Indice di contingenza quadratica media (phi quadro)

$$\Phi^2 = \frac{\chi^2}{n}$$

- Proprietà
  - L'influenza del numero di unità n è eliminata
  - Assume valore 0 se X e Y sono perfettamente indipendenti

Indice di Cramer

$$V = \sqrt{\frac{\Phi^2}{\min\{(H-1),(K-1)\}}}$$

- Proprietà
  - Assume valori compresi tra 0 e 1
  - Assume valore 0 se X e Y sono perfettamente indipendenti
  - Assume valore 1 quando X e Y sono perfettamente dipendenti o interdipendenti

Dipendenza o indipendenza ?

Situazione osservata

Sesso	Statura			Totale
	Basso	Medio	Alto	
f	19	5	0	**24**
m	6	6	4	**16**
Totale	**25**	**11**	**4**	**40**

Situazione teorica di indipendenza

Sesso	Statura			Totale
	Basso	Medio	Alto	
f	15	6.6	2.4	**24**
m	10	4.4	1.6	**16**
Totale	**25**	**11**	**4**	**40**

Contingenze

$$n_{ij} - n_{ij}^{*}$$

Sesso	Statura		
	Basso	Medio	Alto
f	4	-1.6	-2.4
m	-4	1.6	2.4

Calcolo di una misura di associazione

Contingenze

Sesso	Statura		
	Basso	Medio	Alto
f	4	-1.6	-2.4
m	-4	1.6	2.4

$$\chi^2 = \sum_{i=1}^{H} \sum_{j=1}^{K} \frac{\left(n_{ij} - n_{ij}^{*}\right)^2}{n_{ij}^{*}}$$

$$\Phi^2 = \frac{\chi^2}{n}$$

$$V = \sqrt{\frac{\Phi^2}{\min\{(H-1),(K-1)\}}}$$

$$\chi^2 = 9.64$$

$$\Phi^2 = \frac{9.64}{40} = 0.241$$

$$V = \sqrt{\frac{0.241}{\min\{(1),(2)\}}} = \sqrt{\frac{0.241}{1}} = 0.491$$

$$\chi^2 = \frac{4^2}{15} + \frac{(-1.6)^2}{6.6} + \frac{(-2.4)^2}{2.4} + \frac{(-4)^2}{10} + \frac{(-1.6)^2}{4.4} + \frac{(-2.4)^2}{1.6} =$$
$$= 1.066667 + 0.387879 + 2.4 + 1.6 + 0.581818 + 3.6 =$$
$$= 9.636364$$

Laureati in Soc. e SDC per cond. Occupazionale e Sesso

Situazione osservata

Attuale condizione occupazionale	Sesso		Totale
	f	*m*	
Non lavora	23	11	**34**
Lavora	190	68	**258**
Totale	**213**	**79**	**292**

Situazione teorica indipendenza

Attuale condizione occupazionale	Sesso		Totale
	f	*m*	
Non lavora	24.8	9.2	**34**
Lavora	188.2	69.8	**258**
Totale	**213.0**	**79.0**	**292**

Contingenze

Attuale condizione occupazionale	Sesso	
	f	*m*
Non lavora	-1.80	1.80
Lavora	1.80	-1.80

$$\chi^2 = \frac{(-1.8)^2}{24.8} + \frac{(1.8)^2}{9.2} + \frac{(1.8)^2}{188.2} + \frac{(-1.8)^2}{69.8} = 0.55$$

$$\Phi^2 = \frac{\chi^2}{n} = \frac{0.55}{292} = 0.00187$$

$$\text{V} = \sqrt{\frac{0.00187}{1}} = 0.04$$

Dipendenza perfetta

Situazione osservata

Attuale condizione occupazionale	Genere		Totale
	f	*m*	
Non lavora	213	0	**213**
Lavora	0	79	**79**
Totale	**213**	**79**	**292**

Situazione teorica indipendenza

Attuale condizione occupazionale	Genere		Totale
	f	*m*	
Non lavora	155,37	57,63	**213,0**
Lavora	57,63	21,37	**79,0**
Totale	**213,0**	**79,0**	**292,0**

Contingenze

Attuale condizione occupazionale	Genere	
	f	*m*
Non lavora	57,63	-57,63
Lavora	-57,63	57,63

$$\chi^2 = 21{,}38 + 57{,}63 + 57{,}63 + 155{,}41 = 292$$

$$\Phi^2 = \frac{\chi^2}{\text{n}} = \frac{292}{292} = 1$$

$$\text{V} = \sqrt{\frac{1}{1}} = 1$$

Laureati per cond. Occupazionale e Facoltà

Situazione osservata

Attuale condizione occupazionale	Facoltà				Totale
	SDC	*Soc.*	*Sc. politiche*	*Economia*	
Lavora	222	276	105	286	**889**
Non Lavora	190	66	31	65	**352**
Totale	**412**	**342**	**136**	**351**	**1241**

