA**

Funzione endoteliale	Proporzione di individui con malattia renale cronica (p)	Odds di malattia renale cronica [p/(1-p)]	Logit (ln odds)
$\geq 400$ (n=111) (funzione endoteliale normale)	0.063	0.067	-2.70
$< 400$ (n=389) (disfunzione endoteliale)	0.170	0.205	-1.58

Il concetto di *odds* è descritto in un precedente articolo di questa serie (3). Una più alta risposta vasodilatatoria denota una funzione endoteliale migliore.

## ODDS, ODDS RATIO E LOGIT

Gli *odds* di malattia renale cronica (Tab. I) sono calcolati con la formula:

$$\text{odds} = [p/(1-p)]$$

Per il gruppo di individui con alterata funzione endoteliale si ha:

$$\text{odds} = 0.170/(1-0.170)=0.205$$

mentre, per il gruppo con normale funzione endoteliale, il calcolo dell'*odds* è:

$$\text{odds} = 0.063/(1-0.063)=0.067$$

Pertanto, l'*odds ratio* (OR) della malattia renale cronica tra i pazienti con e senza disfunzione endoteliale è calcolato come rapporto tra due *odds*:

$$\text{OR} = 0.205/0.067=3.06$$

Il passo successivo è quello di trasformare in *logit* gli *odds* della malattia renale cronica (ultima colonna della Tab. I). Il *logit* si calcola come logaritmo naturale ( $\ln$ ) degli *odds*:

$$\text{logit} = \ln[p/(1-p)]$$

Nella formula precedente  $p$  è la proporzione di individui con malattia renale cronica in ogni categoria di funzione endoteliale. La trasformazione in *logit* degli *odds* nel gruppo con alterata funzione endoteliale è:

$$\text{logit } y = \ln(0.205)=-1.58$$

Come nella regressione lineare, anche nella regressione logistica il rapporto tra variabile dipendente ed indipendente è descritto da un'equazione:

$$\text{logit } y = \beta_0 + \beta_1 x_1$$

Per poter interpretare questa equazione, entrambi i membri devono essere elevati ad esponente "e" ( $e=2.7183$ ).  $\beta_0$  è l'intercetta, cioè il valore del logaritmo naturale degli *odds* della malattia renale cronica, quando la funzione endoteliale è zero.  $\beta_1$  è il coefficiente di regressione (1.118, Tab. II), cioè indica di quanto aumenta in media il logaritmo naturale dell'*odds* della malattia renale cronica nei pazienti con disfunzione endoteliale rispetto a quelli senza questa alterazione. Nella regressione logistica, i coefficienti di regressione sono calcolati con il metodo della massima verosimiglianza (4). Per stimare l'incremento dell'*odds* di malattia renale cronica nei pazienti con disfunzione endoteliale rispetto a quelli senza, dobbiamo calcolare l'esponentiale del coefficiente di regressione. L'*odds ratio* sarà uguale a:

$$\text{OR} = 2.7183^{1.118}=3.06$$

Lo studio documenta che i pazienti con disfunzione endoteliale hanno un

**TABELLA II - REGRESSIONE LOGISTICA SEMPLICE**  
VARIABILE DIPENDENTE: MALATTIA RENALE CRONICA

	Unità di aumento	Coefficiente di regressione ( $\beta_j$ )	Odds ratio (intervallo di confidenza al 95%)	P
Funzione endoteliale	0 ( $\geq 400\%$ )	1.118	1.00 (gruppo di riferimento)	0.007
	1 ( $< 400\%$ )		3.06 (1.35-6.82)	
Intercetta ( $\beta_0$ )=-2.70				<0.001

odds ratio di malattia renale cronica che è circa tre volte più alto rispetto a quello dei pazienti con normale funzione endoteliale. Tale eccesso di rischio è significativo, in quanto l'intervallo di confidenza al 95% non include l'unità (Tab. II).

