



“Lego People”

Foto di Joe Shlabotnik

<http://www.flickr.com/photos/joeshlabotnik/305410323/>

Licenza: Attribuzione 2.0 Generico (CC BY 2.0)



A.1 Indagine statistica

Il termine statistica significa *scienza dello stato*. Questo termine venne usato per la prima volta nel XVI secolo per indicare lo studio dei dati utili al governo degli stati prevalentemente relativi a fenomeni di carattere demografico (nascite, morti, ecc.). Negli anni, la statistica si è estesa ai campi più disparati: fisica, psicologia, ricerca di mercato, indici di gradimento, sondaggi, meteorologia, ... È nata essenzialmente con lo scopo di descrivere i fenomeni (statistica descrittiva), successivamente è divenuta uno strumento utile anche per fare previsioni (statistica inferenziale). A grandi linee si può definire come la scienza che si occupa della raccolta e dell'analisi dei dati relativi ad un certo gruppo di persone, animali o oggetti al fine di descrivere in maniera sintetica un fenomeno che li riguarda e fare eventualmente previsioni sul suo andamento futuro.

Ad esempio, la statistica cerca di rispondere a domande del tipo:

- quanta acqua sarà necessaria in Italia fra 3 anni?
- quanta corrente elettrica sarà necessaria per il fabbisogno nazionale fra 5 anni?
- quale sarà il tasso di disoccupazione nazionale fra 1 anno?

Definizione A.1. L'insieme di elementi oggetto dell'indagine statistica è detta *popolazione* o universo, mentre ciascun elemento della popolazione è detto *unità statistica*.

Sono esempi di popolazione statistica gli abitanti di una città in un certo anno, i prezzi di un determinato bene, le temperature massime registrate in una giornata in un particolare luogo, i ciclomotori circolanti in Italia, gli alunni di una scuola.

Definizione A.2. Per ogni unità statistica si possono studiare una o più caratteristiche ed ognuna di tali caratteristiche costituisce un *carattere* della popolazione oggetto di indagine. I caratteri possono essere di tipo *qualitativo* o *quantitativo*. Si definisce *modalità* del carattere indagato ciascuno dei diversi modi in cui esso può presentarsi.

Sono esempi di carattere qualitativo il colore degli occhi, il colore dei capelli, il tipo di scuola frequentato, il gradimento di un certo programma televisivo. Le modalità di un carattere qualitativo sono espresse mediante nomi o aggettivi. I caratteri qualitativi sono a loro volta suddivisi in *ordinabili*, cioè può essere definita una relazione di ordine tra essi (per ogni coppia di elementi si può stabilire quale dei due è il primo e quale il secondo – es. il tipo di scuola frequentato è ordinabile a partire dalla scuola dell'infanzia fino alla laurea, il gradimento di un programma televisivo è ordinabile a partire dalla completa mancanza di gradimento fino al gradimento massimo) e *non ordinabili* o *sconnessi* (es. colore degli occhi, colore dei capelli).

Sono invece caratteri quantitativi l'età, l'altezza, il numero di auto prodotte da una fabbrica, ..., ovvero le modalità di un carattere quantitativo sono espresse mediante numeri. I caratteri quantitativi possono essere di tipo *discreto*, quando assumono solo valori puntuali, oppure di tipo *continuo*, quando possono assumere tutti gli infiniti valori compresi in un determinato intervallo. Sono esempi di caratteri quantitativi discreti il numero di figli in una famiglia, i pezzi prodotti in una catena di montaggio; sono esempi di caratteri quantitativi continui l'altezza di una persona, il peso di una persona, la lunghezza di un fiume.

L'indagine statistica può riguardare l'intera popolazione (in tal caso si parla di *censimento*) oppure solo una sua parte (in tal caso si parla di *indagine a campione*). Supponiamo di voler effettuare un'indagine relativa alle persone che fumano in Italia. Il fenomeno collettivo in esame è il fumo, la popolazione di riferimento è costituita dalla popolazione italiana in età adulta, l'unità statistica è rappresentata da ogni cittadino oggetto dell'indagine, i caratteri oggetto dell'indagine possono essere "fumatore/non fumatore", "numero di sigarette fumate", che cosa si fuma (es. pipa, sigaro, sigaretta). Data l'elevata numerosità della popolazione di riferimento la tipologia di indagine preferibile è quella a campione.

A sua volta, l'indagine a campione può essere effettuata su un *campione casuale*, quando si scelgono a caso i campioni all'interno della popolazione o su un *campione stratificato*, quando si suddivide la popolazione in *classi* o *strati* senza specifici criteri e per ogni strato si prende a caso un campione.

 *Esercizio proposto:* A.1

A.2 Fasi di un'indagine statistica

Definizione A.3. Dato un carattere oggetto di rilevazione, si definisce *frequenza* il numero delle unità statistiche su cui una sua modalità si presenta.

Affinché un'indagine statistica sia rigorosa (e quindi garantisca un'elevata affidabilità) è necessario che sia strutturata secondo le seguenti fasi:

- a) **Studio del problema e impostazione dell'indagine statistica.** Si individua in maniera precisa lo scopo della ricerca, il fenomeno sul quale indagare, la popolazione statistica di riferimento, le singole unità statistiche ed il carattere, o caratteri, oggetto di indagine.
- b) **Rilevazione dei dati statistici.** La rilevazione non è altro che la raccolta dei dati statistici riguardanti ogni elemento della popolazione e relativi al fenomeno che si vuole analizzare. La rilevazione può avvenire secondo diverse modalità:

rilevazione diretta o globale: viene eseguita direttamente su tutte le unità statistiche che formano la popolazione;

rilevazione indiretta o parziale: eseguita solo su una parte della popolazione. Si deve scegliere in tal caso un sottoinsieme della popolazione, detto *campione*, che deve essere rappresentativo della popolazione di riferimento, ovvero deve essere il più possibile eterogeneo rispetto alle caratteristiche della popolazione e contenere al suo interno un numero non troppo ristretto di unità.

- c) **Spoglio delle schede e tabulazione.** Contemporaneamente o successivamente al rilevamento, i dati raccolti vengono ordinati, suddivisi in classi omogenee e riassunti tramite tabelle dette *tabelle statistiche*.

- d) **Rappresentazione dei dati statistici.** La rappresentazione può avvenire attraverso diversi tipi di grafico:
- diagramma cartesiano:* rappresentazione nel piano cartesiano dei valori della variabile sull'asse orizzontale e delle relative frequenze sull'asse verticale;
 - ideogramma:* si rappresenta un certo numero di dati con un simbolo;
 - diagramma a barre o a colonne:* grafico composto da segmenti o barre (orizzontali o verticali) proporzionali alle frequenze;
 - areogramma:* grafico a forma di cerchio composto da settori circolari con aree direttamente proporzionali alle frequenze;
 - istogramma:* grafico composto da rettangoli aventi area proporzionale alla frequenza.
- e) **Elaborazione dei dati.** Con specifici algoritmi di calcolo, vengono elaborati i dati tabulati al fine di costruire opportuni indici di sintesi.
- f) **Interpretazione dei risultati.** Attraverso i grafici e gli indici è possibile descrivere le caratteristiche peculiari del fenomeno analizzato.

Analizziamo in dettaglio le singole fasi che seguono la raccolta dei dati.

A.2.1 Spoglio delle schede e tabulazione

Dopo aver raccolto i dati per ciascuna modalità del carattere o per ciascuna classe individuata si deve determinare:

- ➔ la *frequenza assoluta*, cioè il numero di volte con cui si presenta una modalità del carattere indagato;
- ➔ la *frequenza relativa*, cioè il rapporto tra la frequenza assoluta e il numero totale dei casi presi in esame;
- ➔ la *frequenza percentuale*, cioè la frequenza relativa moltiplicata per 100.

Si compila poi una tabella di frequenza che sintetizza la raccolta dei dati, come nell'esempio seguente.

Esempio A.1. La tabella seguente fornisce la distribuzione di frequenze assolute degli alunni di una classe rispetto al carattere sesso.

Sesso	Femmine	Maschi	Totale
Numero di alunni	15	12	27

Per costruirla, si è operata la classificazione della popolazione degli alunni della classe rispetto ad un determinato carattere (il sesso), sono state individuate le modalità con cui questo si è manifestato (femmina, maschio) ed è stato effettuato il conteggio delle unità in corrispondenza di ciascuna modalità (frequenza assoluta). Dalle frequenze assolute si ricavano le frequenze relative: 15 alunni su 27 sono femmine: la frazione è di $15/27$ femmine sul totale degli alunni. Quindi Dall'operazione $15 \div 27$ otteniamo 0,56 (approssimando a due cifre decimali) che è la frequenza relativa. La frazione può essere espressa in forma percentuale: 0,56 equivale a dire 56 su 100 ed è consuetudine scriverlo in forma percentuale 56%. Tale valore è la frequenza percentuale.

Ripetendo lo stesso procedimento per i maschi si ottiene la seguente tabella delle frequenze:

Sesso	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
Femmine	15	0,56	56%
Maschi	12	0,44	44%

Si può concludere che la classe è formata per il 56% da femmine e per il 44% da maschi.

Esempio A.2. Supponiamo che i voti elencati di seguito siano quelli riportati in matematica a fine trimestre dagli alunni della tua classe: 5, 4, 6, 8, 8, 7, 7, 6, 5, 5, 6, 7.

