Prerequisiti:

- Saper operare con i numeri razionali.
- Riconoscere e usare correttamente le diverse rappresentazioni dei numeri (frazioni, numeri decimali, percentuali).
- Conoscere i valori approssimati.
- Possedere le nozioni fondamentali della logica.

L'unità è indirizzata agli studenti del primo biennio di tutte le scuole superiori.

OBIETTIVI DI APPRENDIMENTO

Una volta completata l'unità, gli allievi devono essere in grado di:

- *riconoscere le modalità sotto cui si presenta un fenomeno statistico*
- *raccogliere dati e rappresentarli con tabelle e grafici*
- *rappresentare con tabelle e grafici una serie storica*
- *passare dai dati grezzi alle distribuzioni statistiche di frequenze e alle corrispondenti rappresentazioni grafiche*
- *leggere e interpretare una tabella e un grafico*
- *definire e calcolare le medie più significative di un insieme di dati (moda, mediana, medie aritmetica e armonica) e, nel caso di due soli dati, la media geometrica*
- *dimostrare che è nulla la media aritmetica degli scarti dalla media aritmetica dei dati*
- *conoscere e dimostrare la relazione che lega media aritmetica, media geometrica e media armonica di due dati numerici non negativi*
- *definire lo scarto semplice medio e lo scarto quadratico medio*
- *calcolare i valori medi e le misure di variabilità per mezzo di un foglio elettronico*

Statistica descrittiva

Unità 12

12.1 SCOPO DELLA STATISTICA

12.1.1 La **statistica** studia i fenomeni in cui è coinvolto un numero elevato di individui.

Per esempio:
- numero di nati o di morti di una certa comunità in un determinato periodo;
- numero di incidenti avvenuti in un determinato periodo in una data città;
- costo delle merci acquistate da una data azienda in un certo intervallo di tempo;
- voti riportati in prove d'esame dagli studenti di una determinata scuola;
- misure di una data grandezza, ottenute sperimentalmente in laboratorio;

eccetera.

Fenomeni siffatti si dicono ***fenomeni collettivi*** o ***statistici***. Nel loro studio non interessa il comportamento del singolo individuo in quanto tale, ma il suo comportamento come componente dell'insieme di individui che si considera.

Ogni singolo individuo è chiamato ***unità statistica***. L'insieme delle unità statistiche che si vuole studiare si dice ***popolazione*** o ***collettivo statistico***.

Per esempio, se si vuole studiare la statura dei componenti di una certa comunità, ogni altezza rilevata è una unità statistica; il loro insieme è il collettivo statistico.

12.1.2 Il termine "statistica" significa letteralmente "scienza dello Stato" e fu introdotto nel XVII secolo, come attività idonea a raccogliere informazioni utili all'amministrazione pubblica. Oggigiorno non ha perso questa caratteristica, anzi l'ha ampliata.

La statistica si considera di solito sotto due aspetti:
- **statistica descrittiva** (o semplicemente **statistica**),
- **statistica inferenziale** (detta anche **inferenza statistica** e, a volte, **statistica induttiva** o **statistica matematica).**

- Scopo della *statistica descrittiva*, dopo aver definito l'oggetto su cui svolgere l'indagine, è di raccogliere i dati, coerenti con lo scopo prefisso, relativi a tutti gli individui che compongono il collettivo o ad un opportuno campione dello stesso: sono detti ***dati statistici***. Alla raccolta dei dati segue la loro organizzazione e interpretazione.

  Ciò avviene rappresentando i dati medesimi in apposite tabelle ed eventualmente su opportuni grafici; quindi vengono riassunti e descritti per mezzo di alcuni parametri numerici.
- La *statistica inferenziale*, prendendo in esame una parte del collettivo su cui verte l'indagine (detto *campione*), scelta in modo da rappresentare l'intera popolazione, ne ricava informazioni su tutto il collettivo.

  Naturalmente, le conclusioni cui si perviene mediante la statistica inferenziale, non hanno il crisma della certezza ma sono soltanto delle stime, peraltro affette da errore, del quale pure la statistica inferenziale ne valuta la portata.

In questa unità ci occuperemo della statistica descrittiva. La statistica inferenziale, invece, sarà affrontata, anche se non in modo approfondito e completo, da chi proseguirà gli studi nel secondo biennio, ma solo in alcuni indirizzi (1).

---

[1] Cfr.: Unità 81: Cenni di statistica inferenziale.

12.2 DEFINIZIONE DELL'OGGETTO. RILEVAZIONE DEI DATI. RAPPRESENTAZIONE TABULARE E GRAFICA

12.2.1 La prima cosa che deve esser chiara a chi conduce un'indagine statistica è l'***oggetto dell'indagine***; vale a dire su che cosa s'intende indagare e per quale scopo.

Due esempi per chiarire:

- Un insegnante che voglia programmare la sua azione didattica, può pensare di condurre un'indagine statistica sul livello di quelli che giudica i prerequisiti indispensabili che devono essere posseduti dai suoi alunni per il conseguimento di determinati obiettivi.
- Una ditta costruttrice di automobili, che si accinga a produrre una nuova vettura, può pensare di condurre un'indagine sui gusti dei potenziali compratori.

Altri esempi puoi fornirli da te.

12.2.2 Dopo aver precisato l'oggetto dell'indagine, bisogna ***raccogliere i dati*** riguardanti o tutti gli individui che compongono il collettivo (come nel primo degli esempi suddetti) o un campione degli stessi (come nel secondo esempio).

Nel primo caso la rilevazione dei dati si dice *totale*; nel secondo *parziale* (o *campionaria*).

I dati statistici raccolti sono ripartiti e catalogati su apposite ***tabelle statistiche***, le quali possono essere *semplici*, *complesse*, *a doppia entrata*.

Ogni tabella fornisce la ***distribuzione statistica*** relativa al fenomeno studiato.

- Una **tabella semplice**, della quale la tabella 1 fornisce un esempio, si sviluppa su due colonne (o su due righe, se conviene di più): nella prima sono indicati i vari aspetti (detti più esattamente ***modalità***) sotto cui si presenta il fenomeno collettivo su cui s'indaga; nell'altra sono indicate le quantità (o ***frequenze assolute***) degli individui che presentano la modalità corrispondente.

  La tabella 1 fornisce la distribuzione di frequenze dell'esito dello scrutinio della classe considerata.

Tab. 1 – A.S. 2003-04. Scrutinio finale. Classe 1ª A.

Modalità	PROMOSSI senza debito formativo	PROMOSSI con debito formativo	NON PROMOSSI	TOTALE
Frequenza assoluta	**15**	**10**	**2**	**27**

Sei invitato a calcolare, sul totale degli alunni della classe considerata sopra, le percentuali di:
1) promossi senza debito formativo, 2) promossi con debito formativo, 3) non promossi.
Quanto vale la somma di queste tre percentuali trovate?

- Una **tabella complessa** si sviluppa su più di due colonne, mettendo insieme più tabelle semplici (Tab. 2).

Tab. 2 – Distribuzione delle frequenze delle scuole secondarie di 2° grado nell'anno scolastico 1986-87.

Indirizzo di scuola	**Impegnate nella sperimentazione**			**Non sperimentali**	**TOTALE**
	globale	**parziale**	**tutte**		
Classico	94	206	300	1084	**1384**
Tecnico	78	450	528	820	**1348**
Professionale	37	186	223	533	**756**
Artistico	18	---	18	245	**263**
Totale	**227**	**842**	**1069**	**2682**	**3751**

Con riferimento a questa distribuzione, ti invitiamo a calcolare le percentuali delle scuole dell'indirizzo professionale impegnate nella sperimentazione sul totale delle scuole:
a) impegnate nella sperimentazione e dello stesso tipo; b) impegnate nella sperimentazione;
c) dello stesso tipo; d) comunque prese.
Ti proponiamo, inoltre, di sostituire questa distribuzione con quella che da essa si ottiene sostituendo ai valori assoluti i corrispettivi valori percentuali, calcolati rispetto al totale generale e di stabilire se è vera o falsa ciascuna delle seguenti affermazioni:

- le scuole impegnate nella sperimentazione sono poco meno del 30% del totale delle scuole;
- le scuole non sperimentali degli ordini tecnico e professionale sono complessivamente circa il 56% del totale delle scuole;
- la percentuale delle scuole non sperimentali dell'ordine classico è pressoché uguale alla somma delle percentuali delle scuole non sperimentali degli ordini tecnico e artistico.

- La tabella 3 fornisce la distribuzione per aree geografiche delle 1069 scuole secondarie di 2° grado che, nell'anno scolastico 1986-87, erano impegnate nella sperimentazione. Si tratta evidentemente di una *tabella semplice*.

Tab. 3 – Distribuzione per aree geografiche delle scuole secondarie di 2° grado nell'anno scolastico 1986-87.

Area Geografica	**NORD**	**CENTRO**	**SUD**	**TOTALE**
Numero di scuole	488	286	295	**1069**

Può essere interessante conoscere la distribuzione delle stesse scuole contemporaneamente per aree geografiche e per ordine di studi. In questo caso bisogna ricorre ad una *tabella a doppia entrata* (Tab. 4).

Tab. 4 – Distribuzione per aree geografiche e per ordine di studi delle scuole secondarie di 2° grado nell'anno scolastico 1986-87.

Area Geografica / **Ordine di studi**	**NORD**	**CENTRO**	**SUD**	**TOTALE**
Classico	151	94	55	**300**
Tecnico	218	125	185	**528**
Professionale	108	63	52	**223**
Artistico	11	4	3	**18**
Totale	**488**	**286**	**295**	**1069**

In generale: una **tabella a doppia entrata** fornisce la distribuzione dei dati statistici riguardanti due modalità di uno stesso fenomeno collettivo (oppure due fenomeni diversi).
Di questa tipologia di tabelle ci occuperemo in maniera più approfondita dopo il primo biennio [(2)].

Per il momento, con riferimento alla distribuzione precedente, ti invitiamo a calcolare le percentuali delle scuole dell'indirizzo tecnico e del SUD:
a) sul totale delle scuole dello stesso tipo; b) sul totale delle scuole della stessa area geografica;
c) sul totale generale.
Ti proponiamo, inoltre, di sostituire questa distribuzione con quella che da essa si ottiene mettendo al posto dei valori assoluti i corrispettivi valori percentuali, calcolati rispetto al totale generale.

12.2.3 Oltre che mediante tabelle, i dati statistici possono essere rappresentati anche graficamente per

[2] Cfr.: Unità 55: Nozioni di statistica bivariata.

mezzo di diagrammi cartesiani, istogrammi, eccetera. In realtà, la rappresentazione grafica non fornisce maggiori o diverse informazioni rispetto a quella tabulare. Ha il vantaggio, però, di rendere immediatamente comprensibile la distribuzione dei dati.

Un esempio. Un test (30 domande a scelta multipla con 4 alternative, di cui una sola corretta, per un punteggio massimo acquisibile di 30×3=90 punti), è stato somministrato a 80 studenti. Ha dato gli esiti riassunti nella tabella 5, dove le modalità sono state opportunamente "raggruppate in classi".

Tab. 5 – Distribuzione delle frequenze dei punteggi raggruppati in classi

classe	valore	frequenza
C1	≤ 0	0
C2	1 – 10	4
C3	11 – 20	7
C4	21 – 30	8
C5	31 – 40	10
C6	41 – 50	15
C7	51 – 60	14
C8	61 – 70	12
C9	71 – 80	8
C10	81 – 90	2

Questa distribuzione può essere rappresentata per mezzo di un ***istogramma*** (detto anche ***diagramma a barre*** – Fig. 1), che presenta sull'asse orizzontale le classi di punteggi e su quello verticale i numeri degli studenti (frequenze assolute) che li hanno ottenuti.

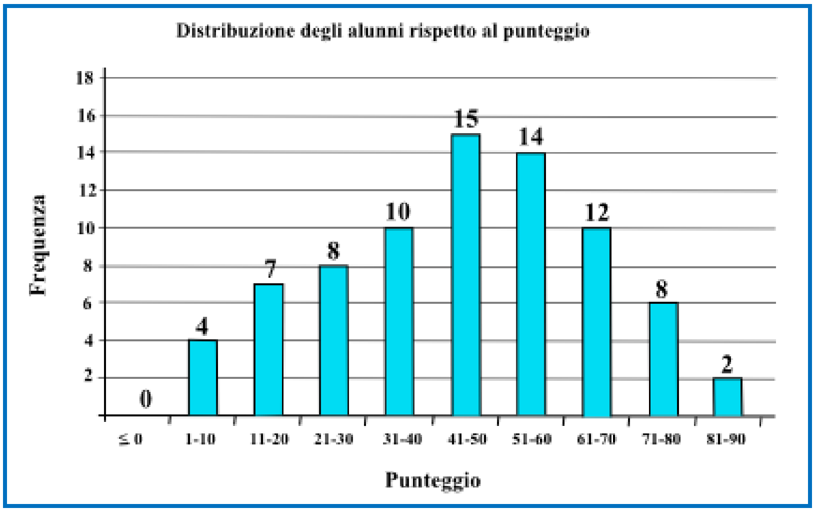


FIG. 1

12.2.4 In altre circostanze, invece di un istogramma, può essere più descrittivo un ***aerogramma circolare*** (detto a volte ***diagramma a torta***).

