

## Cap 1. LA STATISTICA DESCRITTIVA

### Rivedi la teoria

#### La distribuzione di frequenze

La statistica indaga su **fenomeni collettivi** analizzandone uno o più **caratteri**. Ciascun carattere di una indagine statistica si può presentare con diverse **modalità**; il numero di unità statistiche che presentano una particolare modalità è la **frequenza assoluta** di quella modalità.

I caratteri di una indagine statistica possono essere:

- di tipo **qualitativo**, e in questo caso le modalità vengono descritte a parole con aggettivi o nomi,
- di tipo **quantitativo**, e in questo caso le modalità sono espresse mediante numeri o intervalli di numeri.

L'insieme delle coppie ordinate (*modalità, frequenza assoluta*) rappresenta la **distribuzione di frequenze** del carattere e si rappresenta di solito mediante una tabella; in essa, accanto alla colonna delle frequenze assolute, si è soliti riportare anche quella delle frequenze relative, calcolate come rapporto della frequenza assoluta rispetto al totale delle osservazioni.

Per esempio, quella che segue è la tabella delle frequenze che rappresenta il numero di bambini iscritti alla scuola materna al primo, secondo e terzo anno:

Anno di corso	Freq. ass.	Freq. rel.
primo	22	$\frac{22}{65} = 0,338$
secondo	25	$\frac{25}{65} = 0,385$
terzo	18	$\frac{18}{65} = 0,277$
<b>TOTALE</b>	65	1

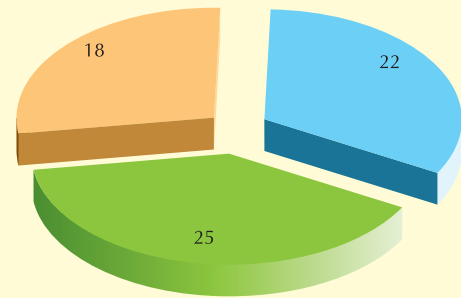
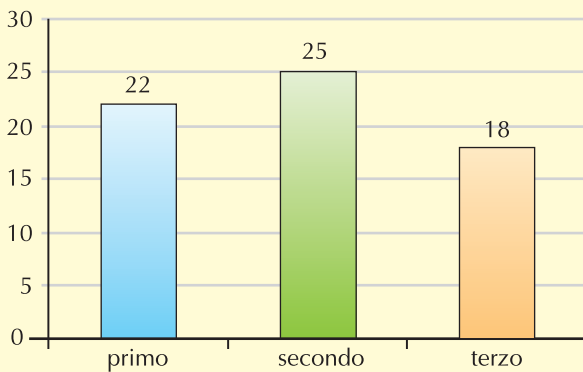
#### La rappresentazione grafica

I dati di una indagine statistica si possono rappresentare graficamente in modi diversi, più o meno adatti al diverso tipo di indagine.

- I **diagrammi a rettangoli** o **diagrammi a barre** sono utili per rappresentare dati di tipo qualitativo; sono i più usati specialmente quando si devono fare confronti fra diverse distribuzioni.
- I **diagrammi circolari** o **diagrammi a torta**, usando la suddivisione di un cerchio in settori, danno una idea dell'incidenza delle diverse modalità del carattere sul totale delle osservazioni.
- Gli **istogrammi** vengono usati per dati di tipo quantitativo; sono rappresentati mediante rettangoli la cui area è proporzionale alla frequenza assoluta.

Altre modalità sono rappresentate dagli ideogrammi, nei quali si utilizzano disegni che ricordano l'oggetto dell'indagine, e dai cartogrammi, nei quali si usano rappresentazioni più o meno dilatate o compresse delle aree geografiche sulle quali si effettua l'indagine.

