

Parte 1: La produzione statistica dei dati

Introduzione

Statistica e statistiche

La **statistica** analizza in termini quantitativi i fenomeni collettivi (Leti 1986).

Per **fenomeno collettivo** si intende un fenomeno che si applica ad una pluralità di osservazioni (unità distinte tra loro oppure osservazioni ripetute su una stessa unità). Il campo di applicazione della statistica è estremamente variegato andando dai fenomeni naturali (fisici, astronomici, botanici, meteorologici) a quelli economici, sociali, demografici, psicologici.

Esempi di fenomeni collettivi:

Il consumo di un bene in un periodo fissato, la distribuzione per anno del numero di infortuni o delle malattie professionali.

Nello studio dei fenomeni, di qualunque natura essi siano e qualunque sia la finalità che si intende perseguire, si presenta spesso la necessità di **compattare** le manifestazioni degli stessi per migliorarne la comprensione. Tali operazioni si espletano generalmente attraverso:

1. l'esecuzione di rappresentazioni tabellari e grafiche che mettano in evidenza certi aspetti generali del fenomeno o dei fenomeni oggetto di analisi;
2. il calcolo di indici che mettano in evidenza quello che c'è di tipico nelle manifestazioni dei fenomeni;
3. il calcolo di indici che mettano in evidenza quello che c'è di mutabile o di variabile nelle manifestazioni dei fenomeni;
4. la misura delle relazioni fra fenomeni mediante appropriati indici;
5. l'introduzione di modelli che esprimano analiticamente l'insieme delle manifestazioni del fenomeno e la relazione tra fenomeni diversi ma logicamente collegati.

Informazioni e decisioni

Tra le ragioni per cui si è interessati alla comprensione dei fenomeni prenderemo in esame quelle legate alla risoluzione di uno specifico problema decisionale.

Qualunque sia il problema decisionale da risolvere, dal più banale al più complesso, è necessario procedere ad una chiara definizione del problema stesso ed all'individuazione delle possibili relazioni che connettono i vari elementi o aspetti che lo caratterizzano.

Esempio: Un'impresa ha interesse nella produzione e commercializzazione di un nuovo prodotto. Dovrà quindi almeno prendere in esame aspetti quali:

- la domanda potenziale del nuovo prodotto
- la presenza sul mercato di prodotti competitivi
- i costi di produzione e di commercializzazione
- la disponibilità di materie prime di macchinari e personale
- la disponibilità ed il costo del denaro

Il quadro logico di riferimento e le informazioni sono gli ingredienti essenziali di ogni processo decisionale

La disciplina che si occupa della raccolta e del trattamento scientifico delle informazioni è la Statistica. La Teoria delle decisioni rappresenta il conteso di riferimento quando le informazioni devono essere utilizzate per risolvere uno specifico problema decisionale.

La Teoria statistica delle decisioni

La Teoria statistica delle decisioni si sostanzia come la fusione della statistica e della Teoria delle decisioni. Tale disciplina può essere vista come generalizzazione ed estensione della statistica che in tal modo risulta meglio caratterizzata nelle sue diverse connotazioni e meglio precisata nei contenuti. In altre parole si può definire la statistica come il fondamento logico per la risoluzione dei problemi decisionali

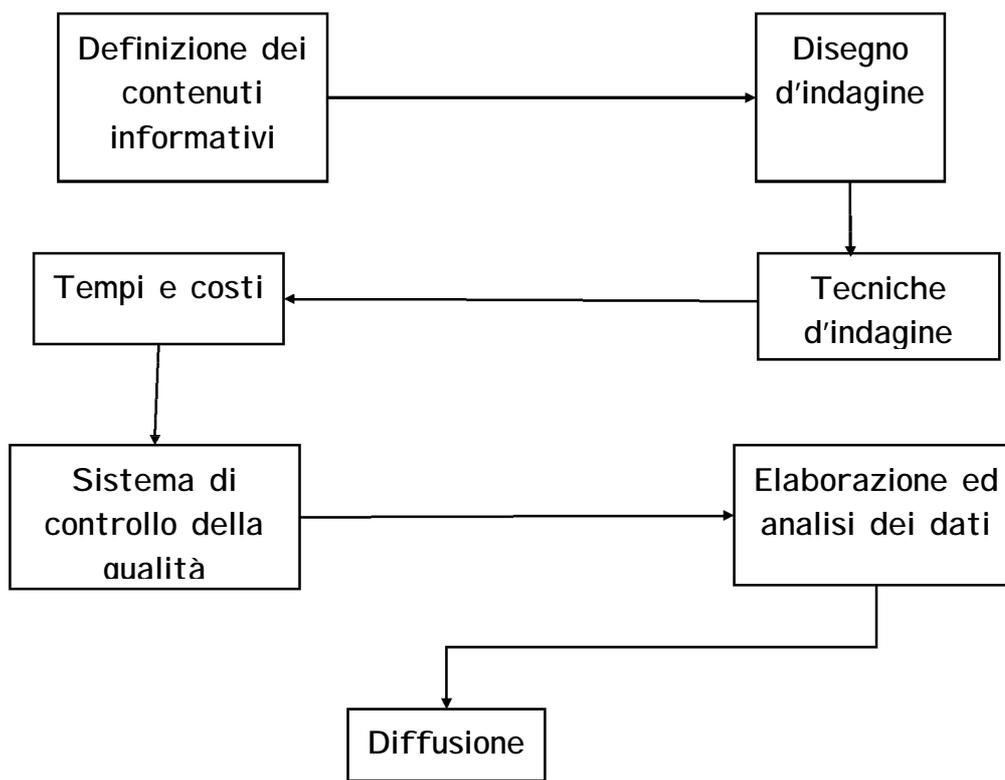
- Cosa sono i dati?

