STATISTICA DESCRITTIVA

svolti dal prof. Gianluigi Trivia

11 dicembre 2025

Raccolta e tabulazione dei dati

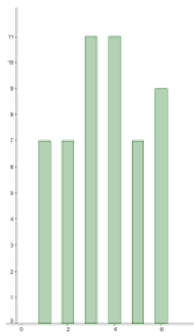
Esercizio 1. Quando si ottiene un volume di dati elevato, è necessario presentarli in una tabella di frequenza. Raggruppare i dati assegnati che rappresentano i numeri che compaiono su un dado dopo 62 lanci

6	6	4	3	5	3	1	4	6	5	6	4	2
6	4	5	6	2	1	6	4	3	4	6	1	5
3	3	6	2	2	4	3	4	5	3	4	1	2
1	1	2	1	5	3	4	3	4	2	3	5	3

Soluzione. Contiamo il numero di volte in cui compare un valore (frequenza) e scriviamo i risultati in un tabella

Numero del dado	Frequenza
1	7
2	7
3	11
4	11
5	7
6	9
Totale	52

Il risultato totale deve corrispondere al numero complessivo di lanci.
I dati della tabella si possono riassumere anche con un grafico opportuno, ad esempio a barre.



La frequenza può essere espressa non solo in valore assoluto, ma anche come

Numero del dado	f. ass.	f. rel.	f. (%)	f. cumulata
1	7	$\frac{7}{52} = 0,134$	13,4%	13,4
2	7	0,134	13,4%	26,8
3	11	0,212	21,2%	48,0
4	11	0,212	21,2%	69,2
5	7	0,134	13,4%	82,6
6	9	0,174	17,4%	100
Totale	52	1,000	100%	

Esercizio 2. La misura di una grandezza fisica ripetuta 20 volte fornisce i seguenti valori, espressi in cm:

2,04	1,84	1,92	2,10	1,78	1,82	1,76	1,94	2,02	1,88
2,05	1,87	2,00	1,83	1,86	2,12	2,02	1,95	2,05	2,01

raggruppare i valori in classi e calcolare tutti i tipi di frequenza dell'esercizio precedente.

Soluzione. I valori presentano un limite inferiore $1,76$ e uno superiore $2,12$. La differenza tra i due valori è $2,12-1,76=0,36$. In teoria la misura poteva fornire qualsiasi valore all'interno di questo intervallo e quindi questo caso caratterizza una variabile continua e non più discreta.

Scegliamo di raggruppare questi valori in classi di ampiezza $0,4$ come nella tabella (con $1,76<x\leq 1,80$) e così per le altre classi

Classi	f. ass.	f. rel.	f. (%)	f. cumulata
1,76 - 1,80	2	$\frac{2}{20}=0,10$	10%	10
1,80 - 1,84	3	0,15	15%	25
1,84 - 1,88	3	0,15	15%	40
1,88 - 1,92	1	0,05	5,0%	45
1,92 - 1,96	2	0,10	10%	55
1,96 - 2,00	1	0,05	5,0%	60
2,00 - 2,04	4	0,20	20%	80
2,04 - 2,08	2	0,10	10%	90
2,08 - 2,12	2	0,10	10%	100
Totale	20	1,000	100%	

Esercizio 3. Una gelateria ha rilevato le vendite giornaliere di cinque gusti di gelato alla frutta che indichiamo con: L=limone; F=fragola; M=Melone; P=Pompelmo; A=Amarena. I dati della rilevazione di un giorno sono riportati nella tabella

L	P	F	F	L	A	A	P	M	M
F	P	M	M	M	L	L	F	P	M
F	A	L	F	F	A	A	A	P	F
A	L	L	F	L	F	F	L	M	M
P	L	M	P	P	F	F	L	A	L

Compilare la tabella delle frequenze

Soluzione. Questa volta mettiamo la tabella in orizzontale

Gusto	Limone	Fragola	Melone	Pompelmo	Amarena	Totale
f. assoluta	12	13	9	8	8	50
f. relativa	$\frac{12}{50}=0,24$	0,26	0,18	0,16	0,16	1,00
f. %	24	26	18	16	16	100
f. cumulata	24	50	68	84	100	

Esercizio 4. Uno studente registra il tempo che impiega per svolgere quindici esercizi di matematica, simili a quelli che deve affrontare nel compito in classe il giorno successivo. I tempi di svolgimento degli esercizi, espressi in minuti, sono i seguenti:

$$4 \quad 5 \quad 6 \quad 12 \quad 10 \quad 8 \quad 3 \quad 4 \quad 7 \quad 14 \quad 18 \quad 15 \quad 19 \quad 13 \quad 8$$

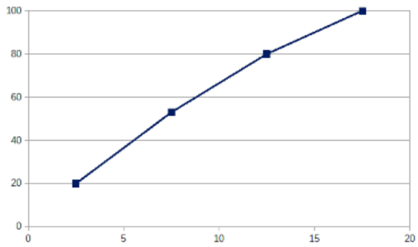
Supponi di suddividere i dati nelle quattro classi: $0 \leq t < 5$, $5 \leq t < 10$, $10 \leq t < 15$, $15 \leq t < 20$ dove t indica il tempo, espresso in minuti, impiegato per risolvere un esercizio.

Compila la tabella della distribuzione in classi.

Soluzione. In questo caso le classi e la loro ampiezza è già indicata, per cui

Classi	f. ass.	f. rel.	f. (%)	f. cumulata
$0 \leq t < 5$	3	$\frac{3}{15} = 0,20$	20%	20
$5 \leq t < 10$	5	0,33	33%	53
$10 \leq t < 15$	4	0,27	27%	80
$15 \leq t < 20$	3	0,20	20%	100
Totale	15	1,000	100%	

Mostriamo il grafico delle frequenze cumulate.



Lettura dei grafici

Esercizio 5. Il grafico in figura riassume i dati di uno studio statunitense sulla diffusione dei social media nel periodo2005-2010, diviso per fasce d'età. **a**). Individua la fascia di età per la quale dall'aprile 2009 al maggio 2010 si è avuto il più consistente incremento percentuale nella diffusione dei social media. **b**). Individua la fascia d'età che dal settembre 2005 al maggio 2010 ha manifestato nei confronti dei social media un interesse che non è risultato sempre crescente. **c**). Individua la fascia d'età per la quale dal settembre 2005 al novembre 2008 si è avuto il più consistente incremento percentuale nella diffusione dei social media.

Soluzione. a). Per la fascia 18-29 anni l'incremento è del 7%; per la fascia 30-49 del 10%; per la fascia 50-64 del 11% e infine per la fascia > 65 del 9%, risultando quindi la fascia con il maggior incremento percentuale. La fascia d'età è quella 50-64 anni (lo si può osservare anche dalla pendenza del tratto di retta)

b). In questo caso sull'intero periodo la fascia over 65 ha mostrato una diminuzione nel 2008.

c). Si vede che nonostante i diversi livelli di partenza l'incremento maggiore è della fascia 18-29 per il grande balzo del primo periodo.

Esercizio 6. Il grafico mostra l'incidenza percentuale sul PIL della spesa pensionistica in Italia negli anni 1981-2013. (Fonte: Italia in cifre, ISTAT)

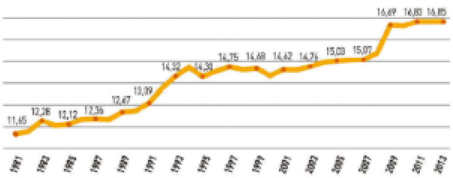
Individua, relativamente alla sequenza degli anni indicati in tabella:

a. In quale anno l'incidenza sul PIL è risultata massima;

b. In quale anno l'incidenza sul PIL è risultata minima;

c. i due anni successivi tra cui si è registrato il massimo aumento nell'incidenza sul PIL;

d. i due anni successivi tra cui si è registrato il minimo aumento nell'incidenza sul PIL;



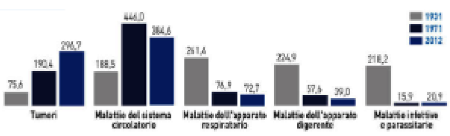
Soluzione. a) Nel 2013; b) Nel 1981; c) Nel 1991-1993 e nel 2007-2009; d) 2005-2007 e 2011-2013.

Esercizio 7. Nel diagramma sono confrontate alcune cause di mortalità nei tre anni 1931, 1971, 2012.(Fonte: Italia in cifre, ISTAT).

a. Qual era la più frequente causa di mortalità, fra quelle indicate, nel 1931? E nel 1971? E nel 2012?

b. Di quanto è aumentata in percentuale la mortalità per tumori dal 1931 al 2012?

c. Di quanto è diminuita in percentuale la mortalità per malattie del sistema circolatorio dal 1971 al 2012?



Soluzione. a) Nel 1931: apparato respiratorio; nel 1071: sistema circolatorio; nel 2012: sistema circolatorio.

b) $(296,7-75,6):75,6\times 100=292,5\%$

c) $|384,6-446,0|:446,0\times 100=13,8\%$

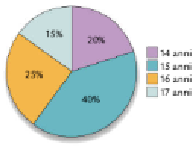
Esercizio 8. Il seguente diagramma a torta rappresenta la suddivisione per età dei 240 allievi di una scuola.

a. Quanti allievi hanno 17 anni?

b. Qual è la frequenza relativa degli allievi che hanno 16 anni?

c. Quanti allievi hanno meno di 16 anni?

d. Quanti allievi hanno più di 15 anni?



