

ESERCITAZIONE SU TABELLE A DOPPIA ENTRATA

In un collettivo di 20 laureati è stato rilevato il numero di colloqui di lavoro effettuati dopo la laurea e il numero di mesi trascorsi dalla laurea e il numero di mesi trascorsi.

Nella parentesi il primo dato si riferisce al numero di colloqui, la seconda cifra al numero di mesi trascorsi.

(5, 2) (5, 4) (5, 4) (5, 4) (5, 4) (5, 3) (5, 3) (5, 3) (5, 2) (5, 4)
(2, 2) (2, 2) (2, 3) (2, 3) (3, 3) (3, 2) (2, 4) (3, 4) (4, 2) (4, 3)

Si richiede:

- 1) la costruzione della tabella a doppia entrata;
- 2) il calcolo del tempo medio trascorso dalla laurea;
- 3) il calcolo del numero medio di colloqui di lavoro sostenuti da coloro che si sono laureati da meno di quattro mesi;
- 4) il calcolo della distribuzione percentuale per riga e commentare il significato

SVOLGIMENTO

Quesito 1

Per il calcolo della tabella a doppia entrata per prima cosa è opportuno identificare il numero e i valori assunti dalle modalità della variabile X (numero di colloqui di lavoro effettuati), in questo caso la variabile ha 4 modalità i cui valori sono 2,3,4,5. Analogamente si tratta la variabile Y (numero di mesi trascorsi), le cui modalità (in totale 3) assumono i seguenti valori 2,3,4. In tal modo si costituisce una tabella che è formata da tante righe quante sono le modalità della X (4) e tante colonne quante sono le modalità della Y (3)

La tabella è composta dalla variabile X e quali valori assumono queste modalità. Lo stesso discorso vale per la variabile Y. In questo caso sono 4 le modalità della X e i valori sono 2, 3, 4 e 5

X\Y	2	3	4	Totale
2	2	2	1	5
3	1	1	1	3
4	1	1	0	2
5	2	3	5	10
Totale	6	7	7	20

La tabella è costituita da n (=4) righe e m (=3) colonne e da un numero di celle pari a mxn (=12).

In ognuna di queste celle vengono inserite le frequenze assolute degli studenti che presentano simultaneamente una modalità della variabile X e una della variabile Y. Per esempio ci sono 2 studenti, laureati da 2 mesi, che hanno sostenuto 2 colloqui

Quesito 2

x_i	n_i	$x_i \cdot n_i$
2	6	12
3	7	21
4	7	28
Σ	20	61

$\mu = 3,05$ cioè il tempo medio trascorso dalla laurea

Quesito 3

Per il numero medio di colloqui di coloro che si sono laureati da meno di 4 mesi

Si isola solo il gruppo dei 13 studenti che si è laureato prima di quattro mesi () si

x_i	n_i	$x_i \cdot n_i$
2	4	8
3	2	6
4	2	8
5	5	25
Σ	13	39

$\mu = 39/13$

Quesito 4

X\Y	2	3	4	Totale
2	2	2	1	5
3	1	1	1	3
4	1	1	0	2
5	2	3	5	10
Totale	6	7	7	20

X\Y	Distribuzioni condizionate di Y x			Totale
	2	3	4	
2	0,40	0,40	0,20	1
3	0,33	0,33	0,33	1
4	0,50	0,50	0,0	1
5	0,20	0,30	0,50	1
Totale	0,30	0,35	0,35	1

ESERCITAZIONE - INDIPENDENZA

Esercizio n. 1

La tabella di seguito (tab. 1) riporta 20 studenti classificati secondo il rendimento scolastico (X) e il livello di impegno nello studio (Y). Si richiede:

- 1) il calcolo - per entrambi i caratteri- delle frequenze relative in caso di indipendenza;
- 2) la costruzione della tabella di contingenza;
- 3) la verifica che la somma algebrica delle contingenze di una riga o colonna sia nulla;
- 4) il calcolo dell'indice Chi Quadrato di Pearson, della contingenza quadratica media Φ^2 , dell'indice C_1 di contingenza di Cramer

Tab. 1 - Studenti secondo il rendimento scolastico e il livello di impegno nello studio

X	Y			Totale
	Basso	Medio	Alto	
Basso	2	2	1	5
Medio	1	1	1	3
Alto	1	1	0	2
Molto alto	2	3	5	10
Totale	6	7	7	20

SVOLGIMENTO

Quesito 1.

