Lezione 9
Inferenza statistica

PROF. ROBERTO COSTA

SCIENZE DELL'EDUCAZIONE - STATISTICA SOCIALE (305SF)

Dove eravamo rimasti

Nella scorsa lezione abbiamo concluso l'argomento delle relazioni tra due variabili.

Abbiamo visto i coefficienti χ^2 di Pearson, Φ^2 e V di Cramer.

Abbiamo visto come possiamo leggere la relazione tra microdati di due variabili quantitative, legate da un rapporto di causa-effetto, attraverso l'analisi della regressione.

Abbiamo poi visto come possiamo valutare la bontà del modello ottenuto attraverso la costruzione di una retta col metodo dei minimi quadrati, attraverso l'indice di determinazione R^2.

Ci sono dei dubbi?

Dalla descrizione all’inferenza

Fino ad oggi abbiamo visto strumenti e tecniche tipiche della statistica descrittiva, ovvero con l’obiettivo di interpretare in modo più immediato dei dati, attraverso la loro sintesi (distribuzione di frequenze, rappresentazione grafica, valori caratteristici centrali e di disuguaglianza, indicatori, ecc.).

Ci troveremo in situazioni in cui la finalità dell’indagine statistica non è (solo) la sintesi dei dati ottenuti dalle unità statistiche coinvolte, ma a conoscere le caratteristiche dell’intera popolazione da cui quelle unità sono state estratte.

Si definisce statistica inferenziale l’insieme delle tecniche che consentono di estendere i risultati ottenuti su una porzione della popolazione (chiamata campione) a tutta la popolazione.

I risultati ottenuti avranno inevitabilmente un grado di incertezza.

Cominciamo con un gioco

Un contenitore con dei sassolini colorati di 4 colori. Se vi chiedessi quanto pesano i sassolini azzurri cosa mi rispondereste?

Senza alcuna ulteriore indicazione è molto difficile fornire una risposta.

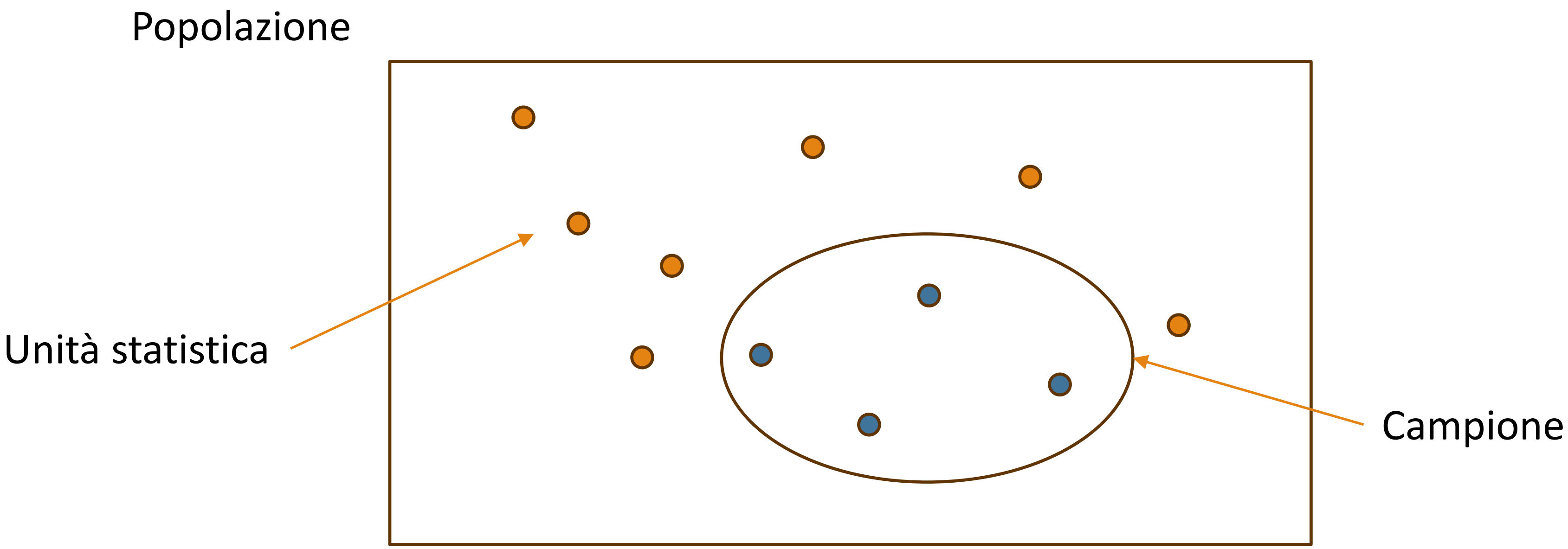
Potremmo cercare di soppesare il contenitore per avere una prima idea e poi fare delle valutazioni.