Soluzione. a) Il 15% cioè $240\times 0,15=36$

b) i ragazzi sono il 25% del totale, pari a 1, e quindi $0,25$.

c) sono gli allievi che hanno 14 e 15 anni, cioè, in percentuale, il 60% che corrisponde a 144.

d) sono gli allievi che hanno 16 e 17 anni, cioè il restante 40% che corrisponde a 96.

Esercizio 9. Il diagramma cartesiano in Fig. a mostra il numero totale di clienti di una filiale di un gruppo bancario che si dichiarano soddisfatti dei servizi offerti negli anni 2009-2012; il diagramma in Fig. b mostra invece la percentuale di clienti della filiale che si dichiarano soddisfatti.

a. Puoi osservare che tra il 2009 e il 2010, così come tra il 2011 e il 2012, sebbene il numero di clienti soddisfatti sia cresciuto, la corrispondente percentuale è diminuita. Spiega perché ciò non costituisce una contraddizione. **b**. Determina il numero complessivo di clienti della filiale in ciascuno degli anni dal 2009 al 2012.

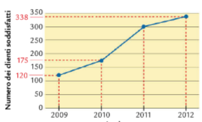


Figura a

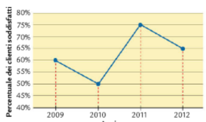


Figura b

Soluzione. a) L'osservazione si può fare attraverso la pendenza dei tratti di retta. si vede che la pendenza massima si ha negli anni dal 2010 al 2011, mentre la pendenza minima negli anni tra il 2011 e il 2012. Anche se il numero dei clienti aumenta sempre la crescita dei soddisfatti non è lineare, cioè non cresce come il numero dei soddisfatti e vi sono due periodi in cui la percentuale cala del 10% rispetto all'anno precedente.

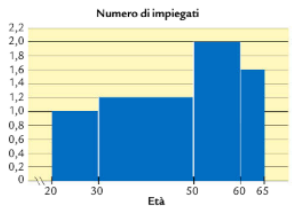
b) 2009: soddisfatti 60% cioè 120 clienti, insoddisfatti il restante 40%; il totale sarà $120 \times \frac{100}{60} = 200$; 2010: $175 \times \frac{100}{50} = 350$; 2011: 400; 2012: 520

Esercizio 10. In un ufficio si rilevano le età degli impiegati; suddivisi i dati nelle classi [20, 30), [30, 50), [50, 60), [60, 65), l'istogramma rappresenta la distribuzione di frequenza delle classi.

a. Quanti sono gli impiegati dell'ufficio aventi età maggiore o uguale a 30 anni ma minore di 50 anni?

b. Quanti sono gli impiegati dell'ufficio che hanno 50 anni o più di 50 anni?

c. Quanti sono complessivamente gli impiegati dell'ufficio?



Soluzione. (In un istogramma con classi di ampiezza diversa l'area di ciascun rettangolo è uguale alla frequenza della classe).

a) il valore è dato dall'area del 2° rettangolo: 24.

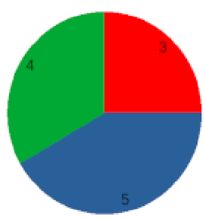
b) il valore è dato dall'area del 3° E 4° rettangolo: $20+8=28$

c) il numero totale degli impiegati è dato dalla somma delle aree di tutti rettangoli: $10+24+28=62$.

Rappresentazione dei dati

Esercizio 11. Un bambino possiede in tutto 12 pennarelli: 4 sono verdi, 5 sono blu e i rimanenti sono rossi. Rappresentare la distribuzione dei colori dei pennarelli con un diagramma circolare.

Soluzione. Il grafico è quello indicato anche come a torta.



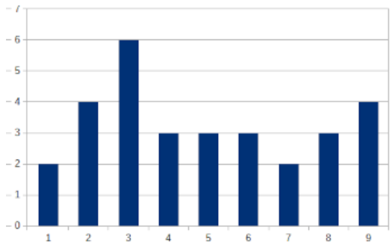
Esercizio 12. Le prime 30 cifre decimali di π sono: 1; 4; 1; 5; 9; 2; 6; 5; 3; 5; 8; 9; 7; 9; 3; 2; 3; 8; 4; 6; 2; 6; 4; 3; 3; 8; 3; 2; 7; 9. **a**. Determina la tabella che rappresenta la distribuzione di frequenze delle cifre. **b**. Rappresenta tale distribuzione di frequenze tramite un diagramma a barre.

Soluzione. Costruiamo la tabella delle frequenze assolute

si e la loro ampiezza è già indicata, per cui

cifre	1	2	3	4	5	6	7	8	9
frequenza	2	4	6	3	3	3	2	3	4

Diagramma a barre



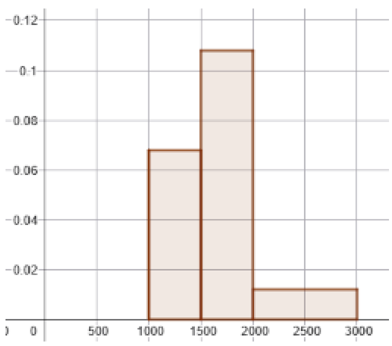
Esercizio 13. Nella tabella è rappresentata la distribuzione di frequenze degli stipendi (in euro), suddivisi in classi, rilevati su un campione di 100 persone. Rappresenta questa distribuzione tramite un istogramma.

Fasce stipendi	1000-1500	1500-2000	2000-3000
frequenza assoluta	34	54	12

Soluzione. Le fasce hanno ampiezze diverse e quindi i rettangoli avranno basi diverse e la loro area dovrà corrispondere alle diverse frequenze. Completiamo la tabella

Fasce stipendi	1000-1500	1500-2000	2000-3000
frequenza assoluta	34	54	12
ampiezza classe	500	500	1000
densità frequenza	$\frac{34}{500} = 0,068$	0,108	0,012

Istogramma (fasce stipendiali - densità di frequenza)



Esercizio 14. Si è rilevato il numero di automobili possedute da un campione di 20 famiglie italiane e si sono ottenuti i dati grezzi riportati nella tabella.

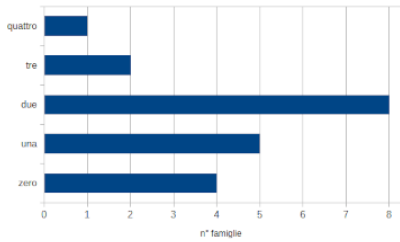
Famiglia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N. auto	0	1	2	0	1	1	2	2	4	2	0	3	1	2	0	2	3	2	2	1

a. Costruisci la distribuzione delle frequenze assolute e relative del numero di automobili possedute dalle famiglie.
b. Rappresenta la distribuzione delle frequenze assolute tramite il diagramma che ritieni opportuno.

Soluzione. Costruiamo una tabella numero di auto per famiglia - numero famiglie

N. auto	0	1	2	3	4
N famiglie (fr. assoluta)	4	5	8	2	1
fr. relativa	$\frac{4}{20} = 0,20$	0,25	0,40	0,10	0,05

Istogramma (fasce stipendiali - densità di frequenza)



Esercizio 15. Nella seconda riga della seguente tabella sono riportate le valutazioni di un titolo azionario il primo giorno dell'anno, per cinque anni consecutivi.

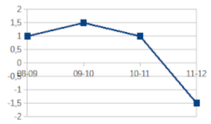
Giorno	1/1/2008	1/1/2009	1/1/2010	1/1/2011	1/1/2012
Valore titolo (€)	10,50	11,50	13	12	10,50

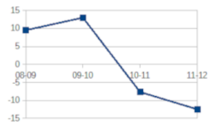
a. Completa la tabella, determinando le variazioni assolute del prezzo del titolo e quelle percentuali (queste ultime arrotondate con una cifra dopo la virgola).
b. Rappresenta l'andamento del valore del titolo negli anni in esame (2008-2012) e le corrispondenti variazioni percentuali rispetto all'anno precedente (a partire dal 2009) tramite opportuni diagrammi cartesiani.

Soluzione. Costruiamo la tabella

Intervallo tempo	2008/2009	2009/2010	2010/2011	2011/2012
Variazione assoluta	$11,50 - 10,50 = 1$	1,5	-1	-1,5
Variazione %	$\frac{1}{10,5} \simeq 9,5$	$\simeq 13,0$	$\simeq -7,7$	$\simeq -12,5$

grafico

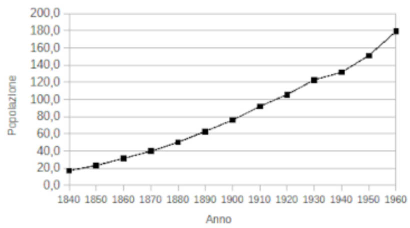




Esercizio 16. La tabella riporta la popolazione degli Stati Uniti (in milioni) dal 1840 al 1960. Rappresentare graficamente tali dati.

