

# Il ciclo produttivo dei dati

### ➤ Cosa sono i dati?

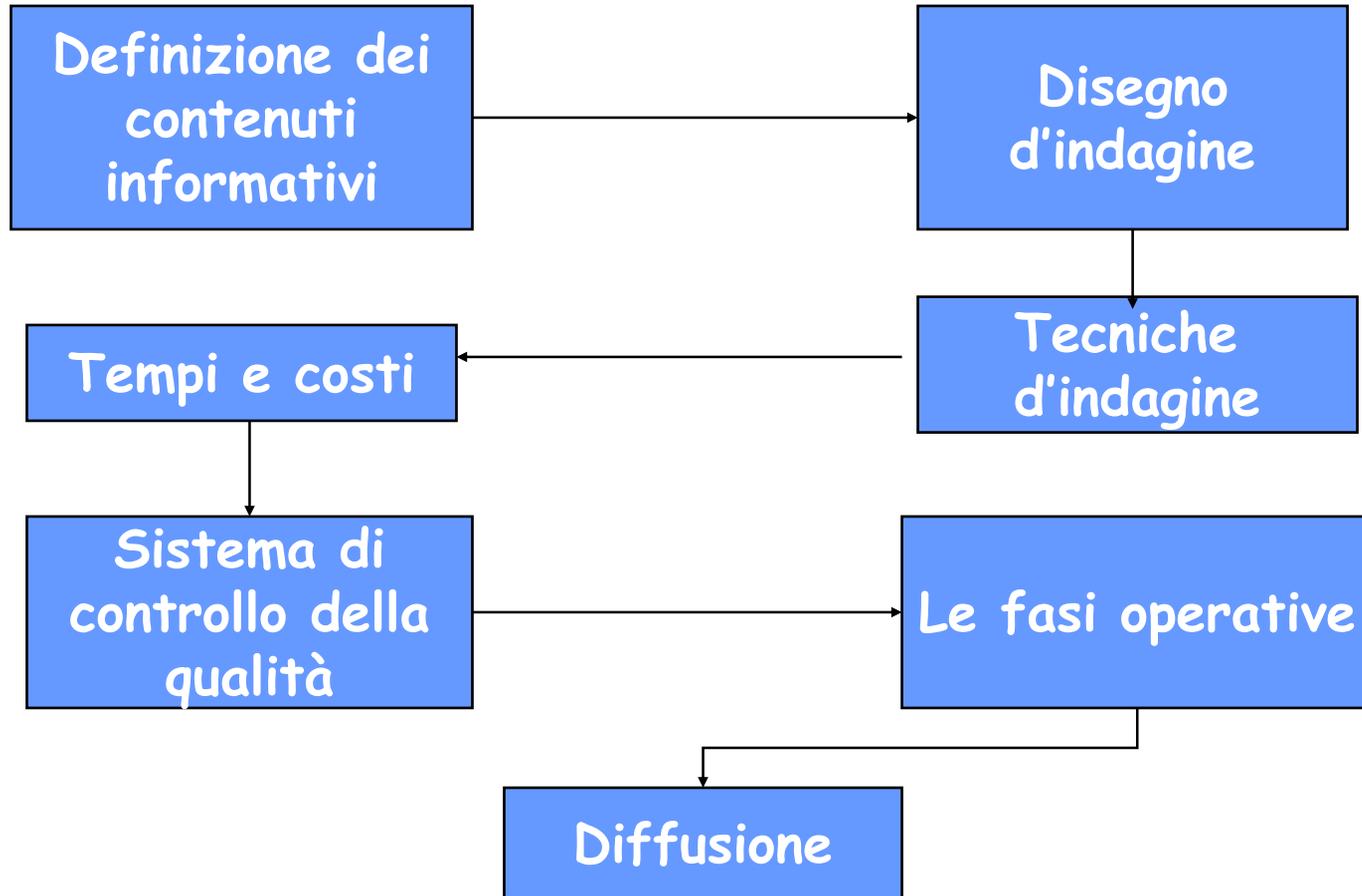
I dati sono dei *prodotti*.