Ne vediamo un esempio. Nell'anno scolastico 2000/01 insegnavano nelle scuole italiane docenti, fra nominati a tempo indeterminato e nominati a tempo determinato, distribuiti in percentuale, tra i vari ordini di scuola, secondo la tabella 6.

Tab. 6 – Distribuzione delle percentuali di docenti per ordine di scuola. A.S. 2000/01					
Tipo di scuola	**Scuola dell'infanzia**	**Scuola elementare**	**Scuola media**	**Scuola superiore**	**Totale**
Percentuale di docenti	10,66	32,44	23,93	32,97	100

Il relativo aerogramma è rappresentato in figura 2.

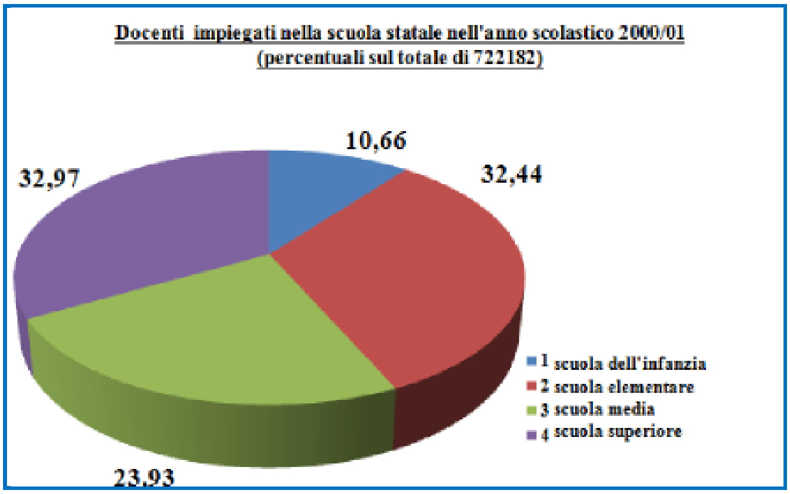


FIG. 2

12.2.5 Esempi interessanti di una successione di dati sono costituiti dalle cosiddette "serie storiche".

Una **serie storica** non è altro che una successione di dati ordinati secondo opportuni intervalli di tempo, come secolo, anno, mese, settimana, giorno, eccetera.

Qualche esempio è sufficiente per capire di cosa si tratti.

Al riguardo consideriamo le percentuali di studenti promossi, sul totale dei candidati ammessi, in una data regione, agli esami di Stato a partire dall'anno 1999 e fino all'anno 2007. La tabella 7, che rappresenta per l'appunto la serie storica di tali percentuali, fornisce i dati in questione:

Tab. 7 – Serie storica della percentuale di studenti promossi.									
anno	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
promossi	94,4	97,6	97,4	97,5	97,9	98,0	98,2	98,1	97,8

Un altro esempio di serie storica è fornito dall'andamento della popolazione italiana com'è stata rilevata nei censimenti nazionali effettuati ogni decennio a partire dal 1951 fino al 2011 (Tab. 8).

Tab. 8 – Serie storica della popolazione italiana residente.							
anno	**1951**	**1961**	**1971**	**1981**	**1991**	**2001**	**2011**
popolazione (in migliaia)	47.516	50.624	54.137	56.557	56.411	56.992	59.464

La rappresentazione può essere fatta mediante istogrammi ma anche per mezzo di diagrammi cartesiani: ti invitiamo a disegnare gli uni e gli altri in entrambe le situazioni descritte.

12.2.6 La compilazione di tabelle, che forniscono la distribuzione della popolazione di una data nazione, distinguendo per età e sesso, è effettuata dai cosiddetti Istituti di Statistica.

Per esempio, l'Istituto Centrale di Statistica Italiano (ISTAT), in seguito al censimento nazionale del 2001, ha fornito la seguente tabella (Tab. 9), in cui figura la distribuzione, per sesso e per classi di età, della popolazione residente in Italia.

Tab. 9 – Distribuzione della popolazione residente in Italia per sesso e classi di età- Censimento 2001

CLASSI DI ETÀ (in anni)	**(Espressi in migliaia)**		
	MASCHI	**FEMMINE**	**TOTALE**
< 5	1.344	1.274	2.618
5-9	1.375	1.304	2.679
10-14	1.441	1.365	2.806
15-19	1.518	1.446	2.964
20-24	1.739	1.685	3.424
25-29	2.138	2.109	4.247
30-34	2.284	2.260	4.544
35-39	2.314	2.310	4.624
40-44	2.025	2.041	4.066
45-49	1.850	1.889	3.739
50-54	1.895	1.954	3.849
55-59	1.620	1.705	3.325
60-64	1.657	1.807	3.464
65-69	1.427	1.653	3.080
70-74	1.229	1.574	2.803
75-79	913	1.373	2.286
80-84	445	790	1.235
85-89	268	574	842
90-94	88	241	329
95-99	13	49	62
100 e più	1	5	6
Totale	**27.584**	**29.408**	**56.992**

Si può trarre da questa tabella una conclusione immediata: su una popolazione di quasi 57 milioni di residenti, le femmine superano i maschi di più di 1,8 milioni di unità. In percentuale, i maschi sono il 48,4% e le femmine il 51,6%. Per cui il numero delle femmine supera quello dei maschi di circa il 3,2%.
La stessa tabella consentirebbe di trarre altre conclusioni interessanti e più dettagliate, ma, al fine di una lettura più immediata dei dati, può rivelarsi utile un istogramma (Fig. 3).
Si può desumere (cosa, lo ribadiamo, che poteva dedursi anche dalla tabella 9) che:

- fino all'età di circa 34 anni il numero dei maschi supera, anche se di poco, quello delle femmine;
- fra i 35 e i 44 anni il numero dei maschi e quello delle femmine sono pressoché uguali;
- a partire dai 45 anni il numero delle femmine supera quello dei maschi e lo scarto aumenta dopo il 60° anno di età.

Il che fa pensare che probabilmente nascono più maschi che femmine, pur con differenze minime, ma la durata media della vita è superiore per le femmine piuttosto che per i maschi.

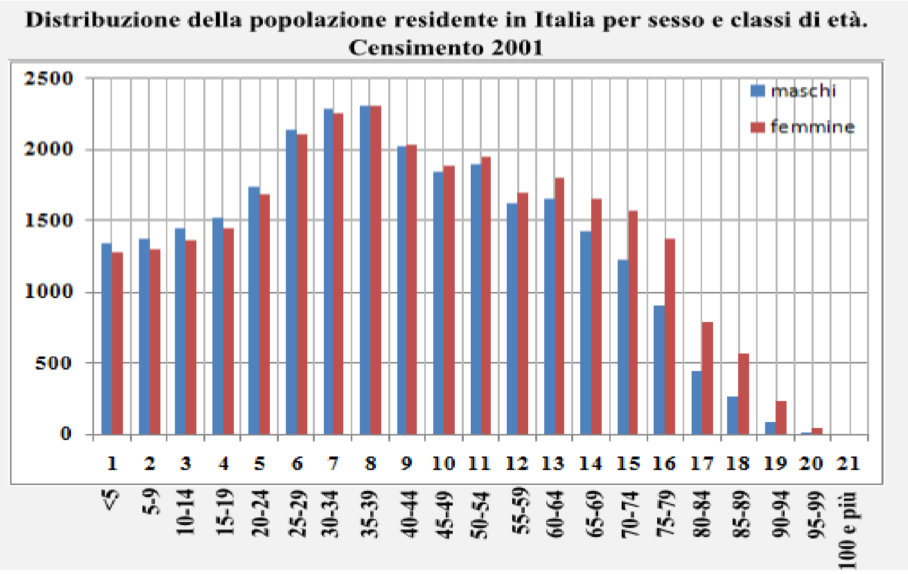


FIG. 3

12.3 VALORI MEDI

12.3.1 Se la tabulazione di una serie di dati statistici o la loro rappresentazione grafica hanno il pregio di descrivere in modo ordinato e comprensibile il fenomeno sul quale s'indaga, qualche volta è comodo, e spesso anche necessario, sintetizzare con un solo dato l'andamento del fenomeno, a condizione naturalmente che questo valore di sintesi sia effettivamente idoneo a riassumere le caratteristiche del collettivo che interessa evidenziare.

Per esempio, se vogliamo avere un'idea dell'altezza delle persone che compongono un collettivo (una squadra di calcio, una città, una nazione, eccetera), non è necessario conoscere le altezze delle singole persone. È sufficiente sintetizzare le varie altezze con un solo valore che riassuma l'altezza media delle persone del collettivo.

Di valori di sintesi di una serie di dati statistici, detti più propriamente **indici di posizione**, ce n'è più d'uno. Si chiamano genericamente **medie**. Ce ne vogliamo occupare diffusamente.

12.3.2 Supponiamo che i seguenti numeri:

[1] $$a_1,\ a_2, \dots, a_n$$

siano i dati statistici relativi ad un certo fenomeno. Si dice pure che essi costituiscono la successione dei valori assunti dalla **variabile statistica** A che descrive il fenomeno.

Tra i valori suddetti indichiamo con:

$$\mathbf{min}(\mathbf{A}) \quad \text{e} \quad \mathbf{max}(\mathbf{A})$$

rispettivamente il valore più piccolo (*minimo*) e quello più grande (*massimo*). Questo significa che, per ogni valore a_i (i=1,2,...,n) della successione [1], risulta:

$$\min(A) \le a_i \le \max(A).$$

- Per esempio, se l'insieme di dati è il seguente (i valori non sono necessariamente in ordine crescente o decrescente):

  $$36\,,\ 18\,,\ 20\,,\ 18\,,\ 20\,,\ 25\,,\ 22\,,\ 24\,,\ 20\,,\ 26\,,\ 23\,,$$

  il suo minimo e il suo massimo sono rispettivamente 18 e 36.

È evidente poi che, per ogni valore x dell'insieme – compresi quindi anche 18 e 36 – risulta:

$$18 \le x \le 36 .$$

12.3.3 Considerato l'insieme di dati [1], <u>ogni numero reale</u> m tale che:

$$\min(A) \le m \le \max(A)$$

si chiama **media** dei valori dell'insieme.

Con riferimento all'esempio precedente, una media dei valori ivi considerati è ogni numero reale m tale che:

$$18 \le m \le 36 ;$$

quindi è una media di quei valori il numero 20 che è uno degli elementi dell'insieme, ma è una media pure il numero 21,3 che non lo è.

Alcune medie particolari sono: la *mediana*, la *moda*, la *media aritmetica*. Trattiamone più in dettaglio.

- Supponiamo di scrivere in ordine non decrescente i valori assunti da una variabile statistica. Il termine di mezzo (se i dati sono in numero dispari) o la semisomma dei due termini centrali (se i dati sono in numero pari) si chiama ***mediana*** (o ***valore mediano*** o ***valore centrale***) dei dati considerati.

  Ritornando ai dati dell'esempio precedente, dopo averli scritti in ordine non decrescente:

  $$18\,,\ 18\,,\ 20\,,\ 20\,,\ 20\,,\ \underline{22}\,,\ 23\,,\ 24\,,\ 25\,,\ 26\,,\ 36\,,$$

  si può notare che sono in numero dispari (=11) e che il termine di mezzo è 22: esso rappresenta quindi la mediana dei dati considerati.

  Se i dati, invece, fossero questi, scritti in ordine non decrescente:

  $$18\,,\ 18\,,\ 20\,,\ 20\,,\ \underline{20}\,,\ \underline{22}\,,\ 23\,,\ 24\,,\ 25\,,\ 26\,,$$

  e quindi in numero pari (=10) la loro mediana sarebbe la semisomma dei due valori centrali; vale a dire il valore $\frac{20+22}{2}=21$.

- La ***moda*** (detta anche ***norma*** o ***valore modale*** o ***valore normale***) di un insieme di dati statistici è il valore che si presenta il maggior numero di volte, cioè con la maggior frequenza.

  In entrambi gli esempi precedenti essa è il valore 20.

  Può capitare che, nell'insieme dei dati, due o più di essi si presentino un uguale numero di volte, sempre con la maggior frequenza. La successione ha allora due o più mode: si dice rispettivamente *bimodale* o *plurimodale*.