Anno	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	196
Popolazione (10^6)	17,1	23,2	31,4	39,8	50,2	62,9	76,0	92,0	105,7	122,8	131,7	151,1	179,

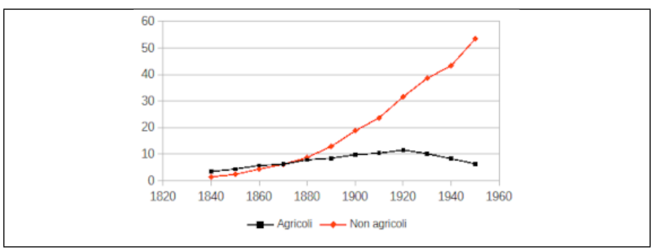
Soluzione. Un simile insieme di dati rappresenta una serie temporale. Mettiamo gli anni sull'asse orizzontale e la corrispondente popolazione sull'asse verticale

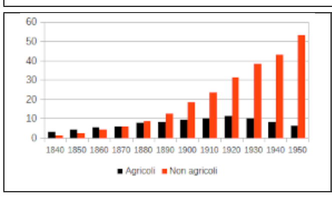


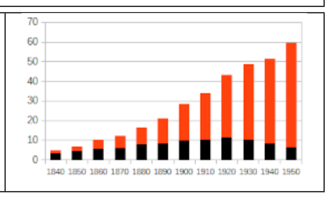
Esercizio 17. La tabella riporta il numero di lavoratori agricoli dagli anni 1840 al 1950. Tracciare il grafico dei dati usando a) grafici lineare, b) a barre, c) a rettangoli composti.

Anno	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950
Lav agricoli (10^6)	3,5	4,4	5,8	6,2	7,9	8,5	9,8	10,4	11,6	10,2	8,3	6,4
Lav non agricoli (10^6)	1,5	2,4	4,4	6,2	8,8	12,9	18,8	23,7	31,6	38,6	43,3	53,4

Soluzione. Un simile insieme di dati rappresenta una serie temporale. Mettiamo gli anni sull'asse orizzontale e la corrispondente popolazione sull'asse verticale



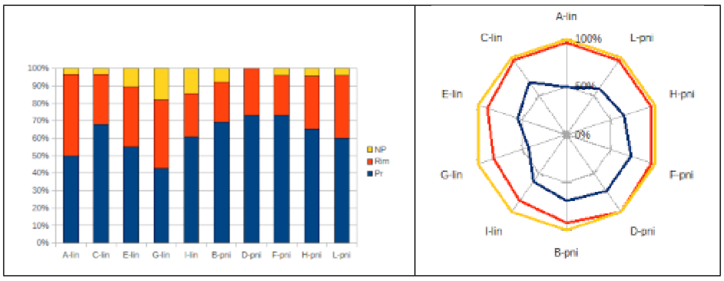




Esercizio 18. La tabella riporta gli esiti finali di voti finali degli studenti di prima liceo per sezione e per tipologia di corso. Confrontare le classi in base al tipo di corso.

Linguistico	Pr %	Rim %	NP %
A	50	46,4	3,6
C	67,8	28,6	3,6
E	55,2	34,5	10,3
G	42,8	39,3	17,9
I	60,7	25,0	14,3
PNI			
B	69,2	23,1	7,7
D	73,1	26,9	0
F	73,1	23,1	3,8
H	65,2	30,4	4,4
L	60,0	36,0	4,0

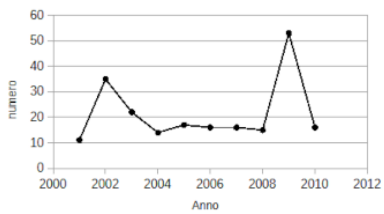
Soluzione. Rappresentiamo con un diagramma a colonne composte e poi con un diagramma radar per confrontare i risultati degli studenti delle due diverse tipologie



Esercizio 19. Nella tabella è riportato il numero di fenomeni sismici di magnitudo maggiore o uguale a 4, registrati in Italia dal 2001 al 2010 (fonte INGV). Rappresenta la serie storica mediante un diagramma cartesiano.

Anno	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
numero	11	35	22	14	17	16	16	15	53	16

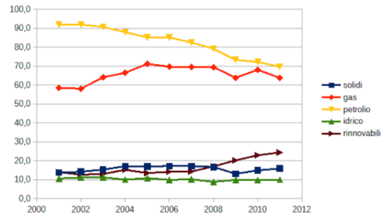
Soluzione. Il diagramma cartesiano è costruito ponendo sull'asse delle ascisse gli anni, a partire, ad es. dall'anno 2000, gli anni, e sull'asse delle ordinate il numero dei fenomeni.



Esercizio 20. Nella tabella è riportato il consumo interno lordo di energia proveniente da diverse fonti. Rappresenta la serie storica mediante un diagramma cartesiano.

anni	consumo interno lordo (milioni di tonnellate di petrolio equivalenti)				
	combustibili solidi	gas naturale	prodotti petroliferi	energia elettrica	fonti rinnovabili
2001	13,7	58,5	91,9	10,6	14,0
2002	14,2	58,1	92,0	11,1	12,6
2003	15,3	64,1	90,8	11,2	13,0
2004	17,1	66,5	88,0	10,0	15,2
2005	17,0	71,2	85,2	10,8	13,5
2006	17,2	69,7	85,2	9,9	14,2
2007	17,2	69,5	82,5	10,2	14,3
2008	16,7	69,5	79,2	8,8	17,0
2009	13,1	63,9	73,3	9,9	20,2
2010	14,9	68,1	72,2	9,7	22,9
2011	15,9	63,8	69,7	10,0	24,4

Soluzione. In questo caso i dati da rappresentare sono numerosi in uno stesso grafico; è necessario pertanto aggiungere una legenda che ne faciliti il riconoscimento.



Si può osservare che l'utilizzo del petrolio è andato sempre diminuendo; le fonti rinnovabili mostrano un incremento negli ultimi anni, mentre le altre fonti si mantengono pressoché costanti.

Indici di posizione

Media aritmetica

Esercizio 21. Trovare la media aritmetica del seguente gruppo di numeri: -2, $2, -4, 4, 5, -5$.

Soluzione. La media aritmetica è il rapporto tra la somma di tutti i valori divisa per il loro numero, cioè

$$m = \sum_{i=1}^{n} \frac{x_i}{n}$$

In questo caso i numeri sono a due a due opposti e la loro somma è zero

$$m = \frac{0}{6} = 0$$

Esercizio 22. Trovare la media aritmetica del seguente gruppo di numeri: $-3, -2, 4, 5, 6$.

Soluzione. In questo caso si hanno cinque valori

$$m = \frac{-3-2+4+5+6}{5} = \frac{10}{5} = 2$$

Esercizio 23. Trovare la media aritmetica del seguente gruppo di numeri: $1, 5, 4, 4, 2, 6$.

Soluzione. In questo caso si hanno sei valori

$$m = \frac{1+5+4+4+2+6}{6} = \frac{22}{6} = \frac{11}{3}$$

Esercizio 24. Trovare la media aritmetica del seguente gruppo di numeri: $-\frac{3}{5}, \frac{7}{5}, \frac{1}{5}, \frac{1}{2}$.

Soluzione. In questo caso si hanno quattro valori

$$m = \frac{-\frac{3}{5}+\frac{7}{5}+\frac{1}{5}+\frac{1}{2}}{4} = \frac{\frac{-6+14+2+5}{10}}{4} = \frac{\frac{15}{10}}{4} = \frac{3}{8}$$

Esercizio 25. Trovare la media aritmetica del seguente gruppo di numeri: $11, 15, 14, 14, 12, 16$.

Soluzione. In questo caso si hanno sei valori

$$m = \frac{11+15+14+14+12+16}{6} = \frac{82}{6} = \frac{41}{3}$$

Esercizio 26. Trovare la media aritmetica del seguente gruppo di numeri: $2 \cdot 10^{-4}, 4 \cdot 10^{-4}, -20 \cdot 10^{-5}$.

Soluzione. In questo caso si hanno sei valori

$$m = \frac{2 \cdot 10^{-4} + 4 \cdot 10^{-4} - 20 \cdot 10^{-5}}{3} = \frac{4}{3} \cdot 10^{-4}$$

Esercizio 27. Determinare per quale valore di a la media aritmetica del tre numeri $a, \frac{1}{2}a, \frac{1}{3}a$ è uguale a 22.

Soluzione. In questo caso si hanno sei valori

$$22 = \frac{a+\frac{1}{2}a+\frac{1}{3}a}{3} = \frac{\frac{6a+3a+2a}{6}}{3} = \frac{11}{18}a$$

da cui

$$a = \frac{22 \times 18}{11} = 36$$

Media aritmetica ponderata

Tale indice viene utilizzato quando i dati sono raggruppati e se ne conoscono le diverse frequenze.

Esercizio 28. Nella tabella si riportano le reti segnate negli incontri del campionato di calcio tedesco 2008/2009. Calcolare la media delle reti segnate per incontro.

reti segnate	0	1	2	3	4	5	6	7	8	9
n. incontri	15	43	77	71	47	31	16	4	1	1

Soluzione. Calcoliamo la media aritmetica ponderata sul numero di incontri

$$\overline{x}=\frac{0\cdot 15+1\cdot 43+2\cdot 77+3\cdot 71+4\cdot 47+5\cdot 31+6\cdot 16+7\cdot 4+1\cdot 8+1\cdot 9}{15+43+77+71+47+31+16+4+1+1}=2,92$$

Esercizio 29. Il peso in kg degli studenti di una classe raggruppato in classi. Trovare il peso medio.

peso	n.studenti
50-55	5
55-60	6
60-65	9
65-70	3
70-75	2

Soluzione. Calcoliamo la media aritmetica ponderata sul numero degli studenti. In questo caso ogni classe ha un'ampiezza uguale a 5 e nel calcolo si prende il valore centrale della classe, cioè la media aritmetica tra i due estremi

$$\overline{x}=\frac{52,5\cdot 5+57,5\cdot 6+62,5\cdot 9+67,5\cdot 3+72,5\cdot 2}{5+6+9+3+2}=60,7\,kg$$

Mediana

La mediana di un insieme di valori ordinati in ordine di grandezza è il valore centrale oppure la media aritmetica dei due valori centrali.