Proviamo a procedere con un esperimento, facendo un’estrazione casuale di 12 sassolini e riportiamo i vari risultati su un foglio.

Dalla descrizione all'inferenza

Quantificare il grado di incertezza significa associare ai risultati campionari un livello di probabilità, ovvero calcolare la probabilità che i risultati ottenuti siano validi per la popolazione di riferimento.

E' pertanto necessario conoscere qualche elemento essenziale della teoria delle probabilità.



I concetti di base della probabilità

Quotidianamente affrontiamo delle situazioni che determinano incertezza nelle decisioni da prendere.

Domani pioverà? Arriverò in tempo in stazione per prendere il treno?

Questi sono esempi dei ragionamenti probabilistici, più o meno consapevoli, che facciamo utilizzando la teoria delle probabilità.

I concetti di base della probabilità

I concetti che dobbiamo tenere a mente sono:

- **La prova** (che indicheremo con i), detto anche esperimento aleatorio, ovvero un esperimento soggetto a incertezza.
- **L'evento** (che indicheremo con E), ovvero uno dei possibili esiti della prova.
- **La probabilità** (che indicheremo con P), un valore numerico (compreso tra 0 e 1) che misura il grado di incertezza sul verificarsi di un evento.

Uniamo questi tre concetti in un'unica frase: «in una data prova i, l'evento E si verifica con una certa probabilità P(E)»

Un esempio

Partiamo dalla frase di prima: «in una data prova i, l’evento E si verifica con una certa probabilità P(E)»

Se lancio un dado qual è la probabilità che esca la faccia contrassegnata col numero 1?

La prova i è il lancio del dado

L’evento E è uscita della faccia contrassegnata dal numero 1

La probabilità P(E) è 1/6

Le definizioni di probabilità

Definizione classica di probabilità

La definizione classica di probabilità è attribuibile a Pier-Simon Laplace (1749-1827) a seguito di uno studio sui giochi d’azzardo.

La probabilità è data dal rapporto tra il numero di casi favorevoli all’evento e il numero di casi possibili, a condizione che questi siano tutti egualmente possibili.

$$P(E) = \frac{Numero\ di\ casi\ favorevoli}{Numero\ di\ casi\ possibili}$$

La probabilità, secondo la definizione classica, viene stabilita prima del verificarsi dell’evento.

Nel nostro esempio del lancio del dado la probabilità è pari a 1 evento favorevole su 6 eventi possibili.

Critiche alla definizione classica di probabilità

Questa definizione può essere utilizzato solo quando i risultati di una prova sono:

- noti,
- finiti,
- equiprobabili.

Il limite della definizione classica di probabilità è che si basa sul concetto di equiprobabilità.

Questo implica anche che si sia a conoscenza del concetto di probabilità, che invece si intende definire.

Le definizioni di probabilità

Definizione frequentista

Dalle critiche all'approccio classico di probabilità nell'800 si sviluppa l'impostazione frequentista (Richard von Mises 1883-1953).

La definizione frequentista è la seguente:

La probabilità è il limite a cui tende la frequenza relativa dell'evento al crescere del numero degli esperimenti.

$$P(E) = \lim_{n \to \infty} \frac{n_\varepsilon}{n}$$

Dove n_ε indica il numero di volte che, su n prove, si è verificato l'evento E.

La probabilità frequentista si applica laddove si disponga di dati statistici.

Critiche alla definizione frequentista di probabilità

Viene individuata una debolezza nella definizione frequentista di probabilità con riferimento alla sua applicabilità pratica.

