Unità

SP

Elementi di statistica e probabilità

© Adriano Schena / tipsimages

1. La statistica

Leggendo un giornale o ascoltando una trasmissione televisiva, può capitare di imbatterci in domande del tipo:

«Quanta energia sarà necessaria per il fabbisogno nazionale fra cinque anni?»
«Quante persone, che attualmente hanno 65 anni, vivranno oltre i 75 anni?»
«Quante aule e quanti professori saranno necessari per i bienni fra 6 anni?»
«Verso quali generi musicali si stanno orientando i giovani dai 14 ai 17 anni?»

Non ci aspettiamo certo che si possano dare risposte esatte a queste domande, tanto meno che sia possibile anche solo valutare i problemi prima ancora di raccogliere informazioni relative alle questioni che i problemi pongono.
Le informazioni che si raccolgono sono spesso numeriche, e vengono chiamate **dati**.
La **statistica** si interessa del rilevamento, dell'elaborazione e dello studio dei dati.

In un paese moderno la politica e l'economia sono guidate dalla statistica. Il termine *statistica* ricorda proprio l'origine di questa scienza, nata come strumento per ben governare lo stato.
Oggi il campo di applicazione della statistica è molto vasto; a essa si fa ricorso nella ricerca scientifica, nei sondaggi dell'opinione pubblica, nel controllo dei prodotti industriali, nella psicologia sperimentale, nella meteorologia, nella ricerca di mercato, nella ricerca degli indici di ascolto o di gradimento ecc.
Per esempio, per chi si occupa di meteorologia, è normale parlare di temperatura, vento, pioggia, durata dell'illuminazione diurna. Queste informazioni vengono tradotte in numeri e registrate nelle stazioni meteorologiche sparse in tutto il mondo; il meteorologo fa la sua previsione solo dopo aver studiato i dati sul tempo raccolti in una regione.
Informazioni numeriche sul volume degli affari, sui prezzi, sugli stipendi, sulla disoccupazione si trovano in giornali e riviste: gli uomini d'affari osservano molto accuratamente queste notizie per decidere, per esempio, quanta merce tenere in magazzino se, quando e dove costruire una nuova fabbrica, o quando preparare un lancio pubblicitario.
Anche le autorità scolastiche devono studiare dati numerici per decidere quando e dove costruire nuove scuole, quanti insegnanti occorrono per i prossimi anni, e quali saranno i costi per ogni studente.
La statistica si può usare in molti modi per cambiare e migliorare il nostro mondo.
I farmacologi studiano i dati relativi a esperimenti per preparare nuove medicine e controllarne l'efficacia.
Le industrie costruiscono laboratori e assumono matematici e scienziati per ottenere informazioni su nuovi prodotti, come materiali da costruzione, sostanze chimiche, motori d'automobile, propellenti per missili, calcolatori ad alta velocità.

Esistono istituzioni specializzate proprio per raccogliere e analizzare dati come l'ISTAT (Istituto Centrale di Statistica), che ha sede a Roma.

2. Rilevamenti statistici

Nelle indagini statistiche si prendono in esame fenomeni di vario tipo (per esempio, demografici, economici, sociali ecc.) che riguardano delle *popolazioni*, ossia insiemi di elementi che presentiamo tutti delle caratteristiche comuni.
Gli elementi di una data popolazione si chiamano **individui** o **unità statistiche**.
Esempi di popolazioni statistiche sono i seguenti:

- l'insieme delle persone che in questo istante vivono in Italia (popolazione in senso demografico);
- i cittadini che hanno diritto al voto nelle elezioni per il Parlamento;
- le aziende agrarie della Lombardia;
- le pile elettriche di una certa ditta;
- le precipitazioni atmosferiche giornaliere in una certa località e misurate in una stazione meteorologica;
- le autovetture in circolazione attualmente in Italia;
- i lanci di una moneta durante un certo intervallo di tempo.

Caratteri di una popolazione statistica

Una popolazione statistica si può classificare in vari modi, in base ai diversi **caratteri** (o **argomenti** o **attributi**) dei singoli elementi.
Per esempio:

- le persone viventi attualmente in Italia possono essere classificate secondo l'età, il titolo di studio, lo stato civile ecc.;
- le pile elettriche di una certa ditta si possono classificare secondo il prezzo, la durata, la tensione ecc.;
- le autovetture circolanti in Italia possono essere classificate secondo la cilindrata, la casa costruttrice, la provincia di immatricolazione, l'anno di fabbricazione ecc.

Uno degli scopi della statistica è *lo studio di come si distribuisce una data popolazione, in relazione a uno o più caratteri.*
I caratteri che formano l'oggetto di una rilevazione statistica possono differire per diverse manifestazioni, dette **modalità** del carattere.
I caratteri di una popolazione statistica che sono oggetto di un'indagine statistica possono essere:

- **qualitativi**, come, per esempio, il colore, la nazionalità, la religione, lo stato civile, l'affidabilità, l'attitudine agli studi ecc.

In tal caso, le modalità di un carattere qualitativo saranno espresse da aggettivi. Per esempio:

- il carattere *colore degli occhi* ha le modalità *celesti*, *grigi*, *neri*, . . .;
- il carattere *stato civile* ha le modalità *celibe*, *nubile*, *coniugato*;
- il carattere *titolo di studio* ha le modalità *laurea*, *diploma*, . . .

- **quantitativi**, come, per esempio, la statura, il peso, il numero di stanze di un appartamento, ecc.

Le modalità di un carattere quantitativo saranno, allora, espresse da numeri, che si chiamano anche **valori** del carattere.

Frequenza e intensità

Nello studio di un fenomeno collettivo, il numero che si determina in corrispondenza a una data modalità del suo carattere si chiama **dato statistico**.
Tale dato può avere due significati. Può esprimere:

- *quante volte* si è manifestata quella modalità, e in tal caso si dice **frequenza** di quella modalità;
- *una misura*, e in tal caso si chiama **intensità** di quella modalità.

Le *frequenze* si indicano con la lettera f, munita di un indice (numerico o letterale) che indichi le modalità cui è riferita, e sono sempre rappresentate da un numero naturale: 0, 1, 2, 3, 4, . . .
Le *intensità* si indicano con I, munite di indici, e possono essere rappresentate anche da numeri non interi.

ESEMPI

1 Se in gruppo di N persone si contano 15 italiani, 20 francesi e 7 inglesi, si dirà che le frequenze assolute sono rispettivamente:

$$f_{it} = 15 \qquad f_{fr} = 20 \qquad f_{in} = 7$$

Si noti che, evidentemente, risulta:

$$N = f_{it} + f_{fr} + f_{in}$$

2 Si pongono successivamente su una bilancia tre oggetti e si trova che il primo pesa 5,48 kg, il secondo 9,85 kg e il terzo 7,05 kg. Si dice che l'intensità del carattere *peso* è rappresentata dai numeri sopra indicati, e si scrive:

$$I_1 = 5{,}48 \qquad I_2 = 9{,}85 \qquad I_3 = 7{,}05$$

In statistica si incontrano continuamente delle **tabelle**, che indicano le unità o le intensità corrispondenti ai diversi caratteri considerati o alle diverse modalità esaminate.
Queste tabelle si chiamano **distribuzioni di frequenza** o **di intensità**, in quanto indicano come il totale si suddivide fra le diverse voci.
Tali tabelle si presentano in modo diverso a seconda che si riferiscano a caratteri qualitativi o quantitativi, a distribuzioni di frequenza o di intensità.
Senza entrare in molti dettagli, diamo alcuni esempi di diversi tipi di tabelle.

ESEMPI

1 *Distribuzione di frequenza di modalità qualitative*
Il numero delle persone, che in un certo periodo hanno varcato un dato posto di frontiera, suddiviso per nazionalità costituisce una distribuzione di frequenza (assoluta) che si presenta come è indicato in tabella.

Nazionalità	Numero persone
Italiani	4219
Francesi	967
Inglesi	85
Svizzeri	103
Totale	5374

Il *carattere*, qualitativo, del fenomeno è: nazionalità; le sue *modalità*: italiani, francesi, inglesi, svizzeri; i *dati statistici* sono i numeri: 4219, 967, 85, 103 e tali numeri rappresentano delle *frequenze*.

2 *Distribuzione di intensità di modalità qualitative*
Le quantità esportate da una certa nazione in un dato anno, suddivise per voci, costituiscono una distribuzione di intensità che si presenta come è indicato in tabella.

Voci	Quantità (in q.li)
Prodotti dell'agricoltura	33 710
Materie prime	14 290
Prodotti finiti	7200
Totale	55 200

Il *carattere*, qualitativo, del fenomeno è: materie esportate; le sue *modalità*: prodotti dell'agricoltura, materie prime, prodotti finiti; i *dati statistici* sono i numeri: 33 710, 14 290, 7 200 e tali numeri rappresentano delle *intensità*.

3 *Distribuzione di frequenza di modalità quantitative*
Studiando la distribuzione delle abitazioni in una data regione, secondo il numero delle stanze, si è trovato, il seguente risultato:

Numero delle stanze	Numero delle abitazioni
1	3028
2	3143
3	4712
4	5836
5	4913
6	2724
7	1329
Totale	25 685

Il *carattere* quantitativo del fenomeno è: numero delle stanze; le sue *modalità* sono i numeri: 1, 2, 3, 4, 5, 6, 7; i *dati statistici* sono i numeri: 3028, 3143, 4712, 5836, 4913, 2724, 1329 e tali numeri rappresentano delle *frequenze*.

Le frequenze assolute di due distribuzioni, anche della stessa specie, non sono confrontabili in quanto si riferiscono, in generale, a un diverso numero di casi complessivi.
Questo inconveniente viene superato introducendo il concetto di frequenza relativa, mediante la seguente definizione.

Definizione

Si chiama **frequenza relativa** di una data modalità il rapporto fra la frequenza assoluta di tale modalità e il numero totale dei casi (numero di elementi della popolazione).

Indicando con p la frequenza relativa di una data modalità, si ha:

$$p = \frac{f}{n}$$

Per esempio, nella tabella qui sotto sono indicati il numero degli incidenti sul lavoro, in un dato periodo, che si sono verificati in un certo stabilimento, classificati secondo i diversi giorni lavorativi della settimana. Da questi dati sono state ricavate le frequenze relative indicate nella terza colonna.

Giorni lavorativi	Frequenze assolute f	Frequenze relative p
Lunedì	52	0,208
Martedì	48	0,192
Mercoledì	45	0,180
Giovedì	52	0,208
Venerdì	53	0,212
Totale	250	1,000

3. Le fasi di una ricerca statistica

Un'indagine statistica si svolge attraverso varie fasi, cui vogliamo brevemente accennare.

1. Studio del problema e impostazione della ricerca statistica

Questa prima fase deve precisare lo *scopo* della ricerca, la *definizione del fenomeno* da rilevare, le *ipotesi* che si vogliono provare e l'*individuazione delle unità statistiche* (popolazioni).
Esige inoltre un determinato *grado di precisione*, in relazione agli eventuali strumenti di misura e macchine che vengono usati, nonché alla preparazione del personale addetto ai rilevamenti stessi.

2. Rilevamento, classificazione e tabulazione dei dati statistici

Si chiama **rilevamento statistico** l'operazione di raccolta dei dati statistici riguardanti una determinata popolazione di individui in possesso di un medesimo carattere, relativo al fenomeno che si vuole studiare.
Un rilevamento statistico può avere forme e caratteristiche diverse: può essere *saltuario* o *continuo*, *pubblico* o *privato*, *parziale* o *totale*, ossia può essere eseguito *sull'intera popolazione* (detta anche **universo**) o soltanto *su una parte di essa* (o **campione**).
Per i rilevamenti statistici sono normalmente necessari appositi *registri*, *moduli*, *questionari*, *schede*, ecc. su cui riportare e ordinare tutte le informazioni raccolte sul carattere o i caratteri del fenomeno in esame, per poter, alla fine, costruire le relative *tabelle statistiche*.
A rilevamento compiuto, si mette insieme tutto il materiale raccolto e si procede a *controlli*, di natura diversa, tendenti a eliminare, per quanto è possibile, inesattezze, errori, lacune.
Un'importante verifica, necessaria principalmente quando la rilevazione è parziale, consiste nell'accertare che tutti gli individui siano stati effettiva-

mente e convenientemente rilevati e i registri, prospetti, questionari, schede regolarmente compilati.
Al giorno d'oggi l'uso dell'informatica rende questi procedimenti assai veloci.
In molti casi, si procede a un *controllo* dei dati raccolti, riguardante la loro attendibilità e compatibilità, confrontandoli, per esempio, con risultati già acquisiti o prevedibili. Se da questo esame *critico* emergono aspetti non accettabili, si ricorre talora a una opportuna *indagine suppletiva*.
Nel corso del rilevamento, o dopo, i dati raccolti vengono **classificati**, ossia suddivisi in classi omogenee. Questa operazione ha lo scopo evidente di far conoscere il numero dei casi che si verificano in corrispondenza ai vari caratteri e alle modalità considerate.
Una volta raccolti e classificati, i dati vengono presentati sotto forma di prospetti, le cosiddette **tabelle statistiche**. Esse possono assumere diverse configurazioni e contenere dati di tipo diverso.

3. Diagrammi e analisi statistica dei dati

Dopo la compilazione delle tabelle, occorre procedere alle seguenti operazioni:

- *Traduzione dei dati in diagramma*

  Una tale visualizzazione può suggerire alcune considerazioni che una semplice lettura dei dati non permette di percepire. Nel paragrafo successivo studieremo alcuni tipi di diagrammi statistici.

- *Elaborazione matematica dei dati*

  È la parte più impegnativa e dà luogo alla cosiddetta *metodologia statistica*.

4. Conclusioni dell'indagine

A esse si perviene dallo studio grafico-matematico dei dati raccolti.

4 Rappresentazione grafica dei fenomeni statistici

Lo studio di un fenomeno statistico viene facilitato notevolmente dalle **rappresentazioni grafiche**. Infatti, esse si prestano bene, in generale, a mettere in evidenza certe caratteristiche delle distribuzioni statistiche (come l'andamento complessivo, la forma, le variazioni, ...) che il semplice esame di una tabella non sempre è in grado di mostrare con immediatezza.
I grafici statistici possono assumere varie forme a seconda del tipo di fenomeno studiato.
Vediamo qualche esempio.

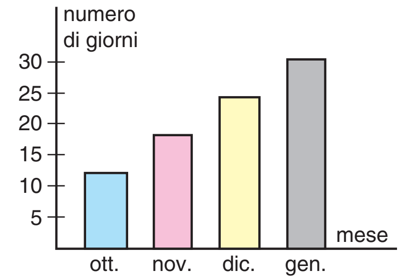
Ortogrammi (o grafici a strisce o a colonne)

Sono grafici costituiti da rettangoli che si innalzano in corrispondenza dei diversi caratteri che si considerano; hanno basi *uguali* e altezze *proporzionali* alle frequenze (o intensità) corrispondenti alle diverse modalità considerate.

ESEMPIO

La tabella presenta dati relativi al *numero mensile di giorni con cielo coperto per quattro mesi, in una certa località*. A fianco è riportato l'*ortogramma* corrispondente.

Mese	Giorni coperti
Ottobre	12
Novembre	19
Dicembre	24
Gennaio	30

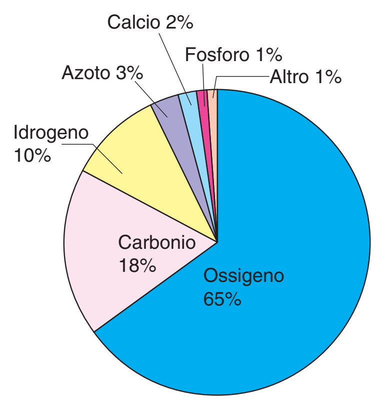


Diagrammi circolari (o areogrammi)

Si ottengono dividendo un cerchio in settori circolari aventi un angolo al centro proporzionale alle frequenze o intensità che rappresentano.

