

## LE INDAGINI STATISTICHE

Una indagine statistica può essere assimilata ad un processo di produzione che sinteticamente può essere individuato nelle seguenti fasi:

1. Definizione degli obiettivi e ipotesi di ricerca;
2. Rilevazione dei dati;
3. Spoglio dei dati;
4. Elaborazione dei dati;
5. Revisione, valutazione, errori e diffusione dei risultati.

### 1 Definizione degli obiettivi e ipotesi di ricerca

La fase principale di ogni indagine statistica è la fase di progettazione della stessa. In questa fase vengono definiti gli obiettivi dell'indagine; viene individuato l'universo statistico di riferimento (totale o campionario); vengono caratterizzati i casi statistici oggetto di osservazione e le relative modalità con cui rilevarne le caratteristiche; vengono analizzate le risorse disponibili (personale, tecnologie, tempi, costi, ecc.); viene predisposto il piano di lavoro; vengono preparate le modalità di rilevazione (schede, questionari, interviste, ecc.); viene predisposto il piano di elaborazione dei dati ed infine vengono definite le modalità di diffusione dei risultati.

Gli *obiettivi dell'indagine*, per essere ben posti, devono soddisfare le caratteristiche: Devono essere definiti in modo chiaro e mirare a scopi ben precisi; devono dipendere solo da chi li ha formulati (responsabile della ricerca); devono essere raggiungibili e compatibili con le risorse disponibili; devono essere misurabili o verificabili. Nella fase di definizione degli obiettivi, di norma, viene individuato anche l'universo statistico di riferimento e, compatibilmente con le risorse disponibili ed in relazione al livello di conoscenza che si vuole avere, si decide se effettuare una indagine generale o campionaria.

La formazione dei *casi statistici* oggetto dell'indagine, costituisce una fase particolarmente critica dell'intero progetto. Infatti una volta determinato l'universo statistico, bisogna selezionare le caratteristiche rilevanti ai fini dell'indagine che si vuole svolgere. E' necessario selezionare solo le informazioni necessarie all'obiettivo prefissato trascurando altre informazioni che, ancorché d'interesse non sono strettamente necessarie. Si tenga conto che la gestione di un numero rilevante d'informazioni da rilevare ed elaborare rallenta l'indagine e fa levitare i costi. Una buona caratterizzazione dei caratteri, costituisce anche una linea per la progettazione ed implementazione del *questionario* mediante il quale rilevare le modalità con cui si presentano i caratteri (rilevazione dei dati).

La progettazione ed implementazione del *questionario*, segue tre linee direttive articolate nel seguente ordine:

1. Definizione degli aspetti generali del questionario (struttura, linguaggio da utilizzare, lunghezza, articolazione delle domande e delle risposte, tipo dei caratteri, colore della carta, ecc.);
2. Redazione del questionario;
3. Verifica su un piccolo campione.

La definizione degli aspetti generali del questionario, richiede una attenta scomposizione dell'obiettivo della ricerca in sub-obiettivi. Per ogni sub-obiettivo vanno ricercati i caratteri statistici da rilevare e quindi procedere alla definizione delle domande da includere nel questionario.

### 2 Rilevazione dei dati

La prima operazione da compiere per analizzare un fenomeno collettivo, è quella della rilevazione, la quale consiste nella raccolta dei dati statistici riguardanti i fenomeni individuali che compongono il fenomeno collettivo oggetto dell'indagine.

Una rilevazione statistica può avere caratteristiche diverse e può essere:

1. *saltuaria* o *continua*
2. *pubblica* o *privata*
3. *parziale* o *totale*
4. *diretta* o *indiretta*
5. *preliminare* o *definitiva* .

Ad esempio un censimento è una rilevazione saltuaria, pubblica e totale. Per avere i dati dei nati in una certa popolazione (o la quotazione di alcune merci) si richiedono, invece, rilevazioni continue, pubbliche e complete.

Una rilevazione statistica richiede, innanzitutto, l'esatta definizione del fenomeno da rilevare ed occorre anche stabilire il modo, il tempo, e lo spazio in cui essa deve essere effettuata e quali sono gli organi ed i mezzi interessati alla rilevazione.

