

**Le prime tre pagine sono di introduzione  
VALUTAZIONE DEL RISCHIO DI MALATTIA**

Una tecnica comunemente utilizzata classifica il rischio mediante una matrice che correla l'entità del danno con la sua potenziale occorrenza.

A ciascun rischio identificato (inteso come fonte potenziale di pericolo) sono associati due valori numerici:

- **P** - probabilità evento (o frequenza di accadimento).
- **D** - entità del danno (o intensità della conseguenza)

<b>Scala della probabilità</b>	Molto probabile	4	5	6	7
	Probabile	3	4	5	6
	Poco probabile	2	3	4	5
	Improbabile	1	2	3	4
		Bassa	Media	Alta	Altissima
<b>Scala della Gravità</b>					

Pertanto viene individuato per ciascun fattore un: **INDICE DI RISCHIO  $R = P \times D$** .

A ciascun INDICE DI RISCHIO deve corrispondere una specifica attività controllo, verifica e/o interventi:

- **R=1** indice di rischio basso
- **R=2-3** indice di rischio medio
- **R= 4-8** indice di rischio alto
- **R ≥ 9** indice di rischio molto alto

**ESEMPIO DI MATRICE DI RISCHIO DI UN TERREMOTO**

Valutazione del rischio. Matrici di rischio



**Valutazione del rischio. Matrici di rischio**

PROBABILITA'				
ELEVATA	7	8	15	16
MEDIO-ALTA	5	6 PROTEZIONE	13	14
MEDIO-BASSA	3	4	11 PREVENZIONE	12
BASSISSIMA	1	2	9	10
	TRASCURABILE	MODESTA	NOTEVOLE	INGENTE
	MAGNITUDO			

Quando nella matrice è presente una probabilità alta si agisce sulla **PROTEZIONE**, in modo tale che se l'evento dovesse accadere procurerebbe meno danni.

Viceversa, se la probabilità è bassa si agisce sulla **PREVENZIONE**, facendo sì che l'evento avverso non accada.

### APPROCCI MEDICI

In ambito medico, il medico può ricorrere a tre diversi tipi di approcci:

- **APPROCCIO CAUSALE**, riconoscendo una patologia dalle sue cause (relazione causa-effetto)  
ESEMPIO. Dati in possesso del medico: tetania, calcio basso e recente tiroidectomia parziale.  
Diagnosi: ipoparatiroidismo. (diagnosi certa)
- **APPROCCIO DEL TIPO RICONOSCIMENTO DEL MODELLO**, basandosi sul riconoscimento dei sintomi  
ESEMPIO. Dati in possesso del medico: febbre, un ginocchio tumefatto e dolente, sfregamento pericardico, alto tasso di antistreptolisine.  
Diagnosi: febbre reumatica.
- **APPROCCIO PROBABILISTICO O BAYESIANO**, alta probabilità che il paziente abbia una determinata patologia in base ai suoi sintomi  
ESEMPIO. Dati in possesso del medico: paziente con dolori colici e clisma opaco che ha evidenziato un restringimento del colon.  
Diagnosi: alta probabilità (non una certezza assoluta) di cancro al colon

Sono detti **ESPERIMENTI CASUALI**, quegli esperimenti che, nonostante vengano eseguiti nelle medesime condizioni, portano a risultati diversi. Infatti, c'è sempre una variabilità dovuta al caso, detta **CASUALITÀ**.

Ad esempio, il risultato ottenuto lanciando un dado o una moneta non è sempre lo stesso.

Un evento, si dice casuale se non viene provocato da nessuna causa bensì è dettato dal caso.

Vi sono tre tipi di approccio alla probabilità:

- **APPROCCIO CLASSICO**, si verifica quando ho un insieme di un numero  $n$  di eventi, detto SPAZIO DEI CAMPIONI, che contiene un sottoinsieme di casi che sono favorevoli all'evento che si verifica. Se un evento si può verificare in  $h$  modi diversi su  $n$  possibili, essendo questi tutti ugualmente possibili, allora la probabilità di tale evento è  $h/n$ .
- **APPROCCIO FREQUENTISTICO**, si attua (quando il verificarsi dei casi non è ugualmente possibile) per calcolare la probabilità dell'evento. Se, dopo avere ripetuto  $n$  volte un esperimento ( $n$  deve essere molto grande) un evento si è verificato  $h$  volte, allora la probabilità di questo evento è  $h/n$ . Questa probabilità è anche detta **probabilità empirica di un evento**.

