

IL CAMPIONAMENTO STATISTICO

Il campionamento è uno degli argomenti fondamentali di tutta la ricerca statistico-epidemiologica. Per campione statistico si intende quel gruppo di unità elementari (non necessariamente persone), sottoinsieme particolare della popolazione o universo, individuato in essa in modo da consentire, con un rischio definito di errore, la generalizzazione all'intera popolazione.

S'intende per popolazione l'insieme, finito o illimitato, di tutte le unità elementari a cui viene attribuito una certa caratteristica che le individua come omogenee, in un contesto che può essere sia reale sia soltanto teorico; ad esempio la popolazione di valori di azotemia in soggetti cirrotici, di sesso maschile e di età 40- 45 anni.

La teoria campionaria costituisce parte integrante e propedeutica dell'inferenza statistica, con le derivanti tecniche di campionamento, e ci permette di individuare le unità le cui variabili si vogliono analizzare.

Lo scopo di condurre indagini nell'ambito sanitario è quello di ottenere dati e informazioni sulle modalità di manifestazioni di particolari fenomeni sanitari al fine di attuare sia nuovi piani di intervento sia migliorare o riprogrammare quelli esistenti. Ad esempio esaminando e valutando l'entità e la direzione della domanda sanitaria si può intervenire sull'attività dei servizi a volte per migliorare la qualità, altre volte per eliminare o ridimensionare quei servizi che risultano non fruiti o poco domandati dall'utenza.

L'acquisizione dei dati relativi al servizio sanitario nella sua interezza può, comunque, risultare un'impresa improba oltre che dispendiosa; si ricorre pertanto all'analisi delle variabili ritenute di maggiore interesse solo su alcune delle unità che compongono l'insieme dei servizi. Le conclusioni a cui si arriva attraverso questo processo possono, non con certezza, ma con una certa probabilità, attraverso procedimenti statistici di inferenza, essere attribuite all'intero universo.

I principali vantaggi di un'analisi su dati campionari, che a prima vista potrebbe apparire limitata e non esaustiva, possono essere sintetizzati nei seguenti quattro punti:

- *Costi ridotti*

Se si osservano le manifestazioni di un fenomeno analizzando un sottoinsieme della popolazione i costi complessivi per l'acquisizione dei dati risultano, evidentemente, inferiori rispetto a quelli che si sosterebbero se si effettuasse il censimento di tutte le unità della popolazione. Di fatto, oggi, il ricorso ai censimenti di tutta la popolazione italiana viene fatto quasi solamente dall'ISTAT, ogni 10 anni, per ottenere un quadro delle principali caratteristiche socio-economico-demografiche dell'intera popolazione. Tutte le altre indagini in ambito sociosanitario sono effettuate quasi sempre su campioni di popolazione.

- *Maggiore rapidità di acquisizione dei dati*

I dati e le informazioni che si intendono raccogliere sono più rapidamente accessibili con rilevazioni parziali piuttosto che con quelle totali. La tempestività nel raccogliere i dati risulta di notevole rilevanza quando le informazioni e i risultati sono necessari nel più breve tempo possibile.

- *Maggiore possibilità di azione*

In alcune indagini poi, la raccolta dei dati, va affidata a gruppi di esperti, a persone altamente qualificate e specializzate che sono frequentemente di difficile reperibilità. In questi casi quindi, l'attuazione di un censimento risulta assolutamente impraticabile e pertanto l'indagine campionaria appare come l'unica via per ottenere le informazioni sul fenomeno che s'intende analizzare.

- *Maggiore accuratezza*

In talune situazioni l'analisi risulta più approfondita in presenza di una numerosità limitata. Il campione permette allora lo svolgimento dell'indagine in maniera più accurata di quanto non lo permetterebbe uno studio complessivo di tutte le unità della popolazione in studio.

Le indagini campionarie possono essere classificate in due tipi:

descrittive e analitiche. Le prime mirano semplicemente ad ottenere informazioni su ampi gruppi di unità (esempio: numero di donne, uomini e bambini che ricorrono ai servizi offerti dalle ASL). Le seconde hanno come obiettivo quello di effettuare confronti tra sottogruppi di una popolazione al fine di scoprire eventuali differenze e di verificare alcune ipotesi o formularne delle altre.