I dati sono dei *prodotti*. Rappresentano, infatti, l'output di una serie di fasi integrate fra loro.

- Cos'è il processo di produzione dei dati?

È il *ciclo produttivo* dei dati

Il ciclo produttivo dei dati



La pianificazione della produzione dei dati

Costituzione del gruppo di progettazione

L'elevato grado di complessità e la multidisciplinarietà richiesta per trattare tutti gli aspetti che attengono alla fase di pianificazione impongono la costituzione di un gruppo di progettazione nel quale siano coinvolti esperti del fenomeno oggetto di studio e delle problematiche che attengono agli ambiti operativi. In particolare, il gruppo di progettazione ha come obiettivo quello di definire gli scopi conoscitivi dell'indagine, adottare definizioni e concetti operativi e pianificare gli aspetti applicativi del processo di produzione. Estrema attenzione dovrà essere posta all'integrazione fra i concetti e le procedure.



Requisiti

- Nomina di un Presidente al quale affidare il coordinamento delle attività
- Contenuta numerosità dei componenti (rischio di disperdere le energie nello sforzo di mediazione; autoesclusione di alcuni membri)

Non è possibile dare indicazioni precise sul numero poiché la scelta ottimale è legata alla complessità del fenomeno allo studio. Nel caso di fenomeni particolarmente complessi una possibile strategia è quella di creare un nucleo stabile di esperti a cui affiancare di volta in volta, e solo per il tempo necessario, altri esperti su temi specifici

- Inserire membri che siano in grado di creare armonia tra tutti i partecipanti all'indagine
- Composizione bilanciata (equilibrio tra membri interni ed esterni)
- Composizione modulare (costituzione, se necessario, di uno o più sottogruppi di lavoro a cui affidare lo studio di argomenti specifici e/o lo svolgimento delle relative attività ed elaborazioni di supporto)
- Costituzione di una segreteria formata da una o più persone interne alle strutture che verranno coinvolte nelle attività operative