  Un esempio di successione bimodale è la seguente:

  $$5, 7, 7, \underline{9}, \underline{9}, \underline{9}, \underline{9}, 12, 13, \underline{16}, \underline{16}, \underline{16}, \underline{16}, 19, 19, 21.$$

  Le due mode sono evidentemente: 9 e 16.

- La ***media aritmetica*** degli n valori [1] è il rapporto tra la somma degli n dati ed il loro numero. In simboli:

  $$\mathbf{M} = \frac{\mathbf{a_1} + \mathbf{a_2} + \cdots + \mathbf{a_n}}{\mathbf{n}} .$$

  Con riferimento al primo dei due esempi precedenti si ha:

  $$M=\frac{36+18+20+18+20+25+22+24+20+26+23}{11} \approx 22{,}9 .$$

Ti proponiamo alcuni esercizi (sono fra l'altro una buona occasione per affinate la tua abilità nel calcolo algebrico).

1. Che la media aritmetica M di n numeri dati sia effettivamente una "media" di quei numeri andrebbe

provato. Il che si fa dimostrando che M è un numero compreso fra il minore ed il maggiore dei numeri dati. Ti proponiamo di effettuare la dimostrazione nel caso di 2 numeri dati, a e b, e di darne anche un'interpretazione geometrica.

2. Constatato che la media aritmetica di due numeri si può considerare come l'operazione "M" che alla coppia a, b di numeri reali, comunque scelti, associa il numero reale $\frac{a+b}{2}$, vale a dire: $aMb=\frac{a+b}{2}$, tale operazione gode della proprietà associativa? In altri termini, è vero che, comunque si prendano i numeri reali a, b, c, risulta $(aMb)Mc=aM(bMc)$? Esprimi a parole.
3. La media aritmetica dei valori della quota di un titolo azionario, calcolata con riferimento alle ultime 28 rilevazioni, è 5,005. La 29ª rilevazione del valore della quota è 4,980. Qual è la nuova media aritmetica?
4. Sono dati 4 numeri qualsiasi a, b, c, d. Prendendoli a 3 a 3 in tutti i modi possibili ma senza ripetizioni, considera la media aritmetica di ciascuna delle terne ottenute (le terne sono 4). Dimostra che la media aritmetica delle medie aritmetiche delle 4 terne è uguale alla media aritmetica dei 4 numeri assegnati.
5. Mediana, moda e media aritmetica di n dati statistici solitamente non coincidono. C'è tuttavia una particolare distribuzione dei dati in cui tali medie coincidono. Sapresti indicare le caratteristiche di una tale distribuzione? Sapresti fornirne un esempio?

- Consideriamo ora il seguente PROBLEMA: Un'azienda paga ad 1/5 del suo personale uno stipendio mensile di € 2100, ai 3/10 del personale uno stipendio di € 1750 ed alla parte rimanente del personale uno stipendio di € 1450. Calcolare lo stipendio medio mensile dei dipendenti dell'azienda.

RISOLUZIONE. Sembrerebbe di poter concludere facilmente che questo stipendio sia la media aritmetica dei valori € 2100, € 1750, € 1450. Ossia:

$$\frac{2100+1750+1450}{3}\approx 1788{,}67\ (€).$$

Ma, a pensarci bene, non può essere così, in quanto l'azienda non paga solo tre stipendi, ma tre "tipi" di stipendi ad un certo numero di impiegati.

Ora, se N è il numero di questi impiegati, $\frac{1}{5}$N percepiscono lo stipendio di € 2100, $\frac{3}{10}$N percepiscono € 1750, mentre $\left[1-\left(\frac{1}{5}+\frac{3}{10}\right)\right]$N, cioè $\frac{1}{2}$N percepiscono € 1450. Quindi la media M degli stipendi è in realtà:

$$M=\frac{\frac{1}{5}N\cdot 2100+\frac{3}{10}N\cdot 1750+\frac{1}{2}N\cdot 1450}{N}=\frac{2\cdot 2100+3\cdot 1750+5\cdot 1450}{10}=1670\ (€)\,.$$

Questa media, che è pur sempre una media aritmetica, si chiama più propriamente ***media aritmetica ponderata***. L'altra media aritmetica, per distinguerla da essa, si chiama anche ***media aritmetica semplice***.

La media aritmetica ponderata viene presa in considerazione allorché i dati statistici risultano opportunamente raggruppati in classi ed ogni classe contiene un numero precisato (detto *peso*) di unità statistiche.

Ebbene, la media aritmetica ponderata di k valori $a_1, a_2, \dots, a_k$, presi rispettivamente con i pesi $p_1, p_2, \dots, p_k$, è il numero:

$$\mathbf{M_p=\frac{a_1p_1+a_2p_2+\cdots+a_kp_k}{p_1+p_2+p_k}}\,.$$

Nell'esempio precedente i valori presi in considerazione sono 3: il primo, 2100, con peso $\frac{1}{5}$N; il secondo, 1750, con peso $\frac{3}{10}$ N; il terzo, 1450, con peso $\frac{1}{2}$ N. La loro media aritmetica, tenendo presente che $\frac{1}{5}N+\frac{3}{10}N+\frac{1}{2}N=N$, è per l'appunto:

$$M=\frac{\frac{1}{5}N\cdot 2100+\frac{3}{10}N\cdot 1750+\frac{1}{2}N\cdot 1450}{N}=1670\,.$$

Ti proponiamo altri esercizi (l'ultimo dei quali è ancora un'occasione per affinare la tua tecnica di calcolo).

1. Un'azienda decide di dare ad 1/4 dei suoi dipendenti una gratifica di € 3000, ad 1/3 dei dipendenti una gratifica di € 1900 ed al resto dei dipendenti una gratifica di € 1500. Calcola la gratifica media dei dipendenti.
2. Una ditta di trasporti pratica i seguenti prezzi: 45 euro al quintale fino a 12 q, 32 euro per ogni quintale oltre i 12. Chiedo alla ditta il trasporto di 20 q di materiale. Quanto mi costa mediamente il trasporto di un quintale? [**R.** 39,8 €/q]
3. Dimostra che la media aritmetica ponderata gode della proprietà associativa. In altri termini, prese tre coppie ordinate qualsiasi di numeri $(a,\ m)$, $(b,\ n)$, $(c,\ r)$ – dove il primo numero di ogni coppia rappresenta un valore della variabile statistica che si considera ed il secondo, positivo, il peso relativo – e detta M_p l'operazione "media aritmetica ponderata", si ha:
$$\big((a,m)M_p(b,n)\big)\,M_p(c,r) = (a,m)M_p\,\big((b,n)M_p(c,r)\big).$$
Bisogna tener presente che:
$$(a,m)M_p(b,n) = \frac{am+bn}{m+n} = (am+bn, m+n).$$
4. Si prendono in considerazione le medie dei voti riportati dai candidati promossi agli esami di Stato, appartenenti alle 5 classi dell'Istituto "Einstein". Dimostra che la media dei voti di tutti gli alunni delle 5 classi è uguale alla media aritmetica ponderata delle medie relative alle singole classi, prendendo come pesi i numeri dei candidati promossi delle classi corrispondenti.

12.3.4 Alcune osservazioni sulle medie:

- La *moda* coincide sempre con uno dei valori dell'insieme che si considera.
- La *mediana* vi coincide se gli elementi dell'insieme sono in numero dispari; può non coincidervi se essi sono in numero pari.
- La *media aritmetica* di solito non coincide con alcuno dei valori dell'insieme; potrebbe, comunque, anche coincidere con uno di essi.

In genere non è possibile fare un elenco esauriente dei casi in cui una media sia più conveniente di un'altra per avere un indice caratteristico di un dato collettivo. Bisogna valutare di volta in volta. Possiamo fare, tuttavia, alcune considerazioni di massima.

Possiamo dire, per esempio, che si ricorre alla media aritmetica quando si cerca la misura più attendibile di una data grandezza, come ad esempio l'altezza media di un gruppo di persone.

Qualche volta è più utile la moda. Per esempio, se si devono confezionare delle divise militari per una truppa, può non interessare la media aritmetica delle stature dei soldati e può essere invece più utile la moda di quelle stature, cioè la statura più frequente.

L'importanza della mediana risiede nel fatto che essa ripartisce i dati rilevati (eccetto, al più, uno di essi) in due insiemi con lo stesso numero di elementi. Un esempio, ancorché banale: i dati rilevati siano i punteggi, in centesimi, riportati dagli studenti che frequentano una certa classe. Conoscere il punteggio mediano può essere utile per capire immediatamente se la maggior parte degli studenti ha ottenuto un risultato positivo o no.

12.3.5 Descriviamo adesso un esempio, che tra l'altro servirà ad evidenziare come siano generalmente diverse le varie medie. In particolare la media aritmetica e la moda, che invece qualche volta vengono erroneamente identificate. E come la media aritmetica di n dati statistici sia in genere diversa dalla media aritmetica del minimo e del massimo dei dati considerati. Ragion per cui le due medie non vanno confuse.

Sono state misurate (con l'approssimazione di 1 cm) le altezze dei 24 alunni di una certa classe. I risultati sono riportati nella tabella 10, dove nella 1ª colonna è indicato il numero d'ordine degli alunni, come figura nel registro di classe, e nella 2ª colonna l'altezza dell'alunno corrispondente (in cm).

Tab. 10 – Distribuzione delle altezze degli studenti.

N°	Altezza		N°	Altezza		N°	Altezza		N°	Altezza
1	164		7	179		13	164		19	165
2	169		8	165		14	177		20	174
3	173		9	172		15	165		21	162
4	161		10	168		16	165		22	170
5	162		11	165		17	179		23	159
6	180		12	179		18	166		24	166

Con riferimento a questa tabella, si trovano i seguenti risultati, che sei invitato a controllare, servendoti, quando occorre, di un foglio elettronico:

- l'altezza mediana è 166 cm ; qual è la sua frequenza relativa?
- la moda è 165 cm ; qual è la sua frequenza relativa?
- la media aritmetica delle altezze è 168,7 cm; mentre la media aritmetica del minimo (159 cm) e del massimo (180 cm) è 169,5 cm .

I dati della tabella 10 possono essere rappresentati con un istogramma (Fig. 4), dopo aver raggruppato le altezze in classi, per esempio di ampiezza 4 cm (6 classi); ossia nelle seguenti classi:

159-162, 163-166, 167-170, 171-174, 175-178, 179-182,

ciascuna contenente rispettivamente i seguenti numeri di studenti:

4, 9, 3, 3, 1, 4.

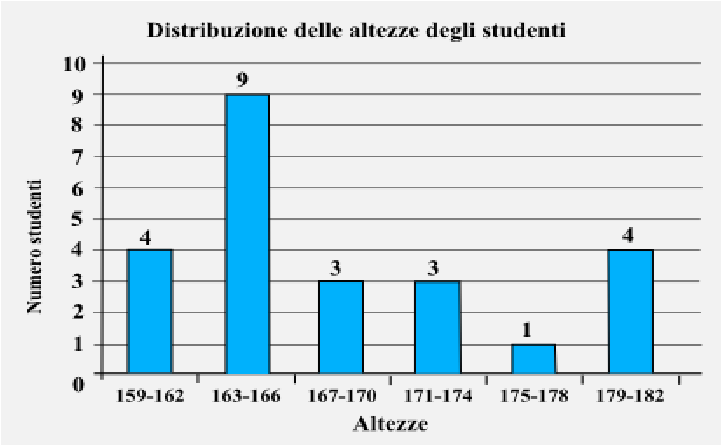


FIG. 4

Prova a rappresentare con un istogramma gli stessi dati dopo averli ripartiti in classi di ampiezza 2 cm (12 classi).

12.3.6 La mediana e la media aritmetica possono essere calcolate utilizzando direttamente la tabella 10, come hai potuto constatare personalmente, ma possono essere calcolate anche mediante la figura 4, che ne fornisce una rappresentazione grafica. In questo secondo caso, però, è prima necessario organizzare una tabella (Tab. 11) in cui, oltre alle singole frequenze, siano riportate passo dopo passo le frequenze che si ottengono sommando all'ultima frequenza considerata quelle precedenti e che si chiamano **frequenze cumulate**.

Tab. 11 – Distribuzione della frequenza delle altezze degli studenti ripartite per classi.