Indipendentemente dallo scopo dell'indagine, va sempre elaborato un "piano di campionamento", che costituisce una delle principali fasi di un'indagine campionaria.

Nel piano di campionamento si stabilisce sia il metodo attraverso cui si estraggono gli elementi che entreranno a far parte del campione, sia la dimensione dello stesso.

Il CAMPIONE (C) statistico è una piccola parte rappresentativa di un tutto (popolazione)

- l'analisi del sangue si fa da un C
 - l'analisi dell' acqua si fa da un C
 - l'analisi del latte si fa da un C
 - l'analisi dell' aria si fa da un C
- Questi sono campioni altamente rappresentativi del “tutto”.
- Ciò che interessa al ricercatore non è tanto il C, quanto la popolazione (P) dalla quale si può considerare estratto il C.
- Le rilevazioni statistiche complete, riguardanti tutti gli elementi di una P, sono sempre più rare; quelle parziali, relative solo ad una parte di casi, si diffondono sempre più, particolarmente nelle Scienze Biomediche.

IL PIANO DI CAMPIONAMENTO E' L'INSIEME DI OPERAZIONI CHE PORTANO A:

- DEFINIRE LA POPOLAZIONE OBIETTIVO NELLA RILEVAZIONE STATISTICA;
- DEFINIRE LE UNITA' CAMPIONARIE;
- SCEGLERE IL O I CAMPIONI SUI QUALI AVVERRA' LA RILEVAZIONE STATISTICA;
- CALCOLARE LA DIMENSIONE DI C;
- DEFINIRE IL METODO O PROCEDIMENTO DI CAMPIONAMENTO;
- FORMULARE, INFINE, UN GIUDIZIO SULLA "BONTA'" DEL CAMPIONE.

DEFINIZIONE DELLA POPOLAZIONE

LA POPOLAZIONE E' DEFINIBILE, IN SENSO GENERALE, COME L'INSIEME DI TUTTE LE UNITA' STATISTICHE CUI COMPETE LA RILEVAZIONE IN OGGETTO.

PER DEFINIRE LA POPOLAZIONE–OBIETTIVO DELLO STUDIO OCCORRE SPECIFICARE:

- LA NATURA DELLE UNITA' STATISTICHE (individui, eventi, ricoverati, deceduti, etc.)
- IL CAMPO DELLA RILEVAZIONE (estensione a particolari categorie o gruppi di soggetti o elementi)

ATTENZIONE PARTICOLARE PER LE:

- POPOLAZIONI “MOBILI”
- UNITA' COLLETTIVE (famiglie, reparti ospedalieri, etc)

METODI DI CAMPIONAMENTO

Nell'ambito del campionamento aleatorio, si distinguono i seguenti metodi.

CAMPIONI PROBABILISTICI



CAMPIONI NON PROBABILISTICI

Non è possibile applicare i precedenti quando:

1. si è costretti a campionare una parte della popolazione che risulta quella realmente accessibile;
2. negli esperimenti di laboratorio su animali in cui il ricercatore frequentemente è costretto ad “usare” quegli animali di cui dispone;
3. in campo sanitario in cui si è costretti a ricorrere a dei campioni ad hoc, i cui elementi sono dei “volontari” o comunque dei soggetti “disponibili” a sottoporsi alla ricerca ad esempio quando si indaga su situazioni delicate;
4. i questionari vengono inviati per posta e la percentuale dei non rispondenti potrebbe comunque alterare il campione;

Nelle sperimentazioni cliniche è possibile introdurre la casualità (randomizzazione) nell'esperimento anziché nei soggetti; (nel confronto di due trattamenti i pazienti selezionati vengono casualmente attribuiti ad uno o all'altro trattamento, in tal caso le implicazioni statistiche (inferenza) saranno applicate ai trattamenti anziché ai soggetti.

