

SCHEDA DI APPROFONDIMENTO

L'elaborazione dei dati continui

Per quanto riguarda l'analisi di un'indagine statistica costituita da dati continui, l'elaborazione dei dati, proprio perché i valori appartengono all'insieme R , deve prevedere sostanzialmente gli stessi passi del paragrafo 3.2. L'unica avvertenza è che, non essendo possibile operare sui singoli valori, si devono sistemare i dati in **classi**, cioè si devono raggruppare in tanti piccoli sottoinsiemi disgiunti. Vediamo come procedere con un esempio. Consideriamo un'indagine relativa all'altezza degli alunni di una certa classe di una scuola. Una volta eseguita la raccolta dei dati (espressi in metri), osserviamo che si tratta di valori appartenenti ad un intervallo continuo.

1,50	1,22	1,45	1,55	1,48	1,60	1,50	1,48	1,63
1,40	1,48	1,53	1,46	1,38	1,35	1,55	1,45	1,48
1,50	1,44	1,47	1,54	1,49	1,42	1,36	1,35	1,33

Le altezze degli alunni della nostra popolazione statistica sono infatti comprese tra 1,22 m (valore minimo) e 1,63 m (valore massimo). Per semplicità consideriamo come intervallo i valori da 1,20 m a 1,70 m, e pertanto il **campo di variabilità** avrà un'ampiezza di $(1,70 - 1,20) \text{ m} = 0,50 \text{ m} = 50 \text{ cm}$.

Dobbiamo dunque suddividere tale intervallo in sottointervalli, detti **classi**, di ampiezza costante. Nel nostro esempio consideriamo intervalli costanti di 10 cm (segui il procedimento sulla tabella seguente).

Bisogna ora contare quante unità statistiche del campione presentano un'altezza compresa entro ogni sottointervallo. Il numero che ne risulta è la **frequenza assoluta**. Dobbiamo fare attenzione al fatto che il valore massimo di ogni classe coincide con il valore minimo della classe successiva. Stabiliamo di considerare il valore minimo incluso, mentre quello massimo escluso; la classe $1,30 \div 1,40$, ad esempio, va intesa da 1,30 m compresi fino a 1,40 m esclusi.

Dopo aver raggruppato i dati raccolti in classi, per meglio interpretarli e comprendere l'entità di una classe rispetto ad un'altra, dobbiamo determinare anche le **frequenze relative** che si ottengono calcolando il quoziente fra la frequenza assoluta ed il numero totale delle osservazioni. Occorre infine calcolare le **frequenze percentuali**, ottenute moltiplicando per 100 quelle relative.

IL LINGUAGGIO DELLA MATEMATICA

Il simbolo \div significa «da a».

Classe	Altezze	Freq. Assoluta	Freq. Relativa	Freq. Percentuale
1	1,20 ÷ 1,30	1	0,04	4%
2	1,30 ÷ 1,40	5	0,19	19%
3	1,40 ÷ 1,50	12	0,44	44%
4	1,50 ÷ 1,60	7	0,26	26%
5	1,60 ÷ 1,70	2	0,07	7%
Totale		27	1	100%

Continuando l'elaborazione dei dati è importante definire la **classe modale**, che è quella dove si concentrano il maggior numero di osservazioni. Nel nostro caso la classe modale è $1,40 \div 1,50$ con ben 12 casi.

Per il calcolo della **mediana** dobbiamo disporre i dati in ordine crescente come nella seguente tabella:

