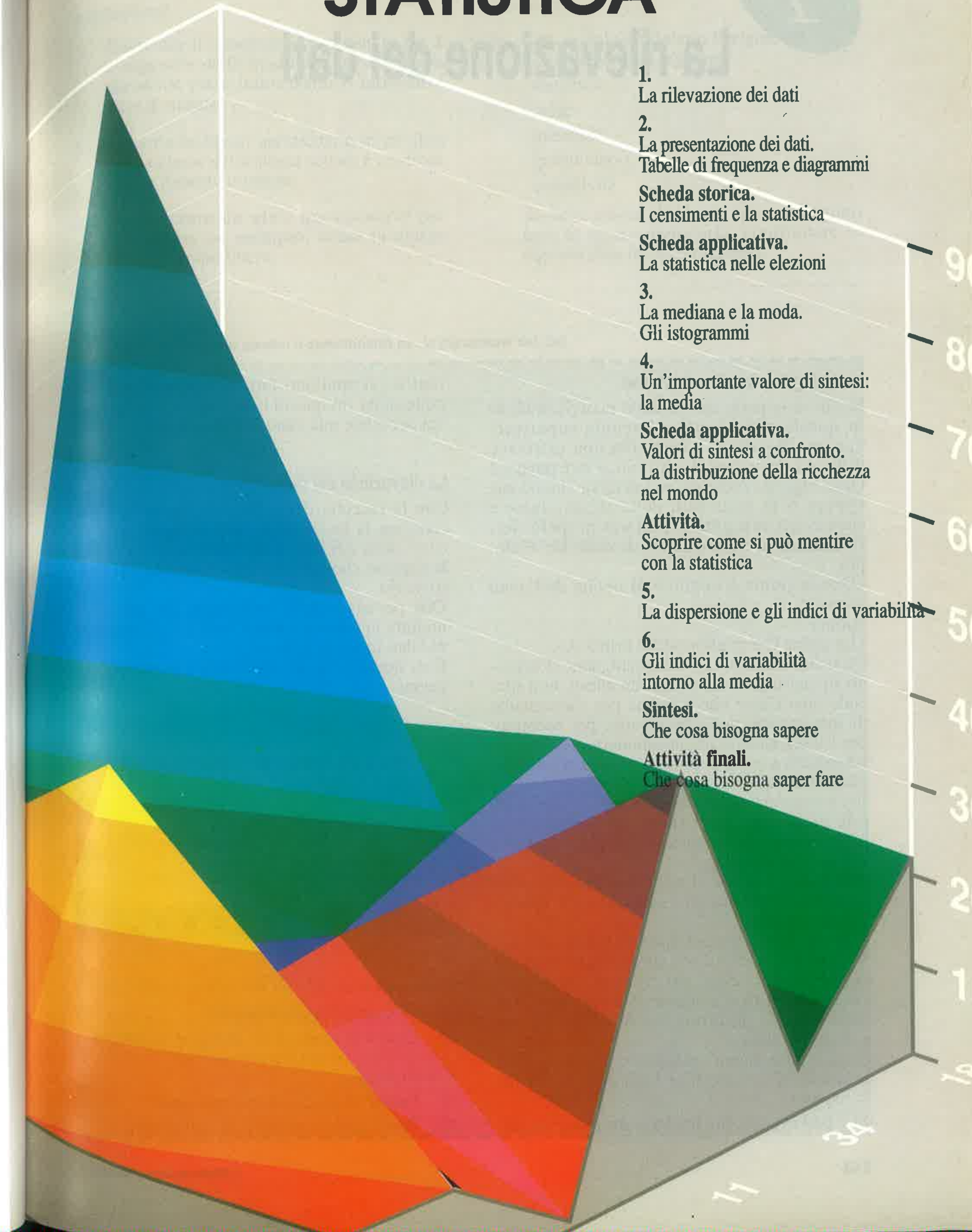
CAPITOLO DECIMO

# STATISTICA



# 1 La rilevazione dei dati

## Un'indagine statistica in classe

In classe si parla spesso delle materie studiate in questo primo anno di scuola superiore: materie che alla scuola media non offrivano difficoltà sembrano ora difficili, e viceversa.
Ogni ragazzo presenta una diversa situazione, eppure si fa parte tutti della stessa classe e spesso gli insegnanti parlano proprio dell'«andamento della classe», dicendo per esempio:
«Questa prima è migliore di quella dell'anno scorso»;
oppure:
«La prima C è migliore della prima A».
Qualche volta tutti gli insegnanti sono d'accordo su questi giudizi; altre volte questo non succede: una classe che è buona per l'insegnante di matematica, può non esserlo, per esempio, per l'insegnante di lingua straniera.
La statistica cerca di portare chiarezza in situazioni come queste.
Ecco come procedere: si esplorano le difficoltà e le preferenze degli alunni, distribuendo un *questionario* come quello che si trova nella pagina seguente (fig. 1).
Come dice il nome, il questionario è una serie di domande (o *questioni*), a cui ogni alunno può rispondere.
Il questionario resterà anonimo, dato che sul foglio non è richiesto di scrivere nome e cognome. Lo scopo infatti non è quello di «incasellare» ogni studente: è invece quello di conoscere le preferenze o le difficoltà più comuni in classe.
Quando tutti hanno completato il proprio questionario, si raccolgono i fogli e si esaminano le risposte.
Si è così organizzata in classe *un'indagine statistica:* si studiano fatti che riguardano una collettività (in questo caso i ragazzi della classe) per avere una visione d'insieme.

## La rilevazione dei dati

Con la raccolta dei questionari compilati si è conclusa la fase iniziale dell'indagine, cioè la *rilevazione dei dati*: i dati sono, in questo caso, le risposte che ogni ragazzo ha scritto sul questionario.
Ora, per avere delle conclusioni che interessino tutta la classe, bisogna ordinare e riassumere i dati in modo efficace.
È di questo che si occupano i paragrafi seguenti.

## Verifiche

*Conoscenze*

① Che cos'è un questionario?

② Qual è lo scopo di un'indagine statistica?

*Comprensione*

① Perché il questionario è anonimo?

② Porta qualche esempio di collettività.



*Applicazioni*

① Esaminare il questionario proposto in fig. 1 e suggerire delle modifiche per eliminare domande poco interessanti o introdurre altre domande.

② Proporre lo stesso questionario in un'altra prima classe della stessa scuola e confrontare le risposte ottenute.

③ Organizzare un altro questionario per conoscere, ad esempio, come la classe passa il tempo libero.

*Vocabolario*

① Cercare sul vocabolario l'origine e il significato dei termini seguenti:
- statistica;
- dato;
- questioni;
- questionario;
- collettività.

Se nella classe gli alunni non posseggono tutti lo stesso dizionario, confrontare le risposte date dai vari dizionari.

**Figura 1**
**Un'indagine statistica in classe: il questionario per la rilevazione dei dati**

**Questionario**

**classe ........** **maschio** ☐ **femmina** ☐

**1. Indica la materia scolastica più difficile**

italiano ☐
lingua straniera ☐
storia ☐
matematica ☐
(altro)................................................ ☐

**2. Indica la materia scolastica più facile**

italiano ☐
lingua straniera ☐
storia ☐
matematica ☐
(altro)................................................ ☐

3. Voto ottenuto nell'ultimo compito di matematica ☐
4. Voto ottenuto nell'ultimo compito di italiano ☐
5. Voto ottenuto nell'ultimo compito di lingua straniera ☐

# 2 La presentazione dei dati. Tabelle di frequenza e diagrammi

## Presentazione di dati

I questionari compilati forniscono un ammasso disordinato di risposte; per trarne delle informazioni utili occorre ordinare i dati e scriverli in forma facilmente comprensibile.
Ecco un esempio di lavoro svolto in una classe di 25 alunni, 15 ragazzi e 10 ragazze.
Per esaminare i dati si può cominciare, per esempio, dalle risposte alla prima domanda, che riguarda le difficoltà offerte dalle materie scolastiche.
Il modo più semplice di descrivere la situazione della classe è il seguente:

- contare gli alunni che ritengono difficile ciascuna materia;
- presentare i dati in una forma espressiva.

Le forme più usate per presentare i dati sono:

- le tabelle;
- i grafici (o diagrammi).

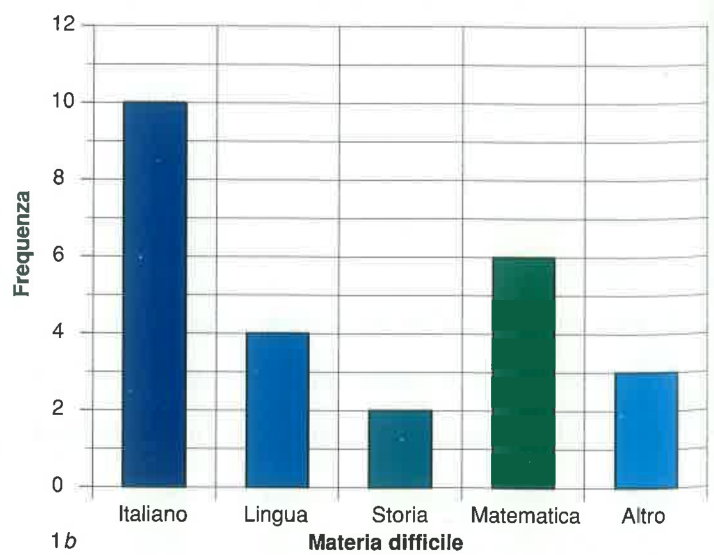
## Le tabelle di frequenza

È facile presentare dati con una tabella come quella di fig. 1*a*: in una colonna sono elencate le materie scolastiche e, a fianco di ogni materia, il numero di alunni che ritiene difficile quella materia. Questo numero prende il nome di *frequenza.*
Il termine frequenza compare spesso nelle indagini statistiche con il seguente significato: *frequenza è il numero di volte che un fenomeno si presenta.*

**Figura 1**
**Tabella di frequenza e diagramma a strisce**

| Materia difficile | Frequenza |
|---|---|
| Italiano | 10 |
| Lingua | 4 |
| Storia | 2 |
| Matematica | 6 |
| Altro | 3 |
| Totale | 25 |

1*a*




1*b*



## I diagrammi a strisce

La tabella è chiara, ma richiede un'attenta lettura; i dati numerici diventano molto più espressivi se sono rappresentati in un grafico o diagramma come quello di fig. 1*b*, organizzato nel modo seguente:

- sull'asse orizzontale sono indicati i nomi delle materie;
- sull'asse verticale il numero di studenti;
- in corrispondenza di ogni materia si è disegnata una *striscia* che ha l'altezza stabilita dal numero di allievi.

Proprio per questo il grafico ottenuto prende il nome di *diagramma a strisce.*

## Un problema che conduce a riflettere sulla frequenza

Spesso si trovano differenze d'opinione fra ragazzi e ragazze; qual è la situazione in questa classe?
Per esaminarla bisogna separare i questionari dei ragazzi da quelli delle ragazze e compilare (fig. 2):

- la tabella di frequenza ed il diagramma a strisce relativi alle ragazze;
- la tabella di frequenza ed il diagramma a strisce relativi ai ragazzi.

Non è facile però confrontare le situazioni descritte dalla figura; per esempio, dalla tabella di frequenza, si rileva che:

- fra i ragazzi, 6 ritengono l'italiano la materia più difficile;
- fra le ragazze, solo 4 hanno la stessa opinione.

È corretto concludere che, in questa classe, i ragazzi incontrano nello studio dell'italiano maggiori difficoltà rispetto alle ragazze?

Certamente questa conclusione si basa sulle frequenze rilevate, ma non tiene conto di tutte le informazioni disponibili; soprattutto viene ignorato che nella classe ci sono più ragazzi che ragazze.