Esercizio 30. Trovare la mediana del seguente insieme di numeri: $-2,2,-4,4,5,-5$

Soluzione. Ordiniamo prima i numeri in forma crescente

$$-5,-4,-2,2,4,5$$

il numero dei termini è pari e non c'è un valore centrale che lascia da entrambe le parti la stessa quantità di valori; prenderemo quindi i due valori centrali e calcoliamo la media tra loro

$$med=\frac{-2+2}{2}=0$$

Esercizio 31. Trovare la mediana del seguente insieme di numeri: $\frac{19}{15}, -\frac{3}{5}, \frac{7}{5}, \frac{1}{5}, \frac{1}{2}, \frac{5}{4}, \frac{13}{10}$

Soluzione. Se vogliamo mantenere la scrittura sotto forma di frazioni è necessario riscrivere tutte le frazioni con lo stesso denominatore

$$\frac{76}{60}, -\frac{36}{60}, \frac{84}{60}, \frac{12}{60}, \frac{30}{60}, \frac{75}{60}, \frac{78}{60}$$

ordiamo ora i numeri

$$-\frac{36}{60}, \frac{12}{60}, \frac{30}{60}, \frac{75}{60}, \frac{76}{60}, \frac{78}{60}, \frac{84}{60}$$

il numero dei termini è dispari

$$med = \frac{75}{60} = \frac{5}{4}$$

Esercizio 32. Trovare la mediana del seguente insieme di numeri: $-a, a, \frac{3}{4}a, \frac{2}{3}a, \frac{1}{2}a$

Soluzione. Manteniamo la scrittura sotto forma di frazioni

$$-\frac{12}{12}a, \frac{12}{12}a, \frac{9}{12}a, \frac{8}{12}a, \frac{6}{12}a,$$

ordiamo ora i numeri

$$-\frac{12}{12}a, \frac{6}{12}a, \frac{8}{12}a, \frac{9}{12}a, \frac{12}{12}a$$

il numero dei termini è dispari

$$med = \frac{8}{12}a = \frac{2}{3}a$$

Moda

La moda di un insieme di numeri è il valore che si presenta con la frequenza maggiore, cioè il valore che compare più volte. La moda può anche non esistere oppure può non essere unica.

Esercizio 33. Trovare la moda del seguente insieme di numeri: $2, 3, 4, 5, 5, 4, 1, 5$

Soluzione. osserviamo che il numero 5 compare tra volte e quindi

$$mod = 5$$

Esercizio 34. Trovare la moda del seguente insieme di numeri: $10, 15, 14, 14, 12, 15, 16$

Soluzione. osserviamo che vi sono due valori che compaiono lo stesso numero di volte, per cui

$$mod = 14, 15$$

Esercizio 35. Trovare la moda del seguente insieme di numeri: $2, 4, 3, 6, 7, 1$

Soluzione. osserviamo che ogni valore compare una volta sola e quindi la moda è data da tutti i numeri presenti

$$mod = 2, 4, 3, 6, 7, 1$$

Esercizio 36. La tabella seguente contiene le altezze di 200 persone

Classi	valore centrale	freq	f. (%)
155,5-158,5	157	2	1,0
158,5-161,5	160	7	3,5
161,5-164,5	163	22	11
164,5-167,5	166	13	6,5
167,5-170,5	169	44	22
170,5-173,5	172	36	18
173,5-176,5	175	32	16
176,5-179,5	178	13	6,5
179.5-182,5	181	21	10,5
182,5-185,5	184	10	5,0
Totale		200	100%

Calcolare media aritmetica, mediana, moda e rappresentare graficamente i dati.

Soluzione. In questo caso ogni valore compare più volte e quindi invece della media aritmetica semplice si deve introdurre la media pesata (in pratica, ad es., invece di scrivere 44 volte il numero 169, si sintetizza il calcolo con 169×44)

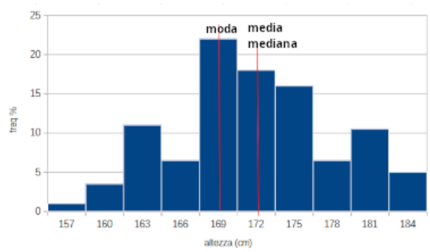
$$m = \frac{157\times2+160\times7+163\times22+166\times13+169\times44+172\times36+175\times32+175\times13+181\times21+184\times10}{200} = 171,6 \simeq 172$$

Il numero di dati è pari (200). Per la mediana osserviamo che i valori centrali sono tutti uguali a 172, per cui

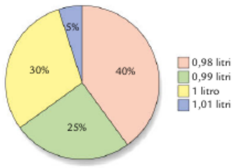
$$med = 172$$

La moda è invece

$$mod = 169$$



Esercizio 37. Si è rilevata, su un campione di bottiglie da 1 litro di acqua minerale, la quantità di acqua effettivamente contenuta in esse. I risultati ottenuti sono riassunti nel seguente diagramma a torta. Calcola la quantità di acqua media contenuta in una bottiglia.



Soluzione. Ogni contenuto misurato ha un peso diverso. Possiamo supporre di riscrivere i dati immaginando che il totale delle bottiglie analizzate sia 100. Avremo

$$m = \frac{0,98 \cdot 40 + 1 \cdot 30 + 0,99 \cdot 25 + 1,01 \cdot 5}{100} = 0,99\,l$$

Esercizio 38. Un ristorante propone 10 diversi primi piatti, i cui prezzi presentano la seguente distribuzione di frequenze.

Prezzo €	10	12	16	18
Frequenza	4	3	2	1

a. Qual è il prezzo medio di un primo?
b. Se il proprietario del ristorante decidesse di raddoppiare il prezzo di ciascun piatto, quale diventerebbe il prezzo medio di un primo?
c. Se il proprietario del ristorante decidesse di aumentare del 10% il prezzo di ciascun piatto, quale diventerebbe il prezzo medio di un primo?

Soluzione. a) Anche qui ogni prezzo ha un peso diverso

$$m = \frac{10 \cdot 4 + 12 \cdot 3 + 16 \cdot 2 + 18 \cdot 1}{10} = \frac{126}{10} = 12,6$$

b) Se il prezzo raddoppia si ha

$$m = \frac{20 \cdot 4 + 24 \cdot 3 + 32 \cdot 2 + 36 \cdot 1}{10} = \frac{252}{10} = 25,2$$

si vede che anche la media raddoppia; infatti, nella frazione raddoppia solo il numeratore. Questa è una proprietà della media.
c) come per il caso b), la media aumenta del 10%, cioè diviene $13,86$ euro.

Esercizio 39. In uno scaffale di una libreria ci sono 20 libri; i primi 10 hanno un numero medio di 300 pagine; i 4 successivi hanno mediamente 360 pagine e la media delle pagine degli ultimi sei è 420. Qual è il numero medio di pagine dei 20 libri?

Soluzione. avremo

$$m = \frac{10 \cdot 300 + 4 \cdot 360 + 6 \cdot 420}{20} = \frac{6960}{20} = 348$$

Esercizio 40. Quattro cubi hanno spigoli di lunghezze 2 cm, 4 cm, 3 cm e 5 cm. a). Qual è la lunghezza media, che chiamiamo m, degli spigoli dei cubi? b). Il volume medio dei quattro cubi è uguale al volume del cubo avente spigolo di lunghezza m?

Soluzione. a) troviamo il valore medio dello spigolo

$$m=\frac{2+4+3+5}{4}=\frac{14}{4}=3,5\,cm$$

b) Calcoliamo il volume di ogni cubo ($V=l^3$); avremo: $8\,cm^3$, $64\,cm^3$, $27\,cm^3$, $125\,cm^3$; pertanto

$$m=\frac{8+64+27+125}{4}=\frac{224/4}{4}=56\,cm^3$$

tale volume è diverso da quello del cubo avente come spigolo quello medio, $V=3,5^3=42,875\,cm^3$.

Esercizio 41. Trovare la media, la mediana e la moda dei seguenti dati non ordinati. Tracciare poi il grafico della distribuzione di frequenze.

7	4	10	9	15	12	7	1	10	8
8	11	4	14	10	5	14	12	6	5

Soluzione. I numeri sono 20 e per trovare la media basta sommarli e dividere il risultato per 20

$$m=\frac{172}{20}=8,6$$

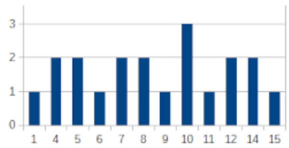
Per trovare la mediana i numeri vanno ordinati

1 4 4 5 5 6 7 7 8 8 9 10 10 10 11 12 12 14 14 15

20 è un numero pari per cui

$$med=\frac{8+9}{2}=8,5$$

la moda è 10, cioè il valore che compare tre volte.



