UNIVERSITA' DEGLI STUDI DI PERUGIA
DIPARTIMENTO DI FILOSOFIA SCIENZE SOCIALI UMANE E DELLA FORMAZIONE
Corso di Laurea in Scienze per l'Investigazione e la Sicurezza

9. ELEMENTI DI TECNICA DEI CAMPIONI

Prof. Maurizio Pertichetti

Statistica sociale

9. Elementi di tecnica dei campioni

Come è stato già detto ad inizio di questo corso introducendo gli aspetti metodologici, la statistica può acquisire le informazioni dall'intera **popolazione** (o universo, o collettivo) di riferimento o da un sottoinsieme di essa detto **campione**.

Per **popolazione** s'intende l'insieme, finito o illimitato, di tutte le unità statistiche di cui vogliamo indagare una certa caratteristica che le individua come omogenee.

Per **campione statistico** si intende un gruppo di unità statistiche, sottoinsieme opportunamente estratto dall'intera popolazione, dal quale trarre, con margini di errore contenuti, indicazioni sulle caratteristiche della popolazione stessa. Ovvero per determinare le caratteristiche fondamentali di una popolazione statistica non sempre è necessario analizzarla *tutta* , ma può essere sufficiente osservarne solo *una parte ridotta* di essa, per l'appunto un campione statistico, **rappresentativa** di tutta la popolazione, ovvero un limitato numero di unità statistiche che ne riproduca le caratteristiche.

Corrispondentemente, le indagini statistiche possono essere classificate, a seconda dell'estensione, in **indagini totali o censuarie**, quando prendono in considerazione tutte le unità statistiche della popolazione, e in **indagini campionarie**, quando si rivolgono ad una parte di essa.

I censimenti raggiungono, ma non sempre, tutta la popolazione ed è evidente che, se questa è quantitativamente molto numerosa e dispersa sul territorio, la loro organizzazione, il loro svolgimento, nonchè l'elaborazione dei dati acquisiti si traducono in fasi che richiedono tempi molto lunghi, materializzando il limite di arrivare a conoscere i risultati della ricerca quando magari la realtà rilevata si è venuta già a modificare. Per non dire poi dei costi elevati. Il ricorso ai campioni consente invece di ridurre questi problemi, sebbene il rischio casuale tipico del metodo induttivo sia quello di pervenire a stime dei parametri inficiate da margini di errori, per quanto definiti con livelli di probabilità.

L'errore di campionamento è ciò che costituisce la differenza tra i valori (le stime) ottenuti con il campione e quelli reali della popolazione e non può mai essere determinato con esattezza, in quanto la vera caratteristica della popolazione è ignota. Della stima, che essendo una valutazione approssimativa non può essere esatta, se ne può tuttavia stabilire la variabilità, ossia i limiti probabili di oscillazione.

Il campionamento è uno degli argomenti fondamentali della ricerca statistica. E' sostanzialmente l'ambito operativo della statistica inferenziale che è volta all'induzione probabilistica delle caratteristiche incognite di una popolazione, ovvero si occupa di risolvere il cosiddetto problema inverso, ossia, sulla base di osservazioni svolte su un campione di unità rappresentative di tutta la popolazione e selezionate con date procedure, perviene a conclusioni che possono essere generalizzate (inferenza), entro dati livelli di probabilità di errore, all'intera stessa popolazione. Alla base della statistica inferenziale vi sono la teoria del calcolo delle probabilità e la teoria dei campioni.

Come detto il corso limita il proprio interesse al filone della statistica descrittiva, tuttavia trattare brevemente aspetti riferiti alla tecnica campionaria deve ritenersi utile come notazione indispensabile per una comprensione più esauriente degli ambiti operativi della statistica.

Lo studio e l'analisi di un campione serve dunque a risalire alle caratteristiche della popolazione cui si riferisce, attraverso la stima dei parametri, cioè dei valori caratteristici (statistici) assunti dalle variabili nell'intera popolazione. Serve, cioè, a dare valori approssimativi della popolazione sulla base dei parametri del campione:

	Popolazione (N)	Campione (n)
Media	μ	$\overline{X}$
Deviazione	σ	s
Varianza	σ^2	s^2

Affidabilità: data la media μ (quella vera) della popolazione e la media $\overline{X}$ proveniente dal campione, la differenza tra queste due medie (supposte non uguali) prende il nome di **errore di campionamento**, che è una misura di affidabilità del campione.

