Esperienza 2: Il foglio elettronico e la distribuzione degli errori casuali

Materiali a disposizione: un pc con installato il programma Microsoft Excel (o in alternativa si può utilizzare i Fogli di Google tramite un account Gmail);

Scopo dell'esperienza: quest'esperienza ha la finalità di introdurre gli studenti all'utilizzo di un foglio elettronico di calcolo per poter essere in grado di maneggiare ed analizzare molteplici dati.

I fogli di calcolo:

Un documento di un foglio di calcolo consente di poter gestire dati e numeri, creare grafici: viene visualizzato come un foglio suddiviso in *colonne*, ciascuna contrassegnata da una lettera dell'alfabeto, e in *righe* numerate: si possono aggiungere colonne e righe cliccando con il pulsante destro sulla lettera che contrassegna una colonna o sul numero della riga e quindi cliccando su *Inserisci*.

L'intersezione di una riga con una colonna forma una *cella*, individuabile attraverso un riferimento composto dalla lettera che contraddistingue la colonna e dal numero della riga.

Le celle:

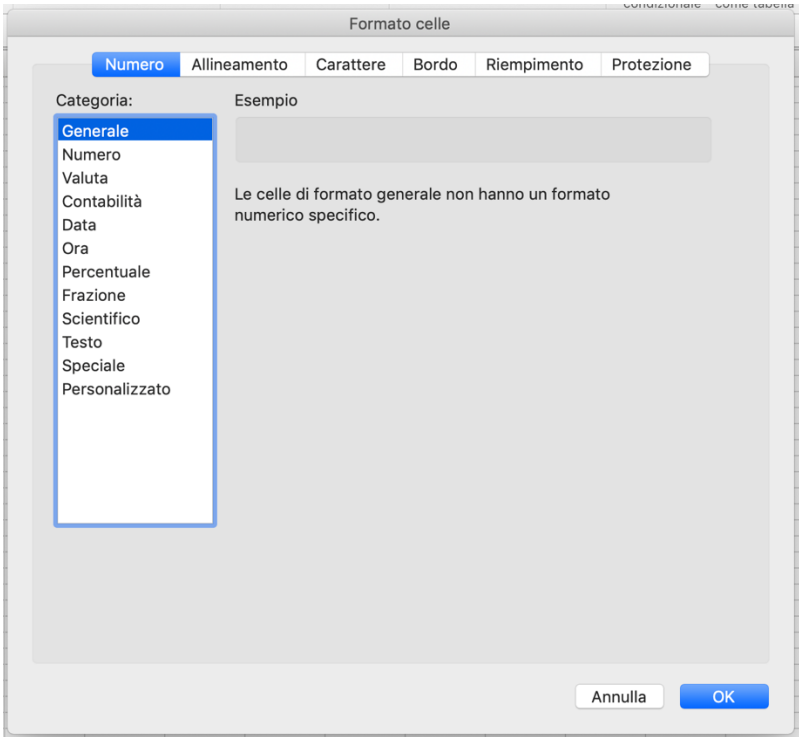
Nelle celle che compongono il foglio elettronico si possono inserire tre tipi di dati fondamentali:

- *dati numerici*: comprendono le 10 cifre, i segni + e –, la virgola decimale. Su questi dati possono essere eseguiti calcoli ed operazioni. Un tipo particolare di valori sono le date che si presentano separando giorno, mese ed anno in diversi formati come ad esempio gg/mm/aa o gg-mm-aa;
- *dati di testo*: comprendono qualunque successione di caratteri e possono contenere, oltre alle lettere dell'alfabeto, tutti i simboli di interpunzione. Hanno funzione di commento per i valori inseriti nelle varie celle, allo scopo di rendere il foglio di facile

consultazione anche per chi non ha partecipato alla sua costruzione. Se la lunghezza del testo supera quella della colonna si visualizza sulla cella vicina, purché sia vuota;

- *formule*: devono iniziare con il simbolo = e possono contenere le parentesi (), le quattro operazioni + - * / e l'elevamento a potenza ^. Vengono calcolate seguendo le normali regole algebriche e visualizzano il risultato numerico nella cella. In una formula si possono indicare altre celle, attraverso il loro riferimento, con l'avvertenza che devono sempre contenere valori numerici (altrimenti viene visualizzato un messaggio d'errore all'interno della cella). Se una formula contiene il riferimento ad una cella, il proprio valore cambia quando cambia quello della cella richiamata.

