



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Dipartimento di Comunicazione
ed Economia

La rappresentazione dei dati - Distribuzioni



UNIMORE Scienze
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA della comunicazione

Analisi dei dati per la ricerca sociale

A. A. 2021/22

Elvira Pelle

Frequenze relative

Dividendo una frequenza assoluta per il numero totale di unità statistiche (n nel nostro caso) otteniamo le cosiddette *frequenze relative* (f_j), ovvero

$$\left(\begin{array}{c} \text{frequenze} \\ \text{relative} \end{array} \right) = \frac{\left(\begin{array}{c} \text{frequenze} \\ \text{assolute} \end{array} \right)}{\left(\begin{array}{c} \text{numero totale di} \\ \text{osservazioni} \end{array} \right)} \text{ cioè } f_j = \frac{n_j}{n}$$

Hanno il vantaggio, rispetto alle frequenze assolute, di permettere di confrontare le distribuzioni di due o più collettivi (con numerosità diversa)



Esercizio: Numero di esami in un a.a. in due collettivi

X	n_j^A	X	n_j^B
0	6	0	13
1	16	1	23
2	18	2	35
3	12	3	17
4	5	4	12
5	3	5	10
Totale	60	Totale	110

- ▶ È corretto affermare che coloro che hanno sostenuto 1 esami sono più numerosi nel collettivo A?
- ▶ È corretto affermare che coloro che hanno sostenuto 3 esami sono più numerosi nel collettivo B?



Esercizio: Numero di esami in un a.a. in due collettivi

X	n_j^A	f_j^A
0	6	0.1
1	16	0.27
2	18	0.3
3	12	0.2
4	5	0.08
5	3	0.05
Totale	60	1

X	n_j^B	f_j^B
0	13	0.12
1	23	0.21
2	35	0.32
3	17	0.15
4	12	0.11
5	10	0.09
Totale	110	1



Frequenze cumulate

N.B.: Ha senso calcolarle se il carattere è almeno ordinale!

La *frequenza assoluta (relativa) cumulata* per la modalità/classe x_j è la somma delle frequenze assolute (relative) per le modalità/classi $\leq x_j$

$$N_j = n_1 + \dots + n_j$$

$$(F_j = f_1 + \dots + f_j)$$

modalità/classe	frequenze cumulate assolute	frequenze cumulate relative
x_1	$N_1 = n_1$	$F_1 = f_1$
x_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
\vdots	\vdots	\vdots
x_j	$N_j = n_1 + \dots + n_j$	$F_j = f_1 + \dots + f_j$
\vdots	\vdots	\vdots
x_K	n	?



Esercizio: Numero di libri letti

Si costruisca la distribuzione di frequenze cumulate per il carattere Numero di libri letti nell'ultimo anno e se ne interpreti quella relativa alla modalità 3

Libri	n_j
0	1
1	3
2	3
3	2
4	5
5	7
6	3
7	3
8	5



Esercizio: Numero di libri letti

Si costruisca la distribuzione di frequenze cumulate per il carattere Numero di libri letti nell'ultimo anno e se ne interpreti quella relativa alla modalità 3

Soluzione. Partendo dalla distribuzione di frequenze assolute, abbiamo

Libri	n_j	N_j
0	1	1
1	3	4
2	3	7
3	2	9
4	5	14
5	7	21
6	3	24
7	3	27
8	5	32



Esercizio: Durata di una lampadina

Si costruisca la distribuzione di frequenze cumulate relative per il carattere Durata di una lampadina (in giorni) e se ne interpreti quella relativa alla seconda classe

Durata	n_j
[0; 75)	11
[75; 190)	32
[190; 365)	45
[365; 680)	82



Esercizio: Durata di una lampadina

Si costruisca la distribuzione di frequenze cumulate relative per il carattere Durata di una lampadina (in giorni) e se ne interpreti quella relativa alla seconda classe

Soluzione. Partendo dalla distribuzione di frequenze assolute, calcoliamo innanzitutto le frequenze relative