Efficienza: è legata al costo, un campionamento è più efficiente di un altro se, a parità di affidabilità, è meno costoso.

Il procedimento dell'inferenza statistica conduce a risultati "esatti" solo se il campione è perfettamente **rappresentativo della popolazione**. Questa rappresentatività è garantita dalla condizione di **casualità** della selezione delle unità della popolazione che faranno parte del campione. Condizione, questa della casualità, che a sua volta si realizza quando tutte le unità della popolazione hanno la stessa probabilità di essere estratte ed incluse nel campione.

Per illustrare il **concetto di casualità** si può ricorrere all'immagine di un'urna dalla quale vengono estratte delle palline. A questa si può aggiungere l'immagine dell'estrazione dei numeri del lotto, per sottolineare come le palline non siano riconoscibili da parte di chi le estrae nel momento in cui le estrae.

Tali immagini rendono evidente il requisito che tutte le unità della popolazione campionata dovrebbero avere ad ogni ciclo di estrazione, ovvero quello della stessa probabilità di essere estratte, così come avviene per le palline dell'urna e i numeri del lotto. Il campione casuale richiede che la probabilità di estrazione sia conosciuta e non nulla.

La teoria statistica fa notare che, se si scelgono gli elementi di un campione in modo casuale (il che equivale a estrarre le palline da un'urna) non solo ogni elemento ma anche ogni combinazione di elementi (di uguale numerosità) ha la stessa probabilità di essere scelta.

Sebbene il termine nella letteratura scientifica è ancora oggi abbastanza vago, ampiamente soggettivo e altrettanto ampiamente discutibile, per **rappresentatività** solitamente si indica l'esistenza di un rapporto proporzionale fra le distribuzioni di uno o di alcuni caratteri, oggetto di studio, nel campione e nella popolazione. Un campione è rappresentativo dell'universo di cui fa parte se ne riproduce, in piccolo, le caratteristiche, con scarti «non significativi» imputabili al «caso».

La teoria campionaria costituisce parte integrante e propedeutica dell'inferenza statistica.

I principali vantaggi di un'analisi su dati campionari, che a prima vista potrebbe apparire limitata e non esaustiva, possono essere sintetizzati fondamentalmente nei seguenti tre 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 sosterrebbero se si effettuasse il censimento di tutte le unità della popolazione. Oggi, il ricorso ai censimenti viene fatto soltanto dall'ISTAT, ogni 10 anni, per ottenere un quadro delle principali caratteristiche socio-economico-demografiche dell'intera popolazione italiana, mentre tutte le altre indagini le svolge quasi sempre su campioni di popolazione. Peraltro va detto che il ricorso alla procedura di campionamento è necessario ogniqualvolta la popolazione di riferimento non è fisicamente raggiungibile nella sua totalità.

- **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 accuratezza**. In presenza di una numerosità limitata l'analisi risulta più approfondita. 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 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 obbiettivo 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 le unità statistiche che entreranno a far parte del campione, sia la dimensione dello stesso.

Il piano di campionamento è l'insieme delle operazioni che portano a definire:
- la popolazione obiettivo della rilevazione statistica;
- le unità campionarie;
- l'ampiezza del campione, cioè la sua numerosità, ovvero il numero di unità di cui deve essere composto;
- il metodo o procedimento di campionamento;

Ampiezza del campione

La scelta della dimensione del campione dipende da 3 elementi fondamentali:
- **la variabilità tra gli elementi della popolazione**. Una popolazione con una variabilità (attitudine di un carattere ad assumere modalità diverse) maggiore richiede un campione più grande, mentre una maggiore omogeneità richiede un campione più piccolo. Esempio estremo: in una popolazione di tutti uguali basterebbe un sola persona a rappresentarla;
- **il livello di precisione** che si vuole raggiungere. Più grande è la precisione richiesta, maggiore dovrà essere la numerosità del campione. La precisione, però, non cresce allo stesso modo (uniforme) con cui cresce il campione, anzi raggiunta un certa dimensione del campione, la precisione aumenta in modo quasi impercettibile che rende inutile aggiungere altre unità.
- **le risorse economiche disponibili e i tempi di esecuzione dati** per lo svolgimento dell'indagine.