In generale cliccando con il tasto destro sulla cella e quindi su *Formato celle*… si possono visualizzare tutti i diversi tipi di celle.



Le celle si selezionano semplicemente andando a cliccarci sopra con il mouse: un contorno verde appare ogni qual volta una cella viene attivata; se si tiene premuto su di essa e ci si sposta con il mouse in orizzontale ed in verticale si può selezionare un'area più estesa del foglio di calcolo. Una cella può essere anche spostata da una posizione ad un'altra andando a posizionare il mouse sul bordo: una volta che sarà comparsa una manina, cliccate e spostate il mouse in orizzontale ed in verticale. La cella si sposta di pari passo e quando si rilascia il mouse la stessa va a posizionarsi dove è stato indicato. Qualora la cella di destinazione sia stata già riempita, il proprio contenuto sarà sostituito da quello della cella di arrivo.

Inserire e modificare dati:

Prima di digitare un dato, di qualunque tipo, occorre rendere attiva la cella in cui si vuole inserirlo; al termine dell'operazione di inserimento, per confermarlo, si utilizza una delle seguenti procedure:

- premere il tasto Invio presente sulla tastiera;
- spostarsi su un'altra cella con le frecce direzionali della tastiera;
- se si sceglie di non confermare la digitazione, occorre fare clic sul pulsante Esc della tastiera.

Per modificare invece i dati inseriti precedentemente, si può:

- fare doppio clic sulla cella da modificare, quindi correggere il suo contenuto e confermare;
- posizionare il cursore sulla cella e digitare il nuovo contenuto che alla conferma sostituirà per intero il precedente;
- posizionare il puntatore del mouse sulla barra della formula nel punto in cui si visualizza il contenuto della cella attiva e apportare le modifiche volute.

Copiare il contenuto di una cella:

Sul foglio elettronico si può copiare il contenuto di celle in altre: se si copiano celle contenenti numeri o testo, questi rimangono invariati. Se la copia riguarda celle contenenti formule con riferimenti ad altre celle, questi possono cambiare o rimanere invariati, a seconda del tipo di riferimento che perciò sarà detto *relativo* o *assoluto*.

Per copiare una cella o una zona di celle si può seguire una delle seguenti procedure:

- indicare con il puntatore del mouse l'angolo in basso a destra della cella; quando la freccia del puntatore diventa una croce, trascinare tenendo premuto il pulsante sinistro del mouse fino alla cella desiderata. Questa procedura è possibile se le celle da copiare sono contigue e si utilizza in special modo quando si vuole copiare una formula da una cella ad un'altra;
- si utilizzano i pulsanti Copia e Incolla o la combinazione Ctrl+C (per Copia) e Ctrl+V (per Incolla) che diventano rispettivamente cmd+C e cmd+V sulla tastiera di un Mac.

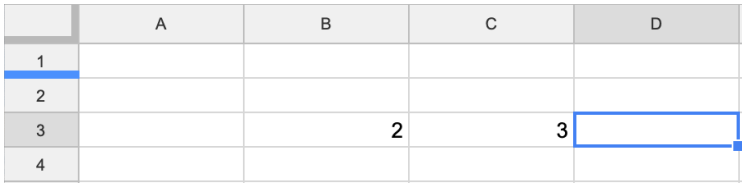
Le formule:

Come già anticipato, le celle del foglio di calcolo possono ospitare formule che consentono di effettuare calcoli. La sintassi prevede che debbano iniziare con il simbolo = . Ad esempio se vogliamo sommare due numeri scriveremo:

=2+3

Ed il foglio restituirà il risultato del calcolo 5.

Le formule possono contenere anche i riferimenti ad altre celle; supponiamo infatti di avere la situazione in figura e di andare a voler calcolare nella cella D3 il risultato di 2+3.



Il foglio di calcolo consente di inserire nella formula i riferimenti alle celle B3 e C3 senza dover andare ad inserire manualmente i valori da sommare. Si possono utilizzare due procedure.