Durata	n_j	f_j
[0; 75)	11	0.065
[75; 190)	32	0.188
[190; 365)	45	0.265
[365; 680)	82	0.482
Totale	170	1



Esercizio: Durata di una lampadina

Si costruisca la distribuzione di frequenze cumulate relative per il carattere Durata di una lampadina (in giorni) e se ne interpreti quella relativa alla seconda classe

Soluzione. Calcoliamo ora le frequenze cumulate relative

Durata	n_j	f_j	F_j
[0; 75)	11	0.065	0.065
[75; 190)	32	0.188	0.253
[190; 365)	45	0.265	0.518
[365; 680)	82	0.482	1
Totale	170	1	





UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Dipartimento di Comunicazione
ed Economia

La rappresentazione dei dati - Le rappresentazioni grafiche



UNIMORE Scienze
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA
della comunicazione

Analisi dei dati per la ricerca sociale

A. A. 2021/22

Elvira Pelle

Rappresentazioni grafiche delle distribuzioni di frequenza



Finalmente un grafico!

Possiamo cercare di visualizzare le distribuzioni di frequenza, rappresentando in qualche modo ciascuna modalità del carattere con la relativa frequenza.

Esempio: Ad un gruppo di 246 studenti è stato chiesto il loro rapporto rispetto all'ultima edizione di Sanremo.

Modalità	freq. ass.	freq. rel.
Non visto	169	0.69
Visto e non piaciuto	12	0.05
Visto e piaciuto	65	0.26
Totale	246	1



Diagramma a torta - Carattere qualitativo

Il grafico è costruito rappresentando ogni modalità con una fetta di torta di superficie proporzionale alla sua frequenza:

$$\text{angolo} = 360 \cdot \text{frequenza assoluta}/n$$

oppure

$$\text{angolo} = 360 \cdot \text{frequenza relativa}$$



Sanremo: Diagramma a torta

Esempio: Ad un gruppo di 246 studenti è stato chiesto il loro rapporto rispetto all'ultima edizione di Sanremo.

Modalità	freq. ass.	freq. rel.
Non visto	169	0.69
Visto e non piaciuto	12	0.05
Visto e piaciuto	65	0.26
Totale	246	1

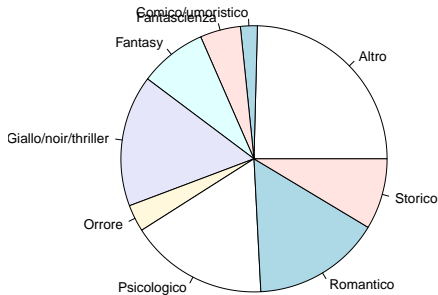


Esempio: distribuzione di frequenza del *genere letterario preferito*

Modalità	Frequenza assoluta
Altro	60
Comico/umoristico	5
Fantascienza	12
Fantasy	20
Giallo/noir/thriller	39
Orrore	8
Psicologico	41
Romantico	38
Storico	21
Totale	244



Esempio: generi letterari preferiti



Notiamo che, se la variabile non è ordinale, l'ordine delle modalità nel grafico è arbitrario.

Grafico a barre (o a nastri) - Carattere qualitativo o quantitativo discreto

- Il grafico è costruito rappresentando ogni modalità con una barra verticale (nastro orizzontale) di altezza (lunghezza) proporzionale alla sua frequenza (assoluta o relativa)
- Le barre (nastri) sono equidistanziate ed hanno tutte la stessa base (altezza)
- Se il carattere rappresentato è qualitativo ordinale o quantitativo (discreto!), le modalità verranno rappresentate rispettando l'ordine