- **APPROCCIO ASSIOMATICO**, dato che sia l'approccio classico che quello frequentistico hanno degli handicap; infatti, del primo non si può mai essere sicuri che le possibilità che si verifichi siano uguali, mentre del secondo poiché "n deve essere molto grande" e tende a infinito. Per queste difficoltà i matematici preferiscono un approccio assiomatico alla probabilità, che si basa su un modello che parte dalla definizione di tre assiomi (DA APPROFONDIRE) ed impiega la teoria degli insiemi.

La statistica è un insieme di tecniche con base matematica che servono per studiare un fenomeno collettivo (ovvero qualsiasi manifestazione misurabile) e si divide in due rami:

- **DESCRITTIVA** (distribuzioni statistiche, rappresentazioni grafiche, medie, variabilità, regressione e correlazione)
- **INFERENZIALE** (probabilità, campionamento, stima, test di ipotesi)

Tra i vari personaggi che hanno contribuito alla formazione della statistica vi sono:

**GALILEO GALILEI** matematico e astronomo italiano che introdusse la sperimentazione (l'esperimento è qualcosa che produce dati e l'analisi di questi ultimi è lo scopo della statistica)

**JOHANN GAUSS** matematico e astronomo tedesco che introdusse l'analisi dei dati e la curva di Gauss.

Le **RILEVAZIONI DEI DATI** possono essere:

- **SPERIMENTALI**, in cui i fenomeni vengono misurati in un laboratorio tramite esperimenti che produrranno dei dati. Sono facilmente controllabili
- **OSSERVAZIONALI**, effettuati tramite sondaggi, indagini di mercato. Non sono controllabili.

## TERMINOLOGIA

Nella STATISTICA un termine molto usato è la popolazione, ovvero un insieme di unità statistiche (oggetto della misurazione del fenomeno es. 1 persona) delle quali verranno studiati uno o più caratteri (es. colore del vestito) tramite dei campioni (unità scelte a caso dalla popolazione). I diversi caratteri presi in considerazione si presentano con modalità differenti (come si presenta un fenomeno al momento della misurazione es. azzurro).

Per rilevazione statistica si intende l'insieme delle operazioni mediante le quali si effettua la misurazione del carattere sulle unità statistiche.

Per estrazione si intende l'insieme delle procedure attuate per misurare un sottoinsieme del carattere di una popolazione.

Per prova si intende un esperimento

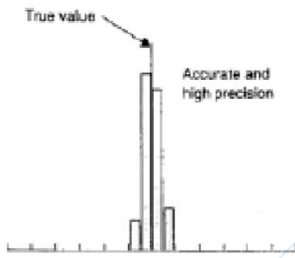
Per evento si intende il risultato di una prova. All'evento certo viene assegnato probabilità 1, all'evento impossibile probabilità 0

Per ampiezza si intende la differenza tra il valore più piccolo e il valore più grande di una classe di modalità

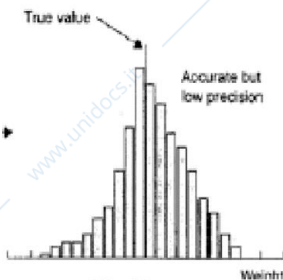
Per quantificare l'incertezza in una misurazione, si usano le parole:

- **PRECISIONE**: una misura è tanto più precisa quanto più i risultati ripetuti della stessa misura sono vicini (bassa variabilità)
- **ACCURATEZZA**: uno strumento è accurato se è adatto a misurare il fenomeno per cui è stato costruito

