

L'acquisizione dei dati

Il ricercatore che voglia esprimere un giudizio quantitativo su certi aspetti di un fenomeno collettivo si può trovare di fronte a due situazioni:

**i dati statistici sono
disponibili
(fonti)**

**i dati statistici
occorrenti non
sono disponibili**



**Raccolta dei dati
(rilevazione ad hoc)**

ALCUNE FONTI STATISTICHE “UFFICIALI”

dati reperibili presso enti o organismi istituzionalmente preposti alla raccolta e diffusione dell'informazione

Enti del SISTAN

(Istat, Uffici di statistica di regioni, province, comuni, Uffici di statistica dei ministeri, ...)

Organismi internazionali

(EUROSTAT, OECD, ONU, FAO, ecc.)

Dati disponibili su:

- pubblicazioni periodiche
- banche dati
- mezzi di informazione di massa

Come trovare le informazioni:

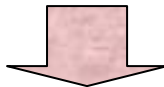
- cataloghi (Es. catalogo SISTAN)
- uffici specializzati (es. Istat)
- Internet
- ...

DATI PROVENIENTI DA FONTI AMMINISTRATIVE

dati reperibili presso enti o organismi che diffondono dati ottenuti come sottoprodotto della loro attività (ministeri, enti territoriali, ecc.). Esempi: registro ditte di Camere di Commercio; archivio bilanci di imprese; archivi Inps; ec

Risorsa ricca, spesso tempestiva, alta frequenza (registri correntemente aggiornati), riferimento territoriale minuto, basso costo

MA



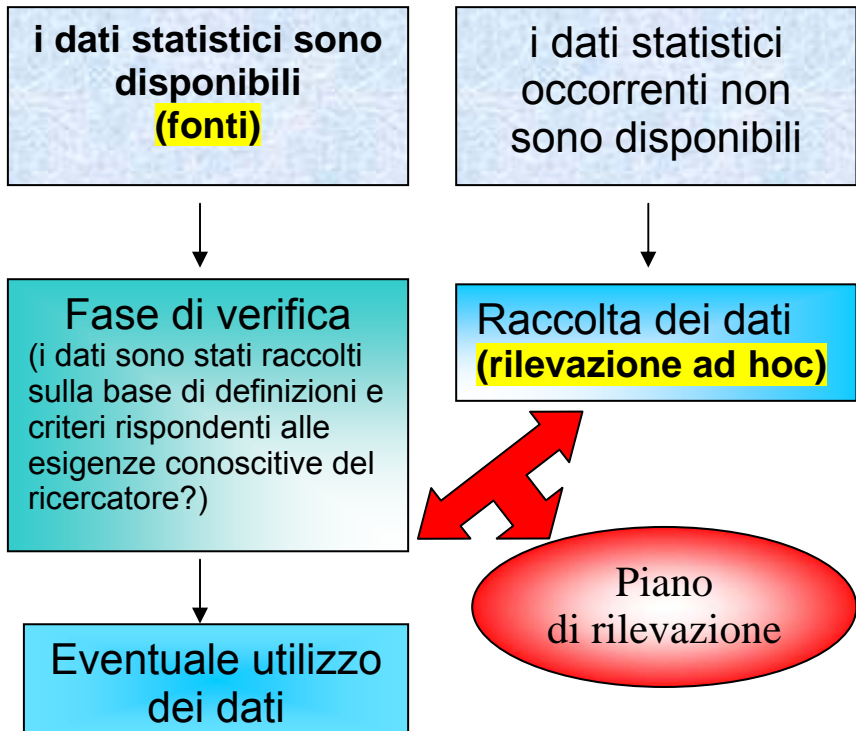
CAUTELA NELL'USO
(verifica della rispondenza rispetto agli obiettivi conoscitivi)

DATI AMMINISTRATIVI

- disponibili solo per la pop. oggetto del 'trattamento' amministrativo (spesso \neq popolazione di interesse)
- possibili problemi di copertura se le finalità amministrative della raccolta creano interesse da parte delle unità a sfuggire (es.: esazione fiscale) o ad introdursi indebitamente (es.: pensioni di invalidità)
- attenzione solo ai dati rilevanti per gli scopi amministrativi (spesso \neq dati rilevanti per l'analisi statistica); es: professione del deceduto.
- concetti, definizioni, classificazioni possono non coincidere con quelli della ricerca statistica (ad esempio una statistica sulle famiglie di fatto, condotta utilizzando dati amministrativi raccolti con riferimento alle famiglie anagrafiche)

Cautele nell'uso dei dati

- Se raccolti da altri occorre verificarne la rispondenza agli obiettivi
- Se indagine ad hoc occorre garantirsi che i dati che si raccoglieranno siano coerenti con gli obiettivi



IL PIANO DI RILEVAZIONE

FASE 1.

