FACOLTÀ DI SOCIOLOGIA – A.A. 2004–2005
ESAME DI RELAZIONI TRA VARIABILI
Appello del 20/12/2005

QUINTO RECUPERO

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

Esercizio 1: Su un collettivo di 28 soggetti (che frequentano un noto collegio di Milano) sono stati rilevati congiuntamente i seguenti due fenomeni: **X**="competenza tecnica in ambiente Linux" (secondo le modalità: SUFFICIENTE/BUONA/OTTIMA) e **Y**="utilizza il programma di scrittura scientifica LaTeX" (secondo le modalità: NO e SÌ). Dalla rilevazione si sono ottenuti i seguenti risultati: un solo individuo dei 6 che hanno una sufficiente competenza tecnica in ambiente Linux non utilizza il programma di scrittura scientifica LaTeX; sui 20 individui che utilizzano il programma di scrittura scientifica LaTeX 10 hanno anche un'ottima competenza tecnica in ambiente Linux; infine 8 individui hanno una buona competenza tecnica in ambiente Linux.

(a) Organizzare i dati in una tabella a doppia entrata e fornire la distribuzione condizionata del fenomeno **X** nella sottopopolazione degli individui che utilizzano il programma di scrittura scientifica LaTeX. Commentare i risultati ottenuti confrontandoli anche con la distribuzione (relativa) marginale di **X**.

(b) Valutare mediante un opportuno indice di associazione se coloro che non hanno un'ottima competenza tecnica in ambiente Linux tendono a non utilizzare il programma di scrittura scientifica LaTeX.

(c) Dopo aver esposto il concetto e di Indipendenza Statistica fornirne la condizione e definire le frequenze (congiunte) teoriche o attese di indipendenza.

(d) Valutare il grado di connessione tra **X** e **Y** e commentare il risultato ottenuto.

(e) Esporre il concetto di indipendenza in media di un fenomeno quantitativo da un fenomeno qualitativo, definire l'indice di dipendenza η^2 e discuterne i valori.

Esercizio 2: Su un collettivo di 12 collegiali (che frequentano il terzo anno di Università) sono stati rilevati congiuntamente i seguenti due fenomeni: **X**="numero medio di esami sostenuti" e **Y**="numero di amici", ottenendo i seguenti risultati:

Soggetto	1	2	3	4	5	6	7	8	9	10	11	12
X	21	8	10	22	21	27	2	8	16	23	26	21
Y	5	4	2	11	1	15	4	6	8	7	14	11

(a) Dopo aver costruito e commentato il diagramma a dispersione, determinare 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 la bontà di adattamento ai dati osservati, commentando i risultati ottenuti. Infine utilizzare la retta di regressione per prevedere il numero di amici di uno studente che ha sostenuto in media 13 esami al terzo anno di Università. Commentare il risultato ottenuto, valutandone anche l'affidabilità sulla base della bontà di adattamento del modello utilizzato per la previsione.

(c) Esporre e discutere il criterio dei minimi quadrati per la determinazione della retta di regressione.

(d) Sia $\overset{*}{\sigma}{}^2_Y = 8$ la varianza residua da un secondo modello, diverso dalla retta di regressione individuata al punto (2a), che spiega il fenomeno **Y** in funzione di **X**. Stabilire se tale modello è da preferire alla retta di regressione calcolata al punto (2a) motivando la risposta.

(e) Definire la covarianza e discuterne anche geometricamente significato, valore, segno e legame con il coefficiente di correlazione lineare ρ.

FACOLTÀ DI SOCIOLOGIA – A.A. 2004–2005
ESAME DI RELAZIONI TRA VARIABILI
Appello del 20/12/2005

TRACCIA DI SOLUZIONE QUINTO RECUPERO[1]

(**1a**)

X \ **Y**	**NO**	**SÌ**	
SUFFICIENTE	1	5	6
BUONA	3	5	8
OTTIMA	4	10	14
	8	20	**28**

X \ **Y**	$p_{X\mid Y=\text{sì}}$	p_X
SUFFICIENTE	0,250	0,214
BUONA	0,250	0,286
OTTIMA	0,500	0,500
	1	**1**

Le distribuzioni sono molto simili pertanto ci si attende una situazione prossima all'i.s.

(**1b**)

X \ **Y**	**NO**	**SÌ**	
SUFFICIENTE/BUONA	4	10	14
OTTIMA	4	10	14
	8	20	**28**

$$\mathrm{E} = \frac{4 \times 10}{4 \times 10 + 10 \times 4} = \frac{40}{80} = 0,5$$

Commento all'indice di Edwards: Il valore di E indica "nessun legame" fra le due modalità, nè "associazione" nè "dissociazione".