Documento di progettazione

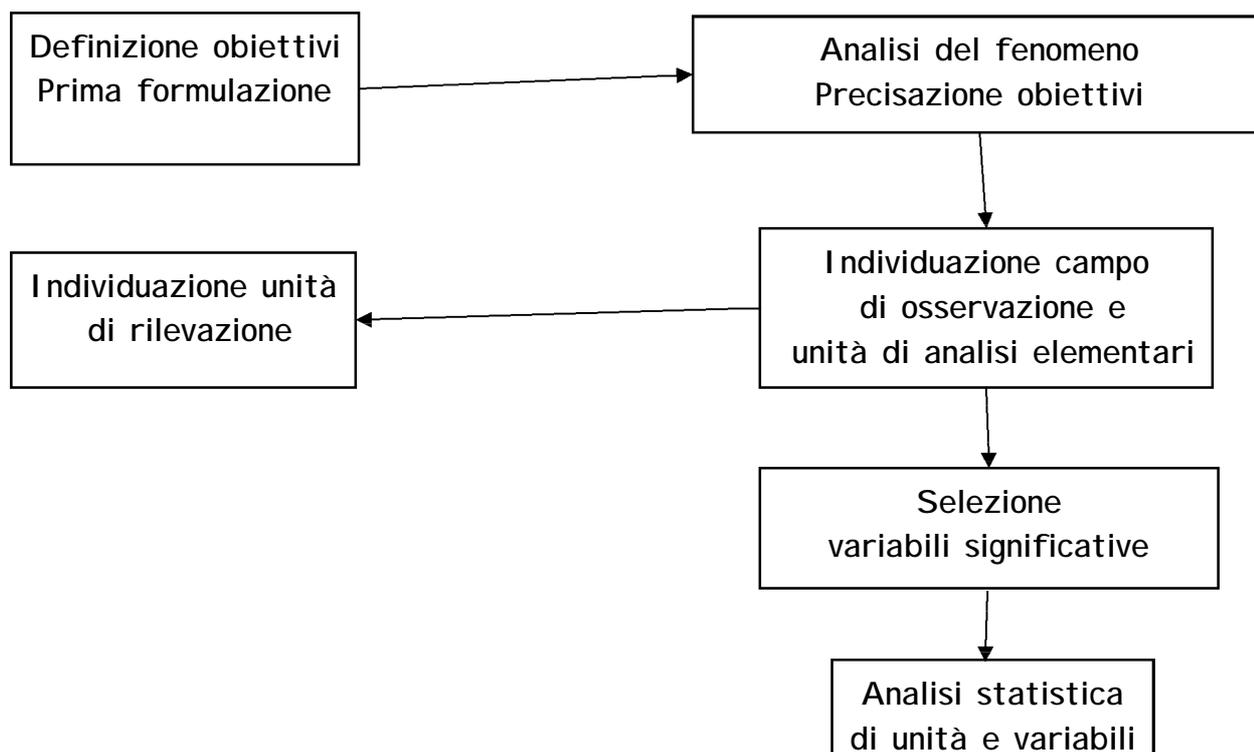
Il documento di progettazione viene redatto dal gruppo di progettazione al fine di descrivere tutti gli aspetti relativi alla progettazione.

È buona norma organizzare tale documento in due sezioni. Una che racchiuda l'analisi concettuale e l'altra che tratti gli argomenti attinenti all'implementazione dell'indagine, non tralasciando di descrivere le relazioni esistenti fra i due ambiti e di evidenziare l'impatto che le definizioni e le procedure individuate hanno sulle componenti della qualità delle informazioni prodotte.

Elementi	Dimensioni
Obiettivi	<ul style="list-style-type: none"> - contestualizzazione del fenomeno oggetto di indagine - analisi delle informazioni già disponibili da altre fonti
Definizioni e concetti	<ul style="list-style-type: none"> - relazione delle definizioni e dei concetti adottati con gli obiettivi - problematiche riguardanti il passaggio dalle definizioni teoriche all'applicabilità pratica
Grado di confrontabilità	<ul style="list-style-type: none"> - confronti fra i dati dell'indagine - confronti tra i dati raccolti e quelli disponibili da altre fonti (problema della continuità delle serie storiche)
Classificazioni	<ul style="list-style-type: none"> - standard adottati - problemi di riconducibilità ad altri standard
Periodicità e tempestività	<ul style="list-style-type: none"> - pianificazione del disegno di indagine (trasversale e/o longitudinale) - tempo intercorrente fra il periodo di riferimento dei dati e la pubblicazione dei dati
Liste ed archivi	<ul style="list-style-type: none"> - criteri di scelta e descrizione delle liste da utilizzare per identificare la popolazione obiettivo - analisi della completezza e della ridondanza delle liste utilizzate - valutazioni del grado di affidabilità delle informazioni disponibili - calcolo dei pesi di riporto all'universo
Campionamento	<ul style="list-style-type: none"> - disegno di campionamento in relazione alle liste disponibili - problemi di applicabilità del disegno teorico alle situazioni pratiche
Strumenti di raccolta	<ul style="list-style-type: none"> - descrizione degli strumenti usati per la raccolta dei dati presso le unità statistiche - descrizione degli eventuali modelli ausiliari di aiuto alle operazioni di contatto o di ritorno sul campo
Rilevazione	<ul style="list-style-type: none"> - organizzazione logistica - interazione fra le strutture coinvolte nelle operazioni - piano dei solleciti
Acquisizione dati	<ul style="list-style-type: none"> - procedure informatiche usate per la registrazione su supporto magnetico - procedure implementate per la codifica dei quesiti aperti - strumenti utilizzati per il controllo dell'errore

