

Una definizione introduttiva

La Statistica è una metodologia per l'analisi e lo studio descrittivo quantitativo di insiemi di dati.

- metodologia e non una teoria: lo scopo non è quello di spiegare ma di descrivere, analizzare ed eventualmente fornire elementi per assumere decisioni;
- quantitativa: ciò che utilizza il numero come mezzo di espressione, si studia la realtà attraverso gli aspetti quantitativi - i dati - che essa offre;
- insiemi: l'obiettivo è quello di comprendere le caratteristiche di collettivi; ciascun elemento è depositario di un'informazione che si rapporta all'intera popolazione in esame.

Analisi esplorative e confermative

- *Approccio esplorativo*: consiste nel descrivere i dati in forma sintetica al fine di evidenziare strutture di relazione implicite e permettere di proporre un modello interpretativo;
- *Approccio confermativo*: consiste nella verifica di alcune ipotesi formulate prima della rilevazione dei dati ed analizzate sulla base di metodi statistico-inferenziali.
- Nota. La scelta fra i due approcci dipende dagli obiettivi dell'indagine e dal grado di conoscenza del fenomeno.

Fenomeni Collettivi

Un *collettivo statistico* è formato da un insieme di elementi omogenei rispetto ad una caratteristica comune.

Gli elementi che costituiscono un collettivo statistico si chiamano *unità del collettivo*.

- *Popolazione*: se il collettivo comprende tutte le unità omogenee rispetto ad una caratteristica comune;
- *Campione*: se il collettivo in esame costituisce un sottoinsieme della popolazione di riferimento.

Classificazione dei fenomeni collettivi

- collettività di casi singoli: fenomeni relativi a singole osservazioni di insiemi di elementi
- collettività di osservazioni: fenomeni relativi a più osservazioni di uno stesso elemento.
- collettivi di stato: fenomeni individuabili in maniera esatta solo se si fissa un preciso istante di tempo; esempio: nascite.
- collettivi di movimento: fenomeni individuabili con riferimento ad un periodo di tempo.

Unità statistiche e Popolazione

- L'oggetto dell'osservazione di ogni fenomeno individuale che costituisce il fenomeno collettivo è detto *unità statistica*;
- L'insieme delle unità statistiche costituisce il *collettivo* o *popolazione statistica*;
- Esempi di unità statistiche: individui, famiglie, aziende, etc.
- Nota importante: Prima di qualunque indagine risulta essenziale delimitare con precisione la popolazione che si vuole esaminare;

Unità statistiche e Unità di rilevazione

- L'*unità di rilevazione* a volte può non coincidere con l'unità statistica che forma il collettivo.
- Censimenti demografici:
unità statistica: individuo; *unità di rilevazione*: famiglia;
- Censimenti industriali:
unità statistica: unità locale; *unità di rilevazione*: impresa
- Nota importante: la definizione corretta dell'unità statistiche e dell'unità di rilevazione è condizione fondamentale per l'esattezza della ricerca. L'inclusione o esclusione errata di numerose unità può portare a cambiamenti significativi nella popolazione e quindi a conclusioni fuorvianti.

Classificazione delle unità statistiche

Le unità statistiche si dicono:

- *Semplici* - esempio: individui, pezzi meccanici;
- *composte* se sono aggregati di unità semplici uguali, esempio: famiglie, collegi;
- *complesse* se sono aggregati di unità statistiche di natura diversa, esempio: azienda.

Caratteri e modalità

Il *carattere* di un'unità statistica è una particolare caratteristica degli elementi della popolazione che si manifesta in essi in forme differenti, che vengono chiamate *modalità*.

Requisiti di un carattere in base a cui effettuare un'analisi di dati:

- il carattere deve poter assumere modalità differenti;
- più elementi della popolazione possono avere la stessa modalità del carattere;
- ogni elemento della popolazione deve possedere un'unica modalità del carattere;
- devono esistere almeno due elementi della popolazione per cui il carattere si presenta con modalità differenti.
- Esempio, il personale di un'azienda può essere descritto in base a: sesso, età, anzianità di lavoro, salario, numero di figli a carico, comune di residenza, etc.