### REGRESSIONE LOGISTICA MULTIPLA

Quando abbiamo più variabili, l'equazione della regressione logistica si presenta nella forma generale:

$$\text{logit } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n$$

In questo studio, una domanda importante che i ricercatori si sono posti è se il rapporto tra funzione endoteliale e malattia renale cronica sia reale o se sia dovuto all'effetto confondente dell'età. L'età può essere considerata un potenziale confonditore, in quanto essa è associata sia al rischio di malattia renale cronica (la variabile di risultato) che all'esposizione (la disfunzione endoteliale). Inoltre, l'età non è un effetto dell'esposizione, perché non vi sono evidenze che essa si trovi nella catena causale tra esposizione e malattia renale cronica. Per escludere un eventuale confondimento dipendente dall'età, i ricercatori hanno introdotto l'età nel modello della regressione logistica multipla (Tab. III).

**TABELLA III - REGRESSIONE LOGISTICA MULTIPLA**  
VARIABILE DIPENDENTE: MALATTIA RENALE CRONICA

	Unità di aumento	Coefficiente di regressione ( $\beta_j$ )	Odds ratio (intervallo di confidenza al 95%)	P
Funzione endoteliale	0 ( $\geq 400\%$ )	0.974	1.00 (gruppo di riferimento)	0.002
	1 ( $< 400\%$ )		2.64 (1.17-6.00)	
Età	1 anno	0.047	1.05 (1.02-1.07)	<0.001
Intercetta ( $\beta_0$ )=-4.91				<0.001

L'aggiustamento per l'effetto confondente dell'età influenza il rapporto tra disfunzione endoteliale e rischio di malattia renale cronica, in quanto l'odds ratio aggiustato (2.64) è minore (-14%) di quello crudo (3.06). Tuttavia, nonostante l'aggiustamento per età, il rapporto tra alterata funzione endoteliale e malattia renale cronica rimane statisticamente

significativo (Tab. III) e, quindi, è lecito concludere che la disfunzione endoteliale è, di per sé, un fattore di rischio per la progressione del danno renale nell'ipertensione.

## **NUMERO DI COVARIATE NELL'ANALISI DI REGRESSIONE LOGISTICA MULTIPLA**

Il massimo numero di covariate che può essere inserito in un modello di regressione logistica multipla è strettamente dipendente dal numero di eventi e/o di casi di malattia. Una semplice regola è quella di includere nel modello di regressione logistica multipla 1 covariata ogni 10 eventi (5). Per esempio, se abbiamo un campione di 1000 individui e 20 casi di malattia, il massimo numero di covariate da includere nel modello logistico multiplo dovrebbe essere pari a 2.

## **CONCLUSIONE**

La regressione logistica è un importante strumento statistico per valutare il rapporto tra fattori di rischio ed esiti clinici (morte, presenza/assenza di malattia, ecc.), di tipo binario (dicotomici), e per controllare l'effetto confondente delle variabili associate ai fattori di rischio e agli esiti clinici in questione.

## **DICHIARAZIONE DI CONFLITTO DI INTERESSI**

Gli Autori dichiarano di non avere conflitto di interessi.

## **Indirizzo degli Autori:**

Dr. Giovanni Tripepi  
CNR-IBIM, Istituto di Biomedicina  
Epidemiologia Clinica e Fisiopatologia  
delle Malattie Renali e dell'Ipertensione Arteriosa  
Via Vallone Petrarà 55/57  
89124 Reggio Calabria  
e-mail: gtripepi@ibim.cnr.it

---

## **BIBLIOGRAFIA**

1. Provenzano F, Zoccali C, Tripepi G. Regressione lineare nella ricerca clinica. [Linear regression analysis in clinical research.] *G Ital Nefrol* 2011; 28 (1): 80-4.
2. Perticone F, Maio R, Tripepi G, Zoccali C. Endothelial dysfunction and mild renal insufficiency in essential hypertension. *Circulation* 2004; 110 (7): 821-5.
3. Provenzano F, Tripepi G, Zoccali C. [How to measure effects in clinical research (corrected)]. *G Ital Nefrol* 2010; 27 (3): 296-300. Erratum in: *G Ital Nefrol* 2010; 27 (4): 422.
4. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied regression analysis and other multivariable methods*. Duxbury Press, Pacific Grove 1998; 639-55.
5. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49 (12): 1373-9.