

**Probabilità condizionata:**  $p(A/B)$  che avvenga A, una volta accaduto B

**Eventi indipendenti:** un evento non influenza l'altro

**Eventi disgiunti:** il verificarsi di un evento esclude l'altro

**Evento prodotto:** Evento in cui si verifica sia A che B ;  $p(A\&B) = p(A) \times p(B/A)$

**Evento somma:** si verifica A o B o entrambi  $p(A) + p(B) - p(A\&B)$

**Evento complementare:**  $p(\tilde{A}) = 1 - p(A)$

**Parametri descrittivi:** non dimostrano nulla servono solo a descrivere il campione

**Media:** parametro descrittivo del campione

**Varianza:** variazione media intorno alla media, somma degli scarti al quadrato fratto "n-1" gradi di libertà. Se si lavora su dati teorici si divide per "n" così come per campioni di grande numerosità dove "n" o "n-1" si equivalgono

**Valore atteso:** prodotto ogni valore per la sua probabilità diviso "n" se equiprobabili → la media viene definita molte volte come valore atteso

**Legge dei grandi numeri:** all'aumentare del numero delle prove la frequenza **tende** al valore della sua probabilità teorica, **tende** a 1 (evento certo)

**Percentile:** 40esimo percentile in altezza = più alto del 40% della popolazione totale

**Mediana:** 50esimo percentile

**Moda:** punta di massima frequenza

**Distribuzione di probabilità:** rappresentatività di tutti i valori della variabile → serve per sapere se supponendo vera  $H_0$  (ipotesi nulla) si ha o no la possibilità di avere un determinato valore → in definitiva se il valore che devo esaminare può essere considerato valore causale oppure dovuto ad una causa specifica incidente.

. Lo strumento di base per capire quali sono le differenze che stanno dentro l'oscillazione casuale e quelle che stanno fuori, che sono quindi ritenute dovute a uno specifico fattore, questo strumento è la distribuzione di probabilità.

**funzione di distribuzione** è la funzione che rappresenta per ogni x la probabilità di ottenere un valore uguale o minore di X. Quindi invece di stabilire la probabilità dei diversi valori io vado a stabilire qual è la probabilità che ho di ottenere un valore minore o uguale a un certo valore X. Se la variabile è discreta ho finiti valori uguali o inferiori a X. Ciò non è possibile per una variabile continua → al denominatore ho infinito e il risultato è 0

**La funzione di distribuzione permette di calcolare un valore finito in una variabile continua.**

In una variabile continua la probabilità è sempre uguale a zero, nella funzione di distribuzione invece no, si hanno infiniti casi al numeratore e infiniti al denominatore e questo dà un numero finito.

Se la variabile è continua la funzione di distribuzione non è più definita dalla somma delle probabilità, ma come l'integrale di una certa funzione. L'integrale è una somma di infiniti termini, ognuno di valore infinitesimo, è la conversione nel caso continuo della sommatoria del caso discreto (sommatoria è somma di elementi definiti). Cos'è quella "f(x)" quando calcolo l'integrale? È la **densità di probabilità**, che rappresenta la probabilità che il valore x cada all'interno di un certo intervallo diviso la larghezza di questo intervallo.

La **distribuzione binomiale** è la distribuzione che compete a una variabile che può essere o SI o NO o successo o insuccesso o testa o croce, compete a una variabile dicotomica (cioè che ha 2 valori).

$$p = \frac{1}{2^n} \binom{n}{s} = \frac{n!}{s!(n-s)!}$$

Questa formula vale solo se la probabilità di successo p e di insuccesso q sono uguali ed ognuna è uguale ad 1/2, quindi la loro somma sia 1

. Nel caso che ciò non sia vero la formula cambia leggermente.

$$p \neq q$$

$$p = p^s q^i \binom{n}{s}$$

La distribuzione di probabilità serve per misurare l'oscillazione casuale → se ottengo risultati al di là di tali parametri ( normalmente fissati al 5% di errore ) non considero il dato frutto del caso. Nella distribuzione binomiale sommo le colonne esterne simmetriche che mi devono dare un risultato inferiore al 5%