Elementi	Dimensioni
Revisione dei dati	<ul style="list-style-type: none"> - identificazione ed imputazione degli errori - valutazione dell'impatto del piano di correzione sulle stime finali
Costi attesi e vincoli di implementazione	<ul style="list-style-type: none"> - piano finanziario per le attività di produzione del dato in senso stretto - piano finanziario per le attività di controllo della qualità dei dati (con analisi dei vincoli posti sull'implementazione delle strategie ottimali in relazione alla qualità ottenibile applicando strategie sub-ottimali)
Misure della qualità	<ul style="list-style-type: none"> - indagini pilota - indagini di controllo - indicatori di qualità del processo di produzione
Diffusione dei dati	<ul style="list-style-type: none"> - modalità - supporti - strategie per la tutela della riservatezza

Definizione dei contenuti informativi



Definizione degli obiettivi

Questa prima fase del processo produttivo dei dati è molto complessa ed articolata poiché deve definire in modo preciso le esigenze conoscitive ed individuare, seppure in forma generica, gli obiettivi che sottendono l'avvio del processo di produzione del dato.

È, però, necessario delimitare, senza ambiguità, cosa interessa e cosa non interessa, ricordando che: più è ampio è lo spettro degli argomenti trattati maggiori divengono le complessità da affrontare sul piano concettuale, statistico ed operativo.

Analisi del fenomeno

È la prima vera fase del processo produttivo dei dati.

Si tratta di definire, con la maggiore precisione possibile, i contorni del fenomeno d'interesse, entrando nel merito ed affrontandone gli aspetti specifici.

Un'analisi insufficiente e superficiale compromette la completezza e l'effettiva utilizzazione delle informazioni

Esempio:

Studio delle cause del calo dei livelli di fecondità

- estendere lo studio a tutte le donne in età feconda o limitare l'analisi alle donne appartenenti alle fasce più giovani
- differenziare l'analisi per numero di figli già avuti
- limitare la raccolta dei dati alle donne coniugate
- considerare la coppia nel suo insieme e/o disgiuntamente
- individuare le cause che si ritengono influenti sul fenomeno (demografiche e sociali)

Un ruolo fondamentale è svolto in questa fase dall'esperto di settore che oltre ad effettuare l'analisi del fenomeno allo studio contribuisce all'individuazione dei canali privilegiati tramite i quali ottenere altre eventuali informazioni per il completamento della fase di analisi.

In generale, sarà molto utile l'analisi di esperienze maturate nell'ambito di indagini relative a fenomeni della stessa natura di quello allo studio.

Estremamente importante è acquisire tutta la documentazione statistica disponibile sui fenomeni che risultano affini a quello in esame e per i quali esistono dati statistici relativi a variabili non specifiche del fenomeno ma comunque su di esso presumibilmente influenti e pertanto da utilizzare anche nel processo produttivo che sta nascendo.

Mettendo opportunamente in relazione i dati statistici del maggior numero di variabili che si ritengono connesse al fenomeno allo studio è possibile definirne in maniera chiara i contorni e trarre

indicazioni utili ai fini sia dell'analisi sia dell'individuazione di ulteriori aspetti relativamente ai quali approfondire il lavoro di documentazione.

Esempio:

Studio delle cause del calo dei livelli di fecondità

- serie temporali e territoriali degli indici di fecondità totale e dei quozienti specifici per età
- dati censuari per ottenere informazioni sulle caratteristiche socio-professionali e sulla distribuzione territoriale delle madri e/o delle coppie da incrociare con il numero di figli già avuti e con i valori dei quozienti specifici per età
- indagine sui comportamenti familiari, in particolare quelli che riguardano la donna e la famiglia
- statistiche demografiche tratti dalle schede di nascita sui nati vivi per diverse caratteristiche del padre e della madre

Un'altra utile strategia per effettuare una completa analisi del fenomeno oggetto di studio è quella di ricorrere a metodi di sondaggio diretto (contatti epistolari, riunioni, interviste) presso i supposti potenziali utenti dei risultati dell'indagine che si vuole impostare o presso operatori specializzati del settore interessato che non occorre coinvolgere stabilmente in attività di consulenza (**testimoni privilegiati**).

Il ricorso a tale forma di sondaggio ha il vantaggio di fare emergere l'eventuale necessità di approfondire particolari aspetti del fenomeno o di prendere in esame elementi non emersi nella prima fase di definizione delle esigenze conoscitive.

Le modalità da utilizzare per l'individuazione dell'utenza dipendono dal tipo di indagine e di fenomeno allo studio. In generale, sono da consultare eventuali archivi dei fruitori di dati statistici e comunque anche in questo caso risulta essenziale il contributo dell'esperto di settore.