CAMPIONAMENTO CASUALE SEMPLICE

(at random – randomizzazione – C. randomizzato)

1. PREPARARE UNA BASE DI SONDAGGIO (SAMPLING FRAME);
2. DETERMINARE LA NUMEROSITA' CAMPIONARIA;
3. SCEGLIERE DELLE UNITA' CAMPIONARIE

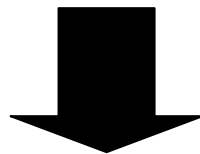


TAVOLA DEI NUMERI CASUALI

46	072	97718	73189	59878	04588	692	66	38434	40272	89771	00686
50	496	39692	24085	71779	15687	662	07	90331	35776	85413	15800
12	600	63927	50359	37540	41359	110	11	41613	65568	75066	27455
48	616	57382	85689	27861	59135	383	60	34711	60409	86253	71462
46	642	76507	83543	17067	24598	993	05	53754	26068	94233	65309
66	476	03290	50341	38126	46627	581	22	05830	88610	37392	49322
10	911	58689	72799	85227	44187	07286		36816	96732	72213	56158
58	369	19998	13711	36703	61578	71593		63712	44037	66954	50747
91	136	89653	40289	30930	80642	32392		14276	79849	44117	93915
66	983	44986	21176	13889	25913	956	01	99590	38581	36491	29925
84	659	22963	46657	93568	45189	159	07	53529	32914	31407	74481
16	370	15896	26540	63456	77955	194	72	90187	18836	38374	50823
51	962	18844	85961	66062	38212	559	51	72211	39045	67240	30163

55289	18202	49671	58143	44912	68204	31391	52319	10633	12146
48023	09540	85628	06281	57154	59317	21691	75261	90525	52591
14798	38093	82618	59617	60396	47863	20462	41026	93710	34491
93722	00368	76492	09912	49084	36237	53900	95067	60697	70823
92215	91249	67934	96279	14949	51163	64807	75102	55473	41117
84980	02710	86329	14559	60912	25090	88126	38326	14818	73918
15786	94713	47704	50971	83736	22432	45223	26815	88408	03901
14510	54630	46853	80698	21089	06777	48303	46988	73035	78366
34300	02413	05981	60951	61023	86349	81066	78329	85623	00742
76204	75471	82181	35454	93086	28888	46459	92850	75617	94005
15161	14287	07003	47560	82243	87339	50047	72822	49783	88463
55376	47645	65456	81222	16625	60816	50228	06328	90450	95496
85252	82677	47132	19295	49170	70543	02715	94290	66588	78247
59173	65864	11685	29750	96562	44030	95630	70747	03598	57338
75755	50084	24656	35590	50263	36487	47167	11955	83545	27189
66704	35388	18479	14849	50527	39960	07032	28009	08718	88005
06825	50142	34241	39373	14363	74470	74684	68779	24274	01953
15205	30117	78412	25253	76498	49688	24256	29259	30967	84152
26513	16234	90329	39268	52305	13938	84976	83552	09715	96443
31216	78892	86748	22159	68678	30969	73137	19875	90005	13434
12738	57513	29679	07678	69125	47869	27664	31192	96509	45292
30264	28122	12021	08704	03626	48134	44675	72004	94533	15530
29137	62351	71997	43811	97560	91770	71973	96267	34254	51623
28166	65149	99684	63318	13423	30061	16421	02000	38288	39696
51309	25148	81989	26320	70166	80126	19348	50146	18261	33036
23212	17788	15066	28825	82313	95166	17890	51844	42524	85515
18162	81375	11775	93908	27783	51448	31913	42555	48292	36972

Si voglia estrarre 10 pazienti (C) da 80 in trattamento (P): dalla tavola di cui sopra si prendono numeri a due cifre <80. Leggendo le prime due cifre della tavola per colonna, le unità da campionare sono quelle a cui corrispondono i numeri evidenziati in VERDE. Oppure quelli in ROSSO etc. Gli elementi di P corrispondenti a tali numeri costituiranno i campioni randomizzati. Questa tecnica assicura la stessa probabilità ad ogni elemento di P di essere estratto e di entrare a far parte del C.

