Statistica Sociale

Le misure di posizione e scala che abbiamo introdotto fino ad ora si riferivano ad una sola variabile. A volte, però, abbiamo in qualche modo dato un'occhiata a due variabili congiuntamente.

Si ha una distribuzione doppia quando si esaminano congiuntamente due variabili sulle unità statistiche del nostro collettivo.

Come nel caso di distribuzioni relative ad una singola variabile, si parlerà di distribuzioni doppie disaggregate quando si elencano le N coppie di modalità e di distribuzioni doppie di frequenze, quando le osservazioni sono aggregate per modalità o classi.

Analisi dell'associazione tra due variabili

L'obiettivo delle indagini statistiche che coinvolgono più variabili è quello di studiare l'associazione tra le variabili in esame

Associazione: quando una variabile cambia il suo valore, l'altra variabile tende ad assumere certi valori

Un'analisi tra due variabili è detta bivariata

C'è associazione tra:

- tipologia familiare e situazione economica?
- sesso e retribuzione?
- numero di ore passate all'aria aperta e l'età?

Utilizzando le tabelle di contingenza è possibile osservare la distribuzione dei soggetti secondo tutte le possibili combinazioni tra le modalità di due variabili

	Orientamento Politico			
Sesso	Democratici	Indipendenti	Repubblicani	Totale
Femmine	573	516	422	1511
Maschi	386	475	399	1260
Totale	959	991	821	2771

La tabella mostra la distribuzione degli intervistati al GSS 2004, secondo il sesso e l'orientamento politico

Tabelle di contingenza

Utilizzando le tabelle di contingenza è possibile osservare la distribuzione dei soggetti secondo tutte le possibili combinazioni tra le modalità di due variabili

Unità	Sesso	Età	Statura	Colore occhi
1	F	24	163	Marroni
2	F	21	165	Azzurri
3	M	34	185	Azzurri
4	F	22	164	Marroni
5	F	21	167	Marroni
6	F	22	175	Verdi
7	M	24	178	Verdi
8	F	21	155	Marroni