Conclusioni

Una definizione dettagliata e circostanziata dei particolari aspetti del fenomeno che interessano permette di tradurre in termini statistici il problema e consente, quindi, di individuare in modo univoco:

- Ø il collettivo o i collettivi da coinvolgere (campo di osservazione);
- Ø le unità di analisi e le unità di rilevazione interessate;
- Ø le variabili che a priori vengono ritenute influenti ai fini della manifestazione del fenomeno.

Campo di osservazione

Definizione: il campo di osservazione è costituito dal collettivo (o collettivi) d'interesse all'interno del quale si manifesta il fenomeno.

Esempio di indagine sulle condizioni abitative delle famiglie:

Il campo di osservazione potrebbe essere individuato dalle famiglie ed anche dalle abitazioni occupate dalle famiglie

Una volta stabilito il campo di osservazione del fenomeno è possibile determinare la o le unità di analisi elementari sulle quali costruire il processo produttivo ed individuare l'unità di rilevazione.

In questa fase del processo di pianificazione della produzione dei dati prevalgono le valutazioni di carattere tecnico-statistico. Le scelte operate condizioneranno pesantemente l'organizzazione della rilevazione. Pertanto le decisioni prese devono essere realistiche e compatibili in relazione alle effettive potenzialità della rete di rilevazione (elenchi anagrafici non informatizzati e/o non aggiornati).

Selezione delle variabili significative

La fase di individuazione degli elementi sui quali fondare il processo produttivo dei dati si conclude con la selezione delle variabili significative ai fini dell'efficace descrizione del fenomeno.

In sostanza si tratta di specificare le *variabili descrittive* e *specifiche* del fenomeno e le *variabili connesse* al fenomeno che lo influenzano e sono da esso influenzate.

Nel primo caso la scelta delle variabili descrittive sarà guidata dalla necessità di mettere in luce le modalità di manifestazione del fenomeno; nel secondo caso la decisione su quali variabili considerare sarà condizionata dalle ipotesi formulate su quelle che si ritiene possano essere le cause e/o gli effetti connessi al fenomeno.

Esempio Indagine sui nuovi comportamenti educativi:

Unità di analisi: singolo genitore

Variabili descrittive

- § tempo dedicato, effettivamente e personalmente, alla cura del figlio nell'arco della giornata
- § modalità con cui tale cura viene esercitata
- § scala per misurare l'atteggiamento che caratterizza il contatto con il figlio

Variabili connesse

- § sesso, età, stato civile
- § titolo di studio
- § professione
- § posizione nella professione

Considerazioni conclusive

La definizione degli elementi che costituiscono le basi del processo produttivo dei dati è caratterizzata da operazioni fortemente legate fra loro. Averle legate nel diagramma con una rigida successione è puramente indicativo. Infatti, sebbene le fasi di analisi del fenomeno, individuazione del campo di osservazione e di selezione delle variabili siano logicamente distinte e successive è ovvio che l'intero ciclo può riavviarsi in presenza di nuovi elementi emersi ad esempio nella fase di selezione delle variabili.

Suggerimenti

§ Ricorrere al quadro di riferimento ottenuto nella fase di analisi del fenomeno per affrontare correttamente il problema di individuazione delle variabili descrittive (una inadeguata definizione di tali variabili renderà incompleta la raffigurazione finale del fenomeno). Inoltre, sarà più agevole rendere esplicite le variabili che rappresentano le forme di manifestazione del fenomeno se il quadro di riferimento è chiaro ed esaustivo.

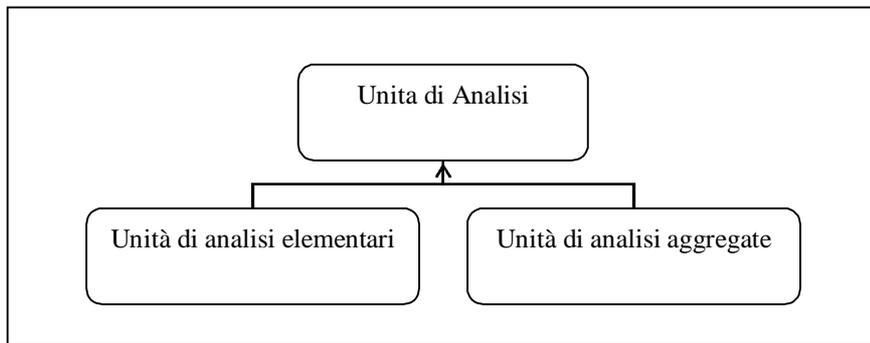
§ Il quadro di riferimento gioca un ruolo fondamentale anche nella scelta delle variabili connesse così come il ricorso alla documentazione statistica acquisita in precedenza (una erronea selezione di tali variabili provocherà difficoltà molto serie nell'interpretazione dei risultati con il rischio di renderli inutilizzabili).