CAMPIONAMENTO SISTEMATICO SEMPLICE

A volte può risultare molto arduo dover numerare, come richiesto dal campionamento casuale semplice, tutti gli elementi della popolazione, specie se questa è molto numerosa. Qualora si disponga di elenchi degli elementi di una popolazione da campionare per individuare le n unità di C tra gli N della P si sceglie dagli elenchi una unità ogni K (N/n) elementi, a partire da una qualunque unità scelta a caso far le prime K .

Questo sistema è molto conveniente quando, come ad esempio, si debba estrarre degenti dagli elenchi relativi ai diversi reparti di un ospedale. Si prelevano quindi da ogni singolo registro le unità sempre ogni K degenti.

Se si debbono estrarre 50 gestanti da 500 che hanno partorito durante l'ultimo anno presso una divisione di Ostetricia e Ginecologia di un ospedale, il rapporto $K=500/50=10$ denota che si deve scegliere, dall'elenco presso il reparto, una donna ogni dieci. La prima donna si individua scegliendo casualmente (es. tramite la tavola dei numeri casuali) un numero compreso tra le prime dieci donne dell'elenco e le altre si sceglieranno, partendo da questa, intervallate di dieci unità.

In definitiva il metodo prevede che:

- o VENGONO SELEZIONATE UNITA' STATISTICHE A INTERVALLI REGOLARI IN UN LISTATO
- o DELLA POPOLAZIONE.

Esempio: volendo estrarre 40 studenti (n) da 200 immatricolati (N) al Corso di Laurea in Medicina e Chirurgia, si calcola $N/n = 200/40 = 5$. Si sceglie uno studente ogni cinque.

N.B. La prima unità tra 1 e 5 si sceglie a bando, poi in progressione aritmetica in ragione di 5. Ciò assicura a tutte le unità la stessa Pro. di far parte di C .

Si hanno tanti C quanti sono i modi di scegliere la prima unità della serie.

Numero di Matricola degli studenti iscritti al I anno.

•	15841		15862
	15842		15863
✚	15843	•	15864
	15844		15865
	15845		15866
	15846		15867
•	15847		15868
	15848	•	15869
✚	15849		15870
	15850	✚	15871
	15851		15872
•	15852		15873
	15853		15874
✚	15854	•	15875
	15855		15876
	15856		15877
•	15857		15878
	15858		15879
✚	15859		15880
	15860		15881
	15861		15882

.....

NECESSITA:

- che l'elenco sia casuale (ad es. nel controllo di laboratorio, negli iscritti a una Facoltà, etc.)
- non elenchi logici (ad es. ordine alfabetico, elenco telefonico)

CAMPIONAMENTO A GRAPPOLI (CLUSTER)

si divide l'universo in gruppi (grappoli).

ES. per le aree geografiche

Mentre nel campionamento stratificato si suddivide la popolazione in sottogruppi detti strati, a volte può essere utile dividerla in un gran numero di sottogruppi detti grappoli (clusters) ed effettuare il campionamento tra i grappoli.

Il metodo quindi non prevede il campionamento diretto degli elementi, ma vengono campionati grappoli di elementi.

Viene spesso fatto ricorso ad un tale campionamento per ridurre il costo della raccolta dei dati, evitando così il costo relativo alla elencazione complessiva di tutti gli elementi della popolazione ed anche le spese di viaggio per gli intervistatori che invece vengono concentrati su un grappolo.

Nel caso più semplice si può usare la scelta casuale per selezionare i grappoli formati e quindi indagare su tutte le unità contenute in quelli campionati. Un tale metodo viene usualmente definito campionamento ad uno stadio, poiché effettuato un'unica volta. Per contro si definisce campionamento a più stadi quello in cui la scelta delle unità campionarie è molto più complessa ed il procedimento campionario interviene più di una volta nella scelta delle unità. La frazione di campionamento deve in questo caso essere calcolata in modo tale che gli elementi della popolazione abbiano la stessa probabilità di essere campionati.

In generale il campionamento a grappoli al vantaggio di una riduzione di costi non associa un'alta precisione, in quanto per ridurre troppo i costi si rischia di non ottenere un campione rappresentativo di tutta la popolazione, ma solo di una parte, quella relativa ai grappoli selezionati.