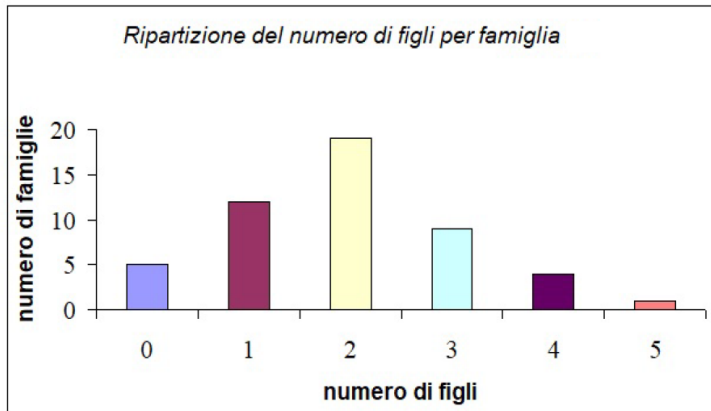
Esempio: *Il numero di figli*

Su un collettivo di 50 famiglie è stato rilevato il carattere *numero di figli*

N figli (X)	n_j	f_j
0	5	0.1
1	12	0.24
2	19	0.38
3	9	0.18
4	4	0.08
5	1	0.02
Totale	50	1



Esempio: Il numero di figli



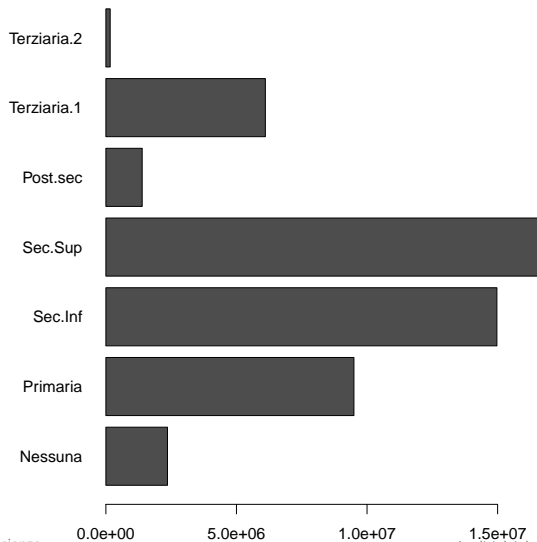
Esempio: titoli di studio

In base al censimento 2011, la distribuzione di frequenza assoluta del *titolo di studio* per la popolazione italiana sopra i 15 anni è

Titolo	Frequenza Italy
Nessuna	2358195
Primaria	9498111
Sec.Inf	14973047
Sec.Sup	16617577
Post.sec	1389813
Terziaria.1	6106337
Terziaria.2	164622



Esempio: titoli di studio



Istogramma - Carattere quantitativo continuo

Si costruiscono, l'uno adiacente all'altro, tanti rettangoli quante sono le classi in cui è suddiviso il collettivo, ponendo

(base rettangoli) = (ampiezza della classe)

(area rettangoli) \propto (frequenze assolute (relative))

Il simbolo \propto significa “proporzionale a”.



Istogramma - Carattere quantitativo continuo

Si costruiscono, l'uno adiacente all'altro, tanti rettangoli quante sono le classi in cui è suddiviso il collettivo, ponendo

(base rettangoli) = (ampiezza della classe)

(area rettangoli) \propto (frequenze assolute (relative))

Il simbolo \propto significa “proporzionale a”.

Dobbiamo, a questo punto, distinguere 2 casi:

1. le classi hanno tutte la stessa ampiezza
2. le classi hanno ampiezze diverse



Istogramma - Classi con la stessa ampiezza

Essendo l'area dei rettangoli uguale a $\text{base} \times \text{altezza}$, se le gli intervalli hanno uguale ampiezza, di fatto l'altezza coincide con (o è proporzionale a) la frequenza assoluta (relativa):

