

ESERCITAZIONE N. 5 - INDIPENDENZA E CORRELAZIONE

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 , dell'indice C_r 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 $n_{ij} = (n_{i.} \cdot n_{.j}) / n$

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		
	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= $\sum (\eta - n_{ij})^2 / n_{ij}$

CALCOLO DEL CHI QUADRATO UTILIZZANDO LA FORMULA ALTERNATIVA^()*

Quadrato delle frequenze osservate

X	Y		
	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 (*): $\chi^2 = (\text{Doppia sommatoria } n_{ij}^2 / n_{i.} n_{.j} - 1) n$

$\chi^2 = 2,6825$

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

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

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

Esercizio n. 2

Calcolare il coefficiente di correlazione lineare tra le variabili X ed Y i cui valori sono riportati nella tabella seguente (tab. 2).

Tab. 2 - Studenti secondo il tempo trascorso dalla laurea e il numero di colloqui di lavoro sostenuti

X	Y			Totale
	2	3	4	
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

Schema di calcolo per la Media (XY)

X	Y			Totale
	2	3	4	
2	8	12	8	28
3	6	9	12	27
4	8	12	0	20
5	20	45	100	165
Totale	42	78	120	240

$$r_{xy} = \text{cov}(XY) / \sqrt{\text{var}(X) \text{var}(Y)}$$

$$\text{cov}(XY) = \text{Media}(XY) - \text{Media}(X)\text{Media}(Y)$$

$$\text{Media}(Y) = 61/20 \quad \text{Media}(Y) = 3,05 \quad \text{tempo medio trascorso dalla laurea}$$

$$\text{Media}(X) = 77/20 \quad \text{Media}(X) = 3,85 \quad \text{numero medio di colloqui dopo la laurea}$$

$$\text{Media}(XY) = 12$$

$$\text{cov}(XY) = 0,257$$

$$\text{Media}(Y^2) = 9,95 \quad \text{Media}(X^2) = 16,45$$

$$\text{Media}(Y)^2 = 9,3025 \quad \text{Media}(X)^2 = 14,82$$

$$\text{Media}(Y^2) - \text{Media}(Y)^2 = 0,6475 \quad \text{Media}(X^2) - \text{Media}(X)^2 = 1,63$$

$$r_{xy} = 1,06 / \sqrt{0,6475} \quad r_{xy} = 0,2494$$

Esercizio sulla costruzione del boxplot

Testo

Siano dati i valori seguenti:

1, 3, 10, 13, 14, 18, 26, 38, 39, 43, 46, 57, 63, 64, 66, 67, 68, 69, 71, 74, 76, 77, 79, 80, 95, 96, 160

Calcolare riassunto a cinque e costruire il box plot

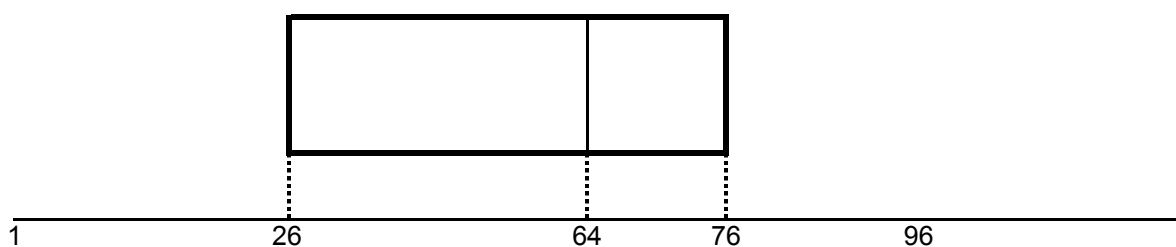
Soluzione

Si calcolino i valori $Q_0 (=x_{\min})$, Q_1 , $Q_2 (= \text{mediana})$, Q_3 e $Q_4 (=x_{\max})$. I valori sono 27

$Q_0 = 1$, $Q_1 = 26$, $Q_2 = 64$, $Q_3 = 76$ e $Q_4 = 160$

In sostanza, il riassunto a cinque è il seguente: 1, 26, 64, 76 e 160

Si disegni una scatola di estremi Q_1 e Q_3 e tagliata sulla mediana



Si calcolano i valori a e b:

$$a = Q_1 - 1,5 \cdot (Q_3 - Q_1) = 26 - 1,5 \cdot (76 - 26) = -49$$

$$b = Q_3 + 1,5 \cdot (Q_3 - Q_1) = 76 + 1,5 \cdot (76 - 26) = 151$$

da cui $\alpha = 1$ e $\beta = 96$

α = minimo dei valori maggiori di a

β = massimo valori minori di b