La distribuzione binomiale funziona solo se io considero il verificarsi di 1 evento dicotomico i cui risultati possono essere solo il verificarsi o il non verificarsi di quell'evento. La realtà è che nella ricerca non è sempre così, ci sono volte in cui i valori possono essere tanti, ad esempio se prendo l'altezza. La distribuzione binomiale è la discretizzazione, cioè la conversione in valori discreti, di un'altra distribuzione relativa ad una variabile continua (che è fondamentale in statistica), ed è la distribuzione normale o distribuzione di Gauss. **La distribuzione gaussiana o normale** è la distribuzione di una variabile continua, è l'estensione della binomiale al caso continuo.

Numero pari di prove ho un picco, dispari due

La distribuzione normale è anche definita come curva degli errori.

**CURVA DEGLI ERRORI**

Condizioni:

- un errore è la somma di molte componenti di uguale ampiezza
- le diverse componenti sono fra loro indipendenti
- ciascuna componente è positiva o negativa con uguale probabilità

allora l'ampiezza dell'errore ha una distribuzione normale

L'entropia in statistica, poiché misura il disordine, è una misura del caso. In statistica più alta è l'entropia, più alto è il disordine, più alta è l'incidenza del caso. La distribuzione gaussiana è la distribuzione di una variabile caratterizzata dal fatto di avere la massima entropia possibile. La distribuzione normale è la distribuzione del caso per eccellenza

Cosa importante: una variabile a distribuzione binomiale ha una media e una varianza ben definite, la media è data da  $n \cdot p$  dove  $n$  è il numero di prove e  $p$  è la probabilità ad es. di successo. La varianza invece è data da  $n \cdot p \cdot q$ , dove  $q$  è la probabilità dell'insuccesso. Questi due valori sono relativi a una distribuzione teorica, siamo in casi teorici. La media sia che riguardi valori teorici che sperimentali si calcola nello stesso modo, la varianza no, se la calcolo per valori teorici va divisa per  $n$  non per  $(n-1)$ .

**Una generica variabile normale con media  $\mu$  e varianza  $\sigma^2$  è indicata con  $N(\mu, \sigma^2)$  e la sua densità**

Per standardizzare la curva gaussiana ottenuta dal mio esperimento e compararla con altre devo trovare il valore "z" normalizzato → si paragona con la curva g. standard  $N(0,1)$

**Zeta =  $X - \mu / \sigma$**

**Il punto cruciale è che per trovare il punto zeta devo sapere la media e la varianza teorica**

Vedi QI media 100 varianza 15

**La distribuzione X quadro** è la distribuzione che compete ad una somma di variabili Z elevate al quadrato → la somma di variabili gaussiane o la loro divisione per una costante produce una variabile gaussiana, **non** il loro prodotto

**La varianza di variabili gaussiane hanno una distribuzione X quadro, compete alla somma di variabili gaussiane elevate al quadrato**

Se nei miei esperimenti non posso trovare il punto Z perché non ho una media e una varianza teorica di confronto uso il **T-Test** usando la media e la varianza che trovo empiricamente

Il t-test ha una forma simile alla curva gaussiana solo più ampia.

È più ampia perché nella distribuzione gaussiana ho solo una variabile, la X, nella T ho tre sorgenti di variabili: la X, la media, la varianza al denominatore

- la distribuzione **t** è la distribuzione che compete direttamente a un rapporto fra medie e deviazioni standard (radice di varianza) se io non conosco il valore teorico della varianza.
- la distribuzione **t** ha una somiglianza molto forte con una distribuzione **z**, con la differenza che ha dei limiti di falsificazione più grandi.
- nel momento in cui io aumento i gradi di libertà, cioè aumento la dimensione del campione, la distribuzione **t** tenderà sempre di più ad una gaussiana e i suoi limiti di falsificazione si avvicinano sempre di più a quelli della distribuzione **z**.