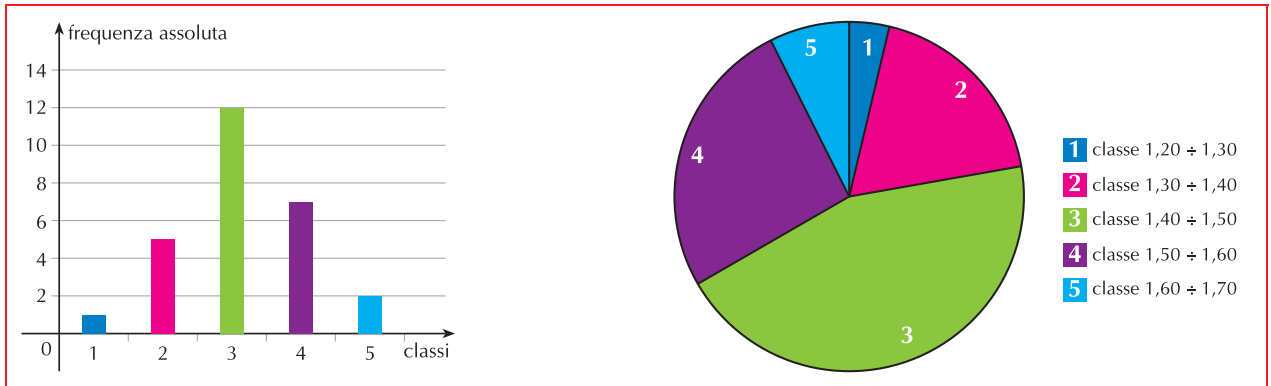
1,22	1,33	1,35	1,35	1,36	1,38	1,40	1,42	1,44
1,45	1,45	1,46	1,47	1,48	1,48	1,48	1,48	1,49
1,50	1,50	1,50	1,53	1,54	1,55	1,55	1,60	1,63

e considerare l'elemento centrale della distribuzione (evidenziato in rosso).

Per valutare la **media aritmetica** dobbiamo impostare il seguente calcolo:

$$\text{Media} = \frac{1,22 + 1,33 + 1,35 + \dots + 1,55 + 1,60 + 1,63}{27} = \frac{39,39}{27} \approx 1,458.$$

Possiamo ora rappresentare i dati secondo le procedure ormai note. Nella figura seguente abbiamo utilizzato un istogramma e un areogramma.



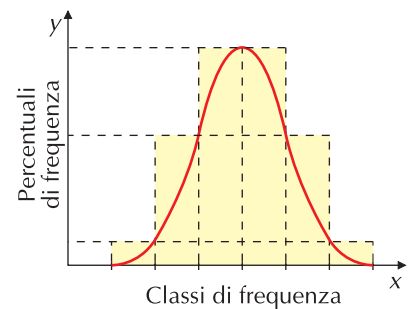
La distribuzione gaussiana

In molti casi i dati relativi ad un certo fenomeno sono distribuiti in modo simmetrico rispetto al valore centrale; questo avviene soprattutto se si considera un numero molto alto di valori. Questo tipo di distribuzione è stata studiata da Karl Friedrich Gauss (1777-1855), un famoso matematico tedesco, ed è rappresentata nella figura a lato. Tale grafico prende il nome di **curva di Gauss** ma si può chiamare anche **curva a campana**, data la sua forma.

Analizzare questa curva dal punto di vista matematico sarebbe troppo complicato. Ci limiteremo a descriverla riferendoci all'esempio precedente relativo alle altezze di tutti i ragazzi presenti in una classe. Già considerando l'istogramma della figura in alto, ottenuto tenendo conto solamente di 27 unità, si nota una certa simmetria rispetto al valore centrale della distribuzione (la classe 3 corrispondente alle altezze comprese fra 140 cm e 150 cm).

Se invece di considerare la popolazione costituita da 27 unità avessimo potuto utilizzare le altezze di tutti i ragazzi presenti nei diversi istituti scolastici della tua regione, avremmo ottenuto, distribuendo i dati raccolti in classi di ampiezza uguale, una situazione simile a quella descritta dai rettangoli gialli dalla figura in alto. Volendo approssimare questo istogramma con una curva (che vuol dire stringere l'ampiezza delle classi in cui facciamo le osservazioni) avremmo ottenuto la linea rossa.

In ogni caso, sia riferendosi all'istogramma che alla curva di Gauss, si capisce che la maggior parte dei ragazzi si concentra nella classe "centrale" della distribuzione. Quelli molto alti e quelli molto bassi sono ovviamente in minor numero.