L'impostazione frequentista si può utilizzare se:

- Le prove che originano gli eventi siano illimitatamente ripetibili
- Le prove siano svolte sempre nelle medesime condizioni

Le definizioni di probabilità

Nella concezione classica la probabilità viene stabilita a priori (prima di osservare i dati).

Nella concezione frequentista la probabilità si ottiene a posteriori, dopo aver esaminato i dati.

L'approccio classico e quello frequentista rientrano nel campo della **probabilità oggettiva**, ovvero che si basa su informazioni oggettive.

Si parla di **impostazione soggettiva** quando si assegna un valore di probabilità in base al nostro grado di fiducia, basandoci sulle conoscenze che abbiamo su quel determinato fenomeno.

La probabilità soggettiva

La probabilità P(E) di un evento E è la misura del grado di fiducia che un individuo coerente attribuisce al verificarsi dell'evento E.

Bruno De Finetti (1906-1985) è uno dei maggiori sostenitori di questa impostazione, che ammette che le probabilità di uno stesso evento possano essere diverse, se fornite da soggetti diversi.

Per riassumere

Per le nostre finalità limitiamoci a ricordare che la probabilità di un evento E è un numero compreso tra 0 e 1, calcolato secondo la definizione classica.

Ci sono due casi estremi:

Evento impossibile: è un evento che non potrà mai verificarsi. La probabilità di un evento impossibile è pari a 0.

Evento certo: è un evento che si verifica sempre. La probabilità di un evento certo è pari a 1.

Evento impossibile: che al prossimo appello uno studente prenda 35 all’esame di statistica sociale

Evento certo: che esca il valore compreso tra 1 e 6 lanciando un comune dado.

L’algebra di Boole

Per formalizzare in modo rigoroso gli eventi utilizziamo l’algebra di Boole.

Il singolo esito di una prova prende il nome di **evento elementare** e si indica con la lettera greca omega minuscolo (ω_i)

L’evento non elementare (E) è un evento che può essere scomposto in due o più eventi elementari.

Ad esempio nella prova lancio del dato ci sono 6 eventi elementari (che esca 1, che esca 2, … , ovvero ω_1= 1, ω_2= 2, …)

Un esempio di evento non elementare è che nel lancio del dado esca un numero minore di 4, ovvero che esca 1, o 2, o 3 (ovvero ω_1= 1, o ω_2= 2, o ω_3= 3).

L’insieme di tutto gli eventi elementari possibili si chiama **spazio campionario**, che si indica con la lettera greca omega maiuscola (Ω).

Le operazioni fondamentali dell’algebra di Boole

Indichiamo gli eventi con le lettere maiuscole (A, B, C, …).

Vediamo le tre operazioni fondamentali dell’algebra di Boole:

Unione tra due eventi A e B ($A \cup B$)

Si verifica quando almeno uno dei due eventi, o entrambi, si verificano.

Intersezione tra due eventi A e B ($A \cap B$)

Si verifica quando entrambi gli eventi si verificano.

Negazione di un evento A ($\bar{A}$)

Si verifica quando non si verifica l’evento A.

