

FACOLTÀ DI SOCIOLOGIA - A.A. 2005-2006
ESAME DI RELAZIONI TRA VARIABILI
15/2/2006

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

Esercizio 1:

100 soggetti sono stati classificati secondo la circonferenza toracica X e la statura Y (entrambe espresse in cm), secondo le seguenti modalità (intervalli):

60- 70	70- 80	80- 90	per X
150- 165	165- 180	180- 195	per Y

La rilevazione ha fornito le seguenti informazioni:

- il 30% dei soggetti ha circonferenza toracica compresa tra 60-|70 cm. Di questi, il 50% ha statura compresa tra 150-|165 cm mentre per il 40% la statura è compresa tra 165-|180 cm;
 - il 25% dei soggetti ha statura tra 150-|165 cm. Di questi nessuno ha circonferenza toracica tra 80-|90 cm;
 - il 50% dei soggetti ha circonferenza toracica tra 70-|80 cm. Di questi, il 62% ha statura tra 165-|180 cm;
 - infine, tra coloro che hanno circonferenza toracica tra 80-|90 cm, il 90% ha statura compresa tra 180-|195 cm.
- (a) Costruire la tabella a doppia entrata relativa ai fenomeni X ed Y e ricavare le distribuzioni condizionate di X rispetto alle tre modalità di Y .
- (b) Stabilire se esiste indipendenza statistica fra X e Y ; in caso di risposta negativa verificare se X è indipendente in media da Y e, in caso di risposta negativa, valutare con l'opportuno indice il grado di dipendenza commentando il risultato.
- (c) Definire e interpretare la proprietà di associatività delle medie condizionate e la proprietà di scomposizione della varianza marginale in varianza "NEI" e "FRA" gruppi.
- (d) Valutare tramite un opportuno indice e commentando il risultato, il grado di associazione per la coppia di modalità: circonferenza toracica > 80 cm e altezza > 180 cm.
- (e) Discutere e dimostrare il valore dell'indice di associazione di Edwards nel caso in cui i due fenomeni congiuntamente osservati siano statisticamente indipendenti.

Esercizio 2:

Nella tabella che segue vengono riportati il numero delle pagine (X) ed il costo di pubblicazione in euro (Y) relativi a 10 volumi editi di una nota casa editrice italiana:

x_i	150	180	295	300	310	330	335	340	350	400
y_i	27	30	28	35	45	40	32	50	55	60

- (a) Costruire e commentare il diagramma a dispersione. Determinare la retta di regressione dei minimi quadrati che interpreta la dipendenza di Y da X .
- (b) Tracciare la retta calcolata sul diagramma a dispersione e valutarne la bontà di adattamento ai dati osservati, commentando il risultato ottenuto. Utilizzare quindi la retta di regressione per prevedere il costo di pubblicazione di un volume di 250 pagine. Commentare il risultato ottenuto, valutandone anche l'affidabilità sulla base della bontà di adattamento del modello utilizzato per la previsione.
- (c) Fornire un esempio (con dati a scelta) di tabella a doppia entrata in cui esista perfetta correlazione lineare positiva ($\rho_{XY} = +1$) e affiancare il corrispondente diagramma a dispersione. Commentare tabella e grafico.
- (d) Sia $DS_2=500$ la devianza spiegata da un secondo modello, diverso dalla retta di regressione individuata al punto (a), per interpretare 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 e discutere il criterio dei Minimi Quadrati per la determinazione della retta di regressione.

Esercizio 1

a)

$X \setminus Y$	150- 165	165- 180	180- 185	
60- 70	15	12	3	30
70- 80	10	31	9	50
80- 90	0	2	18	20
	25	45	30	100

x	$f_{11}/f_{.1}$	$f_{12}/f_{.2}$	$f_{13}/f_{.3}$
60- 70	0,6	0,27	0,1
70- 80	0,4	0,69	0,3
80- 90	0	0,04	0,6
	1	1	1

b)

Dato che $f_{31} = 0 \neq \frac{f_{3.}f_{.1}}{N}$, non vi è indipendenza statistica. Inoltre, dato che le distribuzioni condizionate ricavate al punto precedente non coincidono, possiamo concludere che non sussiste indipendenza in media di X da Y . Valutiamo dunque il grado di dipendenza con l'indice η_X^2 .