ESEMPIO

È noto che le percentuali dei vari elementi che formano il corpo umano sono: ossigeno (65%), carbonio (18%), idrogeno (10%), azoto (3%), calcio (2%), fosforo (1%) e, in più piccole quantità, ferro, rame, potassio, sodio, iodio ecc.
Mediante un areogramma si dà immediatamente la distribuzione dei vari elementi nel corpo umano.



Ideogrammi

ESEMPIO

Gli ideogrammi sono delle rappresentazioni grafiche di dati tramite figure di varie grandezze. Esempi di ideogrammi si trovano nei testi di geografia per illustrare comparativamente i prodotti (agricoli, di allevamenti e industriali) dei diversi paesi.
In una scuola media di una certa località è stata condotta un'indagine statistica sul numero di studenti iscritti anno per anno, dal 2001 a oggi. I dati sono stati raccolti nella seguente tabella:

Anno scolastico	Popolazione scolastica
2001-02	200
2002-03	300
2003-04	375
2004-05	450
2005-06	480
2006-07	560
2007-08	585

Possiamo rendere efficace la rappresentazione visualizzando la situazione nel modo seguente.

Consideriamo una *figurina umana stilizzata* e assumiamola come **unità grafica**, assegnandole un valore numerico corrispondente, ad esempio, a 50 *alunni*; cioè

 → 50 alunni

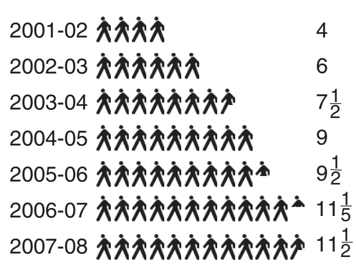
Il numero degli alunni di ogni anno scolastico sarà rappresentato da *tante unità grafiche*, cioè da tante figurine, *quanti sono i gruppi di* 50 *alunni*. Per esempio, nell'anno 2001-02 erano iscritti a quella scuola 200 alunni, cioè quattro gruppi di 50 alunni; il grafico relativo sarà:

Se la divisione del numero degli alunni per 50 dà un resto, questo dovrà essere rappresentato con una parte di unità grafica, corrispondente alla frazione del gruppo non completo.
Per esempio, nell'anno scolastico 2006-07 erano iscritti a quella scuola 560 alunni.

Si ha allora: $560 = 11$ unità grafiche e $\frac{1}{5}$ di unità grafica.

Il grafico relativo sarà dunque:

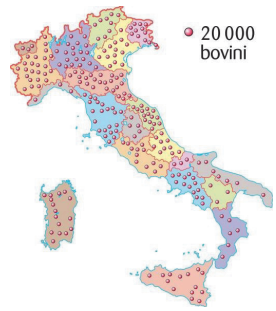
Il grafico completo, relativo a tutti gli anni scolastici considerati, è allora il seguente:

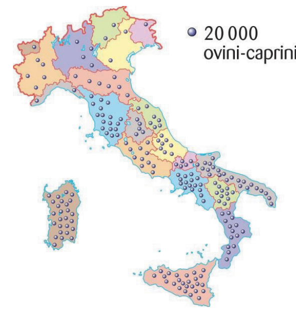


Cartogrammi

Con cartogrammi si intendono delle rappresentazioni grafiche di distribuzioni statistiche territoriali mediante *carte geografiche* relative al territorio considerato, sulle quali le diverse intensità del carattere quantitativo di un fenomeno vengono indicate con segni o colori opportuni, per indicare i valori delle intensità del carattere osservato, nelle singole località o regioni considerate.
La rappresentazione può essere resa di più immediata ed efficace interpretazione visiva dall'uso dei colori, le cui diverse tonalità e gradazioni indicano di solito i vari livelli di intensità.
Ogni cartogramma è accompagnato da tutte le spiegazioni e didascalie necessarie per la sua corretta *lettura* in senso statistico.

ESEMPIO





Questi due cartogrammi sono relativi al numero di bovini e ovini-caprini presenti nelle varie regioni d'Italia.
Nelle varie regioni sono segnati dei puntini: la *didascalia* indica che ognuno di essi rappresenta 20 000 capi di bestiame.
Da tali espressive rappresentazioni grafiche si deducono subito le seguenti interessanti osservazioni, relative a quegli anni:

- l'allevamento dei bovini, che è quello di maggior valore economico, è più sviluppato nelle regioni settentrionali e particolarmente nella pianura padana;
- l'allevamento degli ovini e dei caprini prevale nelle regioni meridionali e nelle isole, con una punta massima in Sardegna;
- nel confronto dei due cartogrammi si ha la conferma di un grave squilibrio tra nord e sud nel grado di intensità dell'economia agraria.

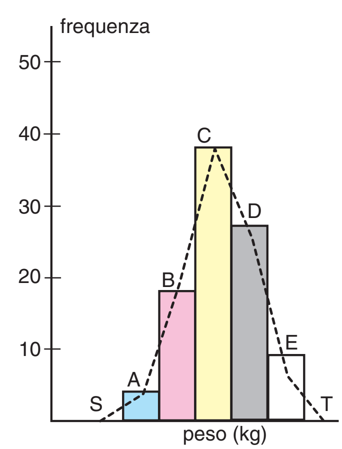
Istogramma

Quando la serie dei valori del carattere è molto numerosa e compresa in un intervallo $[x_1, x_n]$ assai limitato, allora può essere conveniente un'altra rappresentazione grafica.
Prima di tutto si distribuiscono i dati in *classi* e si determina il numero di individui appartenenti a ciascuna classe, detto *frequenza della classe*.
Un ordinamento dei dati in una tabella secondo le classi e secondo le corrispondenti frequenze delle classi, è detta *distribuzione di frequenze*.

ESEMPIO

La tabella riporta la distribuzione di frequenza dei pesi (registrati al chilogrammo più prossimo) di 100 studenti maschi di una data scuola.
A fianco è rappresentato l'istogramma relativo.

Peso (kg)	Numero di studenti
A 60-62	5
B 62-64	18
C 64-66	42
D 66-68	27
E 68-70	8
Totale	100



Una scrittura definente una classe, come 60-62 nell'esempio precedente, è detta *intervallo della classe*. I numeri 60 e 62 sono detti i *limiti* (rispettivamente, *inferiore* e *superiore*) della classe.
Analogamente per le altre classi.
Si chiama *ampiezza* della classe la differenza tra il limite superiore e quello inferiore.
Possiamo allora dire che un **istogramma** consiste in un insieme di rettangoli aventi:
a) *base sull'asse orizzontale, con gli estremi nei limiti di ogni classe*;
b) *aree proporzionali alle frequenze delle classi*.

Poligoni di frequenze

La rappresentazione a istogramma consente di introdurre un'altra rappresentazione di notevole applicazione in statistica.
Consideriamo l'istogramma dell'esempio precedente. Se congiungiamo con dei segmenti i punti medi delle basi superiori dei successivi rettangoli, otteniamo una spezzata che, se è riferita a una distribuzione di frequenze, si chiama *spezzata di frequenze*.
Se aggiungiamo due rettangoli degeneri, di altezza zero, all'inizio e alla fine dell'istogramma, la spezzata si può completare in modo da ottenere un poligono chiuso (nella figura dell'esempio precedente il poligono chiuso è *SABCDET*).
In questo caso è facile vedere che l'area sottesa alla spezzata è eguale all'area complessiva dei rettangoli. Infatti nella figura dell'esempio si vede che, per ciascun rettangolo, l'area inclusa in più dalla poligonale è eguale a quella esclusa rispetto ai rettangoli.
Si perviene così a un principio di *conservazione delle aree* che ha notevole interesse, specie nelle applicazioni pratiche.

5. Medie statistiche

In molti casi, interessa esprimere i diversi valori delle intensità di un fenomeno mediante un **solo** numero, che li possa convenientemente riassumere e che chiamiamo **valore medio**.
Esistono vari tipi di *valori medi*, il cui uso dipende dal particolare tipo di dati statistici in esame e dallo scopo che si vuole raggiungere.
In questo approccio elementare alla statistica, noi ci limiteremo a esaminare solamente tre tipi di valori medi.

Media aritmetica

Consideriamo il seguente esempio.

ESEMPIO

Si vuole controllare la qualità dei televisori a colori prodotti nelle varie catene di montaggio di una data fabbrica, che ha un ritmo produttivo di 1600 televisori giornalieri, cioè di 200 televisori per ciascuna delle 8 ore lavorative.

È risultato che i televisori difettosi sono distribuiti, nelle 8 ore, come esposto nella tabella.

Ora	Televisori difettosi
1ª	7
2ª	5
3ª	5
4ª	8
5ª	6
6ª	9
7ª	7
8ª	9
Totale	56

L'intera distribuzione di frequenza dei pezzi difettosi nelle 8 ore dà un'idea confusa del fenomeno che si vuole studiare. In un caso come questo, per esprimere sinteticamente il fenomeno si fa ricorso a un procedimento di questo tipo: si sommano tutti i pezzi difettosi costruiti nelle 8 ore e si divide tale somma per il numero delle ore:

$$\frac{7+5+5+8+6+9+7+9}{8} = \frac{56}{8} = 7$$

Si vede facilmente che il numero 7 sostituito ai singoli valori di pezzi difettosi, dati dalla tabella, *ne lascia inalterata la somma*:

$$7 \times 8 = 56$$

Per questo motivo si dice che 7 è la *media aritmetica* di tali valori.
Si osservi che il numero effettivo dei televisori difettosi che si sono presentati nelle diverse ore non si discosta molto dalla media aritmetica.
Possiamo perciò dire che *in media* si sono presentati 7 televisori difettosi per ogni ora di lavoro.

Diamo allora la seguente definizione.

Definizione

Si chiama **media aritmetica** (semplice) di n numeri $x_1, x_2, \dots, x_n$, il numero M che si ottiene dividendo la loro somma per il numero n. In simboli:

$$M = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \tag{1}$$

Ad esempio, la media aritmetica dei numeri 10, 12, 5, 3, 8 è:

$$M = \frac{10+12+5+3+8}{5} = \frac{38}{5} = 7{,}6$$

Per vedere a quale concetto intuitivo corrisponde il numero M, consideriamo, insieme alla successione $x_1, x_2, x_3, \dots, x_n$ anche la seguente successione di uguale numero n di termini:

$$M, M, M, \dots, M \tag{2}$$

È facile vedere, tramite la (1), che la somma dei termini $x_1, x_2, \ldots, x_n$ è uguale alla somma dei termini della (2), cioè che risulta:

$$x_1 + x_2 + \ldots + x_n = n \cdot M \tag{3}$$

Pertanto, il numero M rappresenta *il valore costante* che dovrebbe avere ciascuno dei numeri $x_1, x_2, \ldots, x_n$ *perché la somma complessiva* dei loro valori *resti invariata*.
Ne segue che la media aritmetica viene usata in tutti quei problemi in cui deve rimanere costante la somma dei valori $x_1, x_2, x_3, \ldots, x_n$ indipendentemente dal modo con cui questa somma viene ripartita in n parti.
Così la media aritmetica dei redditi di n persone dice quale reddito avrebbe ognuno, se mettessero insieme tutti i loro redditi e ne dividessero la somma in n parti uguali.
Così pure, si userà la media aritmetica per determinare la statura media di un gruppo di persone, o per trovare il consumo medio di un dato prodotto, e così via.
La media aritmetica ha una vastissima diffusione, sia in teoria sia in pratica, ed è per questo che quando si parla semplicemente di media si vuole alludere proprio alla media aritmetica.

FOCUS

La media aritmetica è un valore *sintetico* dell'insieme di n valori $x_1, \ldots, x_2, \ldots, x_n$, che permette di effettuare confronti con medie di altre distribuzioni espresse nella stessa unità di misura. Pertanto si può affermare, per esempio, che: «il salario medio della ditta A è maggiore di quello della ditta B»; «il numero medio di radioascoltatori della rete A è minore di quello della rete B», ecc.

Più in generale, se in una distribuzione di *frequenze* il valore x_1 compare con la frequenza f_1, il valore x_2 con la frequenza f_2, …, il valore x_k con la frequenza f_k, in modo che risulti:

$$f_1 + f_2 + \ldots + f_k = n$$

anziché calcolare la media aritmetica applicando la (9) e scrivendo:

$$M = \frac{\overbrace{(x_1 + x_1 + \ldots + x_1)}^{f_1 \text{ volte}} + \overbrace{(x_2 + x_2 + \ldots + x_2)}^{f_2 \text{ volte}} + \ldots + \overbrace{(x_k + x_k + \ldots + x_k)}^{f_k \text{ volte}}}{n}$$

si può scrivere, più semplicemente:

$$M = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_1 + f_2 + \ldots + f_k} \tag{4}$$

La media aritmetica espressa dalla (4) si chiama **media aritmetica ponderata** dei k valori $x_1, x_2, \ldots, x_k$ di **pesi** rispettivi $f_1, f_2, \ldots, f_k$.

ESEMPIO

Determinare l'età media dei 30 componenti di una banda musicale giovanile, così ripartiti:

Età (anni)	$x_1 = 14$	$x_2 = 15$	$x_3 = 16$	$x_4 = 17$	$x_5 = 18$	$x_6 = 19$
Numero	$f_1 = 4$	$f_2 = 12$	$f_3 = 8$	$f_4 = 4$	$f_5 = 1$	$f_6 = 1$

Si tratta, evidentemente, di una media aritmetica ponderata e quindi applichiamo la relazione (4):

$$M = \frac{14 \cdot 4 + 15 \cdot 12 + 16 \cdot 8 + 17 \cdot 4 + 18 \cdot 1 + 19 \cdot 1}{30} = \frac{469}{30} = 15^{\mathrm{a}}\, 7^{\mathrm{m}}\, 18^{\mathrm{g}}$$

cioè l'età media (ponderata) degli alunni della classe è 15 anni, 7 mesi e 18 giorni.

FOCUS

Particolare attenzione va posta quando si debba calcolare la media aritmetica ponderata su una distribuzione di frequenze di un carattere quantitativo, *le cui intensità siano suddivise in* n *classi*.
In questo caso, i valori $x_1, x_2, \ldots, x_n$ devono essere convenientemente scelti entro ciascuna classe di intensità. Se, ad esempio, si hanno ragioni per ritenere che la distribuzione delle frequenze, entro ogni singola classe, sia *uniforme*, allora si dovranno scegliere come $x_1, x_2, \ldots, x_n$ i valori *centrali* di ciascuna classe.

ESEMPIO

Calcolare la statura media di un gruppo di giovani raggruppati secondo l'altezza in classi di ampiezza uniforme (5 cm), e le cui relative distribuzioni di frequenza sono riportate nella tabella.

Classe di statura (cm)	Frequenza	Valore centrale	Prodotto $x_1 f_1$
155-160	0	157,5	0
160-165	20	162,5	3250
165-170	52	167,5	8710
170-175	30	172,5	5175
175-180	12	177,5	2130
180-185	1	182,5	182,5
Totali	115	–	19 447,5

Le classi di altezza sono di 5 cm e possiamo supporre in ciascuna classe una *distribuzione uniforme* delle altezze osservate.
Attribuiremo così a ciascuna classe di altezza *una sola altezza*, e precisamente l'*altezza centrale* nella classe (riportata nella terza colonna della tabella).
L'altezza media richiesta è pertanto la media aritmetica ponderata dei valori:

$$x_1 = 157{,}5 \quad x_2 = 162{,}5 \quad x_3 = 167{,}5$$

$$x_4 = 172,5 \qquad x_5 = 177,5 \qquad x_6 = 182,5$$

con i pesi rispettivi:

$$f_1 = 0 \qquad f_2 = 20 \qquad f_3 = 52 \qquad f_4 = 30 \qquad f_5 = 12 \qquad f_6 = 1$$

Applichiamo la formula (4):

$$M = \frac{19\,447{,}5}{115} = 169{,}11$$

Moda o valore normale

Un altro importante tipo di media usato in statistica è la *moda*, o *valore normale*, così definita:

Definizione

Data una distribuzione di frequenze di un carattere quantitativo, si chiama **moda** o **valore normale** ogni valore del carattere al quale corrisponde la massima frequenza.