Il modo secondo cui può essere condotta la rilevazione si distingue in:

**automatica** : quando deriva da dichiarazioni provenienti direttamente dalle persone interessate (es. le rilevazioni dell'ufficio di stato civile per le nascite, i morti, i matrimoni, ecc.);

**riflessa** : quando i dati vengono raccolti da appositi rilevatori (es. il censimento).

Riguardo al tempo la rilevazione può essere:

**continua** : quando le rilevazioni vengono registrate man mano che i fenomeni si verificano;

**periodica** : quando viene effettuata ad intervalli regolari di tempo (es. il censimento);

**occasionale** : quando viene compiuta senza alcuna periodicità (es. la rilevazione dei danni provocati da una guerra, oppure i sondaggi politici).

Gli organi che eseguono le rilevazioni statistiche possono pubblici o privati. Le rilevazioni compiute dagli organi pubblici riguardano fenomeni di interesse pubblico come, ad esempio, quelle di carattere demografico ed economico. In Italia il principale organo pubblico dedito agli studi di statistica è l'ISTAT. Le rilevazioni private sono compiute da imprese commerciali su determinati fenomeni che rivestono particolare interesse di ricerca per alcuni privati.

Per quanto riguarda i mezzi con i quali possono essere condotte le rilevazioni statistiche diciamo soltanto che per le rilevazioni automatiche si usano registri, ruoli, ecc., mentre per quelle riflesse si usano dei questionari.

### 3 Spoglio dei dati

Una volta ultimata la raccolta delle unità statistiche si riunisce tutto il materiale e si procede a controlli di natura diversa per cercare di eliminare inesattezze ed errori. Una volta eseguiti i controlli si passa allo spoglio ed alla classificazione dei risultati raggruppando gli elementi raccolti secondo i caratteri prestabiliti formando delle tabelle di spoglio. Queste sono costituite da varie colonne o righe che sono riferite ai diversi caratteri del fenomeno collettivo che sono stati oggetto della rilevazione. In ciascuna colonna (o riga) vengono riportati i rispettivi dati di frequenza che sono stati rilevati. Le tabelle statistiche si dividono in **semplici**, **complesse** e **a doppia entrata**.

Le **tabelle semplici** sono prospetti nei quali sono elencate le modalità qualitative o quantitative del fenomeno in esame ed a fianco le relative frequenze o intensità.

### Esempio.

Riportiamo una tabella semplice riguardante la distribuzione di una popolazione di 10.000 individui secondo la statura suddivisa in classi di intensità di 10cm in 10cm a partire dall'altezza di 120cm:

Statura	Frequenza
120-130	4
130-140	15
140-150	136
150-160	2324
160-170	5604
170-180	1801
180-190	98
190-200	18
TOTALE	10.000

Le tabelle complesse possono ritenersi una composizione di tabelle semplici che presentano dati statistici riguardanti più fenomeni.

Osservatori	Pressione media	Umidità relativa	Sereno	Misto	Coperto
Torino	1016,2	80	104	112	150
Milano	1017,3	77	103	101	161
Venezia	1015,5	76	112	127	127
Bologna	1016,8	72	106	109	150
Firenze	1015,6	71	97	115	154
Roma	1015,8	72	119	147	100
Pescara	1014,9	75	109	127	130
Napoli	1015,2	69	107	127	132
Palermo	1015,1	60	121	141	104
Cagliari	1016,5	69	125	144	96

Quando lo spoglio delle unità statistiche è stato effettuato secondo due caratteri (ad esempio numero di abitanti e numero di stanze per unità abitative), l'osservazione di ogni unità statistica conduce a due risultati. Allora per la rappresentazione di queste distribuzioni statistiche si fa uso delle tabelle a doppia entrata:

Prof.ne padre	Licei Cl. e Sci.	Ist.Tecnici	Ist.Prof.	Totale
Agricoltori	50	160	90	300
Artigiani	20	50	30	100
Commercianti	60	110	30	200
Impiegati	150	440	110	700
Liberi Prof.sti	90	50	10	150
Operai	10	60	330	400
Altre Prof.ni	20	30	100	150
Totale	400	900	700	2000

Le modalità del primo carattere sono rappresentate dal tipo di scuola, mentre quelle del secondo carattere sono rappresentate dalla professione del padre. In questo caso tanto le modalità del primo carattere quanto quelle del secondo sono qualitative.

#### **4 Elaborazione dei dati**

Questa attività permette di organizzare le informazioni raccolte attraverso la precedente attività. E' necessario, anzitutto, procedere all'inserimento dei dati all'interno di una banca dati adeguatamente predisposta. Dopo di che si procederà al calcolo di medie o percentuali e all'eventuale rappresentazione dei risultati ottenuti in grafici che, sicuramente, facilitano la visualizzazione e l'interpretazione. Sono diverse infatti le **tipologie di elaborazione** utilizzabili ed ognuna è coerente e adeguata alle necessità informative e alle scelte metodologiche fatte.