$$\bar{x} = \frac{65 \times 30 + 75 \times 50 + 85 \times 20}{100} = 74$$

$$\bar{x} | y_1 = \frac{65 \times 15 + 75 \times 10 + 85 \times 0}{25} = 69$$

$$\bar{x} | y_2 = \frac{65 \times 12 + 75 \times 31 + 85 \times 2}{45} = 72,78$$

$$\bar{x} | y_3 = \frac{65 \times 3 + 75 \times 9 + 85 \times 18}{30} = 80$$

$$\sigma_X^2 = \frac{65^2 \times 30 + 75^2 \times 50 + 85^2 \times 20}{100} - 74^2 = 49$$

$$\sigma_{FRA}^2 = \frac{(69 - 74)^2 \times 25 + (72,78 - 74)^2 \times 45 + (80 - 74)^2 \times 30}{100} = 17,72$$

$$\eta_X^2 = \frac{\sigma_{FRA}^2}{\sigma_X^2} = \frac{17,72}{49} = 0,36$$

La dipendenza di X da Y è il 36% della perfetta dipendenza, ovvero la statura spiega statisticamente il 36% della variabilità della statura nel collettivo osservato.

d)

Per prima cosa dicotomizzo la tabella:

$X \setminus Y$	150- 180	180- 185
60- 80	68	12
80- 90	2	18

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

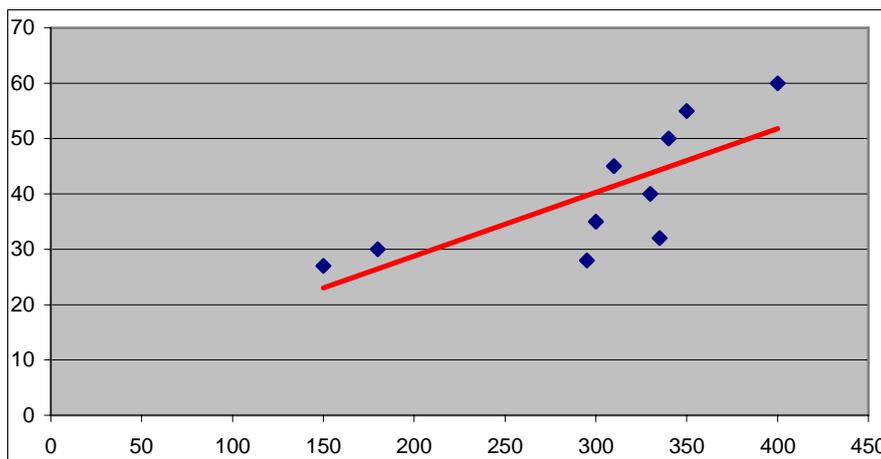
$$E = \frac{f_{11}f_{22}}{f_{11}f_{22} + f_{12}f_{21}} = \frac{68 \times 18}{68 \times 18 + 12 \times 2} = \frac{1224}{1224 + 24} = 0,98$$

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

Esercizio 2

a)

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



$$\bar{x} = \frac{150 + 180 + 295 + 300 + 310 + 330 + 335 + 340 + 350 + 400}{10} = 299$$

$$\bar{y} = \frac{27 + 30 + 28 + 35 + 45 + 40 + 32 + 50 + 55 + 60}{10} = 40,2$$

$$\sigma_x^2 = \frac{150^2 + 180^2 + 295^2 + 300^2 + 310^2 + 330^2 + 335^2 + 340^2 + 350^2 + 400^2}{10} - 299^2 = 5324$$

$$\sigma_y^2 = \frac{27^2 + 30^2 + 28^2 + 35^2 + 45^2 + 40^2 + 32^2 + 50^2 + 55^2 + 60^2}{10} - 40,2^2 = 125,16$$

$$\sigma_{xy} = \frac{1}{10} (150 \times 27 + 180 \times 30 + 295 \times 28 + 300 \times 35 + 310 \times 45 + 330 \times 40 + 335 \times 32 + 340 \times 50 + 350 \times 55 + 400 \times 60) - 299 \times 40,2 = 613,2$$

$$b = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{613,2}{5324} = 0,11$$

$$a = \bar{y} - b\bar{x} = 40,2 - 0,11 \times 299 = 7,31$$

b)

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

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \frac{613,2^2}{5324 \times 125,16} = 0,56$$

Quindi in questo caso la retta dei m.q. spiega il 56% della variabilità di Y , una quota abbastanza scarsa.

Calcolo ora il costo di pubblicazione previsto dal modello per un volume di 250 pagine:

$$\hat{Y} = 7,31 + 0,11 \times 250 = 34,81$$

Questa previsione, dato il valore dell'indice ρ_{XY}^2 determinato in precedenza, risulta affidabile al 56%.

d)

$$DT = N \times \sigma_Y^2 = 10 \times 125,16 = 1251,6$$

$$DS = \rho_{XY}^2 \times DT = 0,56 \times 1251,6 = 700,9$$

Poiché $DS = 700,9 > 500 = DS_2$ la retta di regressione è da preferire al secondo modello.