

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Metodi Statistici per la Ricerca Sociale

Modelli per variabili risposta categoriche

1. La seguente tabella mostra i risultati di un modello logistico, in cui la variabile risposta Y rappresenta l'uso di marijuana (1 = Uso di marijuana). Le variabili esplicative sono il consumo di alcool e di sigarette.

Variabile	Stima	SE
Costante	-5.309	0.4752
Alcool (No)		
Si	1.094	0.4647
Sigarette (No)		
Si	1.047	0.1638

- (a) Scrivere l'equazione che definisce i valori stimati delle probabilità: $\hat{\pi}_i$
(b) Interpretare i coefficienti in termini di odds.

Soluzione:

- (a) Valori stimati delle probabilità

$$\hat{\pi}_i = \frac{\exp\{-5.309 + 1.094 \cdot \text{Alcool}_i + 1.047 \cdot \text{Sigarette}_i\}}{1 + \exp\{-5.309 + 1.094 \cdot \text{Alcool}_i + 1.047 \cdot \text{Sigarette}_i\}}$$

- (b) $e^{1.094} = 2.986$ è rapporto delle quote parziale tra uso di marijuana e consumo di alcool (controllando per il consumo di sigarette). Gli odds a favore di uso di marijuana sono circa il triplo se un soggetto consuma alcool sia se il soggetto fumi sigarette che non fumi sigarette. $e^{1.094} = 2.849$ è rapporto delle quote parziale tra uso di marijuana e consumo di sigarette (controllando per il consumo di alcool). Gli odds a favore di uso di marijuana sono circa il triplo se un soggetto fuma sigarette sia se il soggetto consumi alcool che non consumi alcool.

2. Di seguito si riportano i risultati di un'analisi in cui un modello di regressione logistica è stato stimato su un campione di 54 uomini anziani. La variabile risposta è una variabile binaria che assume valore 1 per soggetti che presentano sintomi di senilità. Come variabile esplicativa si considera il punteggio ottenuto a un particolare test (il punteggio WAIS = Wechsler Adult Intelligence Scale). Valori maggiori della variabile WAIS indicano capacità intellettive migliori.

Variabile	Coefficiente	SE	p -value
Costante	3.3728	1.6239	0.0378
WAIS	-0.4292	0.1621	0.0081

- (a) Scrivere l'equazione che definisce i valori stimati delle probabilità: $\hat{\pi}_i$
(b) Interpretare il p -value relativo al coefficiente della variabile esplicativa WAIS.
(c) Interpretare il coefficiente relativo alla variabile WAIS.

Soluzione:

(a) Valori stimati delle probabilità

$$\hat{\pi}_i = \frac{\exp\{3.3728 - 0.4292 \cdot \text{WAIS}_i\}}{1 + \exp\{3.3728 - 0.4292 \cdot \text{WAIS}_i\}}$$

(b) Il p -value relativo al coefficiente della variabile esplicativa WAIS è piccolo (inferiore a 0.01). Quindi si ha forte evidenza che il coefficiente sia diverso da zero (l'ipotesi nulla $H_0 : \beta_1 = 0$ sarebbe rifiutata a ogni livello $\alpha > 0.0081$).

(c) Interpretazione in termini di odds: $e^{-0.4292} = 0.6510$. Per ogni incremento unitario della variabile WAIS, gli odds a favore della presenza di sintomi di senilità sono moltiplicati per 0.6510 (ossia si riducono del 35% circa).

3. Si consideri il problema dell'abbandono degli studi universitari entro la fine del primo anno di università. Sia Y una variabile binaria con $Y_i = 1$ se uno studente i abbandona gli studi e $Y_i = 0$ se uno studente i non abbandona gli studi. Specificare un modello di regressione logistica per lo studio dell'associazione tra la scelta di abbandonare gli studi e le seguenti variabili esplicative: $x_1 =$ voto di maturità, $x_2 =$ tipo di scuola (scientifica, tecnica, biomedica, socio-economica, umanistica), $x_3 =$ variabile binaria che indica se lo studente riceve una borsa di studio o meno. (Non considerare interazioni tra variabili esplicative e considerare come livello di riferimento del tipo di scuola la categoria "umanistica").

Soluzione: Si indichi con $\pi_i = Pr(Y_i = 1)$.

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}^S + \beta_3 x_{i2}^T + \beta_4 x_{i2}^B + \beta_5 x_{i2}^{SE} + \beta_6 x_{i3}$$

dove

$$x_{i2}^S = \begin{cases} 1 & \text{Tipo di scuola} = \text{Scientifica} \\ 0 & \text{Altrimenti} \end{cases} \quad x_{i2}^T = \begin{cases} 1 & \text{Tipo di scuola} = \text{Tecnica} \\ 0 & \text{Altrimenti} \end{cases}$$

$$x_{i2}^B = \begin{cases} 1 & \text{Tipo di scuola} = \text{Biomedica} \\ 0 & \text{Altrimenti} \end{cases} \quad x_{i2}^{SE} = \begin{cases} 1 & \text{Tipo di scuola} = \text{Socio-economica} \\ 0 & \text{Altrimenti} \end{cases}$$

4. La seguente tabella mostra i risultati di un modello logistico utilizzato per stimare la probabilità di effettuare acquisti on-line, in funzione di una serie di covariate (reddito, età, genere, numero di componenti componenti il nucleo familiare, presenza di una connessione internet domestica, livello di istruzione).

Acquisti on-line	OR	p-value
Reddito	1.000	0.658
Età	0.908	0.000
Genere (Riferimento: Uomini)		
Donne	0.828	0.068
N. componenti (Riferimento: 1)		
2	0.854	0.560
3 o +	0.744	0.003
Connessione (Riferimento: No)		
Si	3.543	0
Istruzione (Riferimento: Bassa)		
Media	4.882	0.031
Alta	5.938	0.002

- (a) Commentare i risultati della stima, tenendo conto anche del p -value.
- (b) Se si volesse esprimere i risultati come coefficienti, cosa si dovrebbe fare? Riportare i valori dei coefficienti

Soluzione:

- (a) Esempi. ‘Età’ – $OR = 0.908$: Controllando le altre variabili, gli odds a favore dell’acquisti on-line vengono moltiplicati per 0.908 per ogni incremento di un anno dell’età, ossia un incremento di un anno nell’età comporta una riduzione negli odds a favore dell’acquisti on-line del 9.2%. ‘Istruzione media’ – $OR = 4.882$: Controllando le altre variabili, avere un livello di istruzione media rispetto a un livello di istruzione bassa moltiplica gli odds a favore dell’acquisti on-line per 4.882, ossia gli odds a favore dell’acquisti on-line di coloro che hanno un livello di istruzione media sono circa 5 volte gli odds a favore dell’acquisti on-line di coloro che hanno un livello di istruzione bassa.
- (b) La stima dei coefficienti è il logaritmo del rapporto degli odds:

Acquisti on-line	Coefficiente
Reddito	0.000
Età	-0.097
Genere (Riferimento: Uomini)	
Donne	-0.189
N. componenti (Riferimento: 1)	
2	-0.158
3 o +	-0.296
Connessione (Riferimento: No)	
Si	1.265
Istruzione (Riferimento: Bassa)	
Media	1.586
Alta	1.781

Esercizi dal libro di testo: Capitolo 15

15.1 15.2 15.3 15.5 15.7 15.8 15.9