Per poter effettuare una lettura più agevole si costruisce una tabella in cui vengono riportati sulla prima colonna i singoli valori rilevati (le modalità del carattere) in ordine crescente, nella seconda la frequenza assoluta, cioè quante volte compare quel determinato voto, nella terza la frequenza relativa e nella quarta quella percentuale (che si ottiene moltiplicando per 100 la frequenza relativa):

Voto riportato	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
4	1	$1/12 = 0,083$	8,30%
5	3	$3/12 = 0,25$	25,00%
6	3	$3/12 = 0,25$	25,00%
7	3	$3/12 = 0,25$	25,00%
8	2	$2/12 = 0,167$	16,70%
Totale	12	$12/12 = 1$	100%

Esempio A.3. Misurando l'altezza di un gruppo di cani di razza pastore italiano si sono ottenute le seguenti misure in cm:

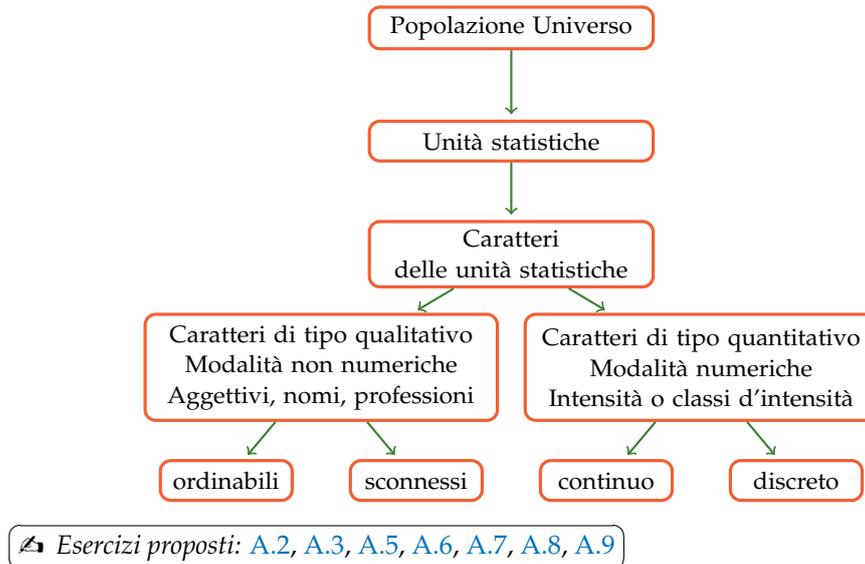
57,1 60,8 60,7 56,2 59,5 62,4 56,1 61,2 54,5 64,5 57,5 58,3 55,2
 58,7 57,2 56,1 58,9 57,7 53,2 59,2 58,9 54,5 55,3 62,1 59,0 58,3
 61,3 60,1 56,4 60,2 61,7 57,3 58,3 59,5 62,6 59,4 58,3 59,4 59,4
 59,3 57,6 60,0 60,7 56,7 61,1 59,8 55,3 63,9 58,0 55,2 54,9 53,8

Il carattere indagato nella popolazione cani pastore italiano è di tipo quantitativo continuo; con questo tipo di dati è praticamente impossibile calcolare le frequenze se le altezze non si raggruppano in classi.

Vediamo come procedere: osservando i dati ottenuti si nota che il valore minore è 53,8 mentre il valore maggiore è 64,7. Possiamo allora suddividere i dati in gruppi partendo da 53,0 cm fino a 65,0 cm, formando classi di ampiezza 1cm e ottenendo la seguente tabella:

Classe (cm)	Frequenza assoluta	Frequenza percent.	Classe (cm)	Frequenza assoluta	Frequenza percent.
53,0 - 53,9	2	3,85%	59,0 - 59,9	9	17,31%
54,0 - 54,9	3	5,77%	60,0 - 60,9	6	11,54%
55,0 - 55,9	4	7,69%	61,0 - 61,9	4	7,69%
56,0 - 56,9	5	9,61%	62,0 - 62,9	3	5,77%
57,0 - 57,9	6	11,54%	63,0 - 63,9	1	1,92%
58,0 - 58,9	8	15,38%	64,0 - 64,9	1	1,92%

Riassumendo



A.2.2 Rappresentazione grafica

La rappresentazione grafica dei dati statistici facilita notevolmente lo studio delle caratteristiche del fenomeno che si sta esaminando; infatti dopo aver impostato l'indagine, raccolto, classificato ed elaborato i dati nelle tabelle, i dati non sempre si presentano in una forma di facile lettura ed il loro significato e la loro interpretazione rimane poco chiara. Attraverso la rappresentazione grafica, i risultati dell'indagine emergono immediatamente, in maniera diretta e sintetica.

La rappresentazione grafica può avvenire utilizzando diversi tipi di grafico a seconda delle caratteristiche da analizzare.

Diagramma cartesiano

La rappresentazione grafica attraverso un diagramma cartesiano dà, in modo immediato, informazioni sull'andamento globale del fenomeno e viene utilizzata prevalentemente per la rappresentazione di serie storiche (per esempio, per rappresentare il numero di auto prodotte per anno da una fabbrica) oppure quando si hanno due caratteri quantitativi e si vuol analizzare il tipo di legame esistente fra di essi.

Esempio A.4. Consideriamo la tabella statistica relativa alla domanda "quante ore al giorno passi al computer?", posta ad un campione di 50 ragazzi dai 16 ai 24 anni.

Rappresentiamo la tabella attraverso un diagramma cartesiano costruito tracciando due rette perpendicolari, gli assi, quello verticale orientato verso l'alto e quello orizzontale orientato verso destra. Riportiamo sull'asse orizzontale il numero di ore e sull'asse verticale il numero di ragazzi e determiniamo i punti aventi come coordinate (numero ore; numero ragazzi).

Il punto A avrà come coordinate (0;4), il punto B avrà come coordinate (1;6) e così via. Uniamo poi i punti con segmenti e otteniamo il diagramma cartesiano (grafico [A.1](#)). Precisamente A(0;4), B(1;6), C(2;12), D(3;16), E(4;8), F(5;4), G(6;2).

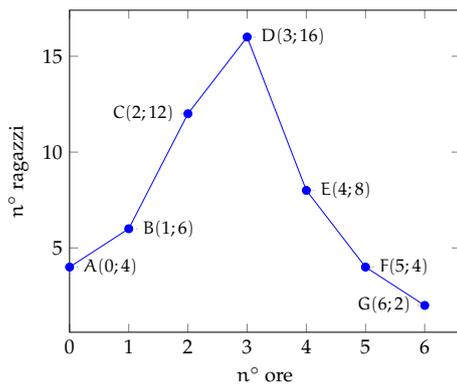


Grafico A.1: Esempio 24.4

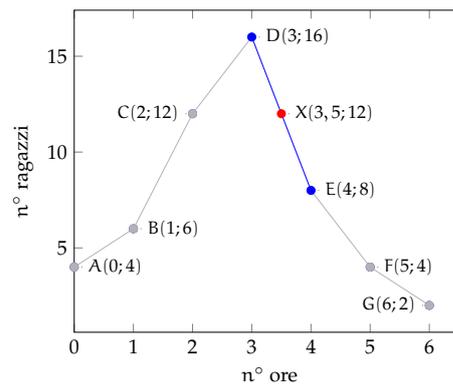


Grafico A.2: Esempio 24.4

Numero di ore	0	1	2	3	4	5	6
Numero di ragazzi	4	6	12	16	8	4	2

Dal grafico A.2 si può notare immediatamente che la maggior parte dei ragazzi trascorre dalle 2 alle 3 ore al computer dato che il picco più alto si ha proprio nei punti C e D. Si può anche notare che, ad esempio, il punto X di coordinate (3,5; 12), appartenente al segmento di congiunzione tra i punti D ed E, non ha significato reale, dato che le sue coordinate non sono riportate nella tabella statistica del fenomeno da studiare.

Ideogramma

Nella rappresentazione grafica attraverso *ideogramma* si rappresenta un certo numero di dati con un simbolo che si assume come *unità grafica*; il simbolo deve richiamare l'oggetto dell'indagine e dare quindi una visione immediata del fenomeno. Ad esempio si può utilizzare un uomo stilizzato per rappresentare un dato riguardante il numero di persone che vivono in un determinato territorio, una macchina per la produzione annua di automobili in una fabbrica, e così via. Tale tipo di rappresentazione è spesso usata in campo pubblicitario perché caratterizzata da un evidente impatto visivo.

Esempio A.5. Un istituto scolastico ha visto aumentare i suoi iscritti, dall'anno scolastico 2003-2004 all'anno 2008-2009 secondo quanto riportato nella seguente tabella:

Anno scolastico	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09
Iscritti	150	200	200	325	375	450

Possiamo rappresentare mediante ideogramma i dati contenuti nella tabella statistica. Consideriamo una faccina stilizzata come unità grafica assegnandole il valore di 50 ragazzi iscritti.

☺ = 50 iscritti

Il numero degli iscritti di ogni anno scolastico sarà rappresentato da tante unità grafiche quanti sono i gruppi di 50 iscritti. Per avere il grafico relativo all'anno 2003-2004 si devono usare tre faccine, in quanto $150 : 50 = 3$.

$$\text{a.s. 2003-2004} = \text{😊😊😊}$$

Se la divisione del numero degli iscritti per 50 dà resto, esso si dovrà rappresentare disegnando solo una parte dell'unità grafica, corrispondente alla frazione tra resto e 50. Ad esempio nell' a.s. 2006-2007 ci sono stati 325 iscritti; $325 : 50 = 6$ col resto di 25, quindi 325 sarà uguale a 6 unità grafiche e $\frac{25}{50} = \frac{1}{2}$ unità grafica, cioè mezza faccina, ovvero $325 : 50 = 6,5$ cioè 6 faccine e mezzo.

$$\text{a.s. 2006-2007} = \text{😊😊😊😊😊😊😊}$$

Il grafico completo sarà:

$$\begin{array}{ll} \text{a.s. 2003-2004} = \text{😊😊😊} & 3 \\ \text{a.s. 2004-2005} = \text{😊😊😊😊} & 4 \\ \text{a.s. 2005-2006} = \text{😊😊😊😊} & 4 \\ \text{a.s. 2006-2007} = \text{😊😊😊😊😊😊😊} & 6 \text{ e } 1/2 \\ \text{a.s. 2007-2008} = \text{😊😊😊😊😊😊😊😊} & 7 \text{ e } 1/2 \\ \text{a.s. 2008-2009} = \text{😊😊😊😊😊😊😊😊😊} & 9 \end{array}$$

Diagramma a barre o a colonne

Questo tipo di rappresentazione, detta anche diagramma a nastri o a bastoni, viene usata quando si vuole fornire un'idea delle frequenze delle diverse modalità di un fenomeno. In genere si usa per caratteri qualitativi o quantitativi discreti. Per poter valutare il significato statistico della lunghezza delle barre (o delle colonne) è necessario scegliere opportunamente una scala di riferimento: la larghezza della barra (o della colonna) è arbitraria ma uguale per tutte le barre (o colonne) e la sua lunghezza è proporzionale alla caratteristica che si deve rappresentare. Le barre (o le colonne) possono inoltre essere suddivise in parti di colori diversi per indicare le singole componenti o i singoli fenomeni che si vogliono analizzare.

La differenza fra la rappresentazione a barre e quella a colonne consiste soltanto nell'orientamento del grafico: nel diagramma a barre si indicano le modalità del carattere sull'asse verticale e le frequenze sull'asse orizzontale, mentre in quello a colonne le modalità del carattere sono riportate sull'asse orizzontale e le frequenze su quello verticale.

