

**Università di Cassino**  
**Corso di Statistica 1**  
**Esercitazione del 15/10/2007**  
**Dott. Alfonso Piscitelli**

**Esercizio 1**

Il seguente *data set* riporta la rilevazione di alcuni caratteri su un collettivo di 20 soggetti.

Soggetto	Sesso	Età	Reddito (Migliaia di €)	Titolo di studio	Nucleo familiare	Statura (cm)	Colore degli occhi
1	M	22	0,7	Diploma	3	173	NERO
2	F	18	0,2	Lic. Media	4	168	MARRONE
3	F	34	1,6	Diploma	2	165	MARRONE
4	M	42	2,5	Laurea	5	180	NERO
5	F	50	3,2	Diploma	3	163	AZZURRO
6	F	12	0,1	Lic. Elementare	4	160	NERO
7	M	46	3,8	Lic. Media	4	177	MARRONE
8	M	72	1,3	Nessun Titolo	2	164	VERDE
9	F	27	1,2	Laurea	3	158	AZZURRO
10	F	48	1,7	Lic. Media	5	170	NERO
11	F	35	1,9	Laurea	1	167	NERO
12	M	84	0,8	Nessun Titolo	1	159	MARRONE
13	F	21	0,4	Diploma	5	174	AZZURRO
14	F	44	1,8	Diploma	4	164	VERDE
15	M	56	1,9	Lic. Media	2	177	NERO
16	F	58	3,2	Lic. Media	3	172	NERO
17	F	37	2,1	Diploma	1	166	MARRONE
18	F	16	0,1	Lic. Media	4	160	MARRONE
19	M	73	1,6	Lic. Elementare	2	170	AZZURRO
20	M	64	2,2	Lic. Elementare	3	184	VERDE

**a.** Costruire la distribuzione di frequenza per il carattere **Sesso** e rappresentarla graficamente con un aerogramma.

**b.** Costruire la distribuzione di frequenza per il carattere **Colore degli occhi** e rappresentarla graficamente prima con un diagramma a barre e poi con il diagramma di Pareto.

**c.** Rappresentare graficamente il carattere **Nucleo familiare** con un diagramma a punti. Successivamente, a partire dalla distribuzione di frequenza costruire un diagramma a linee.

**d.** Costruire la distribuzione di frequenza per il carattere **Età** suddividendo la distribuzione in 4 classi equiampie e rappresentarla graficamente.

**e.** Costruire la distribuzione di frequenza per il carattere **Reddito** suddividendo la distribuzione in 5 classi equifrequenti e rappresentarla graficamente.

f. Costruire la distribuzione doppia di frequenza per i caratteri **Sesso** e **Età**, utilizzando per questo ultimo carattere la suddivisione in classi equiampie operata precedentemente.

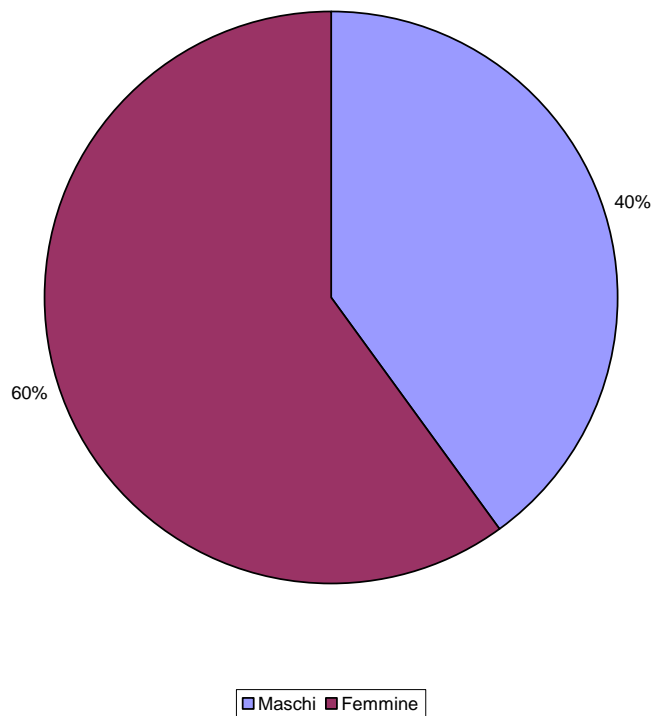
### Soluzioni

a) La **distribuzione di frequenza semplice** associa alle modalità o intensità che può assumere un carattere X, le corrispondenti frequenze assolute. Per costruire una distribuzione di frequenza semplice partendo da una successione di osservazioni è sufficiente assegnare ad ogni modalità la propria frequenza assoluta.