Sesso	Numerosità
F	6
M	2
	8

Colore occhi	Numerosità
Azzurri	2
Marroni	4
Verdi	2

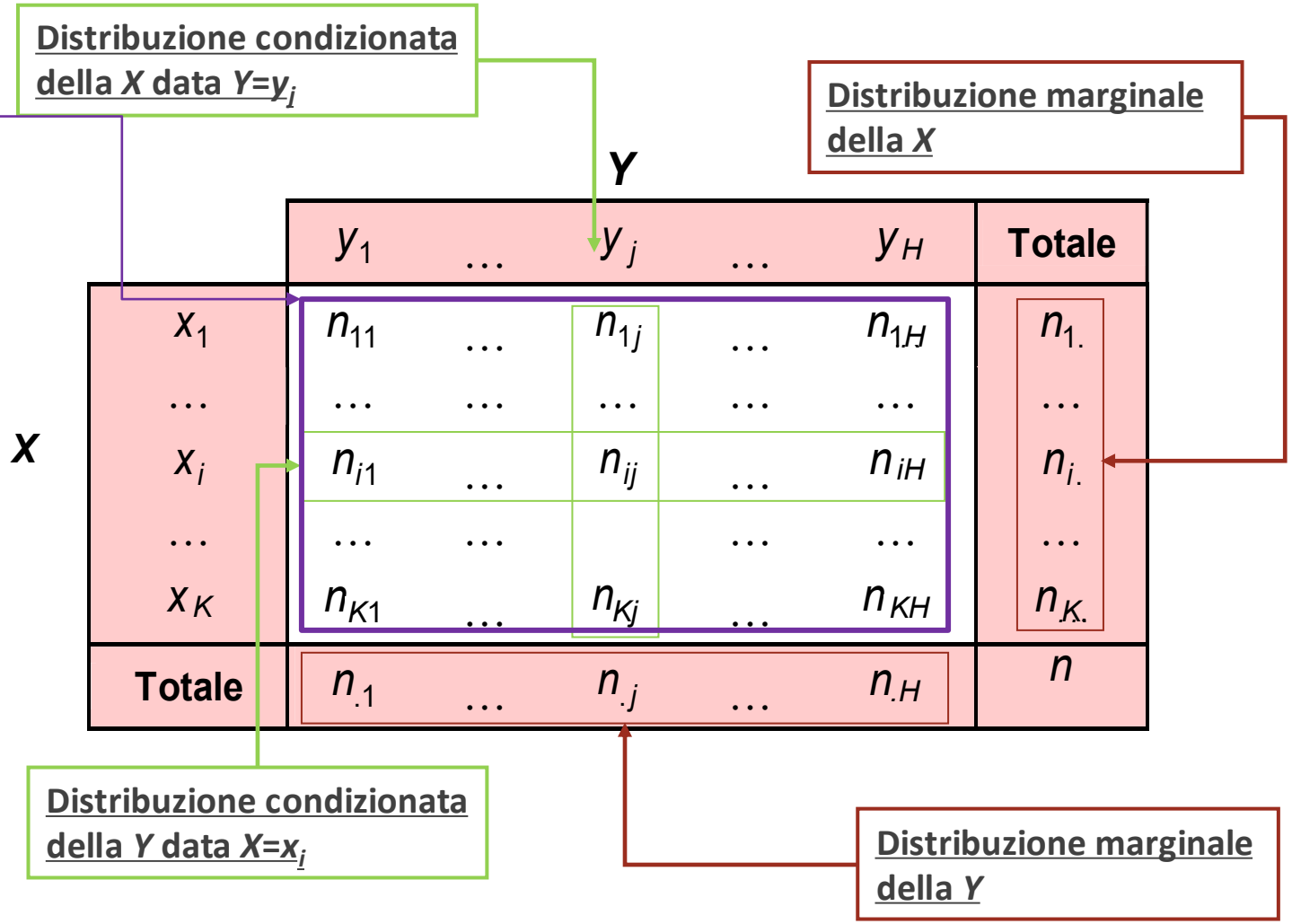
Sesso/Colore occhi	Azzurri	Marroni	Verdi	
F	1	4	1	6
M	1	0	1	2
	2	4	2	8

Dati due caratteri X e Y si definisce distribuzione doppia di frequenze l'insieme delle frequenze congiunte n_{ij}, ovvero le frequenze assolute delle unità che presentano congiuntamente la modalità *i*-esima della variabile X e la modalità j-esima della variabile Y

- Si possono costruire per tutti i tipi di variabili ma per migliorare la leggibilità della tabella è opportuno ricodificare la variabile in categorie o classi (soprattutto nel caso di caratteri quantitativi continui)

Possiamo identificare le distribuzioni marginali e le distribuzioni condizionate

- Una distribuzione marginale di una variabile corrisponde alla distribuzione di frequenza della singola variabile (totali di riga/totali di colonna)
- Una distribuzione condizionata di una variabile corrisponde alla distribuzione di frequenza di una variabile condizionata rispetto ad una o più modalità dell'altra variabile



Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	$n_{i\cdot}$
Licenza media	5	30	15	10	60
Diploma	7	42	21	14	84
Laurea triennale	3	18	9	6	36
Laurea magistrale	2	12	6	4	24
$n_{\cdot j}$	**17**	**102**	**51**	**34**	204

Distribuzione condizionata di X dato Y

"Tra chi vota Sinistra, quanti hanno la licenza media?" Lavoro sulla Colonna "Sinistra" quindi 5/17=0,294 ovvero il 29,4 % di chi vota "Sinistra" ha la licenza media

Distribuzione condizionata di Y dato X

"Tra chi ha la licenza media, quanti votano Sinistra?" Guardo la riga: 5/60=0,083 ovvero l'8,3%

Distribuzione congiunta

"Che proporzione del campione ha allo stesso tempo la licenza media **e** vota Sinistra?" Guardo la coppia (X,Y) quindi 5/204= 0,025 ovvero il 2,5%

Distribuzione doppia, marginale e condizionata

	Orientamento Politico			
Sesso	Democratici	Indipendenti	Repubblicani	Totale
Femmine	573	516	422	1511
Maschi	386	475	399	1260
Totale	959	991	821	2771