Ecco allora come si deve procedere per esaminare correttamente questa situazione:

- si calcola il rapporto F fra il numero dei ragazzi che ritengono l'italiano difficile (la frequenza 6) e il numero totale dei ragazzi della classe (15); si ottiene:

$$F = \frac{6}{15} = 0{,}4$$

- si calcola il rapporto F' fra il numero delle ragazze che ritengono l'italiano difficile (la frequenza 4) e il numero totale delle ragazze della classe (10); si ottiene:

$$F' = \frac{4}{10} = 0{,}4$$

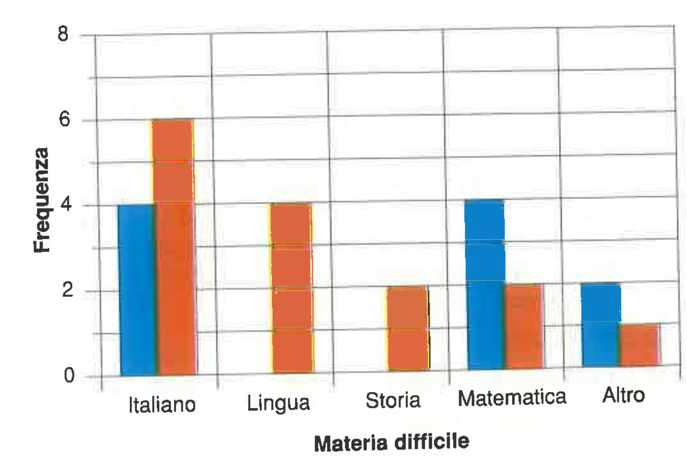
- si confronta F con F', e si trova:

$$F=F'$$

- si conclude che, nella classe esaminata, i ragazzi e le ragazze hanno incontrato le stesse difficoltà nello studio dell'italiano.

**Figura 2**
**Le opinioni delle ragazze e dei ragazzi**

| Materia difficile | Frequenza | |
|---|---|---|
| | ragazze | ragazzi |
| Italiano | 4 | 6 |
| Lingua | - | 4 |
| Storia | - | 2 |
| Matematica | 4 | 2 |
| Altro | 2 | 1 |
| Totale | 10 | 15 |



### Frequenza relativa e frequenza assoluta

Il procedimento seguito è basato su un'idea: il confronto fra le risposte di due collettività deve tener conto del fatto che le due collettività (i ragazzi e le ragazze) sono composte da un numero diverso di elementi (15 e 10).
Per questo si valuta ogni frequenza *relativamente* al numero di elementi della collettività; questa valutazione si ottiene, per esempio nella collettività dei ragazzi, con il rapporto:

$$F = \frac{6}{15} = 0{,}4$$

che è ottenuto calcolando:

$$F = \frac{\text{frequenza}}{\text{numero di dati esaminati}}$$

Questo rapporto prende anche il nome di *frequenza relativa.*
Spesso, per distinguere i due modi di valutare la frequenza, si usa il termine *frequenza assoluta*, invece di frequenza.
Frequenza assoluta e frequenza relativa sono termini molto usati nelle indagini statistiche con i seguenti significati:

- *la* frequenza assoluta *è il numero di volte che un fenomeno si presenta;*
- *la* frequenza relativa *è il rapporto fra la frequenza assoluta e il numero di dati esaminati.*

### Frequenza relativa in forma percentuale

La frequenza relativa è dunque espressa da una frazione o dalla sua espressione sotto forma di numero decimale.
Ma spesso la frequenza relativa viene anche espressa in notazione percentuale, scrivendo, per esempio:

$$0{,}4 = \frac{40}{100} = 40\%$$

La percentuale ha un significato intuitivo: in una collettività composta di 100 persone con le stesse opinioni del gruppo di 15 ragazzi ora esaminati, si troverebbero 40 persone che ritengono difficile lo studio dell'italiano.
Nella tabella di fig. 3 sono presentati i dati relativi alle opinioni dei ragazzi e delle ragazze della classe, valendosi appunto della frequenza relativa espressa in tre modi, e cioè:

- con una frazione;
- con un numero decimale;
- con la notazione percentuale.

### Diagrammi a settori circolari

Le frequenze relative diventano più espressive se vengono rappresentate graficamente. In questo caso conviene valersi dei *diagrammi a settori circolari*, ottenuti con il procedimento seguente (fig. 4):

- si divide un cerchio in 100 settori uguali, ciascuno di ampiezza $\alpha$, data da:

$$\alpha = \frac{360°}{100} = (3{,}6)°$$

- si osserva che ogni settore è $\frac{1}{100}$ del cerchio e perciò rappresenta una percentuale dell'1% dell'intero cerchio;
- si rappresenta una percentuale del $p\%$ con un settore di ampiezza $\beta$, data da:

$$\beta = p \cdot (3{,}6)°$$

Calcolando le ampiezze corrispondenti alle percentuali delle tabelle di fig. 3, si ottengono i diagrammi a settori rappresentati in fig. 5.

**Figura 3**
**Tabelle con le frequenze relative**

**Ragazze**

| Materia difficile | Frequenza relativa | Decimale | Percentuale |
|---|---|---|---|
| Italiano | $\frac{4}{10}=\frac{2}{5}$ | 0,4 | 40% |
| Matematica | $\frac{4}{10}=\frac{2}{5}$ | 0,4 | 40% |
| Altro | $\frac{2}{10}=\frac{1}{5}$ | 0,2 | 20% |
| Totale | 1 | 1 | 100% |

**Ragazzi**

| Materia difficile | Frequenza relativa | Decimale | Percentuale |
|---|---|---|---|
| Italiano | $\frac{6}{15}=\frac{2}{5}$ | 0,4 | 40% |
| Lingua | $\frac{4}{15}$ | 0,27 | 27% |
| Storia | $\frac{2}{15}$ | 0,13 | 13% |
| Matematica | $\frac{2}{15}$ | 0,13 | 13% |
| Altro | $\frac{1}{15}$ | 0,07 | 7% |
| Totale | 1 | 1 | 100% |





## V e r i f i c h e

*Conoscenze*

① Come si dispongono i dati in una tabella di frequenza?
② Che cosa significa il termine «frequenza di un fenomeno»?
③ Come si traccia un diagramma a strisce?
④ Che cosa significano i termini «frequenza assoluta» e «frequenza relativa»?
⑤ Come si determina la frequenza relativa di un fenomeno?
⑥ Come si traccia un diagramma a settori?

*Comprensione*

① La frequenza assoluta può essere una frazione?
② La frequenza relativa può essere un intero?
③ Le frequenze assolute possono essere rappresentate con un diagramma a settori?
④ Le frequenze relative possono essere rappresentate con un diagramma a strisce?

*Applicazioni*

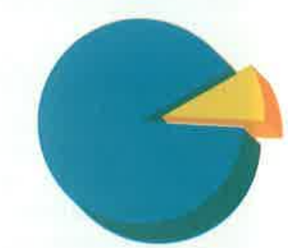
① Riprodurre il diagramma a strisce di fig. 1 in due modi diversi:
  a. dimezzando la scala sull'asse verticale (1 settore=2 persone);
  b. raddoppiando la scala sull'asse verticale (2 settori=1 persona).
  Confrontare «l'impressione» che danno i grafici così ottenuti con quella data dal diagramma del testo.
② Ripetere le considerazioni esposte nel paragrafo per esaminare i dati relativi alla materia più facile.

*Vocabolario*

① Cercare su un vocabolario l'origine e il significato dei termini «tabella», «grafico», «diagramma», «frequenza».

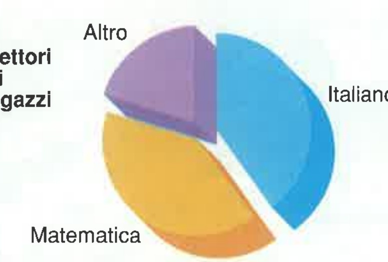
**Figura 4**
**Come si traccia un diagramma a settori circolari**

Settore ampio 3,6° = 1% dell'intero cerchio
99% del cerchio

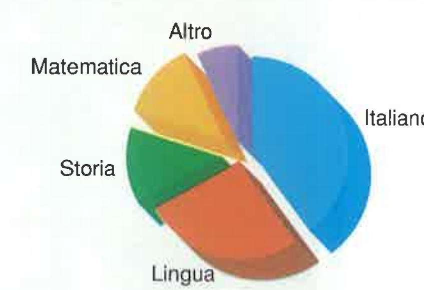


Settore ampio 10 · 3,6° = 10% dell'intero cerchio
90% del cerchio

**Figura 5**
**Diagrammi a settori con le opinioni di ragazze e ragazzi**




| Ragazze | *p* | β |
|---|---|---|
| Italiano | 40% | 40 · (3,6°) = 144° |
| Matematica | 40% | 144° |
| Altro | 20% | 72° |




| Ragazzi | *p* | β |
|---|---|---|
| Italiano | 40% | 144° |
| Lingua | 27% | 97° |
| Storia | 13% | 47° |
| Matematica | 13% | 47° |
| Altro | 7% | 25° |

# I censimenti e la statistica

## Il termine «statistica» è recente

Fino a poche decine di anni fa la *statistica* non faceva parte del linguaggio comune e non compariva come materia di studio nelle scuole secondarie; il problema di organizzare delle *statistiche* era lasciato allo Stato.

Infatti, in origine, il termine statistica si riferiva ai *censimenti*, cioè alle indagini condotte dallo Stato per conoscere la composizione della popolazione: quanti erano gli abitanti di una data età, quanti avevano un dato *censo* (ossia una data situazione economica), etc.

## I censimenti hanno origini antichissime

In Cina, in India, in Assiria, in Egitto, in Grecia, a Roma si effettuavano periodicamente delle inchieste sul numero di abitanti, sui quantitativi annuali di grano o di altri prodotti e, sulla base di questi dati, venivano stabilite le tasse.

Nella Bibbia si trovano alcuni fra i documenti più antichi: si parla, nel Vecchio Testamento, di un censimento indetto da Mosè nel deserto del Sinai e, nel Nuovo Testamento, del censimento organizzato dai romani all'epoca della nascita di Cristo (fig. 1).

**Figura 1**
**Il censimento a Betlemme in un quadro del pittore fiammingo Pieter Bruegel**





## I censimenti a Roma dall'antichità al Medioevo

Anche nei libri di scrittori latini come Tito Livio, Strabone o Giovenale si trovano notizie di antichi censimenti.

Sappiamo così che Roma, poco dopo la sua fondazione, aveva indetto un censimento della popolazione: fu condotto da Servio Tullio nel 555 a.C.

A questo censimento ne seguirono tanti altri; avvenivano ogni cinque anni, ogni *lustro*, ed è per questo che la parola «lustro» indicò sia il periodo di cinque anni, sia il censimento stesso.

Più tardi si trovano i termini *censo*, per indicare il computo della popolazione ed anche la tassa da pagare, e *censore*, per indicare l'addetto al censimento; questi termini puntano l'attenzione sullo scopo principale del censimento: conoscere la composizione e la ricchezza della popolazione per reclutare soldati e anche per imporre il pagamento delle tasse.

Gli scrittori latini riferiscono che il *pater familias*, cioè quello che era allora il capo della famiglia, denunciava spesso meno membri della famiglia e meno ricchezze. Ma i censori romani erano severissimi e conducevano la lotta contro le menzogne nei censimenti infliggendo pene molto gravi: il colpevole poteva anche essere fustigato o venduto come schiavo.

L'ultimo censimento dell'antica Roma è del 260 d.C.; dopo, in seguito alle invasioni barbariche, si interrompe la tradizione dei lustri e non si hanno più notizie certe degli abitanti di Roma per quasi mille anni.

Bisogna arrivare al 1198 per trovare i dati di un censimento nella Roma dei Pontefici: gli abitanti erano 35 000. Successivamente, gli abitanti di Roma diminuiscono fino a diventare 17 000 nel 1377, epoca del trasferimento della sede pontificia ad Avignone.

Dopo il 1500 si trova invece un aumento della popolazione, aumento non sempre uniforme sia per le numerose guerre che per le paurose pestilenze.

## In Italia censimento voleva dire «paura»

I censimenti – come si è visto – avevano due scopi:

- conoscere quanti (e quali) erano gli uomini atti alla guerra;
- conoscere quante (e quali) erano le famiglie che dovevano versare denaro allo Stato, cioè che dovevano pagare le imposte (o tasse).

Perciò un censimento faceva paura: la gente ne diffidava, cercava di sfuggire all'inchiesta e spesso, in quei giorni, abbandonava le case o le capanne per nascondersi.

Ma le abitazioni rimanevano e gli addetti al censimento buttavano giù la porta, entravano e ispezionavano: un focolare, della legna bruciata di recente erano lì a testimoniare che qualche famiglia viveva sotto quel tetto.

Ecco dunque che «un fuoco» segnalava un nucleo familiare e la catasta più o meno grande di legna da ardere dava un'idea di quanti dovevano essere i componenti; così si stabiliva quale tassa si poteva imporre a quella famiglia.

Per questo, a partire dal primo Medioevo, si ritrovano in tutta l'Italia due termini caratteristici dei censimenti:

- *fuoco*, per indicare una famiglia;
- *focàtico*, per indicare la tassa che doveva pagare la famiglia.

Nel 1815, l'anno del Congresso di Vienna, che dette un nuovo assetto territoriale all'Europa, nella penisola vivevano circa 16 milioni e mezzo di persone, divise per tradizioni, governi e politiche diverse, ma unite dal «fuoco» e dal «focatico» e cioè dalla paura dei censimenti.