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.

# Costituzione del gruppo di progettazione

Esperto



# Costituzione del gruppo di progettazione

## Requisiti - 1

- 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

# Costituzione del gruppo di progettazione

## Requisiti – 2

- **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.

# Documento di progettazione

<b>Elementi</b>	<b>Dimensioni</b>
<b>Obiettivi</b>	<ul style="list-style-type: none"><li>-contestualizzazione del fenomeno oggetto di indagine</li><li>-analisi delle informazioni già disponibili da altre fonti</li></ul>
<b>Definizioni e concetti</b>	<ul style="list-style-type: none"><li>-relazione delle definizioni e dei concetti adottati con gli obiettivi</li><li>-problematiche riguardanti il passaggio dalle definizioni teoriche all'applicabilità pratica</li></ul>
<b>Grado di confrontabilità</b>	<ul style="list-style-type: none"><li>-confronti fra i dati dell'indagine</li><li>-confronti tra i dati raccolti e quelli disponibili da altre fonti (problema della continuità delle serie storiche)</li></ul>
<b>Classificazioni</b>	<ul style="list-style-type: none"><li>-standard adottati</li><li>-problemi di riconducibilità ad altri standard</li></ul>
<b>Periodicità e tempestività</b>	<ul style="list-style-type: none"><li>-pianificazione del disegno di indagine (trasversale e/o longitudinale)</li><li>-tempo intercorrente fra il periodo di riferimento dei dati e la pubblicazione dei dati</li></ul>

# Documento di progettazione

<b>Elementi</b>	<b>Dimensioni</b>
<b>Liste ed archivi</b>	<ul style="list-style-type: none"><li>-criteri di scelta e descrizione delle liste da utilizzare per identificare la popolazione obiettivo</li><li>-analisi della completezza e della ridondanza delle liste utilizzate</li><li>-valutazioni del grado di affidabilità delle informazioni disponibili</li><li>-calcolo dei pesi di riporto all'universo</li></ul>
<b>Campionamento</b>	<ul style="list-style-type: none"><li>-disegno di campionamento in relazione alle liste disponibili</li><li>-problemi di applicabilità del disegno teorico alle situazioni pratiche</li></ul>
<b>Strumenti di raccolta</b>	<ul style="list-style-type: none"><li>-descrizione degli strumenti usati per la raccolta dei dati presso le unità statistiche</li><li>-descrizione degli eventuali modelli ausiliari di aiuto alle operazioni di contatto o di ritorno sul campo</li></ul>