Quindi la distribuzione di frequenze del carattere (variabile) **Sesso** è:

Sesso	$n_i$	$f_i$
Maschi	8	0,4
Femmine	12	0,6
<b>Tot:</b>	<b>20</b>	<b>1</b>

Un **aerogramma** o **diagramma a torta** rappresenta i dati sottoforma di fette o sezioni di un cerchio. Ogni fetta rappresenta una categoria e la dimensione della fetta è proporzionale alla frequenza relativa della categoria

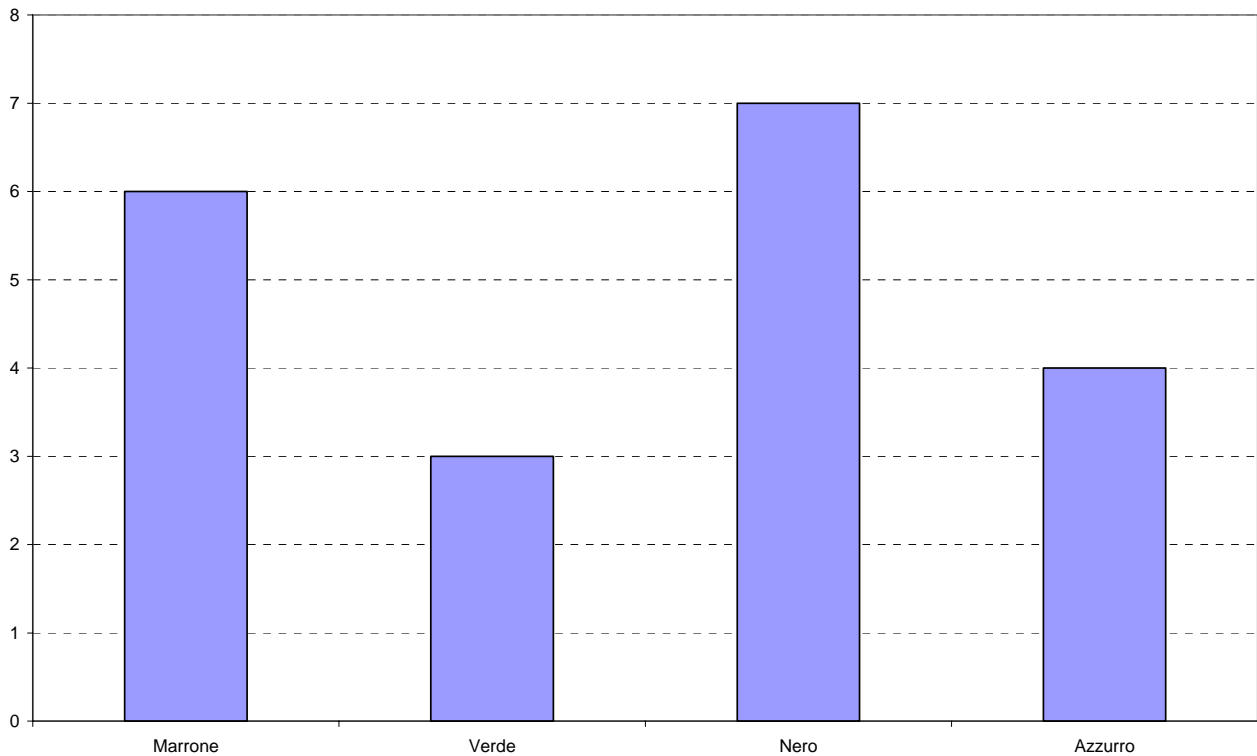


b) La distribuzione di frequenze associata alla variabile **Colore degli occhi** è:

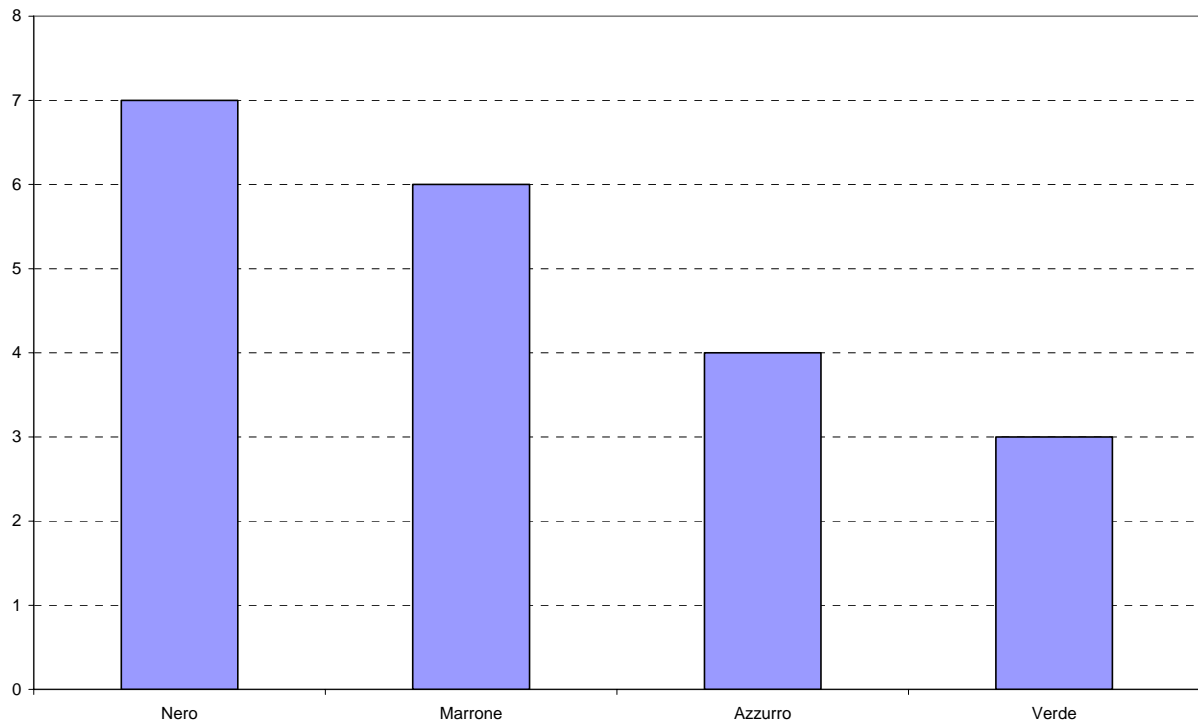
Colore degli occhi	$n_i$
Nero	7
Marrone	6
Azzurro	4
Verde	3

**Tot: 20**

Il **diagramma a barre** rappresenta la frequenza o la frequenza relativa di una tabella di frequenza sottoforma di un rettangolo o barra o colonna.

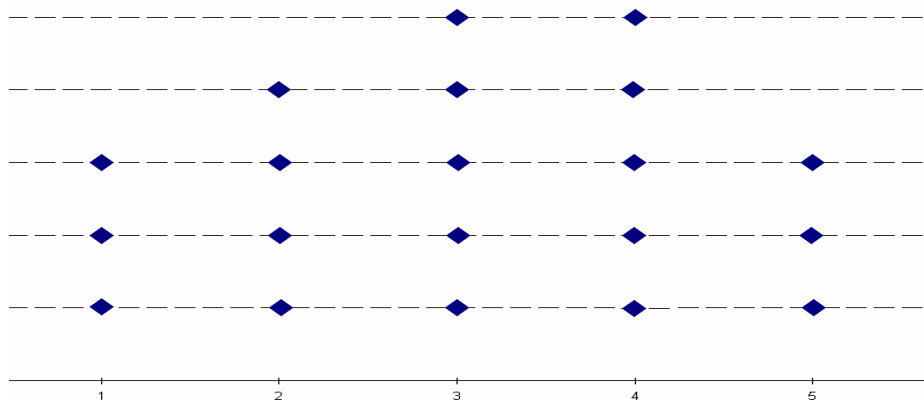


Il **diagramma di Pareto** presenta le stesse caratteristiche del diagramma a barre con l'unica differenza nell'ordine delle barre: le modalità sono ordinate in senso decrescente rispetto alle frequenze



c) A partire dalla successione dei valori del carattere **Nucleo familiare** è possibile effettuare un diagramma a punti.

Il **diagramma a punti** è un grafico usato per piccole quantità di dati in cui ogni osservazione è indicata da un punto su un singolo asse orizzontale. L'asse del grafico è suddiviso in modo tale che ogni punto, rappresentante i dati, occupi un posto univoco sull'asse. Quando più d'una osservazione presenta lo stesso valore, i punti vengono "impilati" l'uno sopra l'altro.