Analisi statistica

Definizione: le *unità di analisi* rappresentano le entità elementari di un particolare collettivo di riferimento (individui, famiglie, abitazioni, ecc.) alle quali viene *associato* un particolare insieme di determinazioni delle variabili prese in esame.

Definizione: le *unità di analisi elementari* sono unità di analisi cui corrispondono determinazioni che hanno un riscontro diretto sul questionario (l'informazione è reperibile sul questionario).

Definizione: le *unità di analisi aggregate* sono a tutti gli effetti delle unità di analisi che però devono essere costruite, a partire dai dati relativi alle unità elementari, mediante procedimenti di riaggregazione (l'informazione in questo caso non è direttamente reperibile sul questionario).



Osservazioni

- E' con riferimento all'unità di analisi elementare che il fenomeno si manifesta e pertanto esse rappresentano l'entità rispetto alla quale si andranno a rilevare le informazioni di interesse
- L'attribuzione di una unità di analisi alla categoria "elementare" o "aggregata" dipende sia dalla natura intrinseca dell'unità (il componente di una famiglia, per sua natura può essere soltanto unità elementare poiché non può essere disaggregabile in alcun'altra unità di analisi) sia dalle scelte effettuate nella fase di impostazione dell'indagine

Definizione: *l'unità di rilevazione* è l'elemento strumentale presso il quale vengono assunte, e conseguentemente riportate sul questionario, le informazioni elementari relative alle unità di analisi.

Osservazioni

- L'unità di rilevazione non sempre corrisponde all'unità di analisi e normalmente è unica per ogni rilevazione (nel caso del censimento si hanno tre distinte unità di rilevazione: famiglie, convivenze, abitazione non occupata)
- La funzione dell'unità di rilevazione si esaurisce nel momento in cui l'attività di rilevazione è terminata
- L'individuazione delle unità di rilevazione può comportare problemi di carattere organizzativo (localizzazione sul territorio o contatto con essa, sensibilizzazione dei rispondenti, ecc.) e di tipo tecnico-statistico (estrazione o individuazione degli archivi, preparazione degli elenchi, sostituzione delle unità non contattate, ecc.)
- La risoluzione dei problemi connessi all'individuazione dell'unità di rilevazione non è univoca (dipende, infatti, dalla natura dell'unità stessa e/o dalla natura dell'indagine. L'organizzazione di un'indagine presso le famiglie è completamente diversa da quella di un'indagine presso le aziende agricole, le imprese o le scuole, ecc.)

Esempi:

1. Per un'indagine sulle aspettative professionali degli studenti iscritti all'università, lo studente è contemporaneamente *unità di analisi elementare e di rilevazione*. L'*unità di analisi aggregata* potrebbe essere il corso di laurea frequentato, la facoltà o l'ateneo di iscrizione (purché sia stato previsto per ciascuno studente di registrare l'informazione su cdl, facoltà ed ateneo).
2. In un'indagine sui consumi in cui il questionario, composto da fogli individuali (*uno per ciascun componente*), viene distribuito alle famiglie, l'*unità di rilevazione* è la famiglia. I singoli componenti sono, invece, le *unità di analisi elementari* dell'indagine. A partire dalle informazioni sui singoli componenti, sarà possibile costruire le *unità di analisi aggregate* relative alla famiglia o al comune, purché ad ogni componente sia associato il codice della famiglia e del comune.
3. Se in un'indagine sui consumi o sulle condizioni abitative delle famiglie italiane l'*unità di rilevazione* è la famiglia ed il questionario è composto esclusivamente da un foglio familiare non sarà possibile registrare notizie attribuibili singolarmente a ciascun componente (che, quindi, non sono *unità di analisi*).
4. Se in un'indagine sulle condizioni abitative delle famiglie italiane l'*unità di rilevazione* è la famiglia e nel questionario (foglio familiare) vengono richieste anche notizie sull'abitazione occupata dalla famiglia è come se avessimo configurato due distinte *unità di analisi elementari*: la famiglia e l'abitazione occupata.

N.B.: Quando ci si trova in presenza di più unità di analisi va posta particolare attenzione all'identificazione del campo di osservazione di ciascuna unità (in questo esempio i collettivi di riferimento sono la popolazione delle famiglie e quello delle abitazioni occupate dalle famiglie).

In questa fase è opportuno effettuare una vera e propria radiografia di ciascuna variabile al fine di stabilire con esattezza natura e caratteristiche, di valutare se e come possa essere misurata con riferimento alle singole unità di analisi ed in che modo, infine, sfruttare dal punto di vista operativo le conclusioni raggiunte.