- Fissazione degli obiettivi
- Scelta/definizione unità statistiche.
- Scelta dei caratteri da rilevare
- Esplicitazione linguaggio osservativo

FASE 2.

- Definire qual è il tipo di indagine più consono a produrre le statistiche che si desiderano

FASE 3.

- Stabilire estensione dell'indagine

FASE 4.

- Determinazione dei mezzi (supporti fisici) per la raccolta dei dati

FASE 4.

- Verifica del materiale raccolto
- Spoglio e tabulazione dei dati

FASE 1. FISSAZIONE OBIETTIVI

La progettazione e l'esecuzione di una indagine è frutto di un impegno multi disciplinare che coinvolge un elevato numero di professionalità.

Scopo dell'indagine è quello di produrre descrizioni riassuntive di carattere quantitativo, riguardanti il *fenomeno di interesse*.

- precisare cosa interessa e cosa non interessa (più ampio l'arco degli argomenti trattati, maggiori le complessità da affrontare sul piano concettuale, statistico ed operativo.
- Definire se interessa descrivere un fenomeno nella sua componente statica o in quella dinamica.
- Specificare se interessa confrontare i risultati con informazioni relative ad altre realtà territoriali.

FASE 1. DEFINIZIONE DELLA COLLETTIVITÀ

Occorre precisare, senza possibilità di incertezze o equivoci, quali sono le unità che la compongono.

- Elencazione unità
- Criterio di appartenenza

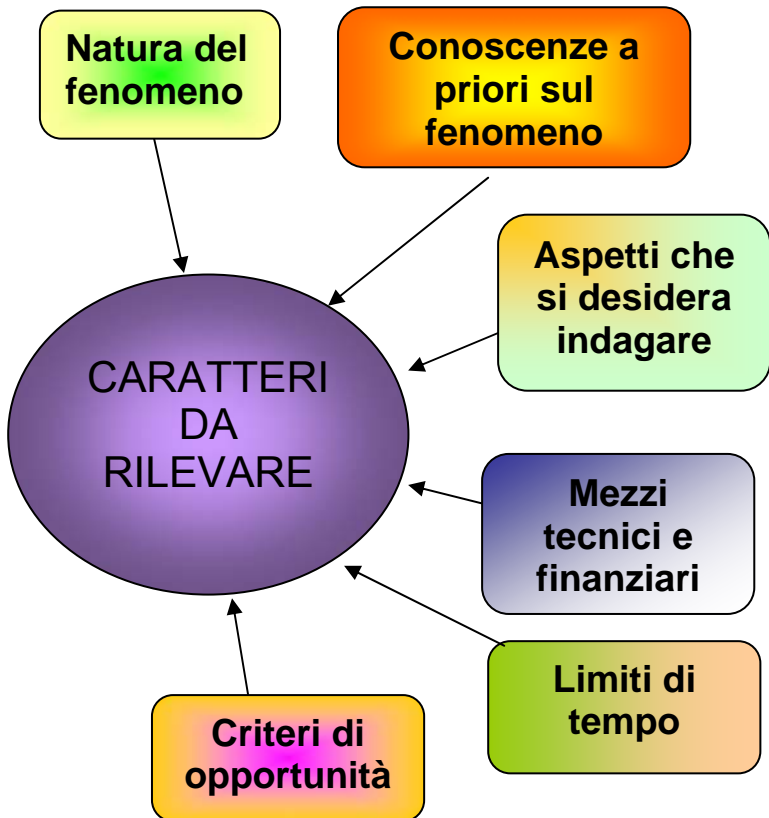
Es.: «insieme delle famiglie dell'area fiorentina»

- significato del termine “famiglia”
- significato di “area fiorentina” (ambito territoriale);
- famiglie residenti nell'area oppure quelle di fatto presenti?;
- data a cui si intende riferirsi (ambito temporale)

OSSERVAZIONE

Talvolta conviene reperire le unità di una collettività per il tramite di unità intermedie, dette *unità di rilevazione*.