Liste di campionamento

Sono gli elenchi che contengono tutti i componenti della popolazione che intendiamo studiare, i quali componenti vi devono comparire una sola volta, affinché ognuno abbia la stessa probabilità degli altri di essere selezionato.

In genere queste liste non sono sempre puntualmente aggiornate specie se riguardano una popolazione molto estesa. Esempi di liste: ANAGRAFE, LISTE ELETTORALI, ELENCO DEL TELEFONO. L'iter migliore prevedrebbe la costruzione di una lista ad hoc per l'indagine da svolgere, cosa che è possibile fare se l'indagine non è molto estesa.

Metodi di campionamento

Vi sono fondamentalmente due tipi di campioni:
- I **campioni probabilistici**, caratterizzati dall'elemento qualificante della **casualità**, dove ciascuna unità della popolazione ha la stessa probabilità **nota** e **diversa da zero** di entrare a far parte del campione. Consentono l'inferenza, ossia la generalizzazione dei risultati a tutta la popolazione.
- I **campioni non probabilistici**, per i quali invece la casualità manca. Le unità vengono scelte in maniera arbitraria o di comodo e la possibilità che ciascuna di esse ha di essere estratta è **non nota**. Non consentono l'inferenza, per cui i risultati hanno validità solo per il campione, che si definisce in tal caso **distorto** Forniscono dati non affidabili e non consentono di calcolare la precisione delle stime.

In altre parole, nel campionamento probabilistico vale la condizione di casualità nella selezione degli elementi della popolazione che faranno parte del campione, che è una condizione necessaria per poter applicare l'inferenza statistica, cioè generalizzare i risultati ottenuti dal campione alla popolazione.

I campioni non probabilistici riflettono, invece, nel bene e nel male l'orientamento di colui che li forma e non consentono di valutare il grado dell'errore che si può commettere perché ad essi non è applicabile la teoria del calcolo delle probabilità.

Ad esempio, se, utilizzando il campione non probabilistico, l'obiettivo è quello di sondare le modalità di utilizzo del cellulare, dopo aver deciso che si vogliono intervistare 900 persone, divisi per genere ed età, in modo da formare sei sottocampioni di numerosità uguale (maschi e femmine tra 15–45 anni, 46-65 anni e oltre 65 anni), si fermano per strada tante persone finché si raggiungono le quote prefissate.

A lavoro svolto avremo certemente ottenuto 900 interviste ma non avremo ragionevolmente un campione rappresentativo di tutta la popolazione in quanto sono rimasti esclusi dalla possibilità di essere intervistati coloro che, nel tempo in cui sono state svolte le interviste, non si sono trovati in quei luoghi. Inoltre la discrezionalità dell'intervistatore nello scegliere le persone può comportare una distorsione nel campione non stimabile.

Riprendendo l'esempio precedente, con il campionamento probabilistico, si procederà invece in maniera diversa. In primo luogo si definiscono le caratteristiche della popolazione con riferimento alla composizione per genere ed età. Si calcolerà quindi la proporzionedi giovani, adulti e anziani tra i maschi e le femmine e si ripartirà il campione con le stesse proporzioni. Così se nella popolazione il 53% sono femmine e il 47% sono maschi, sarà mantenuta la medesima proporzione e avremo 477 femmine (pari al 53% di 900) e 423 maschi (pari al 47% di 900). Allo stesso modo si procederà ripartendo i soggetti, all'interno del genere, in base all'età.

Successivamente ci si procurerà l'elenco nominativo della popolazione di riferimento e si procederà all'estrazione dei nomi delle persone da intervistare (è buona norma formare anche un elenco di riserva qualora i primi non fossero raggiungibili o disponibili a farsi intervistare). Il campione così costruito avrà le caratteristiche di un campione statisticamente rappresentativo della popolazione da cui è stato tratto e consentirà di estendere i risultati dell'indagine alla popolazione da cui proviene.

L'utilizzo di un campione probabilistico comporta tuttavia problemi legati alla sua costruzione in quanto non sempre è possibile disporre (per ragioni di privacy o di mancanza di dati individuali) degli elenchi nominativi da cui estrarre quelli da campionare.