Il valore normale è quindi individuato esaminando la tabella relativa alla distribuzione dei dati; esso è dato dal valore del carattere al quale corrisponde la maggior frequenza.

ESEMPIO

Se la distribuzione delle abitazioni, costruite in una certa regione nel 2002, secondo il numero delle stanze è quella riportata nella tabella, si vede subito che il *valore normale* è 4, perché è questo il valore al quale corrisponde la massima frequenza.
Possiamo quindi dire che gli appartamenti costruiti in quella regione nel 2002 hanno *normalmente* 4 stanze.

Numero delle stanze (intensità)	Numero delle abitazioni (frequenza)
1	12 009
2	35 148
3	73 455
4	127 312
5	109 414
6	70 412
7	28 347

Dall'esempio precedente si può comprendere come in molti fenomeni collettivi possa essere più espressivo il *valore normale* di altri valori medi.
Infatti, per esempio è più espressivo dare il *salario normale* di un gruppo di lavoratori, che non il *salario medio aritmetico*, perché quest'ultimo può essere influenzato da retribuzioni molto alte di un piccolo numero di lavoratori del gruppo. In altre parole, il salario normale dà un'idea immediata dell'intensità di retribuzione di cui *normalmente* godono quei lavoratori.
Analogamente, per un calzaturificio sarà più importante conoscere la lunghezza normale del piede dei suoi possibili clienti, che non la lunghezza media aritmetica.
È bene osservare che la definizione data non esclude che uno stesso carattere quantitativo possa presentare *due* o *più mode*. Cioè, ci possono essere due valori diversi del carattere che presentano entrambi la stessa frequenza massima.
Infine, si osservi che la definizione di *moda* si applica solo ai fenomeni collettivi di *carattere quantitativo* i cui corrispondenti dati statistici sono delle *frequenze*.

Mediana o valore centrale

Un terzo importante tipo di media usato in statistica è la *mediana* o *valore centrale*, così defnita.

Definizione

Data una distribuzione di valori di un carattere quantitativo ordinati dal più piccolo al più grande (o viceversa), si chiama **mediana** o **valore centrale** della distribuzione il valore che vi occupa la posizione di mezzo, ossia la posizione centrale.

Quando il numero dei valori numerici è dispari è facile individuare la mediana; se invece il numero dei valori è pari, solitamente si conviene di assumere come mediana la *semisomma* dei due valori centrali.
Ad esempio, la distribuzione:

$$3 \qquad 10 \qquad 12 \qquad 20 \qquad 40 \qquad 42$$

presenta i due valori centrali 12 e 20 e si assume perciò come mediana la loro semisomma, che è 16, anche se 16 non è un elemento della distribuzione.
Anche la definizione di *mediana* si applica solo ai fenomeni collettivi di *carattere quantitativo*.

Confronto fra i tre valori medi

Consideriamo la seguente tabella relativa alla paga annua netta dei dipendenti di una fabbrica.

Paga annua (euro)	Numero delle persone che ricevono tale paga
120 000	1
50 000	1
35 000	2
27 500	3
20 500	19
17 000	22
13 000	2

Il direttore generale che percepisce € 120 000 all'anno afferma: «La paga annua nella nostra fabbrica è di € 22 240».
Il sindacato che rappresenta i lavoratori della fabbrica dice che il salario *medio* è di € 17 000.
L'agente delle tasse, che rappresenta l'Ufficio delle Imposte, dice che il salario *medio* è di € 20 500.
Si noti che i tre valori citati come salario medio si basano su tre concetti diversi di "valore medio". Si ha infatti:

media aritmetica = € 22 240
moda o valore normale = € 17 000
mediana o valore centrale = € 20 500

Vediamone l'interpretazione.

La *media aritmetica* indica che, se il denaro fosse stato distribuito in modo che ciascuno ricevesse la stessa somma, ciascun dipendente avrebbe avuto € 22 240.
La *moda* ci dice che la paga annua più comune è di € 17 000.
La *mediana* indica che circa metà dei dipendenti hanno meno di € 20 500 e metà ne hanno di più.
Si deve porre ben attenzione all'esempio svolto che dimostra come la media aritmetica, la moda e la mediana siano numeri diversi, che rappresentano cose diverse.
Così quando si parla di *valore medio* bisogna saper con precisione a quale tipo di media ci si riferisce e inoltre bisogna sapere bene quali problemi statistici richiedono un tipo di media piuttosto che un'altra.

La *media aritmetica* si usa, come abbiamo già visto, in molte situazioni:

- in meteorologia, per ottenere la temperatura media o la caduta media di precipitazioni;
- in medicina, per scoprire la durata media di una malattia;
- in antropologia, per scoprire certe caratteristiche medie di un insieme di esseri umani (altezza, ...);
- in economia, per calcolare salari medi, prezzi, ecc.

La *moda*, che si considera come il valore più tipico di tutto un insieme, non tiene conto però degli altri valori dei dati. È facile da determinare, ma spesso in un insieme di dati vi è più di un valore che corrisponde alla definizione di moda.
Tuttavia, anche se la moda è la media meno utile in statistica, qualche volta solo la moda è il valore medio appropriato.
Per esempio, un fabbricante di capi di vestiario sportivi è attrezzato per fare una sola taglia di camicie da uomo e deve scegliere la taglia. Egli decide sulla base della media aritmetica delle taglie delle camicie acquistate dagli uomini. Per vendere più camicie doveva invece scegliere la *moda*, cioè la taglia più comune.

La *mediana* è una valore intermedio e non è influenzato dagli altri valori dei dati, ma soltanto dal fatto che essi siano sotto o sopra il centro dell'insieme dei dati; cioè, se un valore è sotto la media, non importa se è appena al di sotto o se è molto lontano.
La mediana si usa in molte ricerche statistiche, fra cui le seguenti:

- nelle assicurazioni, per provare la lunghezza media della vita;
- nello studio dei medicinali, per misurare l'efficacia media di un medicinale;
- nell'industria, per controllare la qualità di certi prodotti.

6. Probabilità

Si è soliti far risalire la nascita della **teoria della probabilità** al tentativo di dare risposta agli interrogativi che i giocatori d'azzardo avevano per assicurarsi la vincita nel gioco.
Un aneddoto racconta che alcuni giocatori chiesero a Galileo perché gettando tre dadi la somma dei numeri è più spesso 10 che 9.
Nel 1654 un giocatore d'azzardo, il cavaliere di Méré, chiese al matematico

francese Blaise Pascal, quante volte sarebbe uscito un *doppio sei* gettando 24 volte una coppia di dadi.
La soluzione di questi e altri problemi diede origine al calcolo delle probabilità.
Il primo che fece di questo argomento un vero e proprio ramo di ricerca matematica fu lo svizzero Giacomo Bernoulli (1713), mentre la prima sistemazione rigorosa e moderna della nuova scienza fu opera del francese Pierre-Simon de Laplace (1812) nella sua famosa *Teoria analitica della probabilità*.
Al giorno d'oggi, le leggi matematiche della probabilità sono diventate un fattore importante nella nostra vita: il politico le usa per progettare la sua campagna elettorale usando i sondaggi di opinione con le relative rielaborazioni, i genetisti la usano per prevedere la frequenza relativa con cui varie caratteristiche ricorrono in gruppi di individui, gli esperti in telecomunicazioni se ne servono per calcolare la densità del traffico telefonico ecc.
Possiamo, quindi, affermare che oggi le nozioni di probabilità (e di statistica, disciplina gemella), sono presenti in quasi tutte le scienze: in fisica, chimica, biologia, medicina, psicologia, sociologia, scienze politiche, pedagogia, economia, scienze amministrative, ricerca operativa e in tutti i campi dell'ingegneria.

7. Eventi certi, impossibili e casuali

Ognuno conosce l'importanza degli esperimenti nella scienza e nella tecnologia e il fondamentale principio secondo il quale ogni volta che si realizza un insieme di condizioni $\mathcal{C}$ si richiede che si presenti l'evento A.
Ad esempio, quando alla pressione di 1 atm si riscalda dell'acqua a 100 °C (insieme $\mathcal{C}$ delle condizioni), essa si trasforma in vapore acqueo (evento A).
Oppure, data una qualunque reazione chimica fra sostanze che avvenga senza scambi con l'ambiente circostante (insieme delle condizioni $\mathcal{C}$), la quantità totale di sostanza si conserva (evento A).
Un evento che si presenta senza alcuna incertezza, per ogni realizzazione dell'insieme di condizioni $\mathcal{C}$, si dice **certo**.
Un evento che, data la realizzazione di un insieme di condizioni $\mathcal{C}$, non si realizza mai, si dice **impossibile**.
Un evento A che, data la realizzazione di un insieme di condizioni $\mathcal{C}$, può accadere oppure no, si dice **casuale** o **aleatorio**. Ad esempio:

- il numero delle chiamate di servizio di un'ambulanza non è prevedibile con esattezza. Tale numero è soggetto a fluttuazioni importanti e aleatorie;
- il risultato «tre volte testa» nel triplice lancio di una moneta è aleatorio;
- nel tiro a segno si osserva una dispersione di proiettili. Non è possibile indicare a priori lo scarto tra il punto d'incontro del proiettile e il centro dell'obiettivo: tale scarto è una grandezza aleatoria;
- estrarre una pallina rossa da un'urna contenente tredici palline bianche e quattro rosse, è un evento aleatorio.

La semplice constatazione del carattere casuale (o aleatorio) di un evento è di interesse abbastanza limitato: dice semplicemente che l'insieme delle condizioni $\mathcal{C}$ non è sufficiente a garantire la realizzazione o meno dell'evento A.

SAPEVI CHE?

- *Alea*, in latino, significa «dado». L'origine storica del calcolo delle probabilità (i giochi con i dadi), ha lasciato la sua traccia nell'aggettivo «aleatorio», molto usato in matematica come sinonimo di «casuale».

Vi è tuttavia una vasta serie di fenomeni nei quali, dato il *ripetuto* realizzarsi dell'insieme di condizioni C, le percentuali dei casi in cui l'evento A si verifica tende in modo apprezzabile a qualche valore medio.
Per fenomeni di questo tipo si può non solo affermare semplicemente la casualità dell'evento A, ma anche fornire una stima approssimativa della possibilità che esso si verifichi.
Come esprimere quantitativamente questa stima è il compito che ci prefiggiamo.

8. Spazio delle probabilità Eventi

Con il termine *prova* (o *tentativo*) s'intende una singola esecuzione di un ben determinato esperimento.
Da questa prova si ottiene un singolo *risultato elementare* (o *eventualità* o *campione*).

Definizione

> Si chiama **spazio delle probabilità**, associato a un dato esperimento, l'insieme S di tutti i possibili risultati.
> Si chiama **evento** un sottoinsieme A dello spazio S, cioè un insieme di risultati possibili.

- $A \subseteq S$

Un evento si dice:

- **elementare**, se è un insieme con un solo elemento;
- **certo**, se coincide con S;
- **impossibile**, se è l'insieme vuoto $\varnothing$.

Diremo che in una prova **si verifica** (o si realizza) l'evento A, se il risultato a della prova appartiene ad A; cioè se $a \in A$.
In caso opposto si dice che in quella prova *non* si verifica l'evento.
Un risultato che verifica un evento si dice *favorevole* all'evento.
I risultati che, verificandosi in una prova, realizzano l'evento A, si chiamano *modalità* dell'evento A.
Dal momento che gli eventi sono insiemi, ogni affermazione relativa a eventi può essere espressa con il linguaggio degli insiemi, e viceversa. In particolare, avremo un'*algebra degli eventi* corrispondente all'algebra degli insiemi.
Usando le operazioni tra insiemi su eventi di S si possono ottenere nuovi eventi di S.
Così, se A e B sono due eventi (appartenenti a una medesima prova), si chiama:

- **evento somma** di A e B, e si indica con $A \cup B$, l'evento C che risulta dal verificarsi di **almeno** uno degli eventi A o B;
- **evento prodotto** di A e B, e si indica con $A \cap B$, l'evento C che risulta dal verificarsi di **entrambi** gli eventi A e B;
- **evento differenza** tra A e B, e si indica con $A-B$, l'evento che consiste nel fatto che accada l'evento A ma non l'evento B *(fig. 1)*.

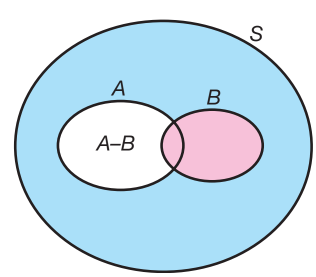


figura 1

Inoltre:

- due eventi A e $\overline{A}$ si dicono **opposti** (o complementari, o contrari) quando uno di essi non si verifica se e solo se si verifica l'altro *(fig. 2)*.

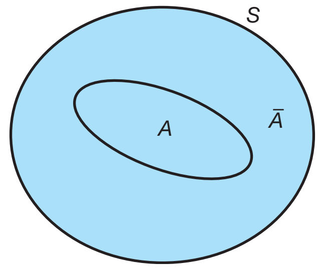


figura 2

Evidentemente:

$$\overline{S} = \emptyset \qquad \overline{\emptyset} = S \qquad A \cup \overline{A} = S \qquad A \cap \overline{A} = \emptyset$$

- due eventi A e B si dicono **incompatibili** se sono disgiunti, cioè se risulta $A \cap B = \emptyset$. In altre parole, A e B sono incompatibili se non possono verificarsi simultaneamente;
- se risulta $A \subseteq B$, si dice che A **implica** B.

ESEMPI

1 È dato l'esperimento «si getti un dato e si osservi il numero che si presenta».

Lo spazio delle probabilità è: $S = \{1, 2, 3, 4, 5, 6\}$.

Sia A l'evento «si presenta un numero pari», B l'evento «si presenta un numero dispari», C l'evento «si presenta un numero primo», D l'evento «si presenta un numero minore di 7» ed E l'evento «si presenta un numero divisibile per 7». Si ha:

$$A = \{2, 4, 6\} \qquad B = \{1, 3, 5\} \qquad C = \{2, 3, 5\} \qquad D = S \qquad E = \emptyset$$

L'evento D è *certo*, l'evento E è *impossibile*, gli eventi A, B e C hanno tre modalità ciascuno. Risulta inoltre:

- $A \cup C = \{2, 3, 4, 5, 6\}$, che è l'evento «si presenta un numero pari o primo»;
- $B \cap C = \{3, 5\}$, che è l'evento «si presenta un numero dispari primo»;
- $\overline{C} = \{1, 4, 6\}$, che è l'evento che «non si presenta un numero primo».

Si noti che A e B sono *incompatibili*: $A \cap B = \emptyset$; in altri termini, un numero pari e un numero dispari non possono presentarsi simultaneamente.

2 Nello stesso esperimento dell'esercizio precedente, sia A l'evento «si presenti il numero 6», B l'evento «si presenta il numero 3», C l'evento «si presenta un numero pari», D l'evento «si presenta un numero multiplo di 3».

Abbiamo:

$$A = \{6\} \qquad B = \{3\} \qquad C = \{2, 4, 6\} \qquad D = \{3, 6\}$$

Gli eventi A e B sono elementari e risulta:

$$A \subseteq C \qquad A \subseteq D \qquad B \subseteq D \qquad A \cup B = D \qquad C \cap D = A$$

3 È dato l'esperimento «si lanci tre volte una moneta e si osservino le sequenza di teste (T) e croci (C) che si presentano».
Lo spazio S è dato da:

$$S = \{TTT, TTC, TCT, TCC, CTT, CTC, CCT, CCC\}$$

Sia A l'evento «due o più teste si presentano consecutivamente» e B l'evento «tutti i lanci diano lo stesso risultato», cioè:

$$A = \{RRR, RRC, CTT\} \qquad B = \{TTT, CCC\}$$

Risulta $A \cap B = \{TTT\}$, che è l'evento elementare «si presenta per tre volte testa». L'evento «si presentano 5 teste» è l'insieme vuoto, cioè l'evento è impossibile.