Per ovviare in parte a ciò si deve cercare di riferirsi a grappoli di numerosità piccola, rispetto alla numerosità totale del campione, per poter ricorrere ad un numero di grappoli piuttosto grande. Ad esempio in una indagine sulla educazione sanitaria nelle scuole di una provincia, non disponendo di un elenco generale di tutti gli alunni delle scuole, si possono estrarre alcune classi (grappoli) di cui si dispone dell'elenco completo e di esse si intervistano poi tutti gli alunni.

Esempio.

Si vuole studiare l'incidenza di reazioni avverse da vaccinazione antimorbillosa, in particolare l'iperpiressia post-vaccinica.

Si stabilisce inoltre di svolgere l'indagine in una grande città.

La tabella che segue è fornita dalla ASL locale e riporta tutti i Medici Pediatri convenzionati con il relativo numero di assistiti.

Stimando in 1000 LA NUMEROSITA' CAMPIONARIA RICHIESTA, e sapendo che circa il 50% dei soggetti in età pediatrica viene sottoposta a vaccinazione, possiamo campionare 7 Pediatri.

Esempio: 3, 9, 11, 16, 24, 26, per un totale di 2675 bambini, tra cui quali identificare quelli vaccinati.

Tabella 3. Numero di bambini assistiti per Pediatria.

Ped.	Ass.	Ped.	Ass.	Ped.	Ass.	Ped.	Ass.	Ped.	Ass.
1	240	7	200	13	344	19	310	25	464
2	441	8	374	14	480	20	290	26	440
3	420	9	333	15	424	21	426	27	287
4	312	10	420	16	380	22	413	28	292
5	346	11	412	17	490	23	398		
6	480	12	307	18	282	24	380		

CAMPIONAMENTO STRATIFICATO

La popolazione viene divisa in “strati” (es. strati di popolosità, età, etc) con conseguente appropriato campionamento.

Si parla di campionamento stratificato proporzionale e di campionamento stratificato “ottimale”.

In quello proporzionale si estrae da ogni strato una certa quantità di unità in proporzione alla numerosità dello strato, tale che $n_1/N_1 = n_2/N_2 = \dots = n_k/N_k = n/N$ e cioè ogni strato contribuisce alla formazione del campione totale nella stessa misura in cui ogni sotto popolazione contribuisce a formare l'intera popolazione. Ad esempio, per accertare l'influenza dell'età sull'incidenza di una certa patologia ed evitare che, mediante un'estrazione casuale semplice, il campione risulti prevalentemente rappresentato o da soggetti giovani o anziani si procede alla stratificazione per età (secondo k classi) della popolazione, poi si calcola una frazione costante per ogni strato.

Nell'ambito di una campagna per la prevenzione delle malattie cardiovascolari è interessante conoscere il livello della colesterolemia in una popolazione al fine di indirizzare verso un più corretto comportamento alimentare. Si concorda di eseguire l'indagine solo sui soggetti di sesso maschile e di età compresa tra i 45 e i 65 anni, cioè la popolazione più a rischio.

Tenendo sempre presente lo scopo che ci si prefigge, cioè quello di migliorare l'efficienza del piano di campionamento, può risultare opportuno ricorrere ad un campionamento stratificato “ottimale”, cioè usare frazioni di campionamento diverse per ogni singolo strato. In alcune situazioni un tal metodo offre maggiore precisione rispetto al campione proporzionale, così quando la variabilità interna degli strati è elevata è più conveniente ricorrere ad un campione ottimale. Allora la frazione di campionamento varierà da strato a strato, sia in funzione della caratteristica che si intende studiare che della variabilità dello strato, applicando quindi il criterio della “ottima ripartizione”.

