

IL CAMPIONAMENTO

I DATI

- I dati si ottengono sia tramite **osservazione** sia tramite **sperimentazione**.
- I dati osservazionali esistono in natura e vengono rilevati direttamente per come si presentano. Sono spesso osservazioni di caratteristiche antropometriche, demografiche, socio-economiche, ma non solo.
- Un'importante classe di dati osservazionali è rappresentata dai risultati di censimenti o di sondaggi d'opinione.

I DATI

Esempi di osservazione:

- rilevazione di età, statura, genere e gruppo sanguigno degli studenti universitari immatricolati presso l'Università di Ferrara nell'A.A. 2013-14
- determinazione dell'ematocrito dei pazienti ricoverati nel reparto di Ematologia
- misurazione del livello di vari inquinanti nell'aria
- monitoraggio delle visite ad un certo sito web in un dato periodo di tempo

Allo scopo di eseguire uno studio è necessario identificare una popolazione sulla quale effettuare alcune misure.



Ad esempio possiamo pensare di eseguire uno studio sulla popolazione italiana

oppure sulla popolazione di una regione o su quella costituita dagli utenti di una ASL.



CONCETTI FONDAMENTALI

1. È impossibile una conoscenza esaustiva della popolazione (**universo**)
2. si deve procedere ad una selezione di un sottoinsieme (**campione**)
3. Tale campione deve essere rappresentativo (**non distorto**) e, per forza di cose, limitato



Sulla base delle informazioni fornite dal campione si cercherà di trarre conclusioni valide a livello di popolazione (universo)

INFERENZA STATISTICA

STUDIO DELLE RELAZIONI TRA
CAMPIONE E POPOLAZIONE



possibilità, sulla base dei risultati
ottenuti su un campione, di fare delle
affermazioni sulla popolazione

Il principale obiettivo di un campionamento è quello di raccogliere dati che consentiranno di generalizzare, con un certo grado di certezza, all'intera popolazione le conclusioni ottenute dal campione (una parte del fenomeno).

Questo processo di generalizzazione è detto «inferenza».

- Nella ricerca, il CAMPIONE (l'esperienza particolare che viene considerata in uno studio) è un mezzo per apprendere e/o approfondire una relazione o un fenomeno che si vuole generalizzare a una POPOLAZIONE.
- La popolazione il più delle volte è puramente astratta, non limitata né nello spazio né nel tempo (universo).

POPOLAZIONE E CAMPIONE

- Raramente in uno studio è possibile esaminare ogni singolo elemento facente parte della popolazione.
- Spesso si è limitati dalle risorse disponibili (economiche, di personale, di laboratori, di tempo ecc.)
- In altre occasioni, anche supponendo di disporre di risorse illimitate, l'intera popolazione da studiare non è fisicamente raggiungibile oppure non è del tutto nota

POPOLAZIONE E CAMPIONE

- Oggetto ottimale di ogni indagine descrittiva è la popolazione statistica e tale ricerca viene definita **esaustiva** in quanto coinvolge tutto lo spazio campione.
- Alcune informazioni a carattere sociale, sanitario ed epidemiologico richiedono l'utilizzo di dati forniti dal censimento della popolazione italiana (era effettuato ogni 10 anni) in grado di fornire valide informazioni su tutta la popolazione residente in Italia.
- Per analogia una ricerca esaustiva viene definita **censuaria**.

POPOLAZIONE E CAMPIONE

- Tipicamente, saremo interessati a una proprietà numerica della popolazione, che chiameremo **parametro**, e baseremo la nostra inferenza sul valore osservato nel campione di una quantità, che chiameremo **statistica**.
- Questo è uno dei problemi più importanti affrontati dalla statistica: come **stimare un parametro di una popolazione usando una statistica calcolata su un campione**.

POPOLAZIONE O UNIVERSO

$$\mu, \sigma$$
$$y = \alpha + \beta x$$

TEORIA DEL CAMPIONAMENTO

Quali soggetti campionare?



CAMPIONE

X_1, X_2, \dots, X_n
 Y_1, Y_2, \dots, Y_n

STATISTICA INFERENZIALE

Cosa possiamo dire dei veri parametri della popolazione?
Qual è il margine di incertezza?

STATISTICA DESCRITTIVA

STATISTICHE

\bar{x}, S
 $Y = a + bx$ 

POPOLAZIONE E CAMPIONE

Variable



1 2 3 4 5

Statistic



Average = 3.75

Parameter



Average = 3.72

Il problema che si deve affrontare è quello della

VARIABILITÀ

I metodi statistici chiamano in causa osservazioni che variano da campione a campione e portano un certo grado di incertezza in ogni analisi.

L'obiettivo primario di pressoché tutti i metodi statistici è quello di comprendere il comportamento di un fenomeno tenendo conto degli effetti di questa variabilità.

PROBLEMI RELATIVI ALL'USO DI UN CAMPIONE

- Se nella popolazione non ci fosse variabilità, ci basterebbe osservare un solo individuo per conoscere l'intera popolazione.
- In presenza della variabilità, invece, le conclusioni tratte dall'osservazione di un **campione** sono soggette ad **incertezza**, a meno che il campione non coincida con l'intera popolazione.

POPOLAZIONE E CAMPIONE

- Una unità statistica può far parte contemporaneamente di più popolazioni:
 - un individuo può essere o meno affetto da una particolare malattia
 - risiedere o meno in un dato Comune
 - può avere o non avere l'abitudine di fumare
- Una popolazione statistica dipende dalla variabile oggetto di indagine e viene definita dallo sperimentatore che deve stabilire un **criterio oggettivo** per assegnare o meno un elemento alla popolazione.

POPOLAZIONE

È la completa collezione degli elementi (soggetti, misure, campioni chimici, ecc.) oggetto di studio.

La collezione è **COMPLETA** quando include assolutamente **TUTTI** gli elementi d'interesse.

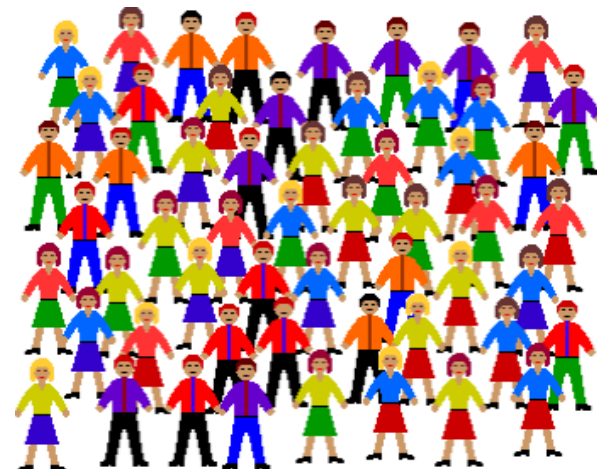
- Insieme che raccoglie tutte le osservazioni possibili, relativamente ad una data variabile o ad un dato fenomeno

- Può essere finita (comunque molto grande) o infinita

POPOLAZIONE

■ Trattiamo come popolazioni anche insiemi che non sono enumerabili e che si realizzeranno nel futuro:

Esempio: quando ci riferiamo ai malati di una certa malattia vogliamo formulare una previsione valida anche per i casi che non sono ancora stati diagnosticati.



POPOLAZIONE

■ **Non rappresenta esclusivamente un gruppo di individui** (popolazione abitante in Italia, popolazione degli iscritti in una ASL) ma piuttosto una serie di informazioni o dati raccolti in un ambito di interesse scientifico, omogenei per metodo di valutazione e raccolta:

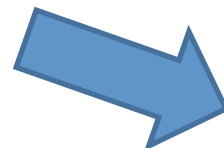
- Soggetti di età > 60 aa. appartenenti a una determinata ASL
- Valori di Ht dei ricoverati nel mese di ottobre nel reparto di ematologia di un ospedale
- Tempi di sopravvivenza dopo trapianto midollo osseo
-

POPOLAZIONE

- La popolazione statistica non corrisponde quasi mai alla popolazione anagrafica.

Esempio:

Un pediatra vuole valutare il valore nutrizionale di un latte sulla crescita dei neonati. Non può effettuare la verifica sulla popolazione dei neonati (insieme troppo vasto e disperso) → studio su un piccolo numero di neonati da cui formulare ipotesi circa l'azione del latte sui neonati in genere



INFERENZA

DEFINIZIONI

CAMPIONE

Un sottoinsieme degli elementi della popolazione

- rappresentativo
- scelto con metodo che ne assicuri la validità
- sufficientemente numeroso



DEFINIZIONI

CENSIMENTO

La collezione di dati da ogni elemento della popolazione

POPOLAZIONE BERSAGLIO

È la popolazione che intendiamo descrivere

Ad es. pazienti cardiopatici, adolescenti della provincia, ...