FASE 1. SCELTA DEI CARATTERI



Operazione che determina il 'taglio' della rilevazione

Es.: «consumi delle famiglie italiane»
molteplici **aspetti** di possibile interesse:

- ammontare complessivo delle spese per consumi;
- incidenza delle spese per consumi alimentari sul totale;
- relazione tra quantità consumata di un determinato bene e reddito
- ecc.

Possibili **caratteri** rilevanti:

- quantità e valore dei beni e servizi del «bilancio» familiare;
- reddito e dimensione del nucleo familiare;
- comune di residenza della famiglia,
- ecc.

FASE 1. ESPLICITAZIONE DEL LINGUAGGIO OSSERVATIVO

- Si decide lo “strumento di misura” per l’osservazione
- Si definiscono conseguentemente i termini con cui sono espresse le differenti manifestazioni di un carattere (risultati della misura → modalità del carattere)

Un **concetto teorico** (carattere di interesse) viene tradotto in un **concetto operativo** (carattere misurabile)

→ **ATTN**: *Es. grado di istruzione*

CLASSIFICAZIONE CARATTERI IN BASE ALL'ESPRESSIONE DELLE MODALITÀ

C. qualitativi

→ Classificazione

ESPRESSIONE VERBALE

C. quantitativi →

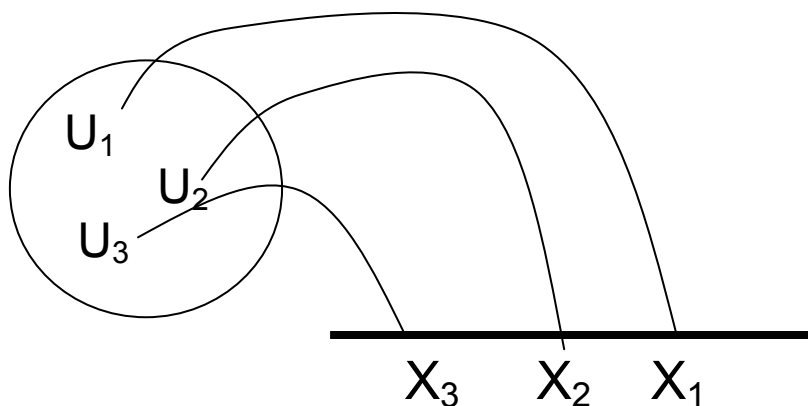
→ misure o conteggi

ESPRESSIONE NUMERICA

NB: *La distinzione dipende dal linguaggio osservativo e non dal carattere in sé*

CLASSIFICAZIONE CARATTERI IN BASE ALLA SCALA DI MISURA

Misura di un carattere = regola che assegna ad ogni unità statistica u_i un numero che esprime la “dose” del carattere x posseduta dall’unità



SCALA = insieme di arrivo

SCALA NOMINALE

confronti possibili

$$x_i = x_j$$

$$x_i \neq x_j$$

→ c. qualitativi sconnessi
(mutabili sconnesse)
es.: sesso, religione, ecc.

NB: nessuna operazione
aritmetica sulle modalità

SCALA ORDINALE

confronti possibili

$$x_i = x_j$$

$$x_i \neq x_j \rightarrow x_i > x_j \text{ oppure } x_i < x_j$$

→ c. qualitativi ordinati
(mutabili rettilinee)
es.: qualifica, giudizi di merito,
ecc.

NB: operazione aritmetiche sulle
modalità?

SCALA DI INTERVALLO

confronti possibili

$$x_i - x_j = x_i - x_j$$

$$x_i - x_j > x_i - x_j$$

$$x_i - x_j < x_i - x_j$$

→ caratteri quantitativi per i quali unità di misura e origine sono arbitrarie

es.: datazione, temperatura, ecc.