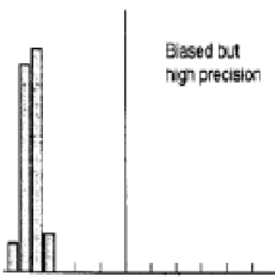
L'ISTOGRAMMA DI FREQUENZA presenta sull'asse delle ordinate il numero di volte in cui si effettua la misura (frequenza) e sull'asse delle ascisse presenta il valore (modalità)



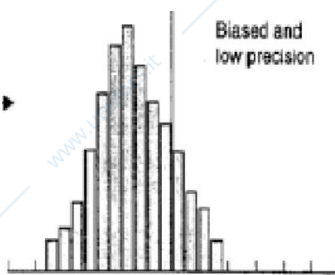
Strumento molto preciso (variabilità bassissima) e molto accurato perché la media di tutti i valori corrisponde al valore vero della grandezza che si sta misurando



Strumento accurato ma con bassa precisione perché le diverse misure sono lontane dal valore vero



Strumento distorto (non accurato) poiché le misure sono lontane dal valore vero ma molto preciso



Strumento non preciso e distorto

Il carattere può essere:

- **QUANTITATIVO**, quando la modalità è espressa da un numero. A loro volta possono essere:
  1. DISCRETI se possono assumere un numero finito o numerabile di valori;
  2. CONTINUI se possono assumere tutti i valori di un intervallo di numeri reali. (ad esempio l'età che può essere espressa in anni, in giorni o minuti).
- **QUALITATIVO**, quando la modalità è espressa in forma verbale. A loro volta possono essere:
  1. SCONNESSI se non esiste un criterio logico secondo il quale ordinare le modalità;
  2. ORDINABILI se esiste un criterio logico secondo il quale ordinare le modalità in livelli

Per la misurazione di un carattere distinguiamo diverse *scale*:

- **SCALE NOMINALI E ORDINALI**, riferite ai caratteri qualitativi
- **SCALE INTERVALLARI E PROPORZIONALI**, riferite ai caratteri quantitativi

Il primo dato che si ha quando si effettua un insieme di misurazioni è la frequenza che rappresenta il numero di volte in cui si presenta un certo fenomeno con la stessa modalità in un campione:

- ASSOLUTA, sommatoria del numero di volte che si presenta un certo dato;

$$\sum_{i=1}^k N_i = N$$

Sommatoria di N, per i che va da 1 a k

- RELATIVA, rapporto tra la frequenza assoluta  $N_i$  (numero di volte che si verifica un evento) e numero totale dei dati N;

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

- PERCENTUALE, frequenza relativa, espressa in percentuale

$$(f_i)\% = f_i * 100$$

- CUMULATA ASSOLUTA, associata ad una modalità o a una classe di modalità, è pari alla somma della sua frequenza assoluta e di quelle delle modalità che la precedono.

$$N_x = \sum_{i=1}^x N_i$$

- CUMULATA RELATIVA, associata ad una modalità o a una classe di modalità, è pari alla somma della sua frequenza relativa e di quelle delle modalità che lo precedono

$$f_x = \sum_{i=1}^x f_i$$

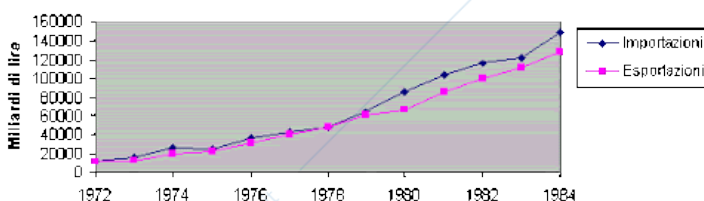
Quando si effettua una misura, c'è la possibilità di commettere un ERRORE che suddividiamo in:

- CAMPIONARIO, quando si osserva un campione, piuttosto che la popolazione
- EXTRACAMPIONARIO, quando l'errore non dipende dall'estrazione (non è un errore di tipo statistico) bensì da altre cause (come lo strumento di misura malfunzionante)
- SISTEMATICO, che deriva dal fatto che si sceglie un campione "sbagliato", ovvero, non rappresentativo
- CASUALE, quando dipende dal caso