DEFINIZIONI

UNITÀ DI CAMPIONAMENTO O UNITÀ DI STUDIO

Sono gli elementi della popolazione

Una persona, una famiglia, una città, ...

POPOLAZIONE DI STUDIO

È la popolazione da cui il campione viene selezionato

DEFINIZIONI

SCHEMA DI CAMPIONAMENTO

È la lista che comprende tutti i membri della popolazione di studio

PARAMETRO

Un numero che descrive una caratteristica della
POPOLAZIONE

STATISTICA

Misura numerica che descrive una qualche
caratteristica del CAMPIONE

POPOLAZIONE FINITA ED INFINITA

- Una popolazione può essere **infinita** quando non si è in grado di identificarne tutte le unità

ESEMPIO

gli ipercolesterolemici: una parte sa di essere in questa condizione ma alcuni, pur avendo un tasso ematico di colesterolo superiore alla norma, non ne sono consapevoli quindi non possono essere identificati

POPOLAZIONE FINITA ED INFINITA

- Una popolazione può essere **finita** come nel caso di:
 - uno studio sui neonati a termine di una determinata area geografica
 - neonati a termine di sesso maschile di uno specifico reparto di ostetricia in un periodo determinato di tempo

POPOLAZIONE FINITA ED INFINITA

POPOLAZIONE INFINITA

Neonati

Neonati a
termine

POPOLAZIONE FINITA

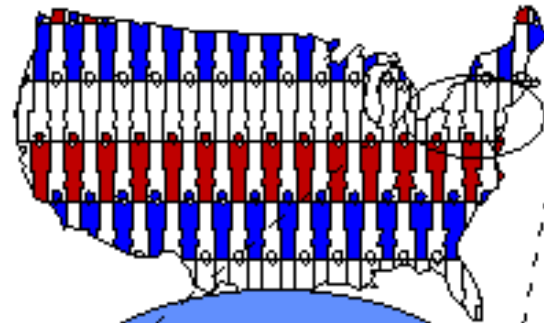
Neonati a
termine
di una
Regione

Neonati a
termine
di un
reparto
di
ostetricia
di un
ospedale

Neonati a
termine
di sesso
maschile
del
reparto

CAMPIONAMENTO

Who do you want to generalize to?



The Theoretical Population

What population can you get access to?



The Study Population

How can you get access to them?



The Sampling Frame

Who is in your study?



The Sample

UTILIZZO DEL CAMPIONE

VANTAGGI

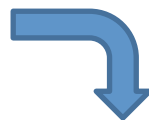
- **risparmio di lavoro e di costi** dell'indagine perché vengono ridotte le unità di osservazione
- **la raccolta dell'informazione può essere più attendibile e più accurata**
- **unica possibilità quando la popolazione, su cui si vogliono fare inferenze, è infinita.**

SVANTAGGI

- **imprecisione delle stime:** le misure calcolate sono solo una approssimazione delle vere misure della popolazione e variano da campione a campione.

CAMPIONI ED ERRORI

- L'utilizzo del campione introduce delle *fonti di errore* nella stima dei parametri incogniti della popolazione, indipendentemente dal tipo di campionamento.
- Un campione *non è mai perfettamente rappresentativo* della popolazione da cui è stato estratto.
- La differenza esistente tra le caratteristiche del campione e quelle vere della popolazione bersaglio è definita



ERRORE DI CAMPIONAMENTO

Fonti di distorsione

Indipendentemente dal tipo di campionamento, quando si seleziona un campione, la fonte potenziale di errore è rappresentata da:

1. bias di selezione

(Campione selezionato differisce in modo sostanziale da popolazione di riferimento)

BIAS = errore in studio epidemiologico che causa valutazione non corretta dell' associazione tra esposizione e malattia

2. bias di informazione

(Gli eventi studiati sono registrati in modo diverso tra i gruppi di studio)

- Di ricordo
- Dell'intervistatore
- Di misclassificazione
- Perdita di dati al follow up

CAMPIONI ED ERRORI

■ Le ragioni per cui il valore del nostro stimatore (calcolato sul campione) può essere diverso dal valore del parametro (nella popolazione) sono essenzialmente due:

1

VARIAZIONE CASUALE

- è dovuta ad un elemento ineliminabile: il CASO
- l'errore che ne deriva può essere stimato

2

SELEZIONE VIZIATA

- è dovuta ad un cattivo campionamento
- l'errore che ne deriva NON può essere stimato

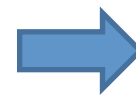
FONTI DI ERRORE (VARIAZIONE CASUALE)

■ La **variazione casuale** è dovuta al caso, cioè a quell'«insieme di fattori o cause, piccole o grandi, che agiscono su un fenomeno senza che noi possiamo o vogliamo controllarli esattamente e prevederne quindi l'azione» (Cavalli-Sforza).

■ La **variazione casuale** fa sì che una misura effettuata su un campione non fornisca un valore identico a quello ottenibile misurando l'intera popolazione: c'è **sempre** un certo errore, che viene detto **errore campionario**.



Errore dovuto all'osservazione solo di una parte della popolazione



Non può essere calcolato con certezza, ma può essere **STIMATO** (calcolo approssimativo)

FONTI DI ERRORE (SELEZIONE VIZIATA)

- Un campione che non è stato ottenuto correttamente fornisce misurazioni e risultati per i quali è **impossibile calcolare il cosiddetto «errore di campionamento»**.
- La selezione viziata fa sì che all'errore campionario si sommi un altro tipo di errore, detto **errore non campionario o bias**.



SCHEMA DELLA PROCEDURA DI CAMPIONAMENTO

**POPOLAZIONE
OBIETTIVO**

Dominio su cui si può fare inferenza



**Base dello
studio**

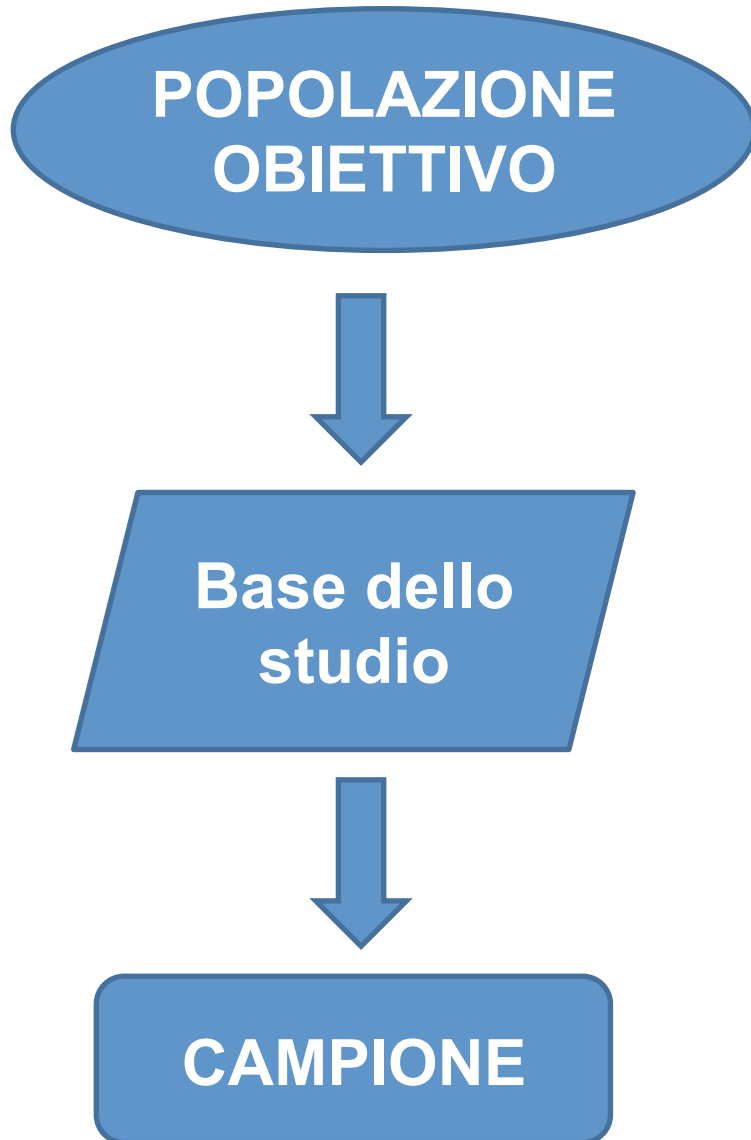
Particolare esperienza concreta, delimitata nello spazio e nel tempo, utilizzata come elemento di conoscenza nella popolazione obiettivo