4 È dato l'esperimento «si lanci due volte una moneta».

Dati gli eventi $A = \{TC, CT, TT\}$ e $B = \{TC, CC\}$, risulta:

$$A \cup B = \{TC, CT, TT, CC\} = S \qquad A \cap B = \{TC\}$$

$$\overline{A} = \{CC\} \qquad A - B = \{CT, TT\}$$

9. Definizione classica di probabilità

Ancora nei primi decenni del secolo, la teoria delle probabilità era fondata sulla cosiddetta definizione classica.
Sebbene la sua importanza quale fondamento di una teoria a carattere deduttivo sia diminuita, essa viene ancora usata per la determinazione dei dati probabilistici e come ipotesi di lavoro.
La definizione classica riconduce il concetto di probabilità a quello di *uguale possibilità* (o *equiprobabilità*) di eventi, che viene considerato come assioma e quindi non necessita di una definizione formale.
Per esempio, nel lancio di un dado, perfettamente regolare e omogeneo, eventi ugualmente possibili sono i lanci i cui risultati sono i numeri 1, 2, 3, 4, 5, 6 segnati sulle facce del dado, visto che, per ragioni di simmetria, nessuna faccia risulta privilegiata rispetto alle altre.
In questa ipotesi, la probabilità $p = P(A)$ di un evento A viene determinata *a priori*, senza far ricorso ad alcuna effettiva prova sperimentale.
Si pone, infatti, la seguente definizione.

Definizione

Si chiama **probabilità** p di un evento aleatorio, previsto da una determinata prova, il quoziente fra il numero dei risultati favorevoli e il numero dei risultati possibili della prova, nell'ipotesi che siano tutti egualmente possibili.

Indicando con A l'evento, con $p = P(A)$ la sua probabilità, con n il numero dei risultati possibili e con m il numero dei risultati favorevoli all'evento, possiamo scrivere:

$$P(A) = p = \frac{m}{n}$$

È evidente che la probabilità di un evento aleatorio è un numero *non negativo* e che non supera l'unità.

Inoltre:
- se è $m = 0$, $p = 0$: l'evento è *impossibile*
- se è $m = n$, $p = 1$: l'evento è *certo*

Inversamente:
- se l'evento è *impossibile*, si ha $m = 0$, e quindi $p = 0$
- se l'evento è *certo*, si ha $m = n$ e quindi $p = 1$

I numeri 0 e 1 si dicono, rispettivamente, le misure della *impossibilità* e della *certezza*.

ESEMPI

1 Un'urna contiene soltanto 7 palline bianche e 5 rosse, tutte uguali e non distinguibili che per il colore. Calcolare le probabilità che, estraendo a caso una sola pallina dall'urna, si verifichino i seguenti eventi:

a) E_1: «pallina bianca»;
b) E_2: «pallina rossa»;
c) E_3: «pallina bianca o rossa».

In base alla definizione di probabilità, si ha:

a) la probabilità che la pallina sia bianca è: $P(E_1) = \dfrac{7}{7+5} = \dfrac{7}{12}$;

b) la probabilità che la pallina sia rossa è: $P(E_2) = \dfrac{5}{7+5} = \dfrac{5}{12}$;

c) la probabilità che la pallina sia bianca o rossa, indifferentemente, è: $P(E_3) = \dfrac{7+5}{7+5} = 1$. Perciò l'evento E_3 è certo.

2 Calcolare la probabilità che, lanciando una volta due dadi regolari, la somma dei numeri presentati dalle facce superiori sia minore di 5.

Il numero dei casi possibili in un lancio è $n = 6^2 = 36$, e sono dati nella seguente tabella.

(1; 1), (1; 2), (1; 3), (1; 4), (1; 5), (1; 6)
(2; 1), (2; 2), (2; 3), (2; 4), (2; 5), (2; 6)
(3; 1), (3; 2), (3; 3), (3; 4), (3; 5), (3; 6)
(4; 1), (4; 2), (4; 3), (4; 4), (4; 5), (4; 6)
(5; 1), (5; 2), (5; 3), (5; 4), (5; 5), (5; 6)
(6; 1), (6; 2), (6; 3), (6; 4), (6; 5), (6; 6)

Il numero dei casi favorevoli all'evento è $m = 6$, e precisamente:

(1; 1) somma = 2 (1; 2) somma = 3
(2; 1) somma = 3
(1; 3) somma= 4 (2; 2) somma= 4
(3; 1) somma = 4

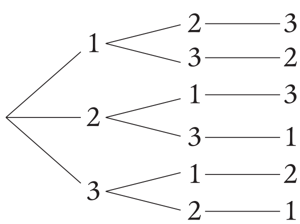
La probabilità cercata è dunque:

$$p = \frac{6}{36} = \frac{1}{6}$$

3 Un'urna contiene tre gettoni numerati con 1, 2, 3. Li si estrae a caso uno dopo l'altro dall'urna e si studia il problema delle coincidenze, cioè quando il numero del gettone coincide con il numero d'ordine della sua estrazione.

Per esempio, se si estraggono, nell'ordine 1, 3, 2, vi è una sola coincidenza (1 è al suo posto, il primo); se, invece, si estraggono 3, 1, 2, non vi è nessuna coincidenza.

Per risolvere il problema, costruiamo il grafico ad albero di tutti i 6 casi possibili.



Ordine di estrazione			*Numero delle coincidenze*
1	2	3	3
1	3	2	1
2	1	3	1
2	3	1	0
3	1	2	0
3	2	1	1

Dal grafico deduciamo la seguente tabella.

Numero delle coincidenze	0	1	2	3
Numero dei casi	2	3	0	1

Con facilità si possono calcolare le probabilità.

La probabilità di avere una coincidenza è dunque $\frac{3}{6} = \frac{1}{2}$; le probabilità di avere due coincidenze è $\frac{0}{6} = 0; \ldots$

FOCUS

Facciamo notare ancora una volta che, quando si deve calcolare la probabilità di un evento aleatorio, si deve essere ben sicuri che i casi possibili della prova siano tutti *ugualmente possibili*. Ad esempio, se in un'urna ci sono 20 palline differenti solo per il colore, delle quali 5 sono bianche e 15 rosse, e se con un opportuno dispositivo si può far uscire dal fondo dell'urna una pallina, allora la probabilità che esca una pallina bianca è $\frac{5}{20} = \frac{1}{4}$, perché tutti i 20 casi possibili sono ugualmente possibili.
Ma se le palline, pur avendo la stessa forma, sono costituite da materiali diversi, leggerissimo quello delle palline bianche e pesante quello delle rosse, allora non si può dire che la probabilità di estrarre una pallina bianca sia $\frac{1}{4}$, perché, mescolando le palline, si disporranno sul fondo dell'urna in prevalenza le rosse anziché le bianche, e quindi tutti i 20 casi possibili *non* sono ugualmente possibili.

Probabilità contraria

Abbiamo detto che dato un evento aleatorio A, si chiama **evento contrario** di A l'evento consistente nel **non verificarsi di** A.
L'evento contrario di un evento A si indica con il simbolo $\overline{A}$, (*non* A).
Premesso ciò, se p è la probabilità di un evento aleatorio A, riferita a una determinata prova, e indichiamo con q la probabilità dell'evento $\overline{A}$, contrario di A, si prova facilmente che risulta:

$$p + q = 1$$

Infatti, se n indica il numero dei casi possibili ed m quello dei casi favorevoli all'evento A, la differenza $n - m$ rappresenta il numero dei *casi sfavorevoli* ad A e quindi favorevoli all'evento contrario $\overline{A}$.
La probabilità q di $\overline{A}$, detta **probabilità contraria ad** A, in base alla definizione è il numero:

$$q = \frac{n - m}{n}$$

Da $q = \dfrac{n - m}{n}$ si ricava:

$$q = \frac{n}{n} - \frac{m}{n} = 1 - p \quad \text{da cui:} \quad p + q = 1$$

La relazione $p + q = 1$ lega la probabilità p di un evento aleatorio alla probabilità q dell'evento contrario ed esprime il fatto che **la somma della probabilità di un evento aleatorio e della probabilità contraria è sempre uguale a 1**.

10. Proprietà additiva della probabilità

Valgono le seguenti importanti proprietà, espresse sotto forma di teoremi.

Teorema 1

Se A e B sono due eventi **incompatibili**, si ha:

$$P(A \cup B) = P(A) + P(B) \tag{1}$$

Dimostrazione

Siano $P(A) = \dfrac{m_1}{n}$ e $P(B) = \dfrac{m_2}{n}$.

Poiché, per ipotesi, gli eventi A e B sono incompatibili, gli m_1 risultati favorevoli ad A sono diversi dagli m_2 risultati favorevoli a B.
Vi sono dunque $m_1 + m_2$ risultati favorevoli a che si verifichi uno degli eventi A e B, cioè favorevoli a che si verifichi l'evento $A \cup B$.
Pertanto:

$$P(A \cup B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B)$$

Questo teorema è detto **proprietà additiva** (o **dell'addizione**) **della probabilità**.

Teorema 2

Se A e B sono due eventi **qualsiasi**, risulta:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Dimostrazione

Scriviamo gli eventi $A \cup B$ e B come somma di eventi incompatibili *(fig. 3)*:

$$A \cup B = A \cup \left(\overline{A} \cap B\right) \qquad B = (A \cap B) \cup \left(\overline{A} \cap B\right)$$

e applichiamo il teorema 1:

$$P(A \cup B) = P(A) + P\left(\overline{A} \cap B\right) \quad \text{e} \quad P(B) = P(A \cap B) + P\left(\overline{A} \cap B\right)$$

Eliminando $P\left(\overline{A} \cap B\right)$ tra queste due eguaglianze si ottiene la relazione cercata:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

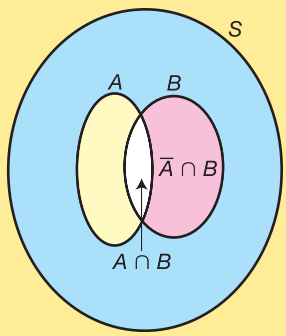


figura 3

11. Legge empirica del caso

Sottoponiamo un evento aleatorio E a un determinato numero n di prove, tutte eguali ed eseguite nelle medesime condizioni, e sia v il numero dei **successi**, cioè il numero delle prove in cui si presenta l'evento R.
Il numero v delle prove riuscite per l'evento E si chiama *frequenza assoluta* dell'evento, nelle n prove eseguite.
Agli effetti del nostro studio, più utile e importante della frequenza assoluta è la nozione di *frequenza relativa*, detta senz'altro *frequenza*, di cui si dà la seguente definizione.

Definizione

Si dice **frequenza relativa**, o semplicemente **frequenza**, di un evento aleatorio E, riferita a n prove, il quoziente tra il numero v dei successi e il numero n delle prove fatte.

Così, se indichiamo con f la frequenza relativa, abbiamo, per definizione:

$$f = \frac{v}{n}$$

Naturalmente, il numero v dei successi non può che assumere uno dei seguenti valori:

$$0, \quad 1, \quad 2, \quad \ldots, \quad (n-1), \quad n$$

e, in corrispondenza di tali valori, la frequenza f vale:

$$0, \quad \frac{1}{n}, \quad \frac{2}{n}, \quad \ldots, \quad \frac{n-1}{n}, \quad \frac{n}{n} = 1$$

La frequenza è dunque un numero certo, nel senso che, su n prove eseguite, assume uno e uno solo di questi valori; ma essa varia, in generale, con il variare del numero n delle prove, e varia pure se si ripete una seconda volta lo stesso numero di prove.

ESEMPIO

Se lanciando una moneta:

a) nei primi 100 lanci si sono avuti 38 successi corrispondenti all'evento «croce», la frequenza f_1, relativa a queste prove, è:

$$f_1 = \frac{38}{100} = 0{,}38$$

b) in una seconda serie di 100 lanci si sono avuti 56 successi, la frequenza f_2, relativa a questo gruppo di prove, è:

$$f_2 = \frac{56}{100} = 0{,}56$$

c) in una terza serie di 1500 lanci si sono avuti 709 successi, la frequenza f_3 relativa è:

$$f_3 = \frac{709}{1500} = 0{,}47 \text{ circa}$$

Intuitivamente siamo portati a pensare che la probabilità p di un evento aleatorio ci possa fornire un'idea sufficientemente esatta di quella che può essere la frequenza (relativa) di quell'evento.
Così, ad esempio, se nel lancio di un dado regolare consideriamo l'evento «faccia contrassegnata col numero 4», di probabilità $\frac{1}{6}$, pensiamo che la frequenza di questo evento, su un gran numero di prove, sia circa $\frac{1}{6}$, e quindi che, ad esempio, su 6000 prove la faccia contrassegnata con il numero 4 esca circa 1000 volte.

Ora, l'esperienza conferma questa nostra intuizione.

Precisamente, sia E un evento aleatorio, di probabilità nota p, che possa venire sottoposto a un numero qualsiasi di prove uguali e nelle medesime condizioni, e si supponga che in tutte queste prove la sua probabilità p resti costante.
In queste ipotesi, e per una classe vastissima di eventi aleatori, si è constatato che, se n è abbastanza elevato, la differenza tra la frequenza relativa $\frac{v}{n}$ e la probabilità p va, *ordinariamente*, decrescendo in valore assoluto.
In base a questi risultati sperimentali, si ammette la seguente **legge empirica del caso**, così definita:

In un grande numero di prove, la frequenza relativa di un evento aleatorio **ordinariamente** si scosta di poco dalla probabilità dell'evento, e l'approssimazione è, **in generale**, tanto maggiore quanto più grande è il numero delle prove ripetute.

Questa legge, quando si precisi meglio il significato delle espressioni *ordinariamente* e *in generale*, è denominata anche **legge dei grandi numeri**.

ESEMPIO

Lanciando 18 000 volte due dadi regolari sempre nelle medesime condizioni, quante volte approssimativamente, in base alla legge empirica del caso, si prevede che si presenteranno due facce superiori recanti numeri di somma otto?
In ogni prova il numero dei casi possibili è, come si è già visto, $n = 36$ e i casi favorevoli all'evento «somma dei numeri delle due facce superiori uguale a 8» sono:

$$(2; 6) \quad (3; 5) \quad (4; 4) \quad (5; 3) \quad (6; 2)$$

perciò si ha: $m = 5$.
La probabilità dell'evento è quindi $p = \dfrac{5}{36}$.

Applicando la legge dei grandi numeri, si potrà porre, senza commettere *ordinariamente* errore apprezzabile:

$$\text{frequenza relativa} \approx \text{probabilità}$$

e quindi: $\dfrac{v}{18\,000} \approx \dfrac{5}{36}$, da cui: $v \approx \dfrac{5}{36} \cdot 18\,000 = 2500$.

È perciò assai probabile che l'evento considerato si presenti circa 2500 volte su 18 000 prove.

12. Altre definizioni di probabilità

Frequentista

Vi sono eventi aleatori per i quali non è possibile determinare con precisione il *numero di casi possibili*, oppure il *numero di casi favorevoli*.
Ad esempio, non è possibile sapere quanti degli italiani tuttora viventi lo saranno anche nel 2020; né è possibile sapere, a priori, quante copie saranno vendute di un dato disco.
Per tali eventi non è possibile perciò determinare la probabilità classica.
Spesso, però, di questi eventi aleatori si riesce a determinare la frequenza relativa per un numero qualsiasi n di prove.
In questi casi, in base alla legge dei grandi numeri, come valore della *probabilità dell'evento*, si è soliti prendere il valore della *frequenza relativa*, per n sufficientemente grande. Cioè si pone:

$$p = \frac{v}{n}$$

Questo procedimento è molto applicato dalle compagnie di assicurazione, per esempio. Infatti, se si vuole determinare la probabilità che un individuo di 40 anni ha di vivere fino a 50, si ricorre a opportune tabelle, compilate da uffici specializzati, che forniscono la frequenza relativa di *sopravvivenza*, in relazione ai vari anni, dedotte statisticamente dal passato.