Situazione teorica indipendenza

Attuale condizione occupazionale	Facoltà				Totale
	SDC	*Soc.*	*Sc. politiche*	*Economia*	
Lavora	295,14	244,99	97,42	251,44	**889**
Non Lavora	116,86	97,01	38,58	99,56	**352**
Totale	**412**	**342**	**136**	**351**	**1241**

Contingenze

Attuale condizione occupazionale	Facoltà			
	SDC	*Soc.*	*Sc. politiche*	*Economia*
Lavora	-73,14	31,01	7,58	34,56
Non Lavora	73,14	-31,01	-7,58	-34,56

$$\chi^2 = \frac{(-73{,}14)^2}{295{,}14} + \frac{(31{,}01)^2}{245{,}99} + \frac{(7{,}58)^2}{97{,}42} + \frac{(34{,}56)^2}{251{,}44} + \frac{(73{,}14)^2}{116{,}86} + \frac{(-31{,}01)^2}{97{,}01} + \frac{(-7{,}58)^2}{38{,}58} + \frac{(-34{,}56)^2}{99{,}56} = 96{,}56$$

$$\Phi^2 = \frac{\chi^2}{n} = \frac{96{,}56}{1241} = 0{,}0778$$

$$V = \sqrt{\frac{0{,}0778}{\min(1,3)}} = \sqrt{\frac{0{,}0778}{1}} = 0{,}279$$

Rappresentazioni grafiche di distribuzioni doppie

- Distribuzione doppia di frequenze
  - Tabella di contingenza
  - Tabella di correlazione

Stereogramma

- Distribuzione unitaria doppia di 2 caratteri quantitativi

Scatter Plot

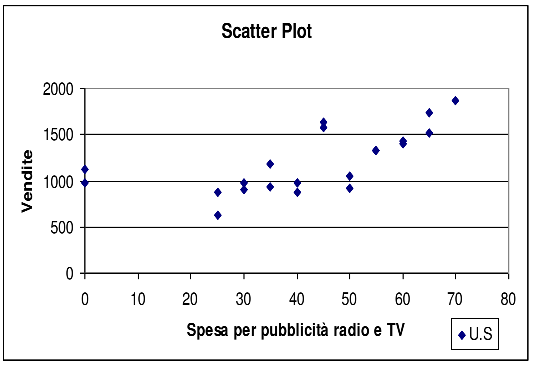
Scatter Plot o Grafico di Dispersione

- *Distribuzione unitaria doppia* di 2 *caratteri quantitativi*
- Sull'asse delle ascisse (X) e su quello delle ordinate (Y) sono riportati rispettivamente i valori numerici delle modalità assunti dalle due variabili rilevate su ogni u.s.
- L'insieme di punti così ottenuto si chiama **nuvola di punti** e consente di studiare la dispersione delle u.s. e la loro somiglianza
- La forma della nuvola può suggerire una relazione funzionale tra i due caratteri

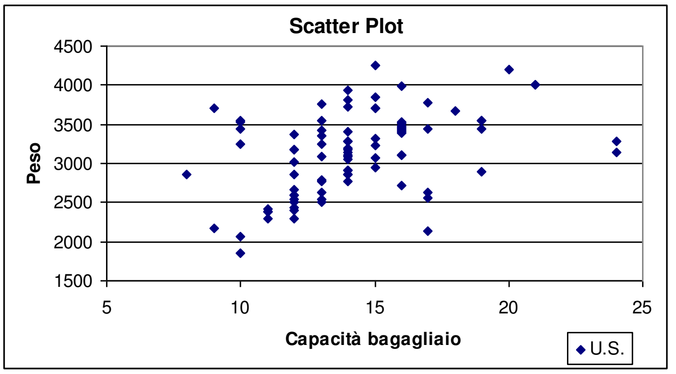
Esempio

Distribuzione Unitaria **Doppia**

Unità Statistica	Vendite	Spesa per pubblicità su radio e TV
1	973	0
2	1119	0
3	875	25
4	625	25
5	910	30
6	971	30
7	931	35
8	1177	35
9	882	40
10	982	40
11	1628	45
12	1577	45
13	1044	50
14	914	50
15	1329	55
16	1330	55
17	1405	60
18	1436	60
19	1521	65
20	1741	65
21	1866	70
22	1717	70



Esempio

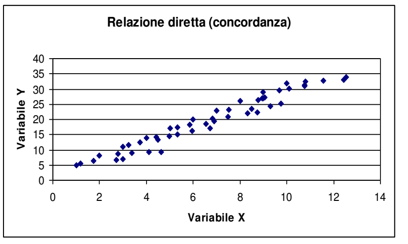


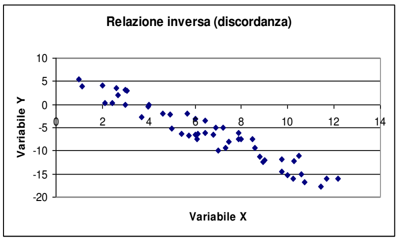
Interdipendenza tra due caratteri quantitativi

- Distribuzione unitaria di 2 caratteri quantitativi X e Y
- Analisi dell'associazione attraverso indici simmetrici che valutano la presenza di
  - *Concordanza*: u.s. con valori piccoli (grandi) di un carattere presentano più frequentemente valori piccoli (grandi) dell'altro carattere
  - *Discordanza*: u.s. con valori piccoli (grandi) di un carattere possiedono più frequentemente valori grandi (piccoli) dell'altro carattere

…continua

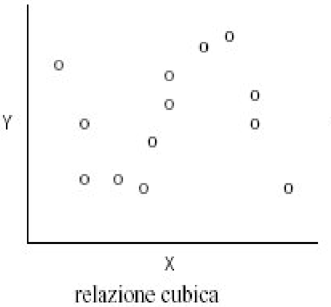
- Per rilevare interdipendenza tra X e Y si può usare lo scatter-plot
- Secondo la forma della nuvola dei punti si ha
  - *Concordanza*: nuvola allungata verso alto a destra
  - *Discordanza*: nuvola allungata verso alto a sinistra
  - *Assenza di <u>interdipendenza lineare</u>*: nuvola pressoché circolare



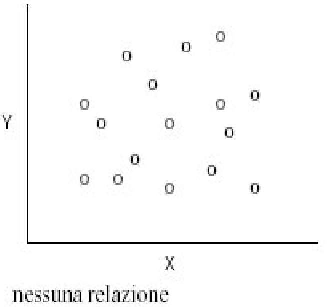


…continua

Assenza di <u>interdipendenza lineare</u>: nuvola pressoché circolare



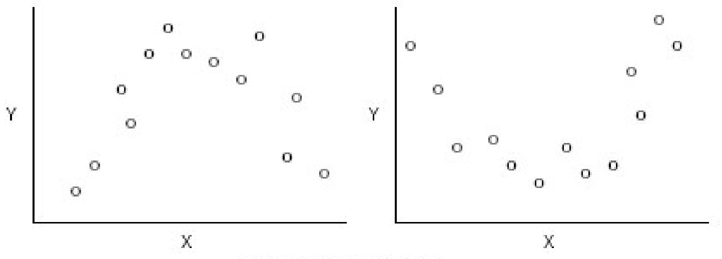
relazione cubica



nessuna relazione

…continua

Assenza di interdipendenza lineare: relazioni quadratiche



relazioni quadratiche

Una misura di interdipendenza: la covarianza

- Media del prodotto degli scarti delle due variabili dalle rispettive medie

$$Cov(X,Y) = \frac{\sum_{j=1}^{N}(x_j - M_X)(y_j - M_Y)}{N}$$

- Proprietà
  - Può assumere sia valori positivi che negativi
  - **Positiva** $\Rightarrow$ in media gli scarti delle due variabili dalle rispettive medie sono concordi $\Rightarrow$ X e Y sono concordi (relazione diretta)
  - **Negativa** $\Rightarrow$ in media gli scarti delle due variabili dalle rispettive medie sono discordi $\Rightarrow$ i caratteri X e Y sono discordi (relazione inversa)
  - Se X e Y sono statisticamente indipendenti $\Rightarrow$ $\sigma_{XY} = 0$
  - Se $\sigma_{XY} = 0 \Rightarrow X$ e Y sono *incorrelati* ma non si può affermare che sono statisticamente indipendenti

Esempio: calcolo della covarianza

Regione	PIL X	TD Y	$(x_j\text{-}M_X)$	$(y_j\text{-}M_Y)$	$(x_j\text{-}M_X)^2$	$(y_j\text{-}M_Y)^2$	$(x_j\text{-}M_X)(y_j\text{-}M_Y)$
PIE	24.778	7.0	3.124	-3.9	9.76	14.82	-12.028
LOM	27.063	4.1	5.409	-6.8	29.26	45.56	-36.512
EMR	25.588	5.0	3.934	-5.9	15.48	34.22	-23.015
LAZ	23.107	11.4	1.453	0.6	2.11	0.30	0.799
CAM	14.219	21.5	-7.434	10.7	55.26	113.42	-79.181
PUG	15.168	16.1	-6.485	5.3	42.06	27.56	-34.051
					153.92	**235.90**	**-183.987**

$$Media_X = \frac{129.923}{6} = 21.65$$

$$Media_Y = \frac{65.1}{6} = 10.85$$

$$VAR(X) = \frac{153.92}{6} = 25.65$$

$$\text{sqm}(X) = \sqrt{25.65} = 5.065$$

$$VAR(Y) = \frac{235.90}{6} = 39.32$$

$$\text{sqm}(Y) = \sqrt{39.32} = 6.27$$

$$Cov(X,Y) = \frac{\sum_{j=1}^{N} (x_j - M_x)(y_j - M_y)}{N}$$

$$Cov(X,Y) = \frac{-183.987}{6} = -30.6645$$