$$(\text{altezza rettangoli}) = (\text{frequenze assolute})$$

o, alternativamente

$$(\text{altezza rettangoli}) = (\text{frequenze relative})$$

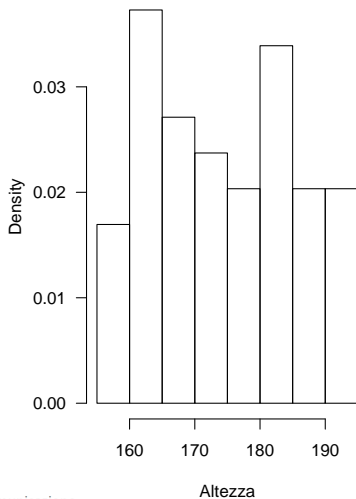


Esempio: altezze

Altezze (X)	n_j	a_j
[155,160]	26	5
(160,165]	36	5
(165,170]	40	5
(170,175]	32	5
(175,180]	41	5
(180,185]	36	5
(185,190]	25	5
(190,195]	7	5



Esempio: altezze



Istogramma - Classi con ampiezza diversa

Dobbiamo calcolare l'altezza del rettangolo in modo tale che la sua area coincida con (sia è proporzionale a) la frequenza assoluta (relativa)

$$(\text{base rettangoli}) = (\text{ampiezza della classe } a_j)$$

$$(\text{altezza rettangoli}) = (\text{densità di frequenza } d_j)$$



Perchè le densità?

Vogliamo che l'area dei rettangoli sia proporzionale alla frequenza (assoluta o relativa) della classe cui si riferisce. Sappiamo che:

- l'area del rettangolo si calcola con la formula $\text{area} = \text{base} \times \text{altezza}$
- la base di ciascun rettangolo è l'ampiezza della classe (a_i)
- l'area dev'essere uguale alla frequenza (n_i , oppure f_i , a seconda se decidiamo di rappresentare le frequenze assolute o relative)

quindi, otteniamo:
$$\underbrace{\text{area}}_{\text{freq}} = \underbrace{\text{base}}_{a_i} \times \underbrace{\text{altezza}}_{h_i(?)}$$

con le frequenze assolute otteniamo

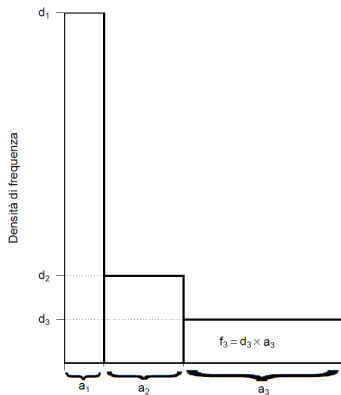
$$n_i = a_i \times h_i \Rightarrow h_i = \frac{n_i}{a_i} \leftarrow \text{densità di frequenza assolute}$$

analogamente, con le frequenze relative

$$f_i = a_i \times h_i \Rightarrow h_i = \frac{f_i}{a_i} \leftarrow \text{densità di frequenza relative}$$



Costruzione di un istogramma

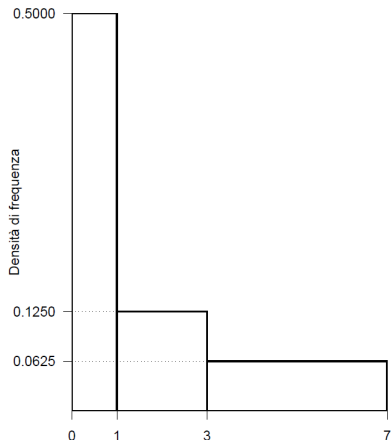


	a_i	n_i	f_i	d_i
[0;1)	1	20	0.50	0.500
[1;3)	2	10	0.25	0.125
[3;7)	4	10	0.25	0.062

$$f_i = \frac{n_i}{\sum_{i=1}^3 n_i}$$

$$d_i = \frac{f_i}{a_i}$$

Costruzione di un istogramma



	a_j	n_j	f_j	d_j
[0;1)	1	20	0.50	0.500
[1;3)	2	10	0.25	0.125
[3;7)	4	10	0.25	0.062