SCALA DI RAPPORTO

confronti possibili

$$\frac{x_i}{x_j} \geq \frac{x_k}{x_h} \quad \frac{x_i}{x_j} \leq \frac{x_k}{x_h}$$

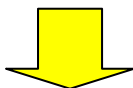
→ caratteri quantitativi per i quali solo l'unità di misura è arbitraria
lo **0** della scala è significativo:
indica assenza del carattere.

es.: misure e conteggi

OSSERVAZIONI SUI CARATTERI QUALITATIVI

Maggiori difficoltà nell'uso di espressioni verbali rispetto ai numeri:

- *trattamento meno agevole*
- *mutabilità del significato nel tempo e nello spazio*



NOMENCLATURE UFFICIALI

+

CODIFICA

(Traduz. modalità in codici numerici)

Es.:

- class. comuni italiani
- class. attività economiche
- class. cause di morte

OSSERVAZIONI SUI CARATTERI QUANTITATIVI

- Un carattere quantitativo può essere sempre trasformato in qualitativo.
 - Una distinzione rilevante
 - C. quantitativi discreti:
modalità come punti isolati
*es.: numero di addetti (numeri naturali), quoziente
mascolinità (numeri razionali)*
 - C. quantitativi continui:
date due modalità è sempre possibile una intermedia
es.: altezza, peso (numeri reali)
- NB:** *le modalità sono sempre espressione di una CLASSE*



- Anche i caratteri discreti possono essere raggruppati in classi.

Utile quando le modalità sono molto numerose rispetto alle determinazioni osservabili
Es. ampiezza demografica comuni

Ovviamente:

maggiore ampiezza classi

→ maggiore sintesi

→ minore informazione

FASE 2. SCELTA DEL TIPO DI INDAGINE (*disegno d'indagine*)

Quali stime interessa produrre?

- Stime di caratteristiche, attività, comportamenti, attitudini in un punto nel tempo
- Stime di caratteristiche basate sull'accumulo di dati nel tempo
- Stime di variazione netta o lorda in due o più punti nel tempo
- Stime di andamenti tendenziali su più periodi temporali
- Stime di durata, transizioni per specifiche tipologie di eventi e specifici sotto-insiemi di popolazione
- Stime di relazioni fra caratteri

a seconda delle informazioni alle quali si è interessati, è necessario fare riferimento a differenti tipi di indagine.

Una prima distinzione

- **indagini trasversali**: si rilevano le u.s. raccogliendo informazioni di interesse riferite ad un particolare momento/periodo di tempo, con l'intento di stimare le caratteristiche riferite allo stato della popolazione oggetto nel momento/periodo di interesse.
- **indagini longitudinali**: l'obiettivo è (principalmente) rivolto a misurare l'evoluzione temporale delle caratteristiche di interesse mediante l'espediente di ricontattare le u.s. per analizzarne i cambiamenti.

NB: con opportune cautele si possono stimare misure di cambiamento con indagini trasversali o misure di stato con indagini longitudinali

Ulteriori specificazioni

- **Indagini occasionali:** pianificate allo scopo di ottenere stime riferite a caratteristiche possedute dalla popolazione in un singolo istante di tempo (es.: distribuzione per età della popolazione in un dato istante) o riferite a un periodo (es.: distribuzione del fatturato realizzato in un anno).

Se nell'indagine occasionale sono raccolte una o più informazioni in comune con altre indagini è possibile ottenere stime di variazione netta.

ATTN: *queste stime di cambiamento possono essere gravemente affette da variazioni indotte dalla diversa tecnica di indagine adottata nei due casi.*

- **Ind. ripetute senza sovrapposizione** fra le unità indagate nelle diverse occasioni: sono spesso chiamate indagini periodiche o ricorrenti. L'indagine viene ripetuta in momenti programmati nel tempo su popolazioni (campioni) diverse
- **Indagini ripetute con una parziale sovrapposizione** delle unità indagate nelle diverse occasioni: sono programmate ad intervalli di tempo regolari con l'uso di *panel* ruotati (le u.s. sono introdotte nel campione, indagate per un prefissato numero di occasioni e quindi escluse (ruotate)).