Classe di altezze	Frequenza	Frequenza cumulata	Frequenza cumulata percentuale	Valori centrali delle classi
1) 159–162	4	4	$4/24 = 16{,}7\ \%$	160,5
2) 163–166	9	4+9=13	$13/24 = 54{,}2\ \%$	164,5
3) 167–170	3	3+13=16	$16/24 = 66{,}7\ \%$	168,5
4) 171–174	3	3+16=19	$19/24 = 79{,}2\ \%$	172,5
5) 175–178	1	1+19=20	$20/24 = 83{,}3\ \%$	176,5
6) 179–182	4	4+20=24	$24/24 = 100\ \%$	180,5

Cerchiamo allora di determinare la mediana in base a questa tabella. Si constata al riguardo che la 1ª classe di altezze occupa il 16,7% di tutte le altezze, mentre la 1ª e la 2ª classe occupano il 54,2%. Se ne desume che la mediana, cioè quell'altezza che ripartisce in due parti uguali l'insieme delle altezze rilevate si trova nella 2ª classe. Possiamo assumerla uguale al valore centrale delle altezze di questa classe, vale a dire uguale a 164,5 cm. Non è il valore esatto della mediana, che sappiamo essere di 166 cm, ma l'errore relativo che si commette assumendo come mediana 164,5 cm, è $\frac{166-164{,}5}{166} \approx 0{,}9\%$; praticamente trascurabile (3).

Occupiamoci adesso della media aritmetica. Essa può essere assunta uguale alla media ponderata dei valori centrali delle classi considerate, presi con pesi uguali alle frequenze delle classi medesime. Vale a dire:

$$M=\frac{160{,}5\cdot 4+164{,}5\cdot 9+168{,}5\cdot 3+172{,}5\cdot 3+176{,}5\cdot 1+180{,}5\cdot 4}{24}=168{,}5\,.$$

Si tratta ancora di un valore diverso da quello effettivo, che sappiamo essere di 168,7 cm, ma si vede subito che la differenza è del tutto irrilevante.

Quanto vale l'errore relativo che si commette nell'assumere questa media al posto di quella effettiva?
Quanto vale la media aritmetica ponderata dei valori centrali delle classi, nel caso che queste siano 12?

- Ora, qualche esercizio per te, ma dopo alcuni preliminari dai quali non si può prescindere.

Le persone di una qualunque comunità possono essere ripartite in 8 categorie, in base al *gruppo sanguigno* e al cosiddetto *fattore Rhesus*. Esistono, infatti, 4 gruppi sanguigni (0, A, B, AB) e due fattori Rhesus (Rh^+ ed Rh^-). Le trasfusioni di sangue possono avvenire tra persone dello stesso gruppo sanguigno ma anche, a determinate condizioni, tra persone di gruppo sanguigno diverso. Le regole per la trasfusione sono sintetizzate nella seguente tabella (Tab. 12), dove abbiamo scritto 0^+ per indicare il "gruppo 0, Rh^+"; e ugualmente per gli altri gruppi.

[3] Quello che abbiamo proposto è un modo grossolano ma efficace di calcolare un'approssimazione della mediana per dati raggruppati in classi. In realtà sarebbe possibile un calcolo più preciso col ricorso ad un'apposita formula, di cui però preferiamo non occuparci.

Tab. 12 – Tabella di corrispondenza dei gruppi sanguigni.		
Gruppo sanguigno	**A quale gruppo può cedere il sangue**	**Da quale gruppo può ricevere il sangue**
0^+	Rh^+	0
0^-	Ogni gruppo	0^-
A^+	A^+, AB^+	0, A
A^-	A, AB	0^-, A^-
B^+	B^+, AB^+	0, B
B^-	B, AB	0^-, B^-
AB^+	AB^+	Ogni gruppo
AB^-	AB	Rh^-

Ed ecco l'esercizio che ti viene proposto, sulla base di quanto detto sopra.
La seguente tabella (Tab. 13) mostra la distribuzione delle frequenze dei gruppi sanguigni in una comunità di 2754 persone.

Tab. 13 – Distribuzione delle frequenze dei gruppi sanguigni.									
Gruppo sanguigno	0^+	0^-	A^+	A^-	B^+	B^-	AB^+	AB^-	Tutti
Numero persone	1102	248	854	193	220	55	56	26	2754

Calcola:
1) a quante persone della comunità possono cedere il sangue gli appartenenti al gruppo 0^+ e a quante quelle appartenenti al gruppo 0^-;
2) da quante persone della comunità possono ricevere il sangue gli appartenenti al gruppo 0^+ e da quante quelle appartenenti al gruppo 0^-.

Indica, inoltre, la moda del gruppo sanguigno delle persone della comunità e rappresenta con un istogramma la distribuzione in questione.
Esprimi, infine, in forma percentuale la distribuzione dei gruppi sanguigni della comunità.

L'istogramma sottostante (Fig. 5) mostra invece la distribuzione delle percentuali di gruppi sanguigni di 8.400 persone.

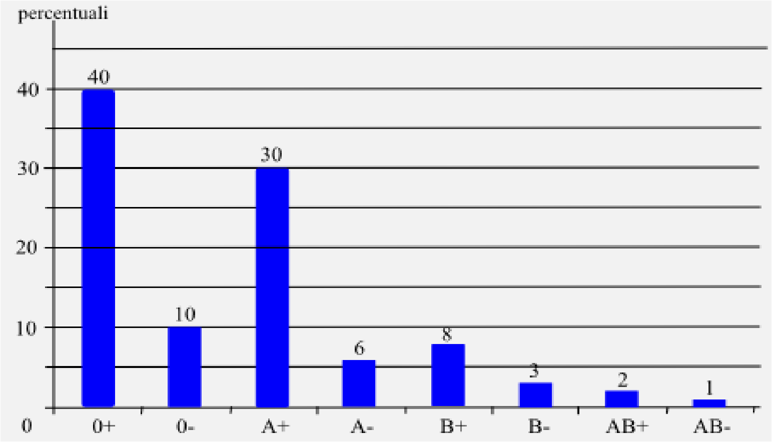


FIG. 5

Calcola quante persone della comunità possono:
a) ricevere il sangue da ogni appartenente al gruppo A^+;
b) cedere il sangue ad ogni appartenente al gruppo B^-.

12.4 INDICI DI VARIABILITÀ

12.4.1 Conoscere un indice di posizione, come per esempio la media aritmetica, è indubbiamente un fatto importante, poiché così è possibile avere informazioni sulle caratteristiche del collettivo che si considera per mezzo di un solo numero.

La sola conoscenza della media aritmetica, tuttavia, non è sufficiente a caratterizzare la dispersione, rispetto ad essa, dei dati statistici considerati. Infatti, due successioni aventi la stessa media aritmetica sono completamente diverse se in una i termini si scostano molto da quella media e nell'altra, invece, si scostano poco.

È nota a tutti la celebre poesia di Trilussa (poeta romanesco, 1871-1950) nella quale si fa riflettere sul fatto che la media aritmetica dei polli mangiati da due persone è sempre "1 pollo" sia se le due persone mangiano un pollo ciascuno sia se una ne mangia due e l'altra non ne mangia affatto.

Ma vediamo un altro esempio. Consideriamo i due seguenti insiemi di dati, i cui elementi sono gli stipendi mensili (in Euro), ciascuno di 5 lavoratori:

- 1° insieme di dati:

$$800\,,\ \ 1100\,,\ \ 1.600\,,\ \ 1.900\,,\ \ 2.600\,;$$

- 2° insieme di dati:

$$1.500\,,\ \ 1.520\,,\ \ 1.620\,,\ \ 1.650\,,\ \ 1.710\,.$$

In entrambi la media aritmetica è 1600 euro. Però, mentre nel primo gli scostamenti dei vari termini da questa media sono molto diversi gli uni dagli altri, nella seconda essi sono più contenuti, come è mostrato qui appresso (il segno "– " indica che il corrispondente valore è minore della media aritmetica, il segno "+" che è maggiore):

- scostamenti dei termini del 1° insieme di dati dalla media aritmetica:

$$-800\,,\ \ -500\,,\ \ 0\,,\ \ +300\,,\ \ +1.000\,;$$

- scostamenti dei termini del 2° insieme di dati dalla media aritmetica:

$$-100\,,\ \ -80\,,\ \ +20\,,\ \ +50\,,\ \ +110\,.$$

Affermare pertanto che la media aritmetica degli stipendi dei lavoratori è di € 1600 è vero in entrambi i casi. Solo che, riferita al primo di essi, l'affermazione sembra essere una ... grossa bugia.

12.4.2 Assieme alla media aritmetica dei dati statistici considerati – che da sola non permette di valutare come i dati stessi si distribuiscano intorno ad essa – è necessario disporre di valori che per l'appunto diano indicazioni su questa distribuzione. Questi valori si chiamano **indici di variabilità** (o **di dispersione**). Ne prendiamo in esame tre: la *dispersione massima*, lo *scarto semplice medio* e lo *scarto quadratico medio*.

La ***dispersione massima*** è la differenza:

$$\boldsymbol{\delta = \max(A) - \min(A)}$$

tra il massimo ed il minimo termine della successione. È chiamata anche ***campo di variazione***.

Nel caso dei polli essa è 0 nella prima situazione (1 pollo a testa) e 2 nella seconda.

Invece nell'altro esempio, essa è: $\delta' = 2.600 - 800 = 1.800$ (€) , nella prima situazione.

Ed è: $\delta'' = 1.710 - 1.500 = 210$ (€) , nella seconda.

Questo indice permette di valutare subito il divario massimo fra i termini della successione. Però, al fine di avere informazioni circa il modo con cui i termini si distribuiscono intorno alla media si assume più spesso la sua metà, che è chiamata ***semidispersione massima***.

12.4.3 Esistono indici di dispersione più affidabili della dispersione massima. Ce ne occupiamo in questo paragrafo. Riprendiamo la solita variabile statistica A, i cui valori sono forniti dalla seguente successione:

[1] $$a_1,\ a_2, \dots ,a_n.$$

Eseguendo la differenza tra ciascuno dei termini della [1] e la loro media aritmetica M, si ottiene quest'altra successione:

[2] $$a_1-M,\ a_2-M, \dots ,a_n-M.$$

I termini di questa seconda successione si chiamano ***scarti*** (o ***scostamenti***) dei dati della successione [1] dalla loro media aritmetica M.

Per esempio:

- con riferimento alla prima delle due successioni considerate sopra, riguardanti gli stipendi percepiti, gli scarti dei suoi termini dallo loro media aritmetica (€ 1600) sono i valori (espressi ovviamente in euro) già calcolati prima, vale a dire i seguenti valori:
$$-800\,,\ -500\,,\ 0\,,\ +300\,,\ +1.000\,;$$
- con riferimento alla seconda successione, essi sono:
$$-100\,,\ -80\,,\ +20\,,\ +50\,,\ +110\,.$$

Ora, come per un insieme di dati numerici è comodo un solo numero che ne descriva la tendenza (per esempio: la media aritmetica dei dati considerati), allo stesso modo è utile un solo numero che dia informazioni sugli scarti dei dati.

Quale idea ti viene in mente per ottenere un solo dato numerico che descriva la tendenza degli scarti?
Hai forse pensato alla media aritmetica degli scarti?

Proviamo con i due esempi considerati prima. Siccome per il calcolo della media aritmetica degli scarti occorre la loro somma, incominciamo a calcolare questa somma:

- con riferimento alla prima successione, essa è:
$$S' = -800 - 500 + 0 + 300 + 1.000 = 0\,;$$
- con riferimento alla seconda, la somma è:
$$S'' = -100 - 80 + 20 + 50 + 110 = 0\,.$$

In entrambi i casi la somma degli scarti è 0.

Ci viene il sospetto che la somma degli scarti dalla media aritmetica sia sempre 0. Controlliamo con riferimento agli scarti [2] di una generica variabile statistica. Si ha:
$$S = (a_1-M)+(a_2-M)+\dots+(a_n-M) = (a_1+a_2+\dots+a_n)-nM;$$
ma dalla formula che fornisce la media aritmetica M si ricava:
$$a_1+a_2+\dots+a_n = nM \text{ e pertanto: } S = nM-nM = 0.$$
Effettivamente: **La somma degli scarti dalla media aritmetica dei dati statistici cui essa si riferisce è nulla.** Ciò è chiaramente causato dalla presenza, negli scarti [2], di termini positivi e di termini negativi che si compensano, in quanto la somma degli uni ha lo stesso valore assoluto di quella degli altri.

Bisogna precisare che **questo è vero solo se gli scarti sono calcolati rispetto alla media aritmetica**.
Per ogni altra media, infatti, la somma degli scarti rispetto ad essa è, in genere, diversa da zero, anche se non obbligatoriamente.

Per esempio, con riferimento alla mediana, nel caso dei due esempi precedenti, si può calcolare che la somma degli scarti è zero nel primo caso (ma solo perché la mediana, che è 1.600, coincide con la media aritmetica), mentre è diversa da zero nel secondo caso (in cui la mediana è 1.620).

12.4.4 Il fatto che la somma degli scarti dalla media aritmetica dei dati rilevati sia nulla ci dice chiaramente che nessuna informazione, circa la dispersione dei dati [1] rispetto alla loro media M, ci può pervenire dal prendere in esame la media aritmetica degli scarti medesimi.