- si scrive in D3:

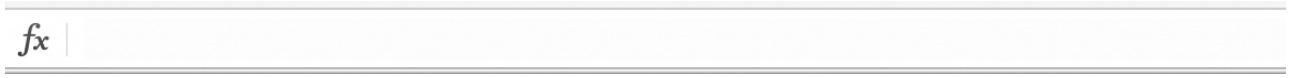
  =B3+C3

- scrivere in D3 il carattere =, cliccare sulla cella B3 il cui bordo ed interno si colorerà, quindi digitare + e cliccare nuovamente sulla cella C3.

Entrambe le procedure restituiscono il valore 5. Il vantaggio di usare questo metodo consiste nel fatto che, qualora il valore della cella B3 o C3 viene cambiato, il risultato nella cella D3 viene aggiornato automaticamente senza doverlo cambiare manualmente ogni volta.

Le formule usano le quattro operazioni matematiche rappresentate dagli operandi di somma +, sottrazione -, moltiplicazione *, divisione / ed elevamento a potenza ^.

Oltre alle formule che consentono di eseguire le operazioni matematiche elementari, il programma mette a disposizione una serie di funzioni standard il cui elenco può essere trovato nella *Barra delle formule* cliccando sul simbolo f_x



Ogni funzione inizia con il simbolo di uguale =, possiede un nome (scritto in maiuscolo) che la identifica e una serie di ***argomenti*** o parametri (posti tra parentesi) su cui agiscono. Alcune funzioni possiedono più argomenti che in tal caso devono essere separati con il *punto e virgola*.

Una funzione di esempio è la funzione SOMMA la cui sintassi è.

$$= SOMMA(arg1;\ arg2;\ \ldots;\ argN)$$

dove $arg1;\ arg2;\ \ldots;\ argN$ sono rappresentati da zone di celle che si vogliono sommare.

Costruire gli istogrammi manualmente:

Per costruire un istogramma dobbiamo prima organizzare gli N dati: in una colonna mettiamo tutti i dati mentre in altre due colonne andiamo a costruire classi e frequenze avendo cura di mettere le classi a sinistra delle frequenze.

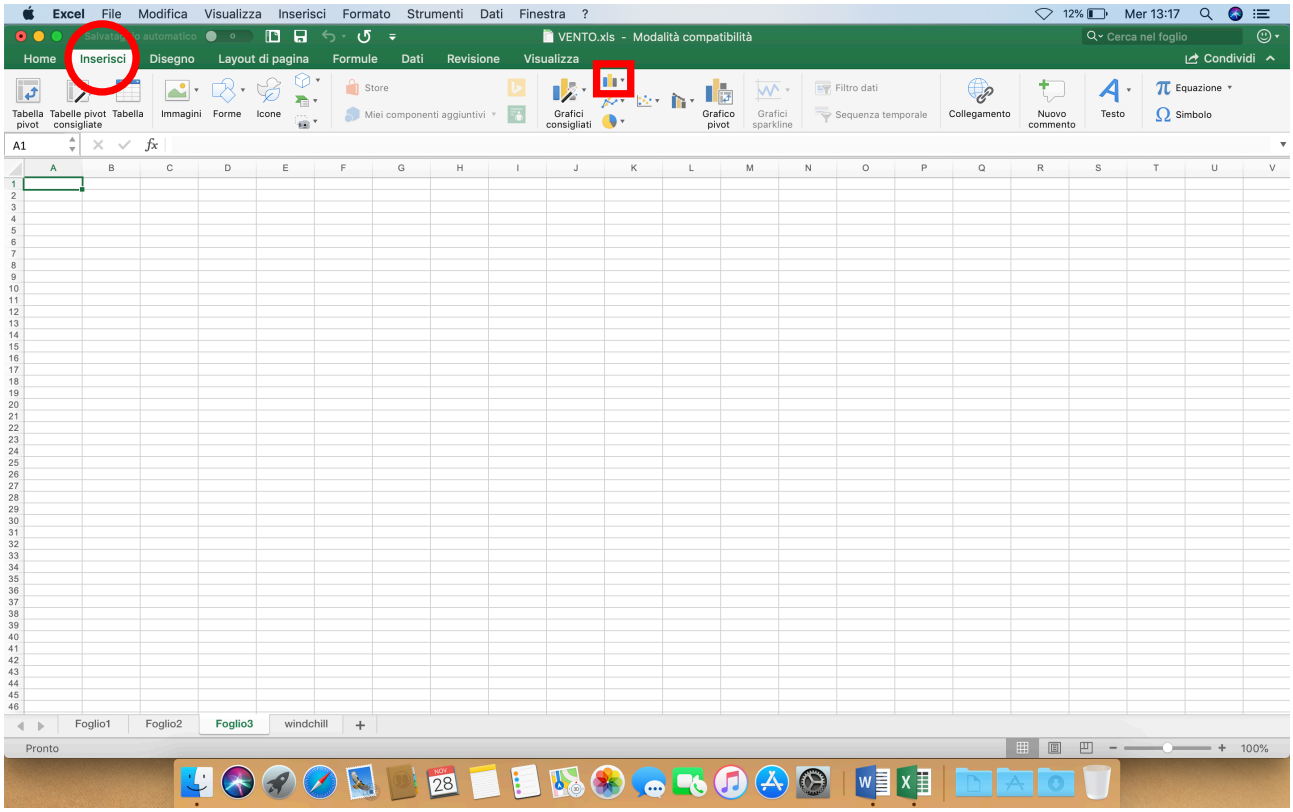
Le prime devono essere create in un numero adeguato n ovverosia tale per cui non ce ne siano tante con troppi dati al proprio interno o viceversa. Esiste un buon criterio empirico che consiste nel rispettare la regola:

$$\frac{N}{n} \geq 10$$

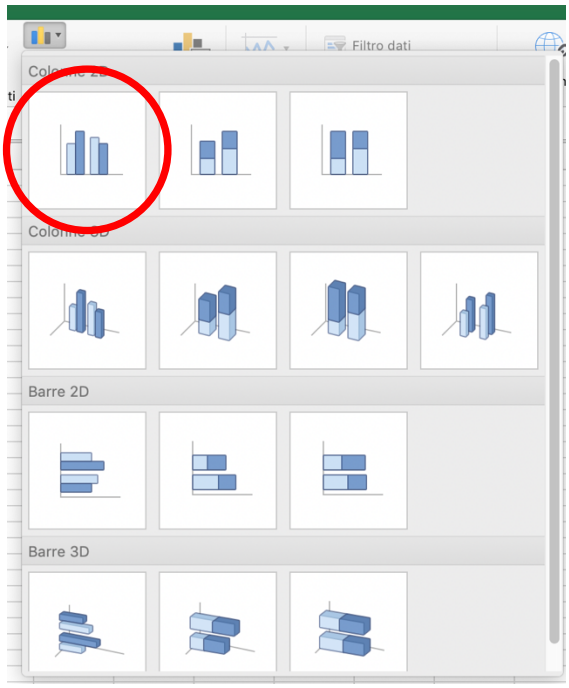
Stabilite le classi, si devono calcolare le frequenze f_i ovvero il rapporto tra la quantità di dati che cadono nella corrispondente classe N_i ed il numero di dati totali N

$$f_i = \frac{N_i}{N}$$

Create classi e frequenze si può andare a costruire l'istogramma: selezioniamo con il mouse le celle delle classi e delle frequenze, avendo avuto cura di averle sistemate in quest'ordine (per selezionare colonne non adiacenti si usa il pulsante CTRL con Windows o CMD con Mac) e andiamo sulla sezione *Inserisci* (cerchio rosso), cliccando nella sezione dei grafici *Istogrammi* (quadrato rosso).