CAMPIONE

Qualsiasi sottoinsieme della base

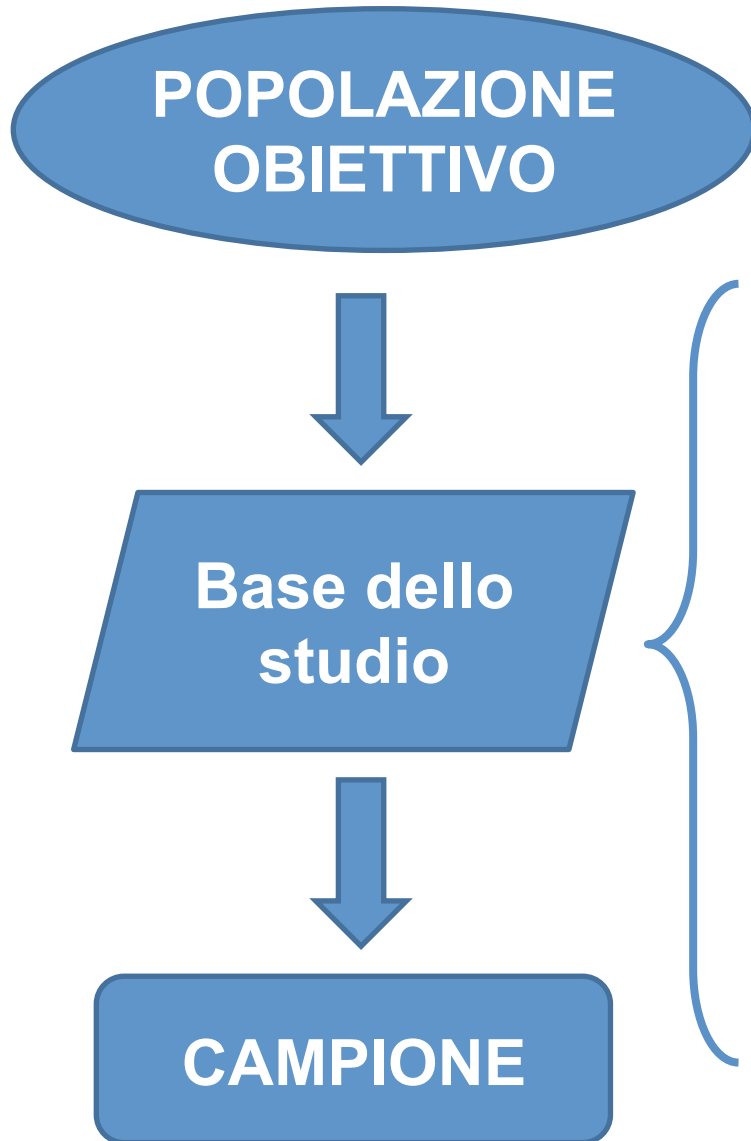
ESEMPIO POPOLAZIONE FINITA



1. Popolazione degli studenti universitari di Ferrara nell'a.a. 2014-15 (finita)
2. Popolazione adulta della città di Ferrara (finita)

1. Elenco degli iscritti
2. Registro aggiornato delle liste elettorali; elenco telefonico

ESEMPIO POPOLAZIONE INFINITA



Popolazione degli ipertesi (infinita)

- Tutti i soggetti ipertesi che si rivolgono al loro medico a seguito di disturbi legati alla malattia in una data area e tempo
- Tutti i soggetti di una data area che hanno valori di pressione superiori a un cut off
- Tutti i soggetti ricoverati con diagnosi di ipertensione

SCelta DELLA BASE DELLO STUDIO

La base dello studio è scelta con criteri logici e di efficienza in funzione della sua idoneità rispetto alla popolazione obiettivo.

In genere:

- Nelle **indagini campionarie o di prevalenza**, mirate alla stima dei parametri della popolazione, deve essere garantita la rappresentatività della popolazione obiettivo.

- Nelle **indagini etiologiche**, mirate allo studio dei fattori responsabili dell'insorgenza di specifiche patologie, deve essere garantita la confrontabilità dei gruppi che hanno esposizioni (fattori di rischio) differenti.

INDAGINI CAMPIONARIE/1

In un'indagine mirata a valutare i consumi alimentari di una Provincia italiana, le possibili basi potrebbero essere:

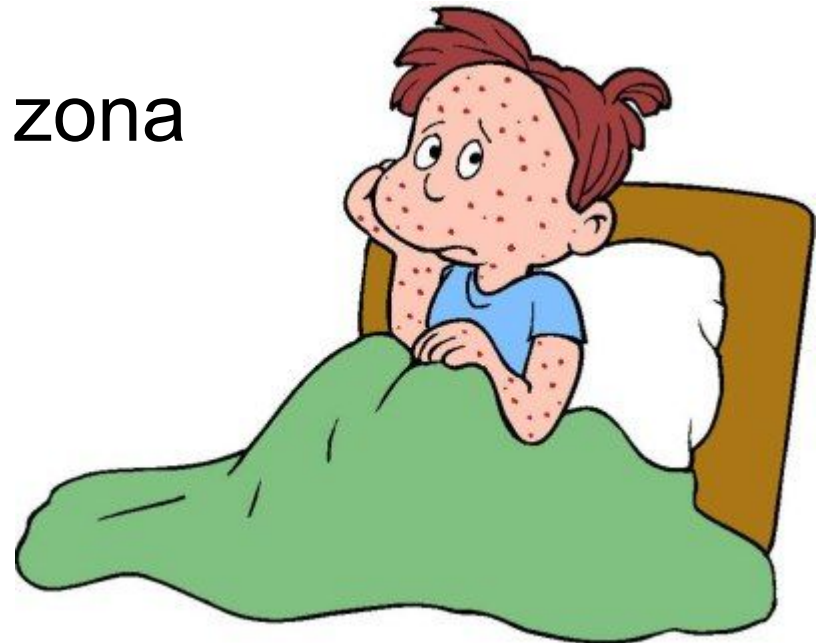
- Elenco telefonico provinciale
- Elenco delle liste elettorali



INDAGINI CAMPIONARIE/2

In un'indagine mirata a valutare le complicanze in bambini tra i 3 e i 5 anni affetti da morbillo:

- Tutti i bambini ricoverati in ospedale con diagnosi di morbillo
- Tutti i bambini iscritti alle scuole materne della zona
- Elenco dei pediatri della zona