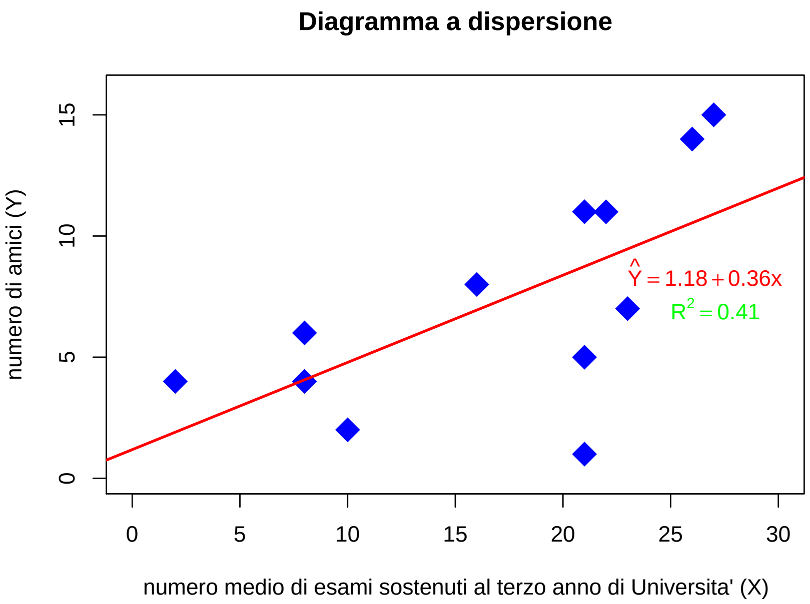
(**1c**)
Si vedano appunti e/o libri di testo.
(**1d**)

$$\tilde{\chi}^2 = \frac{\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{f_{ij}^2}{f_{i\cdot}f_{\cdot j}} - 1}{\min(h-1;k-1)} = \frac{1^2}{6\times 8} + \frac{3^2}{8\times 8} + \frac{4^2}{14\times 8} + \frac{5^2}{6\times 20} + \frac{5^2}{8\times 20} + \frac{10^2}{14\times 20} - 1 \approx 0,03.$$

Commento: siamo molto vicini alla condizione di indipendenza statistica.
(**1e**)
Si vedano appunti e/o libri di testo.

[1] A cura di R. Cirillo.

(**2a**)
N=12



Commento al grafico

La disposizione delle coppie di punti nel diagramma a dispersione indica ($\hat{\mathbf{Y}} = 1,18 + 0,36\mathbf{x}$) una relazione lineare di tipo diretto fra le variabili $\mathbf{X}$ e $\mathbf{Y}$.

$$\begin{aligned}
\mu_X =& \frac{21 \times 3 + 8 \times 2 + 10 + 22 + 27 + 2 + 16 + 23 + 26}{12} = 17,08\\
\mu_Y =& \frac{5 + 4 \times 2 + 2 + 11 \times 2 + 1 + 15 + 6 + 8 + 7 + 14}{12} = 7,33\\
\sigma_X^2 =& \frac{21^2 \times 3 + 8^2 \times 2 + 10^2 + 22^2 + 27^2 + 2^2 + 16^2 + 23^2 + 26^2}{12} - 17,08^2 = 60,58\\
\sigma_Y^2 =& \frac{5^2 + 4^2 \times 2 + 2^2 + 11^2 \times 2 + 1^2 + 15^2 + 6^2 + 8^2 + 7^2 + 14^2}{12} - 7,33^2 = 19,06\\
\sigma_{XY} =& \frac{1}{12}[21 \times 5 + 8 \times 4 + 10 \times 2 + 22 \times 11 + 21 \times 1 + 27 \times 15 + 2 \times 4 + 8 \times 6 +\\
&+ 16 \times 8 + 23 \times 7 + 26 \times 14 + 21 \times 11] - 17,08 \times 7,33 = 21,81\\
\alpha_1^* =& \frac{\sigma_{XY}}{\sigma_X^2} = \frac{21,81}{60,58} = 0,36\\
\alpha_0^* =& \mu_Y - \alpha_1^* \mu_X = 7,33 - 0,36 \times 17,08 = 1,18
\end{aligned}$$

(**2b**)

$$\begin{aligned}
R^2 =& \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \frac{21,81^2}{60,58 \times 19,06} \approx 0,412\\
\hat{Y}_{13} =& \alpha_0^* + \alpha_1^* \times 13 = 1,18 + 0,36 \times 13 \approx 6.
\end{aligned}$$

Si prevede che uno studente che in media ha sostenuto 13 esami al terzo anno di Università abbia 6 amici. Considerando l'indice della bontà di adattamento individuato il risultato è affidabile solo al 41%.
(**2c**)
Si vedano appunti e/o libri di testo.
(**2d**)

$$\overset{*}{\sigma}{}_Y^2 = \sigma_Y^2 \times (1 - \rho_{XY}^2) = 19,06 \times (1 - 0,412) \approx 11,21$$

Dato che $11,21 > 8$ si sceglierà il secondo modello.
(**2e**)
Si vedano appunti e/o libri di testo.