Facoltà di Scienze della Formazione
Corso di laurea in Scienze dell'Educazione
Anno Accademico 2000/2001
Statistica Sociale A
Maria Cristina Migliore

SCHEDA SINTETICA DI PRESENTAZIONE DELLA REGRESSIONE BIVARIATA

L'esistenza e la forza della relazione tra due variabili quantitative si verifica e si misura con il coefficiente di correlazione di Bravais-Pearson. Si ricordi che tale coefficiente è in grado di cogliere solo le relazioni lineari (Migliore 2007).
Il coefficiente di regressione quantifica l'effetto di una variabile su un'altra. In questa scheda si tratta solo la relazione tra due variabili (regressione bivariata), nel caso di relazione approssimativamente lineare. Si distingue la variabile dipendente Y da quella indipendente X[1]. Con il coefficiente di regressione *b* si misura l'effetto di X su Y.

Si supponga di aver condotto una piccola indagine per verificare la relazione tra il grado di volizione e il risultato scolastico in un gruppo di 11 studenti. I dati raccolti sono i seguenti[2]:

Studente	*Volizione*	*Risultato scolastico*
	x_i	y_i
A	9	5
B	13	6
C	15	8
D	15	7
E	14	6,5
F	16	7
G	16	8,5
H	8	4
I	11	6
L	12	5,5
M	7	6

Si intende misurare l'effetto della volizione (X) sul risultato scolastico (Y).
Dall'analisi della correlazione si conosce che la relazione tra le due variabili è lineare e che è approssimata da una retta[3]. In altri termini tale retta rappresenta la relazione tra le due variabili. E' sufficiente conoscere i parametri che definiscono tale retta per conoscere l'effetto della X sulla Y.
La retta è definita dalla seguente funzione:

$$Y_p = a + bX_i$$

dove Y_p sono i nuovi valori di Y nel caso di relazione perfetta tra Y e X, ovvero quando i punti del diagramma (vedi nota 3) sono disposti sulla retta e non intorno alla retta. I parametri *a* e *b*

[1] L'analisi di regressione multipla prende in considerazione due o più variabili indipendenti.
[2] Tratti da Migliore, M. C. (2001). Analisi dei dati per la ricerca socio-educativa. Un'introduzione. Torino, Thélème.

[3] E' sufficiente costruire con i dati forniti un diagramma a dispersione del tipo riportato in Coggi&Calonghi a pag. 200, Fig. 41. Si ponga la variabile X sull'asse delle ascisse (orizzontale) e la variabile Y su quello delle ordinate (verticale). Si provi poi a tracciare a mano libera una retta tra la nuvola di punti, in modo tale che approssimi il più possibile i punti del diagramma.

definiscono rispettivamente l'altezza e la pendenza della retta e sono denominati intercetta (o costante) e coefficiente di regressione (o gradiente o Beta). Il parametro *b* è la misura dell'effetto di X su Y.

Per calcolare *b* si utilizza la seguente formula[4]:

$$b = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

Una volta ottenuto *b* è agevole l'individuazione di *a*:

$$a = \overline{Y} - b\,\overline{X}$$

Nel caso dei dati riportati i calcoli si svolgono come riportato di seguito[5]:

Studente	*Volizione*	*Risultato scolastico*	*Scarti di x_i*	*Scarti di y_i*	*Scarti di x_i al quadrato*	*Prodotti degli scarti*
	x_i	y_i	$x_i - x$	$y_i - y$	$(x_i - x)^2$	$(x_i - x)(y_i - y)$
	P	*Q*	*R*	*S*	*T*	*V*
A	9	5	-3,36	-1,32	11,29	4,44
B	13	6	0,64	-0,32	0,41	-0,2
C	15	8	2,64	1,68	6,97	4,44
D	15	7	2,64	0,68	6,97	1,8
E	14	6,5	1,64	0,18	2,69	0,3
F	16	7	3,64	0,68	13,25	2,48
G	16	8,5	3,64	2,18	13,25	7,94
H	8	4	-4,36	-2,32	19,01	10,12
I	11	6	-1,36	-0,32	1,85	0,44
L	12	5,5	-0,36	-0,82	0,13	0,3
M	7	6	-5,36	-0,32	28,73	1,72
Medie	*12,36*	*6,32*		*Somme*	*104,55*	*33,78*

La somma della colonna V (33,78) rappresenta il numeratore della formula per calcolare *b*. La somma della colonna T costituisce il denominatore della medesima formula. Si risolva il rapporto. Si ottiene che *b* è uguale a 0,32.
Si calcoli ora il parametro *a*. Esso risulta uguale a 2,36.

---

[4] Si tratta di una formula che permette di stimare una retta la più approssimata possibile ai punti del diagramma a dispersione. Essa si basa sul cosiddetto *metodo dei minimi quadrati*.
[5] Per un problema di gestione del foglio elettronico è "saltata" la barra sulle *x* e le *y* senza pedice *i* nelle colonne R, S, T e V. Si considerino tali *x* e *y* simboli di medie.

Il coefficiente di regressione è da interpretarsi nel modo seguente: per ogni incremento di una unità di X, Y aumenta di 0,32 punti. In altre parole per ogni punto in più di volizione, si ottiene 0,32 punti in più come risultato scolastico. Dunque la volizione ha un effetto sull'apprendimento pari a 0,32.

Conoscendo questo valore si può stimare e prevedere per qualsiasi livello di volizione il risultato scolastico. Si supponga che uno studente abbia mostrato un grado di volizione pari a 17. A parità di altre condizioni ci si può attendere un risultato scolastico pari circa a:

$$Y_p = a + bX_i$$

$$Y_p = 2{,}36 + (0{,}32)(17)$$

$$Y_p = 7{,}8$$

Lo studio dell'effetto di una variabile su un'altra richiede una indicazione della capacità complessiva della variabile X di spiegare la variabilità della variabile Y. A questo scopo di calcola il coefficiente di determinazione R-quadro (R^2). Non vi è spazio per illustrare adeguatamente il significato di tale coefficiente. Esso è pari al quadrato del coefficiente di correlazione. Nel caso dell'esempio riportato il coefficiente di correlazione è uguale a 0,809. Il quadrato di tale valore è 0,654. Moltiplicato per 100 diventa 65,4. Esso si interpreta dicendo che la volizione spiega il 65,4% della variabilità dei risultati scolastici.

L'esempio riportato è stato costruito a scopi didattici e la presentazione è a scopo introduttivo[6].

Bibliografia citata

Allison, D. P. (1999). Multiple Regression. A primer. Thousand Oaks, California; London; New Delhi, Pine Forge Press.

Migliore, M. C. (2001). Analisi dei dati per la ricerca socio-educativa. Un'introduzione. Torino, Thélème.

Migliore, M. C. (2007). L'indagine statistica in campo sociale. Variabili e indicatori. Milano, Angeli.

[6] Per successivi approfondimenti e una più completa discussione dell'analisi di regressione si consiglia Allison, D. P. (1999). Multiple Regression. A primer. Thousand Oaks, California; London; New Delhi, Pine Forge Press.