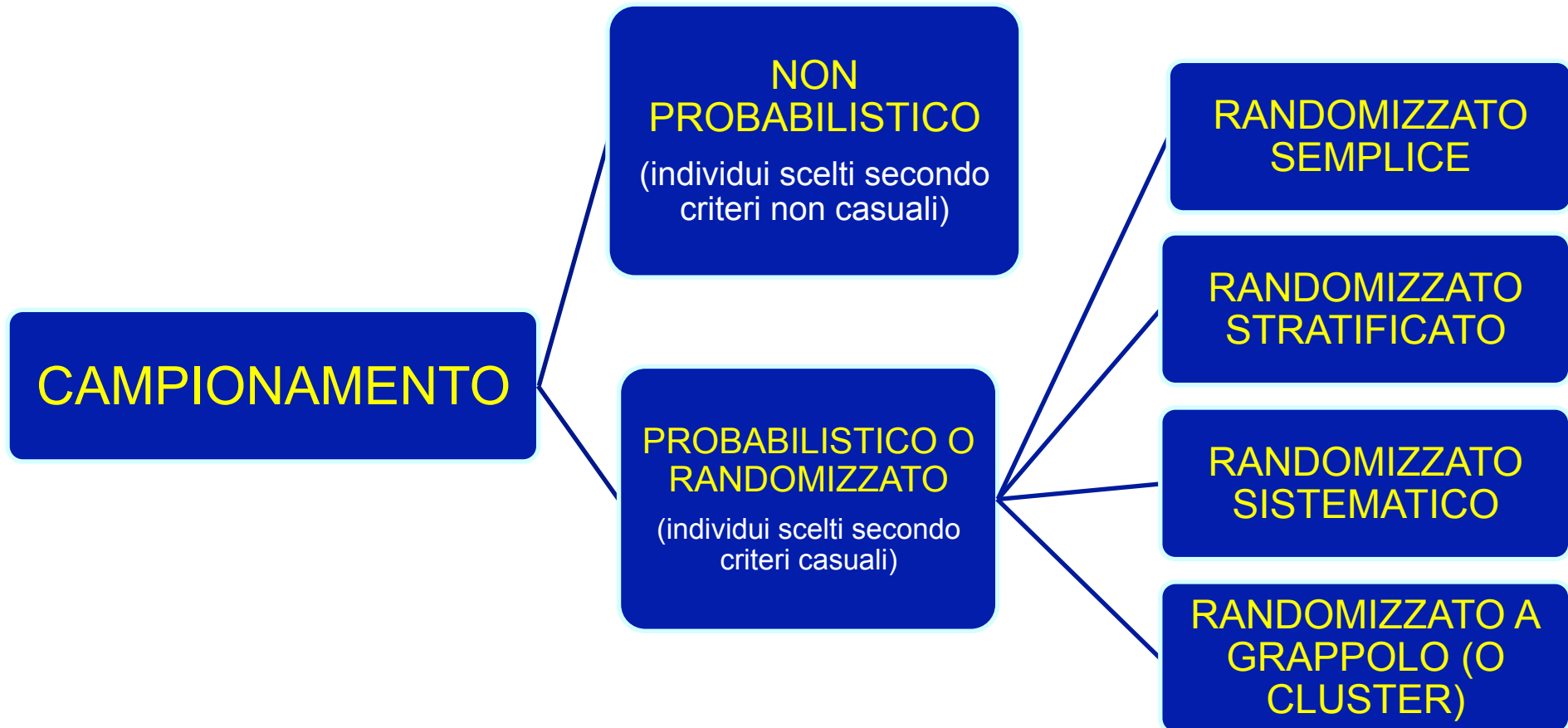
STRATEGIE DI CAMPIONAMENTO

Esistono due tipi principali di campioni

■ **Non probabilistici** → Non sono basati su alcuna teoria statistica

■ **Probabilistici** → Ogni unità della popolazione ha una probabilità nota e diversa da zero di essere incluso

STRATEGIE DI CAMPIONAMENTO



CAMPIONI NON PROBABILISTICI

CAMPIONI DI CONVENIENZA

- Gli individui sono scelti perché facilmente accessibili
- Sono normalmente distorti quindi questa tipologia di campionamento garantisce **poca affidabilità nell'assicurare la generalizzazione dei risultati**



CAMPIONI NON PROBABILISTICI

CAMPIONI DI CONVENIENZA

Esempi:

- gli studenti di una classe
- i residenti in una data via
- popolazione che afferrisce a un ospedale

si corre il rischio di osservare una popolazione che *non è rappresentativa* della popolazione generale

CAMPIONI NON PROBABILISTICI

CAMPIONAMENTO PER QUOTE

- Gli individui sono scelti in modo da “rappresentare” la popolazione per alcune caratteristiche pre-definite (ad es. per età, genere, titolo di studio, etc.)
- La popolazione viene preliminarmente suddivisa in base ad alcune caratteristiche (es. 100 maschi, 100 femmine) lasciando agli intervistatori la possibilità di scegliere le persone da intervistare (nel rispetto delle quote assegnate)

CAMPIONI NON PROBABILISTICI

CAMPIONAMENTO PER SCELTA RAGIONATA

- Consiste nella scelta di persone che si ritiene importante studiare poiché possiedono specifiche caratteristiche e non esiste una base di campionamento opportuna.

Esempi

- i pazienti affetti da una data patologia
- i malati con lo stesso grado di severità della patologia

CAMPIONI NON PROBABILISTICI

QUANDO USARLI?

- Quando è impossibile disporre di una base di campionamento (es. homeless, malattie rare)
- Nelle nuove ricerche in cui i casi di interesse non sono noti o mai stati studiati prima (es. Malattie emergenti).

N.B. I risultati ottenuti da campioni non casuali non permettono generalizzazioni con un definito e noto grado di accuratezza.

In altre parole, i dati rappresentano solo le unità che sono state studiate.

CAMPIONAMENTI PROBABILISTICI

- I campioni probabilistici garantiscono a ciascuna unità della popolazione una definita e positiva probabilità di far parte del campione.
- La scelta dei soggetti è affidata al caso (**randomizzazione**).
- Sono basati sulla teoria della probabilità e **permettono di applicare le tecniche dell'inferenza statistica ai dati ottenuti**.
- Permettono la generalizzazione dei risultati con un **dato e noto margine d'errore**.

CAMPIONAMENTI PROBABILISTICI

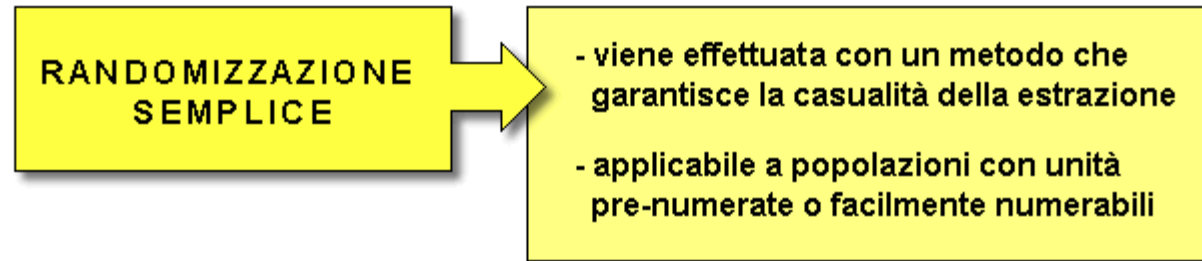
Il campionamento probabilistico utilizza uno schema di campionamento, cioè una lista di tutti i membri della popolazione

- Campionamento casuale semplice (randomizzazione)
- Campionamento sistematico
- Campionamento stratificato
- Campionamento a grappolo (cluster)

CAMPIONAMENTO CASUALE SEMPLICE

- È il tipo più semplice di campione che può essere selezionato dalla popolazione di studio.
- Le unità statistiche sono selezionate indipendentemente una alla volta fino al raggiungimento della dimensione desiderata del campione.
- Ogni unità statistica può essere scelta una volta sola.

CAMPIONAMENTO CASUALE SEMPLICE



- Se ogni unità statistica può essere scelta una volta sola, si tratta di un **campionamento senza reintroduzione**
- Ciascuna unità statistica ha la stessa possibilità di essere inclusa nel campione
- Ogni combinazione di elementi ha una identica probabilità

CAMPIONAMENTO CASUALE SEMPLICE

SVANTAGGI

- richiede la preventiva numerazione di tutti i componenti della popolazione (o una loro facile numerabilità)
- individuare nella popolazione i soggetti corrispondenti al numero estratto

CAMPIONAMENTO CASUALE SEMPLICE

Una modalità per selezionare un campione casuale semplice consiste nell'utilizzare una **tavola di numeri casuali**



Come procedere?