Per ottenere informazioni sulla dispersione dei dati [1], bisogna allora prendere degli indici che non risentano della presenza di scarti negativi.

Ad essi si può arrivare in due modi: o prendendo i valori assoluti degli scarti o prendendo i loro quadrati.

Più esattamente:

- Una misura della dispersione dei dati [1] dalla loro media aritmetica M è la media aritmetica dei valori assoluti degli scarti [2]; vale a dire il numero S tale che:
$$\mathbf{S}=\frac{|\mathbf{a_1}-\mathbf{M}|+|\mathbf{a_2}-\mathbf{M}|+\cdots+|\mathbf{a_n}-\mathbf{M}|}{\mathbf{n}}.$$
Si chiama ***scarto semplice medio***.
- Un'altra misura della dispersione dei dati [1] dalla loro media aritmetica M è la media aritmetica dei quadrati degli scarti [2]; vale a dire il numero σ^2 tale che:
$$\boldsymbol{\sigma}^{\mathbf{2}}=\frac{(\mathbf{a_1}-\mathbf{M})^{\mathbf{2}}+(\mathbf{a_2}-\mathbf{M})^{\mathbf{2}}+\cdots+(\mathbf{a_n}-\mathbf{M})^{\mathbf{2}}}{\mathbf{n}}.$$
Si chiama ***varianza***. Se X è la variabile statistica in gioco, è indicata anche con $\mathbf{var}(\mathbf{X})$.

Per esempio, sempre con riferimento alle due successioni considerate sopra:

- in merito alla prima successione si calcola: S=520 (€), σ^2=3,96·10^8 (€^{2});
- in merito alla seconda si calcola: S=72 (€), σ^2=6,28·10^6 (€^{2}).

Si capisce che, se i dati numerici assegnati sono espressi in una certa unità di misura ("euro" negli esempi considerati, ma potrebbe trattarsi di "metri", "grammi", eccetera), la varianza σ^2 è espressa nel quadrato di quella unità di misura (€^{2} nell'esempio; ma anche m^2, g^2, eccetera). Insomma σ^2 non è omogenea con i dati. Per questo si preferisce ricorrere alla sua radice quadrata, perché per l'appunto è un valore omogeneo con i dati. Si ottiene, in questo modo, una nuova misura della dispersione dei dati [1] dalla loro media aritmetica, la seguente:
$$\boldsymbol{\sigma}=\sqrt{\frac{(\mathbf{a_1}-\mathbf{M})^{\mathbf{2}}+(\mathbf{a_2}-\mathbf{M})^{\mathbf{2}}+\cdots+(\mathbf{a_n}-\mathbf{M})^{\mathbf{2}}}{\mathbf{n}}}.$$
Si chiama ***scarto quadratico medio*** (o ***deviazione standard***). A volte, se X è la variabile statistica in gioco, è indicata anche con $\mathbf{dev}(\mathbf{X})$.

Con riferimento ai due soliti esempi si ha: nel 1° caso: $\sigma \approx 629$ (€); nel 2° caso: $\sigma \approx 79$ (€).

Dunque possiamo concludere che:

- la media aritmetica degli stipendi dei lavoratori considerati nel primo dei due esempi più volte presi in esame, è € 1.600 con uno scarto quadratico medio di circa € 629 (o con uno scarto semplice medio di € 520);
- la media aritmetica degli stipendi dei lavoratori considerati nel 2° esempio, è € 1.600 con uno scarto

quadratico medio di circa € 79 (o con uno scarto semplice medio di € 72).

Ognuno dei due valori S, σ indica di quanto, in media, i dati statistici [1] si discostano dalla loro media aritmetica M. Di solito si preferisce ricorrere al primo valore quando i dati [1] non presentano scostamenti rilevanti rispetto alla media aritmetica, come nel secondo dei due esempi più volte considerati. Altrimenti si ricorre al secondo valore, come nel primo dei due esempi.
La deviazione standard, ad ogni modo, è nella pratica la misura più usata della dispersione dei dati rispetto alla loro media aritmetica.

NOTA BENE. Per determinare gli indici di posizione e di dispersione di una serie di dati statistici e per disegnarne l'eventuale rappresentazione grafica è opportuno servirsi di uno strumento di calcolo automatico, preferibilmente un *foglio elettronico*.

ESERCIZIO. Nella tabella sottostante (Tab. 14) sono riportati i valori delle quote di un particolare "Fondo" relativi agli ultimi 12 rilevamenti. a) Calcolarne la media aritmetica e la deviazione standard. b) Fornire anche una rappresentazione cartesiana dell'andamento dei valori presi in esame.

Tab. 14 – Distribuzione delle quote di un particolare fondo

Rilevamento N°	1	2	3	4	5	6	7	8	9	10	11	12
Valore della quota	6,046	6,554	6,666	6,624	6,542	6,588	6,545	6,512	6,596	6,627	6,577	6,629

12.4.5 Vale un'importante proprietà della media aritmetica e della deviazione standard.

- PROPRIETÀ. Se gli N valori $a_1,\ a_2, \dots, a_N$ di una variabile statistica aumentano o diminuiscono tutti della stessa quantità X, la media aritmetica M aumenta o diminuisce rispettivamente della medesima quantità X, mentre la deviazione standard rimane inalterata.

DIMOSTRAZIONE. Indichiamo con M* la nuova media aritmetica. Si ha:

$$M^*=\frac{(a_1\pm X)+(a_1\pm X)+\dots+(a_1\pm X)}{N}=\frac{(a_1+a_2+\dots+a_N)\pm NX}{N}=\frac{NM\pm NX}{N}=M\pm X\,.$$

Indichiamo con σ* la nuova deviazione standard. Si ha:

$$\sigma^*=\sqrt{\frac{[(a_1\pm X)-(M\pm X)]^2+[(a_1\pm X)-(M\pm X)]^2+\dots+[(a_1\pm X)-(M\pm X)]^2}{N}}=$$

$$=\sqrt{\frac{(a_1-M)^2+(a_1-M)^2+\dots+(a_1-M)^2}{N}}=\sigma\,.$$

12.4.6 Quando i dati sono organizzati per frequenze cumulate, il calcolo degli indici di variabilità può essere fatto ricorrendo proprio alla tabella delle frequenze cumulate.

Forniamo un esempio, calcolando la deviazione standard nel caso dei dati della tabella 10, relativa alle altezze degli studenti.
Ricordando che la media aritmetica, calcolata con riferimento alla medesima tabella delle frequenze cumulate, è 164,5 cm, calcoliamo dapprima la varianza. Si ha:

$$\sigma^2=\frac{(1605-1645)^2\cdot 4+(1645-1645)^2\cdot 9+(1685-1645)^2\cdot 3+(1725-1645)^2\cdot 3+(1765-1645)^2\cdot 1+(1805-1645)^2\cdot 4}{24}=$$

$$=18{,}6666,$$

ovviamente misurata in centimetri quadrati; perciò: $\sigma=\sqrt{18{,}666}\approx 4{,}3$ (cm).
Si tratta di un'approssimazione molto buona della deviazione standard calcolata esattamente, cioè utilizzando i singoli dati.

Prova per esercizio a calcolare questa deviazione standard (vedi la tabella 10 o, se ti fa più comodo, la figura 4) e l'errore relativo che si commette assumendo come deviazione standard quella trovata sopra.

12.4.7 Se distribuiamo in classi le rilevazioni di una variabile statistica, mettendo in ogni classe i valori compresi in un certo intervallo, convenientemente scelto, possiamo costruire un istogramma (Fig. 6) che indichi la variazione dei dati compresi in un intervallo al variare dell'intervallo stesso.

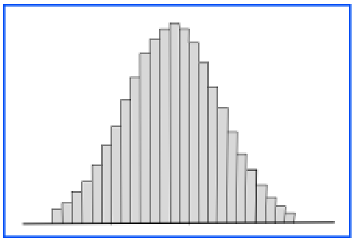


FIG. 6

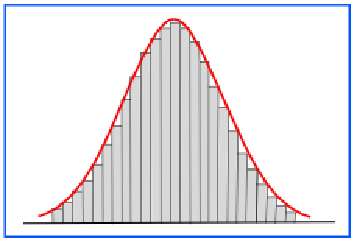


FIG. 7

Se i dati raccolti sono in numero elevato e gli intervalli entro cui vengono collocati sono di ampiezza sufficientemente piccola, l'istogramma che si ottiene inviluppa una certa linea, che in numerosi casi ha la forma di figura 7 ed è conosciuta come ***curva di Gauss*** [4] (o ***gaussiana*** o ***curva a campana***) [5].
Tanto per fare qualche esempio, frequenze che si distribuiscono secondo la curva di Gauss sono: le altezze delle persone, i loro pesi, le misure delle grandezze sperimentali. A condizione che il numero dei rilevamenti sia adeguato (almeno 30 secondo alcuni studiosi, almeno 50 secondo altri).
La curva di Gauss presenta un asse di simmetria in corrispondenza della media aritmetica dei dati. Dall'esame di essa si deduce perciò che i dati si addensano intorno alla loro media aritmetica.

Ti proponiamo di risolvere il seguente esercizio.
Disegna un istogramma che indichi come vari, in un lancio di due dadi, il numero dei casi favorevoli all'uscita di una data somma al variare della somma medesima. Considerate poi, come dati statistici, le suddette somme di punti, calcola:
- la mediana, la moda e la media aritmetica M;
- lo scarto quadratico medio σ.

A conti fatti, dovresti trovare i seguenti risultati, considerati nell'ordine delle domande:

$$7,\ \ 7,\ \ M=7;\ \ \sigma\approx 2{,}4.$$

La curva di Gauss è detta pure **curva degli errori**. Ciò perché, se i dati raccolti sono le misurazioni $x_1, x_2, \dots, x_n$ di una data grandezza, ottenute con uno strumento di misura con un elevato grado di precisione, esse, a causa degli inevitabili *errori accidentali*, non sono uguali fra loro. Questi errori non sono altro che gli scarti $y_1, y_2, \dots, y_n$ delle diverse misure dalla loro media aritmetica. Ebbene, rappresentando questi errori y_i sull'asse delle ascisse di un piano cartesiano ortogonale e i loro numeri N_i sull'asse delle ordinate, detti errori si distribuiscono *mediamente* secondo una curva di Gauss (Fig. 8).

---

[4] **Gauss**, Karl Friedrich, matematico tedesco, 1777-1855.
[5] In realtà, il primo a rappresentare la curva a campana e ad utilizzarla fu il matematico francese Abraham de Moivre (1667-1754) in un articolo del 1733, vale a dire ancor prima che Gauss nascesse.

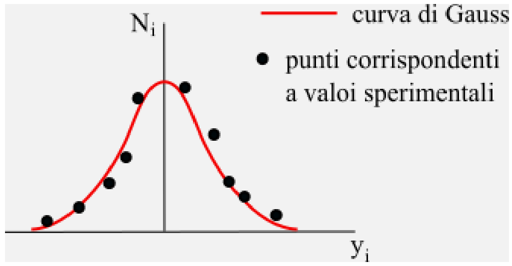


FIG. 8

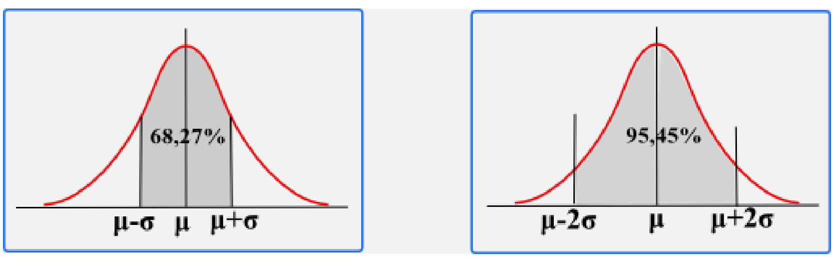
Bisogna dire che anche le misure ottenute, raggruppate convenientemente in classi, si dispongono secondo una gaussiana [6].

In particolare, se sono rispettivamente μ e σ la media aritmetica e la deviazione standard delle misurazioni rilevate, la teoria mostra che di tali misurazioni ne cade:

- il 68,27% nell'intervallo [μ–σ, μ+σ] (Fig. 9a);
- il 95,45% nell'intervallo [μ–2σ, μ+2σ] (Fig. 9b);
- il 99,73% nell'intervallo [μ–3σ, μ+3σ].

In realtà, queste percentuali si presentano ogni volta che le frequenze di un determinato carattere si distribuiscono secondo la curva di Gauss.

Avremo modo di ritornare su questi concetti e di approfondirli nel seguito degli studi [7].



(a) (b)

FIG. 9

Ma adesso ti proponiamo un esercizio.