In particolare bisogna prendere in esame la:

- scala di misura
- tipologia
- classificazione da adottare.

Variabili:

Il termine variabile è spesso sinonimo di **carattere**.

Variabili qualitative o mutabili:

Variabili le cui modalità non assumono valori numerici ma qualità o attributi.

Es. Sesso; stato civile; attività economica; natura della lesione; agenti materiali della lesione.

Variabili quantitative

Variabili le cui modalità assumono valori numerici.

Es. Età; data dell'infortunio; giornate di inabilità; grado percentuale dei postumi permanenti accertati; ammontare complessivo dell'indennità pagata.

Queste a loro volta si dividono in :

Variabili discrete

Le variabili discrete si manifestano in un numero limitato di modalità ed è sempre possibile per esse trovare due modalità distinte all'interno delle quali non ne esiste alcuna altra possibile.

Variabili continue

Le variabili continue possono assumere tutti i valori numerici che rappresentano punti successivi lungo un determinato continuum.

Variabili ad intervalli

Esiste una unità di misura ma non esiste uno zero assoluto cioè le modalità sono dei valori che partono da un'origine arbitraria e pertanto è possibile calcolare la differenza fra due diverse modalità ma non ha senso effettuare il rapporto fra le stesse (es.: la temperatura misurata con un termometro).

Variabili cardinali

le modalità sono valori valutati da un'origine che rappresenta la nullità del fenomeno osservato pertanto le misurazioni di una variabile cardinale sono rapportabili tra loro (es.: reddito, età).

Altre classi di variabili sono:

Variabili temporali

Variabili le cui modalità entrano in relazione con il tempo e riguardo a questo aspetto possono perciò considerarsi:

- invarianti da un certo momento in poi (luogo di nascita; età al primo matrimonio; anno di nascita);
- di stato. Le variabili assumono le rispettive modalità con riferimento ad un determinato istante di tempo (età; stato civile; luogo di residenza);
- di flusso. Le modalità di tali variabili vengono individuate facendo riferimento ad un intervallo di tempo (consumi; reddito; numero degli spostamenti per motivi di lavoro; ammontare mensile dei nati).

Variabili territoriali

Associano all'unità di analisi un riferimento sul territorio. Tale riferimento territoriale viene definito di norma da un'associazione codice-luogo e contribuisce a:

- identificare l'unità di rilevazione sul territorio;
- descrivere l'unità di analisi al pari di qualsiasi altra variabile non territoriale (luogo di lavoro; luogo di nascita; luogo di residenza);
- ricomporre le unità di analisi elementari.

Variabili generiche (o standard o strutturali)

Si definiscono generiche quelle variabili ricorrenti nell'ambito di determinate tipologie di indagini e per le quali è quindi opportuno tenere conto dei vincoli di omogeneità e di standardizzazione introdotti relativamente alle corrispondenti classificazioni. (Sono molto importanti nella definizione di sistemi statistici).

Esempi:

Indagini sulle famiglie:

sex; età; stato civile; luogo di residenza; la professione; la condizione professionale, ecc.

Indagini sugli studenti:

Corso di laurea, titolo di studio, in sede o fuori sede, voto di diploma, ecc.

Variabili specifiche

Le variabili che sono inerenti in misura più o meno stretta al fenomeno oggetto di studio si categorizzano come specifiche. Le variabili di questa natura sono, in generale, meno soggette a vincoli sulle classificazioni da adottare ed in certi casi sono così specifiche che il tipo di

classificazione adottato può anche essere studiato ed impostato *ex-novo* (non è il caso delle classificazioni delle malattie e cause di morte o dei settori produttivi).

Esempi

1. Indagine sulle vacanze:
località prescelta;
2. Indagine sulle forze lavoro:
Numero di ore lavorate;
3. Indagine sul tempo libero:
Tipo di svago prescelto;
4. Indagine sull'uso dei servizi bibliotecari:
Tipologia di argomenti dei testi, i testi più frequentemente consultati, n° di postazioni internet, orari di apertura.

Variabili di sommario

Variabili quantitative discrete di tipo specifico. Se ne distinguono di due tipi:

1. con riscontro diretto sul questionario;
2. ricavate mediante elaborazioni.

Esempi del tipo 1:

- a. Indagine che prevede come unità di analisi la famiglia (foglio familiare)
numero componenti; numero occupati; numero di maschi
- b. Indagine che prevede come unità di analisi l'abitazione
numero stanze; numero di bagni; numero balconi

Esempi del tipo 2:

- c. Indagine che prevede come unità di analisi la famiglia (con foglio individuale)
numero componenti con oltre 65 anni; numero laureati
- d. Indagine che prevede come unità di analisi l'abitazione (con scheda per vano)
numero stanze con balconi; numero bagni con vasca

Classificazione

Definizione: Insieme delle modalità che la variabile assume.

Il caso che prendiamo in esame è quello in cui la classificazione si riferisce al livello di massimo dettaglio con cui la variabile viene espressa cioè alle modalità le cui determinazioni trovano riscontro diretto sul questionario.

I criteri in base ai quali ad una variabile viene associata una certa classificazione dipendono da:

1. la scala di misura della variabile;

2. il grado di articolazione in modalità ritenuto necessario, con riguardo al fenomeno, e richiesto nell'ambito delle esigenze conoscitive precedentemente espresse;
3. eventuale esistenza di vincoli di omogeneità e standardizzazione (nel caso di variabili generiche o standard) e comparabilità anche nel tempo.

L'elemento classificazione interviene in due momenti distinti all'interno del processo di produttivo. Un primo momento fa riferimento alla definizione della classificazione ed alla articolazione in modalità possibili. Tale operazione ha il fine di consentire la predisposizione del questionario e di rendere esplicito il tipo di relazione da instaurare con l'unità di analisi.

Il secondo momento, operativamente successivo e distinto, è quello in cui si procede all'aggregazione dei dati sulla base della classificazione (massima o ridotta, cioè ottenuta per ricomposizione della classificazione massima o meglio attraverso il raggruppamento in classi delle modalità originarie). Nell'operare l'aggregazione si rende spesso necessario il raggruppamento in classi delle modalità originarie (riduciamo cioè la classificazione).

Tale operazione, indispensabile per le variabili quantitative continue, può non essere necessaria per le variabili qualitative e per quelle quantitative discrete.

Esempio:

Lo stato civile non necessita di norma aggregazioni di modalità, ciò dipende naturalmente dallo scarso numero di modalità in cui si articola la variabile. Al contrario la professione viene rilevata al massimo livello di articolazione mentre in fase di aggregazione viene solitamente evidenziata a livello più o meno aggregato.

Avvertenze

Sebbene le due fasi descritte siano operativamente distinte non va sottovalutato che il buon esito della seconda fase dipende fortemente da quanto è stato deciso nella prima fase. È infatti evidente che le classi che possono essere create nell'operazione di aggregazione sono vincolate al livello di articolazione massima individuato nella prima fase.

La scelta di associare ad una variabile una certa classificazione può essere vincolata dalla:

- esistenza di standard internazionali di classificazione;
- esigenza di comparabilità fra i dati nel tempo;
- necessità di rendere comparabili variabili rilevate con indagini differenti.

*N.B. Non sempre è necessario attenersi in maniera rigida alle classificazioni internazionali, l'importante è aver cura che la classificazione prescelta sia comunque ragguagliabile a quella standard. Analoga osservazione vale per analisi di serie storiche, se si decide di modificare nel tempo la classificazione è sufficiente garantire la ricostruzione della classificazione secondo i nuovi criteri.
Inoltre quando si rapporta una classificazione con un'altra si parla di "mappatura"*

Definire una classificazione è un momento particolarmente critico e per alcune variabili, ad esempio quelle che misurano il gradimento, la scelta di una scala a quattro categorie anziché a cinque può fornire risultati addirittura opposti. Analogo effetto si può determinare denominando in modo leggermente diverso le stesse cinque categorie. È, quindi, opportuno valutare con attenzione il tipo di scala da utilizzare, soprattutto se si desidera confrontare i risultati dell'indagine con altre fonti di informazione disponibile.

Se esistono, per le variabili di interesse, classificazioni standard riconosciute a livello internazionale conviene utilizzarle specialmente se le variabili sono particolarmente complesse (attività economiche; professioni; malattie; cause di morte).

Per verificare se scale differenti portano a risultati uguali si effettuano due studi paralleli e quindi si confrontano i risultati.

Conclusioni

L'attività di analisi delle variabili che determinano la configurazione del fenomeno, e delle classificazioni ad esse associate, assume una importanza centrale per tre ragioni principali:

1. consente di definire le modalità di ciascuna variabile e quindi di poter associare alla singola unità di analisi non solo il cosa rilevare (le variabili), ma anche come ciascuna variabile si manifesta nella singola unità (definire una classificazione);
2. mette a disposizione gli elementi per costruire il questionario;
3. garantisce la possibilità di mettere a punto le procedure di aggregazione informatizzate sulla base delle classificazioni prescelte.