Precisamente, si procede nel modo seguente.

Si considera un gruppo di individui, vissuti in passato, tutti aventi l'età di 40 anni: diciamo n il loro numero.
Si osserva poi quanti individui di questo gruppo erano ancora in vita a 50 anni: diciamo v il loro numero.
Le compagnie di assicurazione assumono allora come *probabilità* che un individuo di 40 anni sia ancora in vita a 50, il numero:

$$\frac{v}{n}$$

Soggettiva

Abbiamo detto del debito che la teoria della probabilità ha nei confronti del gioco d'azzardo. Ricorriamo ancora a esso per definire la probabilità di un evento quando non è possibile ricorrere alla definizione classica di probabilità e neppure a quella frequentista.
Supponiamo di assistere alla finale di un torneo di calcio e di voler scommettere sull'esito dell'incontro.
Il giocatore A scommette sulla vittoria della squadra dei «rossi» e il giocatore B scommette sulla vittoria della squadra dei «verdi».
Perché la scommessa risulti equa, occorre che la posta che A è disposto a giocare, S_A, e quella che è disposto a giocare B, S_B, siano tali che:

$$\frac{S_A}{S_B} = \frac{p_\mathrm{r}}{p_\mathrm{V}}$$

dove p_r è la probabilità che **entrambi** i giocatori attribuiscono alla vittoria dei «rossi» e p_V è la probabilità che **entrambi** i giocatori attribuiscono alla vittoria dei «verdi».
Poiché $p_\mathrm{V} = 1 - p_\mathrm{r}$, si avrà:

$$S_A(1 - p_\mathrm{r}) = S_B p_\mathrm{r} \quad \text{da cui:} \quad p_\mathrm{r} = \frac{S_A}{S_A + S_a}$$

Possiamo quindi concludere dicendo che:
la probabilità che attribuiamo a un evento è uguale al rapporto tra la posta che siamo disposti a versare per partecipare alla scommessa e il premio che ritiriamo in caso di vincita.

S.O.S. Sintesi

CARATTERI, DATI STATISTICI E FREQUENZA

Si chiama **statistica** la scienza che si occupa dello studio di un fenomeno attraverso indagini con cui si raccolgono dati che vengono rappresentati, classificati e, infine, analizzati.

- Il fenomeno verso cui si indirizza la ricerca interessa una collettività, detta **popolazione**, ossia un insieme di **individui** o **unità statistiche** che hanno tutte uno o più aspetti (detti **caratteri**) in comune.
- I caratteri possono essere:
  - **qualitativi:** se le loro modalità, ossia i loro modi di manifestarsi, si esprimono con aggettivi o nomi.
  - **quantitativi:** se le loro modalità, ossia i loro modi di manifestarsi, si esprimono con valori numerici, detti valori del carattere.
- Si chiama **dato statistico** il numero che si determina in corrispondenza a una data modalità del carattere e che può esprimere:
  - il numero di volte che si è si è manifestata una determinata modalità $\rightarrow$ **frequenza** f (assoluta) della modalità;
  - una misura $\rightarrow$ **intensità** I della modalità.

In particolare, si chiama **frequenza relativa** di una data modalità, il rapporto fra la frequenza assoluta di tale modalità e il numero totale degli elementi della popolazione:

$$p = \frac{f}{n}$$

- Le persone viventi attualmente in Italia (popolazione) possono essere classificate, secondo i seguenti caratteri:
  - età;
  - professione.

Il primo è un carattere quantitativo: le modalità del carattere «età» sono numeri (da 0 a 110, ad esempio), che esprimono gli anni di vita della singola unità statistica.

Il secondo è, invece, un carattere qualitativo: le modalità del carattere «professione» sono nomi (studente, impiegato, operaio, imprenditore, {\ldots) che esprimono la posizione professionale della singola unità statistica in seno alla società.

Le seguenti tabelle mostrano i risultati di un'indagine compiuta in una classe di 30 studenti relativamente al mese di nascita (tabella 1) e alla produzione di uva e vino in Italia nell'anno 2007 (tabella 2).

Tabella 1

Mese di nascita	
Gennaio	3
Febbraio	2
Marzo	5
Aprile	1
Maggio	5
Giugno	2
Luglio	1
Agosto	0
Settembre	7

– Le modalità «Agosto e Novembre» hanno frequenza 0, le modalità «Aprile», «luglio» e «Ottobre» hanno frequenza 1, le modalità «Febbraio « e «Giugno» hanno frequenza 2 ecc.

– La modalità «Gennaio» ha frequenza relativa uguale a:

$$\frac{3}{30} = \frac{1}{10} = 0{,}1$$

Tabella 2

Uva da vino raccolta - anno 2007 (quintali)	
ITALIA	60 317 099
NORD	27 912 547
CENTRO	9 155 541
SUD	23 249 011

– La modalità «Nord» ha intensità 27 912 547 e la sua frequenza relativa è:

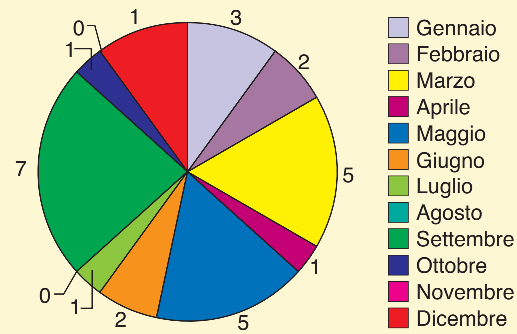
$$\frac{27\,912\,547}{60\,317\,099} \approx 0{,}4627$$

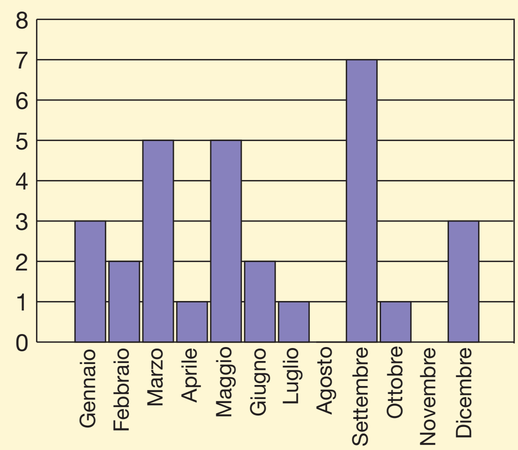
RAPPRESENTAZIONE DI UN FENOMENO STATISTICO

Dopo essere stati raccolti e classificati in tabelle, i dati di un'indagine statistica vengono rappresentati graficamente attraverso diagrammi che mettono in evidenza caratteristiche quali: l'andamento complessivo, la forma, le variazioni della distribuzione statistica (ossia dell'insieme delle coppie ordinate in cui il primo elemento è la modalità e il secondo elemento è la frequenza di tale modalità).

I diagrammi più utilizzati sono:
- gli **ortogrammi** (formati da rettangoli adiacenti di base costante e altezza variabile);
- gli **areogrammi** (formati da un cerchio con settori propozionali alla frequenza delle singole modalità);
- gli **ideogrammi** (costruiti con figure rappresentative del fenomeno su cui si sta indagando: ad esempio un omino stilizzato, se i dati sono demografici);
- i **cartogrammi** (cartine geografiche suddivise in zone colorate con diversa intensità, a seconda delle frequenze rilevate);
- gli **istogrammi** (formati da rettangoli non adiacenti di base costante e altezza variabile).

Riferendoci alla tabella 1, rappresentiamo i dati con un areogramma e con un istogramma.





VALORI MEDI

Chiamiamo **valor medio** di un carattere quantitativo il numero che «riassume» i diversi valori delle intensità di un fenomeno.
I tre tipi fondamentali di valori medi sono:

- la **media aritmetica** (semplice) di n numeri $x_1, x_2, ..., x_n$:

$$M = \frac{x_1 + x_2 + ... + x_n}{n}$$

Osserviamo che se tra gli n valori considerati alcuni coincidono (ad esempio, x_1 si presenta f_1 volte, x_2 si presenta f_2 volte, ..., x_k si presenta f_k volte), è possibile snellire il calcolo della mdia aritmetica calcolando quella che viene definita: **media aritmetica ponderata** di k numeri $x_1, x_2, ..., x_k$:

$$M = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + ... + x_k \cdot f_k}{f_1 + f_2 + ... + f_k}$$

con $f_1 + f_2 + ... + f_k = n$

- la **moda** (o valore normale), il valore del carattere al quale corrisponde la massima frequenza;
- la **mediana** (o valore centrale), il valore che occupa la posizione centrale di una distribuzione di valori di un carattere quantitativo ordinati in ordine crescente (o decrescente).

In particolare:
- se il numero di tali valori è dispari, la mediana è il valore che occupa la posizione centrale;
- se il numero di tali valori è pari, si assume come mediana la semisomma dei due valori centrali.

Consideriamo i dati in tabella relativi al numero di goal segnati da una squadra italiana di calcio di serie A nelle 38 partite giocate nel campionato 2007-2008 (girone a 20 squadre).

Goal segnati	Frequenza
0	8
1	14
2	9
3	6

Abbiamo:
- **media ponderata:**

$$M = \frac{0 \cdot 8 + 1 \cdot 14 + 2 \cdot 9 + 3 \cdot 6 + 4}{38} \approx 1{,}42 \text{ goal}$$

- **moda**: 1 (è il dato con frequenza maggiore)
- **mediana**: scriviamo tutti i valori in ordine crescente:

0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4

Essendo i valori in numero pari, per determinare la mediana dobbiamo calcolare la semisomma dei due valori centrali, ossia:

$$\frac{1 + 1}{2} = 1$$

PROBABILITÀ	**EVENTI**	– Si dice **certo** un evento che, per ogni realizzazione di un insieme di condizioni C, si presenta senza alcuna incertezza. – Si dice **impossibile** un evento che, data la realizzazione di un insieme di condizioni C, non si realizza mai. – Si dice **casuale** o **aleatorio** un evento che, data la realizzazione di un insieme di condizioni C, può accadere oppure no.	– L'estrazione di una pallina rossa da un'urna che contiene 15 palline rosse è un evento certo. – L'estrazione di una pallina rossa da un'urna che contiene 15 palline bianche è un evento impossibile. – L'estrazione di una pallina rossa da un'urna che contiene 5 palline rosse, 8 palline bianche e 6 palline nere è un evento aleatorio.
	SPAZIO DEGLI EVENTI	Si chiama **spazio delle probabilità** associato a un dato esperimento, l'insieme S di tutti i possibili risultati. Si chiama **evento** un sottoinsieme A dello spazio S, cioè un insieme di risultati possibili. Un evento si dice: – elementare, se è un insieme con un solo elemento; – certo, se coincide con S; – impossibile, se è l'insieme vuoto $\emptyset$. Si dice che un risultato è **favorevole** a un evento se verifica l'evento. Si dice che due eventi A e B sono **incompatibili** se: $A \cap B = \emptyset$.	Nell'esperimento «lancio di una dado», consideriamo i seguenti eventi: A: «si presenta il numero 2»; B: «si presenta un numero maggiore di 4»; C: «si presenta un numero minore di 8»; D: «si presenta un numero maggiore di 8». – Lo spazio S delle probabilità è: $S = \{1, 2, 3, 4, 5, 6\}$. – A è un evento elementare. – B ha due modalità (l'uscita del numero 5, o del numero 6, sono favorevoli a B). – C è un evento certo. – D è un evento impossibile – Gli eventi A e B sono incompatibili.
	DEFINIZIONI DI PROBABILITÀ	**Definizione classica** $p(A) = \frac{m}{n}$ con m numero dei risultati favorevoli ed n numero dei risultati possibili Si ha: **1.** $0 \leq p(A) \leq 1$ In particolare: – se $m = 0 \rightarrow p(A) = 0$, allora A è l'evento impossibile; – se $m = n \rightarrow p(A) = 1$ allora A è l'evento certo. **2.** $p(\overline{A}) = 1 - p(A)$, con $\overline{A}$ evento contrario di A. **3.** $p(A \cup B) = p(A) + p(B) - p(A \cap B)$. In particolare, se A e B sono incompatibili, allora: $p(A \cup B) = p(A) + p(B)$	Si estrae una pallina da un'urna che contiene 15 palline numerate da 1 a 15. Calcola la probabilità che, estraendo a caso una sola pallina dall'urna, si verifichino i seguenti eventi: a) A: «la pallina è contrassegnata da un numero primo». b) B: «la pallina è contrassegnata da un numero pari». c) C: «la pallina è contrassegnata da un numero dispari» d) D: «la pallina è contrassegnata da un numero primo o da un numero dispari». I risultati favorevoli ad A sono: 2, 3, 5, 7, 11, 13. I risultati favorevoli a B sono: 2, 4, 6, 8, 10, 12, 14. I risultati possibili sono: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15. a) $p(A) = \frac{6}{15} = \frac{2}{5}$ b) $p(B) = \frac{7}{15}$ c) Osservando che $C = \overline{A}$, abbiamo: $p(C) = 1 - p(A) = 1 - \frac{2}{5} = \frac{3}{5}$ d) Osservando che: – $D = A \cup C$ – i risultati favorevoli ad $A \cap C$ sono: 3, 5, 7, 11, 13, abbiamo: $p(D) = p(A) + p(C) - p(A \cap C) = \frac{2}{5} + \frac{3}{5} - \frac{5}{15} = \frac{2}{3}$
		Definizione frequentista $p(A) = f = \frac{v}{n}$ con v numero dei successi ed n numero delle prove fatte	Se, lanciando un dado 800 volte, otteniamo 389 volte l'uscita del numero 6, la probabilità di tale evento A è $p(A) = \frac{389}{800} = 0{,}48$
		Definizione soggettiva La probabilità di un evento è uguale al rapporto tra la posta S che un individuo è disposto a versare scommettendo sull'esito dell'evento e il premio S' che ritirerà in caso di vincita: $p(A) = \frac{S}{S'}$	Se, scommettendo su un evento A, siamo disposti a versare 10 euro per ricevere, in caso di vincita, 16 euro, significa che riteniamo che la probabilità di tale evento sia uguale a: $p(A) = \frac{10}{16} = 0{,}625$

START

1 Indicare se le seguenti affermazioni sono vere o false. V F

a) La statistica è l'insieme delle tecniche che permettono di raccogliere, analizzare, interpretare informazioni su un determinato fenomeno.

b) Si chiama serie ogni tabella che si riferisce a modalità qualitative o quantitative

c) Si definisce frequenza relativa di una modalità il rapporto tra la frequenza assoluta e il numero delle unità statistiche che formano il campione considerato

d) Si chiama areogramma una rappresentazione di un fenomeno statistico che utilizza le superfici di figure

e) Indicata con $M(x)$ la media aritmetica dei numeri $x_1, x_2, ..., x_n$ e con $M(ax)$ quella dei numeri $ax_1, ax_2, ..., ax_n$ con $a \in R$ si ha $M(ax) \neq M(x)$

f) $M(b + x) = b + M(x)$ con $b \in R$

g) La mediana è quella modalità quantitativa o qualitativa che si presenta con la frequenza maggiore

h) Il campo di variazione è il numero dei valori di un insieme di dati

i) Per la media aritmetica vale la proprietà per cui la somma degli scarti positivi è uguale, in valore assoluto, alla somma degli scarti negativi.

[a) V; b) F; c) V; d) V; e) F; f) V; g) F; h) F; i) V]

2 La tabella rileva la frequenza relativa del colore dei capelli di un gruppo di 40 persone, determinare le frequenze assolute.