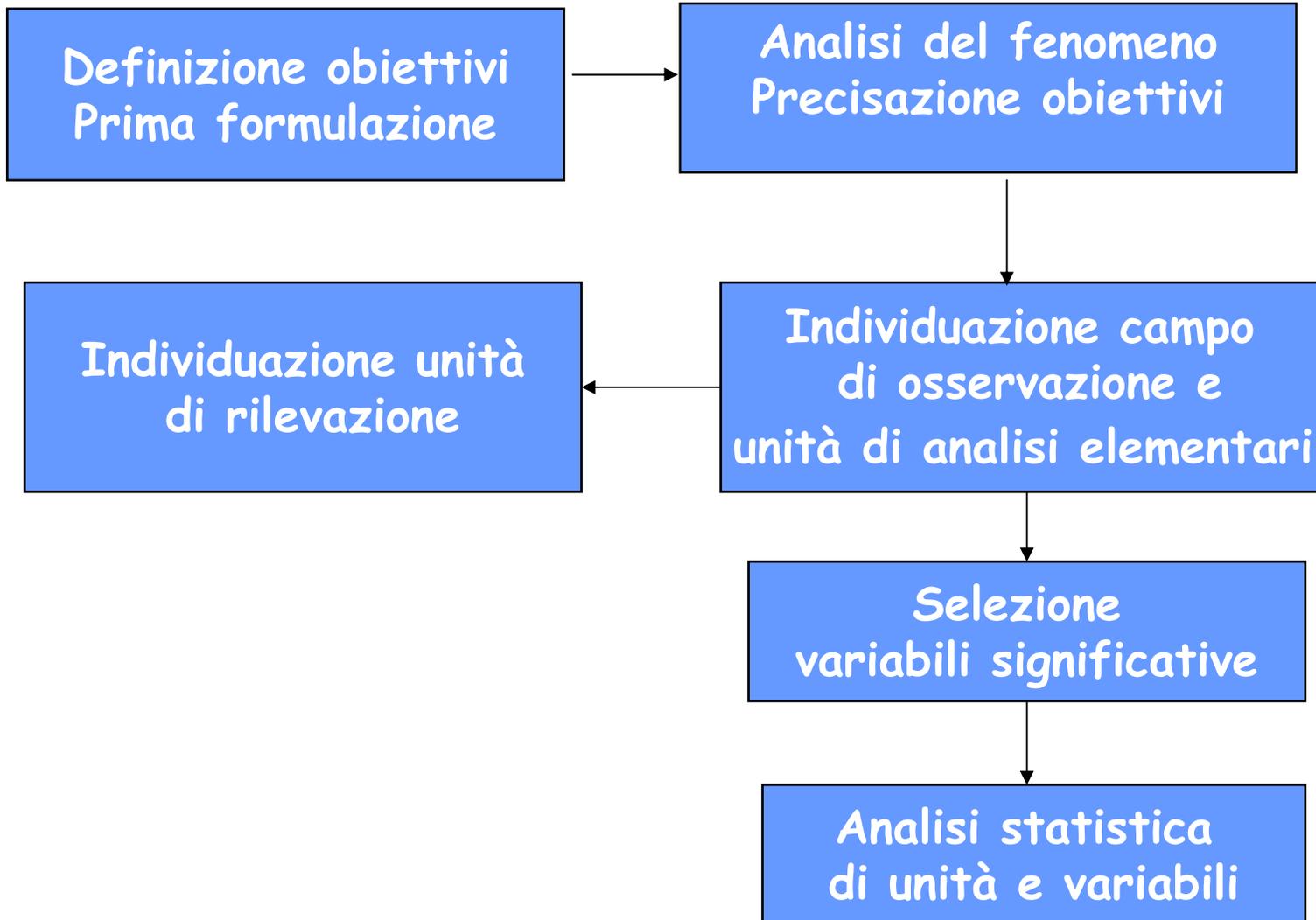
# Documento di progettazione

<b>Elementi</b>	<b>Dimensioni</b>
<b>Rilevazione</b>	<ul style="list-style-type: none"><li>-organizzazione logistica</li><li>-interazione fra le strutture coinvolte nelle operazioni</li><li>-piano dei solleciti</li></ul>
<b>Acquisizione dati</b>	<ul style="list-style-type: none"><li>-procedure informatiche usate per la registrazione su supporto magnetico</li><li>-procedure implementate per la codifica dei quesiti aperti</li><li>-strumenti utilizzati per il controllo dell'errore</li></ul>
<b>Revisione dei dati</b>	<ul style="list-style-type: none"><li>-identificazione ed imputazione degli errori</li><li>-valutazione dell'impatto del piano di correzione sulle stime finali</li></ul>
<b>Costi attesi e vincoli di implementazione</b>	<ul style="list-style-type: none"><li>-piano finanziario per le attività di produzione del dato in senso stretto</li><li>-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)</li></ul>

# Documento di progettazione

<b>Elementi</b>	<b>Dimensioni</b>
<b>Misure della qualità</b>	<ul style="list-style-type: none"><li>-indagini pilota</li><li>-indagini di controllo</li><li>-indicatori di qualità del processo di produzione</li></ul>
<b>Diffusione dei dati</b>	<ul style="list-style-type: none"><li>-modalità</li><li>-supporti</li><li>-strategie per la tutela della riservatezza</li></ul>

# 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)

## Analisi del fenomeno

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.