Le modalità più adatte per la rappresentazione della precedente distribuzione di frequenze sono quella a rettangoli e quella circolare:



## Fai gli esercizi

### 1 ESERCIZIO GUIDA

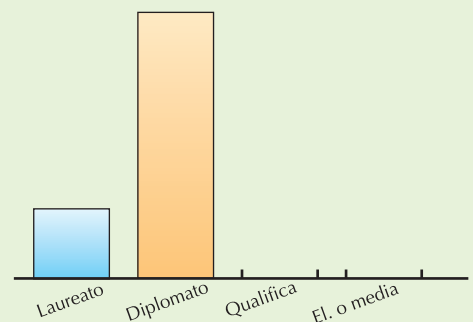
Da un'analisi condotta in una grande azienda risulta che dei 1225 dipendenti 122 sono laureati, 463 sono diplomati, 312 hanno una qualifica professionale ed i rimanenti hanno un titolo di studio elementare o di media superiore.

Costruisci la tabella delle frequenze percentuali e rappresenta graficamente i dati.

Il carattere da analizzare è il titolo di studio dei dipendenti, quindi un carattere qualitativo; si tratta perciò di una mutabile statistica.

La tabella delle frequenze si costruisce mettendo in una colonna le modalità con cui compare il carattere (laureato, diplomato, ecc.), in una seconda colonna la frequenza assoluta di ciascuna modalità; in una terza colonna si possono poi rappresentare le frequenze percentuali dividendo le frequenze assolute per il totale delle osservazioni e moltiplicando il risultato per 100. Completa dunque la tabella e il grafico che seguono:

Titolo di studio	Frequenza assoluta	Frequenza percentuale
Laureato	122	9,96%
Diplomato	463	.....
Qualifica professionale	312	.....
Elementare o media	.....	.....
<b>TOTALI</b>	<b>1225</b>	<b>100%</b>



Per rappresentare i dati conviene usare un diagramma a rettangoli; completa il grafico a lato costruendo i rettangoli di altezza proporzionale alle frequenze.

- 2 Rileva da un quotidiano la valutazione dell'euro nei confronti delle principali monete mondiali (dollaro, euro, sterlina, yen, ...) e riporta i dati in un diagramma a rettangoli.

### 3 ESERCIZIO GUIDA

Un rilevamento fatto per 30 giorni consecutivi sul numero di vetture che transitano per un dato incrocio ha fornito i seguenti dati:

N. auto transitate	0-500	501-1000	1001-1500	1501-1800	1801-2000	2001-2500	2501-3000
Numero giorni	1	3	5	8	8	3	2

Rappresenta i dati con un istogramma.

Fissato un sistema di assi cartesiani ortogonali, devi riportare sull'asse delle ascisse le classi e sull'asse delle ordinate il valore che ottieni dividendo la frequenza per l'ampiezza della classe. Poiché i valori ottenuti sono molto piccoli, devi usare una scala adeguata per rappresentarli.

[altezze dei rettangoli 0,002; 0,006; 0,01; 0,027; 0,04; 0,006; 0,004]

4 Da una indagine sul numero di volte che una persona va al cinema nell'arco di un anno si sono rilevati i seguenti dati:

15	18	20	25	50	15	20	30	18	25	40	25	18
25	30	25	15	20	30	20	18	20	25	40	20	25

Costruisci la distribuzione delle frequenze e rappresenta graficamente i dati.

N. di volte al cinema	15	18	20	25	30	40	50
Frequenza	3	4	6	7	3	2	1

### 5 ESERCIZIO GUIDA

In una scuola, allo scrutinio di giugno, è risultato che il 40% degli studenti è stato promosso, il 7% è stato respinto e ai rimanenti studenti è stato sospeso il giudizio perché insufficienti in alcune materie. Di questi ultimi, il 25% deve riparare una sola materia, il 43% due materie e i rimanenti tre materie. Costruiamo il diagramma a torta della distribuzione e riproduciamo in uno spaccato le percentuali degli studenti con uno, due e tre materie da riparare.

Gli studenti con sospensione del giudizio rappresentano il 53% dell'intera popolazione studentesca.