Sono campionamenti probabilistici:

- il campionamento casuale semplice;
- il campionamento stratificato;
- il campionamento a più stadi;
- il campionamento a grappoli.

Sono campionamenti non probabilistici:

- il campionamento a scelta ragionata;
- il campionamento per quote;
- il campionamento di convenienza;
- il campione di esperti;
- il campionamento telefonico.

Il campionamento casuale semplice

E' la procedura di scelta casuale più semplice. Con il campionamento casuale semplice ciascuna unità della popolazione ha la stessa probabilità di entrare a far parte del campione. Se l'aggregato è concreto, ovvero la popolazione è costituita da un numero finito di unità tutte individuabili, la procedura consiste nell'assegnare un numero d'ordine progressivo distinto ad ognuna delle unità costituenti il colletivo statistico, ed impostando una corrispondenza biunivoca con palline aventi numerosità pari a quella del collettivo. Tali palline sono inserite in un urna dalla quale si estrae il predeterminato numero di unità. Quì soccorre la tavola dei numeri casuali o semplici applicativi di pc, vedi Excel, in grado di generare numeri casuali.

Con riferimento all'importanza data all'ordine di estrazione si hanno due tipologie di campioni:

- **Campioni ordinati**, sono quelli che, pur essendo eventualmente costituiti da identici elementi, differiscono tra loro per l'ordine con cui sono disposti gli elementi stessi;
- **Campioni non ordinati**, sono quelli che, pur presentando uguale numerosità, differiscono tra loro per qualche elemento.

In questa procedura di selezione casuale si distinguono due modalità di estrazione dei campioni: con ripetizione (o bernulliani) e senza ripetizione (o in blocco); a seconda che vi sia reimmissione o no delle unità estratte.

L'insieme dei campioni di prefissata numerosità che si possono estrarre da una data popolazione tramite un procedimento casuale è denominato universo dei campioni. Il numero di elementi di tale insieme varia a seconda dell'ampiezza **N** della popolazione oggetto di osservazione, dell'ampiezza (o numerosità) **n** dei campioni, dell'importanza data all'ordine di estrazione e, infine, della modalità di estrazione.

Ricorrendo al calcolo combinatorio, pertanto si ha:

- Campionamento ordinato senza ripetizione. Il numero dei campioni costituenti l'universo è pari al numero di disposizioni senza ripetizioni di N elementi presi a n a n. *Ciascuna disposizione differisce dalle altre o per gli oggetti o per il loro ordine* , ossia:

$$D_{N,n} = \quad N*(N-1)*(N-2)*....*(N-n+1) = \quad \frac{N\,!}{(N-n)\,!}$$

- Campionamento ordinato con ripetizione. Il numero dei campioni costituenti l'universo è pari al numero di disposizioni con ripetizioni di N elementi presi a n a n. *Ciascun oggetto può essere ripetuto più volte (fino ad un massimo di n volte) e ciascuna disposizione differisce dalle altre o per gli oggetti o per il loro ordine* , ossia:

$$_rD_{N,n} = \quad N*N*....*N = N^n$$

- Campionamento non ordinato senza ripetizione. Il numero dei campioni costituenti l'universo è pari al numero di combinazioni senza ripetizioni di N elementi presi a n a n. *Ciascuna combinazione differisce dalle altre soltanto per gli oggetti e non per il loro ordine* , ossia:

$$C_{N,n} = \quad \binom{N}{n} \quad = \quad \frac{N\,!}{n!*(N-n)!}$$

- Campionamento non ordinato con ripetizione. Il numero dei campioni costituenti l'universo è pari al numero di combinazioni con ripetizioni di N elementi presi a n a n. *Ciascun oggetto può essere ripetuto più volte (fino ad un massimo di n volte) e ciascuna combinazione differisce dalle altre soltanto per gli oggetti e non per il loro ordine* , ossia:

$$_rC_{N,n} = \quad \binom{N+n-1}{n} \quad = \quad \frac{(N+n-1)\,!}{n!*(N-1)!}$$

Esempio

Sia data una popolazione costituita da **N**=4 unità statistiche: **a**, **b**, **c**, **d** . Volendo estrarre da essa campioni, con una procedura di campionamento casuale semplice, di numerosità **n**=2, determinare l'universo dei campioni:

Ordinati estratti senza ripetizione;

a b	b a	c a	d a
a c	b c	c b	d b
a d	b d	b d	d c

$D_{4,2} = 4*3 = 12$

Ciascuna disposizione differisce dalle altre o per gli oggetti o per il loro ordine

Odinati estratti con ripetizione;

a a	b a	c a	d a
a b	b b	c b	d b
a c	b c	c c	d c
a d	b d	c d	d d

$_rD_{4,2} = 4*4 = 16$

Ciascun oggetto può essere ripetuto più volte (fino ad un massimo di n volte) e ciascuna disposizione differisce dalle altre o per gli oggetti o per il loro ordine

Non ordinati estratti senza ripetizione;

a b		
a c	b c	
a d	b d	c d

$$C_{4,2} = \quad \binom{4}{2} \quad = \quad \frac{4\,!}{2!*(4-2)!} = \frac{4*3*2}{2!*2!} = \frac{24}{4} = \quad 6$$

Ciascuna combinazione differisce dalle altre soltanto per gli oggetti e non per il loro ordine

Non ordinati estratti con ripetizione;

a a			
a b	b b		
a c	b c	c c	
a d	b d	c d	d d

$$_rC_{4,2} = \quad \binom{4+2-1}{2} \quad = \quad \frac{5\,!}{2!*(4-1)!} = \frac{5*4*3*2}{2!*3!} = \frac{120}{2*3*2} = \frac{120}{12} = \quad 10$$

Ciascun oggetto può essere ripetuto più volte (fino ad un massimo di n volte) e ciascuna combinazione differisce dalle altre soltanto per gli oggetti e non per il loro ordine

Il 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 delle unità di una popolazione da campionare si può procedere come segue: si calcola *l'intervallo di campionamento* (o *passo di estrazione*)

$$k = \frac{N}{n}$$

e si arrotonda k all'intero più vicino.

Si individua a caso un numero **r** compreso fra 1 e k e si procede scegliendo le unità che corrispondono alle posizioni della lista:

$$r, r+k, r+2k, \ldots\ldots, r+(n-1)k$$

Il numero **r** identifica la prima unità, dopodiché se ne estraggono sequenzialmente una ogni k.
Se il modo in cui le unità sono elencate nella lista può considerarsi casuale, il campionamento sistematico può considerarsi a tutti gli effetti analogo al campionamento casuale semplice. E' il metodo utilizzato dall'ISTAT per estrarre dalle liste anagrafiche.

Esempio
Volendo estrarre 25 pazienti (n) da 250 ricoverati (N) in un dato ospedale, si calcola N/n = 250/25 = 10. Si sceglie un paziente ogni dieci. La prima unità tra 1 e 10 si sceglie a caso, poi in progressione aritmetica in ragione di 10. Ciò assicura a tutte le unità la stessa probabilità di far parte del campione. E' evidente che si hanno tanti campioni quanti sono i modi di scegliere la prima unità della serie.

Il rapporto tra dimensione del campione /dimensione della popolazione è invece detto *ragione di campionamento* e rappresenta la proporzione di elementi della popolazione selezionati per il campione, nel caso esposto n/N = 1/10.

Il campionamento stratificato

E' una procedura di campionamento utilizzata quando si studia un carattere influenzato da un fattore presente nella popolazione. La popolazione viene suddivisa in un numero determinato di strati o classi il più possibile omogenei al loro interno rispetto al carattere indagato e successivamente si procede all'estrazione di un campione casuale semplice di numerosità prefissata da ciascuno strato.

Nel caso di campionamento stratificato ***proporzionale*** , che si utilizza per avere migliori risultati in termini di rappresentatività, si estrae da ogni strato una certa quantità di unità in proporzione alla numerosità, che deve essere **nota**, dello strato, tale che n1/N1= n2/N2= …= nk/Nk= 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.

Il campionamento a più stadi

E' una procedura di campionamento combinato, nel senso che somma metodi campionari diversi, che prevede l'individuazione di una struttura gerarchica dell'universo che si deve indagare.
Se si vuole, ad esempio, rilevare una qualche caratteristica delle famiglie italiane si può estrarre un campione in cui le unità di primo stadio (o unità primarie) sono i comuni e, successivamente all'interno di questi, selezionare le unità di secondo stadio (o unità secondarie) rappresentate dalle famiglie facendo ricorso agli elenchi anagrafici.