Per l'elaborazione dei dati sono disponibili alcuni strumenti che forniscono diverse indicazioni a supporto della realizzazione di questa attività: schema di sintesi per l'elaborazione dei dati; calcolo delle medie; correlazione; correlazione per ranghi; indicatori di variabilità .

#### **5 Revisione, valutazione, errori e diffusione dei risultati**

Prima della fase di interpretazione dei dati importante è la revisione cioè la fase in cui si accertano l'aspetto quantitativo, come la completezza e la non ridondanza delle unità rilevate e l'aspetto qualitativo, quali lacune o errori di compilazione. Un aspetto cruciale della revisione riguarda il trattamento delle mancate risposte, dovute a coloro che non forniscono (in parte o in tutto) risposte al questionario.

Successivamente si passa a dare un giudizio sulla qualità dei dati forniti da una rilevazione statistica che comporta la valutazione di un complesso di elementi: specificazione degli obiettivi della ricerca, schema di campionamento adottato (nel caso di rilevazioni campionarie), completezza e aggiornamento delle liste della popolazione oggetto di studio, adeguatezza del modello di rilevazione, preparazione dei rilevatori. Infatti una rilevazione può dirsi di elevata qualità quando risponda alle esigenze dei soggetti interessati e, nello stesso tempo, rispecchi il rigore metodologico che dà consistenza ai risultati ottenuti. Per assicurare uno standard soddisfacente bisogna quindi porre la massima attenzione verso alcuni fattori fondamentali che caratterizzano il percorso della rilevazione. Si fa riferimento in particolare all'errore campionario (frutto dello schema di campionamento) e all'errore non campionario (direttamente correlato alle modalità e allo strumento di rilevazione).

L'errore non campionario è strettamente legato ai seguenti fattori:

1. Copertura delle liste (coverage error) : le liste rappresentano gli archivi su cui si pianifica la rilevazione totale o da cui è possibile campionare la popolazione. Esse possono consistere in archivi anagrafici (come quelli comunali), telefonici, banche dati costruite ad hoc (archivi delle camere di commercio) eccetera. Quando si individua una lista è importante assicurarsi che essa risponda esattamente alle esigenze dell'analista, che ricopra interamente la popolazione di riferimento e che soprattutto sia quanto più possibile aggiornata. Il mancato rispetto di una, o più, di queste condizioni porterebbe all'impossibilità di includere nella rilevazione tutte e sole le unità appartenenti alla popolazione di interesse o, nel caso di rilevazione campionaria, all'estrazione di un campione non rappresentativo della popolazione, in quanto non a tutte le unità statistiche verrebbe attribuita una uguale probabilità di estrazione.
2. Misurazione e raccolta dei dati (response error) : gli errori di misurazione sono relativi al questionario in quanto strumento di rilevazione. Questo deve essere strutturato in maniera tale da assicurare una corretta e chiara formulazione delle domande. Gli errori di raccolta

riguardano le modalità di acquisizione delle risposte. L'intervistatore deve svolgere l'intervista in modo corretto, mettendo a suo agio l'intervistato e spiegando in modo chiaro e semplice le domande e le informazioni che si richiedono. Il lavoro dell'intervistatore è utile, inoltre, per evitare il fenomeno delle mancate risposte dovute a imprecisione o incomprensione.

3. Spoglio e codifica dei dati (data processing error) : questi errori sorgono nella fase di digitazione delle informazioni raccolte, al momento in cui vengono inserite nella banca dati che sarà utilizzata successivamente per le analisi statistiche. I metodi di rilevazione che utilizzano il computer, come ad esempio il Cati (Computer assisted telephone interviewing), permettono di ridurre notevolmente gli errori e velocizzare la fase di data processing.

La diffusione dei dati conclude il processo di produzione dell'informazione statistica, mettendo a disposizione dei cittadini, delle imprese e delle istituzioni i risultati delle rilevazioni. Il rilascio delle informazioni avviene nel rispetto delle leggi e dei regolamenti, fissati dagli organismi nazionali e internazionali, per garantire alla collettività un'informazione tempestiva, corretta e completa, e per tutelare la riservatezza degli individui e delle imprese. La diffusione dell'informazione statistica attraverso internet e l'utilizzo crescente del web per la consultazione dei dati comportano un adeguamento delle modalità di rilascio alle regole stabilite a livello nazionale e internazionale sull'accessibilità e sull'usabilità.