Classificazione dei caratteri

● Nomenclatura italiana

1. caratteri *qualitativi*: sconnessi o ordinati (attributi);
2. caratteri *quantitativi* (misure).

● Nomenclatura anglosassone

1. un carattere costituisce una *scala nominale* se per le sue modalità è possibile affermare soltanto se sono uguali o diverse;
2. un carattere costituisce una *scala ordinale* se è possibile soltanto dare un ordine alle sue modalità;
3. un carattere costituisce una *scala a rapporti o proporzionale* se il carattere è (quantitativo) misurabile a partire da uno zero assoluto.

Corrispondenze fra le due nomenclature

carattere qualitativo sconnesso	$\Leftrightarrow$	scala nominale
carattere qualitativo ordinato	$\Leftrightarrow$	scala ordinale
carattere quantitativo	$\Leftrightarrow$	scala a intervalli

Studio di una popolazione 1/2

- *Studiare una popolazione* secondo un dato carattere significa esaminare come si distribuiscono in essa le modalità del carattere in esame.
- I caratteri più semplici sono quelli che comportano solo due modalità, chiamati anche *caratteri dicotomici*:
 1. un salariato può essere di sesso *maschile* o *femminile*;
 2. un individuo può avere *più* o *meno* di 65 anni;
 3. un pezzo meccanico può essere *funzionante* o *difettoso*.

Studio di una popolazione 2/2

- Il numero di modalità di un carattere varia a seconda del grado di dettaglio dell'informazione disponibile. Ad esempio il carattere “tipologia dell'utente Enel” può comportare i casi seguenti:

1. due modalità: *privato, azienda*;
2. tre modalità: *privato, professionista, azienda*;
3. quattro modalità: *privato, professionista, azienda, ente pubblico*;

Rilevazione dei dati

Alla base delle ricerche statistiche, ci sono le informazioni che provengono dall'osservazione e rilevazione di fenomeni individuali. La conoscenza della tecnica di rilevazione è importante per:

- Esigenza Scientifica: l'accettazione della loro validità oggettiva deriva dalla conoscenza della tecnica adoperata per la rilevazione dei dati;
- Addestramento del Personale che dovrà procedere alla rilevazione in maniera adeguata.
- Due fasi della rilevazione:
 - i. Scelta del Piano di Rilevazione,*
 - ii. Raccolta dei Dati.*

Fonti Statistiche

L'analisi dei dati ed i risultati cui si perviene dipendono fortemente dal modo in cui i dati sono stati rilevati.

- ISTAT: Annuario Statistico Italiano, Compendio Statistico Italiano, Annuari Specifici, etc.
- Banca d'Italia, ISCO, CNEL, CENSIS
- Istituti privati: *Banche, Il Sole 24 Ore*, etc.
- *Eurostat, ONU, etc.*
- Istituti di Sondaggi: *Doxa, Abacus, Nielsen*, etc.
- www.istat.it, europa.eu.int/comm/eurostat, www.bancaditalia.it, etc.
- Importante: Problema dell'autorevolezza della fonte.

Piano della Rilevazione

Premessa: Definire con esattezza lo scopo della rilevazione.

- Individuare con precisione la *popolazione*, l'*unità statistica* e l'*unità di rilevazione*;
- Stabilire i *caratteri* quantitativi e/o qualitativi da rilevare per ogni unità;
- Indicare i *mezzi tecnici* per raccogliere le informazioni su tali caratteri;
- Fissare l'*estensione della rilevazione* in ordine al territorio, al periodo, alle disponibilità finanziarie.

Rilevazioni Statistiche

Due tipi di rilevazioni statistiche:

1. *Rilevazioni Totali o Censimenti*

quando vengono osservate tutte le unità della popolazione in merito al fenomeno in esame;

2. *Rilevazioni parziali o Campionarie*

quando viene osservato un sottoinsieme delle unità che costituiscono la popolazione:

- consentono di ridurre sensibilmente il tempo ed il lavoro necessario per la raccolta e la elaborazione dei dati;
- si ottiene una riduzione dei costi;
- problema fondamentale: rappresentatività del campione.

