

# FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006

## ESAME DI RELAZIONI TRA VARIABILI

Ultimo recupero - 12/01/2007

**Avvertenza:** Fornire le formule utilizzate e tutti i passaggi dei calcoli eseguiti.

Utilizzare almeno 2 cifre decimali.

### Esercizio 1:

In una piccola indagine sulla figura femminile nella pubblicità condotta su 20 donne, dopo aver rilevato la loro età ( $Y$ ), è stato chiesto un giudizio ( $X$ ) sul "degrado" dell'immagine della donna rappresentata nella pubblicità in Italia negli ultimi cinque anni. Le risposte possibili erano: "Molto degradata", "Abbastanza degradata", "Poco degradata" e "Per niente degradata". I risultati sono riportati nella seguente tabella:

Id. donna intervistata	Giudizio (variabile X)	Età in anni (variabile Y)	Id. donna intervistata	Giudizio (variabile X)	Età in anni (variabile Y)
1	Poco degradata	32	11	Abbastanza degradata	40
2	Abbastanza degradata	55	12	Poco degradata	20
3	Abbastanza degradata	30	13	Molto degradata	35
4	Molto degradata	45	14	Per niente degradata	30
5	Per niente degradata	20	15	Per niente degradata	25
6	Per niente degradata	18	16	Poco degradata	20
7	Molto degradata	65	17	Poco degradata	26
8	Molto degradata	78	18	Per niente degradata	29
9	Molto degradata	66	19	Molto degradata	81
10	Abbastanza degradata	51	20	Per niente degradata	19

- Organizzare i dati in una tabella a doppia entrata utilizzando per l'età, i seguenti 2 intervalli: Fino a 35 anni, Oltre 35 anni. Fissando l'attenzione sul fenomeno  $X$ , costruire, confrontare e commentare la distribuzione di frequenza marginale e quelle condizionate.
- 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.
- Esporre il concetto di fenomeno statistico, fornirne una classificazione discutendo i principali problemi di rilevazione e di analisi statistica.
- Utilizzando i dati di cui al punto a), valutare, tramite un opportuno indice, il grado di associazione per la coppia di modalità: "Molto degradata" ed età maggiore di 35 anni. Commentare il risultato ottenuto.
- Dimostrare, fornendo tutti i passaggi, che il valore massimo dell'indice di connessione  $\chi^2$  è  $N \cdot \min\{h-1, k-1\}$ , dove  $h$  e  $k$  sono rispettivamente il numero di righe e di colonne della tabella a doppia entrata.

### Esercizio 2:

Su un collettivo di 10 donne è stato rilevato il numero ( $Y$ ) di ore settimanali mediamente dedicate alla lettura e il numero ( $X$ ) di ore settimanali mediamente dedicate alla gestione della propria abitazione (attività domestiche). Si sono ottenuti i seguenti risultati:

Ore di lettura $y_i$	4.5	6	5.5	10	0.5	0	7.5	3.5	5	6.5
Ore di attività domestica $x_i$	32	25	30	15	20	38	5	28	46	11

- Costruire e commentare il diagramma a dispersione per la coppia di fenomeni  $X$ : ore di attività domestiche e  $Y$ : Ore di lettura. Determinare quindi i parametri della retta di regressione dei minimi quadrati che interpreta la dipendenza di  $Y$  da  $X$ .
- 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 lettura alla settimana di una donna che dedica di 29 ore del proprio tempo alle attività domestiche, valutandone anche l'affidabilità sulla base della bontà di adattamento del modello utilizzato per la previsione.
- Definire e interpretare i concetti di Varianza Fra e Nei gruppi e discuterne il ruolo nell'analisi della Dipendenza (statistica) di un fenomeno dall'altro.
- Sia 5 la varianza spiegata da un secondo modello, diverso dalla retta di regressione individuata al punto (a), che interpreta 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.
- Esporre i concetti di Indipendenza Statistica e di Incorrelazione; dimostrare e discutere che la prima implica la seconda ma non viceversa.

**FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006**  
**ESAME DI RELAZIONI TRA VARIABILI**  
**Soluzioni appello del 12.01.2007<sup>1</sup>**

**Esercizio 1**

a)

$X$	$Y$	Fino a 35 anni	Oltre 35 anni	Totale
Molto degradata		1	5	6
Abbastanza degradata		1	3	4
Poco degradata		4	0	4
Per niente degradata		6	0	6
Totale		12	8	20

$X$	
Molto degradata	6
Abbastanza degradata	4
Poco degradata	4
Per niente degradata	6
Totale	20

$X$	Fino a 35 anni
Molto degradata	1
Abbastanza degradata	1
Poco degradata	4
Per niente degradata	6
Totale	12

$X$	Oltre 35 anni
Molto degradata	5
Abbastanza degradata	3
Poco degradata	0
Per niente degradata	0
Totale	8

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:

---

<sup>1</sup> A cura di Giancarlo Manzi

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

$$= 20 \left( \frac{1^2}{12 \times 6} + \frac{5^2}{8 \times 6} + \frac{1^2}{12 \times 4} + \frac{3^2}{8 \times 4} + \frac{4^2}{12 \times 4} + \frac{6^2}{12 \times 6} - 1 \right) = 13.4$$

$$\frac{\chi^2}{N \times \min\{(h-1), (k-1)\}} = \frac{13.4}{20 \times 1} = 0.67$$

Tale risultato mette in evidenza una discreta connessione fra età e giudizio sull'immagine della donna nella pubblicità.

c) Vedere appunti e/o libri di testo.

d)

Per prima cosa deve essere dicotomizzata la variabile  $x$ :

$X$	$Y$	>35	<=35
Molto degradata		5	1
Altro		3	11

Il grado di associazione deve essere valutato tramite l'indice di Edwards:

$$E = \frac{f_{11}f_{22}}{f_{11}f_{22} + f_{12}f_{21}} = \frac{11 \times 5}{11 \times 5 + 1 \times 3} = \frac{55}{55 + 3} = 0,95.$$

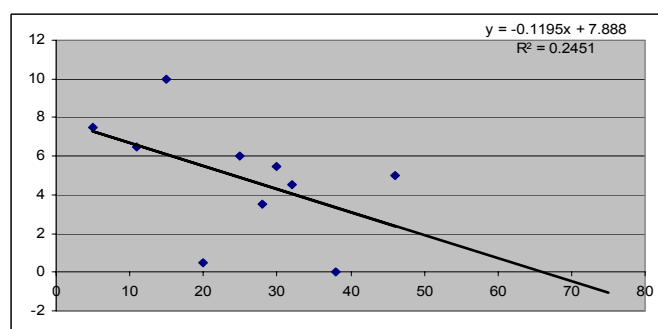
La coppia di modalità considerata è quindi caratterizzata da un'alta associazione.

e) Vedere appunti e/o libri di testo.

## Esercizio 2

a) - b)

Nel grafico seguente sono rappresentati contemporaneamente il diagramma di dispersione e la retta di regressione dei minimi quadrati, la sua equazione ed il coefficiente di determinazione.



In particolare, l'indice di bontà di adattamento ai dati trovato indica un adattamento abbastanza scarso della retta di regressione ai dati.

$$\hat{Y}_{29} = 7,888 - 0,1195 \times 29 = 4,4225$$

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

c) Vedere appunti e/o libri di testo.

d)

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

$$\sigma_Y^2 \times \rho_{XY}^2 = 8,24 \times 0,2451 = 2,0196$$

Dato che  $2,0196 < 5$ , si sceglie il secondo modello.

e) Vedere appunti e/o libri di testo.