1. Elencare e numerare le unità in studio
2. Selezionarle una ad una fino al raggiungimento della dimensione campionaria desiderata utilizzando una tavola dei numeri casuali o un PC

TAVOLA DEI NUMERI CASUALI

83760	31255	71609	89887	00940	54355	44351	89781	58054	65813
66280	56046	50526	33649	87067	02697	06577	16707	96368	47678
70218	28376	98535	34190	96911	81578	97312	20500	48030	27256
02349	88955	52760	73696	91510	38633	38883	90419	26716	98215
93606	21415	34843	12969	84847	06280	95916	12991	08262	58385
24274	18747	37327	06780	08032	98544	24902	81607	87914	22721
67778	70496	57588	89813	71211	83848	93494	27946	79722	70315
89134	06458	40897	73025	04191	77144	49340	89446	71852	80854
83625	00097	71092	12009	63223	37993	50067	25688	98179	34628
03324	68196	72460	55616	27006	50790	28629	88726	97143	63218
84392	36623	91964	03505	46525	40490	77787	68545	02795	72676
76926	10866	39734	50512	04181	78012	78705	86194	28371	54535
06612	60200	49085	85108	71438	10099	99027	65081	82492	77584
76721	02889	95600	07984	31925	59685	91510	40039	43205	37149
64599	51953	55612	89088	58436	21501	86219	74528	59805	65020
79440	99677	49530	55291	34867	54774	52449	23294	94815	95124
35839	00177	57742	09502	42624	29017	94284	81409	36904	54329
83013	94568	75490	12138	24067	86954	00910	61171	82982	87191
19980	47085	46064	19102	26297	79745	99611	04555	52501	32088
55716	10350	67645	62922	81919	47925	91448	36025	20611	38939

GENERAZIONE DI UN CAMPIONE CASUALE SEMPLICE

Sia definita una base composta di N individui, e sia n la dimensione prescelta del campione:

1. Ad ogni individuo della base da 1 a N è assegnato un numero progressivo
2. Si leggono le tavole (tavole dei numeri casuali) a gruppi di tante cifre quante sono le cifre di N
3. I soggetti corrispondenti ai primi n numeri letti (non ripetuti e tali per cui $n < N$) saranno selezionati a far parte del campione

La quantità n/N è definita

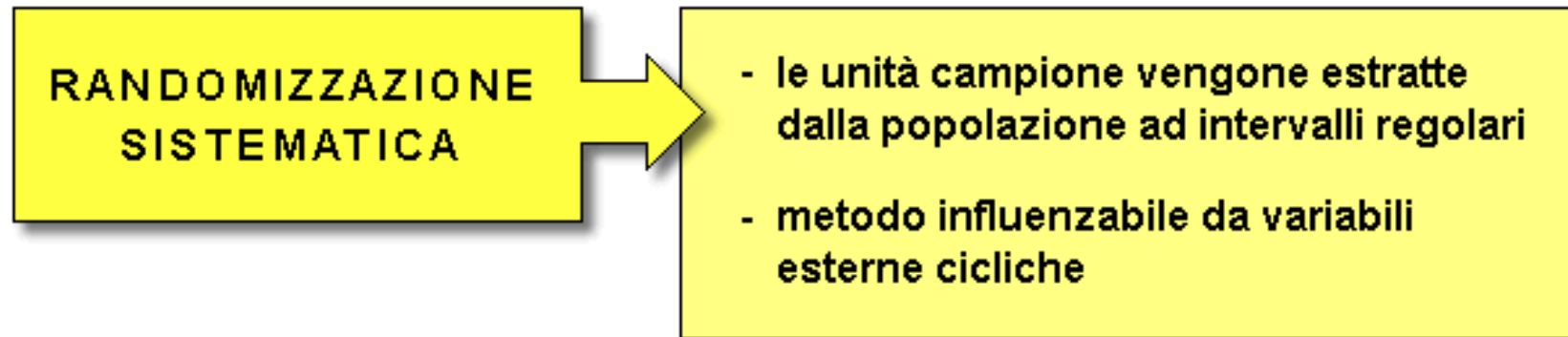
FRAZIONE DI CAMPIONAMENTO

- Il campionamento viene di solito condotto **predefinendo la dimensione del campione.**
- Si **calcola quindi la frazione di campionamento** (cioè la probabilità che un dato individuo sia estratto ed entri a far parte del campione).
- Data una popolazione con N individui ed un campione di C individui (dove N è molto grande rispetto a C) **la probabilità per l'i-esimo individuo è C/N.**

Frazione di campionamento $\psi = \frac{\text{dimensione del campione}}{\text{dimensione della popolazione}}$

- Nel campionamento casuale semplice la stessa frazione di campionamento viene applicata a tutta la popolazione.
- Se la frazione di campionamento è piccola ($C \ll N$), ψ si mantiene praticamente costante anche se i soggetti campionati escono dalla popolazione.
- Altrimenti ψ varia nel corso del campionamento ed occorre tenerne conto applicando una correzione (correzione per la popolazione finita)

CAMPIONAMENTO SISTEMATICO



Caratteristiche:

- Ciascuna unità di studio ha la stessa probabilità di essere selezionata
- Combinazioni di elementi hanno differenti probabilità

CAMPIONAMENTO SISTEMATICO

- Il campionamento sistematico può essere utilizzato se è disponibile una lista completa degli elementi N di una popolazione.
- È simile al campionamento casuale semplice, ma più semplice da applicare.
- A differenza del campionamento casuale semplice, in questo caso si richiede la selezione di un unico numero casuale.

CAMPIONAMENTO SISTEMATICO

- Le unità che costituiranno il campione sono estratte dalla popolazione con un intervallo regolare
- Viene inizialmente stabilito il passo di campionamento preferibilmente associato per comodità alla percentuale di unità da selezionare (passo 10 per campione costituito dal 10% della popolazione)

PASSO DI CAMPIONAMENTO



- In maniera casuale si procede alla scelta della prima unità

Il metodo di selezione **deve essere rispettato fino al completamento del campione** per mantenere la casualità della scelta

Esempio:

- inserimento nel campione di un caso ogni dieci parti cesarei effettuati in un reparto di ostetricia
- la selezione di un emocromo ogni cento in base al numero progressivo di archiviazione in un periodo di tempo stabilito

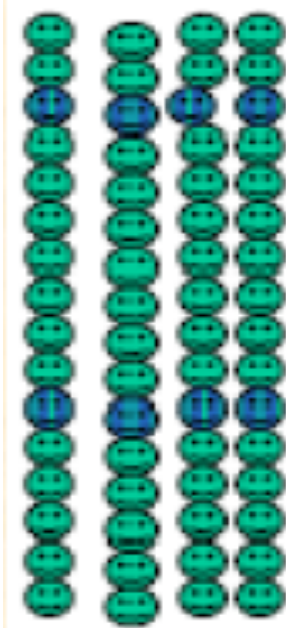
CAMPIONAMENTO SISTEMATICO

Come procedere?

1. Utilizzare una lista degli individui della popolazione
2. Estrarre **casualmente** un numero tra 1 e k da cui cominciare la selezione, con k scelto in modo da ottenere un campione della dimensione voluta ($k=N/n$)
3. Selezionare sistematicamente un individuo ogni k

CAMPIONAMENTO SISTEMATICO

ESEMPIO



$N = 100$

want $n = 20$

$N/n = 5$

**select a random number from 1-5:
chose 4**

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

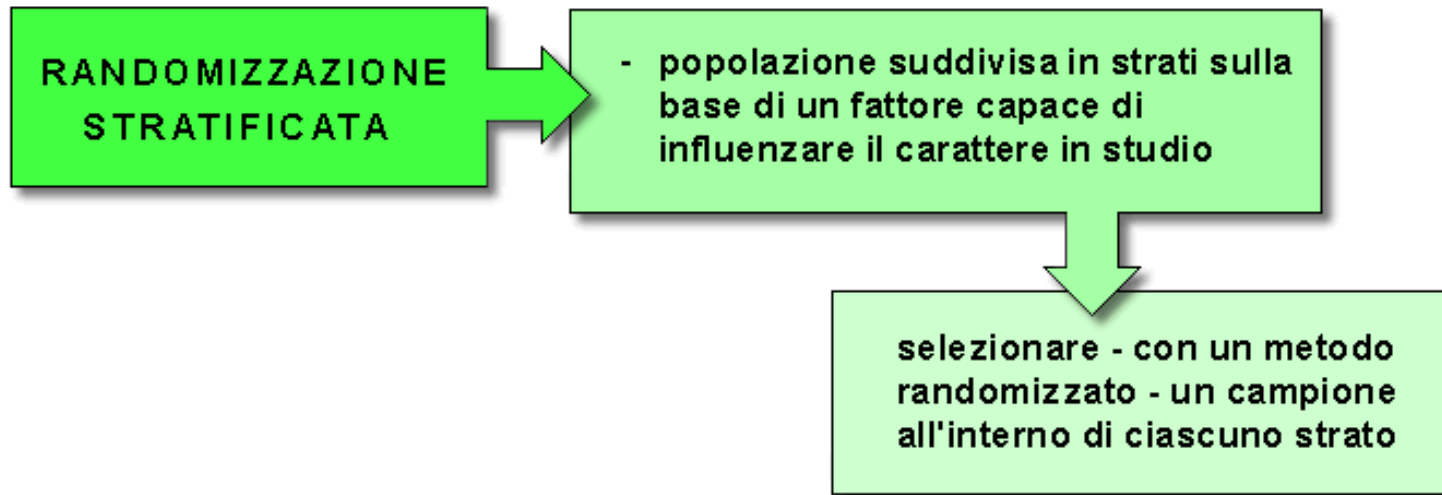
CAMPIONAMENTO STRATIFICATO

■ Il campionamento per randomizzazione stratificata viene effettuato quando si studia un carattere che, presumibilmente o notoriamente, è influenzato da un certo fattore presente nella popolazione.