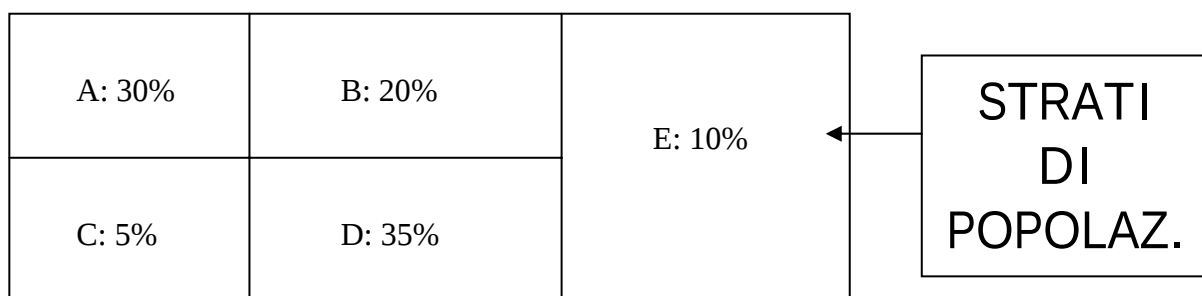
In seguito alla recrudescenza dell'infezione tubercolare registrata in Italia in questi ultimi anni, una ASL vuole studiare la situazione sul proprio territorio attraverso un'indagine campionaria. Si vuole stimare la prevalenza della reattività alla tubercolina con particolare riguardo ai giovani, indicatori più fedeli di una recrudescenza infettiva. La Tabella 1 mostra la composizione per classi di età della popolazione in studio.

Procediamo quindi ad un campionamento stratificato ottimale, con frazione di campionamento maggiore per le classi di età più basse e precisamente: 5% per la prima classe, 4% per la seconda, 3% per la terza, 2% per la quarta e 1% per la quinta.

Tabella 1.

Classi età	N Popolazione	n Campione
<15	27315	5% = 1366
15-24	18751	4% = 750
25-39	24890	3% = 747
40-64	36847	2% = 737
>65	16298	1% = 163
Tot.	124101	3763

La tabella seguente schematizza la composizione percentuale degli abitanti di sesso maschile per Comune di residenza.



Si decide di campionare 500 maschi nei 5 Comuni in proporzione alla percentuale di residenti. Quindi si campionano:

150 dal Comune A;
 100 dal Comune B;
 25 dal Comune C;
 175 dal Comune D;
 50 dal Comune E.

Si vuole campionare 600 donne da una popolazione così strutturata per età:

20-36 anni=30%
 37-65 anni=65%
 >65 anni=5%

Il campione sarà così composto:

180 donne di età 20-36 anni
 390 donne di età 37-65 anni
 30 donne di età >65 anni

CAMPIONAMENTO MULTIFASICO

E' una tecnica di campionamento che risulta molto vantaggiosa quando la popolazione da studiare è molto numerosa e gli elementi possono essere raggruppati in diversi sottoinsiemi. Essa consiste in una prima fase caratterizzata da uno o più campionamenti a grappolo, seguita da una seconda fase in cui, all'interno dei "clusters" selezionati, si procede al campionamento delle unità seguendo una delle metodiche precedentemente enunciate (es. campionamento casuale semplice e campionamento stratificato).

Questa tecnica viene spesso utilizzata nei sondaggi di opinione, allorquando si procede al successivo campionamento delle città, quindi dei rioni ed in ultimo dei soggetti da intervistare. Il campionamento viene definito a n stadi in riferimento al numero dei campionamenti successivi (nel caso città-rioni-persone si parla di campionamento a 3 stadi).

Selezione di campione multifasico

L'Assessorato alla Sanità di una Regione è interessato ad indagare sui comportamenti sessuali di ragazzi e ragazze iscritti alle scuole medie superiori. Si decide di condurre lo studio attraverso la compilazione di un questionario e un colloquio con uno psicologo.

Nella seguente Tabella 1 è riportata la distribuzione per tipo di istituto delle 133 scuole statali e private della Regione.

Tabella 1. Distribuzione delle scuole di una Regione.

SCUOLE	N
Liceo Classico (LC)	9
Liceo Scientifico (LS)	18
Istituto Tecnico Commerciale (RAG)	21
Istituto Tecnico per Geometri (GEO)	16
Istituto Tecnico Industriale (ITI)	21
Scuola d'Arte	11
Istituto Professionale (PROF)	17
Altre	20
Totale	133

Il ricercatore decide che un campionamento multifasico si adatta a questo tipo di studio.