## Analisi del fenomeno

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

## Analisi del fenomeno

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 la 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:

- il collettivo o i collettivi da coinvolgere (campo di osservazione);
- le unità di analisi e le unità di rilevazione interessate;
- 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

### **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.

## Campo di osservazione

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).

## Le unità statistiche

**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

## Le unità statistiche

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.

# Le unità statistiche

## 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.)

## Le unità statistiche - 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).

## Le unità statistiche - Esempi

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.

## Le unità statistiche - Esempi

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*).

## Le unità statistiche - Esempi

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)*

## 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.

# Selezione delle variabili significative

## 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

# Selezione delle variabili significative

## **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.

# Selezione delle variabili significative

## 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 la 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

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.

## Analisi statistica

### 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.

## 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 alcun'altra possibile.

## Variabili continue

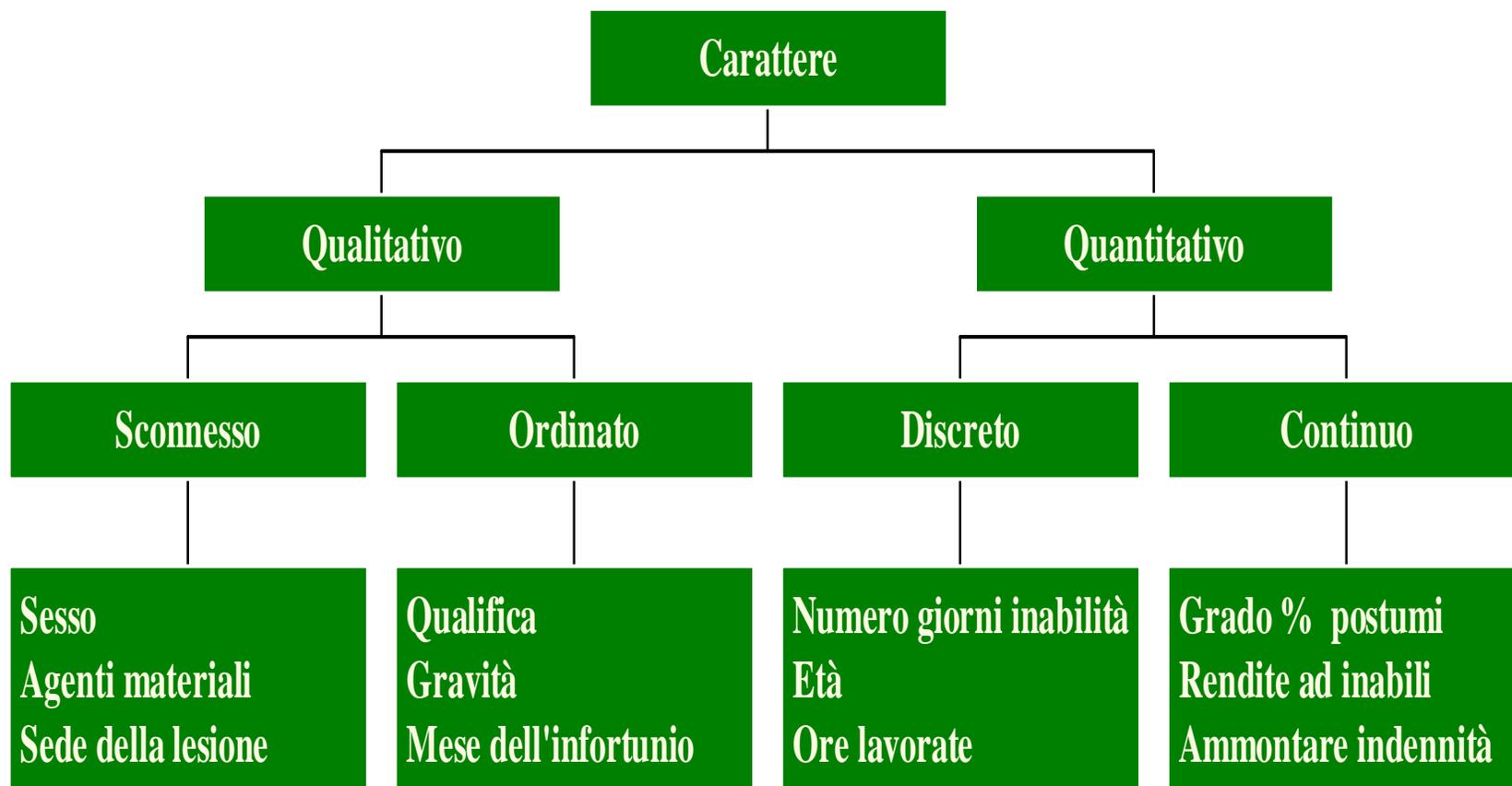
Le variabili continue possono assumere tutti i valori numerici che rappresentano punti successivi lungo un determinato continuum.