Di seguito vengono riportate le due tipologie di grafico accompagnate dalla tabella di riferimento:

Materia	Italiano	Storia	Geografia	Matem.	Scienze	Ed. Fisica	Totale
Maschi	5	4	4	2	6	5	26
Femmine	3	7	2	3	4	5	24

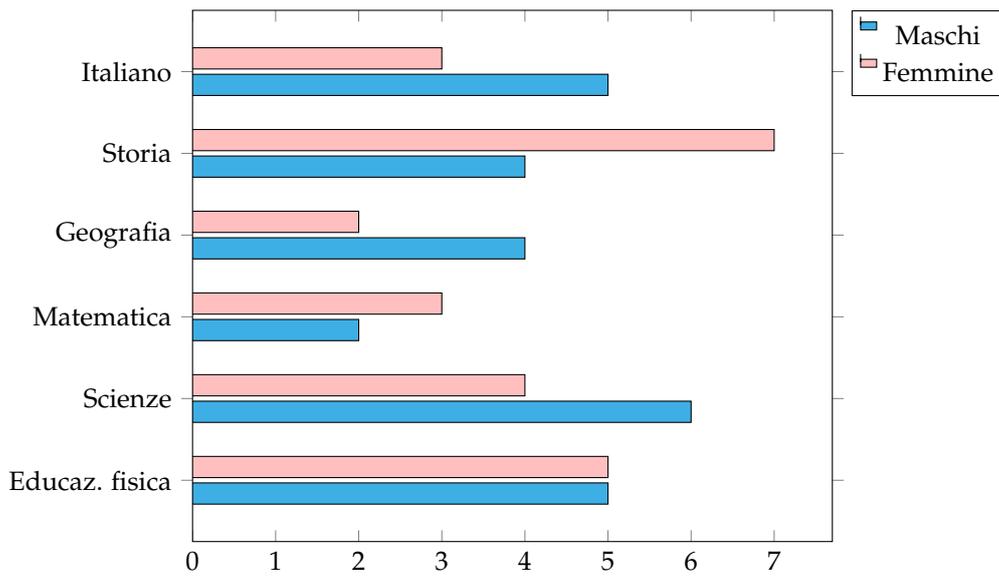


Figura A.1: Diagramma a barre

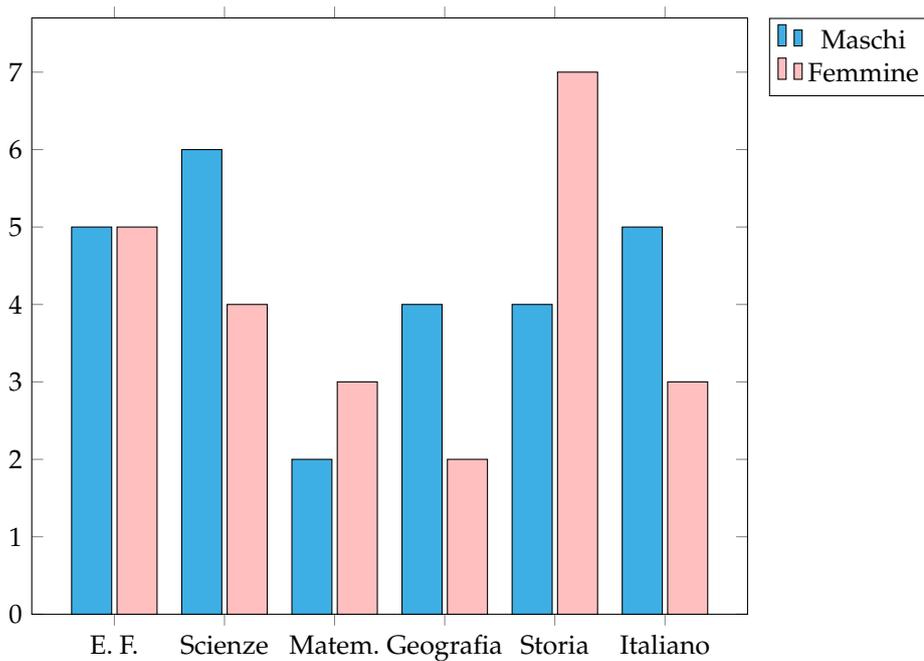


Figura A.2: Diagramma a colonne

Areogramma

Questo tipo di rappresentazione, detta anche grafico a torta, viene utilizzato quando si vogliono evidenziare le parti che compongono un fenomeno, per esempio per indicare come

si dividono gli alunni di una classe in maschi e femmine, o per rappresentare in che modo le varie voci di spesa incidono sul bilancio familiare. Il grafico si ottiene dividendo un cerchio in settori circolari con aree direttamente proporzionali alle frequenze che rappresentano. Per disegnare l'areogramma, si disegna una circonferenza di diametro arbitrario e si fa corrispondere l'angolo al centro di 360° , con il 100% di frequenza percentuale; per ottenere l'angolo corrispondente ad una certa frequenza percentuale f_x si risolve la proporzione $360^\circ : X^\circ = 100 : f_x \Rightarrow X^\circ = 3,6 \cdot f_x$. Si suddivide così la circonferenza negli angoli ottenuti e si evidenziano in maniera differente tra loro i settori circolari ottenuti.

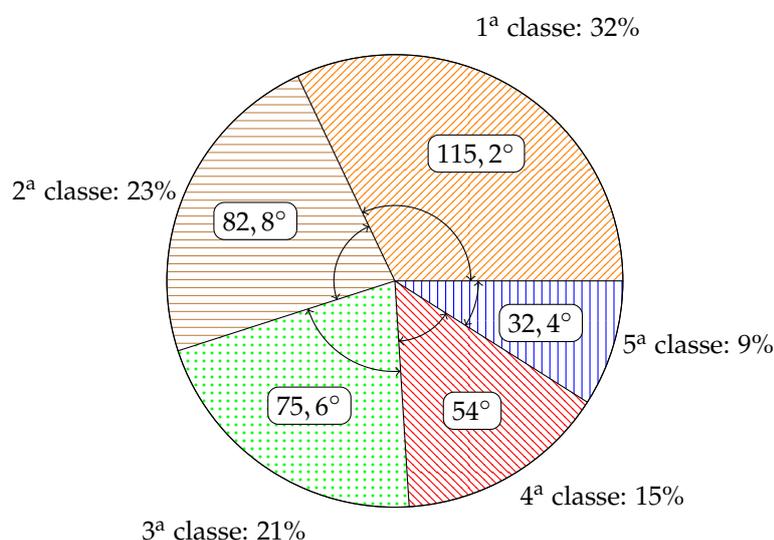
Esempio A.6. Consideriamo la seguente tabella statistica che indica gli studenti di un dato istituto scolastico divisi per classe frequentata, in un dato anno.

Classe	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	Totale
Studenti	320	230	212	152	96	1010

Nella tabella sono indicate le frequenze assolute; calcoliamo ora le frequenze percentuali degli studenti. Per la 1^a classe si ha: $\frac{320}{1010} = 0,32$ arrotondato alla seconda cifra decimale, che equivale al 32% e così via per le classi successive.

Classe	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	Totale
Frequenze percentuali	32%	23%	21%	15%	9%	100%

Rappresentiamo graficamente mediante areogramma i dati contenuti nella tabella precedente.



Per ottenere l'angolo relativo alla frequenza percentuale della 1^a classe si fa: $3,6 \cdot 32 = 115,2^\circ$ e per la 2^a classe: $3,6 \cdot 23 = 82,2^\circ$ e così via per le altre classi.

Dal grafico si può notare immediatamente che la classe più frequentata è la prima.

Istogramma

Si utilizza la rappresentazione grafica attraverso istogramma quando il carattere analizzato è di tipo quantitativo ed i dati sono raggruppati in classi.

Prima di tutto si distribuiscono i dati in classi o gruppi e si determina il numero di unità appartenenti a ciascuna classe; questo numero è detto *frequenza della classe*. Riportando tali dati in una tabella si ottiene la distribuzione delle frequenze. Poiché le classi potrebbero avere ampiezze diverse si calcola la *densità di frequenza*, definita come il rapporto fra la frequenza della classe e la relativa ampiezza.

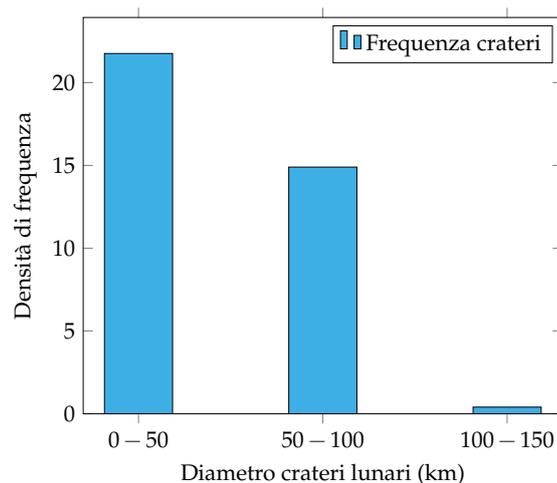
Per disegnare un istogramma si tracciano due assi; sull'asse verticale, orientato verso l'alto, si fissa un segmento unitario e vi si riportano le densità di frequenza. L'asse orizzontale, orientato verso destra, è invece suddiviso in tanti segmenti la cui ampiezza è pari a quella delle singole classi. Il grafico consiste in un insieme di rettangoli aventi per base ogni classe e altezza la densità di frequenza corrispondente. In tal modo l'area di ogni rettangolo rappresenta la frequenza corrispondente a ciascuna classe.

Esempio A.7. Costruiamo un istogramma a partire dalla distribuzione di frequenza riportata nella seguente tabella:

Diametro crateri lunari (km)	Numero di crateri
0 – 50	1 088
50 – 100	745
100 – 150	20

Innanzitutto per ogni classe dobbiamo determinare la densità di frequenza, che si ottiene dividendo la frequenza assoluta per l'ampiezza della classe:

Diametro crateri lunari (km)	Densità di freq.
0 – 50	$1\,088/50 = 21,76$
50 – 100	$745/50 = 14,9$
100 – 150	$20/50 = 0,4$



Esempio A.8. Consideriamo la seguente tabella statistica che riporta i giorni di pioggia di ogni mese, in un dato anno e in una data città.