Esercizio 42. Una famiglia spende per il riscaldamento in due anni consecutivi rispettivamente 1425 euro e 2730 euro. Se la spesa per il combustibile è di $0,57\,€/m^3$il primo anno e $0,65\,€/m^3$ il secondo anno, calcola, arrotondando ai centesimi, il costo medio sostenuto dalla famiglia nei due anni per un metro cubo di combustibile.

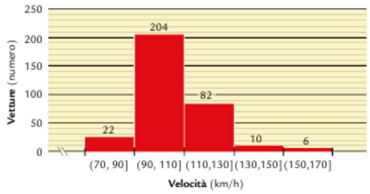
Soluzione. Il medio sostenuto è dato dal rapporto tra la spesa totale e il consumo totale

$$Spesa=1425+2730=4155\,€\quad V_T=\frac{1425}{0,57}+\frac{2730}{0,65}=2500+4200=6700\,m^3$$

il costo medio è pertanto

$$\overline{C}=\frac{4155}{6700}=0,62\,\frac{€}{m^3}$$

Esercizio 43. Da un controllo delle velocità delle auto in un tratto autostradale, con limitazione a 110 km/h si ottengono i valori delle frequenze delle velocità rilevate (raggruppate in classi) rappresentati nell'istogramma. **a.** Qual è la velocità media degli automobilisti che sono stati oggetti del controllo (arrotonda il risultato a meno di una unità)? **b.** Per uno sforamento del limite di velocità superiore ai 40 km/hè previsto il ritiro e la sospensione della patente. Qual è la percentuale di automobilisti che incorre in questa sanzione (arrotonda il risultato a meno di un centesimo)?



Soluzione. Il grafico mostra che 226 automobilisti hanno rispettato il limite e 98 no. Di questi ultimi 6 hanno superato il limite di oltre 40 km/h. Il numero totale degli automobilisti controllati è 324.

a. Calcoliamo la velocità media complessiva (prendiamo come velocità rappresentativa di ogni classe il valore centrale)

$$\overline{v} = \frac{80\cdot 22 + 100\cdot 204 + 120\cdot 82 + 140\cdot 10 + 160\cdot 6}{324} = \frac{34360}{324} = 106\,\frac{km}{h}$$

tale media è anche detta media pesata, cioè ogni velocità è riferita al numero di quelli appartenenti alla classe.

b. gli automobilisti sono 6, per cui

$$\frac{6}{324}\cdot 100 = 1,85\%$$

Esercizio 44. In una classe di 27 alunni, la media dei voti nell'ultimo compito in classe di matematica è stata 6,5. La media dei voti delle 12 ragazze è stata 7. a. Qual è la somma dei voti di tutti gli alunni? b. Qual è la somma dei voti di tutte le ragazze? c. Qual è la media dei voti dei ragazzi?

Soluzione. a) La somma dei voti è

$$s = 27 \times 6,5 = 175,5$$

b) la somma dei voti delle sole femmine è

$$s_f = 12 \times 7 = 84$$

c) gli studenti maschi sono 15 e la somma dei loro voti è $175,5 - 84 = 91,5$; la media dei loro voti è

$$v_m = \frac{91,5}{15} = 6,1$$

Esercizio 45. I dipendenti di un'azienda sono per il 40% donne e per il 60% uomini. Il salario medio mensile delle donne è di 1400 euro, quello medio degli uomini è di 1600 euro. Qual è il salario medio mensile dei dipendenti dell'azienda?

Soluzione. In questo caso abbiamo non dei valori assoluti per il numero di dipendenti ma valori in percentuale; consideriamo quindi come totale il 100%, avremo

$$\overline{sal} = \frac{1400 \times 40 + 1600 \times 60}{100} = 1520\,€$$

Esercizio 46. I tempi di percorrenza da parte di un ciclista in tre tappe ciclistiche sono stati: $6^h36^m40^s$, $4^h40^m50^s$, $5^h10^m58^s$. Qual è il tempo medio di percorrenza sulle tre tappe?

Soluzione. Trasformiamo i tempi di percorrenza in secondi, avremo

$$\begin{aligned} t_1 &= 6 \times 3600 + 36 \times 60 + 40 = 23800\,s \\ t_2 &= 4 \times 3600 + 40 \times 60 + 50 = 16850\,s \\ t_3 &= 5 \times 3600 + 10 \times 60 + 58 = 18658\,s \end{aligned}$$

il tempo medio per la percorrenza delle tre tappe è

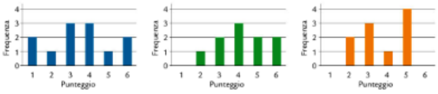
$$\bar{t} = \frac{23800 + 16850 + 18658}{3} = 19769\,s = 5^h29^m29^s$$

Esercizio 47. Sono stati effettuati tre diversi esperimenti consistenti nel lancio successivo di un dado per un certo numero di volte (non uguale nei tre esperimenti) e nella registrazione del punteggio ottenuto. Le distribuzioni di frequenze dei punteggi ottenuti nei tre esperimenti sono rappresentate nei seguenti tre diagrammi.

a. Calcola quante volte è stato lanciato il dado in ciascuno dei tre esperimenti.

b. Calcola media, mediana e moda di ciascuna delle tre distribuzioni di frequenze rappresentate.

c. Calcola il punteggio medio corrispondente ai lanci eseguiti complessivamente in tutti e tre gli esperimenti.



Soluzione. a) Leggiamo le frequenze sull'asse verticale nei tre casi:

$$f_1 = 12 \quad f_2 = 10 \quad f_3 = 10$$

b) Il 1° esperimento mostra un grafico simmetrico rispetto a un'ipotetica retta passante per il punto medio tra i rettangoli relativi ai numeri del dado 3 e 4, per cui la mediana è 3,5;

anche la moda è immediatamente osservabile ed è rappresentata dei numeri 3 e 4; per la media osserviamo che nei dodici lanci si è ottenuto un punteggio complessivo di 42, per cui

$$\overline{m} = \frac{42}{12} = 3,5$$

otteniamo un'informazione generalizzabile, cioè che in una distribuzione simmetrica la media e la mediana sono uguali e si trovano sull'asse di simmetria.

Nel 2° esperimento si ha ancora un punteggio totale di 42, ma solo 10 lanci, per cui la media è $4,2$; la mediana e la moda coincidono e sono uguali a 4.

Nel 3° esperimento il punteggio totale è 37 e i lanci 10, per cui $\overline{m} = 3,7$; la mediana è la media tra il punteggio del 5° e 6° lancio, cioè $3,5$ e la moda è 5.

c) i lanci complessivi sono stati 32 e il punteggio totale è 121, per cui

$$\overline{m}_{123} = \frac{121}{32} \simeq 3,8$$

Esercizio 48. In una scuola il numero di ritardi giornalieri degli studenti, registrati per quindici giorni, è raccolto nella tabella

giorno	n. ritardi	giorno	n. ritardi	giorno	n. ritardi
1	4	6	9	11	11
2	13	7	6	12	17
3	11	8	5	13	3
4	7	9	3	14	5
5	7	10	2	15	11

a. Trova la media, la mediana e la moda dei ritardi nella prima settimana (cioè nei primi sei giorni) e nel periodo complessivo di 15 giorni preso in esame.

Soluzione. Nei primi sei giorni si sono avuti in totale 51 ritardi e la media è

$$\overline{m} = \frac{51}{6} = 8,5$$

la mediana richiede l'ordinamento dei ritardi, $4,7,7,9,11,13$ e pertanto

$$med = \frac{7+9}{2} = 8$$

la moda sempre nei primi sei giorni è 7 che compare due volte.

Calcoliamo ora per l'intero periodo:

$$\overline{m} = \tfrac{114}{15} = 7,6 \quad med = 7 \quad moda = 11$$

Esercizio 49. La ripartizione degli stipendi in un'azienda è rappresentata nella tabella. Determina: **a.** lo stipendio medio; **b.** lo stipendio mediano; **c.** la classe modale.

Stipendio S	frequenza
$1000 \le S < 1400$	6
$1400 \le S < 1800$	10
$1800 \le S < 2200$	4

Soluzione. Per calcolare la media prendiamo il valore centrale di ogni classe, cioè $1200, 1600, 2000$. Avremo

$$\overline{S} = \frac{1200 \cdot 6 + 1600 \cdot 10 + 2000 \cdot 4}{20} = 1560$$

la mediana cade nella seconda classe, per cui è 1600; la moda è sempre la seconda classe $1400 \le S < 1800$.

Varianza, deviazione standard e coefficiente di variazione

In un insieme di dati si chiama campo di variazione la differenza tra il valore più grande e quello più piccolo. La tendenza di certi dati numerici a disporsi intorno ad un valore medio è detta dispersione o variazione dei dati.