## I censimenti dopo l'unità d'Italia

A partire dal 1861, primo anno dopo l'unità d'Italia, il censimento è organizzato ogni dieci anni.

I risultati dei censimenti vengono pubblicati dall'ISTAT (Istituto Centrale di Statistica) e sono a disposizione di chiunque voglia consultarli.

Ma l'ISTAT non si occupa solo dei censimenti che si effettuano ogni dieci anni, organizza ed elabora anche altri dati statistici (fig. 2). Perciò fra le pubblicazioni dell'ISTAT si trovano, per esempio:

- le statistiche dell'istruzione, che forniscono ogni anno i dati sugli alunni dei vari ordini di scuola;
- le statistiche degli incidenti stradali, che forniscono ogni anno i dati sui tipi e la frequenza degli incidenti stradali;
- le statistiche sanitarie, che forniscono ogni anno i dati sulle cause di morte, sulle malattie infettive e sull'attività degli ospedali.

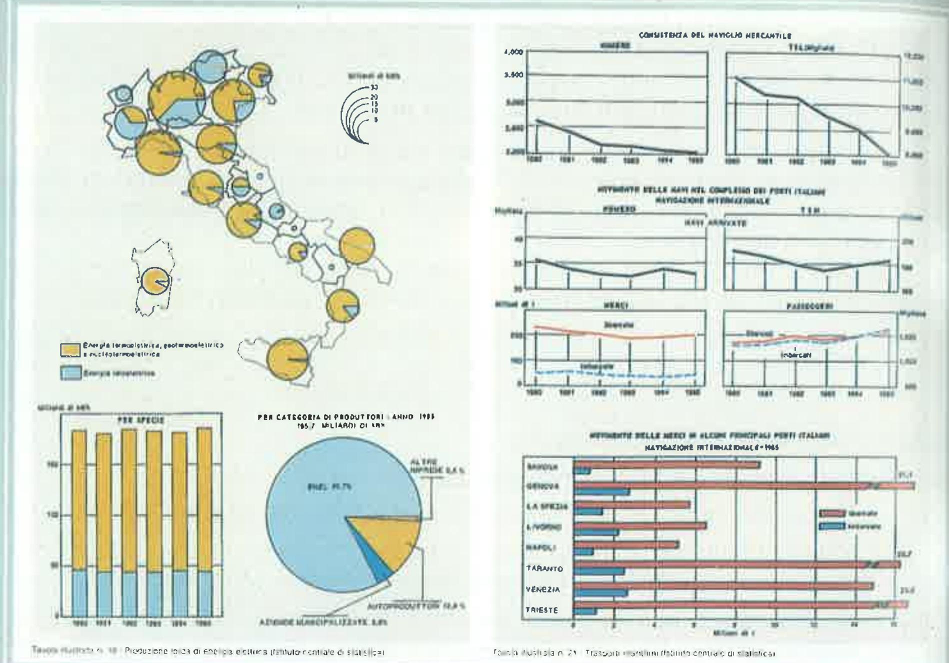


**Figura 2**
**Alcuni dati statistici tratti da pubblicazioni ISTAT**

## Le origini della statistica come scienza

La storia dei censimenti in Italia, tracciata per grandi linee, è dunque arrivata ai giorni nostri, trascurando un importante processo che avveniva parallelamente allo svolgersi dei censimenti: dal semplice computo della popolazione si passava ad indagini complesse riguardanti i fatti più vari, calcolando rapporti, percentuali, medie.

Proprio da questi calcoli si è sviluppata la statistica, diventando poi una scienza a sé.

Molti attribuiscono l'origine di questa scienza a John Graunt, un inglese vissuto a Londra nel Seicento.

Graunt, forse impressionato dalle tante morti dovute alle pestilenze, confrontò i bollettini di mortalità nelle varie zone di Londra, pubblicati dal 1592 in poi; dallo studio di questi bollettini passò quindi a confrontare il numero delle morti con quello delle nascite nei vari decenni del suo secolo.

Ma la grande novità introdotta da Graunt è lo scopo di questi confronti: determinare le leggi fondamentali che regolano la natalità e la mortalità di una popolazione.

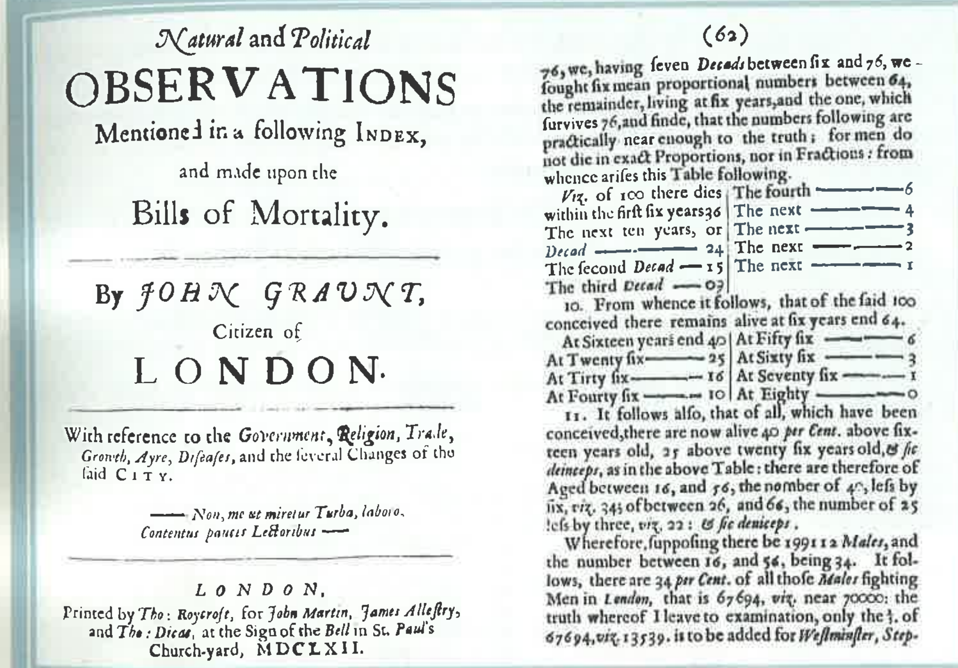
Questo scopo è basato su un'idea a quell'epoca originale: la nascita, la morte di una singola persona sembrano fenomeni dovuti solo al caso; invece, quando si studiano questi stessi fenomeni di una grande popolazione, si scoprono delle regolarità che si ritrovano anche in altre popolazioni.

Ecco due regolarità scoperte da Graunt e ancora oggi valide:

- il numero di nati maschi è di poco superiore al numero di nate femmine;
- invece, nella popolazione adulta si trova un numero circa uguale di maschi e femmine.

Questo ed altri interessanti risultati si trovano in un volumetto pubblicato da Graunt nel 1662 ed intitolato *Osservazioni naturali e politiche fatte sui bollettini di mortalità* (fig. 3).

Il lavoro di Graunt fu accolto con grande interesse dalla Royal Society, un'accademia culturale e scientifica sorta in quell'epoca; ma incontrò anche molte ostilità, soprattutto da una parte della società del suo tempo, quella parte descritta causticamente nella prefazione del volumetto con le parole seguenti:



*Natural* and *Political*

OBSERVATIONS

Mentioned in a following INDEX,

and made upon the

Bills of Mortality.

By *JOHN GRAUNT*,

Citizen of

LONDON.

With reference to the *Government*, *Religion*, *Trade*, *Growth*, *Ayre*, *Diseases*, and the several Changes of the said CITY.

—— *Non, me ut miretur Turba, laboro.*
*Contentus paucis Lectoribus* ——

*LONDON,*
Printed by *Tho: Roycroft*, for *John Martin*, *James Allestry*, and *Tho: Dicas*, at the Sign of the *Bell* in St. *Paul's* Church-yard, MDCLXII.

(62)

76, we, having seven *Decads* between six and 76, we sought six mean proportional numbers between 64, the remainder, living at six years, and the one, which survives 76, and finde, that the numbers following are practically near enough to the truth; for men do not die in exact Proportions, nor in Fractions: from whence arises this Table following.

| | | | |
|---|---|---|---|
| *Viz.* of 100 there dies within the first six years | 36 | The fourth | 6 |
| The next ten years, or *Decad* | 24 | The next | 4 |
| The second *Decad* | 15 | The next | 3 |
| The third *Decad* | 09 | The next | 2 |
| | | The next | 1 |

10. From whence it follows, that of the said 100 conceived there remains alive at six years end 64.

| | | | |
|---|---|---|---|
| At Sixteen years end | 40 | At Fifty six | 6 |
| At Twenty six | 25 | At Sixty six | 3 |
| At Tirty six | 16 | At Seventy six | 1 |
| At Fourty six | 10 | At Eighty | 0 |

11. It follows also, that of all, which have been conceived, there are now alive 40 *per Cent.* above sixteen years old, 25 above twenty six years old, *& sic deinceps*, as in the above Table: there are therefore of Aged between 16, and 56, the nomber of 40, less by six, *viz.* 34; of between 26, and 66, the number of 25 less by three, *viz.* 22: *& sic deniceps*.

Wherefore, supposing there be 199112 *Males*, and the number between 16, and 56, being 34. It follows, there are 34 *per Cent.* of all those *Males* fighting Men in *London*, that is 67694, *viz.* near 70000: the truth whereof I leave to examination, only the ⅕. of 67694, *viz.* 13539. is to be added for *Westminster*, *Step-*

**Figura 3**
**Il frontespizio e una pagina del libro di Graunt pubblicato nel 1662**

«Essendo nato ed essendo stato allevato nella città di Londra, ho sempre osservato che la maggior parte di coloro che regolarmente ricevono i Bollettini settimanali di mortalità non se ne servono in realtà che per guardare l'accrescimento o la diminuzione del numero di sepolture ... in modo da poterne fare oggetto di conversazione nella prima riunione mondana...».

Le riflessioni politico-aritmetiche di Graunt apparivano dunque strane ed audaci ai lettori del tempo; tuttavia si fecero strada, anche se lentamente. Ecco che cosa ne scrive il grande statistico inglese Pearson alla fine dell'Ottocento:

«Là dove nessuno aveva visto a quali fini potevano essere utili i "Bollettini di morte", John Graunt, col suo pratico senso comune, colse la loro importanza e comprese come si doveva estenderli, migliorarli ed analizzarli».

Si può dire dunque che l'opera di Graunt segna un importante passaggio: i metodi tipici delle scienze naturali vengono applicati anche nelle scienze sociali, dando origine alla statistica come scienza.

# La statistica nelle elezioni

## Un primo modo di esaminare i dati elettorali

Quando si organizzano le elezioni, la statistica e le rappresentazioni grafiche sono presenti più del solito nei giornali e nei programmi televisivi.

I mezzi d'informazione presentano in queste occasioni tabelle come quella di fig. 1, relative a recenti elezioni politiche per la Camera dei deputati; accanto ad ogni partito sono riportati i seguenti dati:

- il numero di voti ottenuti, cioè la frequenza assoluta del voto a quel partito;
- le percentuali che esprimono la frequenza relativa del voto ad ogni partito.

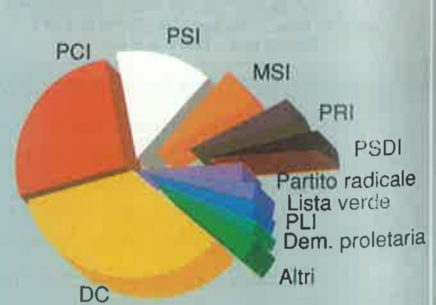
Nella stessa figura la tabella delle frequenze relative è stata visualizzata con il diagramma a settori.

È importante capire subito un fatto: rispetto a quale collettività sono calcolate le percentuali?

**Figura 1**
**I risultati elettorali valutando tutti i voti validi**

| Partiti | Voti | % |
|---|---|---|
| DC | 13 241 188 | 34,3 |
| PCI | 10 254 591 | 26,6 |
| PSI | 5 505 690 | 14,3 |
| MSI | 2 282 256 | 5,9 |
| PRI | 1 429 628 | 3,7 |
| PSDI | 1 140 910 | 2,9 |
| Partito radicale | 988 180 | 2,6 |
| Lista verde | 969 330 | 2,5 |
| PLI | 810 216 | 2,1 |
| Dem. proletaria | 642 161 | 1,7 |
| Altri | 1 328 233 | 3,4 |
| Totale voti validi | 38 592 383 | 100 |

Elezioni politiche del 1987. Fonte: ISTAT




Il cerchio rappresenta tutti i voti validi.