In una tabella a doppia entrata $n_{i.}$ e $n_{.j}$ sono le frequenze marginali di riga e di colonna n è il totale delle unità statistiche.

L'indipendenza tra due caratteri si manifesta quando tutte le frequenze congiunte soddisfano la seguente relazione $n_{ij}=(n_{i.}n_{.j})/n$

Tabella delle frequenze teoriche calcolate in caso di indipendenza

X	Y			Totale
	Basso	Medio	Alto	
Basso	1,50	1,75	1,75	5
Medio	0,90	1,05	1,05	3
Alto	0,60	0,70	0,70	2
Molto alto	3,00	3,50	3,50	10
Totale	6	7	7	20

Quesito 2.

Variabile X: nel caso di indipendenza i profili condizionati sono uguali tra di loro e uguali a quello marginale

X	Y			Totale
	Basso	Medio	Alto	
Basso	0,25	0,25	0,25	0,25
Medio	0,15	0,15	0,15	0,15
Alto	0,10	0,10	0,10	0,10
Molto alto	0,50	0,50	0,50	0,50
Totale	1	1	1	1

Variabile Y

X	Y			Totale
	Basso	Medio	Alto	
Basso	0,30	0,35	0,35	1
Medio	0,30	0,35	0,35	1
Alto	0,30	0,35	0,35	1
Molto alto	0,30	0,35	0,35	1
Totale	0,30	0,35	0,35	1

Quesiti 3 e 4

La contingenza esprime la differenza tra la frequenza osservata e la frequenza teorica ricavata con l'ipotesi di indipendenza

Differenza tra osservate e teoriche

X	Y			Totale
	Basso	Medio	Alto	
Basso	0,50	0,25	-0,75	0
Medio	0,10	-0,05	-0,05	0
Alto	0,40	0,30	-0,70	0
Molto alto	-1,00	-0,50	1,50	0
Totale	0	0	0	0

Quadrato delle differenza tra osservate e teoriche

X	Y			Totale
	Basso	Medio	Alto	
Basso	0,2500	0,0625	0,5625	
Medio	0,0100	0,0025	0,0025	
Alto	0,1600	0,0900	0,4900	
Molto alto	1,0000	0,2500	2,2500	

Quadrato delle contingenze/frequenza teorica

X	Y			Totale
	Basso	Medio	Alto	
Basso	0,1667	0,0357	0,3214	0,524
Medio	0,0111	0,0024	0,0024	0,016
Alto	0,2667	0,1286	0,7000	1,095
Molto alto	0,3333	0,0714	0,6429	1,048
Totale	0,778	0,238	1,667	2,6825

Formula per il calcolo del Chi Quadrato= $\sum (n_{ij}-n_{ij}')^2/n_{ij}'$

CALCOLO DEL CHI QUADRATO UTILIZZANDO LA FORMULA ALTERNATIVA (*)

Quadrato delle frequenze osservate

X	Y			Totale
	Basso	Medio	Alto	
Basso	4	4	1	
Medio	1	1	1	
Alto	1	1	0	
Molto alto	4	9	25	

Quadrato freq. oss. / prod. riga e colonna

X	Y			Totale
	Basso	Medio	Alto	
Basso	0,133333	0,114286	0,028571	0,27619
Medio	0,055556	0,047619	0,047619	0,150794
Alto	0,083333	0,071429	0	0,154762
Molto alto	0,066667	0,128571	0,357143	0,552381
Totale				1,134127
(Totale-1)*N				2,68254

Nota (*): $X^2 = (\text{Doppia sommatoria } n_{ij}^2/n_i \cdot n_j - 1)n$

$X^2 = 2,6825$

Φ^2 (indice di contingenza quadratica media) = $X^2/n = 0,1341$

C_1 (indice di contingenza di Cramer) = $\Phi^2/\max\Phi^2$ dove con $\max\Phi^2$ si intende il minimo ((rig,col)-1)

$C_1 = 0,1341/(3-1) = 2,6825/20/2 = 0,06706$