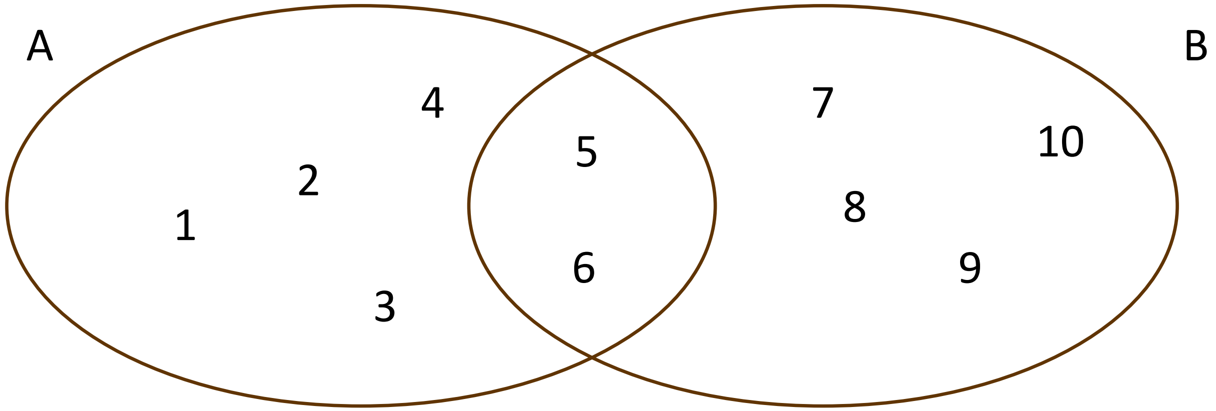
Facciamo un esempio

Supponiamo di avere un’urna con dieci palline, numerate da 1 a 10.

Consideriamo due eventi non elementari:

A – esce una pallina numerata da 1 a 6

B – esce una pallina numerata da 5 a 10

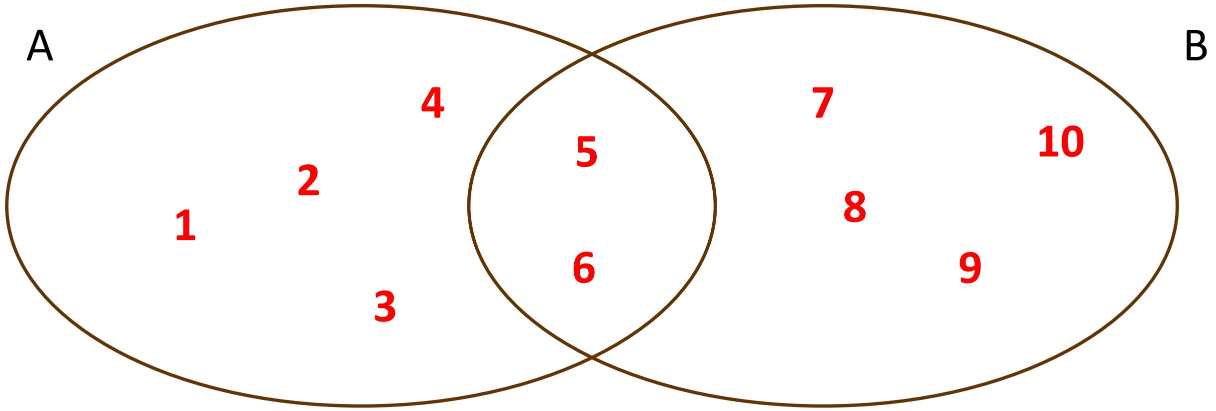


Facciamo un esempio

A – esce una pallina numerata da 1 a 6 - B – esce una pallina numerata da 5 a 10

A ∪ B = si verifica l'evento A oppure l'evento B

In questo caso l'evento A ∪ B corrisponde all'intero spazio campionario Ω

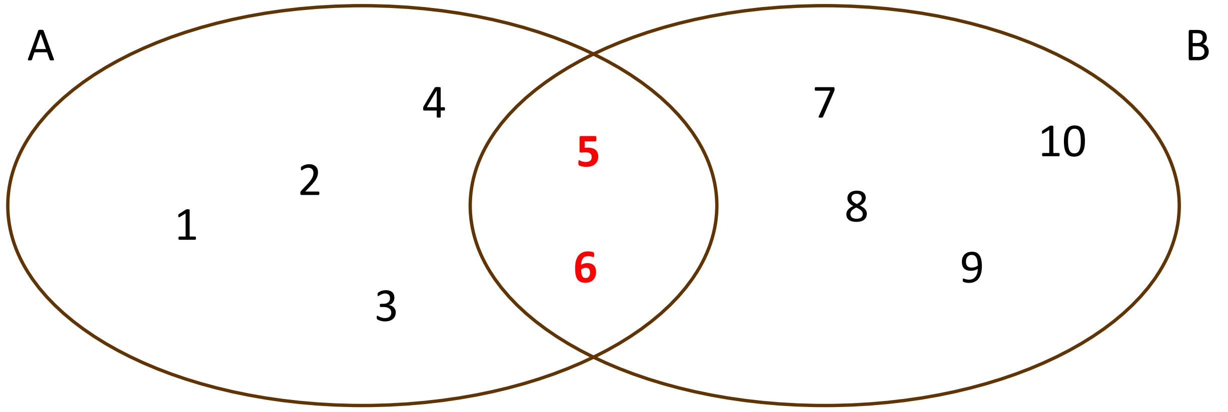


Facciamo un esempio

A – esce una pallina numerata da 1 a 6 - B – esce una pallina numerata da 5 a 10

A ∩ B = si verifica l’evento A e l’evento B

In questo caso l’evento A ∩ B si verifica se si verificano contemporaneamente gli eventi A e B, ovvero se esce 5 o 6.