Con riferimento alla tabella, individuare

- la distribuzione congiunta della variabile Sesso e Orientamento Politico:
- la distribuzione marginale della variabile Sesso:
- la distribuzione condizionata della variabile Orientamento Politico rispetto alla modalità Femmine:
- la distribuzione condizionata della variabile Sesso rispetto alla modalità Repubblicani:

Distribuzioni doppie

Distribuzione marginale della X

$$n_{i.} = \sum_{j=1}^{H} n_{ij} = n_{i1} + n_{i2} + \cdots + n_{iH}, per\ i = 1, \ldots, K$$

Distribuzione marginale della Y

$$n_{.j} = \sum_{i=1}^{K} n_{ij} = n_{1j} + n_{2j} + \cdots + n_{Kj}, per\ j = 1, \ldots, H$$

Numerosità campionaria

$$n = \sum_{i=1}^{H}\sum_{j=1}^{K} n_{ij} = \sum_{i=1}^{K} n_{i.} = \sum_{j=1}^{H} n_{.j}$$

Distribuzione marginale del Sesso

$n_{1.} = 6$

$n_{2.} = 2$

Distribuzione marginale del Colore degli occhi

$n_{.1} = 2$

Sesso/Colore occhi	Azzurri	Marroni	Verdi	
F	1	4	1	6
M	1	0	1	2
	2	4	2	8

Frequenze doppie relative

5/204

Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	$f_{i\cdot}$
Licenza media	0,025	0,147	0,074	0,049	0,294
Diploma	0,034	0,206	0,103	0,069	0,412
Laurea triennale	0,015	0,088	0,044	0,029	0,176
Laurea magistrale	0,010	0,059	0,029	0,020	0,118
$f_{\cdot j}$	**0,083**	**0,500**	**0,250**	**0,167**	1,000

Distribuzioni marginali relative

Orientamento	$n_{\cdot j}$	$\div\, n$	$f_{\cdot j}$
Sinistra	17	17/204	**0,083**
Centro-sin.	102	102/204	**0,500**
Centro-des.	51	51/204	**0,250**
Destra	34	34/204	**0,167**
Totale	**204**		**1,000**

Titolo di studio	$n_{i\cdot}$	$\div\, n$	$f_{i\cdot}$
Licenza media	60	60/204	**0,294**
Diploma	84	84/204	**0,412**
Laurea triennale	36	36/204	**0,176**
Laurea magistrale	24	24/204	**0,118**
Totale	**204**		**1,000**

Distribuzioni condizionate Y|X

Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	Totale
Licenza media	0,083	0,500	0,250	0,167	1,000
Diploma	0,083	0,500	0,250	0,167	1,000
Laurea triennale	0,083	0,500	0,250	0,167	1,000
Laurea magistrale	0,083	0,500	0,250	0,167	1,000
Marginale Y	**0,083**	**0,500**	**0,250**	**0,167**	1,000

5/60

Condiziono sulla riga

Distribuzioni condizionate X|Y

5/17

Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	Marg. X
Licenza media	0,294	0,294	0,294	0,294	**0,294**
Diploma	0,412	0,412	0,412	0,412	**0,412**
Laurea triennale	0,176	0,176	0,176	0,176	**0,176**
Laurea magistrale	0,118	0,118	0,118	0,118	**0,118**
Totale	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**

Condiziono sulla colonna

Frequenze doppie relative: come le interpreto?

Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	$f_{i\cdot}$
Licenza media	0,025	0,147	0,074	0,049	0,294
Diploma	0,034	0,206	0,103	0,069	0,412
Laurea triennale	0,015	0,088	0,044	0,029	0,176
Laurea magistrale	0,010	0,059	0,029	0,020	0,118
$f_{\cdot j}$	**0,083**	**0,500**	**0,250**	**0,167**	1,000

Ci dice com'è distribuito il campione nelle 16 “combinazioni possibili”. Ad esempio, il 20,6% del campione ha il diploma e vota centro-sinistra.

Distribuzioni marginali relative: come le interpreto?