Mesi	Giorni di pioggia	Mesi	Giorni di pioggia
Gennaio	15	Luglio	1
Febbraio	10	Agosto	3
Marzo	14	Settembre	3
Aprile	8	Ottobre	5
Maggio	5	Novembre	9
Giugno	2	Dicembre	11

Dividiamo i mesi dell'anno in classi, raggruppandoli in stagioni. Ad esempio, Luglio, Agosto e Settembre appartengono alla classe dell'Estate e la frequenza di questa classe è data dalla somma delle frequenze di ogni mese, cioè $1 + 3 + 3 = 7$. Si prosegue in questo modo per ogni classe ottenendo così la distribuzione delle frequenze riportata nella tabella.

Stagioni	Estate	Autunno	Inverno	Primavera
Giorni di pioggia	7	25	39	15

Costruisci ora l'istogramma corrispondente alla tabella precedente riportando sull'asse orizzontale le classi (stagioni) e su quello verticale le densità di frequenze.

🔗 *Esercizi proposti:* [A.10](#), [A.11](#), [A.12](#), [A.13](#), [A.14](#), [A.15](#), [A.16](#), [A.17](#), [A.18](#), [A.19](#), [A.20](#)

A.3 Indici di posizione

Nel caso in cui il carattere considerato nell'indagine sia di tipo quantitativo, l'andamento dei dati raccolti può essere sinteticamente descritto per mezzo di opportuni indici. Gli *indici di posizione* vengono utilizzati per dare un'indicazione sulla distribuzione delle frequenze per mezzo di un solo numero. A seconda del carattere oggetto dell'indagine statistica possono essere utilizzati indici differenti.

A.3.1 Moda

Definizione A.4. La *moda* è la modalità del carattere indagato che si presenta più frequentemente.

In una successione di n modalità x_1, x_2, \dots, x_n con le relative frequenze f_1, f_2, \dots, f_n , la moda è la modalità che ha la frequenza maggiore. Questo valore può essere calcolato per qualunque tipo di carattere, sia qualitativo che quantitativo. Se il carattere è quantitativo continuo con dati raggruppati in classi non è possibile determinare con esattezza la moda, ci si limita ad individuare la *classe modale* definita come la classe cui è associata la massima densità di frequenza.

Esempio A.9. Nella tabella seguente sono riportati i numeri degli studenti, divisi per classe, della sezione A di un dato istituto, in un dato anno. Si può osservare che la 1^a classe presenta la frequenza massima di 320 studenti, quindi la moda è la classe prima.

Classe	1°	2°	3°	4°	5°	Totale
Studenti	320	230	212	152	96	1010

Esempio A.10. La tabella raccoglie i dati relativi alla domanda “quante ore alla settimana pratici sport?”, posta ad un campione di 50 ragazzi dai 18 ai 25 anni. Si può osservare che 12 e 18 ore presentano la frequenza massima 14, quindi si hanno due mode 12 ore e 18 ore. In questo caso la distribuzione è bimodale.

Numero di ore	0	4	8	12	16	18	22	Totale
Numero di ragazzi	4	1	3	14	8	14	6	50

Esempio A.11. La tabella seguente è relativa alla distribuzione delle classi di altezza di un gruppo di studenti.

Altezza (cm)	160-165	165-170	170-175	175-185	185-200	Totale
Numero di studenti	5	8	15	10	2	40

Poiché le classi hanno ampiezza diversa è necessario calcolare la densità di frequenza.

Altezza	160-165	165-170	170-175	175-185	185-200
Densità di frequenza	1	1,6	3	1	0,13

La massima densità di frequenza si ha in corrispondenza della classe 170-175, essa rappresenta quindi la classe modale.

A.3.2 Media aritmetica

Definizione A.5. La *media aritmetica* (semplice) è il valore ottenuto sommando tutti i dati e dividendo tale somma per il numero dei dati.

Se abbiamo n dati x_1, x_2, \dots, x_n , la media aritmetica semplice M è data da:

$$M = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Esempio A.12. Riprendiamo in esame la tabella relativa agli studenti, divisi per classe frequentata di un dato istituto scolastico, in un dato anno e calcoliamone la media aritmetica semplice.

Classe	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	Totale
Studenti	320	230	212	152	96	1010

Per calcolare la media aritmetica semplice degli studenti, sommiamo tutti gli studenti delle cinque classi e dividiamo tale somma per il numero delle classi:

$$M = \frac{320 + 230 + 212 + 152 + 96}{5} = \frac{1010}{5} = 202.$$

Possiamo dire che si hanno mediamente 202 studenti per ogni classe.

Definizione A.6. Si definisce *scarto dalla media* (aritmetica) la differenza tra i valori osservati e la media.

Se x_1, x_2, \dots, x_n sono i valori osservati e M la loro media aritmetica, gli scarti sono $s_1 = x_1 - M, s_2 = x_2 - M, \dots, s_n = x_n - M$.

Esempio A.13. Calcoliamo gli scarti dalla media per la distribuzione “studenti per tipologia di classe frequentata”, la cui media è $1010/5 = 202$.

Classe	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	Totale
Studenti	320	230	212	152	96	1010
Scarto	118	28	10	-50	106	0

Si può osservare che vi sono solo valori superiori alla media e altri inferiori, tanto che lo scarto è rappresentato in alcuni casi da un numero positivo, in altri da un numero negativo. Si può verificare che la somma degli scarti dalla media è nulla, cioè gli scarti positivi compensano sempre quelli negativi.

Definizione A.7. La *media aritmetica ponderata* è il valore ottenuto moltiplicando ciascuna modalità del carattere dato con la propria frequenza, sommando tutti i prodotti fra loro e dividendo poi per la somma delle frequenze (che equivale al numero totale n delle unità statistiche considerate).

La media aritmetica ponderata si usa nel caso in cui le unità statistiche sono molte ed è già stata fatta la tabella delle frequenze. Avendo quindi le modalità del carattere m_1, m_2, \dots, m_k e le relative frequenze f_1, f_2, \dots, f_k , la media aritmetica ponderata M è data da:

$$M = \frac{m_1 \cdot f_1 + m_2 \cdot f_2 + \dots + m_k \cdot f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k m_i \cdot f_i}{\sum_{i=1}^k f_i} = \frac{1}{n} \sum_{i=1}^k m_i \cdot f_i.$$

Esempio A.14. Riprendiamo la tabella dell'esempio precedente relativa alla domanda “quante ore al giorno passi al computer?”, posta ad un campione di 52 ragazzi dai 16 ai 24 anni. Calcoliamo la media aritmetica ponderata.

Numero di ore	0	1	2	3	4	5	6	Totale
Numero di ragazzi	4	6	12	16	8	4	2	52

Considerando le 7 modalità del carattere “Numero di ore” riportate nella tabella, si ha:

$$M = \frac{0 \cdot 4 + 1 \cdot 6 + 2 \cdot 12 + 3 \cdot 16 + 4 \cdot 8 + 5 \cdot 4 + 6 \cdot 2}{4 + 6 + 12 + 16 + 8 + 4 + 2} = \frac{142}{52} = 2,73.$$

Possiamo dire che, in media, ciascun ragazzo passa circa 3 ore al giorno al computer.

Il valore della media aritmetica semplice effettuata sulle singole unità statistiche coincide con quella ponderata effettuata sul raggruppamento dei dati per modalità del carattere considerato (tabella delle frequenze).

A.3.3 Mediana

Definizione A.8. La *mediana* di una successione di dati disposti in ordine crescente è il valore equidistante dagli estremi, cioè è

- il dato che occupa la posizione centrale, se il numero dei dati è dispari;
- è la media aritmetica dei dati della coppia centrale, se il numero dei dati è pari.

Poiché per calcolare la mediana i dati devono essere ordinati, è bene sottolineare che tale indice non può essere calcolato se il carattere in esame è di tipo qualitativo non ordinabile.

Esempio A.15. Supponiamo di avere 7 dati disposti in ordine crescente: 5, 8, 10, 14, 18, 20, 25. Allora la mediana è il valore centrale, quello che occupa la quarta posizione, cioè il 14.

Esempio A.16. Supponiamo di avere 8 dati disposti in ordine crescente: 1, 5, 8, 10, 14, 18, 20, 25. La mediana è la media aritmetica dei dati che occupano la 4^a e la 5^a posizione, cioè $\frac{10+14}{2} = 12$.

Esempio A.17. Supponiamo di avere la distribuzione di frequenza riportata nella tabella. Il numero di osservazioni è pari, quindi la mediana è il valore della variabile che corrisponde alla media dei due valori centrali, rispettivamente quelli che nella serie ordinata occupano il 13° e il 14° posto.

È necessario in questo caso determinare le *frequenze cumulate*. Esse si ottengono sommando le frequenze che hanno un valore della variabile minore o uguale alla modalità considerata. La frequenza cumulata relativa al voto 3 rimane 2, quella relativa al voto 4 si ottiene sommando la frequenza del 3 e la frequenza del 4, cioè $2 + 2 = 4$, la frequenza cumulata relativa al voto 5 si ottiene dalla somma della frequenza del 3, del 4 e del 5 e così via. Il 14° posto corrisponde al voto 6, mentre il 15° posto è il voto 7. La mediana è quindi 6,5.

Voto	Frequenza	Frequenza cumulata
3	2	2
4	4	4+2=6
5	3	3+4+2=9
6	5	5+3+4+2=14
7	7	7+5+3+4+2=21
8	2	2+7+5+3+4+2=23
9	2	2+2+7+5+3+4+2=25
10	1	1+2+2+7+5+3+4+2=26
Totale	26	

🔗 *Esercizi proposti:* A.21, A.22, A.23, A.24, A.25, A.26, A.27, A.28, A.29, A.30, A.31

A.32

A.4 Indici di variabilità

Gli *indici di variabilità* vengono calcolati per analizzare in che modo i termini di una distribuzione si concentrano intorno ad un valore medio.

Definizione A.9. Il *campo di variazione* è la differenza fra il valore massimo ed il valore minimo assunti dalla variabile: $CVar = x_{\max} - x_{\min}$.

Tale indice dà un'informazione molto grossolana perché tiene conto solo del primo e dell'ultimo termine della distribuzione e non tiene conto di tutti i valori intermedi. Si considerino, ad esempio, le seguenti distribuzioni di stature:

Gruppo A (statura in cm)	150	155	155	160	165	180	175
Gruppo B (statura in cm)	150	160	175	170	170	170	180

Entrambe le distribuzioni hanno lo stesso valore massimo e lo stesso valore minimo e quindi lo stesso campo di variazione, ma mentre nella prima i valori sono concentrati verso il valore minimo nella seconda si concentrano intorno al valore massimo.

