*Media aritmetica del **PRODOTTO** di 2 variabili statistiche*

Sia X una variabile di media M(x)=m_x e Y una variabile di media M(y)=m_y. Sia Z=X·Y la *variabile PRODOTTO* delle due variabili.

La media aritmetica di Z risulta:

$$M(Z) = M(X \cdot Y) = m_x \cdot m_y + Cov(x, y)$$

ESERCIZIO (Variabile PRODOTTO)

Date le variabili X e Y, calcolare la media del prodotto Z=X·Y.

X	Y
1	8
2	7
3	6
4	5
5	4

Metodo diretto di calcolo:

X	Y	Z
1	8	8
2	7	14
3	6	18
4	5	20
5	4	20
		80

M(Z)=80/5=16

Metodo indiretto di calcolo:

X	Y	$x-m_x$	$y-m_y$	$(x-m_x)(y-m_y)$
1	8	-2	2	-4
2	7	-1	1	-1
3	6	0	0	0
4	5	1	-1	-1
5	4	2	-2	-4
15	30			-10

M(x)=15/5=3

M(y)=30/5=6

M(z)=3*6+(-10/5)=18-2=16

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:

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

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

ESEMPIO di due caratteri indipendenti

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

	Y bibita preferita			
X sesso	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:

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

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.

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

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

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

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

χ^2 =52,1818

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: $\mathbf{0 \leq C \leq 1}$. 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=0,416=41,6%