Orientamento	$n_{\cdot j}$	$\div\, n$	$f_{\cdot j}$
Sinistra	17	17/204	**0,083**
Centro-sin.	102	102/204	**0,500**
Centro-des.	51	51/204	**0,250**
Destra	34	34/204	**0,167**
Totale	**204**		**1,000**

Titolo di studio	$n_{i\cdot}$	$\div\, n$	$f_{i\cdot}$
Licenza media	60	60/204	**0,294**
Diploma	84	84/204	**0,412**
Laurea triennale	36	36/204	**0,176**
Laurea magistrale	24	24/204	**0,118**
Totale	**204**		**1,000**

Ci dice come si distribuisce ciascuna variabile presa singolarmente (perché guardo i totali di riga e di colonna).
Ad esempio: il 41,2% ha il Diploma, il 50% vota Centro-sinistra.

Distribuzioni condizionate Y|X: come le interpreto?

Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	Totale
Licenza media	0,083	0,500	0,250	0,167	1,000
Diploma	0,083	0,500	0,250	0,167	1,000
Laurea triennale	0,083	0,500	0,250	0,167	1,000
Laurea magistrale	0,083	0,500	0,250	0,167	1,000
Marginale Y	**0,083**	**0,500**	**0,250**	**0,167**	1,000

Condiziono sulla riga

Ci dice se l'orientamento politico varia a seconda del titolo di studio. Nel nostro caso le righe sono tutte identiche e coincidono esattamente con la distribuzione marginale di Y.

Distribuzioni condizionate X|Y: come le interpreto?

Titolo di studio	Sinistra	C.-sin.	C.-des.	Destra	Marg. X
Licenza media	0,294	0,294	0,294	0,294	**0,294**
Diploma	0,412	0,412	0,412	0,412	**0,412**
Laurea triennale	0,176	0,176	0,176	0,176	**0,176**
Laurea magistrale	0,118	0,118	0,118	0,118	**0,118**
Totale	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**

Condiziono sulla colonna

Ci dice se il titolo di studio cambia a seconda dell'orientamento politico. Le colonne sono tutte identiche e coincidono con la marginale di X. Sapere come vota una persona non cambia nulla sulla distribuzione del titolo di studio.

Tabella 6.2.3 Laureati del 2011 in lauree magistrali per condizione occupazionale nel 2015 (fonte ISTAT)

Condizione occupazionale	Gruppo di corsi di laurea			Totale
	Medico	Economico-statistico	Letterario	
Occupati	9.090	14.787	7.361	31.238
In cerca di lavoro	126	1.534	2.146	3.806
Non cercano lavoro	202	350	522	1.074
Totale	9.418	16.671	10.029	36.118

Tabella 6.2.4 Distribuzioni doppie percentuali e distribuzioni percentuali condizionate

Condizione occupazionale		Gruppo di corsi di laurea			Totale (%)
		Medico	Economico-statistico	Letterario	
Occupati	(% totale)	25,2	40,9	20,4	86,5
	(% riga)	29,1	47,3	23,6	
	(% colonna)	96,5	88,7	73,4	
In cerca di lavoro	(% totale)	0,3	4,2	5,9	10,5
	(% riga)	3,3	40,3	56,4	
	(% colonna)	1,3	9,2	21,4	
Non cercano lavoro	(% totale)	0,6	1,0	1,4	3,0
	(% riga)	18,8	32,6	48,6	
	(% colonna)	2,1	2,1	5,2	
Totale		26,1	46,2	27,8	100,00

Distribuzione percentuale congiunta

Distribuzione condizionata della condizione occupazionale rispetto al gruppo di laurea

Distribuzione condizionata del gruppo di laurea rispetto alla condizione occupazionale