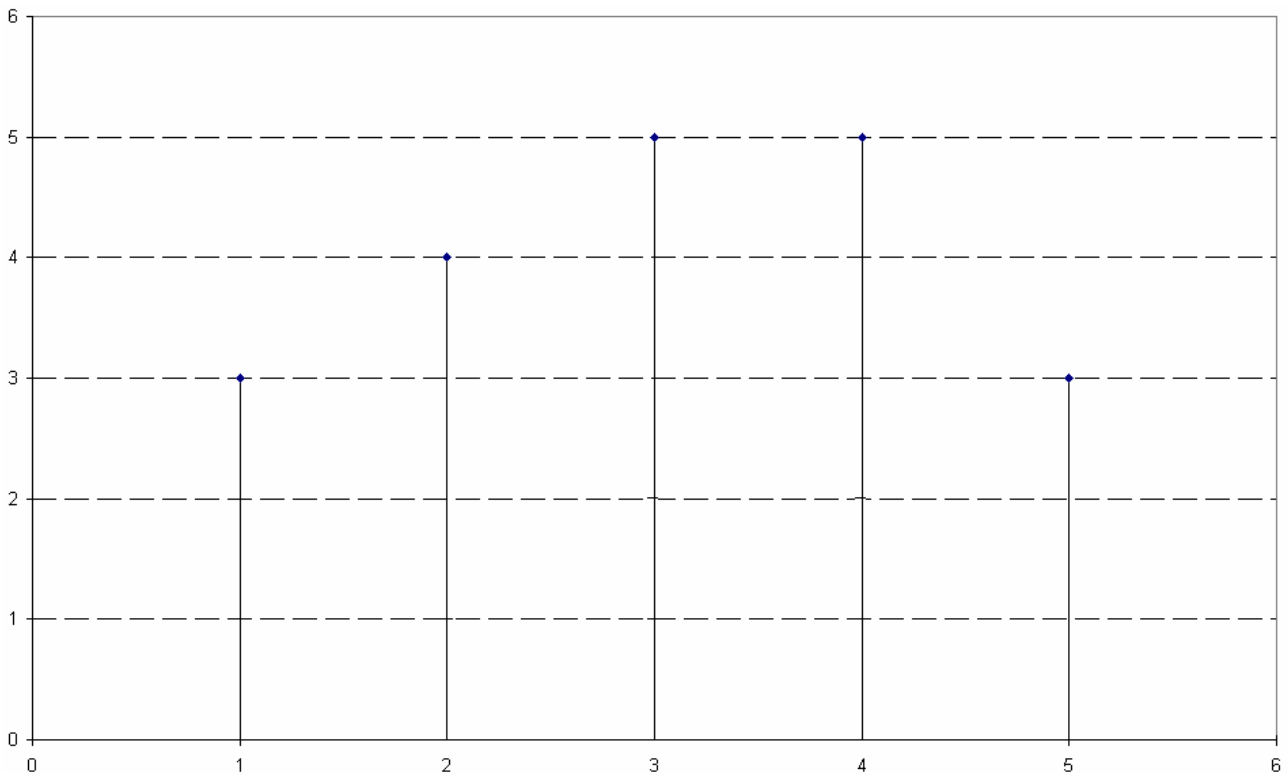


Infine, a partire dalla distribuzione di frequenze della variabile **Nucleo familiare** è possibile realizzare un diagramma a linee.

Nucleo familiare	$n_i$
1	3
2	4
3	5
4	5
5	3

**Tot: 20**

Per costruire il **diagramma a linee** si riportano sull'asse delle ascisse le modalità (intensità) e sull'asse delle ordinate i corrispettivi valori di frequenza, si tracciano quindi i punti e i segmenti corrispondenti.