Rilevazioni Campionarie

Due tipi di rilevazioni statistiche:

- Scelta casuale delle unità statistiche
- Scelta ragionata delle unità statistiche.

Se da una popolazione di N elementi si estrae un campione di n elementi, si definiscono le seguenti quantità:

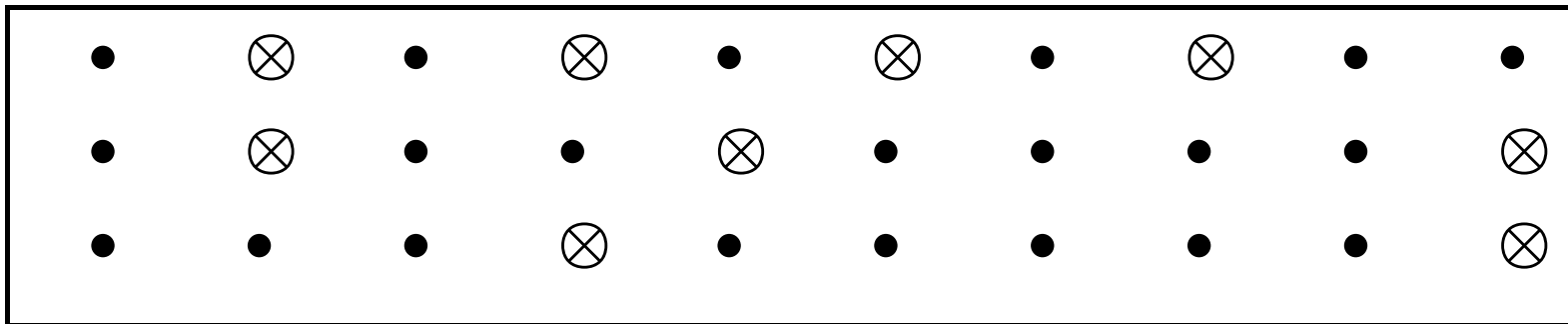
$$\text{Frazione di campionamento} := \frac{n}{N} \cdot 100\%$$

$$\text{Intervallo di campionamento} := \frac{N}{n}$$

Criteri di scelta del campione e le loro proprietà vengono studiati nell'ambito della *teoria dei campioni*.

Campionamento casuale semplice 1/3

Per selezionare un campione casuale, si assegna preliminarmente un numero a ciascun elemento della popolazione; successivamente il campione viene selezionato mediante procedimenti che assicurino l'imprevedibilità di un risultato fra i tanti possibili e garantiscano uguale probabilità per tutte le unità che possono essere scelte.



Campionamento casuale semplice 2/3

Vantaggi

- E' il tipo di campionamento più semplice che consente di fare riferimento ai modelli più elementari di calcolo delle probabilità;
- Richiede una minima conoscenza a priori delle caratteristiche della popolazione;
- Garantisce una scelta obiettiva delle unità da rilevare e tale da escludere qualunque distorsione nei risultati;
- Le stime dei parametri della popolazione e dei loro errori medi sono molto facili;

