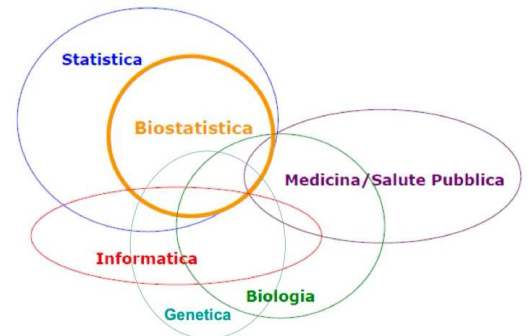


# VARIABILITÀ STATISTICA

**Statistica** = scienza che studia fenomeni collettivi soggetti a variabilità

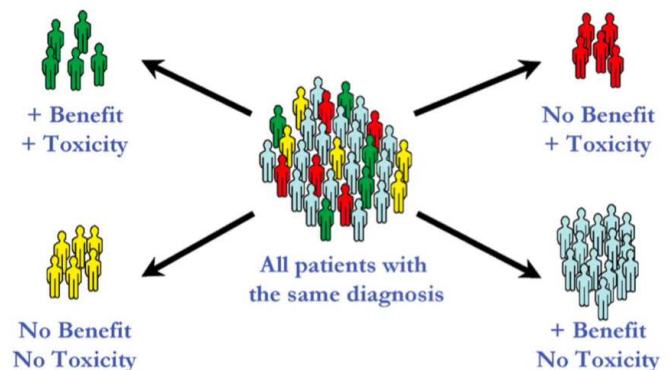
**Biostatistica** = disciplina che si occupa dell'osservazione sviluppo e dell'applicazione dei metodi statistici per lo studio delle problematiche legate alla salute



## Variabilità

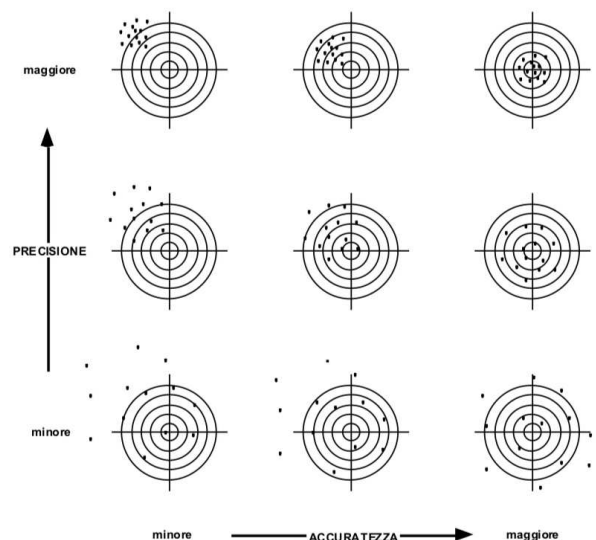
= variabilità di un fenomeno è l'attitudine di questo a manifestarsi in diversi modi, cioè con diverse modalità

- obiettivo dei metodi statistici è comprendere andamento di fenomeno tenendo conto degli effetti di questa variabilità
- **variabilità strumentale** => variabilità può essere legata a procedura di misurazione, a operatore o a strumentazione
- può riguardare rilevazioni riferite a soggetti o ripetizione di una misura su uno stesso soggetto
- **variabilità biologica** = variabilità propria di un fenomeno
- tanti fattori differenziano i soggetti in studio
- c'è una variabilità anche in uno stesso individuo dovuta a tempo o altri fattori **emozionali**, **metabolici**...



## Accuratezza e precisione

- **accuratezza** = capacità di misurare fenomeno in modo corretto controllando variabilità strumentale. l'errore introdotto da una misura poco accurata è **errore sistematico** (rilevabile e correggibile)
- **precisione** = capacità di contenere la variabilità biologia studiando fenomeno in un contesto più omogeneo. errore introdotto da imprecisione è **errore casuale** (quantificabile ma non eliminabile)



## Terminologia

**Variabili** = caratteristiche delle unità statistiche che possono assumere una pluralità di modalità al variare dell'unità su cui sono rilevate

- indicate con lettere maiuscole tra le ultime X, Y, Z

**Dati** = valori numerici o modalità categoriche assunti dalle variabili

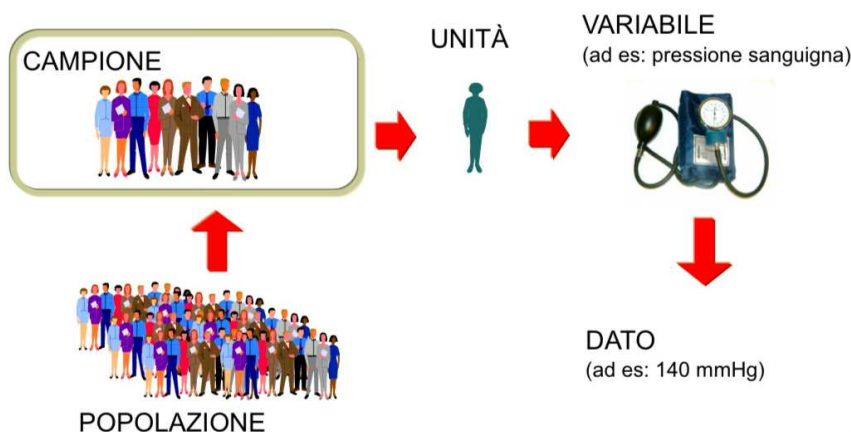
- indicati con lettere minuscole e differenziati con un indice che distingue le diverse unità fra loro x, y...

