

## UD1 - LA STATISTICA

### 1.1 Introduzione

La statistica si occupa di studiare fenomeni naturali, sociali, economici, ecc.

L'**unità statistica** è l'unità elementare, rispondente a determinate caratteristiche, oggetto di studio. Un insieme di unità statistiche forma una **popolazione**.

La caratteristica della popolazione che viene osservata è definita **carattere**.

Il carattere si manifesta secondo diverse **modalità**.

Se le modalità sono espresse attraverso parole si parla di **caratteri qualitativi**, se sono espresse con valori numerici, di **caratteri quantitativi**.

### 1.2 Il rilevamento dei dati

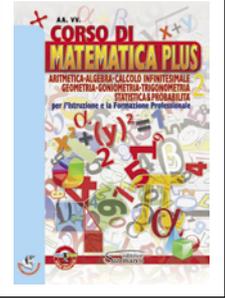
Le informazioni che si raccolgono sono dette **dati statistici**.

Il rilevamento dei dati può essere effettuato in modo completo o per **campione**.

Nel primo caso si parla di statistica descrittiva. Nel secondo caso di statistica inferenziale.

### 1.3 L'elaborazione e l'interpretazione dei dati

I dati rilevati durante un'indagine statistica vengono elaborati e sintetizzati in tabelle chiamate **matrici dati**. Queste tabelle contengono uno o più caratteri e l'indicazione della loro frequenza.



La **frequenza assoluta** indica il numero delle volte in cui una certa modalità del carattere considerato si manifesta.

La **frequenza relativa** ( $f_r$ ) permette di interpretare correttamente i dati anche quando si riferiscono a popolazioni statistiche con un numero di componenti diversi.

$$f_r = \frac{\text{numero di volte che si osserva la modalità}}{\text{numero componenti la popolazione}}$$

Molto spesso si preferisce utilizzare la **frequenza percentuale** ( $f_p$ ).

$$\frac{\text{numero di volte che si osserva la modalità}}{\text{numero componenti la popolazione}} = 100$$

## 1.4 La sintesi dei dati

Per riassumere i dati ci si avvale di **indici di posizione** (media, mediana, moda) e di **indici di dispersione** (campo di variazione, scarto).

### MEDIA ARITMETICA

La media è molto utile nella descrizione sintetica dei fenomeni, ma può essere influenzata dalla presenza di dati estremi.

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

Dando maggior peso ai dati che compaiono più volte si ottiene la media aritmetica ponderata.

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

### MEDIANA

La mediana tiene conto del dato centrale, quindi, non è influenzata dai valori estremi.

Per calcolarla è necessario disporre i dati in ordine crescente o decrescente.

Se i dati sono in numero **dispari**, la mediana corrisponde al valore che occupa il posto centrale.

Se i dati sono in numero **pari**, la mediana corrisponde alla **semisomma dei due valori centrali**.

### MODA

La moda è il dato che si presenta con la massima frequenza. Una stessa distribuzione può avere più mode.

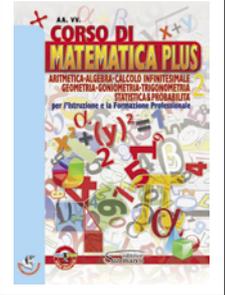
### CAMPO DI VARIAZIONE

Il campo di variazione è l'indice di dispersione più semplice. Si ottiene dalla differenza tra i valori estremi dei dati considerati.

### SCARTO SEMPLICE MEDIO

Lo scarto semplice medio permette di valutare l'affidabilità della media: più è piccolo, più la media è un buon indicatore. Per calcolarlo è necessario:

- 1) calcolare la media dei dati;
- 2) calcolare la differenza tra ogni valore e la media (scarto assoluto);
- 3) calcolare la media degli scarti assoluti.

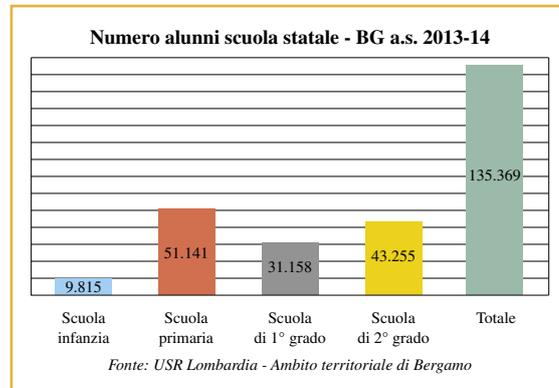


## 1.5 La presentazione dei dati

I risultati di un'indagine statistica possono essere rappresentati mediante grafici che rendono immediatamente chiara l'esposizione dei dati.