Viene detto scarto la differenza tra ogni dato e il valore medio dell'insieme di dati. Si definisce altresì lo scarto medio cioè la somma di tutti gli scarti divisa per il numero dei dati. Questo indice non è però molto significativo in quanto la somma degli scarti positivi è uguale a quella degli scarti negativi, dando quindi una media nulla. Si definisce allora la media degli scarti in valore assoluto

$$Scarto\ medio\ assoluto = \sum_{i=1}^{n} \frac{|x_i - \overline{x}|}{n}$$

il modo però preferito per superare il problema dei segni è elevare gli scarti al quadrato e si definisce allora

$$varianza(\sigma^2) = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n} = \sum_{i=1}^{n} \frac{x_i^2}{n} - \overline{x}^2 = (x_1 - \overline{x})^2 f_1 + (x_2 - \overline{x})^2 f_2 + ... + (x_n - \overline{x})^2 f_n$$

se i dati sono raggruppati indicandone la frequenza, la formula diviene

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2 \cdot f_i}{f_i} = (x_1 - \overline{x})^2 f_{r_1} + (x_2 - \overline{x})^2 f_2 + ... + (x_n - \overline{x})^2 f_n$$

Si definisce deviazione standard (o scarto quadratico medio) la radice quadrata della varianza

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}}$$

Il calcolo di questi indici è di solito lungo; lo svolgeremo sempre utilizzando un foglio di calcolo elettronico quando i dati saranno numerosi.

La deviazione standard è un indice particolarmente importante, perché consente di dedurre informazioni sulla distribuzione dei dati intorno alla media. Infatti, più la deviazione standard è piccola, più raggruppati e vicini alla media sono i dati. Si può infatti dimostrare che, qualsiasi sia la distribuzione dei dati, detta $\overline{x}$ la loro media e σ la loro deviazione standard:

- almeno il 75% dei dati cade nell'intervallo $(\overline{x} \pm 2\sigma)$;
- almeno l'89% dei dati cade nell'intervallo $(\overline{x} \pm 3\sigma)$.

Si chiama infine coefficiente di variazione, C_V, il rapporto tra la deviazione standard e la media aritmetica dei dati

$$C_V = \frac{\sigma}{\overline{x}}$$

Esercizio 50. Trovare la varianza, la deviazione standard e il coefficiente di variazione (se esiste) del seguente gruppo di numeri: $-2, 2, -4, 4, 5, -5$. (Arrotondare i risultati alla seconda cifra decimale).

Soluzione. I numeri sono tra loro a due a due opposti e la loro media è zero. Troviamo la varianza

$$\sigma^2 = \frac{4+4+16+16+25+25}{6} = \frac{90}{6} = 15$$

$$\sigma = \sqrt{15} \simeq 3,87$$

essendo la media uguale a zero C_V non è definito.

Esercizio 51. Trovare la varianza, la deviazione standard e il coefficiente di variazione (se esiste) del seguente gruppo di numeri: $-3, -2, , 4, 5, 6$.

Soluzione. Calcoliamo la media aritmetica dei numeri

$$\bar{x} = \frac{-3-2+4+5+6}{5} = 2$$

calcoliamo gli scarti e inseriamo il tutto in una tabella

x	$x-\bar{x}$	$(x-\bar{x})^2$
-3	$-3-2=-5$	25
-2	$-2-2=-4$	16
4	$4-2=2$	4
5	$5-2=3$	9
6	$6-2=4$	16

avremo

$$\sigma^2 = \frac{25+16+4+9+16}{5} = 14$$

e

$$\sigma = \sqrt{14} \simeq 3,74 \quad C_V = \tfrac{3,74}{2} = 1,87$$

Esercizio 52. Calcola media $\bar{x}$ e varianza V dei numeri x_i, con $i = 1, 2, 3$, completando la tabella.

x_i	f_i
2	3
3	4
6	1
Totale	8

Soluzione. Completiamo la tabella con tutti i calcoli per ottenere quanto chiesto

x_i	f_i	$x_i \cdot f_i$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2 f_i$
2	3	6	1	3
3	4	12	0	0
6	1	6	9	9
Totale	8	$\bar{x} = \frac{24}{8} = 3$		$\sigma^2 = \frac{12}{8} = 1,5$

nei manuali di statistica si trova che $\sigma^2 = \sum_{i=1}^{n} \frac{(x_i-\overline{x})^2}{n-1} = \frac{12}{7} = 1,71$; questo valore è quello che viene restituito dai fogli elettronici o da sw che analizzano i dati statistici e riguarda l'analisi statistica di grandi popolazioni che svengono studiate tramite dei campioni opportunamente scelti.

Esercizio 53. Determinare per quale valore di $a > 0$ la varianza dei numeri $a, 2a, 3a, 4a$, è uguale a 45.

Soluzione. calcoliamo i diversi indici che intervengono nel calcolo

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
a	-1,5a	$2,25a^2$
2a	-0,5a	$0,25a^2$
3a	0,5a	$0,25a^2$
4a	1,5a	$2,25a^2$
$\overline{x} = \frac{10}{4}a = 2,5a$		$\sigma^2 = \frac{5a^2}{4}$

se la varianza deve essere 45, allora

$$\frac{5}{4}a^2 = 45 \quad a^2 = 36 \quad a = 6$$

Esercizio 54. Uno studente ha ottenuto, in quattro compiti in classe, i seguenti voti: 5, 7, 6, 8. **a.** Calcolare la media e la deviazione standard dei voti. **b.** Se il professore decidesse di aumentare i voti di tutti i compiti in classe di 1 punto, quali diventerebbero la media e la deviazione standard dei voti?

Soluzione. calcoliamo i diversi indici che intervengono nel calcolo

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
5	-1,5	$2,25$
7	0,5	$0,25$
6	-0,5	$0,25$
8	1,5	$2,25$
$\overline{x} = \frac{26}{4} = 6,5$		$\sigma^2 = \frac{5}{4} = 1,25$

la deviazione standard è

$$\sigma = \sqrt{\frac{5}{4}} = 1,12$$

b) se i voti aumentano di 1, la somma di tutti quanti aumenta di 4 e si ha $\overline{x} = \frac{30}{4} = 7,5$, cioè aumenta di 1, mentre la deviazione standard non cambia perché gli scarti si mantengono uguali.

,

Esercizio 55. Si sono registrati i tempi di attesa a un call center, considerando un campione di 100 clienti, e si sono ottenuti i dati riportati nella tabella. Determinare la media, la varianza e la deviazione standard dei tempi di attesa.

tempo t (sec)	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$
Numero clienti	12	16	22	28	14	8

Soluzione. Distribuzione in classi di cui prenderemo il valore centrale

$$\bar{t} = \frac{5 \cdot 12 + 15 \cdot 16 + 25 \cdot 22 + 35 \cdot 28 + 45 \cdot 14 + 55 \cdot 8}{100} = 29\,s$$

t valori centrali	5	15	25	35	45	55
$(x_i - \bar{x})^2$	576	196	16	36	256	676
$(x_i - \bar{x})^2 f_i$	6912	3136	352	1008	3584	5408
$\sigma^2 = \frac{20400}{100} = 204$	$\sigma = \sqrt{204} \simeq 14,3$					

Esercizio 56. In due ospedali, A e B, si è registrato il seguente numero di nascite mensili. Si sono ottenuti i dati riportati nella tabella.

Mese	Gen	Feb	Mar	Apr	Mag	Giu
Ospedale A	22	28	18	12	24	20
Ospedale B	58	62	56	52	44	48

a. Calcolare la media, la varianza e il coefficiente di variazione delle nascite mensili nell'ospedale A.
b. Calcolare la media, la varianza e il coefficiente di variazione delle nascite mensili nell'ospedale B.
c. Nel primo semestre c'è stata più variabilità del numero di nascite mensili nell'ospedale A o nell'ospedale B?

Soluzione. Calcoliamo le medie

$$\bar{x}_A = \frac{124}{6} = 20,7 \quad \bar{x}_B = \frac{320}{6} = 53,3$$

costruiamo la tabella degli scarti assoluti

Ospedale A	Gen	Feb	Mar	Apr	Mag	Giu
$(x_i - \bar{x})$	1,3	7,3	$-2,7$	-8,7	3,3	$-0,7$
$(x_i - \bar{x})^2$	1,69	53,29	7,29	75,69	10,89	0,49
Ospedale B	Gen	Feb	Mar	Apr	Mag	Giu
$(x_i - \bar{x})$	4,7	8,7	2,7	-1,3	-9,3	$-5,3$
$(x_i - \bar{x})^2$	22,09	75,69	7,29	1,69	86,49	28,09

Calcoliamo la varianza e il coefficiente di variazione

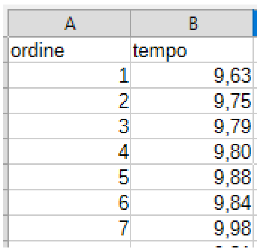
$$\sigma_A^2 = \frac{148,74}{6} = 24,8 \quad \sigma_B^2 = \frac{221,34}{6} = 36,9$$
$$C_{V_A} = \frac{\sqrt{24,8}}{20,7} = 0,24 \quad C_{V_B} = \frac{\sqrt{36,9}}{53,3} = 0,11$$

la variabilità maggiore si è avuta nell'ospedale A.

Esercizio 57. Nella tabella sono riportati i tempi dei finalisti della gara dei 100 metri piani maschili ai giochi olimpici del 2012. Calcolare la media e la deviazione standard dei tempi.