## Come considerare anche i cittadini che non votano

Nella tabella e nel diagramma di fig. 1 le percentuali sono calcolate a partire da tutti i voti validi, ma in ogni elezione si trovano tre situazioni particolari e cioè:

- i cittadini che non vanno a votare;
- le schede bianche, cioè le schede su cui non è scritto nulla;
- le schede nulle, cioè le schede compilate in modo errato.

Per conoscere l'opinione di tutti i cittadini in età di voto, bisogna esaminare più attentamente i dati e compilare una tabella come quella di fig. 2, dove le frequenze relative sono calcolate rispetto alla totalità dei cittadini aventi diritto al voto.

## Esaminare i risultati elettorali da vari punti di vista

Nella fig. 2 si trova anche il diagramma a settori che visualizza la tabella, così si può confrontare questo diagramma con quello relativo alle percentuali di voti ottenuti, dove l'intero cerchio rappresenta i cittadini che hanno espresso il loro voto per un partito (fig. 1).

Questi confronti fanno capire che la stessa situazione può essere descritta in modi diversi; per esempio, per ogni partito si possono considerare:

- il numero di voti ricevuti;
- la percentuale di voti rispetto ai cittadini che hanno votato;
- la percentuale di voti rispetto a tutti i cittadini in età di voto.

E la stessa situazione può apparire anche molto diversa se guardata da punti di vista differenti.

***Spunti di discussione***

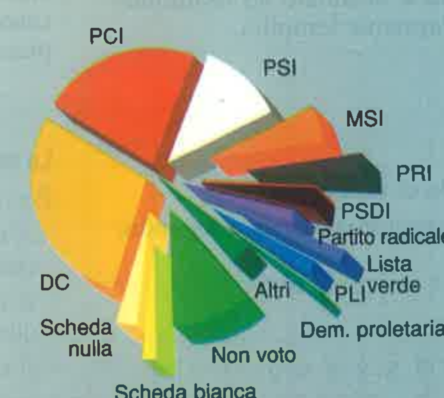
① Cercare i dati relativi alle elezioni più recenti ed esaminarli secondo i punti di vista presentati nella scheda.

② Cercare i dati più recenti relativi alle elezioni degli organi collegiali nella scuola ed esaminarli secondo i punti di vista presentati nella scheda.

**Figura 2**
**I risultati elettorali valutando tutto il corpo elettorale**

| Preferenza | Numero di cittadini | Frequenza |
|---|---|---|
| Non voto | 5 090 339 | 11% |
| Scheda bianca | 774 615 | 2% |
| Scheda nulla | 1 232 492 | 3% |
| DC | 13 241 188 | 29% |
| PCI | 10 254 591 | 22% |
| PSI | 5 505 690 | 12% |
| MSI | 2 282 256 | 5% |
| PRI | 1 429 628 | 3% |
| PSDI | 1 140 910 | 3% |
| Partito radicale | 988 180 | 2% |
| Lista verde | 969 330 | 2% |
| PLI | 810 216 | 2% |
| Dem. proletaria | 642 161 | 1% |
| Altri | 1 328 233 | 3% |
| Totale elettori | 45 689 829 | 100% |

Elezioni politiche del 1987. Fonte: ISTAT




Il cerchio rappresenta tutti i cittadini con diritto di voto.

# 3 La mediana e la moda. Gli istogrammi

Nel terzo quesito del questionario è richiesto di scrivere il voto che ciascun ragazzo ha ottenuto nell'ultimo compito di matematica; ecco, per esempio, i voti rilevati in una classe di 25 alunni:

$9\ 6\ 3\ 4\frac{1}{2}\ 6\ 6\frac{1}{2}\ 6\frac{1}{2}\ 4\ 8\ 4\ 9\ 6\ 4$

$6\ 4\ 5\ 7\ 5\ 7\ 4\ 8\ 4\ 8\ 3\ 6$

I voti sono vari e non è facile esprimere un giudizio complessivo: gli alunni che hanno ottenuto un voto insufficiente diranno che «il compito era difficile ed è andato male», mentre gli altri diranno che «il compito era facile ed è andato bene».
Si può riassumere l'andamento dell'intera classe? La statistica propone vari metodi per riassumere i dati di una collettività con un unico numero; questo paragrafo è destinato ad esaminare due metodi particolarmente semplici.

### La mediana

Ecco un primo modo di procedere:
- si scrivono i voti in ordine crescente, ottenendo:

$3\ 3\ 4\ 4\ 4\ 4\ 4\ 4\ 4\frac{1}{2}\ 5\ 5\ 6\ 6$

$6\ 6\ 6\ 6\frac{1}{2}\ 6\frac{1}{2}\ 7\ 7\ 8\ 8\ 8\ 9\ 9$

- si considera il dato che sta al centro dell'elenco (il 6 in colore);
- si dice che metà classe ha ottenuto un voto inferiore o uguale a 6, l'altra metà ha ottenuto un voto uguale o superiore a 6;
- così si riassumono le abitudini della classe con il numero 6, che divide a metà i dati disposti in ordine crescente e perciò prende il nome di *mediana*.

Queste considerazioni si possono ripetere per riassumere anche altri dati numerici; si procederà nel modo seguente:
- si scrivono i dati in ordine crescente;
- si considera il termine centrale, chiamato *mediana*;
- si dice che metà dei dati non supera la mediana, mentre l'altra metà la supera.

In conclusione, *la mediana è il dato che occupa il posto centrale fra i dati disposti in ordine crescente.*
Se i dati sono in numero pari, si trovano due dati che occupano la posizione centrale; in questo caso, si può considerare come mediana il più piccolo dei due dati centrali.

### La moda

Ecco un secondo modo di sintetizzare dei dati:
- si riportano i dati in una tabella di frequenza come quella di fig. 1*a*;
- si considera il dato che compare con la frequenza massima, cioè 4;
- si conclude che il voto più frequente è 4;
- questo numero 4, a cui corrisponde la frequenza massima, prende il nome di *moda*.

Per ripetere queste considerazioni a partire da altri dati numerici, si procede allora nel modo seguente:
- si scrivono i dati in una tabella di frequenza;



- si considera il dato che si presenta con la massima frequenza; questo dato prende il nome di *moda*.

In conclusione, *la moda è il dato che si presenta con la massima frequenza.*

### Gli istogrammi

La moda diventa particolarmente espressiva se si riportano i dati su un diagramma di frequenza come quello di fig. 1*b*:
- sulla retta orizzontale si riportano i voti in ordine crescente;
- in corrispondenza di ogni dato, si disegna un segmento che ha l'altezza stabilita dalla frequenza.

Il grafico visualizza in particolare la moda: si nota il picco corrispondente al voto 4.

Tuttavia i segmenti sono troppo numerosi e «sparsi» e il grafico è poco espressivo.
Per rendere il diagramma più espressivo si possono riunire i dati in *classi* nel modo seguente:
- nella prima classe si inseriscono tutti i voti compresi fra 2 e 3;
- nella seconda classe i voti compresi fra 3 e 4;
- nella terza classe i voti compresi fra 4 e 5;
- ........
- nell'ultima classe tutti i voti compresi fra 9 e 10.

Ma, per riunire i dati correttamente, bisogna stabilire un criterio chiaro per decidere dove inserire ciascun voto; per esempio il voto 4 va inserito nella seconda o nella terza classe?
Il criterio può essere scelto liberamente, purché sia sempre lo stesso per tutte le classi; le due scelte possibili sono indicate nella tabella A, dove la lettera $v$ indica un voto da inserire.

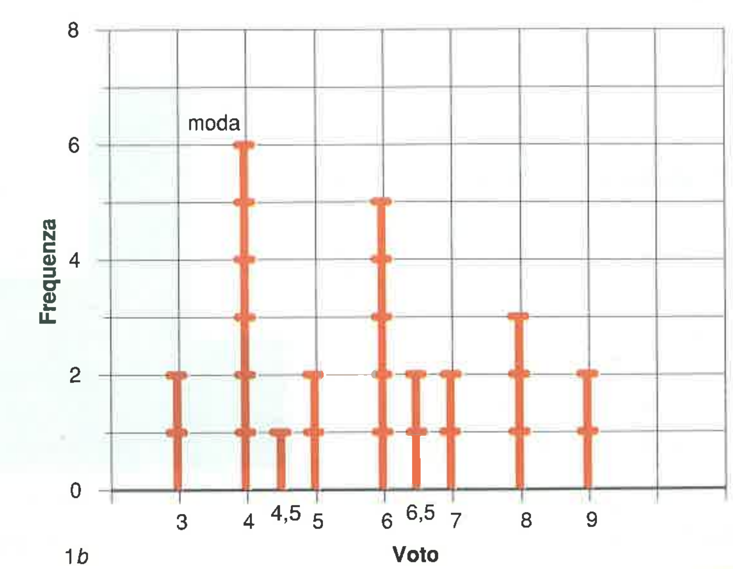
**Tabella A**

| Classe | Primo criterio | Secondo criterio |
|---|---|---|
| Prima | $2<v\leq 3$ | $2\leq v<3$ |
| Seconda | $3<v\leq 4$ | $3\leq v<4$ |
| Terza | $4<v\leq 5$ | $4\leq v<5$ |
| ...... | ...... | ...... |

**Figura 1**
**Tabella di frequenza e diagramma di frequenza dei voti**

| Voto | Frequenza |
|---|---|
| 3 | 2 |
| 4 | 6 |
| 4,5 | 1 |
| 5 | 2 |
| 6 | 5 |
| 6,5 | 2 |
| 7 | 2 |
| 8 | 3 |
| 9 | 2 |

1*a*




1*b*

Basandosi, per esempio, sul primo criterio, si ottengono le classi mostrate nella tabella di fig. 2*a*.
A questo punto si può tracciare il diagramma di frequenza relativo alle classi; si ottiene così un grafico come quello di fig. 2*b*.
Il grafico ottenuto prende il nome di *istogramma*.
Uno dei problemi posti dagli istogrammi è il seguente: come scegliere l'ampiezza delle classi?
Una regola pratica che spesso viene data è la seguente: stabilire l'ampiezza delle classi in modo da avere una decina di colonne.
Il grafico di fig. 2*b*, che è basato su classi ampie 1 voto e presenta 7 colonne, risulta infatti abbastanza espressivo.

## Verifiche

*Conoscenze*

① Come si trova la mediana di più dati?
② Che cosa indica la mediana?
③ Come si trova la moda di più dati?
④ Che cosa indica la moda?
⑤ Come si traccia un istogramma?

*Comprensione*

① La parola «moda» ha un suo significato nel linguaggio comune; si vede un collegamento fra la statistica ed il linguaggio comune?
② Riprendere la fig. 1 del paragrafo 2; si trova anche in quel caso la moda?
③ Riprendere la fig. 1 del paragrafo 2 e spiegare perché in quel caso non è possibile indicare la mediana.

*Applicazioni*

① Ripetere il procedimento indicato nel paragrafo per esaminare i voti ottenuti nell'ultimo compito di italiano.
① Ripetere il procedimento indicato nel paragrafo per esaminare i voti ottenuti nell'ultimo compito di lingua straniera.

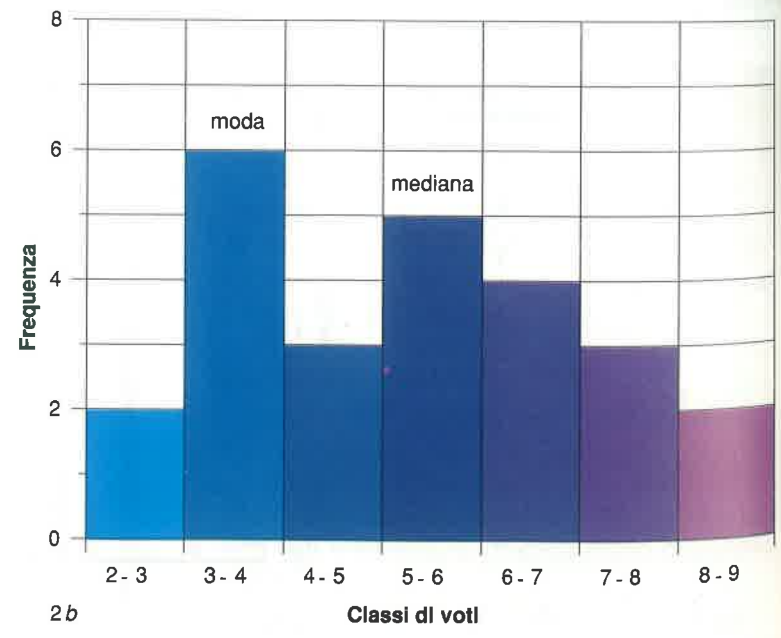
*Vocabolario*

① Cercare su un dizionario l'origine ed il significato dei termini «moda», «mediana», «istogramma».
Se nella classe gli alunni non posseggono tutti lo stesso dizionario, confrontare le risposte date dai vari dizionari.
② Elencare altri significati della parola «mediana». Hanno qualcosa in comune i vari significati di «mediana»?