**Popolazione** = totalità degli elementi che hanno certe caratteristiche

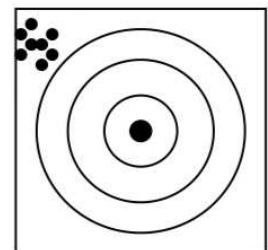
- raramente si studiano tutte le singole unità di una popolazione perché (esame di un numero ridotto consente di trarre conclusioni su popolazione superando seguenti problemi) :
  - popolazione può essere molto grande e lo studio lungo e costoso
  - unità presentano variabilità ridotta o assente
  - non tutte unità sono disponibili

**Campione** = sottoinsieme opportunamente estratto dall'intera popolazione (con campionamento) dal quale si possono ricavare, con margini di errore contenuti, le caratteristiche dell'intera popolazione

- deve riprodurre caratteristiche della popolazione soprattutto per gli aspetti (le variabili) ritenuti rilevanti ai fini della ricerca
- caratteristiche campione sono tanto più simili a quelle della popolazione quanto maggiore la sua numerosità



**Errore di selezione / bias di selezione** = errore sistematico che deriva dal fatto che il campione non è rappresentativo della popolazione (dovrebbe rappresentare immagine in scala ridotta della popolazione)



# Campionamento

= processo di estrazione di un sottoinsieme rappresentativo di unità statistiche da una popolazione di interesse più ampia.

Studio di un numero limitato di unità permette di trarre conclusioni generali su popolazione, la cui validità dipende da corretta applicazione dei metodi di campionamento

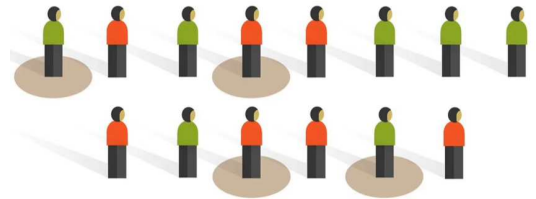
**Campionamento casuale / random sampling** => per avere campionamento corretto : se campione selezionato in modo casuale e intuitivo supporre che sia rappresentativo della popolazione di interesse

metodi di campionamento:

## a) Campionamento casuale semplice

= estraendo una certa quota di unità da popolazione in modo casuale

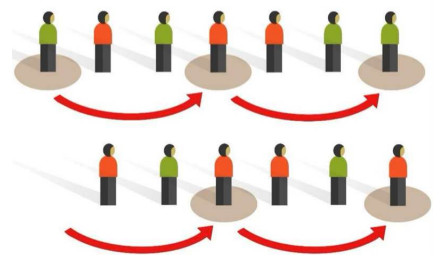
- schema di campionamento in cui ogni possibile sottoinsieme di una dimensione stabilita ha la stessa possibilità di essere estratto da popolazione
- casualità ottenuta con sistema sorteggio randomizzato di unità a partire da elenco (**base di campionamento**) in cui presenti tutte le unità della popolazione



## b) Campionamento sistematico

= unità che costituiscono campioni estratte da popolazione a intervalli regolari (**passo di campionamento**)

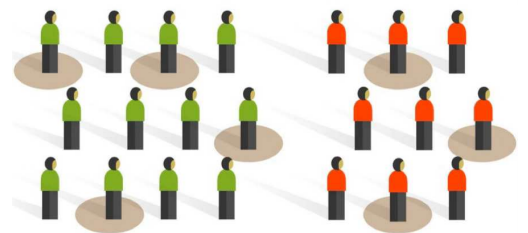
- più pratico rispetto a casuale semplice
- assicura che singole unità del campione siano distribuite uniformemente all'interno della popolazione
- occorre verificare che passo di campionamento garantisca casualità e non sia influenzato da fattori esterni



## c) Campionamento stratificato

= per studiare un carattere che, presumibilmente o notoriamente, può essere influenzato da un certo fattore presente nella popolazione

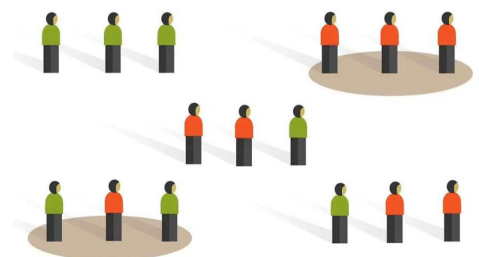
- popolazione suddivisa in categorie (**strati**) definite su base del fattore influenzante e all'interno di ogni strato si estrae un campione randomizzato
- campione ottenuto rappresenta meglio popolazione da cui è stato estratto
- la ridotta numerosità degli strati può rendere poco attendibili stime riferite agli stessi



## d) Campionamento a grappolo

= popolazione divisa in gruppi (**grappoli / cluster**) ciascuno dei quali è una rappresentazione della popolazione e si estraggono i grappoli

- tutte unità possono far parte del campione oppure solo parte (**campionamento a due stadi**)
- rispetto a altri metodi, facilità reclutamento diminuendo costi e tempi indagine
- errore può essere più elevato



## ERRORE CASUALE

- anche senza errore di selezione si può ottenere campione non rappresentativo
- tecniche dell'inferenza permettono di quantificare potenziale errore introdotto da campionamento casuale
- c'è relazione tra errore sistematico, errore casuale e numerosità campionaria :
  - errore casuale è sempre presente
  - errore casuale diminuisce con un'adeguata numerosità campionaria