L'indice non dà quindi alcuna indicazione su quest'ultima informazione. Né può essere utilizzato come indice di variabilità la media degli scarti fra le singole osservazioni e la loro media aritmetica perché tale valore è sempre uguale a zero.

A.4.1 Scarto medio assoluto

Definizione A.10. Si definisce *scarto medio assoluto* la media aritmetica dei valori assoluti degli scarti; esso indica quanto i valori rilevati si disperdono intorno al valore medio della distribuzione:

$$s = \frac{|s_1| + |s_2| + \dots + |s_n|}{n} = \frac{1}{n} \sum_{i=1}^n |x_i - M|.$$

Facendo riferimento alla distribuzione

Classe	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	Totale
Studenti	320	230	212	152	96	1010

si ha che lo scarto medio assoluto è 62,4. Si può allora affermare che in ogni tipologia di classe si hanno in media $202 \pm 62,4$ iscritti.

A.4.2 Varianza e scarto quadratico medio

L'indice di variabilità più utilizzato è la varianza o lo scarto quadratico medio.

Definizione A.11. La *varianza* è la media dei quadrati degli scarti fra le singole osservazioni e la loro media aritmetica:

$$\text{Var} = \frac{[(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2]}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - M)^2.$$

Lo *scarto quadratico medio* è la radice quadrata della varianza: $\sigma = \sqrt{\text{Var}}$.

Se i dati si presentano sotto forma di distribuzione di frequenza, la media deve essere ponderata con le singole frequenze, cioè:

$$\begin{aligned} \text{Var} &= \frac{[(m_1 - M)^2 \cdot f_1 + (m_2 - M)^2 \cdot f_2 + \dots + (m_k - M)^2 \cdot f_k]}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^k (m_i - M)^2 \cdot f_i}{\sum_{i=1}^k f_i} = \\ &= \frac{1}{n} \sum_{i=1}^k (m_i - M)^2 \cdot f_i. \end{aligned}$$

La varianza assume valore zero quando tutti i valori coincidono con la media ed è tanto più grande quanto più i singoli valori si discostano dalla media. Poiché tale indice è influenzato sia dal valore della media che dall'unità di misura utilizzato, spesso si utilizza un indice detto coefficiente di variazione.

A.4.3 Coefficiente di variazione

Definizione A.12. Il *coefficiente di variazione* è il rapporto fra lo scarto quadratico medio (radice quadrata della varianza) e la media aritmetica:

$$CV = \frac{\sigma}{M} = \frac{\sqrt{\text{Var}}}{M}.$$

Tale indice risulta di particolare utilità per confrontare distribuzioni diverse.

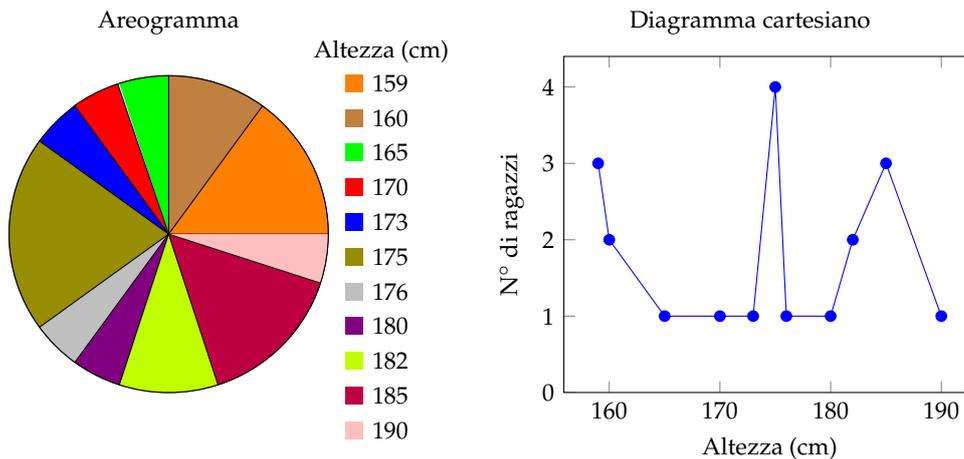
Esempio A.18. È dato l'elenco delle stature, in cm, dei ragazzi di una classe: 165, 182, 159, 173, 160, 175, 185, 190, 175, 180, 159, 185, 176, 170, 175, 160, 175, 182, 159, 185.

- Ordina i dati in una tabella delle frequenze;
- rappresenta i dati graficamente;
- calcola la media, la mediana e la moda;
- calcola la varianza e il coefficiente di variazione.

Tabella delle frequenze

Dati	Frequenze assolute	Frequenze relative	Frequenze percentuali
159	3	0,15	15%
160	2	0,10	10%
165	1	0,05	5%
170	1	0,05	5%
173	1	0,05	5%
175	4	0,20	20%
176	1	0,05	5%
180	1	0,05	5%
182	2	0,10	10%
185	3	0,15	15%
190	1	0,05	5%
Totale	20	1	100%

- La somma delle frequenze assolute è pari al numero totale degli studenti;
- la somma delle frequenze relative è 1 (a meno di approssimazioni nel calcolo delle frequenze relative delle singole modalità del carattere);
- la somma delle frequenze percentuali è 100 (a meno di approssimazioni nel calcolo delle frequenze percentuali delle singole modalità del carattere).

Grafici**Calcolo della media, mediana e moda**

Calcoliamo la media aritmetica:

$$\text{Media} = \frac{1}{20} \cdot (165 + 182 + 159 + 173 + 160 + 175 + 185 + 190 + 175 + 180 + 159 + 185 + 176 + 170 + 175 + 160 + 175 + 182 + 159 + 185) = 173,5.$$

Per determinare la mediana si devono ordinare in modo crescente i dati: 159, 159, 159, 160, 160, 165, 170, 173, 175, 175, 175, 175, 176, 180, 182, 182, 185, 185, 185, 190. Essendo i dati in numero pari si calcola la media dei due dati centrali: $\text{Mediana} = (175 + 175)/2 = 175$. Se i dati sono molti è possibile individuare qual è o quali sono i dati centrali utilizzando la tabella delle frequenze opportunamente costruita, cioè con i dati scritti in ordine crescente.

La moda è la modalità del carattere altezza che è più ricorrente, cioè quello con la frequenza più alta: $\text{Moda} = 175$.

Esercizi proposti: [A.33](#), [A.34](#), [A.35](#), [A.36](#), [A.37](#), [A.38](#)

A.5 Esercizi

A.5.1 Esercizi dei singoli paragrafi

A.1 - Indagine statistica

A.1. In una indagine su alcune famiglie si sono rilevati i seguenti caratteri; indicane il tipo ponendo una crocetta nella casella opportuna; per i caratteri quantitativi indica se sono discreti o continui, per i caratteri qualitativi indica se sono ordinabili o sconnessi:

Carattere	quantitativo		qualitativo	
	discreto	continuo	ordinabile	sconnesso
Reddito mensile del capofamiglia				
Titolo di studio del capofamiglia				
Familiari a carico				
Settore lavorativo				
Luogo di nascita del capofamiglia				
Tempo impiegato per raggiungere il luogo di lavoro				

A.2 - Fasi di un'indagine statistica

A.2. Compila una tabella relativa alla distribuzione degli studenti della tua classe in relazione a:

- colore dei capelli (nero, castano, biondo, rosso);
- anno di nascita;
- città di residenza.

A.3. In una certa nazione in un dato anno si sono vendute 10 540 biciclette, 7 560 scooter, 2 300 moto e 6 532 automobili. Completa la tabella:

Mezzi di trasporto venduti	Freq. assoluta	Freq. relativa	Freq. percentuale
Biciclette			
Scooter			
Moto			
Automobili			
Totale			

A.4. Da un'indagine sulla distribuzione delle altezze in un gruppo di studenti sono stati rilevati i seguenti dati grezzi (espressi in cm):

175 168 169 173 160 165 170 172 177 172 170 173 182
 164 174 185 188 164 175 160 177 176 184 180 176 168
 174 175 177 183 174 166 181 173 166 172 174 165 180
 190 175 176 188 171 172 181 185 184 183 175 173 181

Raggruppa i dati in classi di ampiezza 5cm e costruisci la distribuzione di frequenza. Calcola poi frequenza relativa e percentuale.

A.5. Dall'analisi delle paghe settimanali dei dipendenti di un'industria automobilistica si è ottenuta la seguente distribuzione di frequenza, suddivisa in classi (la parentesi quadra indica che l'estremo della classe considerato è incluso nella classe stessa, la parentesi tonda indica che l'estremo della classe considerato è escluso dalla classe). Determina per ogni classe di reddito frequenza relativa e percentuale.

Classi di reddito (€)	Freq. assoluta	Freq. relativa	Freq. percentuale
[50, 100)	50		
[100, 200)	70		
[200, 300)	30		
≥ 300	50		

A.6. Data la seguente distribuzione dei risultati dei test d'ingresso di matematica in una scuola media, sapendo che l'indagine è stata svolta su 200 alunni, determina frequenze assolute e relative.

Voto	3	4	5	6	7	8	9
Frequenza percentuale	5%	10%	25%	40%	15%	3%	2%
Frequenza assoluta							
Frequenza relativa							

A.7. Osserva la seguente tabella:

	Freq. assoluta	Freq. relativa	Freq. percentuale
Infanzia	950 000		
Primaria	2 538 000		
Secondaria di 1° grado	1 700 000		
Secondaria di 2° grado	2 425 000		
Totale			

- Quale fenomeno descrive la tabella?
- Qual è la popolazione statistica oggetto dell'indagine?
- Quante sono le unità statistiche?
- Qual è stato il carattere indagato?
- Completa la tabella calcolando frequenza relativa e frequenza percentuale.

A.8. In un campione di ginnaste di livello agonistico si è rilevata l'altezza in metri. Questa frase è sufficiente per indicare la popolazione oggetto di indagine e il carattere rilevato? Il carattere analizzato è di tipo qualitativo o quantitativo?

L'indagine ha dato i seguenti risultati:

Altezza (m)	1,49	1,50	1,55	1,58	1,61	1,64	1,67	1,70	1,71
Numero ginnaste	1	6	11	4	6	4	2	2	3

Quante sono le unità statistiche? Determina in percentuale il numero delle ginnaste la cui altezza è non inferiore a 1,60 m.