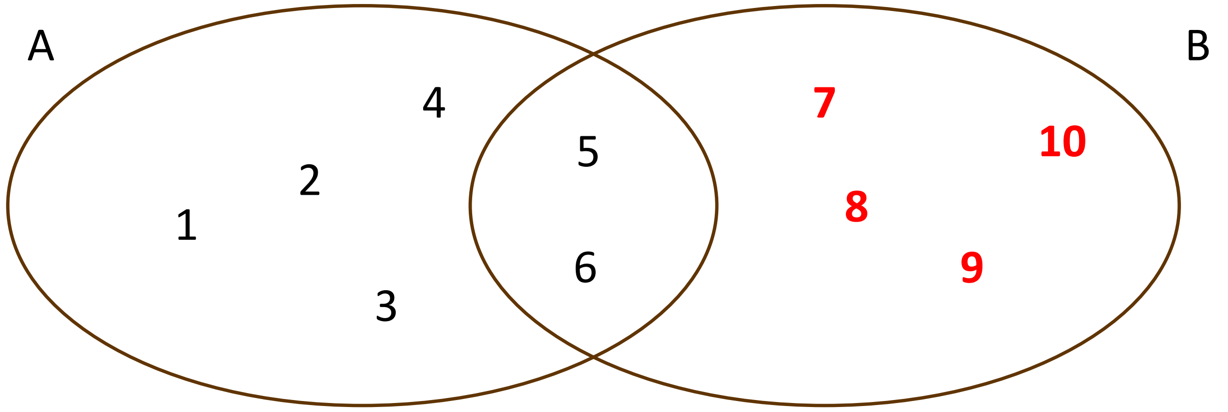
Facciamo un esempio

A – esce una pallina numerata da 1 a 6 - B – esce una pallina numerata da 5 a 10

$\bar{A}$ = non si verifica l’evento A

In questo caso l’evento $\bar{A}$

si verifica se esce una palline il cui numero non è contenuto nell’ovale A (quindi 7, 8, 9 o 10).



I tre postulati sulla probabilità

Postulato 1: $P(A) \geq 0$

La probabilità di un evento A è sempre maggiore o uguale a 0. E' pari a a 0 se A è un evento impossibile.

Postulato 2: $P(\Omega) = 1$

La probabilità che si verifichi uno qualsiasi degli eventi dello spazio campionario Ω è pari a 1. Si tratta dell'evento certo.

Postulato 3: $P(A) + P(B) = P(A \cup B)$ se $A \cap B = 0$

Qualora due eventi siano incompatibili, ovvero non possano verificarsi contemporaneamente, la probabilità dell'evento unione è paria alla somma delle probabilità dei singoli eventi.

Dati due eventi A e B compatibili $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (probabilità totali per eventi compatibili)

Probabilità condizionata

Spesso ci troviamo di fronte a eventi che possono dipendere da altri eventi che si verificano precedentemente. Si parla di probabilità condizionata e si scrive: $P(A|B)$

Si definisce **probabilità condizionata** dell'evento A rispetto all'evento B la probabilità che si verifichi l'evento A dopo che si è verificato l'evento B.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{e } P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Se i due eventi sono indipendenti

$P(A|B) = P(A)$ e $P(B|A) = P(B)$

Da qui si ottiene che la probabilità dell'intersezione di due eventi indipendenti è data dal prodotto delle rispettive probabilità: $P(A \cap B) = P(A)\,P(B)$

La variabile casuale

Si definisce **variabile casuale** (o variabile aleatoria o variabile stocastica) una variabile che può assumere valori diversi in funzione di un fenomeno aleatorio.