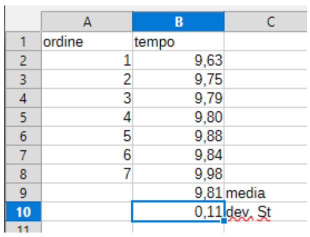
Ordine d'arrivo	Atleta	nazione	tempo
1	Bolt	Giamaica	$9,63$
2	Blake	Giamaica	$9,75$
3	Gatlin	USA	$9,79$
4	Gay	USA	$9,80$
5	Balley	USA	$9,88$
6	Martina	Paesi Bassi	$9,94$
7	Thompson	Trinidad Tobago	$9,98$

Soluzione. Risolviamo utilizzando le opportunità offerte dai fogli elettronici che consentono una rapida esecuzione del calcolo. Scriviamo i valori numerici nelle diverse caselle come mostrato in figura



A	B
ordine	tempo
1	9,63
2	9,75
3	9,79
4	9,80
5	9,88
6	9,84
7	9,98

per calcolare la media nella casella B9, inseriamo la funzione $= media\,(b2:b8) = 9,81$; per calcolare la deviazione standard, la funzione è $= dev.st\,(b2:b8) = 0,11$



	A	B	C
1	ordine	tempo	
2	1	9,63	
3	2	9,75	
4	3	9,79	
5	4	9,80	
6	5	9,88	
7	6	9,84	
8	7	9,98	
9		9,81	media
10		0,11	dev. St

per ottenere le due cifre decimali basta cliccare con il tasto destro sulle due caselle B9 e B10 e selezionare "Formatta celle" e nella finestra che si apre impostare "cifre decimali" uguale a 2.

Esercizio 58. Un grande magazzino ha rilevato le taglie di pantaloni vendute in una settimana. La tabella riporta la distribuzione delle frequenze relative dei dati raccolti.

Taglia	44	46	48	50	52	54
freq. rel.	$0,08$	0,15	$0,20$	—-	0,15	$0,09$

a) Trovare la frequenza relativa della taglia 50 e la moda della distribuzione.

Supposto che nella settimana in cui sono stati rilevati i dati siano stati venduti complessivamente 200 pantaloni:

b. Determinare quanti pantaloni della taglia 50 sono stati venduti.

c. Calcolare la taglia media e la taglia mediana.

d. Calcolare la deviazione standard delle taglie dei pantaloni.

Soluzione. La somma di tutte le frequenze relative deve dare l'unità, 1, per cui

$$f.r_{50} = 1 - (0.08 + 0.15 + 0.20 + 0.15 + 0.09) = 0,33$$

questo valore fa sì che la taglia modale sia 50.

b) Se i pantaloni venduti sono 200, quelli della taglia 50 sono $200 \times 0,33 = 66$

c) calcoliamo la taglia media utilizzando i dati assegnati

$$\bar{t} = \frac{44 \cdot 0,08 + 46 \cdot 0,15 + 48 \cdot 0,20 + 50 \cdot 0,33 + 52 \cdot 0,15 + 54 \cdot 0,09}{1} = 49,18$$

la taglia mediana, cioè quella centrale, è 50, perché in questa taglia cade il centesimo pantalone.

Esercizio 59. Analisi guidata dei voti di una classe di studenti. Partiamo dalla tabella che presenta i voti ottenuti dagli studenti al termine dell'anno scolastico nelle singole discipline. (il foglio elettronico usata è "calc" della suite di Libreoffice (i comandi sono comunque pressoché identici anche in altre applicazioni)

n°	taliano	Latino	Greco	St/geo	Ing	Mat	cienze
1	7	7	7	7	6	5	7
2	6	6	6	6	6	6	7
3	6	6	7	6	7	6	7
4	5	4	4	5	5	4	4
5	6	6	6	7	6	6	7
6	7	8	7	7	9	6	7
7	6	6	6	6	6	6	6
8	7	7	7	7	7	7	8
9	8	8	8	7	8	7	8
10	6	6	6	7	6	6	7
11	8	9	10	9	8	9	9
12	6	6	7	7	7	6	8
13	6	8	7	7	4	5	8
14	7	7	8	8	8	6	8
15	8	8	8	8	8	7	8
16	6	6	6	7	6	6	6
17	6	6	8	6	7	7	8
18	6	7	7	7	6	6	7
19	6	4	4	6	5	5	6
20	7	7	8	8	9	6	9
21	6	6	6	6	7	6	7
22	6	6	6	6	6	7	8
23	7	7	8	7	8	8	8
24	5	5	4	6	4	5	6

- Evidenziamo i voti negativi con un colore in modo da renderli facilmente visibili. Selezioniamo tutte le caselle contenenti i voti e nel menu scegliamo `Formato/Condizionale/Condizione/Minore`, inseriamo minore di 6 e il colore da assegnare alle caselle, avremo

n°	taliano	Latino	Greco	St/geo	nglese	Mat	cienze
1	7	7	7	7	6	5	7
2	6	6	6	6	6	6	7
3	6	6	7	6	7	6	7
4	5	4	4	5	5	4	4
5	6	6	6	7	6	6	7
6	7	8	7	7	9	6	7
7	6	6	6	6	6	6	6
8	7	7	7	7	7	7	8
9	8	8	8	7	8	7	8
10	6	6	6	7	6	6	7
11	8	9	10	9	8	9	9
12	6	6	7	7	7	6	8
13	6	8	7	7	4	5	8
14	7	7	8	8	8	6	8
15	8	8	8	8	8	7	8
16	6	6	6	7	6	6	6
17	6	6	8	6	7	7	8
18	6	7	7	7	6	6	7
19	6	4	4	6	5	5	6
20	7	7	8	8	9	6	9
21	6	6	6	6	7	6	7
22	6	6	6	6	6	7	8
23	7	7	8	7	8	8	8
24	5	5	4	6	4	5	6

- Calcoliamo la media dei voti di ogni studente, di ogni materia e dell'intera classe. Selezioniamo le celle contenenti i valori sulla riga (per la media studente), e nella casella a fianco digitiamo `=media(b2:h2)`. Per estendere il risultato della media a tutti gli studenti basta selezionare la casella della media e, puntando il mouse nel piccolo quadrato nero in basso a destra della casella, trascinare fino all'ultimo studente; selezioniamo ora le celle di una materia per tutti gli studenti e nella casella sottostante digitiamo `=media(b2:b25)`. Estendere il risultato a tutte le materie come prima.

Per calcolare infine la media complessiva di tutti i volti della classe, in una casella vuota digitare `=media(b2:h25)`.

Studente	media				
1	6,6		6,637	media	classe
2	6,1				
3	6,4				
4	4,4				
5	6,3				
6	7,3				
7	6,0				
8	7,1				
9	7,7				
10	6,3				
11	8,9				
12	6,7				
13	6,4				
14	7,4				
15	7,9				
16	6,1				
17	6,9				
18	6,6				
19	5,1				
20	7,7				
21	6,3				
22	6,4				
23	7,6				
24	5,0				

Potrebbe essere interessante raggruppare i voti di tutti gli studenti in classi, cioè determinare la percentuale di voti insufficienti e non relativi a tutte la materie; per questo usiamo la funzione Frequenza. Riferendoci alla figura successiva

	Distribuzione voti						
	Ita	Lat	Gre	St/Ge	Ing	Mat	Sci
5	8,3%	12,5%	12,5%	4,2%	16,7%	20,8%	4,2%
7	79,2%	66,7%	58,3%	79,2%	54,2%	70,8%	50,0%
8	12,5%	16,7%	25,0%	12,5%	20,8%	4,2%	37,5%
10	0,0%	4,2%	4,2%	4,2%	8,3%	4,2%	8,3%

Le classi sono quelle indicate in figura. Nella seconda colonna, digitiamo a fianco di 5 "`=FREQUENZA(A2:A25;BN$6:BN$9)/$A$25`" (BN è la colonna in cui sono inseriti. ad es., i voti 5,7,8,10 e \$A\$25 è la casella in cui è inserito il numero progressivo dell'ultimo studente. Il simbolo del \$ indica al foglio elettronico che quel riferimento deve considerarsi come fisso e rimane invariato anche nelle celle successive.

Il risultato riempirà le quattro celle della colonna Ita e trascinando tali celle si otterranno anche gli altri valori. La divisione per 24 (A25) serve per calcolare direttamente la percentuale formattando le celle indicando che devono contenere numeri percentuali.

Variabile standardizzata

È possibile introdurre una nuova variabile che misura le deviazioni dalla media dove l'unità di misura è la deviazione standard; tale variabile detta, standardizzata, è una quantità adimensionale, è data da

$$z = \frac{x - \bar{x}}{\sigma}$$

Esercizio 60. Uno studente ha ottenuto il voto medio 84 all'esame finale di matematica, per il quale la media generale è stata 76 e la deviazione standard 10. All'esame di fisica, per il quale la media generale è stata uguale al voto 82 e la deviazione standard 16, lo studente ha ottenuto il voto 90. Stabilire in quale materia il voto è stato relativamente più alto.

Soluzione. Calcoliamo la variabile standardizzata in entrambi e casi:

$$z_M = \frac{84-76}{10} = 0,8 \quad z_F = \frac{90-82}{16} = 0,5$$

si vede che il voto più alto relativamente ai risultati dell'intera classe è quello in matematica.