A.9. La tabella mostra dati relativi ad una popolazione di 20 famiglie italiane; le informazioni in essa contenute stabiliscono alcuni aspetti o caratteri dei membri della popolazione: numero di componenti, reddito annuo (in migliaia di euro), titolo di studio del capofamiglia, residenza per area geografica. Osserva la tabella e rispondi alle domande che seguono.

Famiglia	Numero componenti	Reddito annuo	Titolo di studio	Residenza
1	2	28	Elementare	Nord
2	1	35	Media inferiore	Centro
3	3	50	Media inferiore	Nord
4	1	45	Media superiore	Nord
5	1	40	Laurea	Sud
6	2	30	Media inferiore	Sud
7	3	55	Media inferiore	Centro
8	4	80	Media superiore	Centro
9	5	60	Laurea	Sud
10	6	85	Laurea	Nord
11	7	90	Laurea	Nord
12	1	52	Media superiore	Centro
13	2	62	Media superiore	Sud
14	3	75	Media superiore	Sud
15	5	60	Elementare	Nord
16	4	45	Media inferiore	Nord
17	3	42	Media inferiore	Centro
18	2	28	Elementare	Nord
19	8	70	Media superiore	Sud
20	2	38	Laurea	Sud

- ➔ Cosa si intende, in statistica, per popolazione?
- ➔ Quali sono le unità statistiche di cui sono trascritti i dati nella tabella precedente?
- ➔ Quali caratteri riportati nella tabella sono qualitativi e quali quantitativi?
- ➔ Quali sono le modalità dei caratteri qualitativi indagati?
- ➔ Le informazioni della precedente tabella sono sufficienti per stabilire:
 - dove risiede la maggior parte delle famiglie oggetto di questa indagine? Se sì, come lo stabilisci?
 - il numero di famiglie il cui capofamiglia ha come titolo di studio quello di Scuola Media Superiore? Se sì, come lo stabilisci?
- ➔ costruire la tabella:

Titolo di studio	Elementare	Media inferiore	Media superiore	Laurea
Numero di famiglie				

- ➔ È vero che $1/4$ dei capifamiglia, cioè il 25%, è laureato?
- ➔ Costruire un'altra tabella, sul modello della precedente, in cui è riportato il numero di famiglie aventi 1, 2, 3, ecc. componenti. È vero che $1/3$ delle famiglie è costituito da più di 5 persone?

- Individua il reddito minimo e quello massimo, completa la seguente tabella delle frequenze in modo che il carattere reddito sia suddiviso in classi di ampiezza 5, come indicato.

Classe di reddito	Frequenza assoluta
26-30	
31-35	
...	

- Quante famiglie hanno un reddito compreso tra 46 e 90 mila euro? Indica la risposta anche in percentuale.

A.10 (Fonte Wikipedia). Rappresenta con un diagramma cartesiano la seguente serie storica relativa alla produzione di olio di oliva in Puglia, scegliendo un'opportuna unità di misura:

Anno	2006	2005	2004	2003
Produzione olio (quintali)	1 914 535	2 458 396	2 678 201	2 508 084

A.11 (Fonte ISTAT). Rappresenta con un diagramma cartesiano la seguente serie storica, relativa al numero di società quotate in borsa, dal 1975 al 1984:

Anno	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
Società	154	156	156	148	145	141	141	148	150	155

A.12. Rappresenta graficamente, mediante diagramma cartesiano, la seguente tabella che riporta le temperature misurate a Lecce durante una giornata invernale.

Ore	0	2	4	6	8	10	12	14	16	18	20	22
Temperatura (°C)	5	5,5	5,5	6	7,5	10	16	18	16,5	12	8	6,5

A.13. Rappresenta attraverso un ideogramma la seguente tabella statistica, che indica le ore di studio giornaliere di uno studente, usando 2 ore come unità di misura. Scegli un simbolo opportuno.

Giorno	Lunedì	Martedì	Mercoledì	Giovedì	Venerdì	Sabato	Domenica
Ore studio	2	6	5	2	3	4	0

A.14. Costruisci un ideogramma a partire dai dati della seguente tabella:

Regione	Produzione vino (quintali)
Toscana	20 500
Veneto	18 000
Puglia	15 500
Campania	14 500
Molise	8 000

A.15. La seguente tabella rappresenta i risultati di un'indagine sulla capitale europea preferita da un gruppo di studenti universitari. Rappresenta i dati utilizzando un diagramma a nastro.

Capitale preferita	Frequenza
Amsterdam	28
Londra	30
Parigi	25
Roma	42
Vienna	10

A.16. Rappresenta con un diagramma a colonne i dati riportati nella seguente tabella relativi alla vendita di automobili da un concessionario nell'anno 2009.

Marca automobile	Auto vendute
Alfa Romeo	30
Fiat	270
Ford	120
Renault	50
Toyota	40

A.17. Consideriamo la seguente tabella statistica che indica le frequenze percentuali di forza lavoro per settore economico rilevata nel 2006 in Italia:

Forza lavoro per settore economico	Frequenza percentuale
Forza lavoro occupata nell'agricoltura	4,20%
Forza lavoro occupata nell'industria	30,70%
Forza lavoro occupata nei servizi	65,10%
Tasso di disoccupazione	8,00%

Rappresentare graficamente mediante areogramma i dati contenuti nella tabella.

A.18. Rappresentare attraverso un areogramma la seguente tabella statistica, che indica le altezze di 100 studenti maschi di una data scuola dopo aver calcolato le frequenze percentuali:

Altezza (m)	Numero di studenti	Frequenze percentuali
1,50 - 1,59	11	
1,60 - 1,69	18	
1,70 - 1,79	42	
1,80 - 1,89	22	
1,90 - 1,99	6	
Totale	100	

A.19. Rappresentare attraverso un istogramma la seguente tabella statistica, che indica le altezze di 100 studenti maschi di una data scuola:

Altezza (m)	1,50 - 1,59	1,60 - 1,69	1,70 - 1,79	1,80 - 1,89	1,90 - 1,99
Numero di studenti	11	18	42	22	6

A.20. Uno studente universitario di Matematica ha superato 28 esami con queste valutazioni:

18 25 26 23 30 21 24 20 29 28 24 21 23 28
28 24 22 25 24 27 24 21 23 28 18 25 26 23

Organizza i dati in una tabella suddividendoli in classi e rappresentali tramite un istogramma.

A.3 - Indici di posizione

A.21. Un concessionario vende delle moto di diversa cilindrata come descritto nella tabella. Determinare la moda.

Cilindrata	250	350	500	750	1000
Numero moto vendute	34	30	45	100	42

A.22. Calcolare la moda della distribuzione rappresentata attraverso la seguente tabella statistica:

Modalità del carattere	3	6	8	9	12	24
Frequenza	23	78	67	78	89	100

A.23. Calcolare la classe modale della seguente distribuzione:

Abitanti	0 - 999	1 000 - 1 999	2 000 - 4 999	5 000 - 9 999	10 000 - 19 999
Numero comuni	750	1 100	950	2 500	3 000

A.24 (*). Trovare la media aritmetica semplice delle seguenti serie di osservazioni:

a) 3, 4, 6, 7, 10;

c) 34, 53, 45, 67, 87, 90, 100, 123.

b) 6, 7, 8, 12, 15, 22;

A.25. In una classe di 15 ragazzi sono stati rilevati i seguenti pesi in kg: 50, 43, 62, 41, 70, 55, 76, 43, 46, 50, 78, 62, 49, 55, 48. Calcola la media aritmetica semplice del peso dei ragazzi. Costruisci la tabella delle frequenze. Calcola la media aritmetica ponderata del peso dei ragazzi. Che cosa osservi?

A.26 (*). In un insieme di numeri compaiono quattro volte il 3, cinque volte il 5, tre volte il 6, due volte il 10, due volte il 15. Calcolare la media aritmetica.

A.27 (*). Calcola la media della seguente distribuzione di frequenza.

Punteggio	2	4	6	7	12	14
Frequenza assoluta	2	4	5	4	3	2

A.28. Una rivista di auto fornisce i seguenti punteggi per tre diversi modelli di automobili.

	Funzionalità	Volumetria	Prestazioni	Sicurezza	Economia
Modello 1	2,5	4	3,2	3,5	2,5
Modello 2	2,5	3	4	3,5	2
Modello 3	2,7	3	3,5	3,8	2,5

d) calcola la varianza.

A.37. Una ditta paga 5 persone € 165 alla settimana, 4 persone € 199 alla settimana e 2 persone € 218 alla settimana. Trova media aritmetica, moda e mediana. Che percentuale di persone ha la retribuzione che si discosta, sia in positivo che in negativo, di € 20 dalla media?

A.38. È stata effettuata un'indagine statistica fra le persone presenti in una libreria riguardo al numero di libri letti nella scorsa estate. I dati sono raccolti nella seguente tabella:

N° libri letti	0	1	2	3	4	5	6	7
N° persone	20	35	9	6	3	0	1	1

- Organizza i dati in una tabella e calcola la frequenza assoluta, quella relativa e quella percentuale;
- rappresenta i dati in un grafico scelto a piacere;
- calcola moda, media e mediana dandone una semplice interpretazione;
- calcola varianza e coefficiente di variazione.