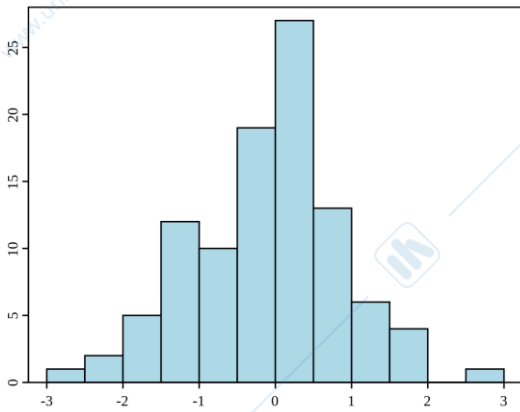
La statistica può essere suddivisa in:

- METEOROLOGICA, un insieme di tecniche matematiche:
  1. per descrivere i fenomeni (statistica descrittiva) tramite tabelle o rappresentazioni grafiche in cui vengono sintetizzati tramite indici (media aritmetica, moda e mediana)
  2. per descrivere l'inferenza dei fenomeni (statistica inferenziale)
- APPLICATA, studio di vari fenomeni come ad esempio l'economia, demografia (studio della popolazione)

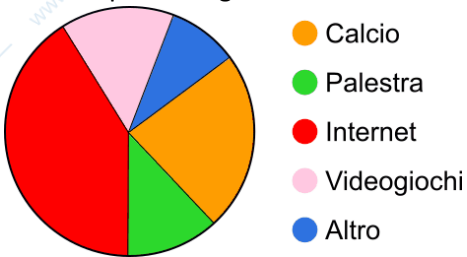
Per avere un'idea visuale sui dati si attua il GRAFICO che offrono il vantaggio di una descrizione del fenomeno in forma visiva, soprattutto nelle serie storiche (rappresentazione nel tempo)  
un esempio di diagramma cartesiano:



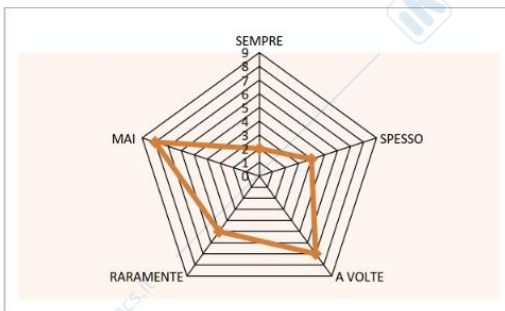
Un esempio di istogramma:



Un esempio di diagramma circolare:

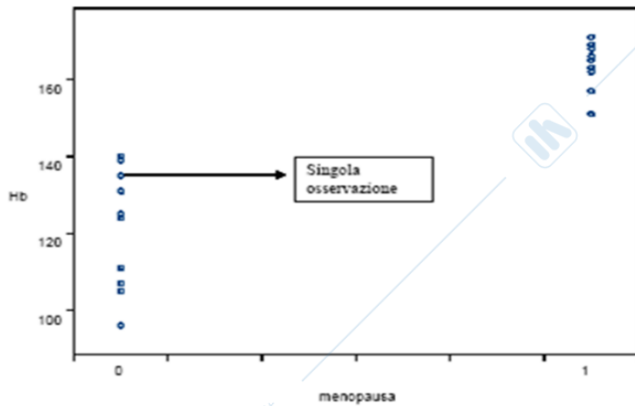


Un esempio di diagramma polare:

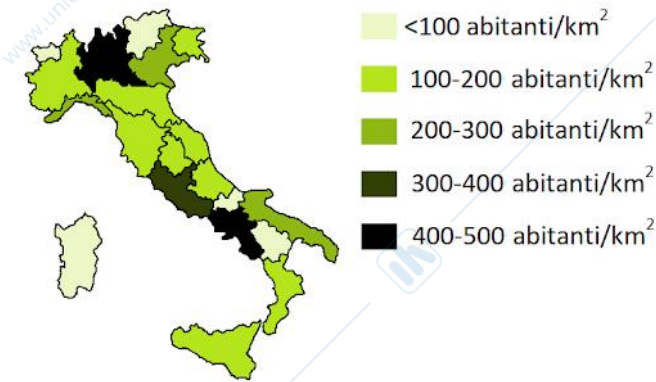


Un esempio di diagramma di dispersione:

ES: Distribuzione di una variabile continua (Concentrazione di Emoglobina), separatamente per i due sottogruppi definiti dalla variabile Menopausa (0= no, 1= si).