Si visualizza un Menù a tendina come quello mostrato nella figura seguente e selezioniamo in *Colonne 2D* il primo grafico sulla sinistra *Colonne raggruppate*:

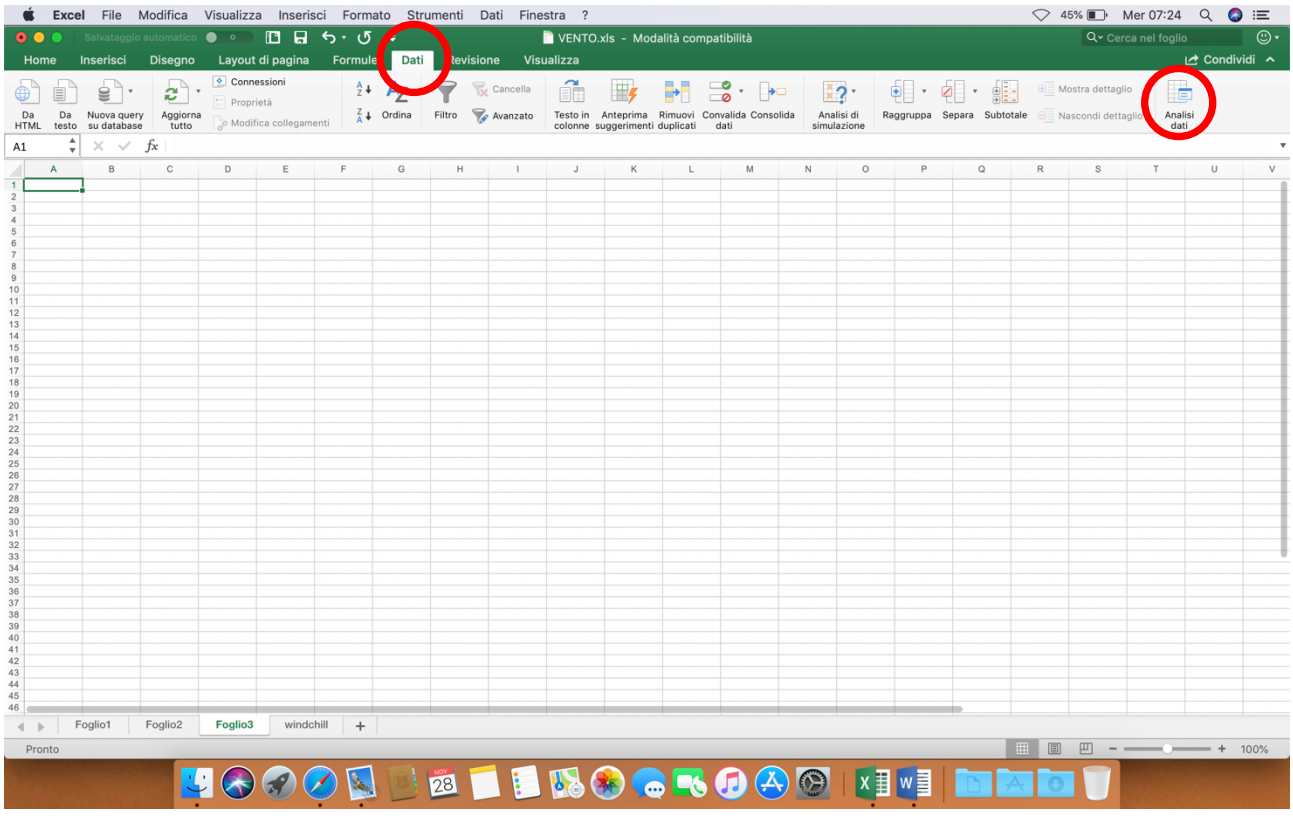


Oltre agli istogrammi si possono creare altri grafici che possiamo trovare nella figura precedente a questa vicino il quadratino rosso.

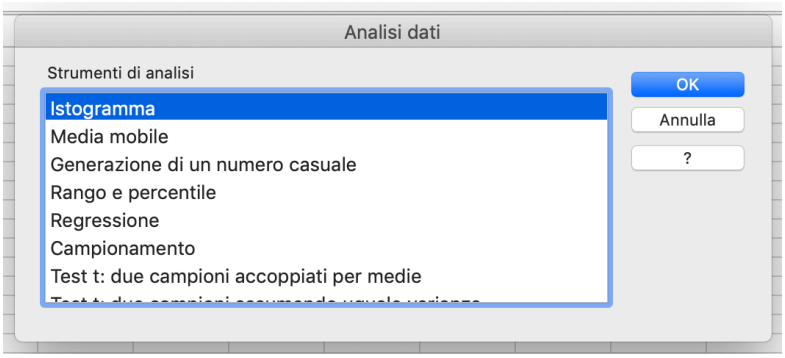
Costruire gli istogrammi in automatico:

Abbiamo un elenco di dati che vogliamo rappresentare sotto forma di istogramma.

