

**FACOLTA' DI SOCIOLOGIA – A.A. 2004-2005**  
**ESAME DI RELAZIONI FRA VARIABILI**  
**Appello del 14.09.05**

**QUARTO RECUPERO**

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

**Esercizio 1:** Su un gruppo di 10 bambini è stato condotto uno studio per studiare l'eventuale relazione tra i fenomeni X= 'numero di ore quotidiane dedicate ai giochi all'area aperta' e Y = 'forma fisica', (F = "in forma", B = "Soprapeso"). La rilevazione ha fornito i seguenti risultati:

Bambino	1	2	3	4	5	6	7	8	9	10
X	≥4	0-1	1-2	≥4	0-1	2-3	2-3	1-2	1-2	0-1
Y	F	F	S	F	S	F	S	S	F	S

- Dopo aver organizzato i dati in una tabella a doppia entrata, fornire la distribuzione condizionata di X nella sottopopolazione dei bambini soprapeso, confrontandola con la distribuzione marginale di X.
- Valutare un opportuno indice, il grado di associazione tra le modalità 'soprapeso' di Y e 'n.ro di ore dedicate al gioco superiore o uguale a 4' di X.
- Valutare il grado di connessione tra X e Y. Successivamente fornire una tabella teorica di massima connessione tenendo fisse entrambe le marginali.
- Commentare la relazione esistente fra due fenomeni quantitativi A e B nel caso dei seguenti valori sintetici: Chi Quadrato (normalizzato)=1 e  $\eta^2(B)=1$ .
- Dimostrare che se fra due fenomeni quantitativi esiste indipendenza statistica, allora la covarianza è nulla.

**Esercizio 2:** su un gruppo di 50 persone sono stati rilevati congiuntamente i fenomeni: X = "Età" e Y = "numero di cene fuori casa in un mese". I risultati sono stati organizzati nella seguente tabella a doppia entrata:

	Y	1	3	5
X				
15- 25		5	8	10
25- 35		1	3	8
35- 45		10	4	1

- stabilire se esiste indipendenza in media di Y da X. In caso negativo, valutare il grado di dipendenza, commentando il risultato.
- Costruire e commentare il diagramma a dispersione e la spezzata di regressione che interpreta la dipendenza di Y da X. Determinare quindi i parametri della retta di regressione dei minimi quadrati. Valutare la bontà di adattamento ai dati e commentare i risultati ottenuti.
- Sia  $\sigma_Y^{2*} = 3$  il residuo lasciato da un secondo modello, diverso dalla retta di regressione dei minimi quadrati individuata al punto b). Decidere se tale modello è da preferire alla retta di regressione calcolata al punto b) e motivare la risposta.
- Fornire le definizioni (formule) di varianza NEI e FRA gruppi; commentare e interpretare.
- Esporre e discutere il criterio dei minimi quadrati per la determinazione della retta di regressione.

**Esercizio 1:**

a) N=10

X	Y	F	S
0- 1		1	2
1- 2		1	2
2- 3		1	1
>4		2	0

X	Y	p	p
0- 1		2/5	3/10
1- 2		2/5	3/10
2- 3		1/5	2/10
>4		0	3/10

b)

Y	S	F
X		
>4	0	2
<4	5	3

$$E = (0 * 3) / (0 * 3) + (5 * 2) = 0$$

c) Si veda libro e/o appunti

$$d) \tilde{\chi}^2 = \frac{1}{\min\{4-1; 2-1\}} \left( \sum_{i=1}^4 \sum_{j=1}^2 \frac{f_{ij}^2}{f_{i.} f_{.j}} - 1 \right) = \left( \frac{1^2}{3 \times 3} + \frac{2^2}{3 \times 7} + \dots - 1 \right) / 1 = 0,26$$

Il livello di connessione fra i due eventi è piuttosto basso.

e) Si veda libro e/o appunti

## Es 2:

a)

	Y	1	3	5	Marginali
X					
20		5	8	10	23
30		1	3	8	12
40		10	4	1	15
Marginali		16	15	19	50

Per soddisfare indipendenza in media deve essere:

$$\mu_Y = \mu_{Y|X=20} = \mu_{Y|X=30} = \mu_{Y|X=40}$$

$$\mu_Y = (1 * 16 + 3 * 15 + 5 * 19) / 50 = 3.12$$

ad esempio

$$\mu_{Y|X=20} = (1 * 5 + 3 * 8 + 5 * 10) / 23 = 3.43$$

Basta trovare una media condizionata differente dalla media di Y per poter affermare che **non** esiste indipendenza in media.

$$\mu_{Y|X=30} = (1 * 1 + 3 * 3 + 5 * 8) / 12 = 4.16$$

$$\mu_{Y|X=40} = (1 * 10 + 3 * 4 + 5 * 1) / 15 = 1.8$$

$$\sigma^2_Y = (1^2 * 16 + 3^2 * 15 + 5^2 * 19) - 3.12^2 = 12.52 - 9.7344 = 2.7856$$

$$\sigma^2_{FRA} = 1/50 [ (3.43 - 3.12)^2 * 23 + (4.16 - 3.12)^2 * 12 + (1.8 - 3.12)^2 * 15 ] = 41.32 / 50$$

$$= 0.82651$$

$$\eta^2 = 0.82651 / 2.7856 = 0.296$$

Esiste una dipendenza piuttosto debole tra X e Y.

b)

$$\mu_x = (20 * 23 + 30 * 12 + 40 * 15) / 50 = 28,4$$

$$\sigma^2_x = 880 - 28,4^2 = 73,44$$

$$\sigma_{xy} = (4160 / 50) - 28,4 * 3,12 = -5,408$$

$$\beta = -5,408 / 73,44 = -0,073$$

$$\alpha = 3,12 - (-0,073) * 28,4 = 5,21$$

$$\rho^2 = (-5,408)^2 / 73,44 * 2,78 = 0,143$$

Dal valore emerge uno scarso adattamento ai dati.

c) si veda libro/appunti

d)

$$\sigma_{Y \text{ (retta)}}^2 = \sigma_Y^2 - \bar{\sigma}_Y^2 = \sigma_Y^2 - \sigma_Y^2 \rho_{XY}^2 = \sigma_Y^2 (1 - \rho_{XY}^2) = 2,78(1 - 0,143) = 2,382$$

dato che  $2,382 < 3$  si sceglierà il primo modello calcolato al punto b).

e) si veda libro/appunti