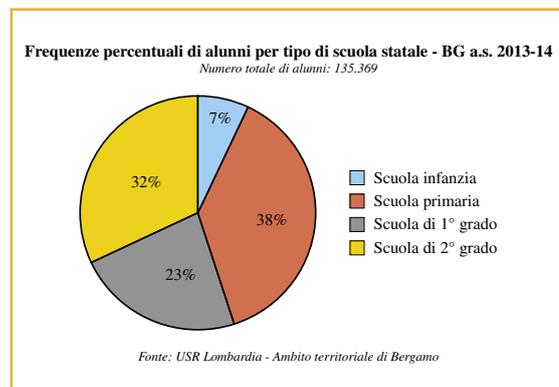
### GRAFICI A BARRE

In figura è proposto un ortogramma a colonne con le modalità sulle ascisse e la frequenza sulle ordinate. Scambiando ascisse e ordinate si otterrebbe un ortogramma a nastri.



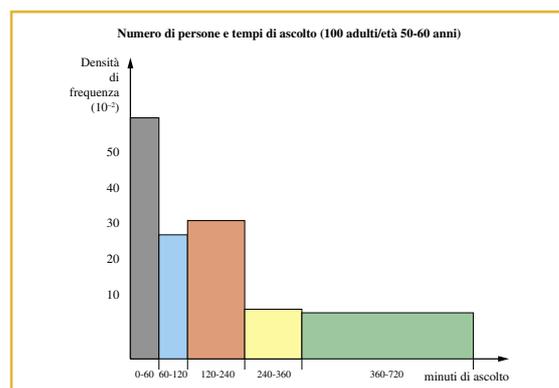
### DIAGRAMMI A TORTA

Sono usati quando si vuol evidenziare il contributo della singola parte rispetto al totale.



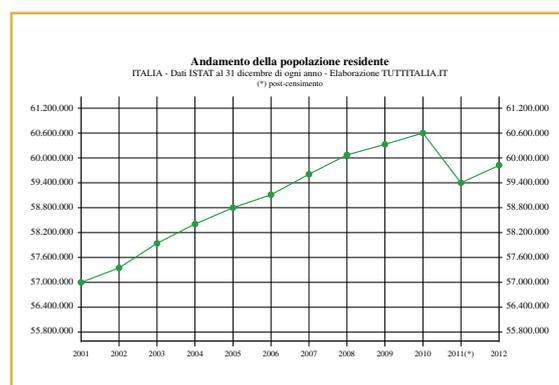
### ISTOGRAMMI

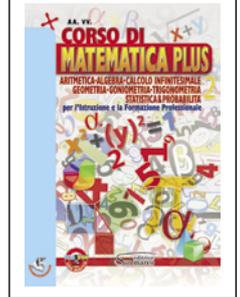
Sono usati per rappresentare fenomeni relativi a classi di grandezze quantitative continue.



### DIAGRAMMI CARTESIANI

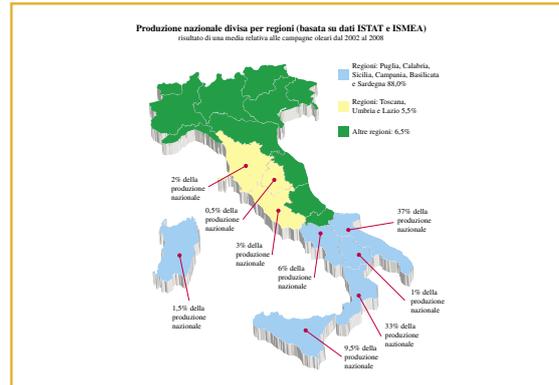
Sono usati per rappresentare l'evoluzione nel tempo di fenomeni statistici.





## CARTOGRAMMI

I dati vengono inseriti all'interno di cartine geografiche (colori graduati, simboli, ecc.).