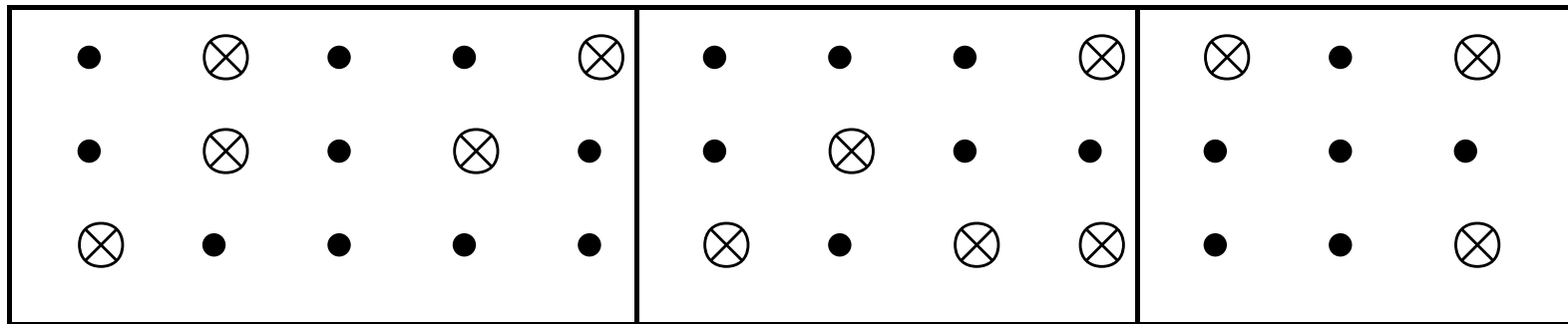
Campionamento casuale semplice 3/3

Svantaggi

- comporta spesso costi di rilevazione più elevati rispetto ad altri procedimenti di scelta delle unità campionarie;
- a parità di dimensione – ossia del numero di osservazioni – può condurre a stime dei parametri della popolazione meno precise di quelle ottenute con altri tipi di campionamento;
- per i motivi sopra detti, nella pratica non viene molto impiegato, specie nelle indagini di grandi dimensioni;
- in generale non vengono utilizzate tutte le informazioni che si posseggono sulla popolazione.

Campionamento casuale stratificato 1/3

Una tecnica di campionamento è quella di suddividere la popolazione di partenza in k gruppi, ciascuno con elementi il più possibile omogenei fra di loro, ed estraendo quindi un campione casuale di opportune dimensioni da ciascuna classe. I gruppi o le classi in cui si suddivide la popolazione si chiamano *strati* ed il procedimento di ripartizione in gruppi si chiama *stratificazione*.



Campionamento casuale stratificato 2/3

Ad esempio si vuole fare un'indagine sui consumi di una popolazione costituita da $N = 1000$ famiglie di cui sia nota la distribuzione secondo le classi di reddito mensile:

Gruppi di famiglie	Reddito mensile (in migliaia di euro)	Numero di famiglie
A	1 - 2	400
B	2 - 4	280
C	4 - 6	220
D	6 - 10	100

sulla base di un campione di $n = 50$ unità.

E' allora opportuno suddividere la popolazione in quattro strati, e quindi consegue che un certo numero di famiglie con reddito elevato sarà certamente incluso nel campione globale, mentre ciò poteva non accadere nel caso di scelta casuale semplice.

Campionamento casuale stratificato 3/3

Vantaggi

- La stratificazione consente, in generale, di aumentare la precisione delle stime senza accrescere la dimensione totale del campione.
- La stratificazione è molto conveniente quando la distribuzione statistica della variabile da rilevare è fortemente asimmetrica.

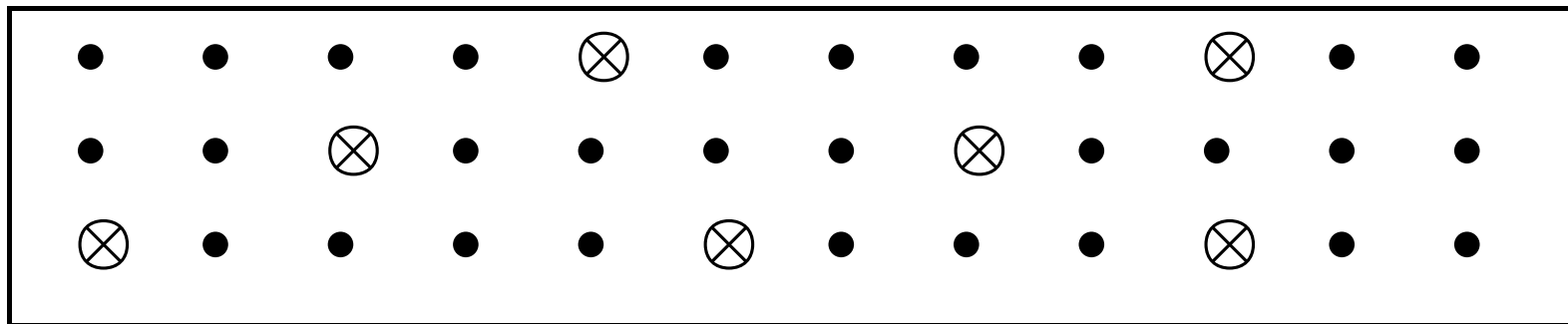
Svantaggi

- Se non si hanno sufficienti informazioni a priori, la costruzione degli strati può risultare alquanto costosa; inoltre se la stratificazione è errata si possono ottenere risultati fuorvianti.

