

Università degli Studi di Cassino
Anno accademico 2003-2004
Corsi di Statistica 1, II (Prof. G. Prozio) e Statistica 1, IV (Dott. D. Vistocco)

Esercitazione del 16/2/2004
Dott. Claudio Conversano

Esercizio 1

Su un campione di 15 studenti della Facoltà di Economia dell'Università di Cassino sono state rilevate le seguenti caratteristiche:

- a) E' stato chiesto allo studente se svolgeva un'attività lavorativa oltre allo studio e se tale attività era di tipo abituale o saltuario
- b) E' stato chiesto allo studente il numero di esami sostenuti

Si riportano di seguito le variabili rilevate e la relativa codifica:

Tipo di lavoro	A = Abituale, S = saltuario, NL = nessuna attività lavorativa
N° di esami sostenuti	≤ 5 = non più di 5 esami, 5- 10 = da 6 a 10 esami, >10 = più di 10 esami

La successione delle risposte fornite dai 15 studenti è riportata di seguito:

(A, 5-|10), (S, ≤ 5), (S, 5-|10), (S, >10), (A, ≤ 5), (NL, 5-|10), (NL, >10), (S, 5-|10) (A, ≤ 5), (S, >10), (NL, >10), (A, ≤ 5), (NL, >10), (A, ≤ 5), (NL, >10)

A partire dalla successione delle risposte fornite:

- a) costruire la distribuzione doppia "Tipo di lavoro – N° esami sostenuti" e rappresentarla in una tabella a doppia entrata;
- b) Indicare quali sono le distribuzioni marginali e le distribuzioni condizionate;
- c) Misurare la connessione (dipendenza) tra i due caratteri osservati.

Svolgimento

Costruzione della distribuzione doppia "Tipo di lavoro – N° esami sostenuti" e rappresentazione in una tabella a doppia entrata.

La tabella a doppia assumerà la forma seguente:

N° esami sostenuti	≤ 5	5- 10	>10
Tipo di lavoro			
Saltuario			
Abituale			
Nessun lavoro			

Sistemando opportunamente ciascuna osservazione, individuata dalla coppia di modalità dei due caratteri osservati si ottiene la seguente distribuzione doppia:

N° esami sostenuti	<=5	5- 10	>10	Totale
Tipo di lavoro				
Saltuario	1	2	2	5
Abituale	4	1	0	5
Nessun lavoro	0	1	4	5
Totale	5	4	6	15

b) distribuzioni marginali e distribuzioni condizionate

Le distribuzioni marginali sono le distribuzioni dei totali di riga e di colonna

Distribuzione marginale del tipo di lavoro:

Tipo di lavoro	Frequenza assoluta
Saltuario	5
Abituale	5
Nessun lavoro	5
Totale	15

Distribuzione marginale del numero di esami superati:

N° esami sostenuti	Frequenza assoluta
< 5	5
5 - 10	4
>10	6
Totale	15

Le distribuzioni marginali si ottengono fissando una riga o una colonna della tabella e considerando la distribuzione di un carattere avendo fissato una modalità dell'altro carattere.

Distribuzioni marginali del carattere "Tipo di lavoro"

Distribuzione del "Tipo di lavoro" per gli studenti che hanno sostenuto fino a 5 esami:

Tipo di lavoro	Frequenza assoluta
Saltuario	1
Abituale	4
Nessun lavoro	0
Totale	5

Distribuzione del “Tipo di lavoro” per gli studenti che hanno sostenuto fino da 5 a 10 esami:

Tipo di lavoro	Frequenza assoluta
Saltuario	2
Abituale	1
Nessun lavoro	1
Totale	4

Distribuzione del “Tipo di lavoro” per gli studenti che hanno sostenuto più di 10 esami:

Tipo di lavoro	Frequenza assoluta
Saltuario	2
Abituale	0
Nessun lavoro	4
Totale	6

Distribuzione del numero di esami sostenuti per gli studenti che lavorano saltuariamente:

N° esami sostenuti	Frequenza assoluta
< 5	1
5 - 10	2
>10	2
Totale	5

Distribuzione del numero di esami sostenuti per gli studenti che lavorano abitualmente:

N° esami sostenuti	Frequenza assoluta
< 5	4
5 - 10	1
>10	0
Totale	5

Distribuzione del numero di esami sostenuti per gli studenti non lavoratori:

N° esami sostenuti	Frequenza assoluta
< 5	0
5 - 10	1
>10	4
Totale	5

Connessione.

La connessione si misura mediante l'indice chi-quadro:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$

in cui n_{ij}^* indica la frequenza teorica associata alla frequenza osservata n_{ij} calcolata dividendo il prodotto delle frequenze marginali per il numero di osservazioni n .

$$n_{ij}^* = \frac{n_{i.} \cdot n_{.j}}{n}$$

Per calcolare l'indice chi-quadro si riporta dapprima la tabella delle frequenze teoriche:

N° esami sostenuti	<=5	5-10	>10
Tipo di lavoro			
Saltuario	1,67	1,33	2,00
Abituale	1,67	1,33	2,00
Nessun lavoro	1,67	1,33	2,00

Dalla tabella delle frequenze teoriche si calcolano i termini $\frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$ e si riportano nella tabella seguente:

N° esami sostenuti	<=5	5-10	>10
Tipo di lavoro			
Saltuario	0,27	0,33	0,00
Abituale	3,27	0,08	2,00
Nessun lavoro	1,67	0,08	2,00

La somma degli elementi di quest'ultima tabella è uguale all'indice chi-quadro:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} = 9,70$$

Poiché l'indice chi-quadro è compreso tra zero e “+ infinito”, se il calcolo dell'indice fornisce un valore elevato (come in questo caso) significa che è molto forte la connessione (o dipendenza tra i due caratteri osservati).