# UD2 - CALCOLO DELLE PROBABILITÀ

## 2.1 La probabilità: introduzione

Un evento che si verifica sicuramente è detto **certo**.

Un evento che può verificarsi o meno è detto **aleatorio**.

Un evento che non si verifica sicuramente è detto **impossibile**.

## 2.2 La probabilità classica di un evento

Il calcolo delle probabilità indica il grado di fiducia della realizzazione di un evento aleatorio.

Si definisce probabilità di un evento  $E$  e si indica con  $p(E)$  il **rapporto** fra il numero dei casi favorevoli  $m$  e quello dei casi possibili  $n$ :

$$p(E) = \frac{m}{n}$$

Due eventi  $E$  ed  $\bar{E}$  (non  $E$ ) sono **complementari** se uno è la negazione dell'altro.

$$p(E) + p(\bar{E}) = 1$$

## 2.3 Calcolare la probabilità totale in caso di eventi incompatibili e compatibili

### EVENTI INCOMPATIBILI

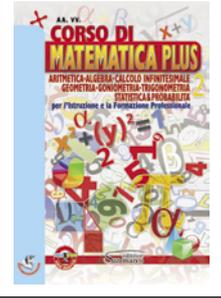
Più eventi sono **incompatibili** se non possono verificarsi simultaneamente in una data prova.

$$\text{Probabilità totale: } p(E_1 \cup E_2) = p(E_1) + p(E_2)$$

### EVENTI COMPATIBILI

Più eventi sono **compatibili** se c'è la possibilità che si verifichino simultaneamente in una data prova.

$$\text{Probabilità totale: } p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$



## 2.4 Calcolare la probabilità composta in caso di eventi compatibili indipendenti e dipendenti

### EVENTI INDIPENDENTI

Due eventi  $E_1$  e  $E_2$  sono **indipendenti** se la probabilità del verificarsi dell'evento  $E_2$  non dipende dal fatto che l'evento  $E_1$  si sia verificato o no (e viceversa).

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2)$$

### EVENTI DIPENDENTI

Due eventi  $E_1$  e  $E_2$  sono **dipendenti** se la probabilità del verificarsi dell'evento  $E_2$  dipende dal fatto che l'evento  $E_1$  si sia verificato o meno.

$$p(E_1 \cap E_2) = p(E_1) \cdot p(E_2)$$

ma qui  $p(E_2)$  indica la probabilità dell'evento  $E_2$ , nell'ipotesi che  $E_1$  si sia verificato.

## 2.5 La probabilità frequentista (statistica) di un evento

Se non si possono conoscere il numero di casi possibili e di casi favorevoli, si ricorre a una diversa definizione di probabilità: la **probabilità frequentista**.

Tale definizione si ottiene dopo aver effettuato un numero, anche molto elevato, di prove relative al verificarsi o meno dell'evento  $E$  in esame.

$$f(E) = \frac{\text{numero di volte in cui } E \text{ si è verificato}}{\text{numero di prove effettuate}}$$

Può capitare che un successivo esperimento porti a risultati diversi, sempre però racchiusi nell'intervallo:  $0 < f < 1$ .

Secondo la **legge dei grandi numeri**, con un numero elevato  $n$  di prove, la probabilità dell'evento nella teoria frequentista tende ad avvicinarsi alla probabilità in senso classico. Diventa quindi calcolabile.

## 2.6 La probabilità soggettiva di un evento

Quando, dato un evento, non sono determinabili casi possibili e favorevoli e neppure sono possibili prove ripetute, la valutazione della probabilità di  $E$  dipende dall'individuo.

La **probabilità soggettiva** di un evento  $E$  è il prezzo  $p$  ( $0 < p < 1$ ) che un **individuo coerente** ritiene equo pagare per ricevere 1 se l'evento si verifica e 0 se non si verifica.

### TEORIA ASSIOMATICA

La probabilità di un evento  $E$  è un numero unico, maggiore o uguale a 0:  $p(E) \geq 0$

La probabilità dell'evento certo è sempre 1:  $P(\text{evento certo}) = 1$

La probabilità che, dati due eventi incompatibili  $E_1$  e  $E_2$ , si verifichi uno o l'altro è la somma delle singole probabilità di  $E_1$  e  $E_2$ :  $p(E_1 \cup E_2) = p(E_1) + p(E_2)$