

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
necessari sono
disponibili
(fonti)

i dati statistici
necessari non
sono disponibili



Raccolta dei dati
(rilevazione ad hoc)

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, 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; ecc.

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

MA



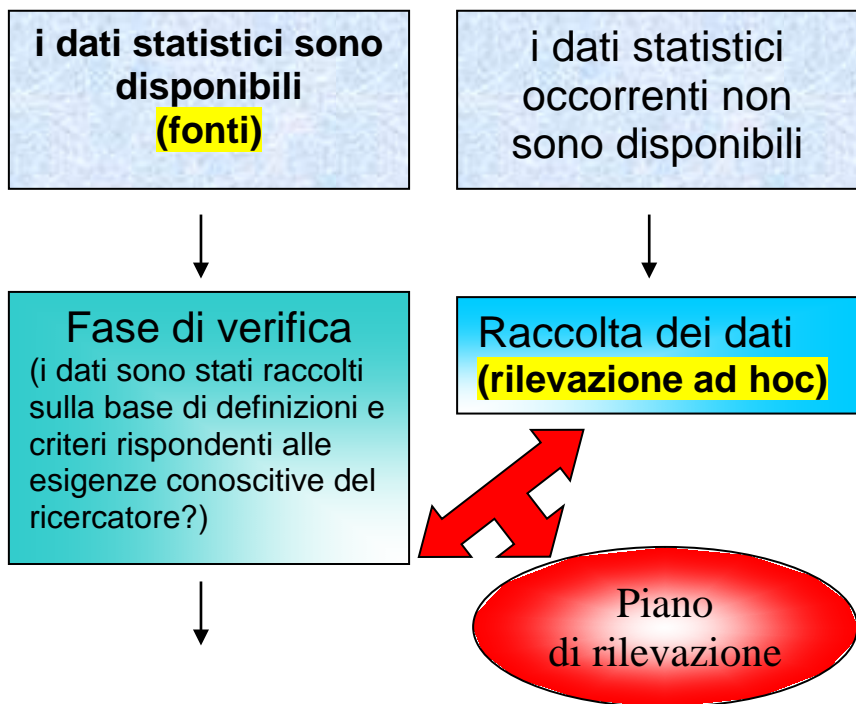
Occorre verificare la rispondenza rispetto agli obiettivi conoscitivi

DATI AMMINISTRATIVI

- disponibili solo per la popolazione oggetto del 'trattamento' amministrativo
non sempre coincide con popolazione di interesse (es: indagine sulla disoccupazione basata sugli iscritti alle liste anagrafiche dei centri per l'impiego)
- possibili problemi di copertura
se le finalità amministrative della raccolta creano interesse 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 dai 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 riferiti alle famiglie anagrafiche

Cautele nell'uso dei dati

- Se raccolti da altri, occorre verificarne la rispondenza agli obiettivi
- Se indagine ad hoc, 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.

- Scelta del tipo di indagine più consono a produrre le statistiche che si desiderano

FASE 3.

- Fissazione dell'estensione dell'indagine

FASE 4.

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

FASE 4.

- Verifica del materiale raccolto
- Spoglio e tabulazione dei dati

FASE 1. FISSAZIONE OBIETTIVI

Progettare e condurre un'indagine è frutto di un impegno multi disciplinare che coinvolge un elevato numero di professionalità.

Scopo dell'indagine è 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. SCELTA DEFINIZIONE DELLE UNITA' STATISTICHE

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

Due alternative

- Elencazione unità
- Criterio di appartenenza

Es.: INDAGINE SULLE «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,*
- *età componenti il nucleo familiare*
- *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**: si usano convenzioni
- *età, peso, altezza, temperatura*
 - *grado di istruzione*
 - *intelligenza*

CLASSIFICAZIONE CARATTERI IN BASE A COME SONO ESPRESSE LE MODALITÀ

C. qualitativi

→ Classificazione

ESPRESSIONE VERBALE

C. quantitativi →

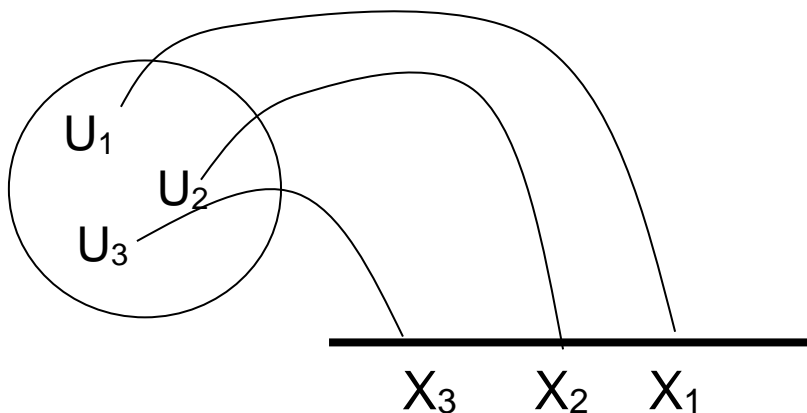
→ misure o conteggi

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

quale scala?