Il campionamento a grappoli

Mentre nel campionamento stratificato si suddivide la popolazione in sottogruppi detti strati, a volte può essere più opportuno dividerla in sottogruppi, detti grappoli o clusters, ed effettuare l'estrazione casuale di questi. Il metodo quindi non prevede il campionamento diretto delle unità, ma quello dei grappoli, per cui a far parte del campione sono le unità appartenenti al grappolo estratto.

In generale il campionamento a grappoli ha il vantaggio di una riduzione dei tempi e dei costi, cui però si associa un'alta imprecisione dei risultati ottenuti, in quanto l'errore campionario può risultare più elevato di quello che si registra con gli altri metodi.

A ciò si ovvia in parte facendo riferimento a grappoli quanto più eterogenei e di numerosità ridotta, rispetto alla numerosità totale del campione, così da disporre di un maggior numero di grappoli. Ad esempio in una indagine che abbia come riferimento gli alunni della scuola media superiore di una provincia, non disponendo di un elenco generale di tutti i frequentanti, si possono estrarre alcune classi (grappoli) di cui si dispone dell'elenco completo e di esse si intervistano poi tutti gli alunni.

Il campionamento a scelta ragionata

E' una procedura per cui sono selezionate quelle unità statistiche che, sulla base di alcune loro caratteristiche o dell'esperienza e del giudizio del ricercatore, meglio rispondono alle finalità dell'indagine. In genere si utilizza quando non è possibile accedere alla lista delle unità della popolazione e l'ampiezza del campione è limitata. Mancando la casualità non permette la valutazione dell'errore campionario e diventa rischioso fare l'inferenza sulla popolazione.

Il campionamento per quote

E' una procedura (simile al campionamento stratificato) che consiste nell'affidare al rilevatore il compito di selezionare, per cui è esclusa la casualità, le unità del campione nel rispetto di quote di popolazione prefissate, di cui si è a conoscenza per esempio dai censimenti, che presentano determinate caratteristiche (età, sesso, ecc.). Si adotta nel caso di indagini su una popolazione distribuita su un territorio molto vasto e per la quale non si possiede una lista completa dei componenti. E' la tecnica di campionamento non probabilistico più utilizzata in particolare nelle indagini di mercato e nei sondaggi di opinioni. E' questa una tecnica che conduce inevitabilmente a diverse distorsioni a causa della libertà di scelta concessa al rilevatore.

Un **parametro**, come già accennato, è un valore numerico di riferimento di una popolazione che misura una caratteristica di essa. Per **distribuzione campionaria** si intende l'insieme di tali valori numerici estratti da campioni di uguale dimensione dalla popolazione. E' evidente che mentre i valori numerici di riferimento di una popolazione risultano unici, ovvero quelli sono e tali rimangono, per quanto non noti, quelli desunti con il campionamento differiscono da campione a campione. Distribuzioni campionarie sono la distribuzione campionaria delle medie, la distribuzione della varianza campionaria, la distribuzione della proporzione campionaria, ecc.

Una delle distribuzioni campionarie più importante è la distribuzione campionaria delle medie. Se in astratto supponessimo di estrarre da una popolazione tutti i possibili campioni casuali di una data numerosità e di calcolare per ciascuno di essi la media

$$\overline{X} = \frac{\Sigma X_i}{n}$$

avremo una distribuzione delle medie dei campioni. Se queste medie le consideriamo come singole osservazioni, otteniamo appunto la **distribuzione della media campionaria** di campioni di quella data numerosità.

Per il teorema del limite centrale si ha che *la media della distribuzione campionaria delle medie è uguale alla media della popolazione:*

$$M\left(\overline{X}\right) = \mu$$

e che nel caso di campioni numerosi e abbastanza ampi si può determinare la probabilità di estrarre campioni i cui risultati, distribuendosi secondo una curva di tipo gaussiana, si dispongono intorno al vero valore medio della popolazione collocandosi entro determinati intervalli.