Colore	Neri	Castani	Rossi	Grigi
Frequenza relativa	0,4	0,35	0,15	0,1

3 Calcolare il campo di variazione, la media aritmetica, la mediana e la moda dei voti riportati da uno studente nel primo quadrimestre: 6, 5, 6, 6, 7, 6, 8, 6, 4, 9.

4 Calcolare la media ponderata dei valori espressi nella tabella con i relativi pesi

x	21	23	25	29	30
Pesi	2	5	12	8	9

[26,64]

5 Calcolare la media aritmetica ponderata di un fenomeno la cui modalità è stata rilevata per classi

x	40-50	50-60	60-70	70-80
n	16	47	23	14

[58,5]

6 Indicare se le seguenti affermazioni sono vere o false. V F

a) Nel calcolo delle probabilità l'insieme di tutti i possibili risultati di un esperimento prende il nome di spazio degli eventi. ☐ ☐

b) Da un punto di vista insiemistico lo spazio degli eventi è l'insieme universo. ☐ ☐

c) Due eventi A e B per i quali $A \cap B = \varnothing$ si dicono compatibili. ☐ ☐

d) Dati due eventi A e B in generale si ha $p(A \cup B) = p(A) + p(B) - p(A \cap B)$. ☐ ☐

e) Indicati con A e $\overline{A}$ un evento e il suo contrario si ha $p(A) = p(\overline{A})$. ☐ ☐

f) Se A e B sono eventi dipendenti si ha $p(A \cap B) = p(A) \cdot p(B)$. ☐ ☐

g) Due eventi contrari sono anche incompatibili. ☐ ☐

h) Due eventi incompatibili sono anche contrari. ☐ ☐

[a) V; b) V; c) F; d) V; e) F; f) F; g) V; h) F]

7 Qual è la probabilità di colpire un bersaglio di forma circolare più vicino al centro che alla circonferenza?

$\left[\frac{1}{4}\right]$

8 Nell'esperimento lancio di due dadi si considerino gli eventi:
A: «il punteggio complessivo è dispari»
B: «il punteggio complessivo è un multiplo di tre».
Calcolare

$p(A)$	$p(B)$	$p(\overline{A})$	$p(\overline{B})$	$p(A \cup B)$	$p(A \cap B)$

$\left[\frac{1}{2}; \frac{1}{3}; \frac{1}{2}; \frac{2}{3}; \frac{2}{3}; \frac{1}{6}\right]$

9 Sapendo che $p(A) = \frac{1}{2}$, $p(B) = \frac{3}{5}$, $p(A \cup B) = \frac{4}{5}$, che cosa si può dire degli eventi A e B?

[Sono compatibili]

10 Due giocatori di tennis disputano tre incontri. Sapendo che la probabilità del primo di vincere una partita è pari a $\frac{2}{3}$, calcolare quale probabilità ha di aggiudicarsi due incontri.

$\left[\frac{4}{9}\right]$

GO!

La statistica

1 Trovare articoli di giornali o annunci pubblicitari, che usino dati numerici e statistiche, per descrivere il mondo d'oggi:

a) nello sport;
b) nelle condizioni meteorologiche;
c) nelle condizioni economiche;
d) nella politica;
e) nella scienza;
f) nella scuola.

2 «Vi sono tre tipi di menzogne: le menzogne, le menzogne spudorate e la statistica» (Disraeli).
Si dice spesso che servendosi della statistica si può dimostrare qualsiasi cosa. Ciò significa che la statistica può essere usata per trarre in inganno. Che cosa c'è di sbagliato nella conclusione delle frasi seguenti?

a) Morirono più persone in incidenti d'aereo nel 2008 che nel 1935. Dunque è più pericoloso viaggiare in aereo nel 2008.
b) In una fattoria, le mucche pezzate hanno prodotto il 30% in più di latte rispetto alle altre. Dunque le mucche pezzate sono le migliori produttrici di latte.
c) Tutti coloro che hanno usato le pasticche Furlot guarirono dal raffreddore in sette giorni. Dunque le pasticche Furlot curano il raffreddore.
d) Da un'inchiesta risulta che sette dentisti su otto usano il dentifricio Minigel. Dunque il dentifricio Minigel è un ottimo dentifricio.

3 Compilare la tabella relativa alla distribuzione per età degli alunni della propria classe (in classe di ampiezza 1 anno).

4 Compilare la tabella relativa alla distribuzione secondo la statura degli alunni della propria classe (in classi di ampiezza 5 cm).

5 Compilare la tabella relativa alla distribuzione secondo il peso degli alunni della propria classe (in classi di ampiezza 4 kg).

6 Compilare la tabella relativa alla distribuzione dei componenti la classe, secondo i voti riportati nell'ultimo compito di matematica (in classi di ampiezza 1 voto).

7 Compilare la tabella relativa alla distribuzione degli alunni della propria classe:

a) secondo il colore degli occhi;
b) secondo il colore dei capelli;
c) secondo il numero dei fratelli;
d) secondo la materia scolastica preferita;
e) secondo le distanze dalla sede della scuola (in classi di ampiezza conveniente);
f) secondo la materia scolastica più antipatica;
g) secondo lo sport preferito;
h) secondo il tipo di musica preferito;
i) secondo le assenze fatte nel primo quadrimestre.

8 Compilare una tabella di distribuzione delle assenze nella propria classe secondo i giorni della settimana in un dato periodo.

9 Si lanci un dato da gioco 25 volte e si compili la tabella dei risultati.

10 Si lanci una moneta 20 volte e si compili la tabella dei risultati.

11 Estrarre 20 volte una carta da un mazzo, rimettendola ogni volta nel mazzo stesso, e compilare la tabella dei risultati secondo il *colore* e una tabella secondo i *semi*.

12 Estrarre per 20 volte una carta da un mazzo, senza rimettere la carta estratta nel mazzo, e compilare la tabella dei risultati secondo il numero della carta.

13 Compilare la tabella di distribuzione dei cognomi degli alunni della propria classe secondo la lettera iniziale.

14 Compilare la tabella di distribuzione dei cognomi dei propri professori (o dei compagni) secondo il numero delle sillabe nei cognomi.

15 Compilare la tabella di distribuzione delle parti del discorso (nomi, aggettivi, verbi ecc.) contenute nei primi 5 versi di una data poesia.

16 Compilare la tabella di distribuzione dei rappresentanti dei diversi partiti politici nel consiglio comunale della propria città.

17 Compilare una tabella di spoglio relativa all'opinione dei propri compagni di classe su un film (espressa in modo sintetico, mediante risposte a un questionario del tipo: *non mi è piaciuto*; *mi è piaciuto*; *ho delle riserve*; *non lo conosco*).

18 Come nell'esercizio precedente, in relazione a un libro in voga e a uno spettacolo teatrale o musicale.

19 Ecco alcune cifre relative ai morti sulle strade di una nazione.

Mese	2004	2005	2006
Gennaio	381	507	552
Febbraio	372	458	415
Marzo	415	493	581
Aprile	456	500	534
Maggio	492	504	580
Giugno	541	537	525
Luglio	588	634	607
Agosto	546	589	543
Settembre	568	619	612
Ottobre	655	670	672
Novembre	625	695	613
Dicembre	881	764	674

a) Vi sono differenze notevoli tra i vari mesi dell'anno?
b) Queste differenze sono le stesse da un anno all'altro?
c) C'è un mese particolarmente «brutto»?
d) Vi sono differenze notevoli da un anno all'altro?
e) La situazione va peggiorando o migliorando?
f) È facile rispondere a simili domande con una tabella come questa?

20 Si prenda un prospetto dei risultati di una giornata del campionato di calcio di serie A e si compili una tabella di distribuzione delle squadre, secondo il numero delle reti segnate.

21 Si prenda il bollettino del lotto di un determinato sabato e si compili la tabella di distribuzione di frequenza dei numeri usciti ripartiti in classi di decine.

22 Si consideri la tavola delle radici quadrate a 5 decimali per i numeri da 100 a 199. Si compili la tabella atta a determinare quante radici terminano con ciascuna delle 10 cifre.

23 Servendosi di dati raccolti nel «Radiocorriere» (o simili) rispondere ad alcune domande. Quante ore della settimana sono dedicate:
a) ai notiziari;
b) allo sport;
c) alla musica popolare;
d) quale frazione del giorno è dedicata alle trasmissioni in genere;
e) quanto tempo riserva la televisione, ogni settimana, ai programmi «seri».

24 Nella seguente tabella sono riportati i dati relativi alla quantità di pioggia (in centimetri) caduta mese per mese, durante gli anni 2000-2004, in una data regione.

Quantità totale di pioggia in centimetri

Mesi	2000	2001	2002	2003	2004
Gennaio	12,7	6,7	9,1	13,7	11,2
Febbraio	1,8	9,0	6,7	0,8	5,6
Marzo	1,7	6,0	3,9	12,3	4,0
Aprile	3,3	0,7	2,5	7,9	4,4
Maggio	0,8	3,9	8,6	3,6	3,6
Giugno	4,7	4,5	8,3	5,9	8,9
Luglio	7,2	8,4	7,1	4,9	9,1
Agosto	12,1	8,1	6,6	4,7	9,2
Settembre	10,7	9,8	12,5	0,3	9,2
Ottobre	4,8	6,6	7,4	5,8	21,2
Novembre	2,6	6,2	7,2	8,3	12,5
Dicembre	11,7	9,5	9,2	20,3	12,2

a) Qual è, di solito, il periodo dell'anno meno piovoso?
b) C'è stato un anno eccezionale?
c) Qual è stata la migliore stagione per il calcio?
d) In quale mese possono esservi state inondazioni?

Rappresentazione grafica dei fenomeni statistici

Rappresentare graficamente, nel modo indicato, le distribuzioni statistiche raccolte nelle tabelle di spoglio degli esercizi precedenti.

25 Istogramma della distribuzione per classi di età degli alunni della classe.

26 Istogramma della distribuzione per classi di statura degli alunni della classe.

27 Diagramma relativo alla distribuzione dei voti di italiano dell'ultimo compito.

28 Diagramma di distribuzione degli alunni secondo il numero dei fratelli.

29 Diagramma di ripartizione dei componenti la classe secondo il colore degli occhi o dei capelli.

30 Diagramma di distribuzione delle assenze secondo i giorni della settimana per due o più settimane consecutive.

31 Istogramma o ideogramma di distribuzione dei diversi partiti politici nel consiglio comunale, o nel parlamento della Repubblica.

32 Istogramma di distribuzione di frequenza dei numeri del bollettino del lotto in un sabato, ripartiti in classi di decine.

Date le seguenti tabelle, farne un'opportuna rappresentazione grafica.

33 Temperature misurate in una certa località durante una giornata.

Ore	0	2	4	6	8	10	12	14	16	18	20	22
Temperature (°C)	1	0	0,5	1,5	3	5	7	7	6	5	3,5	20

34 Andamento delle vendite di un'azienda.

Anni	Quantità vendute	Anni	Quantità vendute
1995	8650	2000	12 910
1996	9310	2001	13 140
1997	10 250	2002	13 720
1998	10 380	2003	14 030
1999	11 500	2004	14 450

35 Velocità del vento misurata a Trieste durante una giornata.

Ore	0	3	6	9	12	15	18	21	24
Velocità (km/h)	50	70	60	80	70	90	85	95	100

36 Temperature misurate in una certa località durante una giornata invernale.

Ore	5	6	7	8	9	10	11	12	13	14
Temperature (°C)	1	0,5	1,5	3	3,8	4,5	6	6,8	9	9,5

37 Produzione di patate, in milioni di tonnellate, in una data nazione.

Anni	2000	2001	2002	2003	2004	2005	2006	2007	2008
Patate	10	13,3	10,6	15,2	10,9	11,8	8,9	9,6	10,5

38 Temperature massime e minime verificatesi a Napoli nei 12 mesi di un dato anno.

Mesi	Temperatura massima	Temperatura minima
Gennaio	20°	2°
Febbraio	22°	1°
Marzo	21°	0°
Aprile	25°	5°
Maggio	30°	7°
Giugno	32°	13°
Luglio	35°	15°
Agosto	37°	14°
Settembre	31°	10°
Ottobre	28°	5°
Novembre	23°	4°
Dicembre	18°	0°

39 Numero dei giorni piovosi in Italia, dal 1986 al 1994.

Anno	Numero dei giorni piovosi	Anno	Numero dei giorni piovosi
1986	85	1991	86
1987	90	1992	90
1988	88	1993	106
1989	86	1994	85
1990	81		

40 Numero dei clienti secondo la quantità acquistata di una data merce.

Quantità (q)	Numero dei clienti	Quantità totale (Q)
0-200	123	19 850
200-400	218	65 320
400-600	116	49 350
600-800	65	41 180
800-1000	11	10 250
oltre 1000	6	7 240
	...	...

41 Prezzi di vendita rilevati per un prodotto nei vari mesi dell'anno.

Mesi	Prezzi	Mesi	Prezzi
Gennaio	185	Luglio	204
Febbraio	192	Agosto	211
Marzo	194	Settembre	196
Aprile	186	Ottobre	199
Maggio	189	Novembre	194
Giugno	198	Dicembre	197

42 Servendosi di pubblicazioni dell'Istituto Centrale di Statistica o di altri uffici statistici, rappresentare graficamente fenomeni statistici, come i seguenti.

a) Numero dei matrimoni in un determinato periodo di tempo.
b) Ripartizioni delle nascite e dei decessi, per sesso.
c) Ripartizione degli studenti universitari secondo i tipi di corso di laurea.
d) Sviluppo edilizio.
e) Movimento negli esercizi alberghieri.
f) Sviluppo della rete stradale dello stato.
g) Sviluppo dei trasporti marittimi e aerei.
h) Popolazione presente in Italia.
i) Produzione annua del frumento e del granturco.

Valori medi

MEDIA ARITMETICA

43 Trovare la media aritmetica dei seguenti numeri.

a) 4 5 7 31 32 b) 3 5 8 12 15 21 c) -9 -4 1 0 2,7 5,5

44 Sono state fatte dieci misure, in centimetri, del diametro di un cerchio:

3,88 4,09 3,92 4,02 3,95 3,97 4,03 3,92 4,06 3,98

Trovare la media aritmetica delle misure. [3,98]

45 Trovare la media aritmetica dei numeri:

5 6 5 3 4 2 5 8 6 4 5 8 3 5 4 4 2 8 [4,8]

46 Calcolare la media aritmetica ponderata dei seguenti:

numeri	*con i rispettivi pesi*	
a) 1; 5; 8; 11	3; 6; 9; 2	[6,35]
b) 1; 2; 3; 4; 5; 6; 7; 8	3; 6; 8; 13; 5; 3; 1; 1	[3,725]
c) 3; 5; 8; 10; 12; 6	1; 6; 10; 10; 7; 2	[8,583]
d) -3; -1; 0; 5; 10	3; 4; 7; 6; 2	$\left[\dfrac{37}{22}\right]$
e) 10; 15; 20; 25; 30	2; 3; 8; 4; 3	[20,75]

47 Quattro gruppi di studenti, composti da 15, 20, 10 e 18 persone, hanno un'altezza media, rispettivamente di 162, 148, 153 e 140 cm. Trovare l'altezza media di tutti gli studenti. [150 cm]

48 Data la seguente tabella:

Classi di intensità	Frequenze	Classi di intensità	Frequenze
40 – 45	8	55 – 60	40
45 – 50	28	60 – 65	12
50 – 55	60	65 – 70	2

determinare la media aritmetica, prendendo per ogni classe il valore centrale. [53,37]

49 In un insieme di numeri compaiono sei volte il 6, sette volte il 7, otto volte l'8, nove volte il 9 e dieci volte il 10. Trovare la media aritmetica di questi numeri. [8,25]

50 Lo stipendio medio annuo pagato a tutti gli impiegati di una ditta è di € 20 000. Lo stipendio medio annuo pagato agli impiegati è di € 20 800 mentre quello pagato alle impiegate è di € 16 800. Determinare la percentuale di impiegati e di impiegate nella ditta. [80%; 20%]

MODA O VALORE NORMALE

51 Calcolare il valore normale, o moda, dell'intensità del carattere delle seguenti distribuzioni statistiche.

Intensità	1	3	5	7	12	16
Frequenza	5	20	14	9	4	2

[3]

Intensità	2	5	6	8	12	15
Frequenza	7	2	21	15	3	9

[6]

52 Determinare la moda dei dati riportati in alcune tabelle degli esercizi precedenti.

53 Un fabbricante di copertoni ne prova 24 per determinare il numero di chilometri che si possono percorrere prima che si consumino e ottiene i risultati seguenti:

Chilometri percorribili	Numero dei copertoni
100 000	1
90 000	1
55 000	6
50 000	8
40 000	8

Quale valore medio è il più appropriato a descrivere la vita media dei copertoni prodotti da quel fabbricante? Perché?