Campionamento sistematico 1/3

Dopo aver ordinato e numerato a partire da 1 le unità della popolazione si sceglie una unità campionaria ogni k unità della popolazione. Generalmente la prima unità viene scelta a caso fra 1 ed il numero più vicino alla frazione di campionamento N/n .

Così se la prima unità scelta è t , le unità campionarie sono quelle contrassegnate dai numeri $t, t + k, t + 2k, \dots$. Ad esempio scegliendo un'unità ogni 20 si ottiene un campione contenente il 5% del totale delle unità.



Campionamento sistematico 2/3

Vantaggi

- Le operazioni di estrazione, rilevazione e controllo del campione sono, in generale, più facili e rapide rispetto a quelle dei campionamenti casuale semplice e di quello stratificato.
- Le stime sono più precise di quelle di un analogo campione casuale se la variabilità entro il campione è maggiore di quella dell'intera popolazione.

Campionamento sistematico 3/3

Svantaggi

- Se la popolazione ordinata varia con tendenza lineare, il campionamento sistematico è più efficiente del campionamento casuale ma meno del campionamento stratificato.
- Se la popolazione ordinata ha un andamento periodico, l'efficienza di un campione sistematico dipende dal valore k e dalla relazione fra k ed il periodo dell'oscillazione.
- Quando N non è un multiplo intero di k , le stime, a rigore, si considerano corrette soltanto per campioni di dimensione maggiore di 50.

Campionamento sistematico - Esempio 1/3

In un sondaggio eseguito per conto dell'Istituto Doxa tra i visitatori della Fiera di Milano, furono date le seguenti istruzioni agli intervistatori:

1. Il visitatore sarà avvicinato nella fase finale della visita, quando sta per uscire dalla Fiera. Perciò Ella si aggirerà in un punto che non sia troppo lontano dall'uscita assegnatale, e tuttavia in un punto che non molto vicino alla stessa da rendere scomoda o sgradita l'intervista.
2. Intervisti soltanto i visitatori adulti sia maschi che femmine. Vanno esclusi i bambini ed i ragazzi.

Campionamento sistematico - Esempio 2/3

3. Non intervisti il personale addetto agli stands che si trovano nel recinto della Fiera per una ragione che non sia quella della visita alla Fiera stessa (vigili, guardiani, camerieri, rivenditori, etc.).
4. Le persone da intervistare dovranno essere scelte puramente a caso. A tale scopo le suggeriamo di procedere nel modo seguente: a partire da un certo istante, conti le persone che le passano davanti ed avvicini, ad esempio, la decima. Compiuta la prima intervista, ricominci a contare e avvicini la decima persona. In caso di rifiuto di questa, fermi la prima persona che le passi vicino.

Campionamento sistematico - Esempio 3/3

5. Non si devono intervistare i visitatori che stanno seduti sulle panchine o che sono ferme all'entrata dei padiglioni, sugli orli delle aiuole, etc., perchè si tratta di regola di persone che hanno visitato solo una parte della Fiera.
6. Eviti di intervistare le persone che sono state presenti ad un'altra intervista. Perciò, se l'intervistato si trova in compagnia di una o più persone, queste ultime non dovranno essere mai intervistate, neppure se si offrono spontaneamente.
7. Non si lasci mai guidare da simpatie o antipatie istintive nella scelta delle persone da intervistare.

Errori di Rilevazioni 1/2

1. *Errori nell'assunzione dei dati*

- (a) imperfezioni nei questionari;
- (b) mancanza di professionalità intervistatore;
- (c) risposte (volutamente) errate.

2. *Errori nello spoglio dei dati*

- (a) interpretazione dei dati;
- (b) trascrizione;
- (c) di riferimento;
- (d) elaborazione.

Errori di Rilevazioni 2/2

3. *Errori di misura*

- (a) accidentali;
- (b) costanti;
- (c) sistematici;
- (d) distribuzione.