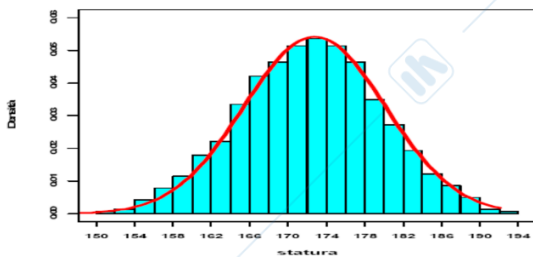
Un esempio di cartogramma:



### DISTRIBUZIONE STATISTICA

In ogni popolazione una modalità può presentarsi più volte. Il numero di volte in cui una modalità si presenta prende il nome di FREQUENZA ASSOLUTA  $N_i$ .

L'insieme delle modalità (asse delle ascisse) e delle frequenze (ordinate) costituisce una DISTRIBUZIONE STATISTICA. La linea rossa rappresenta la DISTRIBUZIONE DI PROBABILITÀ (o VARIABILE ALEATORIA) ed è una funzione matematica che esprime un fenomeno determinato dalla casualità.



Nel caso di un carattere qualitativo la distribuzione si chiama SERIE.

Ad esempio lo STATO CIVILE è un carattere qualitativo sconnesso

Stato civile	$n_i$
Celibe/nubile	34
Coniugato/a	51
Divorziato/a	12
Vedovo/a	18
<b>Totale</b>	<b>115</b>

Nel caso di un carattere quantitativo la distribuzione si chiama SERIAZIONE.

Ad esempio il NUMERO DI FIGLI è un carattere quantitativo discreto

Numero di figli	$n_i$
0	22
1	89
2	56
3	11
4	8
5	4
<b>Totale</b>	<b>190</b>

#### TABULAZIONE DEI DATI

I dati raccolti vengono riportati in apposite tabelle che si distinguono in:

- TABELLE SEMPLICI, relative ad un unico carattere

Carattere	Frequenza $n_i$
$x_1$	$n_1$
$x_2$	$n_2$
$x_3$	$n_3$
...	...
$x_i$	$n_i$
...	...
$x_k$	$n_k$
<b>Totale</b>	<b>N</b>

- TABELLE DOPPIE, relative a 2 caratteri "incrociati"

In questo caso i 2 caratteri sono l'ESITO DELLA CURA (modalità: POSITIVO o NEGATIVO) e il TIPO DI FARMACO (modalità: ASPIRINA o PLACEBO)

Esito cura	Tipo di farmaco		Totale
	Aspirina	Placebo	
Positivo	33	14	47
Negativo	17	41	58
Totale	50	55	105

- TABELLE MULTIPLE, relative a 3 o più caratteri

Un altro esempio di tabella doppia può essere rappresentato dai seguenti caratteri:

- NUMERO DI SIGARETTE FUMATE (carattere quantitativo discreto) contenente le modalità: nessuna/ 1-5/ 6-10/ oltre 10

- CLASSI DI ETÀ (carattere quantitativo continuo) contenente le modalità: 12-13/ 14-15/ 16-17

Numero di sigarette fumate	Classi di età		
	12 - 13	14 - 15	16 - 17
nessuna	90.0	81.2	76.4
1 - 5	5.1	10.8	12.8
6 - 10	4.9	6.7	8.8
oltre 10	0	1.3	2.0
Totale	100	100	100

Valori percentuali

Esempio di tabella della frequenza relativa:

Carattere	Frequenza $n_i$	$n_i/N$
$x_1$	$n_1$	$n_1/N$
$x_2$	$n_2$	$n_2/N$
$x_3$	$n_3$	$n_3/N$
...	...	...
$x_i$	$n_i$	$n_i/N$
...	...	...
$x_K$	$n_k$	$n_k/N$
Totale	N	1

Esempio di tabella della frequenza relativa percentuale:

Carattere	Frequenza $n_i$	$n_i/N$	$\%_i$
$x_1$	$n_1$	$n_1/N$	$n_1 \cdot 100/N$
$x_2$	$n_2$	$n_2/N$	$n_2 \cdot 100/N$
$x_3$	$n_3$	$n_3/N$	$n_3 \cdot 100/N$
...	...	...	...
$x_i$	$n_i$	$n_i/N$	$n_i \cdot 100/N$
...	...	...	...
$x_K$	$n_k$	$n_k/N$	$n_k \cdot 100/N$
Totale	N	1	100

### SINTESI NUMERICA

Un insieme di dati numerici è descritto e sintetizzato da tre informazioni principali:

- La TENDENZA CENTRALE è costituita dalle modalità più rappresentative e viene espressa tramite il calcolo delle medie (sintesi numerica dei dati)
- La VARIABILITÀ, tramite opportuni indici viene espressa la diversità che sussiste:
  - Fra i singoli valori
  - Fra i singoli valori e la tendenza centrale di riferimento (media)
- La FORMA DELLA SUA DISTRIBUZIONE

### LA STATISTICA (Trilussa)

Sai ched'è la statistica? È na' cosa  
che serve pe fà un conto in generale  
de la gente che nasce, che sta male,  
che more, che va in carcere e che  
spósa.

Ma pè me la statistica curiosa  
è dove c'entra la percentuale,  
pè via che, lì, la media è sempre eguale  
puro co' la persona bisognosa.

Me spiego: da li conti che se fanno  
seconno le statistiche d'adesso  
risurta che te tocca un pollo all'anno:  
e, se nun entra nelle spese tue,  
t'entra ne la statistica lo stesso  
perch'è c'è un antro che ne magna  
due.

**DA LEGGERE, LA CHIEDE PER  
INTRODURRE LE MEDIE**

### MEDIE

Le MEDIE sono indici di tendenza centrale e quindi sintetizzano in un solo valore tutti i dati raccolti

**DA COMPLETARE SECONDA LEZIONE**

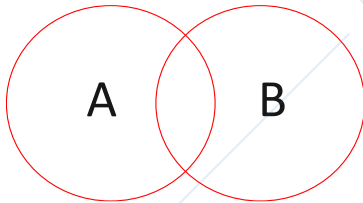
**PROBABILITÀ**

La **PROBABILITÀ** misura la propensione di un evento aleatorio a verificarsi.

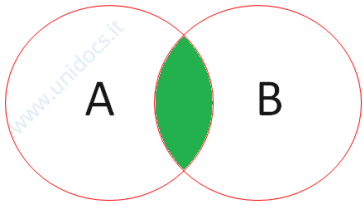
Per trattare la probabilità di eventi aleatori (o casuali) viene utilizzata la **TEORIA DEGLI INSIEMI**.

**TEORIA DEGLI INSIEMI**

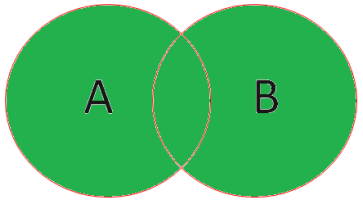
Se ho due insiemi A e B



$A \cap B \rightarrow$  A intersecato a B (" $\cap$ " si legge "e")



$A \cup B \rightarrow$  A unito a B (" $\cup$ " si legge "o")



Se A e B sono due eventi allora posso avere la **PROBABILITÀ** di A intersecato a B [ $P(A \cap B)$ ]

oppure la **PROBABILITÀ** di A unito a B [ $P(A \cup B)$ ]

- $P(A \cap B) \rightarrow$  significa che sto cercando la probabilità che i due eventi si verifichino insieme.