Riepilogo dei simboli:

$\sigma^2$  è la varianza teorica

$s^2$  è la varianza sperimentale

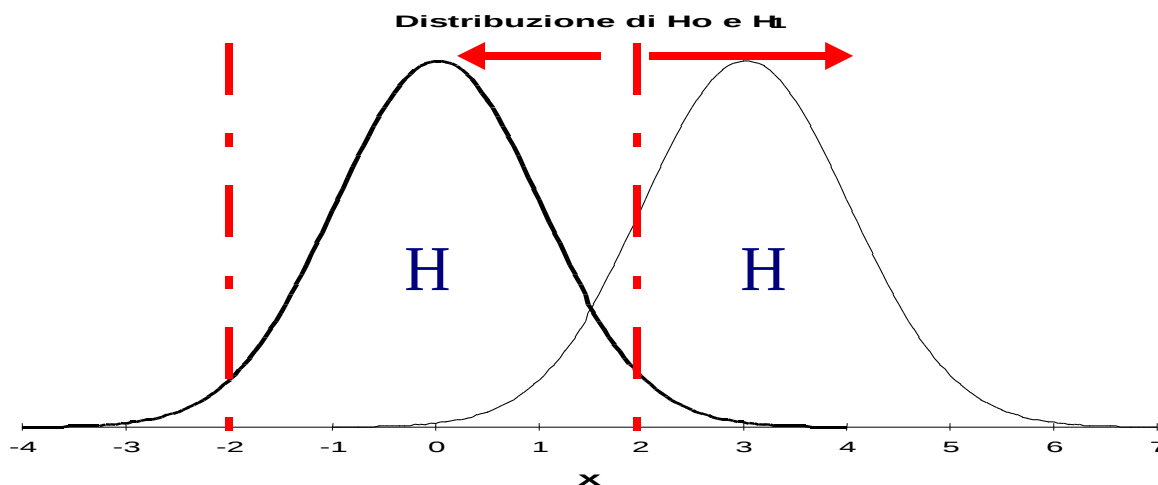
$\sigma$  è la deviazione standard teorica

$s$  è la deviazione standard sperimentale

La varianza teorica fra individui è  $\sigma^2$  mentre la varianza fra medie è  $\sigma^2/n$ , perché la varianza o meglio la deviazione standard è una misura dello scarto medio tra il soggetto e la media del gruppo

**La distribuzione f di Fisher** è il rapporto fra due o più X quadro e ha una distribuzione asimmetrica come la variabile X quadro  
Questa distribuzione viene usata per comparare la varianza di due variabili → per comparare delle varianze devo usare la **F**





	<i><b><math>H_0</math> vera</b></i> <i><b><math>H_1</math> falsa</b></i>	<i><b><math>H_0</math> falsa</b></i> <i><b><math>H_1</math> vera</b></i>
<i><b>Respingo <math>H_0</math></b></i>	<b>errore I tipo</b> $\alpha$	<b>corretto</b>
<i><b>Non respingo <math>H_0</math></b></i>	<b>corretto</b>	<b>errore II tipo</b> $\beta$

Sull'errore di 1° tipo abbiamo visto che è la probabilità che essendo vera  $H_0$  io la falsifichi.

L'errore di 2° tipo o  $\beta$  è la probabilità di non falsificare  $H_0$  pur essendo questa falsa. Quindi è il fallimento di dimostrazione della mia idea di base che è vera.

Si utilizza il parametro  $1 - \beta$  che è la **potenza del test**. Cioè è la probabilità di falsificare  $H_0$  quando questa è falsa.

La potenza del test dipende da questi parametri:

- ▽ da  $H_0$  e da  $H_1$  (dalla loro distanza)
- ▽ dalla numerosità del campione
- ▽ dalla minima differenza apprezzabile
- ▽ dalla varianza casuale