Ad esempio, il risultato del lancio di un dado a sei facce può essere visto come una variabile casuale che può assumere uno dei sei possibili valori (1, 2, 3, 4, 5, 6).

Ogni valore ha probabilità 1/6 di presentarsi.

La variabile casuale

Distinguiamo:

Variabile casuale discreta: ovvero una corrispondenza tra gli eventi di uno spazio campionario Ω e un insieme discreto di numeri reali $\Re$.

Si definisce funzione di probabilità di una variabile causale discreta X, la funzione che associa ad ogni possibile valore di x_i una probabilità $P(x_i)$.

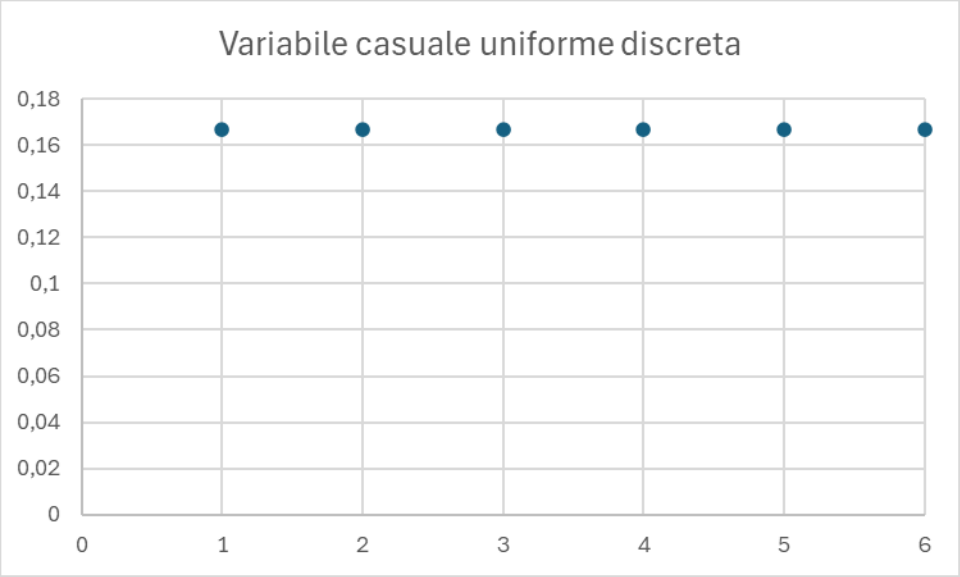
Variabile casuale continua: ovvero una corrispondenza tra gli eventi di uno spazio campionario Ω e un intervallo di numeri reali $\Re$.

Si definisce funzione di densità di una variabile causale continua X la funzione matematica f(x) tale per cui l’area sottostante alla funzione è uguale alla probabilità che X assuma un valore in quell’intervallo.

Un esempio

La **distribuzione discreta uniforme** è una distribuzione di probabilità discreta che è uniforme su un insieme, ovvero che attribuisce la stessa probabilità ad ogni elemento dell'insieme discreto S su cui è definita.

Ad esempio la distribuzione discreta uniforme della prova «lancio di un dado a sei facce» ha sei elementi, ognuno dei quali ha una probabilità pari a 1/6.



Distribuzione discreta uniforme

La **distribuzione discreta** non presenta un valore modale e la media e la varianza sono date da:

E(X) = $\frac{m+1}{2}$

Var(X) = $\frac{m^2-1}{12}$

Dove m è il numero di valori interi che la variabile casuale può assumere.

Distribuzioni discrete

Altre **distribuzioni discrete** sono

Distribuzione di Bernoulli

Deriva da una prova dove interessa esclusivamente stabilire se l'evento E si sia verificato o meno (prove con esito dicotomico: testa o croce, superamento o meno di un esame, presenza o assenza di una certa caratteristica, ecc.).

Distribuzione binomiale

Si. ottiene come somma di n variabili casuali di Bernoulli, indipendenti e identicamente distribuite. Ad esempio lancio ripetuto di una moneta.