**Figura 2**
**Istogramma relativo ai voti**

| Classi di voti | Frequenza |
|---|---|
| 2 - 3 | 2 |
| 3 - 4 | 6 |
| 4 - 5 | 3 |
| 5 - 6 | 5 |
| 6 - 7 | 4 |
| 7 - 8 | 3 |
| 8 - 9 | 2 |

2*a*




2*b*



# 4 Un importante valore di sintesi: la media

### Valori di sintesi

In fig. 1 è di nuovo riportata la tabella di frequenza che descrive i voti riportati nel compito di matematica.
Nel paragrafo precedente i dati della tabella sono stati sintetizzati con un solo numero e questa sintesi è stata effettuata in due modi:

1. si è indicato il dato centrale (6), cioè la mediana;
2. si è indicato il dato più frequente (4), cioè la moda.

Moda e mediana prendono anche il nome di *valori di sintesi*, cioè valori che sintetizzano (o riassumono) più dati.
Si ha dunque che: *valore di sintesi di più dati è un numero che sintetizza tutti i dati.*

### La media

Ecco un terzo modo di ragionare per riassumere più dati: si immagina che i 25 ragazzi della classe abbiano tutti lo stesso rendimento; così il voto in matematica deve essere lo stesso per tutti.
È facile trovare questo voto «di classe» uguale per tutti procedendo nel modo seguente:

- si addizionano i voti dell'ultimo compito e si trova la somma S, data da:

$$S=143,5$$

- si «distribuisce egualmente» la somma S fra i 25 ragazzi della classe, dividendola per 25; si trova il numero $m$, dato da:

$$m = \frac{143,5}{25} = 5,74$$

Il numero $m$ così ottenuto prende il nome di *media* dei dati.
Così si può dire che la media ottenuta dalla classe nel compito di matematica è 5,74, cioè un numero vicino a 6.

Queste considerazioni si possono ripetere per sintetizzare anche altri dati numerici: se i dati fossero uguali, avrebbero tutti il valore dato dalla media, che si calcola nel modo seguente:

$$\text{media} = \frac{\text{somma dei dati}}{\text{numero dei dati}}$$

Più brevemente, indicando con S la somma dei dati, con $n$ il numero di dati e con $m$ la media, si può scrivere:

$$m = \frac{S}{n}$$

**Figura 1**
**I voti nel compito di matematica**

| Voto | Frequenza |
|---|---|
| 3 | 2 |
| 4 | 6 |
| 4,5 | 1 |
| 5 | 2 |
| 6 | 5 |
| 6,5 | 2 |
| 7 | 2 |
| 8 | 3 |
| 9 | 2 |

### Moda, mediana e media: i tre valori di sintesi a confronto

In questo paragrafo e nel precedente sono stati dunque descritti tre valori di sintesi:

- la moda, cioè il dato che si presenta con la massima frequenza;
- la mediana, cioè il dato che occupa il posto centrale fra i dati disposti in ordine crescente;
- la media, cioè il valore che avrebbero i dati se fossero tutti uguali.

In fig. 2 è riportato il diagramma di frequenza che visualizza i dati, completato con l'indicazione dei tre valori di sintesi.

Moda, mediana e media hanno dunque lo stesso scopo: sintetizzare con un unico numero una collettività di dati.

Tuttavia questa sintesi viene effettuata da tre punti di vista differenti:

- la moda punta l'attenzione sul gruppo più numeroso;
- la mediana punta l'attenzione sull'ordine crescente dei dati;
- la media elimina la diversità dei dati, trattandoli come se fossero tutti uguali.

E così la statistica porta chiarezza nelle situazioni, ma non per questo riesce ad eliminare le discussioni, dovute all'importanza che ciascuno può dare ai differenti punti di vista.

## Verifiche

*Conoscenze*

① Che cosa significa valore di sintesi?

② Come si calcola la media di più dati?

*Comprensione*

① Spiegare come si possono sintetizzare più dati con la media, la mediana o la moda.

② La media può essere un numero che non fa parte dei dati?

③ La moda può essere un numero che non fa parte dei dati?

④ Commentando i voti riportati in fig. 2 un ragazzo dice: «Il compito doveva essere difficile, visto che tanti alunni hanno avuto 4». A quale valore di sintesi si riferisce il ragazzo?

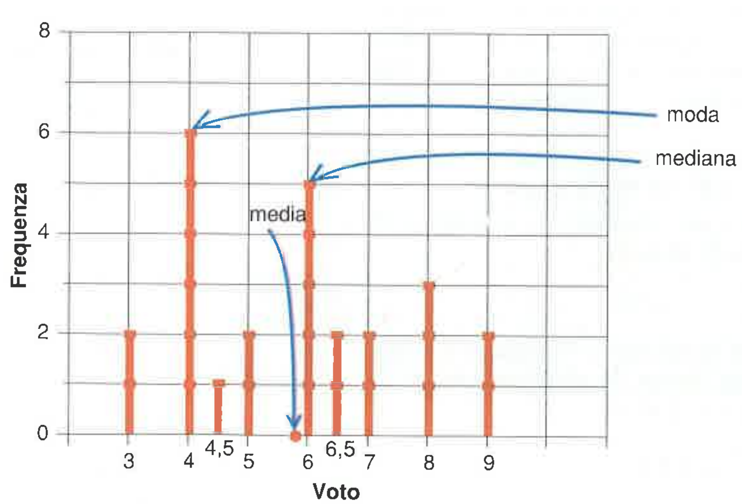
⑤ Commentando gli stessi dati, l'insegnante dice: «Più di metà classe ha ottenuto la sufficienza, perciò il compito non era difficile». A quale valore di sintesi si riferisce l'insegnante?

*Applicazioni*

① Esaminare la situazione della classe nell'ultimo scritto di italiano, basandosi sui tre valori di sintesi.

② Esaminare la situazione della classe nell'ultimo scritto di lingua straniera, basandosi sui tre valori di sintesi.

**Figura 2**
**La rappresentazione grafica dei voti e i tre valori di sintesi**








# Valori di sintesi a confronto. La distribuzione della ricchezza nel mondo

### Numero di abitanti e prodotto nazionale lordo

Per capire come si può esaminare la distribuzione della ricchezza nel mondo si può riprendere una ricerca del 1980 e considerare i dati presentati nella tabella di fig. 1:

- nella prima colonna si trovano alcune aree geografiche;
- nella seconda colonna è indicato il numero di abitanti;
- nella terza colonna si trova il valore di quanto veniva prodotto in un anno (alimenti, macchine, etc.), cioè *il prodotto nazionale lordo*.

Osservando questi dati, si notano subito dei forti squilibri, visualizzati dai diagrammi a settori di fig. 1: ad esempio, l'Europa Occidentale produceva oltre il triplo dell'Asia, pur avendo solo un sesto degli abitanti.

**Figura 1**
**Distribuzione della popolazione e della ricchezza nel mondo**

| Area geografica | Numero di abitanti (in milioni) | Prodotto nazionale lordo (in miliardi di dollari) |
|---|---|---|
| Europa Occidentale | 360 | 3070 |
| Europa Orientale | 110 | 330 |
| Unione Sovietica | 250 | 970 |
| Stati Uniti | 240 | 2580 |
| Giappone | 105 | 1010 |
| America Latina | 280 | 400 |
| Africa | 350 | 180 |
| Asia | 1900 | 970 |
| Australia | 20 | 120 |
| Totale | 3615 | 9630 |

Fonte: da uno studio di Maurice Guernier per il Club di Roma


Il cerchio è la popolazione mondiale.

Il cerchio è il prodotto totale.

## Il reddito annuo per abitante

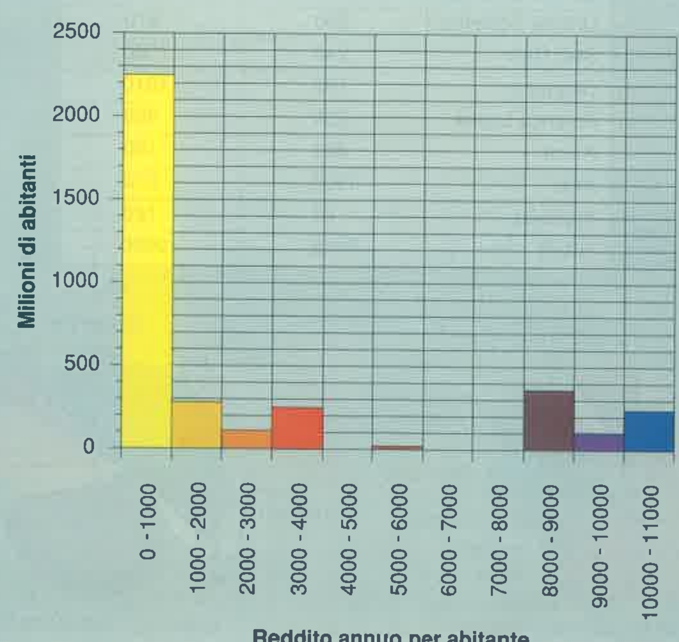
Per rendersi meglio conto degli squilibri esistenti nel mondo conviene dividere il prodotto nazionale lordo di ciascuna area geografica per il numero dei suoi abitanti: si ottiene in questo modo *il reddito per abitante*, riportato nella tabella di fig. 2.

Questi ultimi dati trascurano del tutto le differenze, spesso enormi, fra il livello economico dei singoli abitanti, ma mettono in evidenza la situazione globale di ciascuna zona geografica e rendono gli squilibri ancora più evidenti: il reddito annuo, che era negli Stati Uniti di circa 16 milioni di lire italiane, arriva in Africa e in Asia a sole 750 000 lire italiane.

Davanti ad una tale situazione, la prima domanda che ci si pone è questa: se la ricchezza prodotta nel mondo fosse stata distribuita ugualmente fra tutti gli abitanti del pianeta, quale sarebbe stata la situazione?

**Figura 2**
**Reddito annuo per abitante nel mondo**

| Area geografica | Numero di abitanti (in milioni) | Reddito annuo per abitante (in dollari) |
|---|---|---|
| Europa Occidentale | 360 | 8528 |
| Europa Orientale | 110 | 3000 |
| Unione Sovietica | 250 | 3880 |
| Stati Uniti | 240 | 10750 |
| Giappone | 105 | 9619 |
| America Latina | 280 | 1429 |
| Africa | 350 | 514 |
| Asia | 1900 | 510 |
| Australia | 20 | 6000 |






## I valori di sintesi per esaminare il reddito annuo per abitante

Per rispondere alla precedente domanda bisogna calcolare *la media*, cioè si deve calcolare la ricchezza totale prodotta ogni anno nel mondo e dividere tale valore per il numero di abitanti.

Dalla tabella di fig. 1 si ricava che:

- il prodotto totale mondiale T era 9630 miliardi di dollari, cioè:

$$T=9630\cdot 10^9$$

- la popolazione mondiale P era 3615 milioni di uomini, cioè:

$$P=3615\cdot 10^6$$

- la media M era dunque:

$$M=\frac{9630\cdot 10^9}{3615\cdot 10^6}\cong 2{,}664\cdot 10^3=2664$$

Si trova dunque un reddito annuo medio M, che era di circa 2664 dollari, cioè circa 4 milioni di lire l'anno.

Questo vuol dire che, se tutta la ricchezza fosse stata distribuita in parti uguali, ognuno avrebbe avuto 4 milioni l'anno.

Se, poi, a partire dai dati esposti nella tabella di fig. 2, si determina mediana e moda, si può confrontare questa «equidistribuzione ideale» con la realtà; l'istogramma di fig. 2 visualizza la situazione:

- la moda è 500 dollari l'anno, perché il gruppo più numeroso (i 1900 milioni di abitanti dell'Asia) aveva proprio questo reddito annuo;
- anche la mediana è 500 dollari l'anno, perché i 1900 milioni di abitanti dell'Asia erano più della metà della popolazione mondiale.

La mediana in particolare mette in luce lo stato di povertà del mondo: metà della popolazione mondiale aveva ogni anno non più di 500 dollari per vivere.

Però, valutando solo mediana e moda, resta nascosta la ricchezza delle altre parti del mondo.