I Fase: si procede al campionamento delle 133 scuole attraverso l'uso di numeri casuali o con estrazione da bussolotti; vengono campionate 10 scuole, ad esempio: 1 LC, 2 LS, 2 RAG, 1 GEO, 3 ITI e 1 PROF.

II Fase: si procede al campionamento dei ragazzi all'interno delle 10 scuole scelte e poi per ognuna di queste ad un campionamento sistematico fino al raggiungimento del numero prestabilito.

CAMPIONI NON PROBABILISTICI

In molte indagini epidemiologiche risulta impossibile o impraticabile ricorrere al campionamento casuale in quanto non risultano valide alcune implicazioni procedurali statistiche. Di fatto si è costretti a campionare una parte della popolazione che risulta quella realmente accessibile. Negli esperimenti di laboratorio sugli animali ad esempio il ricercatore frequentemente è costretto ad “usare” quegli animali di cui dispone; altrimenti se dovesse selezionare casualmente le cavie, difficilmente riuscirebbe a fare la ricerca.

In molti progetti di ricerca in campo sanitario si è costretti a ricorrere a dei campioni ad hoc, i cui elementi sono dei “volontari” o comunque soggetti “disponibili” a sottoporsi alla ricerca. Ciò accade per esempio quando si indaga su situazioni delicate, oppure quando questionari vengono inviati per posta e la percentuale dei non rispondenti potrebbe comunque alterare il campione.

In altre situazioni è possibile introdurre la casualità (randomizzazione) nell'esperimento anziché nei soggetti; nel confronto di due trattamenti ad esempio i pazienti selezionati vengono casualmente attribuiti ad uno o all'altro trattamento; in tal caso le implicazioni statistiche (inferenza), saranno applicate ai trattamenti anziché ai soggetti.

Sotto le giuste condizioni quindi i campioni non probabilistici possono dare utili risultati. Essi però non sono trattabili con la teoria campionaria perché non si basano su alcun principio di casualità e pertanto la loro validità è strettamente legata alla situazione cui si riferiscono e non si ha nessuna garanzia della loro validità in circostanze diverse.

Questi metodi “non corretti” sono meno impegnativi da applicare e, sebbene l'accuratezza è discutibile, trovano frequentemente pratica applicazione. Sono i metodi molto usati da organizzazioni per le ricerche di mercato, per sondare gli orientamenti elettorali nella popolazione ecc.

Un esempio in ambito sanitario può essere il seguente. Si vuole valutare l'efficacia di un nuovo vaccino contro l'AIDS. Tutti i soggetti a rischio che volontariamente si presentano ai centri sieroprofilattici e che risultano sieronegativi vengono sottoposti alla vaccinazione. Avendo stimato in 1000 la numerosità campionaria richiesta, si procede al reclutamento di tutti i soggetti in campo nazionale fino al raggiungimento del numero previsto.

Illustrate le principali tecniche per l'estrazione delle unità che formeranno il campione, bisogna determinare la numerosità da dare allo stesso.

CRITERI DI CAMPIONAMENTO PROBABILISTICO

PANEL:

campione continuativo (o permanente) di unità statistiche contattate in successivi periodi di tempo (a meno di sostituzioni dovute ad esigenze tecniche)

Scopo: studio delle dinamiche e delle determinanti dei comportamenti

Inconvenienti:

- difficoltà di costituzione iniziale
- mortalità spontanea
- effetto panel

PANEL RUOTATO:

campione continuativo con sostituzione a rotazione delle unità statistiche sotto osservazione

Scopo: studio dei flussi o delle transizioni (cambiamenti di stato)

Vantaggi:

- migliora la rappresentatività del campione
- riduce il carico di lavoro dei rispondenti
- aumenta l'efficienza delle stime

