

Statistica Descrittiva

Soluzioni 8. Dipendenza

Esercizio A.

- a) La distribuzione marginale relativa del tipo di occupazione è data da: 0,0529, 0,3401, 0,6069.
 b) Le distribuzioni condizionate relative richieste sono le seguenti:

Regione	Agricoltura	Industria	Altre attività	Totale
Piemonte	0,0514	0,3993	0,5493	1
Veneto	0,0541	0,4143	0,5316	1
Emilia Romagna	0,0714	0,3495	0,5791	1
Toscana	0,0421	0,3422	0,6157	1
Lazio	0,0440	0,1992	0,7568	1

Essendo queste distribuzioni non tutte uguali tra loro, possiamo affermare che esiste dipendenza tra i due caratteri.

- c) Per il calcolo degli indici richiesti, si considerino le seguenti frequenze teoriche di indipendenza ottenute come $f_{ij}^* = f_{i.} \cdot f_{.j} / N$:

Regione	Agricoltura	Industria	Altre attività	Totale
Piemonte	89,625	575,813	1027,562	1693
Veneto	97,883	628,871	1122,246	1849
Emilia Romagna	89,678	576,153	1028,169	1694
Toscana	71,625	460,175	821,200	1353
Lazio	96,189	617,988	1102,823	1817
Totale	445	2859	5102	8406

Con queste si ricava l'indice χ^2 che risulta pari a 266,645. Il coefficiente di contingenza in questo caso risulta pari a

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N} \cdot \frac{k}{k-1}} = \sqrt{\frac{266,645}{266,645 + 8406} \cdot \frac{3}{3-1}} = 0,215,$$

dove $k = \min(\text{numero di righe; numero di colonne})$. L'indice C indica una dipendenza bassa del tipo di occupazione dalla regione.

Esercizio B.

- a) La distribuzione marginale della reazione al test è pari a: 25, 36, 31, 20. Quindi, la distribuzione marginale relativa e le distribuzioni condizionate relative della reazione al test sono date da

	elevata	discreta	moderata	negativa	Totale
Gruppo 1	0,326	0,279	0,302	0,093	1
Gruppo 2	0,167	0,458	0,250	0,125	1
Gruppo 3	0,143	0,095	0,286	0,476	1
Marginale Y	0,223	0,321	0,277	0,179	1

b) Le distribuzioni condizionate relative del tipo di patologia data la risposta al test sono le seguenti

	elevata	discreta	moderata	negativa	Marginale X
Gruppo 1	0,56	0,33	0,42	0,20	0,38
Gruppo 2	0,32	0,61	0,39	0,30	0,43
Gruppo 3	0,12	0,06	0,19	0,50	0,19
Totale	1	1	1	1	1

Ovviamente, essendo le distribuzioni condizionate relative non tutte uguali tra loro, c'è dipendenza tra i caratteri X ed Y .

c) Per una tabella di dimensione 3 righe (X) per 4 colonne (Y) ci può essere perfetta dipendenza di X da Y , ma non di Y da X .

d) Per il calcolo dell'indice di dipendenza χ^2 consideriamo la seguente tabella

$\frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}$					
	2,0187	0,2400	0,1013	1,7623	
	0,6876	2,7989	0,1244	0,7714	
	0,6075	3,3426	0,0060	10,4167	
					22,8776

dove $f_{ij}^* = f_{i.} \cdot f_{.j} / N$. Quindi, l'indice di dipendenza χ^2 è pari a 22,8776, mentre il coefficiente di contingenza C risulta pari a 0,5044. Secondo questo indice, nella tabella c'è un grado di dipendenza intermedio.