È il confronto fra media e mediana che fa emergere con chiarezza i contrasti: la media era 2664 dollari l'anno, mentre più di metà della popolazione mondiale aveva un reddito di soli 500 dollari l'anno; quindi l'altra metà della popolazione deve essere molto ricca per far salire tanto la media.

### *Spunti di discussione*

① Cercare dati recenti relativi alla popolazione ed ai consumi annui di energia elettrica per usi domestici in Italia, divisi per regione. Ripetere le considerazioni svolte in questa scheda per individuare:
- il consumo annuo di energia elettrica per abitante nelle varie regioni;
- moda, mediana e media del consumo di energia elettrica per abitante.

② Collezionare i dati relativi al consumo settimanale di carburanti (gasolio, benzina e miscela) per i veicoli privati usati dagli alunni della classe e dalle loro famiglie. Ripetere le considerazioni svolte in questa scheda per individuare, per ogni carburante:
- il consumo settimanale per persona;
- moda, mediana e media del consumo per persona.

# Attività

## Scoprire come si può mentire con la statistica

### Mentire con i grafici

*Attività 1*

Leggere e discutere la seguente storia di «fantapolitica», che avviene in un paese immaginario.

Il ministro del commercio con l'estero riflette sull'aumento delle esportazioni negli ultimi tre mesi e, valutando le esportazioni in milioni di dollari, trova i seguenti valori: in gennaio 151; in febbraio 159; in marzo 165.

Il giorno seguente il ministro afferma alla televisione: «La nostra economia va molto bene; le nostre esportazioni sono aumentate in febbraio ed hanno avuto un ulteriore aumento in marzo».

Per illustrare i dati, la televisione mostra il grafico di fig. 1.

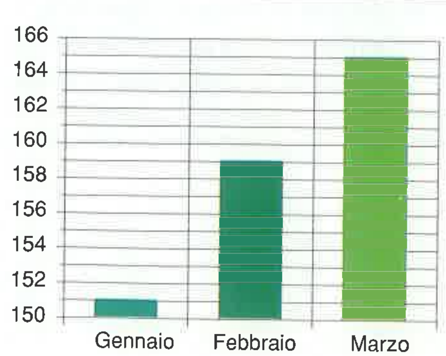
Lo stesso giorno, il ministro del lavoro commenta nel modo seguente il crescente numero di operai in cassa integrazione: «Per quanto riguarda gli operai messi in cassa integrazione in questi ultimi mesi si hanno i seguenti dati in migliaia: in gennaio 151; in febbraio 159; in marzo 165. Il numero di operai in cassa integrazione è dunque aumentato, ma il grafico dimostra che l'aumento del fenomeno non è poi così alto come qualcuno vorrebbe far credere».

Per illustrare i dati, la televisione mostra il grafico di fig. 2.

### Mentire ignorando la frequenza relativa

*Attività 2*

Dalle statistiche delle compagnie di assicurazione risulta che in Italia sono più



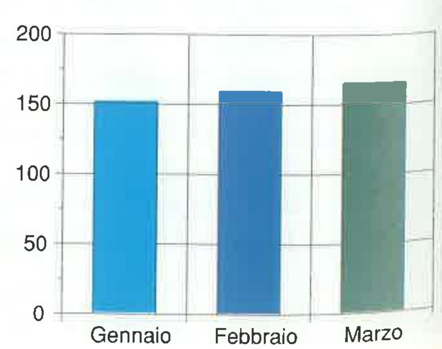




**Figura 1** (a sinistra) **Le esportazioni aumentano molto**

**Figura 2** (a destra) **Gli operai in cassa integrazione aumentano poco**



numerosi gli incidenti provocati da vetture che vanno a velocità moderata rispetto agli incidenti provocati da vetture che viaggiano ad una velocità superiore a 150 km/h.

A partire da queste informazioni, si può concludere che gli italiani guidano meglio a forte velocità?

*Attività 3*

Dalle stesse statistiche risulta che in Italia sono molto più numerosi gli incidenti causati da guidatori che hanno più di venti anni, rispetto agli incidenti provocati da giovani di età da diciotto a venti anni.

A partire da queste informazioni, si può concludere che in Italia i giovani guidano meglio degli adulti con più di venti anni?

### Mentire con i valori di sintesi

*Attività 4*

In fig. 3 sono rappresentati i 25 impiegati di una piccola industria ed il loro stipendio mensile. Esaminare le seguenti frasi che descrivono la situazione economica degli impiegati:

a. «In quell'industria si guadagna bene: lo stipendio medio è di 3,4 milioni al mese»;
b. «Non ti conviene andare a lavorare in quell'industria: i giovani guadagnano appena 1 milione al mese»;
c. «Quell'industria offre buoni stipendi: la metà degli impiegati arriva a guadagnare almeno 3 milioni al mese».

Scoprire in ogni frase qual è il valore di sintesi che descrive la situazione; spiegare perché un solo valore di sintesi descrive la situazione in modo inadeguato.

### Ricercare altre menzogne statistiche

*Attività 5*

Scoprire qualche menzogna statistica in un quotidiano, su una rivista o in un programma televisivo recente.

*Attività 6*

Inventare una storia o dei dati che presentino delle menzogne statistiche.

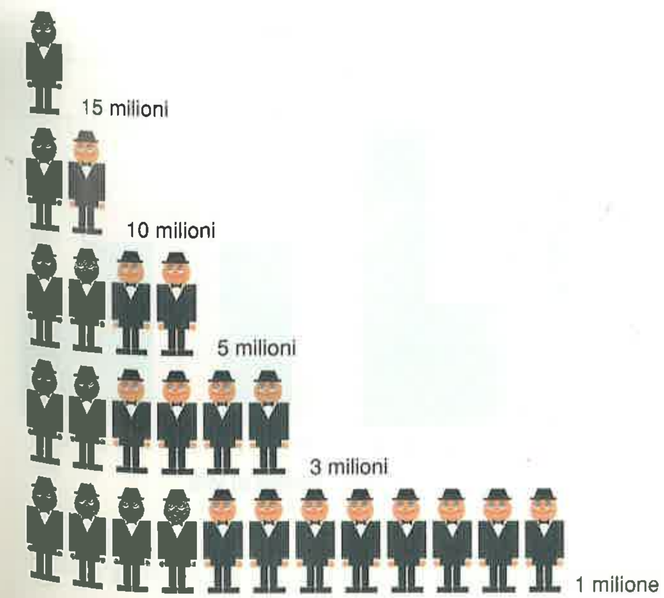



**Figura 3**
**Gli stipendi di una piccola industria**

# 5 La dispersione e gli indici di variabilità

### Un problema che conduce ad osservare la dispersione dei dati

I voti del compito di matematica possono suggerire qualche altra riflessione: i voti delle ragazze e dei ragazzi hanno lo stesso andamento?
Per capire la situazione nella classe di 25 alunni esaminata nei paragrafi precedenti, sono stati separati i voti delle ragazze da quelli dei ragazzi, ottenendo i dati seguenti:

Ragazze: $4 \; 4\frac{1}{2} \; 5 \; 5 \; 6 \; 6 \; 6\frac{1}{2} \; 6\frac{1}{2} \; 7 \; 7$

Ragazzi: 3 3 4 4 4 4 4 6 6 6 8 8 8 9 9

Questi dati possono essere esaminati con i procedimenti mostrati in precedenza; in particolare, si può considerare:

- la rappresentazione grafica con due istogrammi (fig. 1);
- la media, che in entrambi i casi vale circa 5,7;
- la mediana, che in entrambi i casi è 6.

Si osserva subito che, anche se media e mediana sono le stesse, i due istogrammi mostrano una diversa distribuzione dei dati:

- nell'istogramma relativo alle ragazze i dati sono vicini fra loro;
- nell'istogramma relativo ai ragazzi questo non succede e i dati sono più *dispersi*.

Il fatto che i voti delle ragazze siano vicini fra loro ha un significato interessante: vuol dire che le ragazze hanno tutte un rendimento scolastico molto simile.

Invece i voti dei ragazzi molto lontani fra loro segnalano il fatto che ci sono nella classe due gruppi diversi: un gruppo con un rendimento scadente, compensato però da un gruppo con un rendimento molto buono.
Il solo calcolo della media e della mediana trascura completamente questo aspetto della classe; in questo modo si ignora la *dispersione dei dati*.

### Un primo modo di misurare la dispersione: il campo di variazione

Si è così individuato un altro importante aspetto di un insieme di dati: la dispersione, detta anche *variabilità*, che studia quanto i dati variano all'interno di un insieme.
La valutazione di questo aspetto non può essere lasciata all'impressione che si può ricavare da un istogramma.

Per questo la statistica suggerisce vari metodi per esprimere la dispersione per mezzo di un numero.
Fra questi metodi il più semplice è il seguente: calcolare la differenza fra il dato più grande e il dato più piccolo; si ottiene così *il campo di variazione* dell'insieme di dati. Si ha dunque:

campo di variazione=dato massimo–dato minimo

Nel caso dei due insiemi esaminati prima, si ottiene per esempio:

- per i voti dei ragazzi:

  campo di variazione=9–3=6

- per i voti delle ragazze:

  campo di variazione=7–4=3

Così, anche senza osservare gli istogrammi, si trova che i dati dei ragazzi sono più dispersi di quelli delle ragazze.

Tuttavia il campo di variazione non riesce a descrivere come si distribuiscono i dati che si trovano fra il minimo ed il massimo.
In fig. 2 sono rappresentati gli istogrammi relativi ad altri due insiemi di 10 voti che, pur essendo molto diversi dai voti delle ragazze, presentano tutti:

- mediana 6;
- media 5,7;
- campo di variazione 3.

I voti sono i seguenti:

(*a*) 4 4 4 5 6 6 7 7 7 7
(*b*) 4 4 6 6 6 6 6 6 6 7

**Figura 1**
**I voti delle ragazze e dei ragazzi**

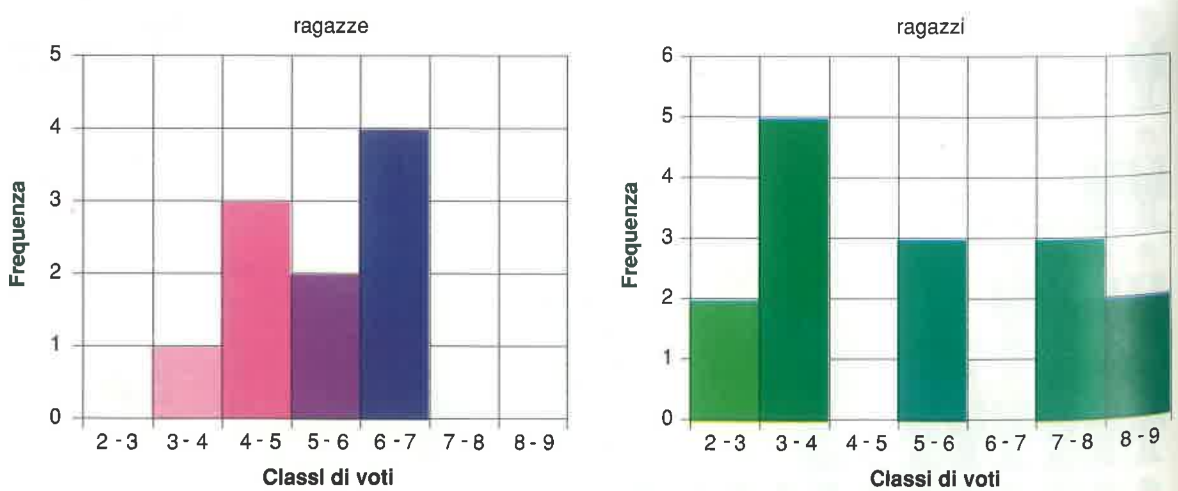
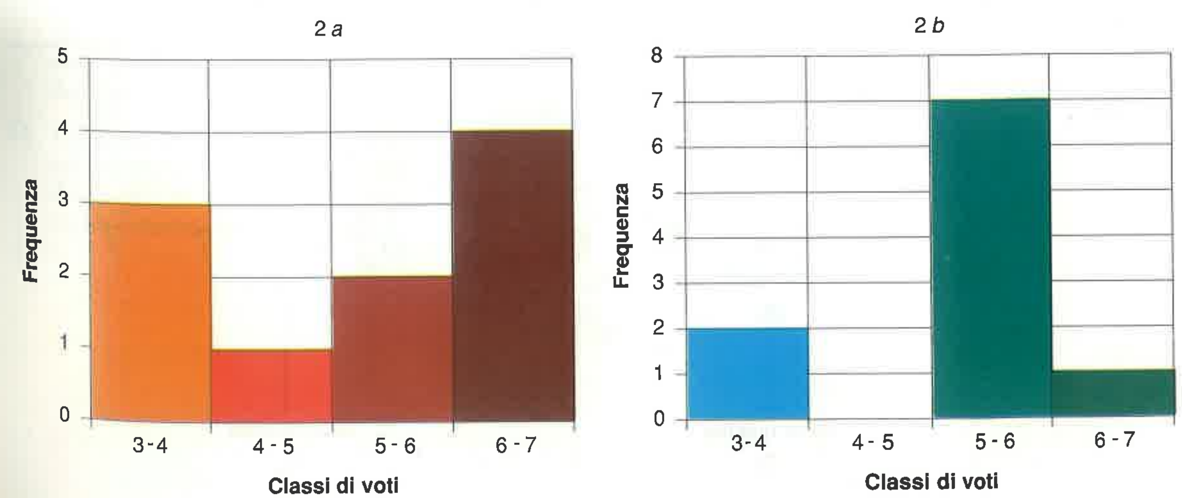



**Figura 2**
**Due insiemi di voti da esaminare**




campo di variazione 3
mediana 6
media 5,7

campo di variazione 3
mediana 6
media 5,7

### Valutare la dispersione intorno alla mediana a partire dai quartili

Una misura della dispersione che elimini l'inconveniente segnalato prima si può ottenere adottando il seguente punto di vista: misurare la dispersione dei dati intorno ad uno dei valori di sintesi.