$$f_j = \frac{n_j}{\sum_{i=1}^3 n_i}$$

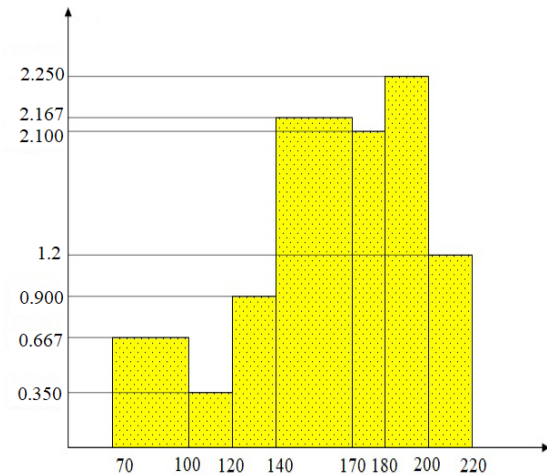
$$d_j = \frac{f_j}{a_j}$$

Esempio: altezze 2

Altezze (X)	n_j	a_j	d_j
[70;100)	20	30	0,667
[100;120)	7	20	0,350
[120;140)	18	20	0,900
[140;170)	65	30	2,167
[170;180)	21	10	2,100
[180;200)	45	20	2,250
[200;220)	24	20	1,2



Esempio: altezze



Grafici a barre multiple

Vengono utilizzati per mettere a confronto le distribuzioni dello stesso carattere su collettivi differenti



Titoli di studio

Potrei essere interessato a confronti tra paesi

Titolo	Frequenza			
	Italy	Sweden	France	Spain
Nessuna	2358195	0	943760	4381305
Primaria	9498111	813963	8309752	5873435
Sec.Inf	14973047	1048392	9952880	11208975
Sec.Sup	16617577	3348458	20902738	7775165
Post.sec	1389813	436909	0	0
Terziaria.1	6106337	1916300	12567952	9815605
Terziaria.2	164622	73725	223497	268390

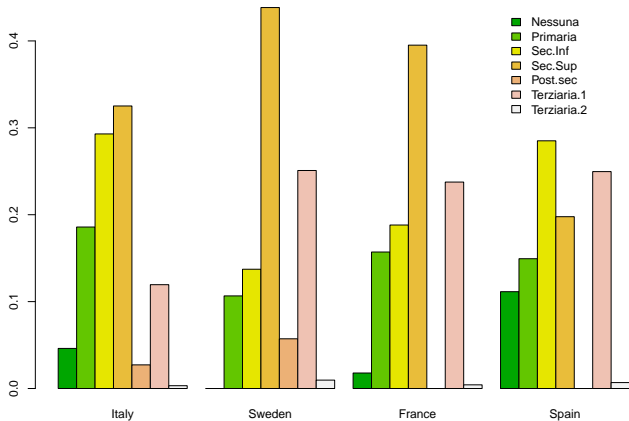


Titoli di studio

Per rendere possibile il confronto si passa alle frequenze relative

Titolo	Frequenza			
	Italy	Sweden	France	Spain
Nessuna	0.0461	0.0000	0.0178	0.1114
Primaria	0.1858	0.1066	0.1571	0.1494
Sec.Inf	0.2930	0.1373	0.1881	0.2850
Sec.Sup	0.3251	0.4384	0.3951	0.1977
Post.sec	0.0272	0.0572	0.0000	0.0000
Terziaria.1	0.1195	0.2509	0.2376	0.2496
Terziaria.2	0.0032	0.0097	0.0042	0.0068





Titoli di studio, osservazioni

- ▶ Se la domanda è “in quali paesi si studia di più”, le frequenze assolute non consentono un confronto agevole perché le popolazioni di riferimento sono molto diverse.
- ▶ Si passa allora alle frequenze relative per ciascun paese.
- ▶ Il confronto può essere fatto affiancando dei diagrammi a barre, il modo in cui le si affianca mette in evidenza cose diverse.
- ▶ Le barre possono anche essere sovrapposte per mettere in luce le diverse composizioni delle popolazioni.



Esempio: titoli di studio