APPROFONDIMENTI SUL CAMPIONAMENTO

STRATIFICATO

CAMP. STRATIFICATO CON ALLOCAZIONE PROPORZIONALE (o A FRAZIONE DI SONDAGGIO COSTANTE, o AUTOPONDERANTE):

vi si ricorre quando le deviazioni standard degli strati sono pressoché costanti; in tal caso

$$\frac{n_h}{n} \propto \frac{N_h}{N}$$

CAMP. STRATIFICATO CON ALLOCAZIONE OTTIMA (o A FRAZIONE DI SONDAGGIO VARIABILE):

vi si ricorre quando le deviazioni standard degli strati sono notevolmente diverse tra loro; in tal caso

$$\frac{n_h}{n} \propto \sqrt{Var_h}$$

n = numerosità campionaria

n_h = numerosità campionaria nello strato h -esimo

N = numerosità della popolazione

N_h = numerosità della popolazione nello strato h -esimo

Var_h = varianza della variabile in studio misurata all'interno dello strato h

CRITERI DI CAMPIONAMENTO NON PROBABILISTICO

CAMPIONAMENTO RAGIONATO:

presuppone un certo grado di informazione sulla popolazione; va preferito quando si può disporre di poche unità campionarie, ma può attuarsi solo per caratteri dei quali sono già note alcune tendenze distributive; utile quando si vogliono informazioni in tempi brevi, in particolare su elementi “tipici” o “estremi”.

CAMPIONAMENTO PER QUOTE:

le unità campionarie vengono individuate rispettando la frequenza delle diverse tipologie di individui nella popolazione;
frequentemente utilizzato nei sondaggi e nelle indagini di mercato perché presenta un'organizzazione snella, ha tempi rapidi di realizzazione e costi contenuti.

CRITERI DI CAMPIONAMENTO:

1) CAMPIONAMENTO CASUALE SEMPLICE, AT RANDOM – RANDOMIZZAZIONE.

MINIMIZZA L'ERRORE SISTEMATICO (BIAS)
DA' LA STESSA PROB. AD OGNI COMPONENTE
DELLA P DI ENTRARE A FAR PARTE DEL C
CAMPIONI PROBABILISTICI: per ogni unità
appartenenti a $P = \text{Prob.}$

ES. uso delle tab. NUMERI CASUALI

2) CAMPIONAMENTO SISTEMATICO

Si estraggono unità da una lista a intervalli regolari nella stessa. La prima unità va scelta casualmente poi si sceglie ad es. 1 per ogni 10 unità.

3) CAMPIONAMENTO CASUALE STRATIFICATO

si divide la popolazione in strati o sottogruppi nei quali si scelgono casualmente le unità.

4) CAMPIONAMENTO A GRAPPOLI (CLUSTER)

si divide l'universo in grossi gruppi (grappoli).

ES. per le aree geografiche

5) CAMPIONAMENTO RAGIONATO: presupposto che si conosce abbastanza P si costruisce poi C con ragionamento: mancando la casualità è rischioso fare inferenza su P.

CAMPIONAMENTO STRATIFICATO

SI UTILIZZA QUANDO LA POPOLAZIONE E' ARTICOLATA IN SOTTOGRUPPI, O **STRATI**, CHE DIFFERISCONO RISPETTO ALL'OGGETTO DI RILEVAZIONE E SONO ESSI STESSI DI INTERESSE.

ESEMPIO CORRENTE E' LA STRATIFICAZIONE PER ETA', SESSO, RESIDENZA, ECC. IN INDAGINI EPIDEMIOLOGICHE:

- DA OGNI STRATO SI ESTRAE UN SOTTOCAMPIONE CASUALE SEMPLICE;
- LA NUMEROSITA' DEI SOTTOCAMPIONI (FRAZIONE DI CAMPIONAMENTO) PUO' ESSERE CALCOLATA CON DUE DIVERSI PRINCIPI:
 - a) VIENE ESTRATTA LA STESSA PROPORZIONE DI INDIVIDUI PER CIASCUNO STRATO;
 - b) VIENE ESTRATTA UNA PROPORZIONE DI INDIVIDUI IN FUNZIONE DIRETTA DELLA VARIABILITA' DELLO STRATO DELLA POPOLAZIONE (METODO PIU' CORRETTO).