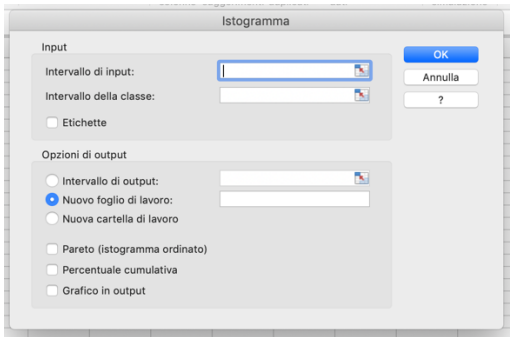
Nella Barra degli Strumenti andiamo su *Dati* e selezioniamo *Analisi Dati*: lo strumento crea in automatico la tabella delle frequenze e l'istogramma.



Dal Menù che compare selezioniamo *Istogramma*

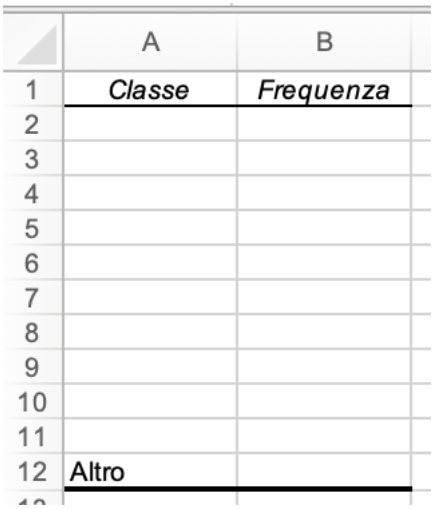


Viene visualizzata la seguente finestra con diverse opzioni:



- ***Intervallo di input***: cliccando riquadro a destra nella casella rettangolare si può andare a selezionare nel foglio di calcolo l'intervallo dei dati da andare ad analizzare.
- ***Intervallo della classe***: consente di andare a selezionare nel foglio di calcolo i valori che definiscono l'estremo sinistro delle classi entro le quali racchiudiamo i nostri dati.
- ***Intervallo di output***: permette di selezionare nel foglio di calcolo la destinazione della tabella delle frequenze.
- ***Nuovo foglio di lavoro***: crea la tabella delle frequenze in un nuovo foglio della cartella (cioè del documento).
- ***Nuova cartella di lavoro***: crea la tabella delle frequenze in una nuova cartella.
- ***Pareto***: permette di presentare i dati nella tabella di output in ordine di frequenza decrescente.
- ***Percentuale cumulativa***: permette di generare nella tabella di output una colonna con le percentuali cumulative e per includere nel grafico una riga per la percentuale cumulativa.
- ***Grafico in output***: crea l'istogramma nel foglio di destinazione.

Se vogliamo che il foglio di calcolo crei in automatico sia le classi che il grafico spuntiamo l'opzione *Grafico in output* e non definiamo l'*Intervallo della classe*. Nel foglio di destinazione viene creato l'istogramma ed una tabella come quella mostrata nella seguente figura:



	A	B
1	*Classe*	*Frequenza*
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12	Altro	

Nelle celle sotto la casella A1 *Classe* vengono create un certo numero di classi secondo un determinato criterio: la prima classe è un intervallo *chiuso a sinistra* ossia racchiude tutti i dati che sono minori o uguali al dato presente nella cella; l'ultima classe *Altro* raccoglie tutti i dati che non rientrano a far parte dell'ultima classe determinata. Fra queste due classi ve ne sono altre che sono costituite da intervalli *aperti a sinistra* e *chiusi a destra* del tipo $(a \div b]$ (dove *a* e *b* sono due numeri): se ad esempio supponiamo che nella cella A4 ci sia il numero 20 ed in A5 il numero 30, ciò significa che nella cella A5 l'intervallo è del tipo $(20 \div 30]$ e raccoglie tutti i dati maggiori di 20 (escluso) fino a 30 (incluso).