1. Prima di effettuare l'estrazione del campione la popolazione viene **suddivisa in strati** basati sul fattore che influenza il carattere da studiare.

2. All'interno di ciascuno strato **si sceglie un campione con un metodo che garantisca la casualità** come, ad esempio, il metodo della randomizzazione semplice o sistematica.

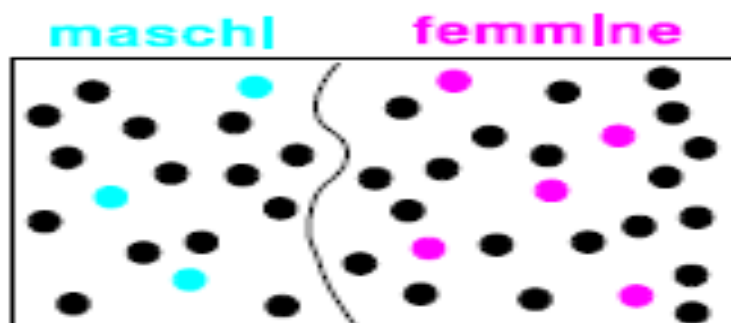
CAMPIONAMENTO STRATIFICATO



- Allestire un campione che riproduca la struttura della popolazione
- Avere sottogruppi omogenei di popolazione che permettano stime migliori dei parametri della popolazione

CAMPIONAMENTO STRATIFICATO

- Questo metodo garantisce che ogni strato sia rappresentato nel campione generale.

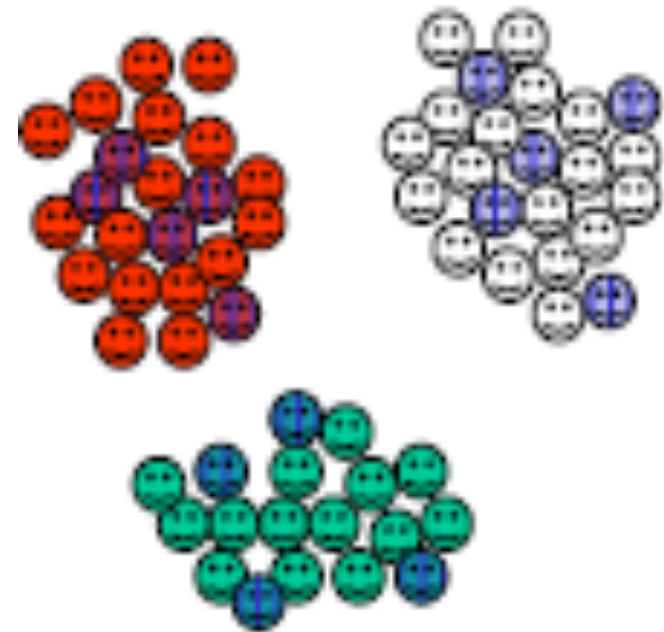


- Se la dimensione dei campioni dei vari strati è selezionata correttamente, la media stimata di un campione stratificato ha una **VARIANZA PIÙ PICCOLA** ed è **perciò più precisa** della media di un campione casuale semplice.

CAMPIONAMENTO STRATIFICATO

■ Il vantaggio principale è rappresentato dal vantaggio di conoscere la struttura di una popolazione, in base a fattori che possono favorire vizi di campionamento e tenerne conto nella selezione campionaria.

■ La casualità viene rispettata in quanto, nell'ambito di ciascuno strato, la selezione segue i criteri del campionamento casuale semplice.



CAMPIONAMENTO STRATIFICATO

- La probabilità di essere incluso nel campione può essere diversa tra i gruppi
 - *Utilizzare una frazione di campionamento più grande (oversampling) per i sottogruppi più piccoli migliora la possibilità di confronti tra gruppi*

- Se gli strati tendono ad essere omogenei al loro interno e diversi tra loro rispetto alla variabile di interesse, il campionamento stratificato è più efficiente del campionamento casuale semplice
 - *Stime più precise a parità di numerosità*

SCOPI DELLA STRATIFICAZIONE

Sulla statistica inferenziale:

- Allestire un campione che riproduca la struttura della popolazione
- Avere sottogruppi omogenei che permettano stime migliori dei parametri della popolazione

Su un'indagine:

- Individuare insiemi di unità con caratteristiche specifiche:
 - Reparti omogenei di un ospedale
 - Pazienti affetti da una data patologia
 - Soggetti esposti a un particolare rischio
 - Ricoverati di un reparto in un dato anno
 - Iscritti ad un ordine professionale

CAMPIONAMENTO STRATIFICATO

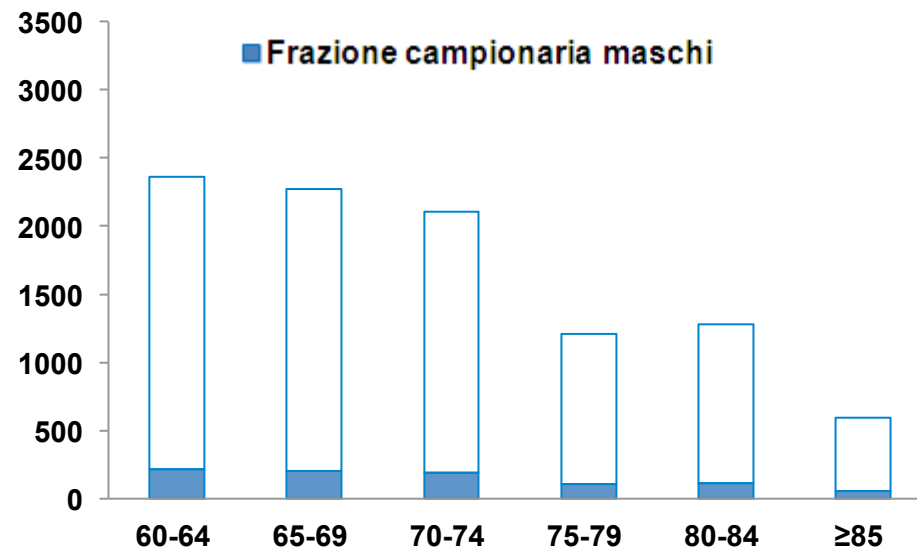
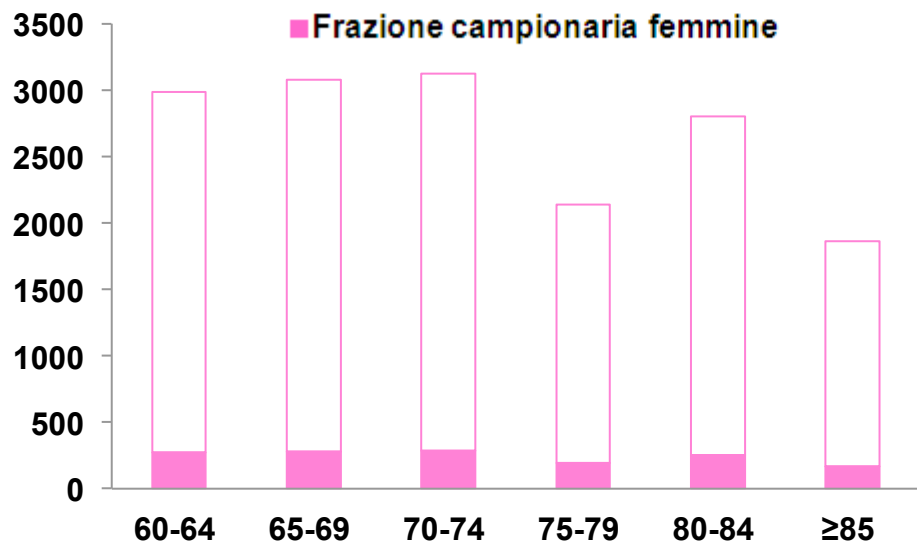


ESEMPIO

Indagine sulla pressione arteriosa delle persone >60 aa. residenti in una determinata località.