Sono state misurate le altezze di 400 persone e si sono trovati i seguenti valori di sintesi:

media aritmetica = 173,5 cm, deviazione standard = 2,8 cm.

Quante persone ritieni che abbiano un'altezza compresa fra (173,5–2,8) cm e (173,5+2,8) cm?

Quante fra (173,5–2×2,8) cm e (173,5+2×2,8) cm?

Quante fra (173,5–3×2,8) cm e (173,5+3×2,8) cm?

Quante persone, infine, pensi che rimangano fuori dall'intervallo compreso fra (173,5–3×2,8) cm e (173,5+3×2,8) cm?

[Ricorda che le frequenze delle altezze delle persone si distribuiscono … come?]

12.5 MEDIA GEOMETRICA E MEDIA ARMONICA

12.5.1 La media aritmetica di n numeri, la mediana e la moda non sono le sole medie interessanti. Altre ce ne sono ed in particolare la *media geometrica* e la *media armonica*. Ne vogliamo fare un breve cenno.

---

[6] Si suggerisce la lettura posta in chiusura di questa unità.

[7] Cfr.: Unità 80: Distribuzione normale. Distribuzione di Poisson.

Considerati i due numeri a, b, essi possono essere interpretati come le misure (espresse rispetto alla medesima unità di misura) di due segmenti adiacenti AB e BC (Fig. 10).

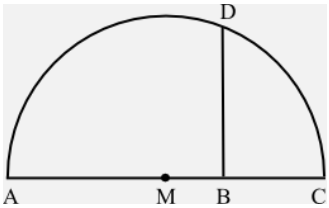


FIG. 10

Se assumiamo AC come il diametro di un semicerchio e tracciamo la corda BD perpendicolare ad AC, accade che la misura h di questa corda, in virtù del 2° teorema di Euclide applicato al triangolo rettangolo ADC, è $\sqrt{ab}$. Essa è una media dei due numeri a, b, dal momento che risulta: $a \leq \sqrt{ab} \leq b$. Infatti:

$$a \leq b \rightarrow aa \leq ab \leq bb \rightarrow a^2 \leq ab \leq b^2 \rightarrow a \leq \sqrt{ab} \leq b.$$

Si chiama *media geometrica*.

Si dice anche che il segmento BD è *medio proporzionale* fra i segmenti AB e BC.

In geometria si trovano molti esempi di grandezza media proporzionale fra due altre.

Proprio con riferimento ai due teoremi di Euclide si hanno due di tali esempi. Altri ne incontrerai nel seguito degli studi.

In generale, dati n numeri reali (positivi) $a_1, a_2, \dots, a_n$, si definisce loro **media geometrica** il numero m_g tale che:

$$\mathbf{m_g = \sqrt[n]{a_1 \cdot a_2 \cdot \ldots \cdot a_n}}\,.$$

12.5.2 Prendiamo ora in esame il seguente problema.

PROBLEMA. Un corpo si muove dalla posizione A alla posizione B, in linea retta, alla velocità media v_1 e percorre il tratto inverso da B ad A alla velocità media v_2. Qual è la velocità media sull'intero percorso di andata e ritorno?

RISOLUZIONE. Sono molti coloro che concludono che la velocità media è la media aritmetica delle due velocità. Non è così e andiamo a spiegarlo.

Chiamiamo per comodità s la distanza AB, t_1 il tempo di percorrenza di AB all'andata e t_2 quello del ritorno. Si ha evidentemente:

$$v_1 = \frac{s}{t_1}\,,\quad v_2 = \frac{s}{t_2}\,.$$

La velocità media sull'intero percorso è allora: $v = \dfrac{2s}{t_1 + t_2}$.

Sembra, ad un'analisi superficiale, che non si possa calcolare v senza conoscere s, t_1, t_2. Ma basta un minimo di riflessione per capire che non è così. Si ha infatti:

$$v = \frac{2s}{t_1 + t_2} = \frac{2}{\frac{t_1}{s} + \frac{t_2}{s}} = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}}\,.$$

Questo valore di v è una media di v_1 e v_2. Ma, di nuovo, questo va provato. In altri termini, posto che sia $v_1 \leq v_2$, deve essere dimostrato che:

$$v_1 \leq \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} \leq v_2.$$

Questa relazione può essere scritta più convenientemente in questo modo:

$$v_1 \le \frac{2v_1v_2}{v_1+v_2} \le v_2$$

e sia ha in successione:

$v_1 \le v_2 \ \rightarrow \ v_1v_1 \le v_1v_2 \le v_2v_2 \ \rightarrow \ v_1v_1+v_1v_2 \le v_1v_2+v_1v_2 \le v_2v_2+v_1v_2 \ \rightarrow$

$$\rightarrow \ v_1(v_1+v_2) \le 2v_1v_2 \le v_2(v_1+v_2) \ \rightarrow \ v_1 \le \frac{2v_1v_2}{v_1+v_2} \le v_2.$$

Il numero: $v=\dfrac{2}{\dfrac{1}{v_1}+\dfrac{1}{v_2}}$ è dunque una nuova media di v_1 e v_2: si chiama *media armonica*.

In generale, dati n numeri reali (positivi) $a_1, a_2, \dots, a_n$, si definisce loro **media armonica** il numero m_{ar} tale che:

$$\mathbf{m_{ar}} = \frac{\mathbf{n}}{\dfrac{\mathbf{1}}{\mathbf{a_1}}+\dfrac{\mathbf{1}}{\mathbf{a_2}}+\dots+\dfrac{\mathbf{1}}{\mathbf{a_n}}}.$$

Si constata agevolmente che ***la media armonica di* n *numeri (positivi) è tale che il suo valore inverso è la media aritmetica dei valori inversi degli* n *numeri***.

12.5.3 Abbiamo così fatto la conoscenza di tre medie interessanti: la media aritmetica, la media geometrica e la media armonica.

Per esempio, dati i due numeri 2 e 8, si calcola facilmente che si ha: $m_a=5,\ m_g=4,\ m_{ar}=3,2$.

E si constata subito che risulta: $m_a>m_g>m_{ar}$. È un caso o questa relazione è generale?

Ebbene sì, è proprio una relazione generale e vale non solo per due numeri assegnati (positivi e diversi fra loro) ma per un numero n qualsiasi di numeri (positivi e non tutti uguali fra loro).

L'affermazione può ovviamente essere dimostrata, ma noi proponiamo una dimostrazione solo nel caso di 2 numeri. In verità la spiegazione della relazione $m_a>m_g$, valida se i due numeri assegnati sono disuguali, è veramente intuitiva: basta riferirsi all'interpretazione geometrica delle due medie. Noi vogliamo comunque fornire una spiegazione formale sia per questa relazione sia per quella che coinvolge la media armonica. Siano allora assegnati due numeri positivi a,b, con a≠b. Si ha in successione:

$(a-b)^2>0 \ \rightarrow \ (a-b)^2+4ab>4ab \ \rightarrow \ (a+b)^2>4ab \ \rightarrow$

$$\rightarrow \left[\begin{array}{l} a+b>2\sqrt{ab} \ \rightarrow \ \dfrac{a+b}{2}>\sqrt{ab} \ \rightarrow \ m_a>m_g \\ 1>\dfrac{4ab}{(a+b)^2} \ \rightarrow \ ab>\dfrac{4(ab)^2}{(a+b)^2} \ \rightarrow \ \sqrt{ab}>\dfrac{2ab}{a+b} \ \rightarrow \ m_g>m_{ar} \end{array}\right.$$

In definitiva: $m_a>m_g>m_{ar}$. [c.v.d.]

Chiaramente, se a=b, risulta: $m_a=m_g=m_{ar}$.

Il fatto che, presi due qualsiasi numeri reali positivi, la loro media geometrica non supera mai la loro media aritmetica permette di fornire una rapida spiegazione della seguente proprietà geometrica (compito che deleghiamo a te): L'area di un rettangolo non supera quella di un quadrato il cui lato ha una lunghezza uguale alla media aritmetica delle dimensioni del rettangolo, quali che siano queste dimensioni. In quale caso le due aree sono uguali?

VERIFICHE

1. Quella volta l'indice borsistico ebbe l'andamento sintetizzato dalla tabella sottostante, dove sono riportati gli indici di un'annata borsistica, poco fortunata, rilevati ogni 15 giorni:

681	726	732	691	677	701	715	743	767	745	700	688	719
677	699	657	595	624	640	665	588	476	523	488	493	

Fornisci una rappresentazione cartesiana dell'andamento del fenomeno. Calcola inoltre quanto l'indice ha perso, in percentuale, nel corso dell'annata. [**R.** … ; 27,6 %]

2. Le percentuali di studenti di scuola superiore promossi a punteggio pieno o con debiti formativi rispetto agli studenti scrutinati o esaminati in quell'anno scolastico, sono distribuite nella tabella sottostante in base all'area geografica di appartenenza:

Area geografica	**Nord**	**Centro**	**Sud e isole**	**Totale**
Promossi a punteggio pieno	50,48	54,12	52,76	52,07
Promossi con debito formativo	34,76	32,31	30,75	32,63
Totali	**85,24**	**86,43**	**83,51**	**84,70**

Fornisci una rappresentazione grafica della situazione.

3. Costruisci una tabella statistica delle altezze degli studenti della tua classe, distinguendo tra maschi e femmine. Quindi calcola la moda e la media aritmetica delle altezze dei maschi, quelle delle femmine e quelle di tutti gli studenti della classe.
La media aritmetica delle altezze di tutti gli studenti è uguale alla media aritmetica delle altre due medie, quella dei maschi e quella delle femmine?

5. Con riferimento alla medesima distribuzione trovata nel precedente esercizio disegna un istogramma che rappresenti il numero degli alunni (indipendentemente dal sesso) che hanno una data altezza, raggruppando però queste altezze in classi ciascuna di ampiezza 2 cm.
Organizza una tabella in cui siano riportate le frequenze cumulate e calcola la mediana e la moda delle altezze rilevate riferendoti però alla tabella delle frequenze cumulate. Calcola infine l'errore relativo che si commette assumendo queste come valori effettivi della mediana e della media aritmetica.

5. LABORATORIO DI MATEMATICA. Conduci – assieme ai tuoi compagni di classe – un'indagine, nella scuola da te frequentata, orientata a determinare le percentuali degli alunni che, nei 5 anni precedenti l'anno in corso, sono stati respinti. Quindi compila una tabella a doppia entrata con le indicazioni di dette percentuali per anno e per classi.

6. Uno studente, interrogato 5 volte, riporta i seguenti voti: 7 - 6 - 6 - 8 - 7 .
Qual è la media aritmetica dei voti riportati?
Si può affermare che, effettuando la media aritmetica dei primi tre voti e quella degli ultimi due, la media aritmetica di queste due medie coincide con la media aritmetica dei 5 voti?

7. Una persona ha realizzato, in 30 giornate di lavoro, i seguenti guadagni:

€ 140 al giorno per 8 giorni; € 110 al giorno per 10 giorni;
€ 95 al giorno per 9 giorni; nessun guadagno per 3 giorni.

Quale delle seguenti espressioni numeriche, il cui valore è calcolato in euro, rappresenta il guadagno medio giornaliero?

[A] $\dfrac{140+110+95+0}{4}$. [B] $\dfrac{140+110+95}{3}$.

[C] $\dfrac{140\times8+110\times10+95\times9}{4}$. [D] $\dfrac{140\times8+110\times10+95\times9}{3}$.

Una sola alternativa è corretta. Individuarla e motivare la scelta effettuata.

8. La signora Giulia ha comprato ciliegie in tre negozi diversi: nel primo ne ha comprate 6 hg pagando € 3,75; nel secondo ne ha comprate 5 hg spendendo € 2,95; nel terzo ne ha comprate 4 hg pagando € 3,20. Quanto ha pagato mediamente per 1 kg di ciliegie la signora Giulia? [**R.** 6,60 €/kg]

9. Durante 30 giorni di scuola sono state registrate in una classe di 24 alunni le seguenti assenze:

2 alunni assenti per 7 giorni; 4 alunni assenti per 3 giorni; 3 alunni assenti per 2 giorni;
1 alunno assente per 12 giorni; 0 alunni assenti per 6 giorni.

Calcola la percentuale media giornaliera di assenze. [**R**. 6,1 %]

10. Un'azienda decide di dare una gratifica di € 900 alla metà dei suoi dipendenti, di € 1200 ad 1/4 di essi, di € 1500 ad 1/5 e di € 2000 alla parte rimanente. Quant'è la gratifica media pagata dall'azienda? [**R.** € 1150]

11. Lancia per 200 volte 2 dadi con le facce numerate da 1 a 6 (l'esperimento – forzando un po' le cose – può essere eseguito ripartendo i compiti fra un gruppo di studenti, per esempio 20, per cui ciascuno effettua 10 lanci) e, dopo aver raccolto i dati relativi al numero delle volte in cui la somma dei numeri usciti ha un dato valore, compila la seguente tabella:

Somma dei punti	2	3	4	5	6	7	8	9	10	11	12
Frequenze assolute											

Disegna quindi un istogramma che rappresenti come varia la frequenza assoluta al variare della somma ottenuta. Infine calcola la mediana e la moda delle somme ottenute.