54 Un concessionario automobilistico vende delle vetture di diversa cilindrata, come è rappresentato nella seguente tabella.

Cilindrata (intensità del carattere)	Numero delle automobili vendute (frequenza)
1000	250
1100	290
1600	70
2000	50

Determinare la moda della distribuzione.

55 Un negozio d'abbigliamento ha venduto vestiti delle varie taglie, come è rappresentato nella seguente tabella.

Taglia (intensità del carattere)	Vestiti venduti (frequenza)
40	15
42	34
44	85
46	70
48	53
50	30

Per il rifornimento riguardante la stagione successiva, a quale taglia di vestito darà la preferenza? Perché?

56 Un negoziante di strumenti musicali vende delle chitarre a prezzi diversi, come è rappresentato nella seguente tabella.

Prezzo della chitarra, in € (intensità del carattere)	Numero delle chitarre vendute (frequenza)
55	10
62	25
67	15
75	25
90	9
115	2

Determinare la moda della distribuzione.

MEDIANA

57 Calcolare la mediana delle seguenti successioni di dati.

a) 4 6 7 15 24 [7]

b) -4 -1 0 $+3$ [$-0,5$]

c) 3 5 7 18 20 [7]

d) -10 -4 -1 5 $\left[-\dfrac{3}{2}\right]$

58 Calcolare la mediana della successione:

2 6 8 12 17 26 con le frequenze rispettive: 1 3 2 7 4 6 [12]

59 Calcolare la mediana della successione:

-4 2 7 12 18 con le frequenze rispettive: 2 1 4 3 5 [12]

60 Determinare la mediana nella distribuzione dei voti di un compito in classe.

61 Determinare la mediana nella distribuzione delle età degli alunni della propria classe.

62 Determinare la mediana nella distribuzione delle stature degli alunni della propria classe.

La probabilità

63 Siano A e B eventi. Trovare un'espressione e tracciare il grafico di Venn per i seguenti eventi:
a) si verifichi A ma non B;
b) si verifichi A oppure B, ma non entrambi.
$\left[\text{a) } A \cap \overline{B};\ \text{b) } \left(A \cap \overline{B}\right) \cup \left(B \cap \overline{A}\right)\right]$

64 Si estrae una carta a caso da un mazzo di 52 carte. Sia A l'evento «è uscito un *re*» e B l'evento «è uscito un *picche*». Si descrivano gli eventi:

$$A \cup B \quad A \cap B \quad A \cup \overline{B} \quad \overline{A} \cup \overline{B} \quad A - B \quad \overline{A} - \overline{B} \quad (A \cap B) \cup \left(A \cap \overline{B}\right)$$

65 Siano A, B, C eventi. Trovare un'espressione e tracciare il grafico per i seguenti eventi:
a) si verifichino A e B ma non C;
b) si verifichi soltanto A.
$\left[\text{a) } A \cap B \cap \overline{C};\ \text{b) } A \cap \overline{B} \cap \overline{C}\right]$

66 Si lancino insieme una moneta e un dado. Determinare:
a) lo spazio S dei risultati;
b) gli eventi:
A: «si presentano testa e un numero pari»;
B: «si presenta un numero primo»;
C: «si presentano croce e un numero dispari».
Inoltre, esprimere i seguenti eventi:
E: «si verifichi A oppure B»;
F: «si verifichino B e C»;
G: «si verifichi soltanto B».
Infine, dire quali fra gli eventi A, B, C sono incompatibili.

PROBABILITÀ CLASSICA

67 Si lanciano contemporaneamente una moneta e un dato. Quali sono i casi possibili che si possono ottenere in tale lancio?

68 Dire qual è la probabilità delle seguenti combinazioni in una famiglia che abbia due figli.
a) Due maschi
b) Un maschio e una femmina
c) Due femmine
Qual è la somma di queste probabilità?
$\left[\text{a) } \frac{1}{4};\ \text{b) } \frac{1}{2};\ \text{c) } \frac{1}{4};\ \text{d) } 1\right]$

69 Dire qual è la probabilità di ciascuno dei seguenti eventi.
a) In una famiglia con tre figli questi siano tutti maschi.
b) Nel lancio di due monete escano due teste.
c) Nel lancio di due dadi si abbia (1, 1).
d) Nel lancio di due dadi si abbia un totale inferiore a 5.
$\left[\text{a) } \frac{1}{8};\ \text{b) } \frac{1}{4};\ \text{c) } \frac{1}{36};\ \text{d) } \frac{1}{6}\right]$

70 Un'urna contiene 10 palline numerate da 1 a 10. Estraendo a caso una pallina dall'urna, qual è la probabilità che il numero estratto non superi 10? [1]

71 Un'urna contiene 15 palline, di cui 5 bianche e 10 nere. Qual è la probabilità di estrarre dall'urna una pallina rossa? [0]

72 Qual è la probabilità che lanciando due volte di seguito una moneta regolare, si presenti due volte «faccia»?
$\left[\frac{1}{4}\right]$

73 Si gettano due dadi. Qual è la probabilità che la somma dei risultati sia 7? $\left[\frac{1}{6}\right]$

74 In una classe di 32 alunni l'insegnante estrae a sorte il nome di un alunno per interrogarlo. Calcolare la probabilità che ha ciascun alunno:
a) di essere interrogato;
b) di non essere interrogato. $\left[\text{a) } \frac{1}{32};\ \text{b) } \frac{1}{32}\right]$

75 Qual è la probabilità che lanciando due monete si ottenga «testa e croce». $\left[\frac{1}{2}\right]$

76 Calcolare la probabilità che lanciando due monete si presenti almeno una «testa». $\left[\frac{3}{4}\right]$

77 Si gettano contemporaneamente due dadi regolari. Calcolare la probabilità di ottenere:
a) la somma dei punti uguale a 6;
b) la somma dei punti uguale a 2;
c) la somma dei punti uguale a 12;
d) i due dadi con lo stesso punteggio. $\left[\text{a) } \frac{5}{36};\ \text{b) } \frac{1}{36};\ \text{c) } \frac{1}{36};\ \text{d) } \frac{1}{6}\right]$

78 Si lancia un dado regolare. Calcolare la probabilità di ottenere:
a) un 2 o un 5 o un 6;
b) una faccia pari;
c) una faccia dispari;
d) una faccia qualunque diversa da 6;
e) un numero non divisibile per 3. $\left[\text{a) } \frac{1}{2};\ \text{b) } \frac{1}{2};\ \text{c) } \frac{1}{2};\ \text{d) } \frac{5}{6};\ \text{e) } \frac{2}{3}\right]$

79 Calcolare la probabilità che lanciando due dadi si presenti una somma di punti minori di 7. $\left[\frac{5}{12}\right]$

80 Un'urna contiene 30 palline di cui 15 bianche, 8 rosse e le rimanenti di altri colori. Determinare la probabilità che, in una estrazione a caso, la pallina estratta sia:
a) bianca;
b) rossa;
c) né bianca né rossa. $\left[\text{a) } \frac{1}{2};\ \text{b) } \frac{4}{15};\ \text{c) } \frac{7}{30}\right]$

81 Da un mazzo di 40 carte regolare e ben mescolate si estrae a caso una carta. Calcolare la probabilità che la carta estratta sia:
a) l'asso di denari;
b) un asso qualunque;
c) una carta di denari;
d) una figura. $\left[\text{a) } \frac{1}{40};\ \text{b) } \frac{1}{10};\ \text{c) } \frac{1}{4};\ \text{d) } \frac{3}{10}\right]$

82 In una classe ci sono 22 alunni. Per interrogarli, il professore di lettere estrae a sorte da un sacchetto contenente 30 palline, numerate da 1 a 30. Se il numero estratto non supera 22, allora viene interrogato il ragazzo che ha sul registro quel numero; se il numero supera 22, allora si fa la somma delle cifre e viene interrogato il ragazzo che corrisponde al numero ottenuto.
È giusto il procedimento adottato dal professore? Qual è la probabilità che ciascun ragazzo ha di essere interrogato?

83 In un'urna ci sono 10 palline rosse. Si possono aggiungere palline azzurre in modo che la probabilità di estrarre una pallina azzurra sia $\frac{1}{3}$? Se sì, determinare il numero e calcolare la probabilità di estrarre una pallina rossa.

$\left(\text{Se } x \text{ e il numero delle palline azzurre da aggiungere, deve risultare } \frac{x}{10+x} = \frac{1}{3}, \text{ da cui } \ldots\right)$

$\left[5;\ \frac{2}{3}\right]$

84 Se un'urna contiene 5 palline verdi, è possibile completarla con palline bianche in modo che la possibilità di estrarre una verde sia $\frac{1}{3}$? Se si aggiungono 3 palline verdi, si possono aggiungere altre palline bianche in modo che la probabilità resti la stessa?

85 Un'urna contiene palline bianche e palline rosse. La probabilità di pescare una pallina bianca è $\frac{3}{10}$. Si può avere un'urna con 70 palline? E con 75? Quali sono le confezioni possibili?

[Sì; no; confezioni contenenti un numero di palline multiplo di 10]

86 Sei persone intervengono a una cerimonia ufficiale; ciascuna stringe la mano alle altre. Una sola stretta di mano, a caso, verrà fotografata. Che probabilità ha una qualsiasi persona di essere fotografata?

$\left[\frac{1}{3}\right]$

PROBABILITÀ FREQUENTISTA

87 Lanciando 2400 volte una moneta regolare, quante volte, in media, si può prevedere che si presenti «croce»? [1200]

88 Lanciando 1800 volte un dado regolare, quante volte, in media, si può prevedere che si presenterà la faccia «uno»? [300]

89 Lanciando a caso due dadi 18 000 volte, quante volte, in media, si può prevedere che si presenterà una somma di numeri uguale a 8? [2500]

90 Se si estrae per 1600 volte una carta da un determinato mazzo di 40, quante volte in media si può prevedere che si presenti un determinato «re»? [40]

91 Con lo stesso numero di prove dell'esercizio precedente, qual è la frequenza media assoluta prevedibili dell'evento «figura qualsiasi»? [480]

92 Lanciando due dadi regolari a caso 9000 volte, quale sarà la frequenza media assoluta dell'evento «somma minore di 8»? [5250]

93 Un'urna contiene 50 palline, di cui 30 rosse e 20 bianche. Si ripeta l'estrazione a caso di una pallina 800 volte di seguito (rimettendo ogni volta la palla estratta nell'urna). Calcolare il numero di volte che prevedibilmente si presenterà, in media, una pallina rossa. [480]

TIME OUT RECUPERO

ESERCIZI SVOLTI

1 Da un'indagine sul prestito dei libri della biblioteca di una scuola è risultato quanto segue.

prestiti per classe

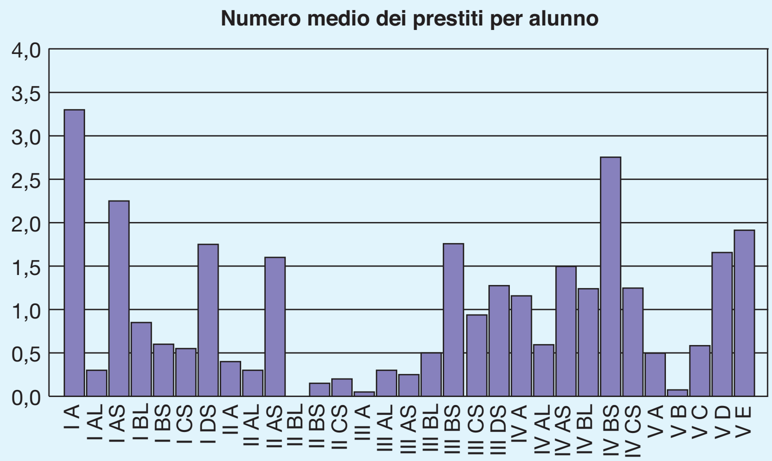
Classe	Totale classe	Numero prestiti	Maschi	Femmine
I A	25	84	37	47
I AL	26	7	4	3
1 AS	26	59	23	36
I BL	25	22	10	12
I BS	26	16	16	0
I CS	24	13	8	5
I DS	26	46	14	32
II A	25	8	7	1
II AL	22	6	3	3
II AS	14	23	12	11
II BL	25	0	0	0
II BS	17	3	3	0
II CS	21	5	2	3
III A	22	1	0	1
III AL	22	6	2	4
III AS	20	5	1	4
III BL	14	7	0	7
III BS	21	37	17	20
III CS	22	21	4	17
III DS	18	22	5	17
IV A	22	26	9	17
IV AL	22	13	5	8
IV AS	20	30	27	3
IV BL	22	27	2	25
IV BS	14	40	25	15
IV CS	16	21	15	6
V A	20	9	2	7
V BD	19	1	0	1
V C	16	10	0	10
V E	16	27	13	14
V F	14	26	7	19

Calcolare la media dei prestiti, la media dei prestiti per alunno e classe, la media dei prestiti per alunno e anno di corso, la percentuale dei prestiti per sesso. Rappresentare i dati trovati con opportuna rappresentazioni grafiche (istogrammi, ortogrammi o aerogrammi).
Per calcolare la media dei prestiti dobbiamo sommare tutti i prestiti e dividere la somma ottenuta per il numero totale degli allievi.
Il numero totale dei prestiti è 621 (somma il numero dei prestiti di ogni classe), mentre il numero totale degli allievi è 642. La media dei prestiti sarà quindi $621 : 642 = 0{,}97$.

Per calcolare la media dei prestiti per classe dobbiamo fare il rapporto fra il numero dei prestiti di ogni classe con il numero totale degli allievi di quella classe. Troviamo quindi la seguente tabella.