E' opportuno utilizzare il diagramma a linee quando le modalità (intensità) rappresentano intervalli temporali. Infatti, il diagramma a linee è particolarmente usato quando si vuole enfatizzare il trend nel tempo.

d) Dopo aver suddiviso la distribuzione del carattere **Età** in classi di uguale ampiezza,

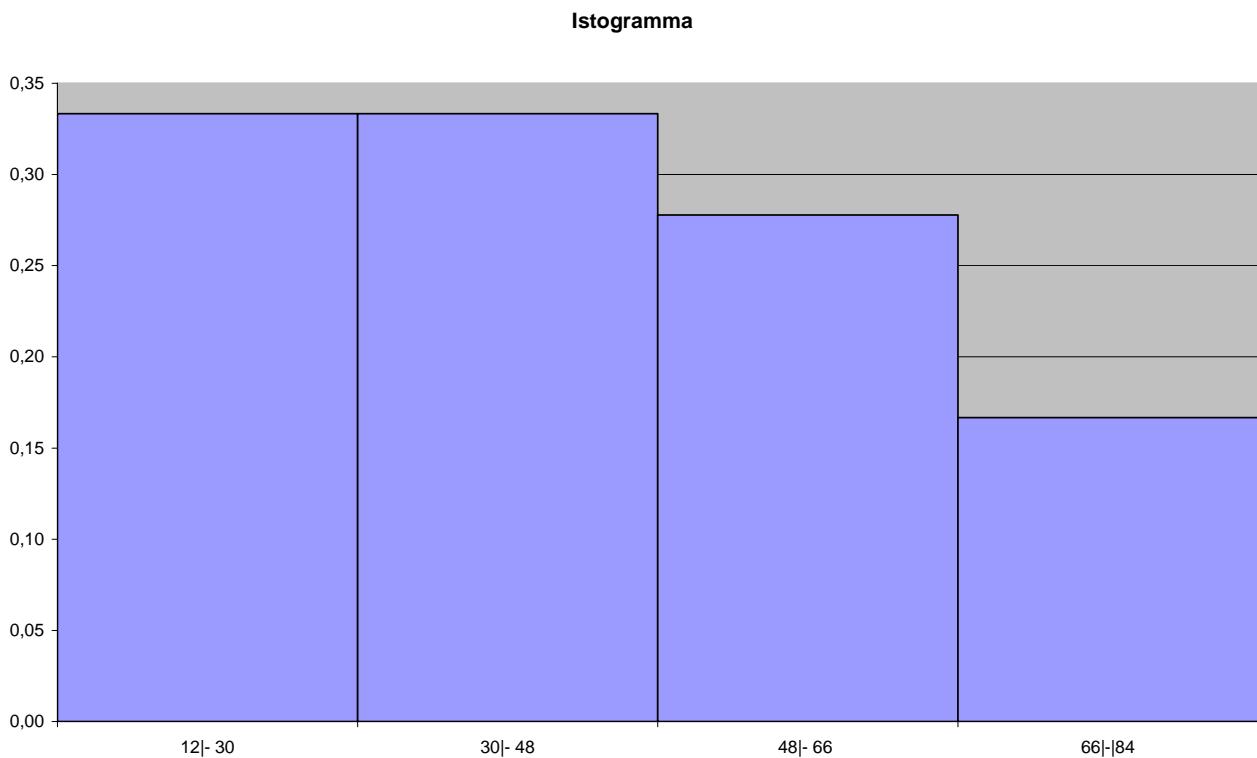
Età	$d_i$	$n_i$	$f_i$	$F_i$	$h_i$
12  - 30	18	6	0,30	0,30	0,33
30  - 48	18	6	0,30	0,60	0,33
48  - 66	18	5	0,25	0,85	0,28
66  - 84	18	3	0,15	1	0,17

**Tot: 20 1**

per poter costruire l'istogramma, è necessario calcolare la densità di frequenza, cioè il rapporto tra la frequenza della i-esima classe e la sua ampiezza:  $h_i = \frac{n_i}{d_i}$ .

Dopodichè è possibile realizzare **l'istogramma** associato alla nostra distribuzione di frequenze.

L'istogramma normalizzato presenta sulle ascisse gli estremi delle classi e sulle ordinate la densità di frequenza. Essendo le classi di ampiezza uguale le basi dei rettangoli associati alle diverse classi avranno tutte la stessa lunghezza.