Invece sotto la cella B1 *Frequenza* vengono visualizzate le frequenze f_i.

Una volta create le classi si spunta nella finestra l'opzione *Intervallo della classe* e si selezionano nel foglio l'intervallo delle celle interessate.

La distribuzione degli errori casuali:

Una ***funzione di distribuzione*** è un modello matematico che collega i valori di una variabile alla probabilità o alla frequenza (a seconda che rispettivamente si stia considerando un caso teorico o sperimentale) che tali valori possano essere osservati.

Queste funzioni vengono utilizzate per studiare un campione di dati appartenente ad una popolazione ovvero all'insieme dei possibili casi che uno sperimentatore potrebbe osservare. Il fine è creare dei modelli per il campione: in tal modo si può generalizzare il comportamento di quest'ultimo a tutta la popolazione non direttamente osservabile dello sperimentatore.

La variabile di interesse è detta ***variabile casuale*** o ***aleatoria***, in base quale si possono distinguere due tipi di distribuzioni:

- ***distribuzioni discrete***: la variabile assume valori discreti (come ad es. la distribuzione dei risultati del lancio di un dado);
- ***distribuzioni continue***: la variabile assume valori continui.

Da un punto di vista strettamente formale, le distribuzioni di probabilità vengono espresse da una legge matematica detta:

- ***funzione di probabilità***, indicata con $p(x)$ per le distruzioni discrete;
- ***funzione di densità di probabilità***, indicata con $f(x)$ per le distribuzioni continue.

In una distribuzione di dati, le caratteristiche più importanti sono il ***valore aspettato*** (o ***valore vero***) μ che corrisponde al valore intorno al quale si dispone la distribuzione e la ***varianza*** σ^2 (o la sua radice, σ, la ***deviazione standard***) che quantifica la variabilità dei dati ossia quanto si discostano dal valore aspettato.

Di funzioni di distribuzione ne esistono diverse in funzione del tipo di popolazione che si sta analizzando.

Di maggior interesse dal punto di vista fisico è la distribuzione degli errori casuali: supponiamo di aver eseguito N misure di una grandezza fisica, ognuna delle quali differirà dalle altre a causa della presenza di errori casuali ovvero ad errori che variano in modo imprevedibile da una misura all'altra e che ne influenzano il risultato qualche volta per eccesso, altre per difetto.

Ogni dato sperimentale $x_1, x_2, x_3, \dots, x_N$ (in generale ogni singolo dato è etichettato con x_i con i che va da 1 a N) rappresenta una stima del valore vero μ della grandezza fisica che stiamo misurando mentre l'insieme N delle misure fatte è il campione di una popolazione formata da tutte le misure possibili (che quindi teoricamente sarebbero infinite).

Utilizzando il campione sperimentale a nostra disposizione possiamo costruire un istogramma con le frequenze che è una stima della distribuzione della popolazione ovvero di come tutti i nostri infiniti dati si dispongono all'interno del grafico.
Si può dimostrare che la miglior stima che possiamo fare del valore vero della grandezza misurata a partire da *N* misure soggette ad errori casuali è la ***media aritmetica*** $\bar{x}$:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

A questa dobbiamo associare un'indeterminazione che nasce dalle fluttuazioni casuali dei risultati che può essere stimata con la ***deviazione standard della media*** $s_{\bar{x}}$:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N(N-1)}}$$

dove σ è la radice della varianza σ^2 delle singole misure x_i:

$$\sigma^2 = \frac{\sum_i (x_i - \bar{x})^2}{N-1}$$

Ricapitolando, le singole misure e la media aritmetica hanno lo stesso valore aspettato μ ovvero sono entrambi una stima del valore vero μ: ciò che li differenzia è l'accuratezza con la quale approssimano il valore vero.
Infatti, se σ^2 è la varianza delle singole misure x_i, la varianza del valore medio ossia quanto la media aritmetica differisce dal valore vero μ è pari a σ^2/N, cioè è *N* volte più stretta: ha un'indeterminazione minore. Perciò $\bar{x}$ è il miglior stimante del valore vero di una distribuzione di misure soggette ad errori casuali.