Distribuzione marginale percentuale

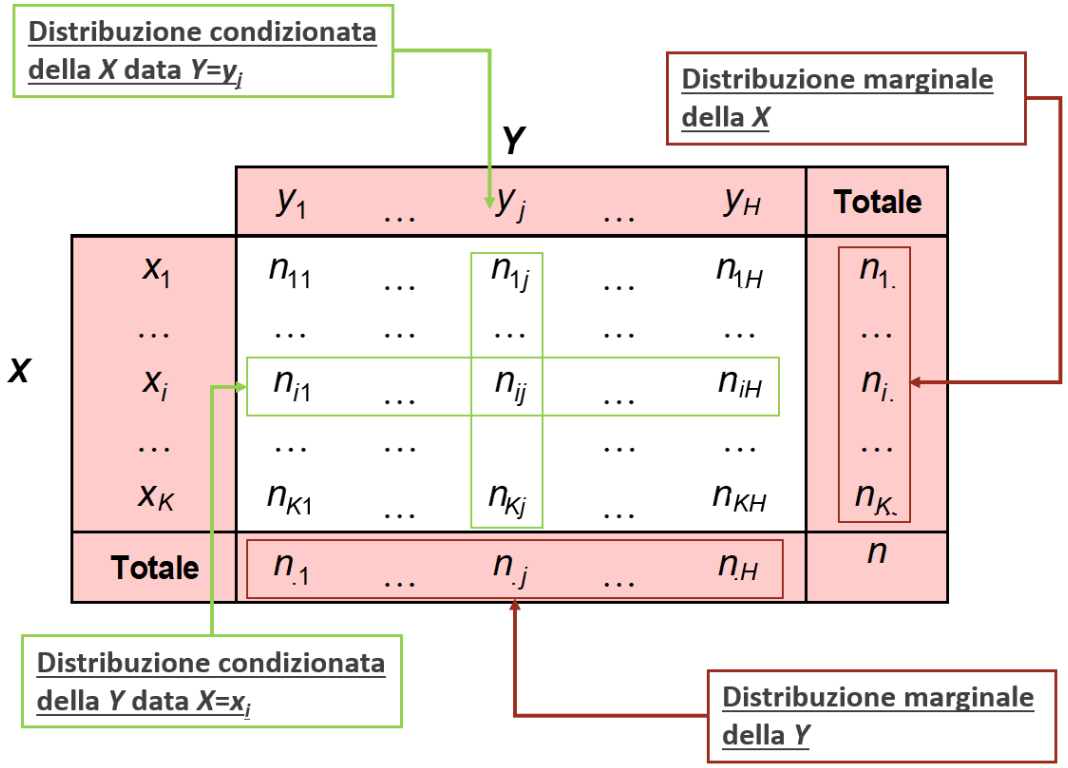
Medie condizionate

Per ogni distribuzione condizionata di un carattere quantitativo si può calcolare la media aritmetica condizionata

$$\bar{x}_{Y=y_j} = \frac{1}{n_{.j}} \sum_{i=1}^{K} x_i n_{ij} \qquad \bar{y}_{X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^{H} y_j n_{ij}$$

Il numero di medie condizionate aumenta all'aumentare delle modalità del carattere

Nel caso di carattere quantitativo suddiviso in classi si può calcolare la media aritmetica approssimata utilizzando il valore centrale di ogni classe



$$\bar{y}_{X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^{H} y_j n_{ij}$$

		Y Numero di case			
		1	**2**	**3**	
X	**1**	21	8	0	29=n_1.
Numero di auto	**2**	12	11	1	24
	3	7	6	2	15
		40	25	3	68

$\bar{y}_{X=1} = 1{,}28$
$\bar{y}_{X=2} = 1{,}54$
$\bar{y}_{X=3} = 1{,}67$

$\bar{x}_{Y=1} = 1{,}65$
$\bar{x}_{Y=2} = 1{,}92$
$\bar{x}_{Y=3} = 2{,}67$

$$\bar{y} = \frac{1}{68}(1 * 40 + 2 * 25 + 3 * 3) = 1{,}46$$

$$\bar{x} = \frac{1}{68}(1 * 29 + 2 * 24 + 3 * 15) = 1{,}79$$

$$\bar{y}_{X=1} = \frac{1}{29}(1 * 21 + 2 * 8 + 3 * 0) = 1{,}28$$

$$\bar{y}_{X=2} = \frac{1}{24}(1 * 12 + 2 * 11 + 3 * 1) = 1{,}54$$

$$\bar{x}_{Y=1} = \frac{1}{40}(1 * 21 + 2 * 12 + 3 * 7) = 1{,}65$$

Calcola la media condizionata della variabile numero di case rispetto alle modalità 3 della variabile numero di auto:

Calcola le medie condizionate della variabile numero di auto rispetto alle modalità 2 e 3 della variabile numero di case: