

## **Disegno d'indagine**

Nella fase di definizione del disegno di indagine si individuano le modalità di rilevazione che soddisfano esigenze conoscitive su:

1. Particolari tipologie di stime (ad esempio stime di variazione netta o lorda in due o più punti nel tempo; stime di andamenti tendenziali su più periodi temporali; stime di durata o frequenze di accadimento per specifiche tipologie di eventi; stime di relazioni fra caratteristiche; ecc.);
2. Tutta o parte della popolazione oggetto di studio.

## **Considerazioni di carattere generale**

- Scegliere il disegno di indagine non opportuno può pregiudicare del tutto o in parte gli scopi della ricerca
- Raccogliere informazioni su tutta la popolazione implica un aumento notevole dei costi ed un incremento del numero di errori non campionari. Pertanto la scelta di un disegno di indagine che preveda la raccolta dei dati di tutte le unità statistiche appartenenti alla popolazione è da limitare a casi particolari (censimenti; indagini amministrative; ecc.)
- Il metodo di selezione del campione è determinato in base alla struttura degli archivi di base ed alle informazioni in esso contenute al fine di massimizzare l'efficienza delle stime prodotte
- Fissare la dimensione del campione in modo da garantire il calcolo di stime della precisione desiderata.

La prima grande distinzione tra i possibili disegni d'indagine può essere fatta tra indagini trasversali ed indagini longitudinali.

Nelle **indagini trasversali** si rilevano le informazioni sulle unità statistiche relativamente ad un determinato momento o periodo di tempo, con l'obiettivo di stimare le caratteristiche riferite allo stato della popolazione allo studio nel momento o nel periodo di interesse.

Nelle **indagini longitudinali** lo scopo è quello di misurare l'evoluzione nel tempo delle caratteristiche d'interesse ricorrendo all'espedito di ricontattare le unità per analizzarne i cambiamenti.

*N.B. Questa distinzione non è rigida. Infatti utilizzando delle opportune accortezze è possibile stimare i cambiamenti ricorrendo alle indagini trasversali o stimare misure di stato con indagini longitudinali.*

## **Indagini occasionali**

La scelta di un piano di indagine occasionale è fatta quando l'obiettivo è quello di calcolare stime riferite a caratteristiche possedute dalla popolazione in un singolo istante di tempo (distribuzione per

età della popolazione in un dato istante) o riferite ad un intervallo di tempo (distribuzione del fatturato realizzato nell'arco di un anno).

### **Indagini ripetute (periodiche o ricorrenti)**

L'organizzazione dell'indagine viene ripetuta in momenti programmati nel tempo. Si può prevedere di non sovrapporre le unità indagate nelle diverse occasioni oppure si può decidere di usare dei panel ruotati. In pratica nelle indagini ripetute con parziale sovrapposizione del campione le unità statistiche sono introdotte nel campione, indagate per un prefissato numero di occasioni e quindi escluse (**ruotate**). Comunque non si persegue l'obiettivo di calcolare stime di transizione di stato. (Es. gli exit pool nelle giornate elettorali).

### **Indagini longitudinali senza rotazione**

Con tale piano di indagine si risponde all'esigenza conoscitiva di studiare le modifiche avvenute in un collettivo durante il tempo, analizzando i cambiamenti avvenuti sui singoli individui. Viene quindi seguito un particolare gruppo di unità nel tempo e si crea un archivio longitudinale per ogni unità osservata. Va però evidenziato che le stime calcolate sono riferite alla sola popolazione di partenza.

### **Indagini longitudinali con rotazione**

L'obiettivo è quello di seguire un particolare gruppo di unità per un periodo di tempo ma mediante l'introduzione periodica di nuovi soggetti nel campione è possibile garantirne la rappresentatività anche nelle occasioni successive alla prima (tenendo conto del fatto che il collettivo di interesse si modifica nel tempo a causa dell'ingresso di nuove unità, ad esempio nascite ed immigrazioni).

### **Indagini totali**

Si definisce indagine totale quel piano di indagine che prevede di rilevare tutte le unità per le quali si possiede un indirizzo negli archivi di base utilizzati.

Il censimento è un conosciuto ed importante esempio di indagine totale. Oltre al censimento, si definiscono indagini totali quelle in cui la popolazione è costituita da poche unità molto importanti, cioè ciascuna di esse possiede una quantità rilevante delle caratteristiche da studiare (ad esempio il fatturato delle grandi imprese), la cui esclusione introdurrebbe gravi distorsioni nelle stime. Un altro esempio è costituito dalle indagini basate su dati amministrativi in cui la raccolta dei dati è avvenuta per finalità diverse da quella statistica (anagrafi; dati sul superamento degli esami; archivi INPS; registro delle malattie infettive; ecc.).

### **Indagini campionarie**

In molte circostanze si ricorre alle indagini parziali, dove cioè una sola parte della popolazione viene selezionata andando a formare il campione, per ovviare ai problemi che un'indagine totale comporta.

In generale l'indagine campionaria consente di dedicare maggiore attenzione a tutte le attività connesse al miglioramento ed al controllo della qualità dei dati raccolti grazie al diminuito onere che comporta la rilevazione.

Ovviamente non va dimenticato che l'attendibilità delle stime calcolate sarà fortemente influenzata dal tipo di strategia di campionamento scelta. Inoltre, si può agire sul livello di precisione delle stime dimensionando adeguatamente la dimensione del campione a partire dalla variabilità delle grandezze da misurare (qualora sia nota a priori).

### **Archivi di base**

Gli archivi di base sono costituiti dalle liste, mappe o da altre forme di elenchi che raccolgono l'informazione disponibile sulle unità che compongono il collettivo oggetto di studio.

L'informazione disponibile negli archivi deve almeno riportare sufficienti dettagli per consentire la localizzazione delle unità per la rilevazione, e teoricamente potrebbero contenere solo queste informazioni.

Raramente gli archivi possono essere considerati perfetti, in quanto spesso si presentano problemi di:

- ü Incompletezza (non tutte le unità sono rappresentate)
- ü Inadeguatezza (mancano delle informazioni)
- ü Inaccuratezza (sono presenti errori)
- ü Obsolescenza (i dati non sono aggiornati, e presentati in modo non adeguato)
- ü Presenza di duplicati delle unità

Le informazioni disponibili devono, almeno, permettere di raggiungere operativamente le unità sul territorio e devono garantire la possibilità di distinguere l'una dall'altra le unità in modo tale da poterle riconoscere al momento del contatto.

Sono due i tipi di struttura che caratterizzano gli archivi che vengono usati nella pratica e precisamente:

1. **Uno a uno** – c'è una corrispondenza biunivoca tra elemento dell'archivio ed unità appartenente alla popolazione di riferimento (è la stessa unità a far parte dell'archivio)
2. **Uno a molti** – ad un elemento della lista corrispondono uno o più elementi del collettivo allo studio ma ad ogni elemento della popolazione obiettivo corrisponde un solo elemento

della lista (è il caso delle anagrafi di popolazione dalle quali si accede alle famiglie e quindi ai singoli componenti)

### **Raccomandazioni**

1. In fase di progettazione occorre valutare possibili alternative sulla base di quanto in archivio risulta aggiornato e rappresentativo della popolazione obiettivo.
2. Valutare l'affidabilità delle informazioni in esso contenute (ad esempio gli indirizzi per il contatto delle unità) anche attraverso delle analisi di fattibilità basate su studi pilota.
3. Se disponibili, utilizzare informazioni provenienti da altre indagini che utilizzano gli stessi archivi.
4. Verificare la possibilità di ottenere aggiornamenti degli archivi all'epoca di riferimento dell'indagine (ad esempio abbinando archivi indipendenti).
5. Predisporre tutte le misure possibili per misurare gli errori nell'archivio in fase di rilevazione (introdurre nel questionario, o preparare appositi modelli da far compilare ai rilevatori, domande utili per contare il numero di unità non trovate o non più esistenti oppure ad attestare l'affidabilità delle informazioni contenute in archivio, ad esempio numero di addetti delle imprese o loro fatturato. Nel caso in cui le unità possano subire delle trasformazioni nel tempo, come nel caso delle imprese, è utile attivare delle procedure che consentano di registrare tali variazioni).
6. Documentare con dovizia di particolari tutti i problemi emersi sia in fase di rilevazione sia in fase di analisi. Il materiale prodotto potrà, così, essere usato nella ricerca delle cause che hanno generato i problemi e nella predisposizione di contromisure per successive indagini

### **Strategia di campionamento**

Impostare una strategia di campionamento implica la definizione delle modalità di estrazione del campione, delle sue dimensioni e delle funzioni dei dati utilizzate per ottenere, dal campione stime riferite alla popolazione di interesse.

La tecnica di estrazione del campione è scelta sulla base delle caratteristiche degli archivi o liste di base disponibili pertanto occorre valutare attentamente le caratteristiche degli archivi di base disponibili.

Perché un campione sia rappresentativo della popolazione di provenienza occorre che gli archivi di base per l'estrazione siano in buono stato di aggiornamento, che la dimensione del campione sia sufficiente e che le procedure di selezione per lo specifico disegno siano appropriate.

### **Campionamento casuale semplice**

E' la più semplice fra le modalità di campionamento. Essa equivale ad associare ad ogni unità della popolazione una biglia numerata e ad estrarre a caso da un'urna, una per volta e senza riporla, tante biglie quante sono le unità che si vogliono campionare. Affinché si possa applicare tale metodo è necessario disporre di una lista che elenchi tutte le unità statistiche della popolazione.

### **Campionamento sistematico**

E' molto efficiente da realizzare quando si disponga della lista delle unità statistiche della popolazione sotto forma di file. Si estrae un numero a caso tra 1 e N (numerosità della popolazione) e si inserisce nel campione l'unità corrispondente nella lista. Le unità successive sono scelte scorrendo la lista a partire dalla prima unità prescelta e selezionando nuove unità con un passo dato dal rapporto  $N/n$ , dove n è il numero di unità che si vogliono inserire nel campione. Una volta giunti in fondo alla lista delle N unità, occorre proseguire il conteggio a partire dall'inizio della lista. Il procedimento termina quando sono state selezionate tutte le n unità da campionare.

### **Stratificazione del campione**

Prima di procedere all'estrazione si suddivide la popolazione in due o più gruppi secondo una o più caratteristiche conosciute sulle unità statistiche. Si procede, quindi, all'estrazione delle unità indipendentemente per ogni gruppo (strato).

Per poter applicare tale tecnica è necessario che le caratteristiche usate nella formazione degli strati siano disponibili sulla lista per ogni unità della popolazione.

Applicando tale modalità di estrazione si ottengono stime più precise, a parità di dimensione del campione, rispetto al campionamento casuale semplice.

### **Campionamento a più stadi**

Quando non sia disponibile una lista complessiva delle unità della popolazione è possibile ricorrere al campionamento a più stadi. Un esempio di tale situazione è dato dall'anagrafe che non esiste come unico archivio nazionale ma è suddivisa negli 8.103 comuni italiani. In questo caso si procede innanzitutto ad estrarre un campione di comuni (unità di primo stadio) e quindi, per ogni comune selezionato, un campione casuale di famiglie (unità di secondo stadio) da ciascuna lista anagrafica.

Conviene limitare l'uso di tale piano di campionamento alle situazioni di reale necessità poiché le stime con esso ottenibili sono di solito meno efficienti (a causa della maggiore variabilità campionaria) di quelle calcolate applicando il campionamento casuale semplice.

### **Campionamento areale**

Si tratta di una procedura di campionamento utilizzata quando non si dispone di una lista per la selezione delle unità, ma queste sono dislocate sul territorio. In questo caso si procede ad una suddivisione in parti (aree) dell'intero territorio e all'estrazione di un campione di aree. Quindi si esplorano le aree campionate, allo scopo di enumerare esaustivamente le unità presenti al loro interno e produrre delle liste complete. Infine, dalle liste prodotte, si estraggono le unità campione da contattare per la rilevazione vera e propria.

### **Selezione delle unità con probabilità differenti**

È una modalità di estrazione per la quale la probabilità di estrarre una unità nel campione non è la stessa per tutte le unità della popolazione. Si ricorre a questo piano di campionamento quando c'è ragione di ritenere che alcune unità statistiche apportino maggiori informazioni piuttosto che altre e quindi si vuole aumentare la probabilità che siano selezionate.

Ad esempio, in uno studio sul fatturato delle imprese, se si ritenesse che il fatturato delle imprese più grandi sia maggiore di quello delle imprese più piccole allora si potrebbe decidere di selezionare le imprese con probabilità proporzionale al numero di addetti in ciascuna di esse (ovviamente sempre che nell'archivio di base siano disponibili le informazioni sul numero di addetti per ogni impresa).

### **Raccomandazioni**

- Testare, monitorare e validare la strategia di campionamento al fine di valutarne la rispondenza agli obiettivi iniziali e l'adeguatezza rispetto a successive occasioni di indagine (considerare più disegni di campionamento alternativi e valutarli alla luce di informazioni disponibili quali censimenti, indagini precedenti, dati amministrativi o appositi studi pilota)
- Prevedere una rotazione del campione per migliorare l'efficienza delle stime di variazione e per limitare il carico della rilevazione sulle unità statistiche
- Monitorare l'efficienza del disegno di campionamento delle indagini ricorrenti
- Studiare soluzioni per trattare il caso in cui:
  - a. si scopra che alcune delle unità indagate non appartengono allo strato loro assegnato
  - b. non sia stato possibile contattare tutte le unità
  - c. sia stata selezionata una unità non appartenente alla popolazione
  - d. l'unità non accetta di partecipare all'indagine
- Non adottare la strategia più efficiente se si ha ragione di ritenerle difficilmente applicabile per evitare che in fase di selezione del campione siano introdotti errori di cui sia difficile valutare gli effetti sulle stime

## **Tecniche di indagine**

È la fase, nel ciclo produttivo dei dati, in cui si definisce l'insieme delle modalità:

- di contatto delle unità statistiche interessate dalla rilevazione
- di reperimento delle informazioni oggetto di interesse.

La scelta della tecnica di indagine più idonea a raccogliere le informazioni oggetto della ricerca è uno degli aspetti di maggiore importanza nella pianificazione e nell'esecuzione di una indagine ed è strettamente connessa ad altre caratteristiche quali il fenomeno indagato, gli archivi di base, la strategia di campionamento, l'organizzazione del personale sul campo, i costi e i tempi attesi.

È, inoltre, da non sottovalutare l'impatto che la tecnica di indagine scelta ha sulla qualità dei dati in termini di mancate risposte ed errori di misura.

### **Esempi:**

- Intervista diretta (o faccia a faccia)
- Intervista telefonica
- Questionario postale autocompilato
- Diario
- Dati amministrativi
- Osservazione diretta
- Tecniche miste
- Nuove tecnologie

## **Il Questionario**

Il questionario di indagine è lo strumento di misura designato a raccogliere le informazioni sulle variabili qualitative e quantitative oggetto di indagine.

Il questionario deve essere visto come uno strumento di comunicazione finalizzato a facilitare l'interazione fra il ricercatore, il rilevatore (se presente) e il rispondente.

Affinché possa svolgere il suo ruolo occorre che il questionario sia uno strumento standardizzato; ovvero domande e comunicazione devono essere identiche per tutti i rispondenti al fine di garantire il confronto tra le informazioni raccolte.

## **Realizzazione**

1. Definizione degli obiettivi e concettualizzazione
2. Redazione del questionario

### 3. Verifica del questionario

#### **Definizione degli obiettivi**

- Definire esattamente quali sono i temi che interessano l'indagine escludendo quelli che non sono di interesse primario
- Preparare la lista delle variabili (e non direttamente le domande) da raccogliere rispetto ai temi di interesse identificati in precedenza
- Preparare un piano provvisorio delle analisi statistiche da compiere per accertarsi che i contenuti necessari allo studio siano tutti espressi

#### **Redazione**

- Organizzare complessivamente il questionario in aree omogenee per tematica trattata, raggruppando le domande relative allo stesso tema possibilmente nella stessa area
- Adottare accorgimenti particolari a seconda del tipo di somministrazione scelta
- Saper individuare quanto spazio è opportuno dare all'intervistato nella risposta (domande aperte, domande strutturate)
- Adottare un linguaggio comprensibile a tutti gli intervistati, non ambiguo e preciso e chiedere cose a cui l'intervistato può rispondere
- Essere precisi nel riferimento temporale delle domande e fare particolare attenzione ai quesiti retrospettivi
- Fare un uso oculato delle domande delicate, utilizzando a volte domande indirette o tecniche proiettive
- Collocare le domande in modo che non influenzino le risposte alle successive e controllare la sequenza delle risposte ad una stessa domanda
- Saper individuare la lunghezza ottimale del questionario per impegno di tempo, approfondimento di temi, ridondanza di informazioni
- Impostare graficamente il questionario in modo da renderlo una guida efficace per la compilazione ed uno strumento adeguato di comunicazione
- Individuare i codici più adatti per ciascuna modalità di risposta ai quesiti del questionario
- Adottare criteri standard per le variabili strutturali
- Prevedere nel questionario una parte per i codici identificativi ed una per i quesiti sui controlli di qualità

#### **Verifica**



Valutare se:

- risponde alle esigenze conoscitive dell'indagine;
- sono state omesse domande;
- i riferimenti spaziali e temporali dei quesiti sono sufficienti;
- il linguaggio e la struttura delle domande sono adeguati;
- è facilmente comprensibile per gli intervistati e semplice da gestire per gli intervistatori.

Prevedere:

- revisione estesa da parte di esperti del fenomeno;
- pre-test;
- test di alternative: si sperimentano versioni alternative del questionario su piccoli campioni indipendenti di unità statistiche;
- indagine pilota: versione completa dell'indagine su scala ridotta per verificare il grado di integrazione tra le fasi dell'indagine ed effettuare eventuali ultimi ritocchi anche sul questionario.

**Gli errori non campionari generabili con i questionari in cui l'unità di rilevazione è costituita dall'individuo**

**Errori dovuti al ricercatore**

- Scarsa chiarezza sugli obiettivi dell'indagine e sull'unità di rilevazione e/o di analisi
- Non corretta individuazione della tecnica di indagine ottimale
- Non corretta predisposizione del questionario (formulazione, sequenza dei quesiti, lunghezza del questionario, errata scelta del periodo di riferimento e di ricordo, grafica, ecc.)
- Non corretta predisposizione delle istruzioni per i rilevatori
- Imprecisioni nel sistema di codifica

**Errori dovuti al rispondente**

- Problemi di memoria
- Mancanza di informazione
- Scarsa motivazione a rispondere attentamente
- Fraintendimento involontario di alcune domande
- Distorsione volontaria della risposta
- Condizionamento legato alla presenza di altre persone
- Incapacità a rispondere per problemi di salute

### **Errori dovuti al rilevatore**

- Impreparazione sul questionario
- Impreparazione sulla conduzione dell'intervista
- Scarsa motivazione, distrazione

### **Errori dovuti al supervisore**

- Carenza nelle istruzioni
- Carenze nel controllo del lavoro degli intervistatori

### **Errori dovuti all'operatore di registrazione**

- Errori di registrazione del *valore* del dato
- Errori di registrazione della *posizione* del dato nel record finale

### **Nuove tecnologie**

Tecniche CASIC (Computer Assisted Survey Information Collection)

CATI: intervista condotta telefonicamente da un intervistatore

CAPI: intervista faccia a faccia con un intervistatore

CASI: questionario autocompilato dal rispondente

Nella categoria CASI rientrano tutte le tecniche che non prevedono il ricorso ad intervistatori e precisamente:

- CSQA (computerized self-administered questionnaire) - il rispondente utilizza direttamente sul proprio PC il programma, in pratica un questionario informatico, per fornire le risposte;
- TDE (touchtone data entry) e VRE (voice recognition entry) – il rispondente telefona ad un computer e risponde alle domande che il programma gli rivolge mediante i tasti del telefono o pronunciando direttamente la risposta.

Principali vantaggi connessi all'uso di un questionario computer-assisted:

- possibilità di condurre interviste dal percorso più articolato (presenza di molte domande filtro);
- costruzione durante l'intervista di domande personalizzate sulla base delle risposte precedenti (formulazione alternativa delle domande);
- rilevazione durante l'intervista delle incoerenze e verifica immediata con il rispondente (riduzione dell'errore non campionario);

- processo di produzione dei dati più efficiente e forte riduzione dei tempi di completamento dell'indagine (inserimento dati, controlli di coerenza contestuale alla compilazione del questionario ed in alcuni casi sintesi delle informazioni)

Principali svantaggi connessi all'uso di un questionario computer-assisted:

- forte investimento iniziale per la creazione del software (progettazione, realizzazione e pre-test);
- nel caso della tecnica CAPI ci sono maggiori costi derivanti dall'acquisto di computer portatili e dalla formazione per i rilevatori (si ammortizzano solo in indagini su più larga scala);
- scarsa flessibilità in caso di imprevisti;
- possibilità che insorgano problemi con l'hardware ed il software durante l'intervista.

#### Regole per la progettazione di un questionario computer-assisted con intervista condotta da rilevatore

- Prevedere tutte le possibili condizioni che si potranno incontrare
- Garantire la possibilità di acquisire una risposta anche in caso di situazioni non previste (l'intervistato deve sempre poter rispondere)
- Associare ad ogni domanda un campo note
- Consentire agevoli ed efficienti spostamenti tra le domande (ritornare ad una domanda precedente e correggere una risposta)
- Evitare i tempi di attesa nell'esecuzione delle operazioni del programma
- Impostare la struttura del programma secondo un approccio modulare (questionari lunghi e complessi; gruppi di domande logicamente legate; successivo utilizzo del modulo)

#### Regole per la progettazione di un questionario computer-assisted con intervista condotta da rilevatore

- Inserire il maggior numero di controlli
- Prevedere l'accesso diretto alle domande che hanno generato l'incompatibilità
- Integrare le informazioni ausiliarie e/o le istruzioni essenziali nel testo della domanda
- Realizzare un help separato con le informazioni ausiliarie e/o le istruzioni da usare poco frequentemente
- Dedicare una zona dello schermo alla visualizzazione di informazioni sull'intervistato (raccolte nel corso dell'intervista, in rilevazioni precedenti oppure note a priori) che evitino nel rilevatore

l'impressione di "isolamento" della domanda corrente dal resto dell'intervista e che lo aiutino a mantenere il controllo della situazione (questionari complessi)

#### Regole per la progettazione di un questionario computer-assisted autocompilato dal rispondente

- Privilegiare le scelte che rendano l'interfaccia del programma di comprensione immediata per il rispondente tipo senza richiedere alcun addestramento informatico
- Studiare una formattazione chiara di ogni singola videata in modo che siano ben evidenziate le informazioni in essa contenute (testo delle domande; eventuale codifica delle risposte; istruzioni su come si risponde)
- Utilizzare una struttura omogenea per le videate
- Sfruttare al massimo le informazioni già note evitando di sovaccaricare ed annoiare il rispondente inutilmente

#### Prospettive di sviluppo dei sistemi CA

I sistemi CASI rappresentano le tecnologie che con più alta probabilità saranno maggiormente sviluppate. Gli elementi favorevoli che determineranno tale tendenza possono sintetizzarsi in:

- maggiori disponibilità di PC a domicilio e facilità di trasmissione dei dati;
- sviluppo del software che consente di realizzare interfacce grafiche di uso immediato anche per utenti non esperti;
- i costi fissi legati all'acquisizione di infrastrutture informatiche e di rete che incidono sempre meno nel budget di una indagine e che sembrano destinati a ridursi costantemente;
- particolare adattabilità all'impiego nelle indagini longitudinali.

Particolare attenzione merita la modalità CAWI cioè CASI on-line su Internet. Già oggi è utilizzata per indagini su sub-popolazioni precocemente alfabetizzate (Es.: studenti ed aziende).

#### Problematiche non ancora risolte e limitazioni dei sw CASIC

- Controlli di coerenza tra intervista in corso ed informazioni note a priori (ad esempio da interviste precedenti agli stessi soggetti)
- Registrazione dati in formato proprietario, non accessibile con i normali sw applicativi per l'elaborazione statistica (nella maggior parte dei casi è supportata la sola conversione in ASCII)
- Struttura dell'archiviazione dei dati inadeguata che rende difficile l'uso delle informazioni (semplici tabelloni di dati)

- Nessun supporto per la realizzazione di indagini longitudinali
- Supporto minimale per la realizzazione di indagini WEB