Si può cominciare col valutare la dispersione intorno alla mediana M.

Per esempio, a partire dai voti delle ragazze, si può ragionare nel modo seguente: la mediana, che è 6, divide i dati in due parti ugualmente numerose, che sono le seguenti:

I. i dati che arrivano fino alla mediana 6, e cioè:

$$4 \quad 4\tfrac{1}{2} \quad 5 \quad 5 \quad 6$$

II. i dati che superano la mediana 6, e cioè:

$$6 \quad 6\tfrac{1}{2} \quad 6\tfrac{1}{2} \quad 7 \quad 7$$

Di ciascuna di queste parti si può di nuovo calcolare la mediana, individuando:

- nel primo gruppo il dato 5 (in colore);
- nel secondo gruppo il dato $6\frac{1}{2}$ (in colore).

In questo modo i dati vengono suddivisi in quattro parti ugualmente numerose, visualizzate in fig. 3; per questo i valori prima individuati prendono i seguenti nomi:

- la mediana dei dati del primo gruppo (5) è detta *primo quartile*, abitualmente contrassegnata dal simbolo $Q_1$;
- la mediana dei dati dell'altro gruppo ($6\frac{1}{2}$) è detta *terzo quartile*, abitualmente contrassegnata dal simbolo $Q_3$.

Nel caso dei dati rappresentati in fig. 3, si trova dunque:

$$M=6 \qquad Q_1=5 \qquad Q_3=6\tfrac{1}{2}$$

I valori $Q_1$ e $Q_3$ così trovati forniscono una chiara indicazione della dispersione: si può infatti dire che la metà centrale dei dati ha un valore compreso fra $Q_1$ e $Q_3$.

Per esprimere quest'idea con un numero, si calcola *la differenza interquartile*, data da:

$$\text{differenza interquartile}=Q_3-Q_1$$

Si osserva subito che la differenza interquartile (abbreviata nel seguito con Q) permette di quantificare la diversità fra i dati di fig. 2 e quelli di fig. 3.
Per i dati di fig. 3 si ha infatti:

$$M=6 \qquad Q_1=5 \qquad Q_3=6\tfrac{1}{2} \qquad \mathbf{Q=1,5} \qquad (1)$$

Per i dati di fig. 2*a*, si ha invece:

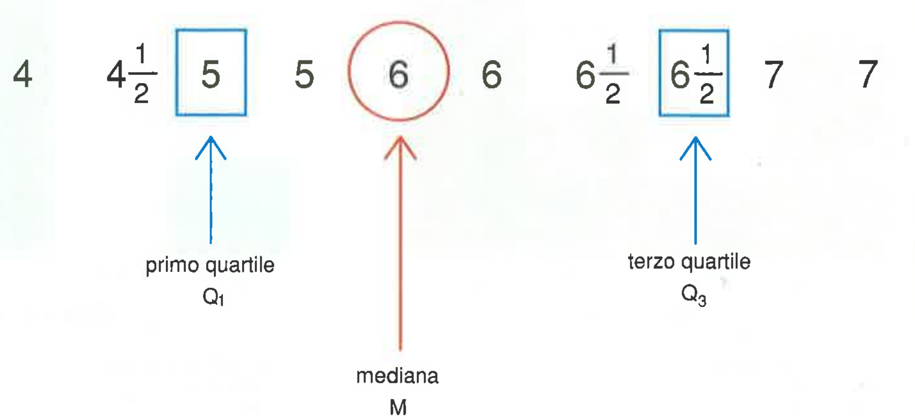
$$M=6 \qquad Q_1=4 \qquad Q_3=7 \qquad \mathbf{Q=3} \qquad (2)$$

E per i dati di fig. 2*b*, si ha infine:

$$M=6 \qquad Q_1=6 \qquad Q_3=6 \qquad \mathbf{Q=0} \qquad (3)$$

Si trova così, senza più pensare all'impressione data dai grafici, che i dati (3) sono i meno dispersi, mentre i dati (2) sono i più dispersi.

**Figura 3**
**I quartili**






Come ultimo esempio, si possono valutare i quartili e la differenza interquartile relativi ai voti dei ragazzi:

$$3 \ 3 \ 4 \ 4 \ 4 \ 4 \ 4 \ \boxed{6} \ 6 \ 6 \ 8 \ 8 \ 8 \ 9 \ 9$$

In questo caso, il numero di dati è dispari e si procede dunque così:

- la mediana M è il dato centrale 6;
- si elimina questo dato centrale per calcolare i quartili, che sono i numeri indicati in colore.

Si trova allora:

$$M=6 \qquad Q_1=4 \qquad Q_3=8 \qquad \mathbf{Q=4} \qquad (4)$$

Si può così concludere che questo insieme di dati è il più disperso di tutti quelli esaminati finora.

### Il campo di variazione e la differenza interquartile sono due indici di variabilità

Le considerazioni ora svolte suggeriscono qualche conclusione di carattere generale.

A. Per riassumere un insieme di più dati si può dare un valore di sintesi come la mediana; questo valore è anche chiamato *indice della posizione centrale*, cioè numero che indica la posizione del «centro» dell'insieme di dati.

B. La mediana da sola non riesce però a riassumere i dati in modo esauriente, perché ignora la dispersione dei dati.

C. La dispersione (o variabilità) dei dati all'interno di un insieme può essere valutata con il *campo di variazione*, che è dunque *un indice di dispersione* (o di variabilità), cioè un numero che misura la dispersione dei dati.

D. Un altro indice di dispersione è la *differenza interquartile*, che è diverso dal precedente perché fissa l'attenzione sulla dispersione intorno alla mediana.

## Verifiche

*Conoscenze*

① Che cosa si intende per dispersione dei dati?

② Dispersione e variabilità hanno un significato diverso?

③ Come si trova il campo di variazione di un insieme di dati?

④ Come si trova la differenza interquartile di un insieme di dati?

*Comprensione*

① Perché il campo di variazione e la differenza interquartile sono detti indici di variabilità?

② Quale aspetto di un insieme di dati viene ignorato indicandone solo la mediana?

③ Si può calcolare la differenza interquartile senza valutare la mediana?

④ Si può calcolare il campo di variazione senza valutare la mediana?

⑤ Quale aspetto di un insieme di dati viene ignorato indicandone solo il campo di variazione?

*Applicazioni*

① Ripetere le considerazioni esposte in questo paragrafo a partire dai voti dell'ultimo compito di italiano.

② Ripetere le considerazioni esposte in questo paragrafo a partire dai voti dell'ultimo compito di lingua straniera.

*Vocabolario*

① Cercare su un dizionario il significato dei termini seguenti:
- variabilità;
- dispersione;
- variazione.

6

# Gli indici di variabilità intorno alla media

### Gli scarti dalla media

Oltre alla mediana, c'è un altro importante indice di posizione centrale: la media.
Anche la media, da sola, ignora la dispersione dei dati e dunque ha bisogno di essere accompagnata da un indice di variabilità.
Per esempio, consideriamo ancora una volta i voti delle ragazze:

$$4 \quad 4\tfrac{1}{2} \quad 5 \quad 5 \quad 6 \quad 6 \quad 6\tfrac{1}{2} \quad 6\tfrac{1}{2} \quad 7 \quad 7$$

La media di questi dati si calcola rapidamente:

$$m=\frac{4+4{,}5+5\cdot 2+6\cdot 2+6{,}5\cdot 2+7\cdot 2}{10}=5{,}75$$

Come valutare la dispersione dei voti intorno alla media?
Un modo semplice di risolvere il problema potrebbe essere organizzato nel modo seguente:

- Valutare la differenza di ogni dato dalla media, differenza che prende il nome di *scarto dalla media*; si ha:

$4-5{,}75=-1{,}75$
$4{,}5-5{,}75=-1{,}25$
$5-5{,}75=-0{,}75$ (2 volte)
$6-5{,}75=0{,}25$ (2 volte)
$6{,}5-5{,}75=0{,}75$ (2 volte)
$7-5{,}75=1{,}25$ (2 volte)

- Calcolare la media di questi scarti, ottenendo l'espressione A.

Si è arrivati dunque ad un risultato molto particolare: il numeratore dell'espressione A vale 0; questo vuol dire che vale 0 la somma degli scarti dalla media.
Questo risultato è un caso legato ai dati esaminati o ha un valore più generale?

### La somma degli scarti dalla media vale sempre zero

C'è un solo modo per capire che cosa si ottiene in generale sommando gli scarti di più dati da un numero fissato: valersi del calcolo letterale.
Si può ragionare nel modo seguente:

- si comincia a considerare un caso semplice, in cui si hanno, per esempio, cinque dati qualunque, indicati con le lettere $a$, $b$, $c$, $d$, $e$;
- si calcolano gli scarti di questi dati da un numero qualunque $p$; si ha:

$$a-p \quad b-p \quad c-p \quad d-p \quad e-p$$

- si sommano questi scarti e si ottiene:

$$a-p+b-p+c-p+d-p+e-p=a+b+c+d+e-5p$$

- si trova che risulta:

$$a+b+c+d+e-5p=0 \quad \text{solo per} \quad p=\frac{a+b+c+d+e}{5}$$

con:

$$p=\frac{a+b+c+d+e}{5}=\text{media dei dati}$$

**Espressione A**

$$M_s=\frac{-1{,}75+(-1{,}25)+2\cdot(-0{,}75)+2\cdot 0{,}25+2\cdot 0{,}75+2\cdot 1{,}25}{10}=\frac{0}{10}=0$$



Questo stesso ragionamento si può ripetere a partire da un numero qualunque di dati e porta dunque a concludere che *la somma degli scarti dalla media vale sempre zero*.
Si tratta di un risultato facile da spiegare: quando si addizionano gli scarti dalla media, gli scarti positivi compensano gli scarti negativi e perciò la somma vale sempre zero.

### Varianza e scarto quadratico medio

Per valutare la dispersione intorno alla media si dovrà dunque eliminare l'inconveniente degli scarti positivi che compensano quelli negativi.
Un metodo che la statistica utilizza molto spesso è il seguente: calcolare la media non più degli scarti, ma dei quadrati degli scarti, quadrati che sono tutti certamente positivi.
Si ottiene, nel caso esaminato, l'espressione B.
Il risultato dell'espressione B prende anche il nome di *varianza*; si ha dunque che *la varianza di più dati si ottiene calcolando la media dei quadrati degli scarti dalla media*.
Per sottolineare la presenza dei quadrati degli scarti, la varianza si indica spesso con il simbolo adottato prima, e cioè:

$$\text{varianza}=s^2$$

Il numero indicato dalla lettera $s$ prende poi il seguente nome:

$$s=\text{scarto quadratico medio}$$

Varianza e scarto quadratico medio sono i più noti e diffusi indici di variabilità intorno alla media.
Così, confrontando ancora una volta i voti dei ragazzi e delle ragazze, si trova:

- voti dei ragazzi:
  media $m=5{,}73$ $s^2\cong 4{,}46$ $s\cong 2{,}11$
- voti delle ragazze:
  media $m=5{,}75$ $s^2\cong 1{,}01$ $s\cong 1{,}01$

e quindi, anche se la media è circa la stessa, si nota subito che i voti delle ragazze sono dispersi intorno alla media meno di quelli dei ragazzi.