Distribuzioni continue

Distribuzione normale

È la più nota delle variabili casuali continue ed è la più utilizzata. È stata introdotta dal matematico Abraham de Moivre e fu riscoperta dal fisico Gauss nel 1796 e successivamente da Quetelet nel 1846.

Si basa sull'osservazione che in natura le misurazioni successive di valori caratteristici sono affette da errori sia positivi che negativi in egual misura e che gli errori più piccoli sono più frequenti di quelli più grandi.

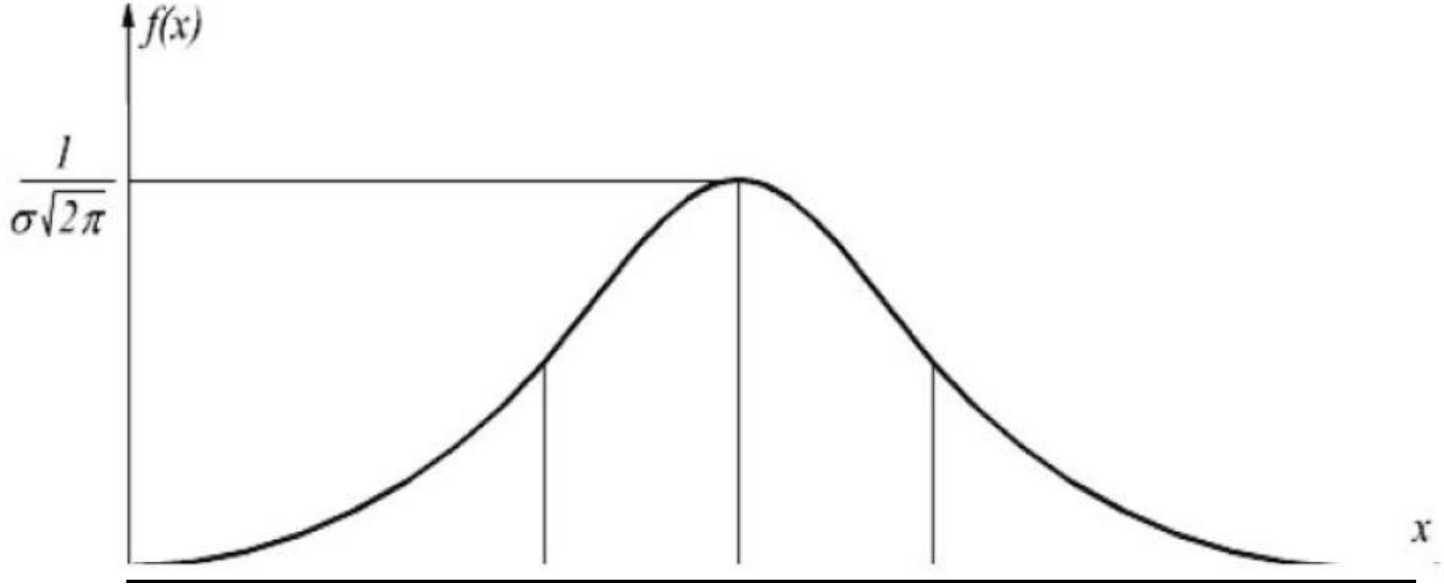
Il termine «normale» si deve al fatto che molti fenomeni nella realtà hanno una distribuzione empirica che si avvicina moltissimo alla curva a forma di campana che quindi rappresenta la «norma».

Caratteristiche della distribuzione normale

- La sua funzione di densità ha una forma a campana
- È simmetrica rispetto al valore $x = \mu$
- Il valore di $x = \mu$ oltre che alla media aritmetica coincide anche con moda e mediana
- È asintotica all'asse delle x da entrambi i lati
- È crescente per $x < \mu$ e decrescente per $x > \mu$
- Possiede due punti di flesso in corrispondenza di $x = \mu + \sigma$ e $x = \mu - \sigma$
- L'area sotto la curva è pari a 1

Caratteristiche della distribuzione normale

La funzione di densità della variabile casuale normale



μ - σ μ μ - σ

Caratteristiche della distribuzione normale

Al variare del valore medio la curva normale si sposta lungo l'asse x, mentre al crescere della varianza si modifica la dispersione e conseguentemente l'appiattimento della curva.