

INDIPENDENZA

Dati due caratteri X e Y, la situazione estrema di assenza di un qualsiasi legame fra di essi si ha quando qualsiasi modalità assunta dal carattere X non modifica la distribuzione di frequenze del carattere Y, e viceversa.

Due caratteri X e Y sono **INDIPENDENTI** se e solo se la distribuzione di frequenze relative condizionate di X è costante e la distribuzione di frequenze relative condizionate di Y è costante:

$$\frac{f_{i1}}{f_{\bullet 1}} = \frac{f_{i2}}{f_{\bullet 2}} = \dots = \frac{f_{ij}}{f_{\bullet j}} = \dots = \frac{f_{ih}}{f_{\bullet h}} = \frac{f_{i\bullet}}{N} \quad (\forall i, i = 1, \dots, k)$$

$$\frac{f_{1j}}{f_{1\bullet}} = \frac{f_{2j}}{f_{2\bullet}} = \dots = \frac{f_{ij}}{f_{i\bullet}} = \dots = \frac{f_{kj}}{f_{k\bullet}} = \frac{f_{\bullet j}}{N} \quad (\forall j, j = 1, \dots, h)$$

ESEMPIO di due caratteri indipendenti

Su una popolazione di N=200 adulti sono state rilevati i due caratteri qualitativi X=sexo e Y=bibita preferita:

X sesso	Y bibita preferita			
	A	B	C	
Maschi	40	28	12	80
Femmine	60	42	18	120
	100	70	30	200

Due caratteri X e Y sono **INDIPENDENTI** se e solo se sussistono le seguenti relazioni:

$$f_{ij} = \frac{f_{i\bullet} \cdot f_{\bullet j}}{N} \quad (\forall (i, j), i = 1, \dots, k; j = 1, \dots, h)$$

DIPENDENZA

Se fra due caratteri esiste un qualche legame, lo si misura mediante una specie di *distanza* fra le frequenze congiunte realmente **osservate** nella popolazione in esame e le frequenze congiunte **ipotetiche** o **teoriche** che si sarebbero dovute osservare nel caso in cui ci fosse stata una perfetta indipendenza fra i caratteri.

ANALISI DELLA CONNESSIONE O DIPENDENZA

L'**Analisi della Connessione** studia il legame esistente fra due caratteri qualitativi o mutabili A e B. Per "connessione" si intende la relazione di dipendenza di B da A e di A da B.

Indice "chi-quadrato" o **Indice quadratico di connessione** (di Karl Pearson):

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

Dove: $f_{ij}^* = \frac{f_{i\cdot} \cdot f_{\cdot j}}{N}$; $0 \leq \chi^2 < +\infty$

Indice "psi":

$$\psi = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^k \sum_{j=1}^h \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}} = \sqrt{\frac{\chi^2}{N}}$$

Dove: $0 \leq \psi < +\infty$

- Se $\chi^2=0$ oppure $\psi=0$, allora i due caratteri sono indipendenti ovvero non esiste connessione;
- Più χ^2 oppure ψ crescono, più i due caratteri sono connessi.

ESEMPIO

A ABBONAMENTO	B SCUOLA			
	Ist. Prof.	Ist.Tecnico	Liceo	
F.S.	48	68	48	164
A.M.T.	20	16	72	108
entrambi	52	36	40	128
	120	120	160	400

Tabella di frequenze congiunte **ipotetiche** o **teoriche**:

A ABBONAMENTO	B SCUOLA			
	Ist. Prof.	Ist.Tecnico	Liceo	
F.S.	49,2	49,2	65,6	164
A.M.T.	32,4	32,4	43,2	108
entrambi	38,4	38,4	51,2	128
	120	120	160	400

$$\chi^2 = 52,1818$$

$$\psi = \sqrt{\frac{52,1818}{400}} = \sqrt{0,1304545} = 0,3612$$

Ma poiché tali **Indici di connessione** non esprimono il grado della relazione di dipendenza fra A e B, è opportuno ricorrere a un indice normalizzato che permetta una migliore interpretazione:

“**Indice di Cramér**” (: Indice normalizzato di connessione)

$$C = \frac{\psi}{\sqrt{\min[(r-1), (c-1)]}}$$

L'*Indice di Cramér* assume valori compresi fra zero e uno: $0 \leq C \leq 1$. Esso assume valore 0 se la connessione non esiste, ovvero i due caratteri sono indipendenti. Esso assume valore 1 se la connessione esiste ed è massima.

NOTAZIONE: nel testo di Olivieri, “Fondamenti di Statistica”, con il simbolo C viene indicato il “*Coefficiente di contingenza*” (: Indice normalizzato di connessione)

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N} \cdot \frac{k}{k-1}}$$

Dove:

$$k = \min(r, c)$$

$$N = \sum_{i=1}^k \sum_{j=1}^h f_{ij}$$

Il *Coefficiente di contingenza* assume valori compresi fra zero e uno: $0 \leq C \leq 1$. Esso assume valore 0 se la connessione non esiste, ovvero i due caratteri sono indipendenti. Esso assume valore 1 se la connessione esiste ed è massima.

Nell'ESEMPIO:

$$C = \frac{0,3612}{\sqrt{\min[(3-1), (3-1)]}} = \frac{0,3612}{\sqrt{2}} = 0,25541 \cong 26\%$$