### Per sintetizzare più dati occorre il valore di sintesi accompagnato da un indice di variabilità

Le considerazioni svolte in questi ultimi due paragrafi suggeriscono di osservare sempre attentamente i dati statistici che tanto spesso sono presentati dai mezzi di informazione.
Per sintetizzare più dati in modo corretto ed esauriente, occorre fornire un indice di posizione centrale, accompagnato da un indice di variabilità; così si ha che:

- la mediana senza la differenza interquartile dà un'informazione incompleta;
- la media può fornire una sintesi scorretta se non è accompagnata dalla varianza o dallo scarto quadratico medio.

### Verifiche

*Conoscenze*

① Come si calcola la varianza?
② Come si calcola lo scarto quadratico medio?
③ Quanto vale sempre la somma degli scarti dalla media?

*Comprensione*

① Perché la somma degli scarti dalla media vale sempre zero?
② La somma degli scarti dalla mediana ha sempre un valore fisso?

*Applicazioni*

① Ripetere le considerazioni svolte in questo paragrafo, a partire dai voti nel compito scritto di italiano.
② Ripetere le considerazioni svolte in questo paragrafo, a partire dai voti nel compito scritto di lingua straniera.
③ Calcolare per un insieme di dati a scelta la somma degli scarti dalla mediana. Commentare il risultato ottenuto.

**Espressione B**

$$s^2=\frac{(-1{,}75)^2+(-1{,}25)^2+2\cdot(-0{,}75)^2+2\cdot 0{,}25^2+2\cdot 0{,}75^2+2\cdot 1{,}25^2}{10}\cong 1{,}01$$

# Che cosa bisogna sapere

### Rilevazione di dati

*Indagine statistica*: studio di fatti che riguardano una collettività.
*Questionario*: una serie di domande.
*Dato*: informazione su ogni componente della collettività.

### Frequenza

*Frequenza assoluta*: il numero di volte che un fenomeno si presenta.

*Frequenza relativa* = $\dfrac{\text{frequenza assoluta}}{\text{numero di dati esaminati}}$

*Tabella di frequenza*: tabella in cui, accanto ad ogni dato, è riportata la corrispondente frequenza assoluta o relativa.

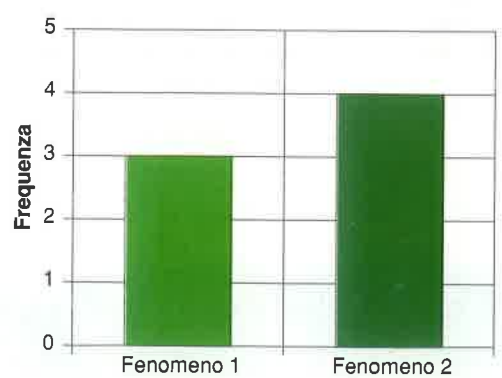
### Rappresentazioni grafiche

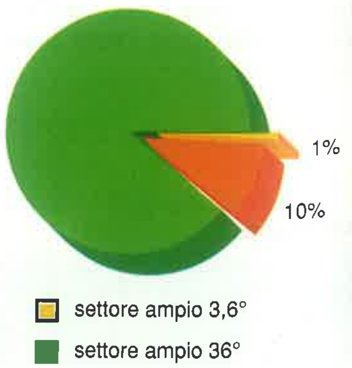
*Diagramma a strisce*: sull'asse orizzontale si indicano i fenomeni osservati e sull'asse verticale la frequenza di ogni fenomeno (fig. 1).

*Diagramma a settori circolari*: un cerchio rappresenta la totalità dei fenomeni osservati; una frequenza relativa espressa da una percentuale del $p\%$ viene rappresentata con un settore ampio $\beta = p(3{,}6)°$ (fig. 2).

**Figura 1** (a sinistra)
**Diagramma a strisce**

**Figura 2** (a destra)
**Diagramma a settori**






*Istogramma*: si suddividono i dati numerici in classi di uguale ampiezza, indicate sull'asse orizzontale; sull'asse verticale si rappresenta la frequenza di ogni classe (fig. 3).

### Valori di sintesi

*Moda*: il dato che si presenta con la massima frequenza.
*Mediana*: il dato centrale fra i dati disposti in ordine crescente.

*Media* = $m = \dfrac{\text{somma dei dati}}{\text{numero dei dati}}$

*Indice della posizione centrale*: numero che indica il centro dell'insieme di dati.

### Indici di variabilità

*Campo di variazione* = dato massimo–dato minimo
*Differenza interquartile* = $Q = Q_3 - Q_1$
con:
$Q_1$ = mediana dei dati fino alla mediana (è il *primo quartile*);
$Q_3$ = mediana dei dati che superano la mediana (è il *terzo quartile*).

*Varianza* = $s^2$ = media dei quadrati degli scarti dalla media
con:
*scarto dalla media* = dato–media

*Scarto quadratico medio* = $s$

### Come rappresentare un insieme di dati con un grafico

- Frequenze relative espresse in forma percentuale si rappresentano efficacemente con un diagramma a settori.
- Frequenze relative o assolute che si ottengono rilevando dati numerici suddivisi in classi di uguale ampiezza si rappresentano con un istogramma.
- Frequenze assolute che si ottengono rilevando opinioni, preferenze o altre caratteristiche non espresse da numeri si possono rappresentare solo con un diagramma a strisce.

### Come sintetizzare un insieme di dati con dei numeri

Si fornisce un valore di sintesi, accompagnato dall'indice di variabilità; per esempio:
- mediana e differenza interquartile;
- media e varianza.

### Proprietà caratteristica della media

La somma degli scarti dalla media vale sempre zero.

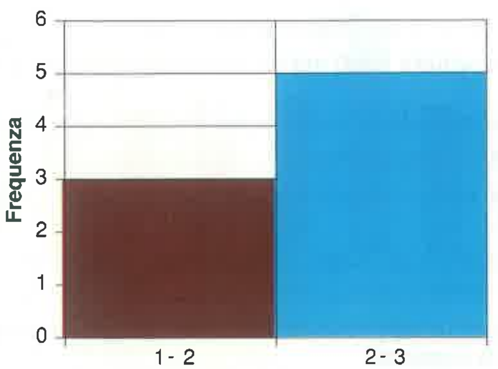



**Figura 3**
**Istogramma**

## Attività finali

# Che cosa bisogna saper fare

### Lavorare con i grafici

*Attività 1*

Ecco i dati relativi al numero di alunni iscritti al primo anno di scuola superiore nell'anno scolastico 1988-89, suddivisi secondo il tipo di scuola.

| Tipo di scuola | Numero di alunni |
|---|---|
| Istituti professionali | 171 385 |
| Istituti tecnici | 330 045 |
| Scuole magistrali | 3 245 |
| Istituti magistrali | 41 585 |
| Licei scientifici | 98 782 |
| Licei classici | 48 728 |
| Licei artistici | 10 021 |
| Istituti d'arte | 18 611 |
| Totale | 722 402 |

Fonte: *Annuario statistico italiano*, ISTAT 1989

- Rappresentare i dati con un diagramma a strisce.
- Si possono rappresentare i dati anche con un diagramma a settori?

*Rispondere alla domanda, seguendo, per esempio, questo schema:*

- *il numero di alunni è una frequenza* ..................................*;*
- *un diagramma a settori rappresenta frequenze* ............*, espresse in forma* ..................................*;*
- *perciò si può disegnare un diagramma a settori solo dopo aver elaborato i dati, in modo da avere* ..........................................................
- Preparare i dati per la rappresentazione con un diagramma a settori, completando la tabella seguente.

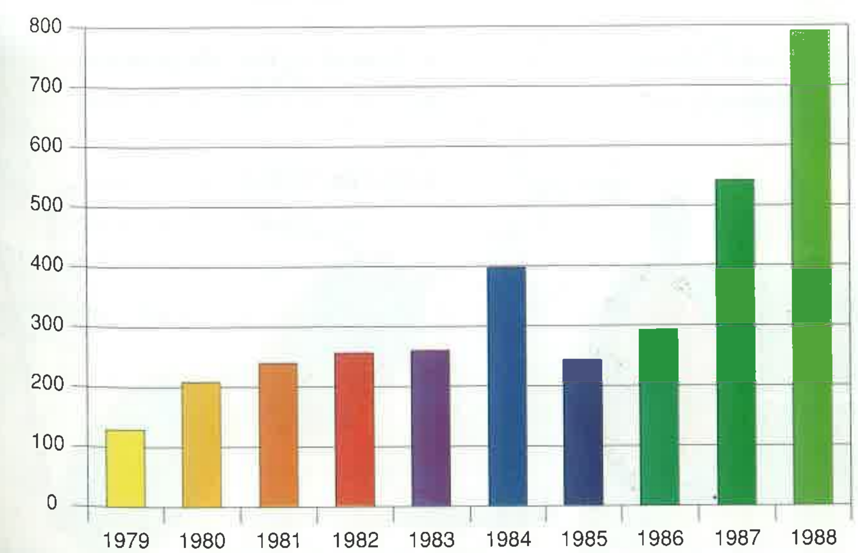


| Tipo di scuola | Frequenza relativa | Percentuale | Ampiezza del settore |
|---|---|---|---|
| Istituti professionali | $\frac{171\,385}{722\,402} \cong 0{,}237$ | 23,7% | $23{,}7 \cdot (3{,}6)° \cong 85°$ |
| Istituti tecnici | | | |
| Scuole magistrali | | | |
| Istituti magistrali | | | |
| Licei scientifici | | | |
| Licei classici | | | |
| Licei artistici | | | |
| Istituti d'arte | | | |
| Totali | 1 | 100% | 360° |

*Attività 2*

Leggere e commentare il grafico di fig. 1, che rappresenta il numero di morti per droga in Italia. Ricostruire, a partire dal grafico, la tabella con i dati numerici.

**Figura 1**
**Decessi per droga in Italia**




Fonte: Ministero dell' Interno

*Attività 3*
Leggere e commentare i grafici di fig. 2, che rappresentano i dati di un'indagine statistica condotta fra 1847 persone in Unione Sovietica. Ricostruire, a partire dai grafici, le tabelle con i dati numerici.

### Riassumere dati con dei numeri

*Attività 4*
Raccogliere i dati relativi alla statura degli alunni della classe.
Elaborare i dati nel modo seguente:

- calcolare moda, mediana e media dei valori ottenuti;
- completare la mediana con la differenza interquartile;
- completare la media con lo scarto quadratico medio.
- rappresentare i dati con un istogramma, scegliendo l'ampiezza delle classi nel modo che sembra più opportuno.

*Attività 5*
La seguente tabella presenta i dati relativi alla statura in centimetri degli iscritti di leva italiani, nati nel 1967.

| <150 | 150-154 | 155-159 | 160-164 | 165-169 | 170-174 | 175-179 | ≥180 | Totale |
|---|---|---|---|---|---|---|---|---|
| 259 | 517 | 4009 | 14 485 | 32 203 | 37 893 | 26 383 | 13 580 | 424 512 |

Rappresentare i dati con un istogramma.
Considerare il valore centrale di ogni classe di statura, completando la seguente tabella:

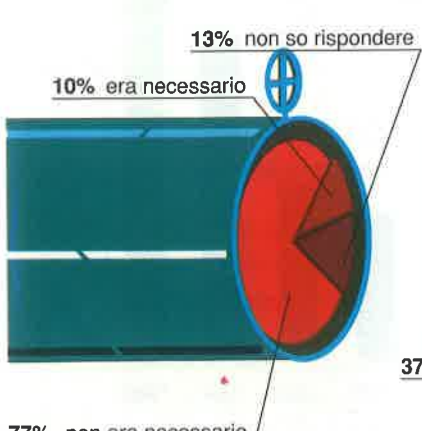
| 148 | 152 | | | | | | 182 | Totale |
|---|---|---|---|---|---|---|---|---|
| 259 | 517 | 4009 | 14 485 | 32 203 | 37 893 | 26 383 | 13 580 | 424 512 |

Elaborare i dati presentati in quest'ultima tabella nel seguente modo:

- calcolare moda, mediana e media dei dati;
- completare la mediana con la differenza interquartile;
- completare la media con lo scarto quadratico medio.

**Figura 2**
**Due grafici da un'indagine statistica in Unione Sovietica**

I. A vostro avviso era necessario che i bolscevichi fucilassero la famiglia dello Zar?




II. Secondo voi, fino a che punto sono responsabili delle sciagure della rivoluzione e della guerra civile:


Fonte: *la Repubblica* 31.10.1990