Per costruire il diagramma a torta bisogna calcolare l'ampiezza dell'angolo al centro di ciascun settore; allora, poiché l'angolo giro (360°) rappresenta il 100%, l'ampiezza di ogni settore si ricava dalla proporzione:

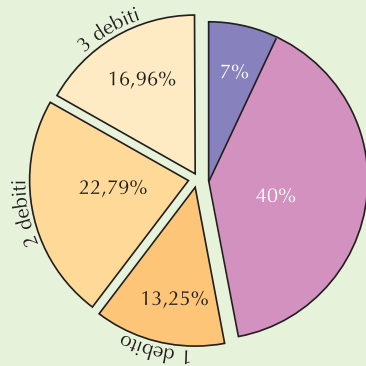
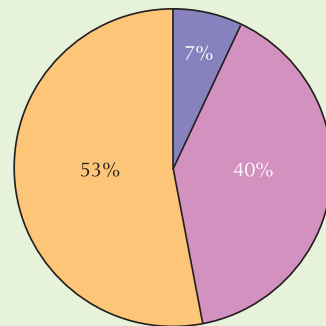
$$\text{ampiezza settore} : 360 = \text{frequenza percentuale} : 100$$

Il settore degli studenti con sospensione del giudizio, che ha ampiezza 190,8°, va poi diviso in tre parti; l'ampiezza in gradi di ciascuna parte si ricava dalla proporzione:

$$\text{ampiezza sottosettore} : 190,8 = \text{frequenza percentuale} : 100$$

In alternativa, si può calcolare la frequenza percentuale relativa al totale delle rilevazioni e usare la formula precedente:

- il 25% del 53% è  $\frac{25 \cdot 53}{100} = 13,25\%$  studenti con una materia



- il 43% del 53% è  $\frac{43 \cdot 53}{100} = 22,79\%$  studenti con due materie
- il 32% del 53% è  $\frac{32 \cdot 53}{100} = 16,96\%$  studenti con tre materie

6 Un'azienda ha rilevato i dati degli ultimi cinque anni relativi alla vendita di tre dei suoi prodotti:

Anni	2006	2007	2008	2009	2010
Prodotto A	1563	1824	1945	1630	845
Prodotto B	7496	8502	8854	9324	9596
Prodotto C	12754	13820	13569	13602	13428

Individua la tipologia dei dati, costruisci la tabella delle frequenze percentuali e rappresenta i dati nel modo che ritieni più opportuno per effettuare un confronto fra le distribuzioni anno per anno e prodotto per prodotto.

Tabella percentuale per anno di produzione.

Anni	2006	2007	2008	2009	2010
Prodotto A	7,17%	7,55%	7,98%	6,64%	3,54%
Prodotto B	34,36%	35,21%	36,33%	37,97%	40,20%
Prodotto C	58,47%	57,24%	55,68%	55,39%	56,26%

Tabella percentuale per prodotto.

Anni	2006	2007	2008	2009	2010
Prodotto A	20,02%	23,36%	24,91%	20,88%	10,82%
Prodotto B	17,13%	19,42%	20,23%	21,30%	21,92%
Prodotto C	18,99%	20,57%	20,20%	20,25%	19,99%

## Rivedi la teoria

Quando si svolge un'indagine statistica, è comodo sintetizzare i dati mediante un solo valore; si ricorre allora al calcolo di un valore medio, che può essere interpretato, a seconda delle situazioni, in diversi modi.

### Le medie ferme

Di un insieme  $x_1, x_2, \dots, x_n$  di  $n$  numeri si possono calcolare diversi tipi di media:

- la **media aritmetica**  $M$ : 
$$M = \frac{x_1 + x_2 + \dots + x_n}{n}$$
- la **media geometrica**  $M_G$ : 
$$M_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$
- la **media quadratica**  $M_Q$ : 
$$M_Q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$
- la **media armonica**  $M_A$ : 
$$M_A = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Per esempio, dati i seguenti numeri: 12 8 24 16 32 le diverse medie, arrotondate al centesimo, sono:

• media aritmetica:  $M = \frac{12 + 8 + 24 + 16 + 32}{5} = 18,4$

• media geometrica:  $M_G = \sqrt[5]{12 \cdot 8 \cdot 24 \cdot 16 \cdot 32} = 16,38$

• media quadratica:  $M_Q = \sqrt{\frac{12^2 + 8^2 + 24^2 + 16^2 + 32^2}{5}} = 20,32$

• media armonica:  $M_A = \frac{5}{\frac{1}{12} + \frac{1}{8} + \frac{1}{24} + \frac{1}{16} + \frac{1}{32}} = 14,55$

### Le medie lasche

Altri due valori sintetici sono:

- la **moda**, che rappresenta il valore di una distribuzione che ha la frequenza maggiore
- la **mediana**, che rappresenta il valore centrale di una distribuzione, vale a dire quel valore che divide in due parti simmetriche i dati, una volta che siano stati disposti in ordine crescente o decrescente.