- Si identificano gli ultra-60enni suddividendoli per sesso e per classe d'età
- La stratificazione permette di strutturare il campione rispettando le proporzioni dei vari strati sesso-età: se la ricerca prevede di operare su un campione del 10% della popolazione per ogni strato viene selezionato casualmente un numero di soggetti corrispondente a tale frazione

	POPOLAZIONE			CAMPIONE		
Età	Maschi	Femmine	Totale	Maschi	Femmine	Totale
60-64	2147	2712	4859	215	271	486
65-69	2064	2797	4861	206	280	486
70-74	1913	2840	4753	191	284	475
75-79	1100	1944	3044	110	194	304
80-84	1162	2546	3708	116	255	371
≥85	543	1693	2236	54	169	224
Totale	8929	14532	23461	893	1453	2346



Esempio: in uno studio epidemiologico sul tumore polmonare si desidera che maschi e femmine siano rappresentati con la stessa numerosità.

- La frequenza relativa nella popolazione dei casi di tumore polmonare è di 10 uomini : 1 donna.

- Con un campione casuale semplice ci si aspetta di trovare solo il 10% di donne.

Si procede quindi ad un campionamento stratificato

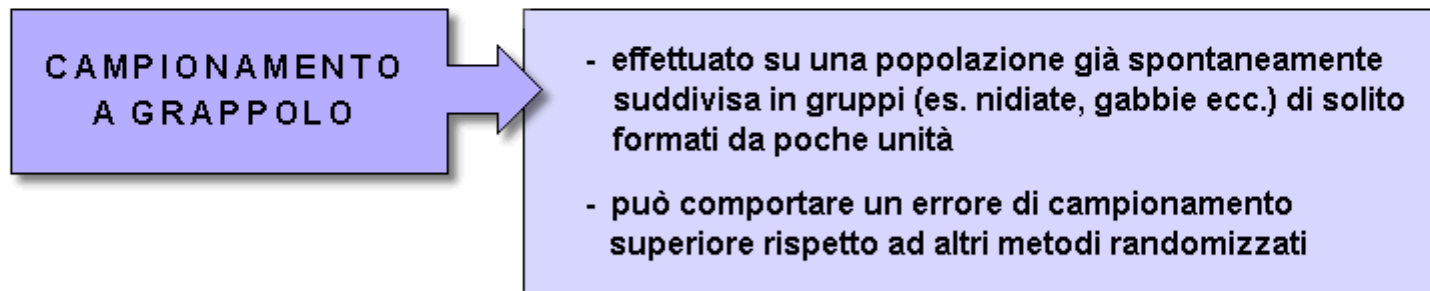
Base di campionamento: i casi di tumore polmonare incidenti (cioè di nuova diagnosi) nella popolazione di Torino negli anni 1993-98

Sono inclusi nel campione 100 uomini e 100 donne.

		N. nella popolazione	N. campione	Frazione di campionamento
Strato 1	Maschi	3355	100	$100/3355=0,0298$
Strato 2	Femmine	847	100	$100/847=0,1181$

CAMPIONAMENTO A CLUSTER

Il campionamento a grappolo (cluster) consiste in un metodo in cui, invece di procedere alla selezione diretta delle singole unità di interesse, **si selezionano dei gruppi (grappoli o cluster) di unità**. Nella pratica, spesso i cluster sono già preformati, e comprendono un numero limitato di unità



CAMPIONAMENTO A CLUSTER

- Può essere preso in considerazione se le unità di studio formano gruppi naturali o se è difficile redigere una lista completa dell'intera popolazione.
- Esso richiede la selezione di un campione casuale di gruppi o cluster e quindi l'inclusione di tutte le unità di studio nei gruppi selezionati
- Il campionamento a cluster è **più conveniente in termini di tempo e costo** rispetto agli altri tipi di campionamento.

CAMPIONAMENTO A CLUSTER

- A volte la popolazione da cui si vuole campionare è aggregata a diversi livelli

Esempio: **comuni, quartieri, scuole, famiglie, etc.**

- Possibilità:

- L'intero cluster è selezionato



Evita che le unità si influenzino tra loro (utile soprattutto per studi sperimentali).

- Le unità vengono selezionate casualmente tra i cluster scelti (in questo caso sono state applicate due campionamenti consecutivi e pertanto è detto «**campionamento a due stadi**»)

- È più comodo del campionamento casuale semplice

CAMPIONAMENTO A CLUSTER

■ L'errore è piccolo se i cluster sono simili tra di loro



■ L'errore può anche essere molto grande se i cluster sono differenti tra di loro



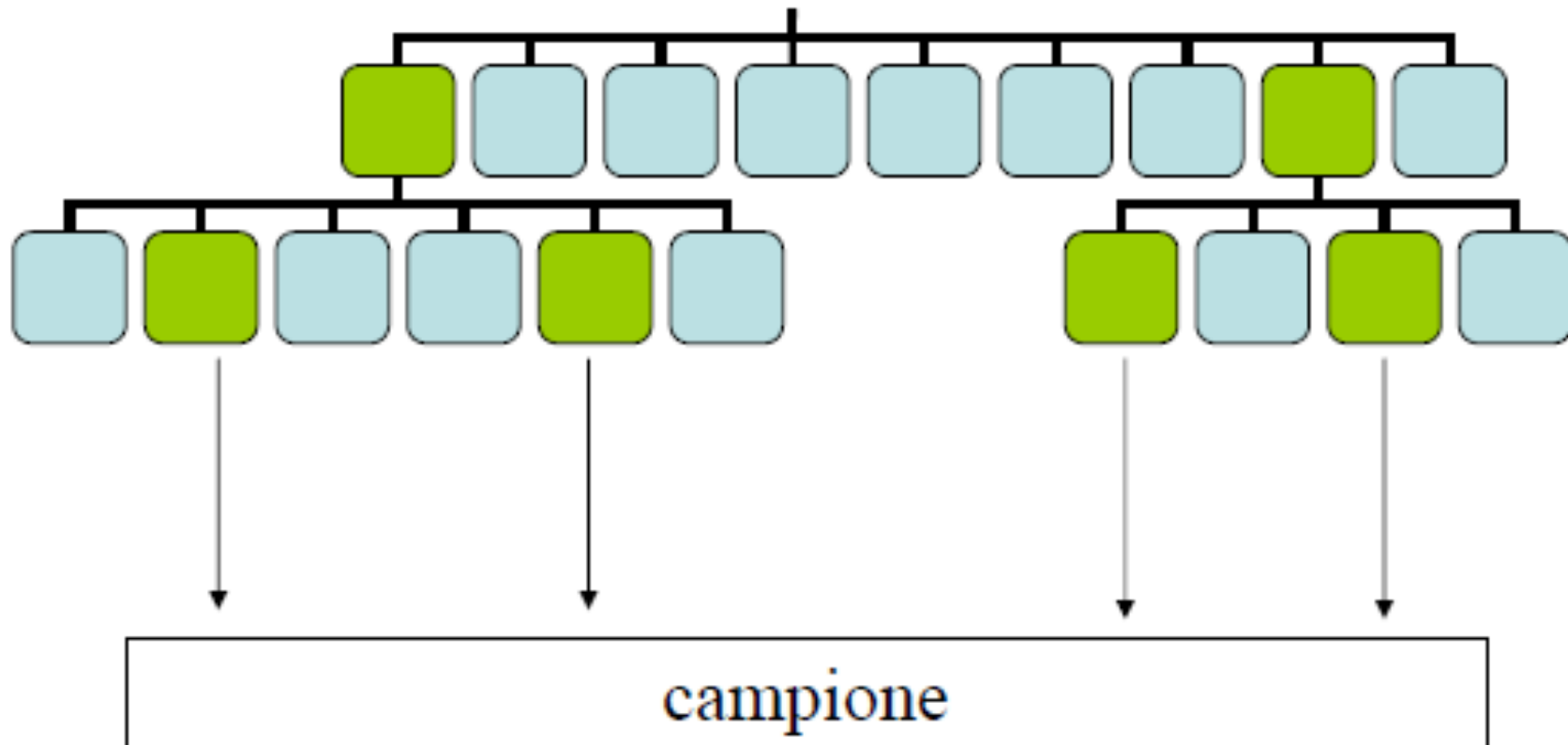
Esempio: voglio verificare l'efficacia di due diversi trattamenti per la disassuefazione dal fumo. Entrambi i trattamenti devono essere proposti dal medico di base.