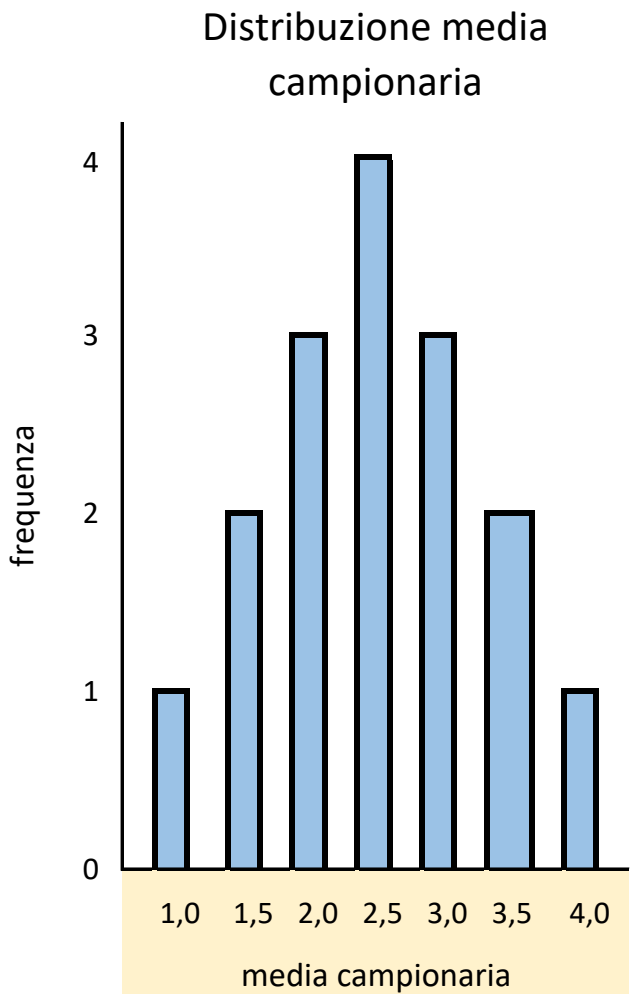
Esempio

Data la popolazione concreta **N** = 1, 2, 3, 4, determiniamo l'universo dei campioni composti di due unità con una procedura di campionamento casuale semplice nella modalità di raggruppamento *ordinati estratti con ripetizione* , cioè *bernoulliana* , e altresì la distribuzione della media campionaria.

Media della popolazione $\mu = \frac{1 + 2 + 3 + 4}{4} = \frac{10}{4} = \mathbf{2,5}$

${}_rD_{4,2} = 4^2$ Numero totale di campioni = 16

numero totale di camp	univ camp di 2 unità		distrib med camp
1	1	1	1,0
2	1	2	1,5
3	1	3	2,0
4	1	4	2,5
5	2	1	1,5
6	2	2	2,0
7	2	3	2,5
8	2	4	3,0
9	3	1	2,0
10	3	2	2,5
11	3	3	3,0
12	3	4	3,5
13	4	1	2,5
14	4	2	3,0
15	4	3	3,5
16	4	4	4,0
	Σ		40,0



$M\left(\overline{X}\right) = 40,0 / 16 = \mathbf{2,5} = \mu$

Ricorrere al campionamento significa utilizzare una procedura per cui ogni campione appartenente all'insieme definito come *universo dei campioni di prefissata numerosità* ha esattamente le stesse probabilità degli altri di essere estratto e che ad essere estratto sarà uno e un solo campione. Riferito all'esempio appena proposto ciò significa, che qualunque estrazione casuale di un campione di due unità dalla popolazione **N** di cui sopra, darà come risultato del sorteggio una soltanto delle 16 disposizioni campionarie.

Nel caso di campioni di numerosità significativa, con **n** > 100, è dimostrato che i valori medi in essi rilevati si ritrovano:

- per il 68,27 % tra il valore vero **μ** della popolazione e ± 1 volta σ_x ;
- per il 95,45 % tra il valore vero **μ** della popolazione e ± 2 volte σ_x ;
- per il 99,73 % tra il valore vero **μ** della popolazione e ± 3 volte σ_x .

σ_x rappresenta l'errore medio di campionamento, ovvero la differenza fra il valore vero della media nella popolazione e quello nel campione, ed è calcolato come $\sigma_x = \sigma / \sqrt{n}$, dove **σ** è lo scarto quadratico medio calcolato nel campione ed **n** la sua numerosità.