Lo scopo principale per introdurre una sovrapposizione del campione è quello di ridurre la varianza campionaria delle stime (nessun tentativo di seguire le unità che si muovono o abbinare le unità rispondenti in occasioni diverse per compiere stime longitudinali (cioè di transizione di stato).

- **Ind. longitudinali senza rotazione:**
predisposte con lo scopo di seguire un particolare gruppo di unità nel tempo, e creare un record longitudinale per ogni unità osservata (storie di vita)

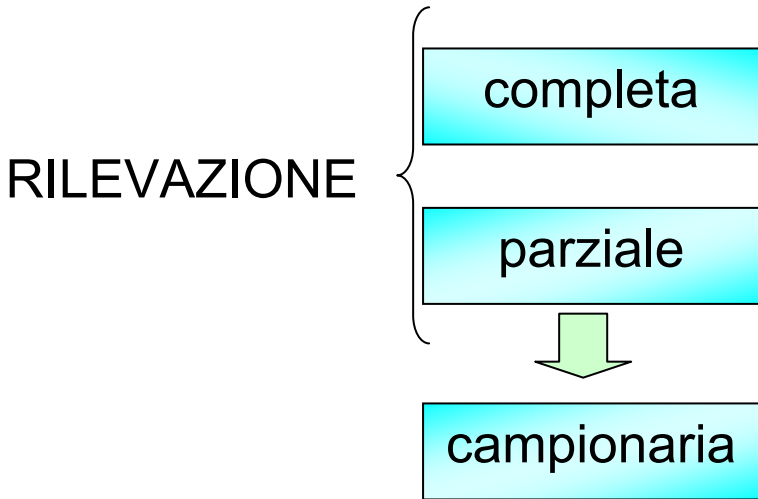
Obiettivo: studiare modificazioni intervenute nel collettivo durante il tempo, utilizzando i cambiamenti avvenuti sui record individuali.

Possibile produrre stime riferite alla sola popolazione di partenza (senza di ingressi di nuove unità, non si riesce a rappresentare eventuali mutamenti nella struttura)

- **Indagini longitudinali con rotazione:**
disegnate per seguire un particolare gruppo di unità per un periodo di tempo, introducendo nuove unità nel campione in occasioni specificate, al fine di creare record longitudinali per ogni unità osservata e produrre analisi longitudinali.

Possibile mantenere il campione stesso rappresentativo della popolazione anche nelle occasioni successive alla prima.

FASE 3. DETERM. L'ESTENSIONE DELL'INDAGINE



Perché un campione?

- Costo
- Tempo
- Impossibilità (popolazioni illimitate, osservazione distruttiva)
- Maggior accuratezza (<dimensioni)

PARLIAMO DI CAMPIONAMENTO

PROBLEMI STRATEGIE

[03_CAMPIONAMENTO.DOC](#)

FASE 3. DETERMINAZIONE DEI MEZZI (SUPPORTI FISICI) PER LA RACCOLTA DEI DATI

modelli di rilevazione

«supporti» fisici che registrano i risultati della rilevazione, per ciascuna u.s. della collettività oggetto di indagine

- Schede
- Questionari cartacei
- Questionari su supporto elettronico
 - CATI
 - CAPI
 - CAWI
 - ...
- Nuove tecnologie (rilevaz. automatiche)

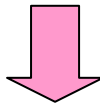
QUESTIONARI qualche considerazione [04 QUESTIONARIO.doc](#)

FASE 4. VERIFICA DEI DATI

VERIFICA:

Analisi dei dati individuali
raccolti per riscontrare:

- omissioni
- incompatibilità
- incongruenze



CONTROLLO DI QUALITÀ:

- imputazioni
- correzione

[05_QUALITA.doc](#)

FASE 4. SPOGLIO E TABULAZIONE

MATRICE DEI DATI

unità	caratteri					
	B	B	...	X	Y	Z
1	a_1	b_1	...	x_1	y_1	z_1
2	a_2	b_2	...	x_2	y_2	z_2
...			...			
i	a_i	b_i	...	x_i	y_i	z_i
...			...			
n	a_n	b_n	...	x_n	y_n	z_n

x_i = modalità del carattere X
osservata sull'unità u_i

SPOGLIO E TABULAZIONE:
classificazione delle unità in
base alle modalità di uno o più
caratteri