Procedo in due fasi:

1. campione dei medici (10 medici tra tutti i medici di base di)
2. campione degli assistiti dei medici campionati nella fase 1 (20 assistiti per ciascun medico)

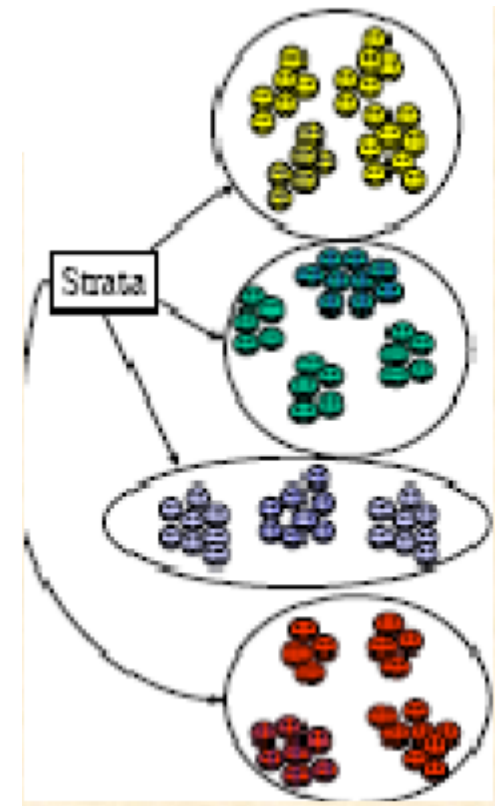
Totale del campione: $10 \text{ medici} \times 20 \text{ assistiti/medico} = 200 \text{ assistiti}$.

SCHEMA DI CAMPIONAMENTO A GRAPPOLO



CAMPIONAMENTO A CLUSTER STRATIFICATO

- L'errore nel campionamento a cluster può essere ridotto creando strati di cluster
- All'interno di ogni strato vengono selezionati uno o più cluster



DIMENSIONI DEL CAMPIONE

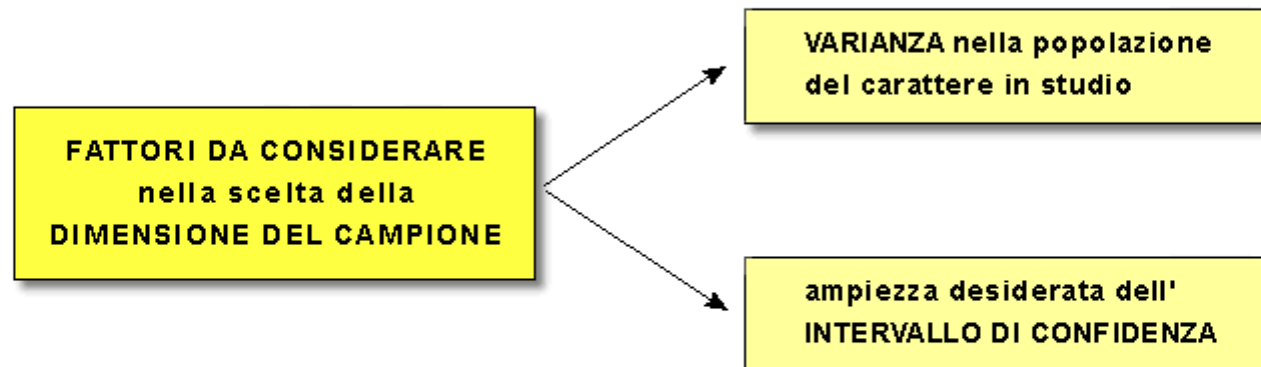
- La dimensione del campione va' tenuta in considerazione al **momento della pianificazione dello studio**
- È essenziale che il **numero delle osservazioni sia sufficiente**
- È intuitivo come la **dimensione del campione aumenti man mano che si richieda una precisione maggiore della stima**

DIMENSIONI DEL CAMPIONE

Il calcolo della dimensione del campione, più propriamente detta **numerosità**, è abbastanza complicato e, soprattutto, richiede la conoscenza di informazioni diverse.

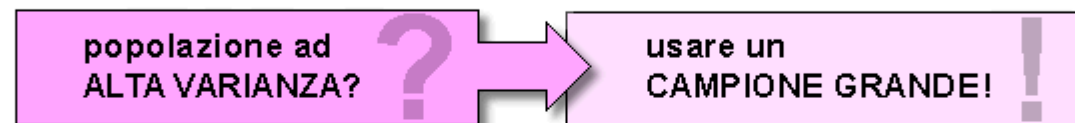
I principali fattori da considerare nell'individuazione della numerosità del campione sono:

- la **varianza**
- l'ampiezza desiderata dell'**intervallo di confidenza**.



DIMENSIONI DEL CAMPIONE

- La **varianza** è una misura del grado di variazioni o oscillazioni presenti, relativamente al parametro che vogliamo stimare, nella popolazione
- È intuitivo che la precisione di un campione è maggiore quando la popolazione da cui è stato estratto è tendenzialmente omogenea, mentre è minore quando la popolazione è eterogenea.
- In una popolazione teorica composta da n individui tutti identici fra loro, lo studio di 1 solo individuo è sufficiente per ottenere una indicazione precisa riguardo alla intera popolazione.



DIMENSIONI DEL CAMPIONE

(IC 95% : 0,51 – 0,73)

■ Il **livello di confidenza** esprime il grado di certezza del risultato. Per convenzione si utilizza generalmente il livello di confidenza al 95%; talvolta si impiegano anche i livelli 90% o 99% o 99,9%.

■ L'**intervallo di confidenza** esprime il margine statistico di errore. Rappresenta una misura della bontà di una stima. Con un certo grado di certezza (dipende dal livello di confidenza scelto) sappiamo che i valori ricadono in questo intervallo.

Esempio: con un livello di confidenza 95% siamo sicuri al 95% che il valore vero cade nell'intervallo trovato. Cioè, se ripetessimo lo studio 20 volte, in media sbaglieremmo 1 volta, ma saremmo nel giusto 19 volte (95%).

[home](#) > [utili per lavorare](#)**home page****utili per lavorare**

- introduzione
- ultimi aggiornamenti
- archivio aggiornamenti

banche dati**motori di ricerca****pubblicazioni interattive****newsletter****tool****indicatori****scuole****utili per lavorare****Sample Size Calculator***(traduzione libera a cura della redazione di EpiCentro)*

Il **Sample Size Calculator** è un servizio pubblico realizzato dalla Creative Research Systems. È possibile utilizzare questo strumento per determinare quante persone è necessario intervistare per ottenere risultati che rispecchino con precisione la popolazione bersaglio richiesta. È anche possibile calcolare il livello di precisione di un campione già esistente.

Prima di utilizzare il Sample Size Calculator, può essere utile definire alcuni termini chiave: l'intervallo di confidenza e il livello di confidenza.

- L'**intervallo di confidenza** esprime il margine statistico d'errore. Ad esempio, in un campione il 47% ha risposto "sì" a una certa domanda. Con un intervallo di confidenza 4 (cioè del 4%) la percentuale di persone che risponderebbero "sì", nell'eventualità di un'intervista a tappeto a tutta la popolazione, sarebbe compresa fra il 43% (47%-4%) e il 51% (47%+4%).
- Il **livello di confidenza** esprime il grado di certezza del risultato. Continuando con l'esempio precedente, porre il livello di confidenza al 95% significa che col 95% di probabilità la percentuale di persone che risponderebbero "sì", nell'eventualità di un'intervista a tappeto a tutta la popolazione, sarebbe compresa fra il 43% (47%-4%) e il 51% (47%+4%).

Mettendo insieme l'intervallo e il livello di confidenza, si è sicuri al 95% che la percentuale reale della popolazione che risponderebbe "sì" è compresa tra il 43% e il 51%.

Fattori che determinano gli intervalli di confidenza

DIMENSIONI DEL CAMPIONE

All'atto pratico, la determinazione della numerosità del campione dipende da considerazioni di tipo

- **NON STATISTICO:**
 - le risorse disponibili

- **STATISTICO:**
 - precisione desiderata
 - prevalenza attesa

DIMENSIONI DEL CAMPIONE

- **PRECISIONE DESIDERATA:** viene espressa attraverso l'errore massimo tollerabile e può essere calcolata in termini assoluti o relativi.

- **PREVALENZA ATTESA:** occorre stimare la prevalenza prima di effettuare l'indagine. Quando non ci sono indizi si ipotizza una prevalenza 0,5 (ossia 50%): questo approccio è di tipo conservativo, nel senso che una prevalenza del 50% fa adottare un campione grande.