Verifica se hai capito

- 1 Il campo di variabilità di una distribuzione statistica è:
 - a. la somma tra il valore più grande e quello più piccolo;
 - b. il rapporto tra il valore più grande e quello più piccolo;
 - c. la differenza tra il valore più grande e quello più piccolo.

2 Qual è il campo di variabilità dei seguenti dati statistici?
35; 23; 55; 32; 50; 65; 29; 25; 70; 66; 44; 37; 41.

3 La frequenza relativa di una serie di dati si ottiene calcolando:
a. il numero di volte in cui il dato si ripete;
b. il rapporto tra la frequenza assoluta ed il totale delle frequenze assolute;
c. il prodotto tra la frequenza assoluta ed il totale delle frequenze assolute;
d. il valore percentuale della frequenza assoluta.

4 Si chiama frequenza percentuale di una serie di dati raccolti in classi di frequenza:
a. il valore ottenuto moltiplicando per 100 il valore della frequenza assoluta;
b. il valore massimo di una serie di dati;
c. il valore minimo di una serie di dati;
d. il valore ottenuto moltiplicando per 100 il valore della frequenza relativa.

5 Si chiama classe modale:
a. la classe in cui si concentra la minor parte dei dati;
b. la classe in cui si concentra la maggior parte dei dati;
c. la classe con i valori più bassi;
d. la classe con i valori più alti.

ESERCIZI E PROBLEMI

1 Il campo di variabilità di una serie di dati è:
a. la semisomma tra il valore più grande e quello più piccolo;
b. la differenza tra l'ultimo dato e il primo;
c. la differenza tra il valore più grande e quello più piccolo;
d. la differenza della media dei dati dal valore più grande.

2 In riferimento alla seguente tabella relativa ai punteggi ottenuti da 45 studenti in un test di matematica, indica quali delle seguenti affermazioni sono vere e quali false:

Classe di	$5 \div 10$	$10 \div 15$	$15 \div 20$	$20 \div 25$
N° Persone	10	8	12	15

a. il campo di variabilità è suddiviso in 4 classi;
b. il campo di variabilità è 25;
c. l'ampiezza di ciascuna classe è 5;
d. il campo di variabilità dei dati è $5 \div 20$.

3 Completa le seguenti affermazioni:
a. la frequenza relativa di una serie di dati è data dal fra la frequenza assoluta e il numero;
b. la frequenza percentuale si ottiene moltiplicando per la frequenza;
c. la classe modale è la classe in cui si concentrano il numero di



I seguenti dati si riferiscono all'età dei 92 dipendenti di un'azienda.

18	38	30	48	55	45	48	47	25	45	54	20	28	41	24	40	28	36	59
34	43	29	51	39	60	57	29	42	46	27	62	46	45	19	41	44	40	47
42	30	35	49	38	30	22	26	34	52	31	50	37	53	40	31	55	50	44
36	50	64	49	23	47	37	42	65	33	48	36	36	40	21	51	50	32	46
51	57	44	56	38	39	48	37	26	43	38	44	39	49	62	61			

Raccogli i dati in una opportuna tabella, calcola le frequenze assolute, relative e percentuali, determina la classe modale e rappresenta i dati con un istogramma e un areogramma.

Determiniamo l'ampiezza del campo di variabilità:

18 → valore minimo

65 → valore massimo

$65 - 18 = 47$ → ampiezza del campo di variabilità

Dobbiamo ora decidere in quante classi dividiamo il campo di variabilità; se stabiliamo di volerlo suddividere in 6 classi, dobbiamo dividere la sua ampiezza per 6:

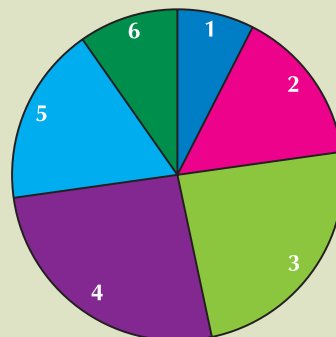
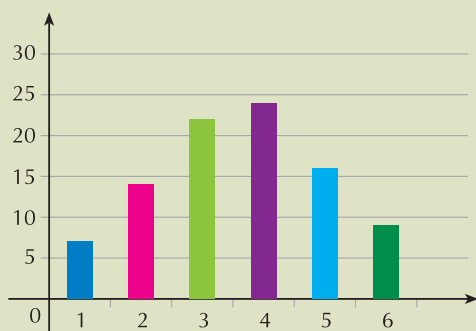
$47 : 6 = 7,833$ → che, arrotondato a 8, costituisce l'intervallo.

Possiamo ora raggruppare i dati in una tabella inserendo nella seconda colonna le classi di ampiezza 8; nella terza colonna il numero di persone che hanno l'età compresa nell'intervallo considerato, ottenendo così le frequenze assolute; in una quarta colonna le frequenze relative calcolate come rapporto tra quelle assolute e il numero totale del campione analizzato; in un'ultima colonna le frequenze percentuali ottenute moltiplicando per 100 quelle relative.

Classe n°	Classi di anni	Freq. Assoluta	Freq. Relativa	Freq. Percentuale
1	$18 \div 25$	7	$7 : 92 = 0,08$	$0,08 \cdot 100 = 8\%$
2	$25 \div 33$	14	$14 : 92 = 0,15$	$0,15 \cdot 100 = 15\%$
3	$33 \div 41$	22	$22 : 92 = 0,24$	$0,24 \cdot 100 = 24\%$
4	$41 \div 49$	24	$24 : 92 = 0,26$	$0,26 \cdot 100 = 26\%$
5	$49 \div 57$	16	$16 : 92 = 0,17$	$0,17 \cdot 100 = 17\%$
6	$57 \div 65$	9	$9 : 92 = 0,1$	$0,1 \cdot 100 = 10\%$

Dalla tabella si capisce che la classe modale della distribuzione è la classe $41 \div 49$.

Possiamo ora rappresentare i dati dell'indagine con un istogramma e un areogramma.



5 Determina il campo di variabilità delle seguenti successioni di dati:

a. 24; 10; 12; 15; 8; 18; 8; 32; 35; 16. [27]

b. 0,75; 1,25; 4,5; 3,6; 0,8; 5,2; 3,4; 2,6; 1,1; 2,2. [4,45]

c. 69; 81; 40; 48; 60; 77; 32; 64; 70; 74; 80; 37. [49]

6 I seguenti dati si riferiscono ad un'indagine statistica relativa al numero di sigarette fumate in media in un giorno da adulti fumatori:

Numero sigarette fumate	5	10	12	14	16	18	20	24	28	30
Frequenza	5	6	9	6	7	8	12	10	3	4

Completa la seguente tabella, determinando l'ampiezza delle classi:

Classe	Frequenza assoluta
1° da a	
2° da a	
3° da a	
4° da a	

7 Completa la seguente tabella relativa al peso in kg dello zaino degli alunni di una scuola.

Classi	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
0 ÷ 3	3		
3 ÷ 6	36		
6 ÷ 9	47		
9 ÷ 12	13		
12 ÷ 15	10		
15 ÷ 18	1		

Rappresenta i dati con un areogramma e un istogramma e indica la classe modale.

● **8** I seguenti dati si riferiscono alla distanza in metri tra casa e scuola di un gruppo di alunni:

120	340	625	270	1050	545	2200	780	1500	50	600	720	240	650	570
640	520	2500	540	610	1200	180	2400	840	310	780	630	570	1900	450

Raccogli i dati in una opportuna tabella, calcola la frequenza assoluta, relativa e percentuale, rappresenta i dati con un istogramma e un areogramma e calcola infine la classe modale, la mediana e la media. [Mediana: 617,5 m; Media: 811 m]

● **9** I seguenti dati si riferiscono all'ampiezza in m² dell'abitazione degli alunni di due classi di una scuola:

60	120	85	200	50	65	90	110	60	80	250	150	80	100	90	95	130	115	75	50
85	90	300	105	90	85	70	95	110	90	120	80	75	115	110	100	95	90	110	260

Raccogli i dati in una opportuna tabella, calcola la frequenza assoluta, relativa e percentuale, rappresenta i dati con un istogramma e un areogramma e calcola infine la classe modale, la mediana e la media. [Mediana: 92,5 Media: 108,25]

● **10** I seguenti dati si riferiscono all'età dei papà degli alunni di seconda di una scuola secondaria di primo grado:

35	60	40	37	45	47	44	46	45	50	52	56	48	46	44	36	40	44	47	48	49	47	41	44	50
51	50	52	58	49	48	51	49	47	52	48	46	51	50	39	40	48	48	50	51	58	50	49	48	46

Raccogli i dati in una opportuna tabella, calcola la frequenza assoluta, relativa e percentuale, rappresenta i dati con un istogramma e un areogramma e calcola infine la classe modale, la mediana e la media. [Mediana: 48; media: 47,4]

- **11** Due classi di una scuola hanno svolto lo stesso compito di matematica. Sono stati raccolti i dati relativi al tempo, espresso in minuti, impiegato dai singoli alunni per completare la prova:

100	92	83	110	105	78	105	102	95	85
75	119	105	100	90	109	98	93	103	120
104	116	95	110	115	80	100	102	108	118
120	95	105	105	96	100	115	107	120	98
115	109	88	104	90	105	102	85	120	115

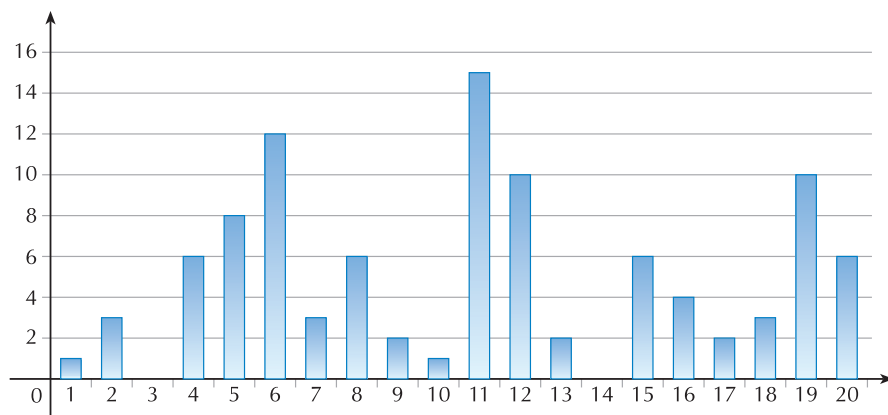
Raccogli i dati in una opportuna tabella e calcola le frequenze assolute, relative e percentuali.

- **12** La seguente tabella riporta il numero di clienti di un negozio nelle prime tre settimane di apertura.

50	37	66	62	46	39	55	58	32	49	50	44	52	40	47	38	62	66	38	37	50
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

- Dopo aver considerato il campo di variabilità e determinato l'ampiezza del raggruppamento, suddividi i dati in opportune classi;
- compila una tabella inserendo per ciascuna classe la frequenza assoluta, relativa e percentuale;
- rappresenta con un areogramma le frequenze percentuali;
- determina la classe modale, la mediana e la moda.

- **13** Il seguente grafico rappresenta i giorni di assenza dei venti alunni di una classe nel corso dell'anno scolastico (gli alunni sono registrati con il numero corrispondente all'ordine alfabetico).



- a. Completa la seguente tabella dopo aver ricavato i dati dal grafico;

Numeri giorni di assenza	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
da 0 a			

- calcola la classe modale, la media e la mediana;
- determina quanti alunni hanno fatto un numero di assenze superiore e quanti inferiore alla media;
- rappresenta con un areogramma le frequenze percentuali.