SCALA NOMINALE

confronti possibili tra 2 modalità

$$X_i = X_j$$

$$X_i \neq X_j$$

→ car. qualitativi sconnessi (mutabili sconnesse)

es.: nazionalità, religione, ecc.

NB: NO operazioni aritmetiche sulle modalità

SCALA ORDINALE

confronti possibili tra 2 modalità

$$X_i = X_j$$

$$X_i \neq X_j \rightarrow X_i > X_j \text{ oppure } X_i < X_j$$

→ c. qualitativi ordinati (mutabili rettilinee)

es.: qualifica, giudizi di merito, ecc.

NB: operazione aritmetiche sulle modalità?

SCALA DI INTERVALLO

confronti possibili tra 2 modalità

$$x_i - x_j = x_i - x_j$$

$$x_i - x_j > x_i - x_j$$

$$x_i - x_j < x_i - x_j$$

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

es.: datazione, temperatura, ecc.

SCALA DI RAPPORTO

confronti possibili tra modalità

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

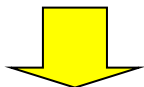
→ caratteri quantitativi (variabili) per i quali solo l'unità di misura è arbitraria. Lo **0** (zero) della scala è significativo: indica assenza del carattere.

es.: misure e conteggi

OSSERVAZIONI SUI CARATTERI QUALITATIVI (scala nominale e intervallo)

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), tasso di mascolinità (numeri razionali)

NB: Quando le modalità sono numerose rispetto alle unità osservate, può convenire raggrupparle in classi. Si guadagna in sintesi e si perde in precisione (es: comuni per ampiezza demografica)

 C. quantitativi **continui**: date due modalità è sempre possibile una intermedia

es.: altezza, peso (numeri reali)

NB: le modalità osservate sono sempre espressione di una CLASSE

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

Quali stime interessa produrre?

- Stime di caratteristiche, attitudini, comportamenti 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 secondo delle informazioni di interesse, è necessario riferirsi a differenti tipi di indagine

Una prima distinzione

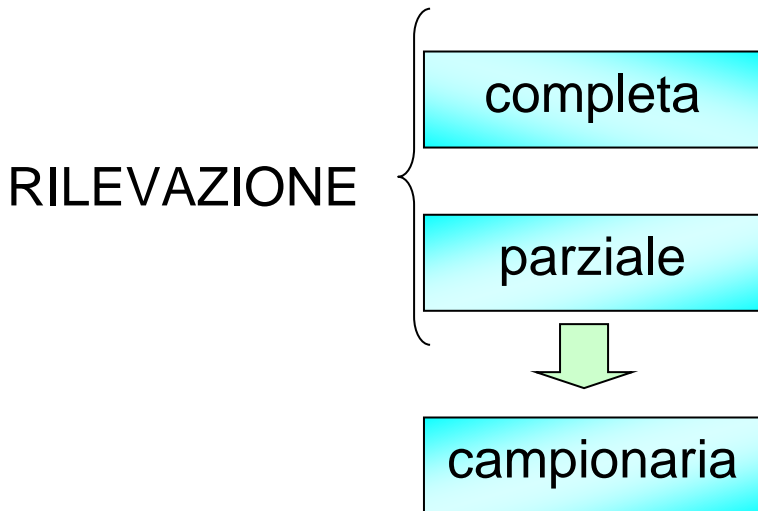
- **indagini trasversali**: si rilevano le u.s. raccogliendo informazioni riferite ad un particolare momento/ periodo di tempo, per stimare caratteristiche riferite a quel momento/periodo.
- **indagini longitudinali**: l'obiettivo è (principalmente) rivolto a misurare l'evoluzione temporale delle caratteristiche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

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

Ulteriori specificazioni

- **Ind. occasionali**
- **Ind. ripetute senza o con parziale sovrapposizione**
- **Ind. longitudinali senza o con rotazione**

FASE 3. SCELTA DELL'ESTENSIONE DELL'INDAGINE



Perché un campione?

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

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)

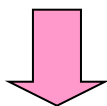
Per considerazioni sulla compilazione vedi il file QUESTIONARIO

FASE 4. VERIFICA DEI DATI

VERIFICA:

Analisi dei dati individuali
raccolti per riscontrare:

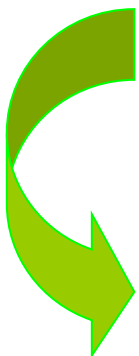
- omissioni
- incompatibilità
- incongruenze



CONTROLLO DI QUALITÀ:

- imputazioni
- correzione

**Per considerazioni sull'argomento
vedi il file QUALITA**



FASE 4. SPOGLIO E TABULAZIONE

MATRICE DEI DATI

unità	caratteri					
	A	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
...
j	a_j	b_j		x_j	y_j	z_j
...
N	a_N	b_N		x_N	y_N	z_N

a_j = modalità del carattere **A** osservata
sull'unità j-esima

x_j = modalità del carattere **X** osservata
sull'unità j-esima

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