	Totale classe	Numero prestiti	Media prestiti per alunno e classe
I A	25	84	3,36
I AL	26	7	0,27
I AS	26	59	2,27
I BL	25	22	0,88
I BS	26	16	0,62
I CS	24	13	0,54
I DS	26	46	1,77
II A	25	8	0,32
II AL	22	6	0,27
II AS	14	23	1,64
II BL	25	0	0,00
II BS	17	3	0,18
II CS	21	5	0,24
III A	22	1	0,05
III AL	22	6	0,27
III AS	20	5	0,25
III BL	14	7	0,50
III BS	21	37	1,76
III CS	22	21	0,95
III DS	18	22	1,22
IV A	22	26	1,18
IV AL	22	13	0,59
IV AS	20	30	1,50
IV BL	22	27	1,23
IV BS	14	40	2,86
IV CS	16	21	1,31
V A	20	9	0,45
V BD	19	1	0,05
V C	16	10	0,63
V E	16	27	1,69
V F	14	26	1,86

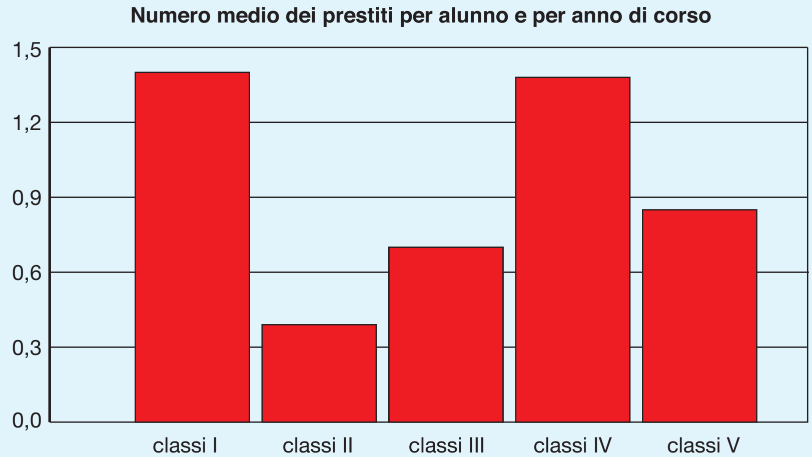
Rappresentando con un ortogramma i dati trovati aabbiamo:



Per calcolare la media dei prestiti per alunno e anno di corso dobbiamo considerare la somma dei prestiti per anno di corso e il totale degli alunni, sempre per anno di corso.

	classi I	classi II	classi III	classi IV	classi V
Numero prestiti	247	45	99	157	73
Numero alunni	178	124	139	116	85
Media	1,38764	0,362903	0,71223	1,353448	0,858824

Rappresentiamo i dati trovati con un ortogramma



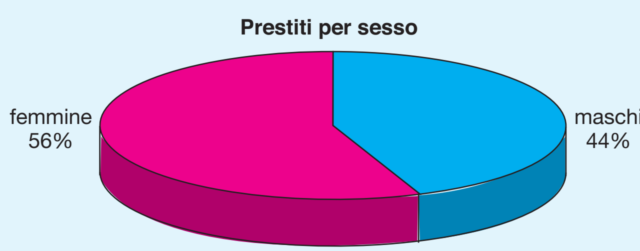
La percentuale dei prestiti per sesso si trova dividendo il numero dei prestiti per sesso per il numero totale dei prestiti e moltiplicando il risultato ottenuto per 100.
Il numero totale dei prestiti per i maschi è 273, quindi:

273 : 621 = 0,4396

da cui ricaviamo la percentuale del 44%

Il numero totale dei prestiti per le femmine è 348, quindi 348 : 621 = 0,5603, da cui ricaviamo la percentuale del 56%.

Rappresentiamo i dati trovati con un aeogramma



2 Calcolare la probabilità che nel lancio di un dado
A: «esca un numero dispari.»
B: «esca un numero minore di 5.»
A e *B*: «esca un numero dispari minore di 5.»
A o *B*: «esca un numero dispari o *n* numero minore di 5.»

La probabilità di un evento *A* è il rapporto fra il numero dei casi favorevoli e il numero dei casi possibili

I casi possibili sono 6 (le facce del dado).

Poiché i numeri dispari sono $\{1, 3, 5\}$ la probabilità dell'evento A è $p(A) = \frac{3}{6} = \frac{1}{2}$.

Poiché i numeri minori di 5 sono $\{1, 2, 3, 4\}$, la probabilità dell'evento B è $p(B) = \frac{4}{6} = \frac{2}{3}$.

I numeri dispari minori di 5 sono $\{1, 3\}$, la probabilità dell'evento A e B è $p(A \cap B) = \frac{2}{6} = \frac{1}{3}$.

$$p(A \text{ o } B) = p(A) + p(B) - p(A \cap B) = \frac{1}{2} + \frac{2}{3} - \frac{1}{3} = \frac{5}{6}$$

Lo stesso risultato si può trovare determinando $A \cup B$:

$A \cup B = \{1, 2, 3, 4, 5\}$, quindi: $p(A \cup B) = \frac{5}{6}$

Poiché A e B sono due eventi qualsiasi, compatibili fra loro $p(A \cup B) = p(A) + p(B) - p(A \cup B)$

ESERCIZI GUIDATI

3 Calcolare la media, la mediana e la moda della seguente serie di numeri:

11,5 12 12,5 11 12,5 11,5 11,5 12 12,5 11,5 10 12,5 11,5 10 12,5

Determina la frequenza di ogni numero.
La frequenza di 10 è ....
La frequenza di 11 è ....
La frequenza di 11,5 è ....
La frequenza di 12 è ....
La frequenza di 12,5 è ....

Per calcolare la media devi applicare la formula

$$M = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots x_n \cdot f_n}{f_1 + f_2 + \ldots f_n}$$

dove $x_1, x_2,\ldots, x_n$ sono i numeri dati e $f_1, f_2,\ldots, f_n$ le relative frequenze. Quindi:

$$M = \frac{10 \cdot \ldots + 11 \cdot \ldots + \ldots\ldots}{\ldots\ldots} = \ldots\ldots$$

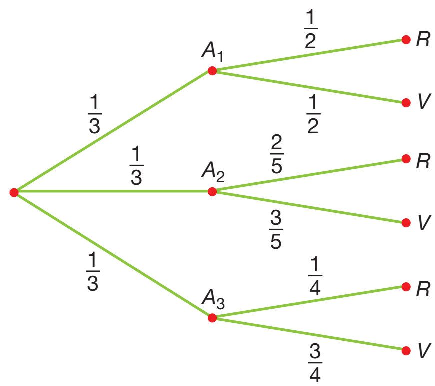
La moda è il valore con frequenza ...., quindi la moda è ....

La mediana è il termine ...., quindi la mediana è ....
(puoi scrivere prima la serie di numeri ordinandola in senso crescente).

4 Si hanno tre urne.

L'urna A_1 contiene 2 palline rosse e 2 palline verdi.
L'urna A_2 contiene 2 palline rosse e 3 palline verdi
L'urna A_3 contiene 1 pallina rossa e 3 palline verdi
Qual è la probabilità che estraendo a caso, da una delle urne, una pallina questa sia rossa?

Rappresentare la situazione con uno schema ad albero



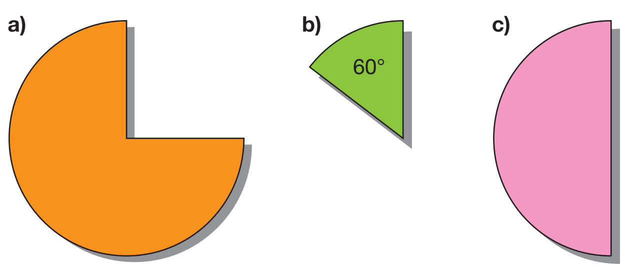
Gli eventi A_1, A_2 e A_3 sono tra loro incompatibili.
Indica con R l'evento «risulta estratta una pallina rossa».

$$\text{prob}(R) = \text{prob}(R \text{ e } A_1) + \text{prob}(R \text{ e } A_2) + \text{prob}\,(R \text{ e } A_3) =$$

$$= \frac{1}{3} \cdot \frac{1}{2} + \frac{\ldots\ldots}{\ldots\ldots} \cdot \frac{\ldots\ldots}{\ldots\ldots} + \frac{\ldots\ldots}{\ldots\ldots} \cdot \frac{\ldots\ldots}{\ldots\ldots} = \ldots\ldots$$

ADESSO TOCCA A TE

5 Calcolare la percentuale rappresentata dalle parti colorate dei seguenti aerogrammi (l'intero è rappresentato da tutto il cerchio).



6 Il libro *Flatlandia*, di Edwin A. Abbott, inizia con questo brano:

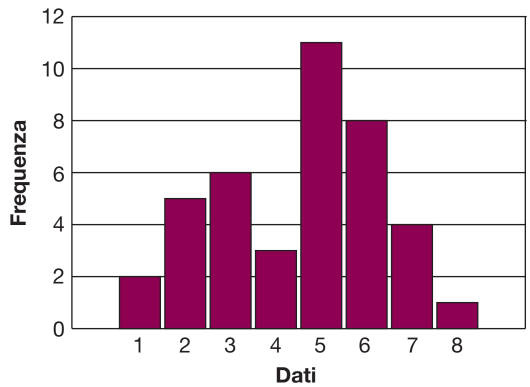
«Chiamo il nostro mondo Flatlandia, non perché sia così che lo chiamiamo noi, ma per renderne più chiara la natura a voi, o lettori beati, che avete la fortuna di abitare nello Spazio.
Immaginate un vasto foglio di carta su cui delle Linee Rette, dei Triangoli, dei Quadrati, dei Pentagoni, degli Esagoni e altre figure geometriche, invece di restare ferme al loro posto, si muovono qua e là, liberamente, sulla superficie o dentro di essa, ma senza potersene sollevare e senza potervisi immergere, come delle ombre, insomma – consistenti, però, e dai contorni luminosi.»
Costruire un optogramma che rappresenti la frequenza delle lettere che vi compaiono.
Qual è la lettera più frequente?

7 In un compito di matematica si sono riportati i seguenti voti:

9,5 5,25 4 8,5 8,5 6,5 8,25 8,5 6,25 7,25 9 5 4,25 6,25 5,5 7,5
3 8 6,25 7,5 7,5 6,5 5,25 6,5 9

Calcolare la media dei voti, la mediana e la moda.

8 Determinare la media e la moda della serie di dati rappresentata dal seguente grafico.



9 Nella seguente tabella sono indicate le temperature minime e massime registrate in una città italiana, durante il mese di agosto 2008.

Giorno	*T* min	*T* max
1	17 °C	29 °C
2	16 °C	29 °C
3	19 °C	29 °C
4	21 °C	29 °C
5	18 °C	31 °C
6	19 °C	27 °C
7	17 °C	26 °C
8	17 °C	29 °C
9	15 °C	27 °C
10	15 °C	27 °C
11	17 °C	25 °C
12	18 °C	24 °C
13	18 °C	27 °C
14	18 °C	27 °C
15	15 °C	21 °C
16	11 °C	25 °C

Giorno	*T* min	*T* max
17	15 °C	24 °C
18	16 °C	25 °C
19	19 °C	26 °C
20	18 °C	27 °C
21	20 °C	27 °C
22	19 °C	25 °C
23	17 °C	27 °C
24	13 °C	25 °C
25	15 °C	25 °C
26	16 °C	26 °C
27	16 °C	27 °C
28	17 °C	28 °C
29	17 °C	29 °C
30	18 °C	29 °C
31	21 °C	26 °C

Calcolare la temperatura media giornaliera, la media delle temperature minime e la media delle temperature massime. Calcolare inoltre, la moda delle temperature minime e quella delle temperature massime.

10 Calcolare la probabilità che nel lancio simultaneo di due dadi si ottengano due numeri la cui somma è 5.

$\left[\frac{1}{9}\right]$

11 In un'urna ci sono 20 palline bianche e 10 palline nere. Calcolare la probabilità di estrarre una pallina nera.

$\left[\frac{1}{3}\right]$

12 Calcolare la probabilità che nel gioco della tombola esca:

a) un multiplo di 7 minore di 50; $\left[\frac{7}{90}\right]$

b) un quadrato multiplo di 4. $\left[\frac{2}{45}\right]$

13 Calcolare la probabilità che estraendo una carta da un mazzo di 40 carte questa:

a) sia di colore rosso; $\left[\frac{1}{2}\right]$

b) sia una figura. $\left[\frac{3}{8}\right]$

14 Nel lancio di due dadi calcola la probabilità che il punteggio complessivo sia

a) pari; $\left[\frac{1}{2}\right]$

b) un numero dispari multiplo di 3. $\left[\frac{1}{6}\right]$

EXTRAMATH

MATEMATICA FUORI DALL'AULA

1 Qual è la media aritmetica dei numeri 1, 2, 3, ...1999,2000?

A) 999 B) 999,5 C) 1000 D) 1000,5 E) 1001

(Giochi di Archimede, 2000)

2 Uno studente ha avuto una media di 6 e 1/2 nei primi quattro compiti. Quale voto deve prendere nel quinto per ottenere la media del 7?

A) 7 e 1/2 B) 8 e 1/2 C) 9 D) 10 E) non ce la può fare.

(Giochi di Archimede, 1997)

3 Michele si prepara all'ultimo compito in classe di matematica dell'anno; lo affronta con tranquillità, sapendo che se prenderà 10 avrà la media del 9, mentre prendendo 5 la media diventerà 8. Quanti compiti ha già fatto quest'anno Michele?

A) 2 B) 3 C) 4 D) 5

E) i dati non sono sufficienti per dare la risposta.

(Giochi di Archimede, 2004) [C]

4 Qual è il più grande degli interi positivi n tali che la media aritmetica dei numeri da 1 a n sia < 2003?
(Nota: la media aritmetica di n numeri è uguale alla loro somma divisa per n.)

A) 2002 B) 2003 C) 4003 D) 4004 E) 4005

(Giochi di Archimede, 2003) [D]

5 In una scatola vi sono quattro sacchetti: il primo sacchetto contiene 4 palline bianche e 3 nere, il secondo 2 palline bianche e 4 nere, il terzo 6 palline bianche e 9 nere, il quarto 5 palline bianche e 10 nere.
Si estrae un sacchetto a caso, e da questo, sempre a caso, si estrae una pallina.
Sapendo che è stata estratta una pallina bianca, quale sacchetto è più probabile che sia stato scelto?

A) il primo
B) il secondo
C) il terzo
D) il quarto
E) tutti i sacchetti hanno la stessa probabilità di essere stati estratti.

(Giochi di Archimede, 1996) [A]

6 In ogni ruota del lotto ci sono 90 numeri. Cinque di essi vengono estratti, uno alla volta, senza rimettere i numeri estratti nell'urna. In una certa ruota viene estratto per primo il numero 27. La probabilità che il secondo estratto sia 28.

A) è $\frac{1}{90}$ B) è $\frac{1}{89}$ C) è $\frac{1}{18}$

D) è minore di $\frac{1}{100}$ perché è improbabile che vengano estratti due numeri consecutivi

E) non si può determinare perche occorre calcolare la probabilità della cinquina.

(Giochi di Archimede, 1999) [B]

7 Qual è la probabilità che, estratti due numeri interi a caso (anche uguali) compresi fra 1 e 12 (estremi inclusi) il loro prodotto sia multiplo di 5?

A) $\frac{1}{5}$ B) $\frac{11}{36}$ C) $\frac{5}{24}$ D) $\frac{1}{4}$ E) nessuna delle precedenti

(Giochi di Archimede, 1999) [B]

8 Quattro squadre di pallacanestro di pari forza disputano un torneo con girone unico all'italiana (ogni squadra incontra ogni altra squadra una sola volta). Qual è la probabilità che ci sia una squadra che alla fine del torneo ha vinto tutte le sue partite? (Le partite di pallacanestro non possono finire con un pareggio.)

A) $\frac{1}{6}$ B) $\frac{1}{\pi}$ C) $\frac{1}{3}$ D) $\frac{1}{2}$ E) $\frac{2}{3}$

(Giochi di Archimede, 1996) [D]

9 Lanciando due dadi regolari con dodici facce, numerate da 1 a 12, la probabilità che la somma dei valori delle facce sia 13 è:

A) $\frac{1}{24}$ B) $\frac{1}{12}$ C) $\frac{13}{144}$ D) $\frac{1}{6}$ E) $\frac{13}{72}$

(Giochi di Archimede, 2001) [B]

10 Quattro ragazzi vogliono telefonare tutti contemporaneamente alle rispettive ragazze. Ogni cellulare può funzionare su quattro frequenze distinte. Se due cellulari si attivano sulla stessa frequenza la comunicazione cade. Se ogni ragazzo non sa che frequenza scelgono gli altri tre, qual è la probabilità che tutti e quattro riescano a parlare con le loro ragazze?

A) $\frac{3}{32}$ B) $\frac{3}{64}$ C) $\frac{1}{256}$ D) $\frac{1}{16}$ E) $\frac{9}{128}$

(Giochi di Archimede, 2003) [A]