Ad esempio, se considero due eventi:

**EVENTO A**  $\rightarrow$  andare a mangiare la pizza

**EVENTO B**  $\rightarrow$  andare al cinema

Se **EVENTO A** e **EVENTO B** si verificano, allora  $P(A \cap B)$  vuol dire andare a mangiare la pizza e andare al cinema.

$P(A \cup B)$  invece significa andare a mangiare la pizza oppure andare al cinema

$P(A \cap B) \rightarrow$  si chiama **TEOREMA DEL PRODOTTO** perché avviene una moltiplicazione

$P(A \cup B) \rightarrow$  si chiama **TEOREMA DELLA PROBABILITÀ TOTALE** e avviene una somma o una sottrazione

La  $P(A)$  (così come  $P(B)$ ...) è un numero compreso tra 0 (l'evento non si verifica) e 1 (l'evento si verifica) [ $0 \leq$

$P(A) \leq 1$ ]

L'unione tra due insiemi ( $A \cup B$ ) è data da:  $A \oplus B \ominus A \cap B$

$\oplus/\ominus \rightarrow$  somma insiemistica

E quindi  $P(A \cup B): P(A) + P(B) - P(A \cap B)$

Invece,  $P(A \cap B): P(A) \bullet P(B|A)$  [" $P(B|A)$ " è una **probabilità condizionata** e significa "dato che si verifica A, qual è la probabilità che si verifica B?"]

**Esistono varie definizioni per quanto riguarda la PROBABILITÀ:**

- DEFINIZIONE CLASSICA  $\rightarrow P(A) = \frac{\text{CASI FAVOREVOLI}}{\text{CASI POSSIBILI}}$   
Ad esempio se lancio un dado, la probabilità che esce la faccia n°6 è 1/6
- DEFINIZIONE FREQUENTISTA, dice che in una serie di n prove, l'evento si verificherà  $N_A$  volte.  
Quando le prove diventano tantissime allora il  $\lim_{n \rightarrow \infty} \left( \frac{N_A}{n} \right) = P(A)$ .  
Questa definizione è anche detta LEGGE DEI GRANDI NUMERI [ $nA = n \cdot P(A)$ ]  
Ad esempio se voglio constatare quante volte esce il n°6 nel lancio dei dadi, in seguito a svariati lanci faccio il rapporto tra il numero di volte che esce la faccia n° 6 e il numero di lanci che ho fatto e ottengo P(A) [quindi è un rapporto tra il numero delle volte che si verifica l'evento sul numero delle prove e all'aumentare del valore di n (numero delle prove) tanto più il rapporto  $nA/n$  si avvicinerà alla probabilità dell'evento P(A)]
- DEFINIZIONE ASSIOMATICA [ $0 \leq P(A) \leq 1$  ;  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  ;  $P(A \cap B) = P(A) \cdot P(B|A)$ ]

**ESEMPIO DI TABELLA DOPPIA**

		MODALITÀ STATO DI SALUTE Y		
		MALATI	SANI	
MODALITÀ ESPOSIZIONE X	ESPOSTI	30	19	49
	NON ESPOSTI	20	10	30
		50	29	79

Se volessi scegliere a caso tra questi 79 soggetti, qual è la probabilità che esca un soggetto esposto?

$$P(E) = 49/79$$

Se invece volessi cercare un soggetto non esposto, allora

$$P(N.E.) = 30/79$$

Probabilità che esca un soggetto malato  $P(M) = 50/79$

Probabilità che esca un soggetto sano  $P(S) = 29/79$

Si può anche calcolare l'intersezione tra due eventi  $P(X \cap Y)$ . Ad esempio,  $P(M \cap E)$  [se scelgo a caso tra i 79 soggetti qual è la probabilità che esca un individuo malato e esposto?] si calcola 30/79.

