

Statistica Descrittiva

Esercitazione 8. Dipendenza

Esercizio A. Relativamente alla Tavola 7.2 del “Compendio Statistico Italiano 1998” (pag. 86) e limitatamente alle regioni Piemonte, Veneto, Emilia Romagna, Toscana e Lazio si consideri la seguente distribuzione doppia di frequenza degli occupati (in migliaia) secondo i caratteri *Tipo di occupazione* (Agricoltura, Industria, Altre attività) e *Regione*:

Regione	Agricoltura	Industria	Altre attività	Totale
Piemonte	87	676	930	1693
Veneto	100	766	983	1849
Emilia Romagna	121	592	981	1694
Toscana	57	463	833	1353
Lazio	80	362	1375	1817
Totale	445	2859	5102	8406

Sulla base di questa distribuzione doppia di frequenza:

- Si ricavi la distribuzione marginale relativa del carattere *Tipo di occupazione*.
- Si ricavino le distribuzioni condizionate relative del carattere *Tipo di occupazione* per le cinque regioni. Si dica se esiste o meno dipendenza tra i due caratteri considerati.
- Si calcoli l'indice di dipendenza χ^2 , il coefficiente di contingenza C e si commentino i risultati ottenuti.

Esercizio B. Si consideri la tavola seguente relativa alla risposta ad un test dermatologico (Y) di tre gruppi di individui affetti da una diversa patologia dermatologica (X):

	Reazione al test dermatologico				totale
	elevata	discreta	moderata	negativa	
Gruppo 1	14	12	13	4	43
Gruppo 2	8	22	12	6	48
Gruppo 3	3	2	6	10	21

- Si ottenga la distribuzione marginale relativa del carattere $Y = \text{Reazione al test}$ e le distribuzioni condizionate relative del carattere $Y = \text{Risposta al test}$ per i tre gruppi.
- Si ottengano le distribuzioni condizionate relative del tipo di patologia (gruppo) data la risposta al test e si dica se esiste dipendenza o meno tra i caratteri X ed Y .
- Si dica se per questo tipo di tabella (di dimensione 3 righe \times 4 colonne) è possibile che ci sia perfetta dipendenza di X da Y , o di Y da X .
- Si calcoli l'indice di dipendenza χ^2 , il coefficiente di contingenza C e si commenti.

Esercizio C. Si consideri la Tavola 11.7 (“Eurostat 2006-2007”, pag. 296), relativamente alle nazioni Grecia, Spagna, Italia, che riporta la produzione di Cereali, Ortaggi e Latte, espressa in milioni di tonnellate

	Produzione (1000 tonnellate)			Totale
	Cereali	Ortaggi	Latte	
Grecia	42	40	67	89
Italia	201	144	101	446
Spagna	133	129	59	321
Totale	376	313	167	856

- Si determinino le distribuzioni marginali dei caratteri *Tipo di produzione* e *Nazione*.
- Si determini la distribuzione condizionata relativa del tipo di produzione rispetto alle nazioni considerate.
- Si valuti se esiste dipendenza tra i due caratteri.

Esercizio D. (Facoltativo) Si considerino due caratteri X e Y , con modalità rispettivamente x_1, \dots, x_r e y_1, \dots, y_c .

- Assumendo $r = 3$ (X presenta solamente tre modalità) e $c = 3$ (Y presenta solamente tre modalità), si riempiano le celle della tabella di frequenza

	y_1	y_2	y_3	Totale
x_1				
x_2				
x_3				
Totale				20

nell'ipotesi che vi sia massima dipendenza tra i due caratteri. Si consideri fissata a 20 la somma delle frequenze assolute.

- Assumendo $r = 3$ (X presenta solamente tre modalità) e $c = 3$ (Y presenta solamente tre modalità), si riempiano le celle della tabella di frequenza

	y_1	y_2	y_3	Totale
x_1				
x_2				
x_3				
Totale				

nell'ipotesi che vi sia indipendenza tra i due caratteri. La scelta del totale delle frequenze assolute è libera.