

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI RELAZIONI TRA VARIABILI
18/05/2006

Avvertenza: Fornire le formule utilizzate e tutti i passaggi dei calcoli eseguiti.
Utilizzare almeno 2 cifre decimali.

Esercizio 1:

I 12 impiegati di una piccola azienda sono stati classificati rispetto al mezzo di trasporto (X) ed al tempo (Y) (espresso in minuti) impiegato per raggiungere il luogo di lavoro, ottenendo i seguenti risultati:

Impiegato	Mezzo	Tempo	Impiegato	Mezzo	Tempo
1	piedi	10	7	auto privata	40
2	auto privata	55	8	piedi	10
3	auto privata	30	9	mezzo pubblico	35
4	mezzo pubblico	45	10	mezzo pubblico	30
5	mezzo pubblico	20	11	mezzo pubblico	25
6	piedi	15	12	auto privata	20

(a) Organizzare i dati in una tabella a doppia entrata utilizzando per il fenomeno Y , i seguenti intervalli:

0-|15, 15-|25, 25-|40, 40-|60

e, fissando l'attenzione sul fenomeno X , costruire e confrontare la distribuzione di frequenza marginale e quelle condizionate dalle modalità "non più di 15 minuti" e "più di 40 minuti".

- (b) Stabilire se fra X e Y esiste indipendenza statistica; in caso di risposta negativa valutare il grado di connessione tra X e Y e commentare il risultato ottenuto.
- (c) Fornire una classificazione dei fenomeni statistici e discutere i principali problemi di rilevazione ed di analisi statistica.
- (d) Utilizzando i dati di cui al punto a), valutare tramite un opportuno indice e commentando il risultato, il grado di associazione per la coppia di modalità: auto privata e tempo > 25 minuti.
- (e) Dimostrare, fornendo e commentando tutti i passaggi, il valore massimo dell'indice di connessione χ^2

Esercizio 2:

Su un collettivo di 10 atleti partecipanti alla maratona nell'ultima Olimpiade sono stati congiuntamente rilevati i fenomeni: X = "età" e Y = "numero medio di ore di allenamento alla settimana", ottenendo i seguenti risultati:

Atleta	1	2	3	4	5	6	7	8	9	10
X	23	31	20	25	26	21	27	32	35	24
Y	8	15	7	23	30	10	10	20	16	18

- (a) Costruire e commentare il diagramma a dispersione. Determinare quindi i parametri della retta di regressione dei minimi quadrati che interpreta la dipendenza di Y da X .
- (b) Tracciare la retta calcolata sul diagramma a dispersione e in seguito valutarne con un indice opportuno la bontà di adattamento ai dati, interpretando il risultato numerico ottenuto. Infine utilizzare la retta di regressione per prevedere il numero medio di ore di allenamento alla settimana che potrebbe sostenere un atleta di 29 anni, valutandone anche l'affidabilità sulla base della bontà di adattamento del modello utilizzato per la previsione.
- (c) Definire e interpretare i concetti di varianza *nei* e *fra* gruppi e il ruolo di questi nell'analisi della Dipendenza di un fenomeno dall'altro.
- (d) Sia 20 la varianza spiegata da un secondo modello, diverso dalla retta di regressione individuata al punto (a), che spiega il fenomeno Y in funzione di X . Stabilire se tale modello è da preferire alla retta di regressione calcolata al punto (a) motivando la risposta.
- (e) Esporre il concetto di Correlazione, definire il coefficiente di correlazione lineare ρ e discuterne i possibili valori.

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI RELAZIONI TRA VARIABILI
Soluzioni appello del 18.05.2006¹

Esercizio 1

a)

$X \backslash Y$	0- 15	15- 25	25- 40	40- 60	
piedi	3	0	0	0	3
auto privata	0	1	2	1	4
mezzi pubblici	0	2	2	1	5
	3	3	4	2	12

X	
piedi	3
auto privata	4
mezzi pubblici	5
	12

X	
piedi	3
auto privata	0
mezzi pubblici	0
	3

X	
piedi	0
auto privata	1
mezzi pubblici	1
	2

b)

E' immediatamente possibile asserire l'assenza di indipendenza statistica in quanto nella tabella a doppia entrata compaiono degli zeri. Per valutare il grado di connessione utilizziamo l'indice chi quadrato:

$$\chi^2 = N \left(\sum_{i=1}^k \sum_{j=1}^h \frac{f_{ij}^2}{f_{i.} f_{.j}} - 1 \right)$$

$$= 12 \left(\frac{3^2}{3 \times 3} + \frac{1^2}{4 \times 3} + \frac{2^2}{4 \times 4} + \frac{1^2}{4 \times 2} + \frac{2^2}{5 \times 3} + \frac{2^2}{5 \times 4} + \frac{1^2}{5 \times 2} - 1 \right) = 12.3$$

$$\frac{\chi^2}{N \times \min\{(h-1), (k-1)\}} = \frac{12.3}{2 \times 12} = 0.51$$