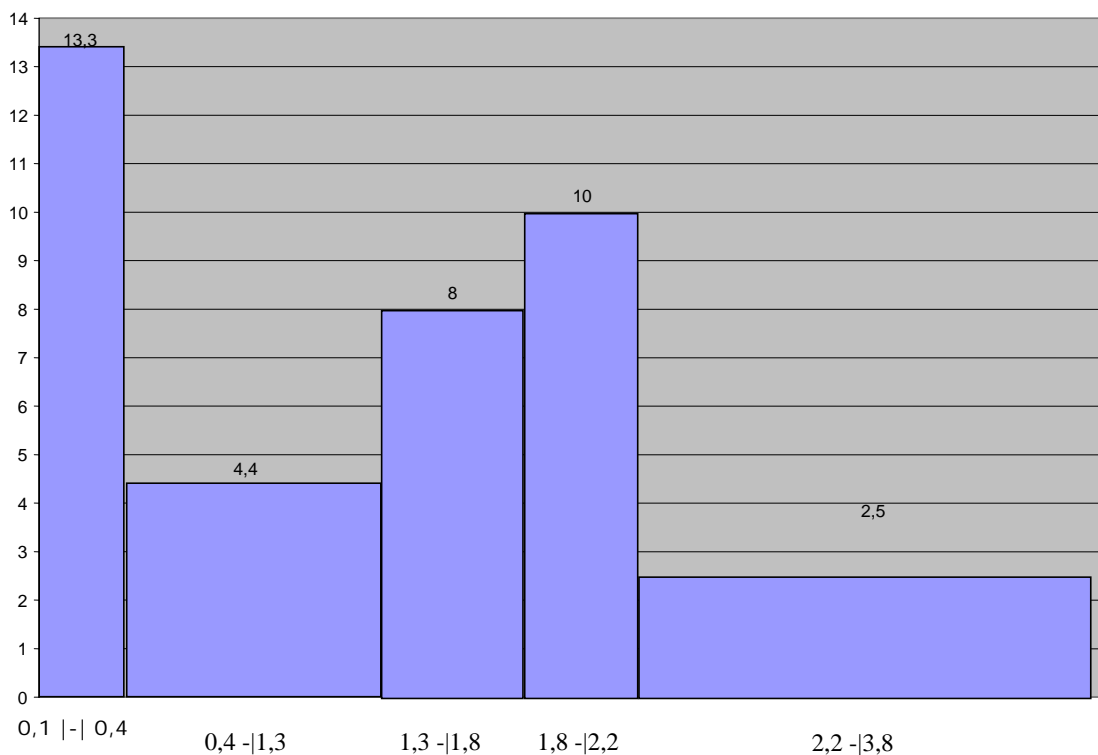
**Si precisa** che nel caso di classi equiampie l'informazione ottenuta dalla densità di frequenza è la stessa che avremmo analizzando la distribuzione di frequenze semplice.

e) Dopo aver suddiviso la distribuzione del carattere **Reddito** in classi di uguale frequenza,

Reddito	$d_i$	$n_i$	$f_i$	$F_i$	$h_i$
0,1   -   0,4	0,3	4	0,02	0,02	13,3
0,4 -   1,3	0,9	4	0,02	0,04	4,4
1,3 -   1,8	0,5	4	0,02	0,06	8,0
1,8 -   2,2	0,4	4	0,02	0,08	10,0
2,2 -   3,8	1,6	4	0,02	1	2,5
Tot:		20	1		

soprattutto in questo caso per poter realizzare l'istogramma, è necessario calcolare la densità di frequenza.

Sulla base dei risultati ottenuti costruire **l'istogramma** associato alla nostra distribuzione di frequenze. Essendo le classi di ampiezza diversa le basi dei rettangoli associati alle diverse classi avranno lunghezza diversa, in particolare maggiore è l'ampiezza della classe più deve essere lunga la base del rettangolo associato a quella classe.



f) Si procede alla suddivisione in quattro classi della variabile **Età**

<b>Età</b>	<b>d<sub>i</sub></b>
12  - 30	18
30  - 48	18
48  - 66	18
66  -  84	18

Per rappresentare la distribuzione doppia di frequenze dei due caratteri **Sesso** e **Età** (suddiviso in classi) è necessario costruirsi una tabella a doppia entrata che ha per righe le due modalità della variabile **Sesso** {Maschio, Femmina} e per colonne le 4 classi in cui è stata suddivisa la variabile **Età** {12|-30; 30|-48; 48|-66; 66|-|84 }.

Si precisa che la stessa informazione si avrebbe da una tabella che ha per righe le classi della variabile **Età** e per colonne le modalità della variabile **Sesso**.

La distribuzione doppia di frequenza delle due variabili è:

	12 -30	30 -48	48 -66	66 - 84	Tot:
Maschio	1	2	2	3	8
Femmina	5	4	3	0	12
Tot:	6	6	5	3	20