



**Analisi dell'associazione tra
due caratteri:
indipendenza e dipendenza**

Introduzione

- ❑ **Analisi univariata, bivariata, multivariata**
- ❑ **Analizzare le relazioni tra i caratteri, per cercare di “prevedere” il valore (sconosciuto) di una variabile a partire da quello (conosciuto) di un'altra**
- ❑ **Distribuzioni doppie di frequenze (tabelle doppie)**
- ❑ **Associazione tra caratteri (dipendenza, indipendenza, ecc..)**
- ❑ **Il χ^2 (Chi-quadrato)**

Distribuzione doppia di frequenze

		Carattere Y						Totale
		Y ₁	Y ₂	...	Y _j	...	Y _k	
Carattere X	X ₁	n ₁₁	n ₁₂	...	n _{1j}	...	n _{1k}	n _{1.}
	X ₂	n ₂₁	n ₂₂	...	n _{2j}	...	n _{2k}	n _{2.}

	X _i	n _{i1}	n _{i2}	...	n _{ij}	...	n _{ik}	n _{i.}

X _h	n _{h1}	n _{h2}	...	n _{hj}	...	n _{hk}	n _{h.}	
Totale		n _{.1}	n _{.2}	...	n _{.j}	...	n _{.k}	n _{..}

Distribuzione di Y condizionata a X₁

Distribuzione Marginale di X

Distribuzione Marginale di Y

Distribuzione di X condizionata a Y₂

Frequenza assoluta congiunta

Distribuzione doppia – 1

a) Distribuzione doppia di frequenze

	Maschi	Femmine	Totale
Economia	75	45	120
Statistica	35	65	100
Matematica	60	20	80
Totale	170	130	300

Distribuzione di Y condizionata a X_1

Distribuzione di X condizionata a Y_1

Frequenza assoluta congiunta

Distribuzione Marginale di X

Distribuzione Marginale di Y

Distribuzione doppia – 2

a) Distribuzione doppia di frequenze

	Maschi	Femmine	Totale
Economia	75	45	120
Statistica	35	65	100
Matematica	60	20	80
Totale	170	130	300

$$75/300 = 0,25$$

$$65/300 = 0,21$$

b) Distribuzione doppia di frequenze relative

	Maschi	Femmine	Totale
Economia	0,25	0,15	0,40
Statistica	0,12	0,21	0,33
Matematica	0,20	0,07	0,27
Totale	0,57	0,43	1,00

Somma=1

Distribuzione doppia – 3

a) Distribuzione doppia di frequenze

	Maschi	Femmine	Totale
Economia	75	45	120
Statistica	35	65	100
Matematica	60	20	80
Totale	170	130	300

$$35/170 = 0,21$$

$$45/130 = 0,35$$

d) Distribuzione di frequenze relative di X condizionate ad Y

	Maschi	Femmine	Totale
Economia	0,44	0,35	0,40
Statistica	0,21	0,50	0,33
Matematica	0,35	0,15	0,27
Totale	1,00	1,00	1,00

Somma=1

Distribuzione doppia – 4

a) Distribuzione doppia di frequenze

	Maschi	Femmine	Totale
Economia	75	45	120
Statistica	35	65	100
Matematica	60	20	80
Totale	170	130	300

$$75/120 = 0,62$$

$$20/80 = 0,25$$

c) Distribuzione di frequenze relative di Y condizionate ad X

	Maschi	Femmine	Totale
Economia	0,62	0,38	1,00
Statistica	0,35	0,65	1,00
Matematica	0,75	0,25	1,00
Totale	0,57	0,43	1,00

Somma=1

INDIPENDENZA di X da Y

a) Frequenze assolute

		Carattere Y		
		Y ₁	Y ₂	Totale
Carattere X	X ₁	10	20	30
	X ₂	20	40	60
	X ₃	30	60	90
	Totale	60	120	180

b) Frequenze relative del carattere X condizionato ad Y

		Carattere Y		
		Y ₁	Y ₂	Totale
Carattere X	X ₁	0,167	0,167	0,167
	X ₂	0,333	0,333	0,333
	X ₃	0,500	0,500	0,500
	Totale	1,000	1,000	1,000

Il carattere X si dirà **indipendente** dal carattere Y se tutte le distribuzioni relative condizionate risultano uguali tra loro e uguali alla distribuzione marginale (e dunque, **al variare della modalità Y la distribuzione relativa di X è la medesima**).

INDIPENDENZA di Y da X

a) Frequenze assolute

		Carattere Y		
		Y ₁	Y ₂	Totale
Carattere X	X ₁	10	20	30
	X ₂	20	40	60
	X ₃	30	60	90
	Totale	60	120	180

b) Frequenze relative del carattere Y condizionato ad X

		Carattere Y		
		Y ₁	Y ₂	Totale
Carattere X	X ₁	0,333	0,667	1,000
	X ₂	0,333	0,667	1,000
	X ₃	0,333	0,667	1,000
	Totale	0,333	0,667	1,000

Il carattere Y si dirà **indipendente** dal carattere X se tutte le distribuzioni relative condizionate risultano uguali tra loro e uguali alla distribuzione marginale (e dunque, **al variare della modalità X la distribuzione relativa di Y è la medesima**).

INDIPENDENZA

E' possibile dimostrare che se il carattere X è indipendente dal carattere Y, allora vale anche la relazione contraria: anche il carattere Y sarà indipendente dal carattere X.

Pertanto: due caratteri X ed Y si diranno indipendenti se le distribuzioni relative condizionate di un carattere rispetto alle modalità dell'altro sono uguali.

$$n'_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

**Frequenze teoriche
di indipendenza**

INDIPENDENZA

a) Frequenze assolute

		Carattere Y		
		Y ₁	Y ₂	Totale
Carattere X	X ₁	10	20	30
	X ₂	20	40	60
	X ₃	30	60	90
	Totale	60	120	180

Ogni volta che non troviamo questa situazione



Dipendenza

$$n'_{11} = \frac{n_{1.} \cdot n_{.1}}{n} = \frac{30 \cdot 60}{180} = \frac{1.800}{180} = 10$$

$$n'_{32} = \frac{n_{3.} \cdot n_{.2}}{n} = \frac{90 \cdot 120}{180} = \frac{10.800}{180} = 60$$

Dipendenza perfetta di Y da X

Un carattere Y dipende perfettamente da X se ad ogni modalità di X è associata una ed una sola modalità del carattere Y

		Carattere Y		
		Y ₁	Y ₂	Totale
Carattere X	X ₁	0	20	20
	X ₂	20	0	20
	X ₃	0	60	60
	Totale	20	80	100

Se $X = X_1 \rightarrow$

$Y = Y_2$

Se $X = X_2 \rightarrow$

$Y = Y_1$

Se $X = X_3 \rightarrow$

$Y = Y_2$

La relazione di dipendenza non è biunivoca!!!

Dipendenza perfetta di X da Y

Un carattere X dipende perfettamente da Y se ad ogni modalità di Y è associata una ed una sola modalità del carattere X

		Carattere Y				Totale
		Y ₁	Y ₂	Y ₃	Y ₄	
Carattere X	X ₁	20	0	0	0	20
	X ₂	0	20	0	0	20
	X ₃	0	0	30	60	90
	Totale	20	20	30	60	130

Se $Y = Y_1$



$X = X_1$

Se $Y = Y_2$



$X = X_2$

Se $Y = Y_3$



$X = X_3$

Se $Y = Y_4$

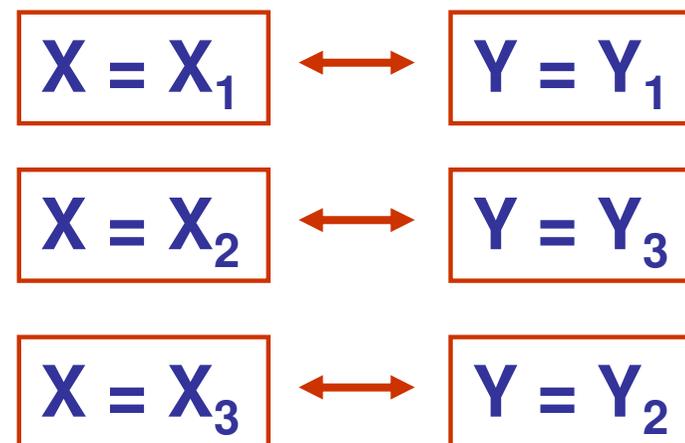


$X = X_3$

Perfetta Interdipendenza

L'interdipendenza perfetta può essere raggiunta solo nel caso di **tabella quadrata**

		Carattere Y			
		Y ₁	Y ₂	Y ₃	Totale
Carattere X	X ₁	25	0	0	25
	X ₂	0	0	30	30
	X ₃	0	45	0	45
	Totale	25	45	30	100



Interdipendenza (perfetta): ad ogni modalità del carattere X corrisponde una ed una sola modalità di Y e, simultaneamente, ad ogni modalità del carattere Y corrisponde una ed una sola modalità di X

Misurare la Dipendenza – il χ^2

$$\chi^2 = \sum_{i=1}^H \sum_{j=1}^K \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$$

Indipendenza



$$n_{ij} = n'_{ij}$$



$$\chi^2 = 0$$

Dipendenza



$$n_{ij} \neq n'_{ij}$$



$$\chi^2 > 0$$

Massima dipendenza



$$\chi^2_{\max} = n \cdot \min[(H - 1); (K - 1)]$$

Misurare la Dipendenza – il χ^2

$$V = \sqrt{\frac{\chi^2}{\chi^2_{\max}}}$$

V di Cramer

Tale indice può variare **tra zero ed uno**, e sarà pari a **zero** nel caso di **indipendenza**, mentre assumerà valore **1** nel caso di **massima dipendenza**.

Esempio

Abbiamo effettuato un'indagine sugli studenti del nostro Ateneo al fine di rilevare il grado di associazione tra il voto da questi riportato nell'esame di matematica e nell'esame di statistica. I risultati sono riportati nella tabella seguente (in riga i voti di matematica, in colonna quelli di statistica):

	18-22	23-26	27-30	
18-22	20	3	2	25
23-26	2	27	6	35
27-30	4	6	30	40
	26	36	38	100

Si calcoli l'associazione tra queste due variabili utilizzando l'indice Chi-quadrato e la V di Cramer

Esempio - 2

b) Frequenze teoriche

	18 - 22	23 - 26	27 - 30	Tot
18 - 22	$\frac{25 \cdot 26}{100} = 6,5$	$\frac{25 \cdot 36}{100} = 9$	$\frac{25 \cdot 38}{100} = 9,5$	25
23 - 26	$\frac{35 \cdot 26}{100} = 9,1$	$\frac{35 \cdot 36}{100} = 12,6$	$\frac{35 \cdot 38}{100} = 13,3$	35
27 - 30	$\frac{40 \cdot 26}{100} = 10,4$	$\frac{40 \cdot 36}{100} = 14,4$	$\frac{40 \cdot 38}{100} = 15,2$	40
Tot	26	36	38	100

c) Differenza (Frequenze effettive - Frequenze teoriche)

	18 - 22	23 - 26	27 - 30	Tot
18 - 22	$(20 - 6,5) = 13,5$	$(3 - 9) = -6$	$(2 - 9,5) = -7,5$	0
23 - 26	$(2 - 9,1) = -7,1$	$(27 - 12,6) = 14,4$	$(6 - 13,3) = -7,3$	0
27 - 30	$(4 - 10,4) = -6,4$	$(6 - 14,4) = -8,4$	$(30 - 15,2) = 14,8$	0
Tot	0	0	0	0

Esempio - 3

**d) Differenza
(Frequenze effettive -
Frequenze teoriche)²**

	18 - 22	23 - 26	27 - 30	Tot
18 - 22	$13,5^2 = 182,25$	$-6^2 = 36$	$-7,5^2 = 56,25$	
23 - 26	$-7,1^2 = 50,41$	$14,4^2 = 207,36$	$-7,3^2 = 53,29$	
27 - 30	$-6,4^2 = 40,96$	$-8,4^2 = 70,56$	$14,8^2 = 219,04$	
Tot				

**e) Differenza (Frequenze
effettive -Frequenze
teoriche)²diviso teoriche**

	18 - 22	23 - 26	27 - 30	Tot
18 - 22	$\frac{182,25}{6,5} = 28,04$	$\frac{36}{9} = 4$	$\frac{56,25}{9,5} = 5,92$	
23 - 26	$\frac{50,41}{9,1} = 5,54$	$\frac{207,36}{12,6} = 16,46$	$\frac{53,29}{13,3} = 4,01$	
27 - 30	$\frac{40,96}{10,4} = 3,94$	$\frac{70,56}{14,4} = 4,9$	$\frac{219,04}{15,2} = 14,41$	
Tot				87,21

Esempio - 4

$$\chi^2 = \sum_{i=1}^H \sum_{j=1}^K \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} = 87,21$$

$$\chi^2_{\max} = n \cdot \min[(H - 1); (K - 1)] = 100 \cdot (3 - 1) = 100 \cdot 2 = 200$$

$$V = \sqrt{\frac{\chi^2}{\chi^2_{\max}}} = \sqrt{\frac{87,21}{200}} = \sqrt{0,436} = 0,66$$

Associazione – competenze acquisite

Cosa abbiamo imparato?

- ❑ Analizzare 2 caratteri
- ❑ Indipendenza, dipendenza, interdipendenza
- ❑ Misurare il grado di dipendenza

Quali strumenti usare?

❑ Per analizzare 2 caratteri



Tabelle doppie

❑ Per misurare la dipendenza



χ^2