12. Delle 200 persone presenti in una sala, 80 hanno 18 anni; 70 ne hanno 17; 40 hanno 16 anni; 5 hanno 32 anni ed altri 5 ne hanno 40. Rappresenta la situazione con un istogramma. Calcola quindi:

a) la mediana e la moda delle età delle 200 persone; b) la media aritmetica di tali età;
c) la semidispersione massima; d) lo scarto semplice medio e lo scarto quadratico medio delle età.

[**R.** a) 17; 18. b) 18,15. c) 12. d) $s \approx 1{,}785$; $\sigma \approx 4{,}25$]

13. Nella sottostante figura 11 è disegnato un istogramma di dati rilevati e raggruppati per classi. Segnare sul grafico le posizioni della mediana, della moda e della media aritmetica dei dati.

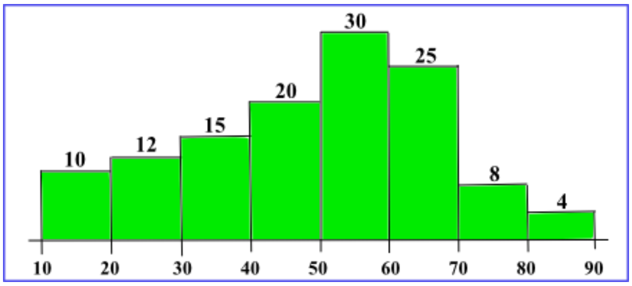


FIG. 11

14. Nella sottostante figura 12 è disegnato un istogramma di dati rilevati e raggruppati per classi. Segnare sul grafico le posizioni della mediana, della moda e della media aritmetica dei dati.

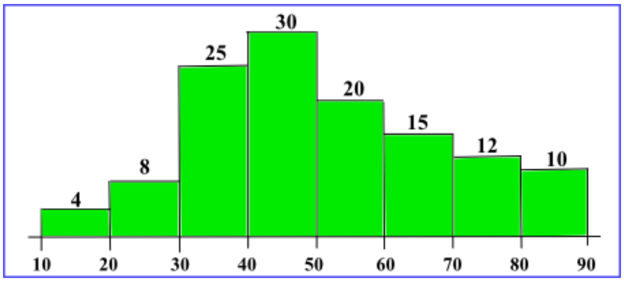


FIG. 12

15. Un corpo effettua un determinato percorso che può essere ripartito in n tratti: il 1° nel tempo t_1 alla velocità media v_1, il 2° nel tempo t_2 alla velocità media v_2, e così via l'n-esimo nel tempo t_n alla velocità media v_n. Dimostrare che la velocità media sull'intero percorso è uguale alla media aritmetica ponderata delle n velocità quando si prendono come pesi i corrispettivi tempi di percorrenza. In quale caso la velocità media sull'intero percorso è uguale alla media aritmetica (semplice) delle n velocità?

16. Nell'arco delle 24 ore di una giornata invernale sono state rilevate da un osservatorio meteorologico le temperature (espresse in °C) sintetizzate nella tabella sottostante.

ora	temperatura	ora	temperatura	ora	temperatura	ora	temperatura	ora	temperatura
0	−3,2	5	−4,2	10	+0,5	15	+6,5	20	+1,2
1	−3,5	6	−4,0	11	+2,1	16	+6,2	21	+0,3
2	−3,7	7	−3,8	12	+4,6	17	+5,7	22	−1,4
3	−3,8	8	−3,2	13	+6,4	18	+5,1	23	−2,1
4	−4,2	9	−1,0	14	+6,6	19	+3,4	24	−2,8

Fornisci una rappresentazione cartesiana dell'andamento della temperatura in funzione del tempo nell'arco delle 24 ore di quel giorno. Inoltre calcola:

a) la variazione massima delle temperature rilevate; b) la loro media aritmetica;

c) gli scarti delle temperature dalla loro media aritmetica;

d) lo scarto semplice medio e lo scarto quadratico medio.

Infine, sullo stesso grafico precedentemente disegnato, disegna la linea che esprime l'andamento della temperatura se questa, nelle 24 ore, si fosse mantenuta costantemente uguale alla temperatura media.
Cosa succede alla media aritmetica ed allo scarto quadratico medio se ciascuno dei dati rilevati aumenta di 1,5°C? [**R.** ... ; b) 0,308; ... ; d) $s \approx 3{,}276$; $\sigma \approx 3{,}751$; ...]

17. LABORATORIO DI MATEMATICA. Prova a rilevare la temperatura di un determinato luogo (magari la stanza dove studi) a intervalli regolari per un certo tempo (per esempio: ogni 15 minuti per 6 ore). Disegna il grafico che rappresenta la variazione della temperatura in funzione del tempo e quindi calcola: a) la temperatura massima e la temperatura minima; b) la temperatura media nell'arco delle 6 ore.
Sullo stesso grafico disegna la linea che esprime l'andamento della temperatura se essa, nel periodo di tempo considerato, si fosse mantenuta costantemente uguale alla temperatura media.
In quali ore dell'intervallo monitorato si ha una temperatura uguale alla media suddetta?
In quali ore la temperatura si scosta maggiormente dal valore medio?

18. In un esperimento, avente per obiettivo la misura di un lato di un bancone, sono stati ottenuti i valori (espressi in cm) elencati nella tabella sottostante. Calcola: a) la media aritmetica delle misure; b) lo scarto semplice medio; c) lo scarto quadratico medio. Disegna anche un istogramma della distribuzione di frequenze e calcola la moda e la mediana delle misure rilevate e le loro frequenze relative.

270,2	270,2	270,4	270,3	270,4	270,4	270,7	270,3
270,4	269,9	270,7	270,8	270,3	270,5	270,5	270,8
270,6	270,6	270,2	270,5	270,3	270,2	270,3	270,2
270,5	270,4	270,1	270,3	270,4	270,3	270,1	270,4
270,3	270,5	270,6	270,1	270,0	270,1	270,3	270,7

19. Con riferimento alla stessa distribuzione dell'esercizio precedente, disegna un istogramma che rappresenti la distribuzione degli scarti delle misure dalla loro media aritmetica, dopo averli raggruppati in classi opportune, ad esempio in 8 classi.

Organizza una tabella in cui sono riportate le frequenze cumulate e, con riferimento a questa tabella, calcola la media aritmetica e lo scarto quadratico medio.

Quali errori relativi si commettono se questi sono assunti come valori effettivi della media aritmetica e dello scarto quadratico medio delle misure rilevate?

Cosa succede alla media aritmetica ed allo scarto quadratico medio se ciascuno dei dati rilevati aumenta di 10 cm?

[**R.** $M \approx 270{,}37$ cm; $s = 0{,}17$ cm; $\sigma \approx 0{,}21$ cm; ...]

20. LABORATORIO DI MATEMATICA. Assieme ai tuoi compagni di classe lascia cadere un oggetto da una certa altezza. Ripeti l'esperimento, sempre con lo stesso oggetto che cade dalla stessa altezza, per 30 volte. Ogni volta misura il tempo di caduta, servendoti di un cronometro di precisione. Dopo aver raccolto i tempi ottenuti su una tabella, disegna un istogramma che rappresenti la distribuzione dei tempi rilevati, dopo averli raggruppati in classi opportune. Organizza quindi una tabella delle frequenze cumulate e, sulla base di questa tabella, calcola la mediana, la media aritmetica e la deviazione standard delle misure rilevate.

21. È stata condotta un'indagine statistica su un campione di 9862 persone, di età superiore ai 18 anni, variamente distribuite sul territorio nazionale, orientata a determinare il numero degli anni di scuola frequentati a partire dalla 1ª elementare ed esclusa l'università, senza contare gli eventuali anni ripetuti; a determinare sostanzialmente l'ultima classe frequentata. I dati rilevati sono sintetizzati nella tabella sottostante.

Anni di scuola	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Frequenze assolute	2	4	11	34	9	1.417	63	352	4.841	218	225	374	14	2.298

Calcola: a) la moda, la mediana, la media aritmetica degli anni di scuola frequentati; b) lo scarto quadratico medio. [**R.** a) 8; 8; 8, b) 2,7]

22. Un'automobile viaggia sull'autostrada del Sole. Precisamente percorre un primo tratto alla velocità media v_1, un secondo tratto alla velocità media v_2, un terzo ed ultimo tratto alla velocità media v_3. Tenendo presente che i tre tratti sono di uguale lunghezza, dimostrare che la velocità media dell'automobile sull'intero percorso è uguale alla media armonica delle tre velocità medie.

23. Un'automobile percorre il tratto di autostrada Roma-Firenze alla velocità media di $120\ \text{km/h}$ e il tratto di ritorno alla velocità media di $130\ \text{km/h}$. Qual è la velocità media sull'intero percorso di andata e ritorno? [**R.** 124,8 km/h]

24. La classe è composta da 18 alunne e 8 alunni. La media aritmetica delle altezze delle prime è 160,2 cm, quella dei secondi è di 168,4 cm. Calcolare la media aritmetica delle altezze di tutti (alunni e alunne). [**R.** ≈ 162,7 cm]

25. LABORATORIO DI MATEMATICA. Nel campionato italiano di calcio 2010/2011 l'Udinese realizzò solo 1 punto nelle prime 5 partite ma alla fine del campionato (38 partite complessive) si classificò quarta realizzando una media di circa 1,737 punti a partita. Quale fu la media punto nelle prime 5 partite? Quale quella nelle 33 partite successive? Discutine con i tuoi compagni e, se è il caso, chiedete l'aiuto del professore. [**R.** 0,2; ≈1,970]

26. PROBLEMA RISOLTO. Alla festa di compleanno di Anna l'età media dei partecipanti è di 22 anni. Se l'età media degli uomini è 26 anni e quella delle donne è 19 anni, qual è il rapporto tra il numero degli uomini e quello delle donne?

[**R.** Tratto dagli esami di Stato 2009, sessione ordinaria, indirizzo sperimentale]

RISOLUZIONE. Se N_U è il numero degli uomini e N_D quello delle donne, indicata con M_U la media aritmetica delle età degli uomini e con M_D quella delle donne e, inoltre, con S_U la somma delle età degli uomini e con S_D quella delle donne, si ha:

$$\frac{S_U}{N_U}=M_U=26,\quad \frac{S_D}{N_D}=M_D=19,\quad \frac{S_U+S_D}{N_U+N_D}=22.$$

Da qui segue rispettivamente:

$$S_U=26\ N_U,\ S_D=19\ N_D,\ S_U+S_D=22\ (N_U+N_D).$$

Pertanto:

$$26\ N_U+19\ N_D=22\ N_U+22\ N_D;\ \text{ossia: } 4\ N_U=3\ N_D \text{ e infine: } \frac{N_U}{N_D}=\frac{3}{4}.$$

27. Fornire un esempio in cui il valore di sintesi più appropriato a descrivere un fenomeno collettivo è la mediana; un esempio in cui tale valore è la media aritmetica; un esempio in cui esso è la moda; ancora un esempio in cui quel valore è la media geometrica; ed infine un esempio in cui esso è la media armonica.

28. La media aritmetica delle altezze degli alunni di ognuna delle 25 classi del Liceo "Galilei" non supera 175 cm. Dimostrare che la media aritmetica delle altezze di tutti gli alunni del Liceo "Galilei" non supera 175 cm o, viceversa, dimostrare che questo non è vero.

29. L'istogramma sottostante (Fig. 13) visualizza la distribuzione delle percentuali di partecipanti ad un concorso in funzione delle percentuali di punteggio conseguite nella prova scritta, calcolate rispetto al punteggio massimo conseguibile.

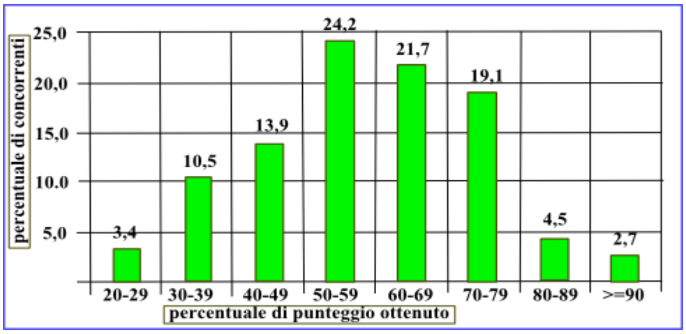


FIG.13

a) Considerato che supera la prova chi ha ottenuto almeno il 60% del punteggio massimo conseguibile, quanti concorrenti hanno superato la prova, posto che essi erano complessivamente 446?
b) Qual è stata la percentuale media di punteggio? Quale la mediana?
c) Quanto vale la deviazione standard?