Esercizio 61. La variabile standardizzata è stata proposta come metodo di valutazione delle verifiche degli studenti. Il voto finale è stabilito non in termini assoluto con un criterio basato su un punteggio prestabilito, ma valutato rispetto alla classe. La tabella presenta un possibile risultato di una verifica di matematica o di fisica che ha un punteggio massimo di 100.

studente	punteggio	voto	studente	punteggio	voto
1	72	7	11	85	8
2	83	8	12	91	9
3	51	5	13	40	4
4	42	4	14	50	5
5	60	6	15	42	4
6	54	5	16	34	3
7	43	4	17	40	4
8	65	6	18	66	6,5
9	74	7	19	56	5,5
10	62	6	20	65	6

si chiede di valutare mediante la variabile standardizzata.

Soluzione. Calcoliamo la media dei punteggi e la loro deviazione standard

$$m = 58,8 \quad \sigma = 16,5$$

la deviazione standard risulta alta indicando una dispersione significativa tra i punteggi. La votazione è costruita nel modo seguente:

punteggio	z	voto	punteggio	z	voto
72	0,80	7,5	85	1,59	8,5
83	1,47	8	91	1,96	9
51	-0,47	6	40	-1,14	4
42	-1,02	4	50	-0,53	5,5
60	0,08	6	42	-1,02	4,5
54	-0,29	6	34	-1,50	4
43	-0,96	5	40	-1,14	4,5
65	0,38	6	66	0,44	6
74	0,93	8	56	-0,17	6
62	0,20	6	65	0,38	6

Interpolazione dei dati

Qualche volta è possibile costruire una relazione matematica che colleghi due variabili statistiche, cioè passare da una serie di dati raccolti in forma discreta ad una equazione che esprime la relazione per tutti i dati possibili di quella variabile, anche se non sono stati raccolti e provare a prevedere quali saranno i valori della grandezza oltre il campione o l'insieme raccolto. Se si ha un campione di dati di due variabili x e y e si vuole studiare la possibile relazione che le lega, si inizia col rappresentare i valori delle due variabili in un piano cartesiano. Avremo quindi una serie di punti, detto diagramma di dispersione. Interpolare tali dati significa la curva che meglio si avvicina a tutti i valori.

Vedremo in particolare i casi di interpolazione lineare mediante rette, cosa per altro non sempre possibile. Le curve possono essere rappresentate da funzioni polinomiali o trascendenti.

Il calcolo per l'interpolazione è basato sul metodo dei minimi quadrati, cioè sulla ricerca della curva per la quale la somma di tutti i quadrati delle differenze tra le ordinate della curva e dei punti trovati è minima. Ovviamente delegheremo tale calcolo ai fogli elettronici, analizzando in particolare le cosiddette serie temporali, cioè i valori di una grandezza in tempi diversi.

Nel caso in cui si effettui una interpolazione lineare, si tratta di ricavare l'equazione di una retta che meglio approssima il numero dei dati disponibili.

L'equazione della retta cercata è del tipo $y = a_1x + a_0$, dove a_1 è il coefficiente angolare e a_0 l'ordinata all'origine. La teoria mostra che per ottenere tale retta basta risolvere il sistema

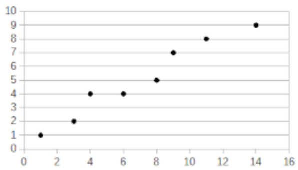
$$\begin{cases} \sum y = a_0 N + a_1 \sum x \\ \sum xy = a_0 \sum x + a_1 \sum x^2 \end{cases}$$

dove le sommatorie sono estese al numero N dei dati.

Esempio. Assegnati i seguenti valori

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

rappresentiamo il diagramma di dispersione



per applicare il sistema e risolverlo, costruiamo la tabella

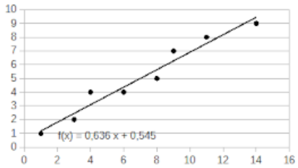
x	y	x^2	xy
1	1	1	1
3	2	9	6
4	4	16	16
6	4	36	24
8	5	64	40
9	7	81	63
11	8	121	88
14	9	196	126
$\sum x = 56$	$\sum y = 40$	$\sum x^2 = 524$	$\sum xy = 364$

il sistema diventa, sapendo che $N=8$

$$\begin{cases} 40 = 8a_0 + 56a_1 \\ 364 = 56a_0 + 524a_1 \end{cases} \quad \begin{cases} a_0 = 5 - 7a_1 \\ 364 = 56(5-7a_1) + 524a_1 \end{cases} \quad \begin{cases} a_0 = 0,545 \\ a_1 = \frac{84}{132} = 0,636 \end{cases}$$

l'equazione della retta interpolatrice è $y = 0,636x + 0,545$.

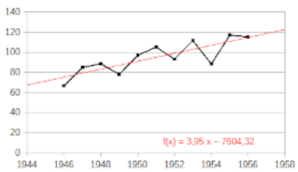
Il grafico sotto mostra quanto ottenuto con un foglio elettronico (Calc)



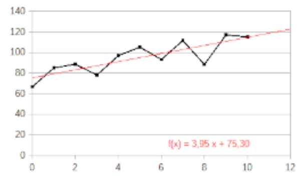
Esercizio 62. Nella tabella è riportata la produzione di acciaio di un certo paese in milioni di tonnellate durante gli anni 1946-1956. a) Tracciare il grafico; b) trovare l'equazione della retta dei minimi quadrati; c) stimare la produzione di acciaio per gli anni 1957 e 1958 e confrontarli con i veri valori, che sono rispettivamente 112,7 e 85,3.

Anno	Produzione
1946	66,6
1947	84,9
1948	88,6
1949	78,0
1950	96,8
1951	105,2
1952	93,2
1953	111,6
1954	88,3
1955	117,0
1956	115,2

Soluzione. Ci affidiamo alle potenzialità di un foglio elettronico (Calc).



l'equazione della retta è $y = 3,95x - 7604,32$. (il valore di q della retta è legato alla scelta dei valori sull'asse x; si potrebbero sostituire assegnando al 1946 il valore 0. Con questa scelta si ottiene



per ottenere la retta interpolante basta cliccare sui dati e selezionare poi ”linea di tendenza” scegliendo in questo caso come curva la retta.

c) Stima dei valori negli anni successivi tramite l’estrapolazione, supponendo che anche i dati si attestino intorno alla retta.

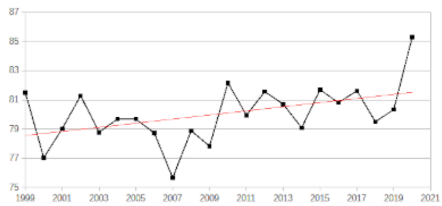
$$y(1957) = 125,83 \quad y(1958) = 129,78$$

come si può vedere, soprattutto il secondo valore è alquanto lontano da quello stimato.

Esercizio 63. Nella tabella sono riportati i punteggi medi ottenuti di tutti gli studenti negli esami di maturità nel corso di diversi anni. Si chiede di rappresentarli graficamente e di trovare la retta interpolante.

Anno	Punteggio medio
1999	81,49
2000	77,03
2001	79,03
2002	81,28
2003	78,78
2004	79,68
2005	79,68
2006	78,72
2007	75,66
2008	78,88
2009	77,83
2010	82,16
2011	79,95
2012	81,56
2013	80,70
2014	79,08
2015	81,69
2016	80,81
2017	81,62
2018	79,50
2019	80,33
2020	85,29

Soluzione. Sempre con foglio elettronico rappresentiamo questi dati in un grafico a dispersione con gli anni sull’asse orizzontale e le votazioni medie sull’asse verticale. Ancora al foglio elettronico chiediamo di rappresentare la linea di tendenza lineare. Il risultato è il seguente:



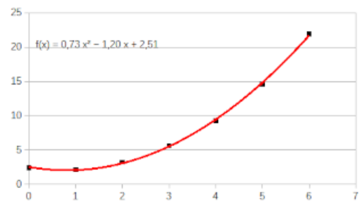
La linea di tendenza mostra un andamento crescente nei risultati e fluttuazioni sensibili al variare delle modalità di svolgimento degli esami di maturità.

Per ridurre il peso delle fluttuazioni è possibile ottenere una dispersione minore utilizzando una media mobile, cioè la media di tre anni in tre anni (il numero degli anni è modificabile a piacere). Per ottenere questo nuovo grafico di dispersione basta cliccare sempre sui dati e scegliere ”media mobile” impostando il numero degli anni su cui calcolare.

Esercizio 64. Interpolare i dati della tabella con una linea di tendenza di secondo grado:

x	0	1	2	3	4	5	6
y	2,4	2,1	3,2	5,6	9,3	14,6	21,9

Soluzione. I dati in tabella mostrano una successione dei valori di y che non segue un andamento descrivibile da una relazione di tipo proporzionalità diretta. La procedura con il foglio elettronico è sempre la stessa, basta scegliere una linea di tendenza polinomiale di secondo grado (si avrà una curva che descrive un tratto di parabola).



Esercizio 65. Nella tabella è riportato il numero di batteri per unità di volume presenti in una coltura dopo x ore. Tracciare il grafico e interpolare i dati con la curva più opportuna.

N. Ore	0	1	2	3	4	5	6
N. batteri	32	50	72	99	150	220	295

Soluzione. Le ore aumentano di uno ogni volta mentre il numero dei batteri cresce assai più rapidamente. Rappresentiamo i dati e scegliamo una linea di tendenza data da una funzione di tipo esponenziale.