A.5.2 Esercizi riepilogativi

A.39. Scegli la risposta corretta:

- se compi un'indagine sul peso degli allievi della tua scuola, la popolazione è costituita
 - dagli allievi della scuola;
 - dai pesi degli allievi della tua scuola;
 - da ciascun allievo della scuola;
 - dal peso di ciascun allievo della scuola.
- nella stessa indagine, da cosa sarà costituita un'unità statistica?
 - dagli allievi della scuola;
 - dai pesi degli allievi della tua scuola;
 - da ciascun allievo della scuola;
 - dal peso di ciascun allievo della scuola.
- un'indagine statistica realizzata intervistando solo una parte della popolazione statistica è definita
 - incompleta;
 - universo;
 - censimento;
 - a campione;
- la frequenza percentuale si ottiene
 - dividendo la frequenza per il totale delle frequenze e moltiplicando il risultato per 100;
 - moltiplicando la frequenza per 100;
 - moltiplicando la frequenza per il totale delle frequenze e dividendo il risultato per 100;
 - dividendo la frequenza per 100.
- la mediana:

- a) è il valore che si ottiene dividendo la somma dei valori delle singole osservazioni per il loro numero;
- b) è il valore equidistante dagli estremi di un insieme di dati ordinati;
- c) è il valore che si presenta con la massima frequenza in un insieme di dati;
- d) è il valore che indica la percentuale di dati al di sopra o al di sotto della media.
6. la media aritmetica:
- a) è il valore che si ottiene dividendo la somma dei valori delle singole osservazioni per il loro numero;
- b) è il valore equidistante dagli estremi di un insieme di dati ordinati;
- c) è il valore che si presenta con la massima frequenza in un insieme di dati;
- d) è il valore che indica la percentuale di dati al di sopra o al di sotto della media.
7. la moda:
- a) è il valore che si ottiene dividendo la somma dei valori delle singole osservazioni per il loro numero;
- b) è il valore equidistante dagli estremi di un insieme di dati ordinati;
- c) è il valore che si presenta con la massima frequenza in un insieme di dati;
- d) è il valore che indica la percentuale di dati al di sopra o al di sotto della media.
8. nella seguente distribuzione di dati 2, 4, 4, 4, 4, 6, 6, 6, 7, 7, quale delle seguenti affermazioni è corretta?
- a) la media aritmetica è 5, la moda è 4, la mediana è 6;
- b) la media aritmetica è 4, la moda è 6, la mediana è 5;
- c) la media aritmetica è 5, la moda è 6, la mediana è 4;
- d) la media aritmetica è 5, la moda è 4, la mediana è 5.
9. nella tua classe la mediana dell'altezza è 152cm. Questo significa che:
- a) non ci sono studenti più bassi di 152cm;
- b) 152cm è l'altezza più comune;
- c) la metà degli studenti ha un'altezza inferiore a 152cm, mentre l'altra metà ha un'altezza superiore;
- d) in media gli studenti sono alti 152cm.
10. nella tua classe la moda dell'altezza è 152cm. Questo significa che:
- a) non ci sono studenti più bassi di 152cm;
- b) 152cm è l'altezza più comune;
- c) la metà degli studenti ha un'altezza inferiore a 152cm, mentre l'altra metà l'ha superiore;
- d) in media gli studenti sono alti 152cm.
11. nella tua classe la media aritmetica dell'altezza è 152cm. Questo significa che:
- a) non ci sono studenti più bassi di 152cm;
- b) 152cm è l'altezza più comune;
- c) la metà degli studenti ha un'altezza inferiore a 152cm, mentre l'altra metà l'ha superiore;
- d) se tutti gli alunni avessero la stessa altezza questa sarebbe di 152cm.

A.40. In un test sulla prova di velocità di lettura i candidati hanno ottenuto i seguenti risultati:

N° pagine lette in 15 minuti	10	12	11	9	14	13	7
N° candidati	2	5	2	1	1	3	4

- Organizza i dati in una tabella indicando frequenza assoluta, frequenza relativa e percentuale;
- rappresenta i dati in un diagramma a bastoni;
- calcola moda, media e mediana;
- quanti candidati in percentuale hanno letto un numero di pagine sopra la media?

A.41. In un gruppo di ragazzi le stature (espresse in centimetri) risultano distribuite nel seguente modo: 163, 169, 171, 165, 173, 165, 163, 168, 168, 169, 171, 169, 181, 165, 168, 169, 169, 163, 169, 168, 150, 168, 172, 181, 165, 169, 172, 169, 192, 173, 163, 168.

- Costruisci una tabella indicando i dati, la loro frequenza, la frequenza relativa e la percentuale;
- suddividi i dati in 4 classi, costruisci la distribuzione di frequenza e rappresentali graficamente con un istogramma;
- calcola la moda, la media e la mediana.

A.42. Sono state misurate le pulsazioni al minuto di 20 persone ottenendo i seguenti dati: 79, 72, 69, 69, 72, 80, 73, 73, 70, 66, 80, 68, 70, 72, 82, 75, 72, 71, 74, 64.

- Organizza i dati in una tabella comprensiva di percentuale di frequenze;
- rappresenta graficamente i dati;
- calcola moda, media e mediana.

A.43. Ventuno ragazzi sono stati sottoposti a una verifica; i dati seguenti esprimono il numero di errori commessi da ciascuno di loro: 3, 4, 1, 3, 6, 6, 3, 1, 4, 7, 3, 1, 1, 3, 7, 7, 1, 3, 7, 3, 3.

- Organizza i dati in una tabella comprensiva di percentuale di frequenze;
- rappresenta graficamente i dati;
- calcola moda, media e mediana;
- quanti alunni, in percentuale, hanno fatto meno di 5 errori?

A.44. I dati riportati in tabella si riferiscono ai giorni di assenza degli alunni di una classe.

Alunno	n° giorni	Alunno	n° giorni	Alunno	n° giorni	Alunno	n° giorni
Mauro	5	Romeo	2	Bruna	7	Silvia	2
Antonio	7	Anna	4	Pietro	2	Alessio	2
Paola	5	Luca	4	Nicola	7	Patrizia	9
Luisa	5	Amedeo	5	Aldo	2	Franca	1
Carla	1	Marco	7	Luigi	2	Chiara	7

- Organizza i dati in una tabella comprensiva di percentuale di frequenze;
- rappresenta i dati con un istogramma;
- calcola moda, media e mediana;
- quanti alunni, in percentuale, hanno fatto meno assenze rispetto alla media?

A.45. Nella tabella sono riportati i punteggi ottenuti da 22 alunni in un test formato da 20 quesiti a scelta multipla e il numero di risposte esatte.

N° ordine	Punteggi	Risposte esatte	N° ordine	Punteggi	Risposte esatte
1	80	26	12	55	11
2	62	12	13	58	11
3	48	9	14	80	16
4	71	14	15	75	14
5	80	16	16	65	12
6	90	18	17	58	11
7	75	15	18	58	10
8	67	13	19	62	12
9	79	15	20	57	11
10	62	12	21	60	12
11	95	19	22	48	8

- Il punteggio medio è stato ... con uno scarto quadratico medio di ...;
- la mediana della distribuzione è il punteggio ...;
- le risposte esatte sono state in media ... con uno scarto quadratico di ...;
- rappresenta ciascuna distribuzione con un istogramma, dopo aver aggregato i dati in classi come indicato nelle tabelle sottostanti.

Carattere ...		Carattere ...	
Punteggio	Frequenza assoluta	Risposte esatte	Frequenza assoluta
$48 \leq p < 58$		$7 \leq \text{r.e.} < 9$	
$58 \leq p < 68$		$9 \leq \text{r.e.} < 11$	
$68 \leq p < 78$		$11 \leq \text{r.e.} < 13$	
$78 \leq p < 88$		$13 \leq \text{r.e.} < 15$	
$88 \leq p < 98$		$15 \leq \text{r.e.} < 17$	
		$17 \leq \text{r.e.} < 19$	
		$19 \leq \text{r.e.} < 21$	
Totale		Totale	

A.46. Una scatola contiene 20 sacchetti di biscotti confezionati da una industria. I pesi rilevati in grammi sono: 380, 365, 371, 375, 376, 369, 376, 377, 381, 383, 384, 377, 370, 375, 374, 376, 373, 378, 383, 378.

- Il carattere rilevato è ..., esso è di tipo ... e si presenta secondo modalità Inserisci nella tabella sottostante nella colonna C1 il carattere rilevato e le sue modalità;
- quanto è il peso totale della scatola? Come lo hai calcolato?
- il peso medio dei sacchetti di biscotti è Media = ...;
- qual è il campo di variazione del peso dei sacchetti? CVar = ...;
- la mediana della distribuzione è ...;
- nella colonna "scarto" riporta, per ciascun valore del carattere indagato, lo scarto dalla media. Verifica la proprietà degli scarti rispetto rispetto alla media: la loro somma è ...;
- completa la colonna |scarto| con il valore assoluto degli scarti e determina lo scarto medio assoluto $s = \dots$;

- h) completa la colonna scarto^2 con il quadrato degli scarti e calcola la varianza $\text{Var} = \dots$ e il coefficiente di variazione $\text{CV} = \dots$;
- i) raggruppa i valori del carattere in classi di ampiezza 5gr e completa la tabella;
- j) metti in evidenza la classe modale e spiega il significato di moda;
- k) costruisci l'istogramma della distribuzione;

	C1	scarto	scarto	scarto^2		C1	scarto	scarto	scarto^2
1					11				
2					12				
3					13				
4					14				
5					15				
6					16				
7					17				
8					18				
9					19				
10					20				
Totale									

- l) organizza i dati in classi:

Classi di peso	Frequenza assoluta
[365, 370)	
...	

A.47. Dai dati di scrutinio del primo quadrimestre in una scuola secondaria di 2° grado, è stata elaborata la seguente tabella in cui compaiono i voti in matematica degli alunni delle classi prime:

Voto	3	4	5	6	7	8	9	10	Totale
Frequenza	1	3	5	7	2	3	1	1	
Frequenza relativa									
Frequenza percentuale									

- a) Indica il numero di unità statistiche oggetto dell'indagine e spiega come lo puoi ottenere;
- b) il carattere rilevato è ...; esso è di tipo ... e si presenta secondo modalità ...;
- c) la tabella assegnata è di dati aggregati o disaggregati?
- d) rappresenta la distribuzione attraverso un grafico a barre (o a nastro);
- e) cosa si intende per frequenza assoluta?
- f) completa la colonna della frequenza relativa;
- g) completa la colonna frequenza percentuale;
- h) determina la moda della distribuzione: $\text{Moda} = \dots$;
- i) il voto medio in matematica alla fine del primo quadrimestre è stato ...;
- j) determina la mediana della distribuzione: $\text{Mediana} = \dots$;
- k) amplia la tabella indicando gli scarti dalla media;
- l) calcola lo scarto medio assoluto e lo scarto quadratico medio;

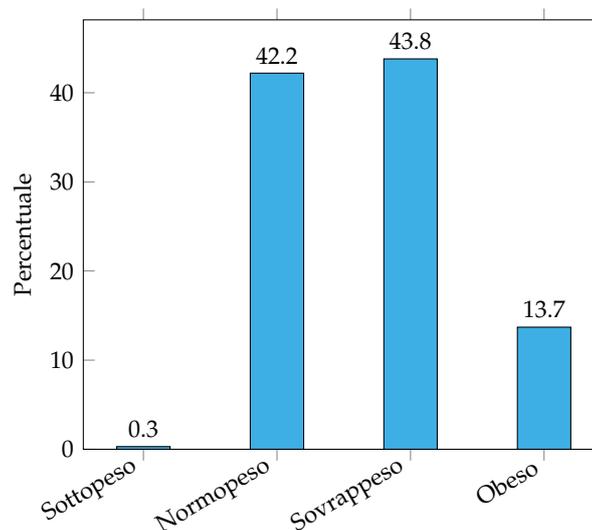
- m) il voto medio dei ragazzi sufficienti è stato . . . , quello dei ragazzi insufficienti è stato . . . ;
 n) rappresenta la situazione con un areogramma distinguendo tra ragazzi sufficienti e ragazzi insufficienti.