¹ A cura di Giancarlo Manzi

Tale risultato mette in evidenza una discreta connessione fra mezzo di trasporto e tempo impiegato per raggiungere il posto di lavoro.

d)

Per prima cosa dicotomizzo la tabella:

X	Y	>25	≤25
auto		3	1
non auto		3	5

Valuto il grado di associazione attraverso l'indice di Edwards:

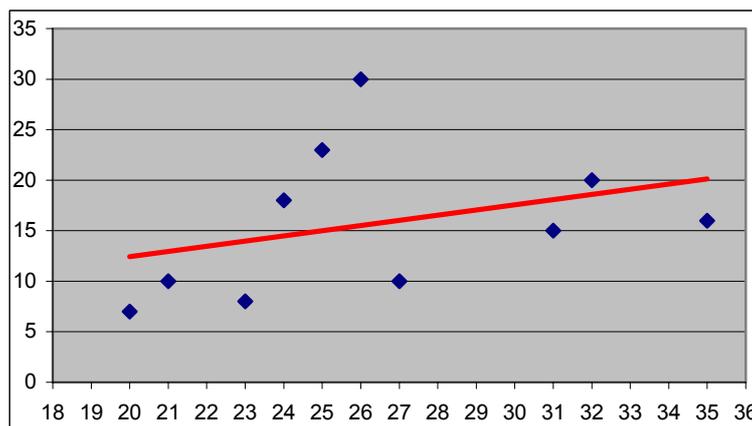
$$E = \frac{f_{11}f_{22}}{f_{11}f_{22} + f_{12}f_{21}} = \frac{3 \times 5}{3 \times 5 + 1 \times 3} = \frac{15}{15 + 3} = 0,83.$$

La coppia di modalità considerata è quindi caratterizzata da una associazione abbastanza forte.

Esercizio 2

a)

Nel grafico sono rappresentati contemporaneamente il diagramma di dispersione e la retta di regressione dei minimi quadrati.



Il diagramma a dispersione evidenzia una correlazione positiva fra le due variabili. Determino i parametri della retta di regressione:

$$\bar{x} = \frac{23 + 31 + 20 + 25 + 26 + 21 + 27 + 32 + 35 + 24}{10} = 26.4$$

$$\bar{y} = \frac{8 + 15 + 7 + 23 + 30 + 10 + 10 + 20 + 16 + 18}{10} = 15.7$$

$$\sigma_x^2 = \frac{23^2 + 31^2 + 20^2 + 25^2 + 26^2 + 21^2 + 27^2 + 32^2 + 35^2 + 24^2}{10} - 26.4^2 = 21.64$$

$$\sigma_y^2 = \frac{8^2 + 15^2 + 7^2 + 23^2 + 30^2 + 10^2 + 10^2 + 20^2 + 16^2 + 18^2}{10} - 15.7^2 = 48.21$$

$$\sigma_{xy} = \frac{1}{10} (23 \times 8 + 31 \times 15 + 20 \times 7 + 25 \times 23 + 26 \times 30 + 21 \times 10 + 27 \times 10 + 32 \times 20 + 35 \times 16 + 24 \times 18) - 26.4 \times 15.7 = 11.12$$

$$b = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{11.12}{21.64} = 0.51 \qquad a = \bar{y} - b\bar{x} = 15.7 - 0.51 \times 26.4 = 2.24$$

b)

Calcolo l'indice della bontà di adattamento della retta ai dati:

$$\rho_{xy}^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{11.12^2}{21.64 \times 48.21} = 0.12$$

Questo valore indica un adattamento molto scarso della retta di regressione ai dati.

$$\hat{Y} = 2.24 + 0.51 \times 29 = 17.03$$

Tuttavia l'affidabilità di questa previsione, dato il valore dell'indice ρ_{xy}^2 determinato in precedenza, risulta molto ridotta.

d)

Per scegliere tra i due modelli confronto le varianze spiegate. La varianza spiegata del primo modello (ossia della retta dei minimi quadrati individuata in precedenza) è pari a:

$$\sigma_y^2 \times \rho_{xy}^2 = 48.21 \times 0.12 = 5.79$$

Dato che $5.79 < 20$, si sceglie il secondo modello.