Un altro esempio può essere la probabilità che esca un soggetto sano ed esposto  $P(S \cap E) = 19/79$

Se invece volessi calcolare  $P(M/E)$  [probabilità di avere un soggetto che, dato che è esposto, è malato] allora si calcola 30/49

Due eventi si dicono indipendenti se il verificarsi di uno non modifica la probabilità del verificarsi dell'altro. Come posso verificare se la probabilità si modifica? Posso verificare se due eventi sono indipendenti calcolando, per lo stesso evento, la probabilità condizionata di un evento a un altro evento [ad esempio  $P(M/E) = 30/49$ ] e la probabilità non condizionata [ad esempio  $P(M) = 50/79$ ].

Negli esempi presi in considerazione, la probabilità condizionata risulterà 0,61 mentre quella non condizionata risulterà 0,63. Sono due risultati molto vicini tra loro, quindi l'esposizione non ha un'influenza sulla malattia ed ha un indice di rischio bassissimo (si abbassa del 2% la probabilità di essere malati se si è esposti).

**CALCOLO ODD**

Per calcolare l'ODD devo, come in questo caso, calcolare la percentuale di malati e sani sia tra gli esposti che tra i non esposti.

Quindi devo effettuare la ricerca dei malati tra gli esposti [ $P(M/E) = 30/49 = 0,61$ ] e la ricerca dei sani tra gli esposti [ $P(S/E) = 19/49 = 0,39$ ].

Allora calcolo l'ODD tra gli esposti:  $0,61$  (PROBABILITÀ EVENTO) /  $0,39$  (PROBABILITÀ NON EVENTO) = **1,56**

In seguito, effettuo la ricerca dei malati tra i non esposti [**P (M/N.E.):** 20/30= 0,67] e la ricerca dei sani tra i non esposti [**P (S/N.E.):** 10/30= 0,33]

Allora calcolo l'ODD tra i non esposti: **0,66(PROBABILITÀ EVENTO) / 0,33 (PROBABILITÀ NON EVENTO) = 2,03**

Infine calcolo il rapporto tra le due ODD e ottengo l'OR= 1,56 : 2,03 = **0,76** (è una percentuale comune sia agli esposti che ai non esposti ed essendo <1 è più probabile che l'evento non si verifichi e non si tratta di un fattore di rischio bensì di protezione).

Per calcolare la percentuale di protezione si sottrae l'1 con l'OR= 1-0,76= 0,24 [24%]. Quindi su 100 persone, esposte a un determinato fattore, ci saranno 24 malati in meno rispetto alle persone non esposte.

Adesso prendiamo in considerazione la variabile X stato di salute, ovvero malati e sani.

$$X \begin{cases} M = 50/79 [P] \\ S = 29/79 [1-P] \end{cases}$$

Nella funzione **P (x) = f (x)**, basta sostituire la x nella f e si ottiene la probabilità.

Dato che le funzioni "lavorano" con i numeri bisogna dare dei valori, quindi:

- M= 1
- S = 0

La funzione che assegna la probabilità **P** quando **x=1** e **1-P** quando **x=0**, venne ricavata da Bernoulli

ed è  $P(x) = P^x \cdot (1-P)^{1-x}$  [prende il nome di VARIABILE ALEATORIA]

Se si vuole cercare la probabilità che si trovi una persona sana:

$$\rightarrow P(X=0) = P^0 \cdot (1-P)^{1-0} = 1-P$$

Per calcolare la variabilità di x la formula è  $V(x) = P(1-P)$

Per calcolare la media della x la formula è  $E(x) = P$

Per calcolare la varianza la formula è  $V(X) = E(x - P)^2 = P(1 - P)^2 + (1 - P)(0 - P)^2 = P(1 + P^2 - 2P) + (1 - P)P^2$

Se volessimo calcolare la varianza dell'essere malato dato che si è esposti:

Se faccio un campione di n unità da una fabbrica per vedere il numero di sani e malati, il numero di malati sarà dato da  $Z = \sum_{i=1}^n X_i$  e z potrà assumere valori {0,1,2,3...n}.

Per calcolare la probabilità di z si usa la BINOMIALE :

$$P(Z) = \binom{n}{z} P^z (1-P)^{n-z}$$