Se riprendiamo la distribuzione dei bambini alla scuola materna della prima nota di ripasso della teoria, possiamo dire che:

- la moda è rappresentata dal secondo anno che ha il maggior numero di bambini iscritti;
- la mediana, poiché il totale delle osservazioni è 65, è rappresentata dal 33-esimo elemento, che, messi in ordine di anno i bambini, appartiene al secondo anno di corso.

Altri indici importanti per l'interpretazione dei dati di una indagine statistica sono gli **indici di dispersione**:

- lo **scarto quadratico medio**  $\sigma$  ci dà informazioni su come si disperdono i dati attorno alla media aritmetica  $M$ . Si calcola con la formula:

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}}$$

- la **varianza**  $\sigma^2$  è il quadrato dello scarto quadratico medio e viene usato in alternativa a  $\sigma$ .

Nel confronto fra distribuzioni con unità di misure differenti o valori medi molto diversi tra loro si usano i **coefficienti di variazione** (CV) definiti come rapporto fra scarto quadratico medio e media aritmetica o fra scostamento medio e mediana:

$$CV = \frac{\sigma}{M} \quad \text{oppure} \quad CV = \frac{S}{M_e}$$

### Fai gli esercizi

7 Un certo prodotto è stato venduto in cinque supermercati diversi ai seguenti prezzi (in euro):

12,54    12,65    11,95    12,70    13,10

Calcola il prezzo medio.

[12,588]

8 In un serie di test di matematica, uno studente ha riportato i seguenti voti: tre volte 6, due volte 7, una volta 5, una volta 4, tre volte 8. Qual è il voto medio?

[6,5]

9 Calcola la media aritmetica della seguente distribuzione:

modalità	2	4	6	8	10
frequenza	8	12	20	24	18

(Suggerimento: si tratta di una media ponderata che si calcola con la formula  $M = \frac{\sum x_i f_i}{\sum f_i}$  [ $\approx 6,78$ ])

## 10 ESERCIZIO GUIDA

Riprendiamo la distribuzione dell'esercizio 4 precedente (numero di volte che una persona va al cinema in un anno) e troviamo la moda e la mediana.

Ricostruiamo la tabella delle frequenze e troviamo anche le frequenze cumulate per il calcolo della mediana.

n. film visti	15	18	20	25	30	40	50
frequenza	3	4	6	7	3	2	1
freq. cumulate	3	7	13	20	23	25	26

La frequenza assoluta più alta è 7 che si trova in corrispondenza del 25, quindi 25 è la moda della distribuzione.

La metà del totale delle frequenze è  $26 : 2 = 13$ , il valore mediano si trova quindi in corrispondenza della modalità che assume il tredicesimo posto, cioè 20; la mediana è il valore 20.

11 Trova la moda e la mediana della seguente distribuzione statistica che riguarda il numero di volte che un gruppo di ragazzi sono stati interrogati in una certa materia:

interrogazioni	2	3	4	5	6	7	8
frequenza	10	15	20	28	18	12	4

[moda: 5; mediana: 5]

## 12 ESERCIZIO GUIDA

Calcoliamo lo scarto quadratico medio della seguente rilevazione statistica che riguarda il numero di pezzi prodotti da una macchina utensile in alcuni giorni consecutivi:

13      14      12      10      15      10      9      13

Calcoliamo innanzi tutto la media aritmetica:  $\frac{13 + 14 + 12 + 10 + 15 + 10 + 9 + 13}{8} = 12$

Calcoliamo gli scarti dalla media: 1   2   0   -2   3   -2   -3   1

Calcoliamo i quadrati degli scarti: 1   4   0   4   9   4   9   1

Calcoliamo la somma dei quadrati degli scarti:  $1 + 4 + 4 + 9 + 4 + 9 + 1 = 32$

Dividiamo per il numero delle osservazioni:  $\frac{32}{8} = 4$

Calcoliamo la radice quadrata del valore ottenuto: 2

Lo scarto quadratico medio è 2.