A.48 (Prove Invalsi 2011). Il reddito medio annuo dei lavoratori agricoli di un certo paese ammonta a 3 500 scudi e quello dei lavoratori dell'industria a 4 500 scudi. È corretto affermare che il reddito medio complessivo ammonta a 4 000 scudi?

A.49 (*Prove Invalsi 2011). La settimana scorsa la mamma chiese ad Aurelia di trascrivere al computer un manoscritto e Aurelia le assicurò che avrebbe battuto 20 pagine al giorno. Per la prima metà del manoscritto andò piuttosto lentamente battendo 10 pagine al giorno e poi, per recuperare il tempo perduto, trascrisse la seconda metà a 30 pagine al giorno. Quando ebbe finito portò a sua madre la trascrizione dicendole: Vedi, ho fatto una media di 20 pagine al giorno, come ti avevo promesso. Infatti $(10 + 30)/2 = 20$. Non è vero, replicò sua madre.

A.50 (*Prove Invalsi 2011). In una indagine sullo stato di salute della popolazione sono state raccolte informazioni relative al peso e alla statura di 1 000 intervistati. Gli intervistati sono stati poi suddivisi in quattro gruppi, come riportato nel grafico seguente. Quante sono le persone in sovrappeso?

- a) Più di 500, ma meno di 600;
 b) più di 600;
 c) meno della somma delle persone sottopeso e obese;
 d) all'incirca tante quante sono le persone normopeso.



A.51. Quattro amici sostengono l'Esame di Stato conseguendo punteggi la cui media aritmetica è $77,5/100$. Se tre di essi hanno conseguito un punteggio, in centesimi, rispettivamente di 70, 76, 80, quale punteggio ha conseguito il quarto studente?

A.52 (Prove Invalsi 2004-2005). La seguente tabella si riferisce alla rilevazione effettuata in una classe prima di un Istituto Tecnico.

Sesso	Scuola media di provenienza			
	Scuola A	Scuola B	Scuola C	Altre scuole
Maschi	5	3	4	2
Femmine	6	3	4	3

Qual è la percentuale di alunni provenienti dalla Scuola B?

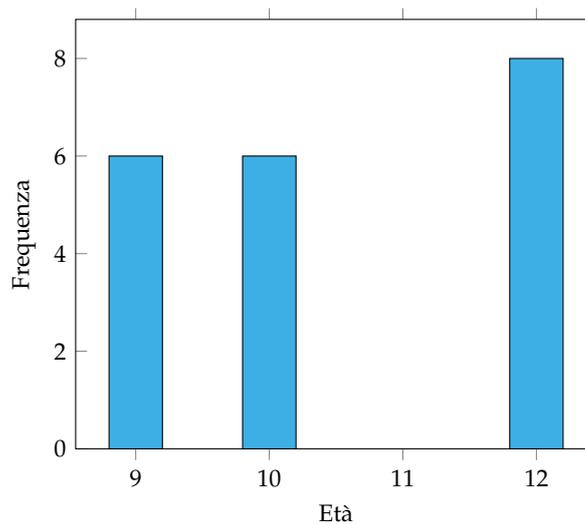
A.53 (Prove Invalsi 2005-2006). In una classe di 25 alunni, i punteggi (abbreviati in tabella con p) ottenuti in un test di matematica risultano distribuiti come indicato nella seguente tabella.

Puntaggio	$0 \leq p < 20$	$20 \leq p < 40$	$40 \leq p < 60$	$60 \leq p < 80$	$80 \leq p \leq 100$
Numero alunni					

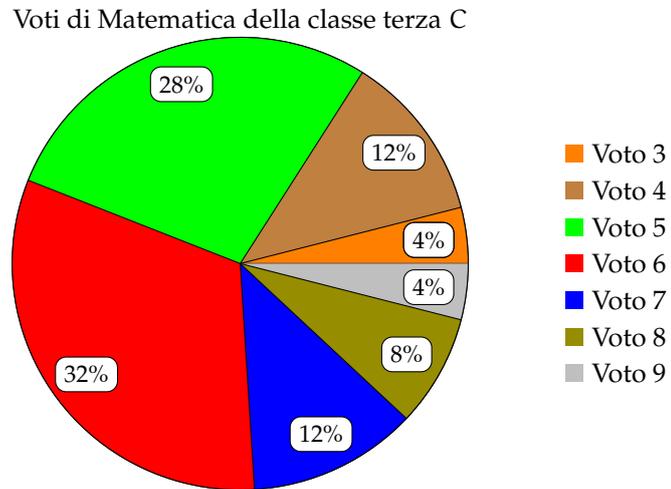
Qual è la percentuale di alunni che ha ottenuto un punteggio inferiore a 60?

A.54 (Prove Invalsi 2005-2006). Un impiegato ha percepito per i primi 3 mesi dell'anno uno stipendio mensile di € 850. Nei 9 mesi successivi ha percepito lo stipendio mensile precedente aumentato di € 200. Quant'è lo stipendio medio nell'anno di quell'impiegato?

A.55 (Prove Invalsi 2005-2006). Nel grafico seguente si riporta l'età dei ragazzi che frequentano una palestra. Qual è la media aritmetica dell'età dei ragazzi se la distribuzione di frequenza è quella indicata nel grafico?



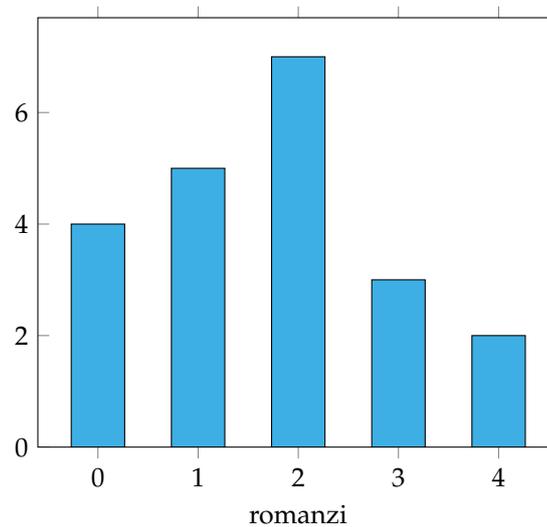
A.56 (Prove Invalsi 2006-2007). I 25 alunni della terza C, dopo aver raccolto i voti conseguiti nella verifica scritta di matematica, hanno costruito il seguente grafico:



Quanti ragazzi hanno conseguito come voto 7?

- a) 12; b) 7; c) 5; d) 3.

A.57. La figura indica quanti romanzi leggono gli alunni di una classe in un mese. Quanti sono gli alunni che leggono almeno 2 romanzi?



A.58 (Prove Invalsi 2004-2005). Il Ministero dell'Istruzione ha diffuso le seguenti informazioni sul numero di alunni stranieri della scuola italiana nell'anno scolastico 2003-2004. La tabella riporta solo le 5 nazionalità più numerose.

Nazionalità più numerose	Numero di alunni	Percentuale di alunni sul totale degli stranieri
Albania	50 000	18,00%
Marocco	42 000	15,00%
Romania	28 000	10,00%
Cina	16 000	6,00%
Ecuador	11 000	4,00%

Cosa si può dedurre da tali dati sugli alunni stranieri di nazionalità russa? Sono ...

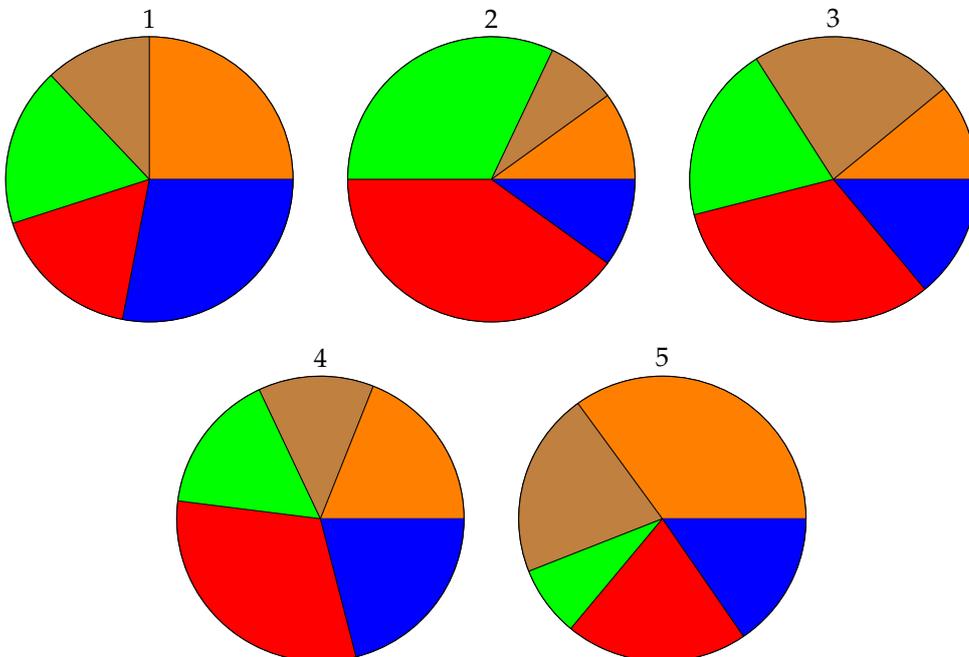
- a) meno di 11 000;
- b) sicuramente meno di 400;
- c) una percentuale compresa fra il 4% e il 18%;
- d) assenti dalle scuole italiane.

A.59. La tabella mostra la superficie delle varie province del Lazio.

Provincia	Frosinone	Latina	Rieti	Roma	Viterbo
Superficie (km ²)	3 240	2 251	2 749	5 352	3 612

Quale dei diagrammi riportati sotto descrive graficamente i dati della tabella?

■ Frosinone
 ■ Latina
 ■ Rieti
 ■ Roma
 ■ Viterbo



A.5.3 Risposte

A.24. a) 6; b) 11, 7; c) 75.

A.26. 21.

A.27. 7,1.

A.30. a) 6; b) 10; c) 89.

A.31. 43.

A.32. 15.

A.49. 15.

A.50. d.