30. La divisa dello Stato di Florilandia è la "lira", quella dello Stato di Lupilandia è il "marco". Il grafico sottostante (Fig. 14) mostra l'andamento delle due valute negli ultimi 12 mesi [8].

a) Il grafico mostra che negli ultimi 12 mesi entrambe le valute si sono apprezzate: quale si è apprezzata di più? In quali percentuali, rispetto ai valori iniziali, si sono apprezzate le due valute?

b) Di quanto, mediamente ogni mese, si sono apprezzate le due valute?

c) In quali periodi la lira è risultata più forte del marco? In quali periodi quest'ultimo è risultato più forte della lira? Ci sono stati dei momenti in cui le due valute hanno avuto lo stesso valore?

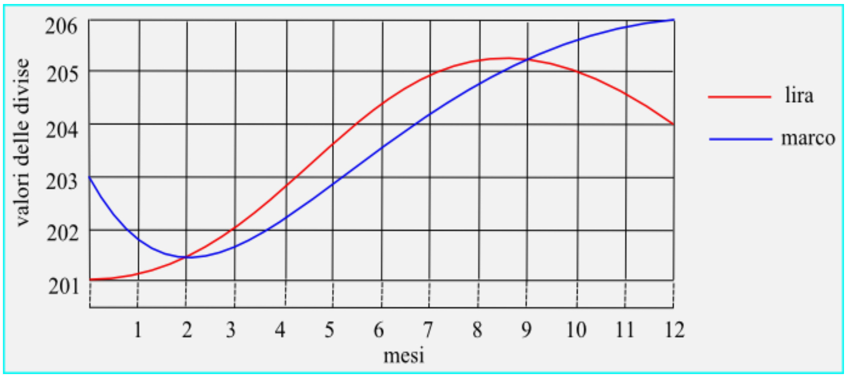


FIG. 14

31. Una certa quantità d'acqua è contenuta in una pentola collocata su un fornello elettrico. La temperatura dell'acqua, che inizialmente è di 15°C, aumenta col trascorrere del tempo e un termometro, che pesca nell'acqua, segnala le temperature registrate nella tabella sottostante.

Tempi (in minuti)	0	5	10	15	20	25
Temperature (in gradi centigradi)	15,0	30,4	47,5	66,2	83,5	100,0

Dopo aver disegnato un diagramma cartesiano, che visualizzi la situazione reale, dire quale dei grafici sottostanti (Fig. 15) si adatta meglio alla situazione.

Dopo aver riprodotto il grafico scelto su carta adeguatamente quadretta, utilizzarlo per calcolare, con approssimazione ad 1°C, le temperature dell'acqua ai minuti 8 e 18.

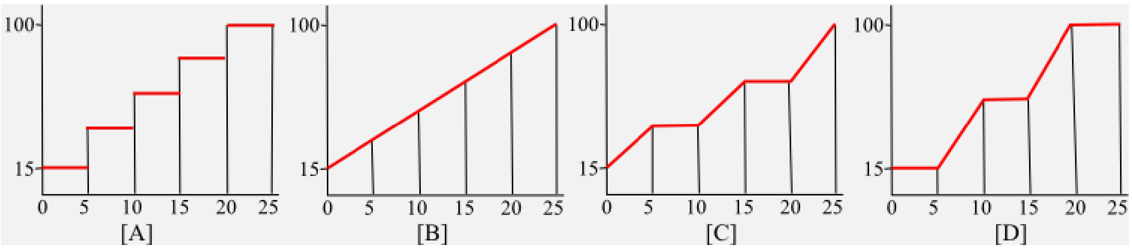


FIG. 15

32. Sono dati tre numeri reali a, b, c, con a>b>c. Si calcola la media aritmetica m' dei primi due numeri e successivamente la media aritmetica m di m' e c. Indicata con M la media aritmetica dei tre numeri a,

[8] Alle volte, come nel caso in esame, è lecita una rappresentazione grafica che prescinda dalla reale distanza dei punti degli assi coordinati dall'origine del sistema di riferimento. Qui, com'è evidente, l'eccezione riguarda solamente l'asse delle ordinate, ma in altre circostanze può riguardare l'asse delle ascisse o addirittura entrambi gli assi coordinati.

b, c, dimostrare che risulta $m<M$.

33. Un automobilista percorre una determinata distanza dividendola in tre settori di uguale lunghezza. Percorre il primo settore alla velocità media di $62\ \mathrm{km/h}$, il secondo alla velocità media di $74\ \mathrm{km/h}$ e il terzo alla velocità media di $80\ \mathrm{km/h}$. Conclude, in modo affrettato, che sull'intero percorso ha tenuto una velocità media pari alla media aritmetica delle velocità medie tenute sui tre settori. Ha in effetti commesso un errore, ma la sua valutazione non è molto distante da quella reale, come mostrano le risposte alle seguenti domande.
a) Qual è l'errore assoluto commesso dall'automobilista? b) Quale l'errore relativo?
[**R.** a) $\varepsilon \approx 0{,}813\ \mathrm{km/h}$; b) $\varepsilon_r \approx 1{,}14\ \%$]

UNA BREVE SINTESI PER DOMANDE E RISPOSTE

DOMANDE.

1. È vero che un fenomeno collettivo è una manifestazione democratica di un gruppo di persone?
2. È vero che una rilevazione di dati statistici si dice totale perché è fatta da tutte le persone addette alla rilevazione stessa?
3. Come può essere definita la statistica (descrittiva)?
4. È vero o è falso che la media aritmetica di due numeri positivi, comunque scelti, è maggiore della loro media geometrica?
5. Si deve rifornire di scarpe una guarnigione militare: serve una media delle misure delle scarpe dei soldati. Qual è il valore di sintesi che meglio si adatta alla situazione?
6. Si vuole valutare il livello di rendimento di una classe in matematica: lo si fa sulla base dei voti conseguiti dagli alunni alla fine del 1° quadrimestre. Qual è il valore di sintesi che meglio si adatta alla situazione?
7. Si vuole valutare se un dato compito in classe è andato bene o no: si suppone che sia andato bene se il numero dei voti da "6" in su supera quello dei voti inferiori al "6". Qual è il valore di sintesi che meglio si adatta alla situazione?
8. Un quotidiano titola a caratteri cubitali: «Scandaloso: i voti riportati dagli studenti negli esami di matematica sono per il 50% al di sotto della mediana». Trovi anche tu che ciò sia scandaloso?
9. Bisogna calcolare la media aritmetica delle altezze degli studenti della tua classe. Basta calcolare la semisomma delle altezze minima e massima. È vero o falso?
10. L'operazione "media aritmetica di due numeri" gode della proprietà associativa. È vero o è falso?
11. Si sa che una serie di dati sono distribuiti simmetricamente rispetto alla loro media aritmetica. Questo è sufficiente per stabilire come sono disposte rispetto a tale media la mediana e la moda?
12. Se gli N valori di una successione di dati aumentano o diminuiscono nella stessa misura, la deviazione standard non varia. È vero o falso?
13. La media aritmetica di una serie di dati statistici calcolata coinvolgendo i singoli valori effettivamente rilevati è uguale a quella che si ottiene basandosi sulla distribuzione dei valori cumulati. È vero o falso?
14. È vero che uno degli indici di dispersione più usati è costituito dalla media aritmetica degli scarti di una serie di dati dalla loro media aritmetica?

15. È vero che la media aritmetica degli scarti di una serie di dati da una loro media è 0?

16. Quel sito web ha fatto registrare una media di N_1 visite al giorno nel mese di marzo e di N_2 viste al giorno nel mese di aprile. È corretto affermate che la media nell'arco dei due mesi è stata di $(N_1+N_2)/2$ visite al giorno, cioè la media aritmetica delle medie giornaliere di ciascun mese?

RISPOSTE.

1. No. Un fenomeno collettivo è un fenomeno in cui sono coinvolti molti individui, detti unità statistiche. Esso è oggetto di studio della Statistica.

2. No. Si dice totale perché coinvolge tutte le unità statistiche.

3. La statistica può essere definita come la scienza che studia la raccolta dei dati (le unità statistiche), la loro organizzazione (mediante grafici e/o tabelle) e la loro interpretazione (per mezzo di indici numerici).

4. È falso. La media aritmetica di due numeri positivi è maggiore della loro media geometrica a condizione che i due numeri non siano uguali.

5. La moda.

6. La media aritmetica dei voti.

7. La mediana.

8. Ciò che è scandalosa è l'ignoranza crassa di chi ha fatto quel titolo. La mediana è per definizione il valore che divide a metà i dati statistici presi in esame ed è pertanto ovvio che il 50% di tali dati si collochi al di sotto di essa.

9. Falso: la media aritmetica semplice di due numeri non gode della proprietà associativa. Gode di tale proprietà invece la media aritmetica ponderata a condizione che ogni dato sia preso assieme al suo peso.

10. È falso: la media aritmetica di due numeri non gode della proprietà associativa. Ne gode invece la media aritmetica ponderata, a condizione che ciascun dato sia considerato assieme al suo peso.

11. La mediana coincide con la media aritmetica. Anche la moda vi coincide, ma a condizione che la distribuzione sia unimodale, cioè presenti una sola moda.

12. Vero, la deviazione standard non varia. Varia invece la media aritmetica, che aumenta o diminuisce nella stessa misura in cui aumentano o diminuiscono gli N dati rilevati.

13. È falso, ma la seconda è un'ottima approssimazione della prima.

14. No. La media aritmetica degli scarti di una serie di dati dalla loro media aritmetica è nulla e perciò non può essere usata come indice di dispersione.

15. No. Questo è vero solo se la media rispetto alla quale si calcolano gli scarti è la media aritmetica dei dati assegnati.

16. No, non è corretto. La media esatta è invece $\frac{N_1 \cdot 31 + N_2 \cdot 30}{31+30}$, vale a dire la media aritmetica ponderata delle medie giornaliere di ciascun mese, assumendo come pesi i numeri dei giorni dei rispettivi mesi.

LETTURA

IL *QUINCONCE* DI GALTON

Nella seconda metà dell'Ottocento, quella che in seguito sarebbe stata chiamata *curva a campana di Gauss* era semplicemente la *curva degli errori*. Ne abbiamo spiegato i motivi nel precedente paragrafo N° 12.4.7.

Non era però ancora ben chiaro e comunque non era accettato da tutti che certe misure come, per esempio, l'altezza delle persone, il loro peso, la loro circonferenza toracica, e altre ancora, si distribuissero normalmente, dando luogo cioè ad una distribuzione dei dati somigliante ad una curva a campana.
Per sostenere questa tesi lo scienziato inglese Francis Galton (1822-1911) costruì nel 1873 un meccanismo che "visualizza" il concetto stesso. Questo congegno si chiama ***macchina di Galton*** o anche ***quinconce di Galton***. È costituito da una tavola verticale (Fig. 16, parte alta) nella quale sono piantati dei chiodi disposti secondo la configurazione del numero 5 in un comune dado (Fig. 17), ossia in modo che ogni chiodo, salvo ovviamente quelli della regione periferica, sia circondato da altri quattro. Da qui il nome di "quinconce" (9).
Da una fessura idonea, situata in cima alla tavola, sono fatte cadere delle palline (di norma almeno 30), le quali, rimbalzando sui chiodi, si muovono casualmente e finiscono per essere raccolte in alcuni contenitori cilindrici posti alla base della tavola.
Ebbene, la disposizione delle palline assomiglia proprio ad una curva a campana (Fig. 16, parte bassa).
In molte scuole, nel laboratorio di Fisica, è presente la macchina di Galton. Eseguire direttamente l'esperimento non fa male alla salute.
Si può avere, comunque, un'idea seppur vaga della macchina di Galton assistendo al gioco televisivo THE WALL, condotto da Gerry Scotti.

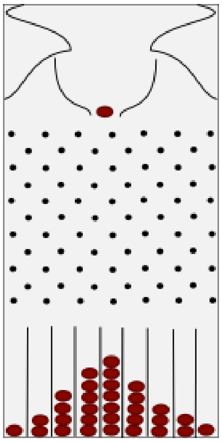


FIG. 16

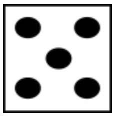


FIG. 17

[9] *Quinque* è il nome latino di cinque. Esistono esempi interessanti di disposizione a quinconce. Ne mostriamo due: la diposizione delle stelle nella bandiera americana, la disposizione iniziale delle pedine nel gioco della dama.