**13** Seguendo l'esempio precedente, calcola lo scarto quadratico medio della seguente distribuzione di dati:

15 39 24 33 12 9 4 25 72 84.

[ $\approx 25,43$ ]

**14** Il Signor Rossi possiede un pacchetto di azioni di due società diverse che indicheremo con *A* e *B*. Esse, durante la settimana appena trascorsa, hanno subito delle variazioni nelle loro quotazioni in Euro come indicato dalla seguente tabella:

	<b>A</b>	<b>B</b>
lunedì	17,24	8,64
martedì	18,21	7,92
mercoledì	19,14	6,87
giovedì	18,95	8,52
venerdì	18,91	8,91

Quale delle due società ha avuto maggiore variabilità?

[ $\sigma_A = 0,70$ ;  $\sigma_B = 0,73$ ;  $CV_A = 0,04$ ;  $CV_B = 0,09$ ]

# Verifica del recupero

- 1 Costruisci la distribuzione di frequenze dei seguenti dati che indicano la temperatura registrata alla stessa ora in 20 giorni consecutivi:

12 15 18 10 12 12 15 15 12 10 15 18 18 15 12 12 15 15 12 15

2 punti

- 2 Rappresenta graficamente nel modo che ritieni più opportuno la seguente distribuzione di frequenze:

modalità	1	2	3	4	5	6	7
frequenza	12	15	16	25	18	10	5

2 punti

- 3 Calcola media aritmetica, moda e mediana della distribuzione di frequenze dell'esercizio 2 precedente.

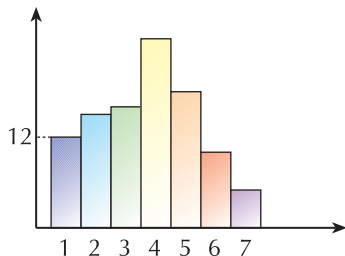
3 punti

- 4 Calcola lo scarto quadratico medio e la varianza della distribuzione dell'esercizio 2 precedente.

3 punti

## Soluzioni

1	temperature	10	12	15	18
	frequenza	2	7	8	3



- 3  $M = 3,71$ , moda = 4, mediana = 4

- 4  $\sigma = 1,66$ ;  $\sigma^2 = 2,76$

Esercizio	1	2	3	4	
Punteggio					

Valutazione  
in decimi





# Math in English

## Glossary

**average value to estimate frequency histogram mean median**

valore medio stimare frequenza istogramma media aritmetica mediana

**mode range standard deviation**

**variance**

moda campo di variabilità deviazione standard, scarto quadratico medio varianza



- 1 Throwing a dice we have obtained the following results:

<b>Score</b>	1	2	3	4	5	6
<b>Frequency</b>	18	16	25	19	22	20

Find:

- a. the mean score      b. the mode      c. the median.

- 2 The following table represents the surface (in square metres) of some small holiday homes in a residential hotel.

<b>Surface</b>	[20 – 40)	[40 – 50)	[50 – 55)	[55 – 60)	[60 – 90]
<b>Frequency</b>	10	12	10	7	9

For the above data:

- a. estimate the mean and the standard deviation  
b. draw a histogram and state the modal class  
c. draw a cumulative frequency curve.

- 3 Two systems for measuring the calcium content of laboratory samples were being considered. A standard sample, containing 100 units of calcium was weighed 16 times using each system. The results were:

### System "A"

105	102	102	105	106	107	108	108
104	105	107	104	103	105	107	105

### System "B"

95	95	99	102	97	107	98	103
101	103	109	100	100	91	98	96

Which system is potentially the "best"?

3  $\sigma_A = 1,91; \sigma_B = 4,59$

2 mean = 50,9; standard deviation = 14,68

1 3,59; 3; 4