## Analisi statistica

**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).

# Classificazione dei caratteri statistici



## 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.

### Esempio

#### **Indagini sulle famiglie**

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

#### **Indagini sugli studenti**

.....

## 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

**Indagine sulle vacanze:** località prescelta

**Indagine sulle forze lavoro:** numero di ore lavorate

**Indagine sul tempo libero:** tipo di svago prescelto

**Indagine sull'uso dei servizi bibliotecari:** .....

## 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

**Indagine che prevede come unità di analisi la famiglia (foglio familiare)**

numero componenti; numero occupati; numero di maschi

**Indagine che prevede come unità di analisi l'abitazione**

numero stanze; numero di bagni; numero balconi

### Esempi del tipo 2

**Indagine che prevede come unità di analisi la famiglia (con foglio individuale)**

numero componenti con oltre 65 anni; numero laureati

**Indagine che prevede come unità di analisi l'abitazione (con scheda per vano)**

numero stanze con balconi; numero bagni con vasca

## Classificazione - 1

**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.

### Classificazione - 2

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).

## Classificazione - 3

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.

### Esempi

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.

## Classificazione - 4

### Avvertenza

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.

## Classificazione - 5

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.*

## Classificazione - 6

### Avvertenze

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).

## 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.

## 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.

# Disegno d'indagine

## 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

# Disegno d'indagine

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.*

# Disegno d'indagine

## **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.

# Disegno d'indagine

## **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).

# Disegno d'indagine

## **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.).

# Disegno d'indagine

## 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 scelto. 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).

Un campione è un sottoinsieme delle unità della popolazione, che la rappresenta con riferimento al problema oggetto di studio.

# Disegno d'indagine

## Archivi di base - 1

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.

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

- incompletezza
- inadeguatezza
- inaccuratezza
- obsolescenza
- presenza di duplicati delle unità

# Disegno d'indagine

## Archivi di base - 2

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)

# Disegno d'indagine

## Archivi di base - 3

### Raccomandazioni

- In fase di progettazione occorre valutare possibili alternative sulla base di quanto in archivio risulta aggiornato e rappresentativo della popolazione obiettivo.
- 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.
- Se disponibili, utilizzare informazioni provenienti da altre indagini che utilizzano gli stessi archivi.
- Verificare la possibilità di ottenere aggiornamenti degli archivi all'epoca di riferimento dell'indagine (ad esempio abbinando archivi indipendenti).

# Disegno d'indagine

## Archivi di base - 4

### Raccomandazioni

- 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).
- 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.

# Disegno d'indagine

## Strategia di campionamento - 1

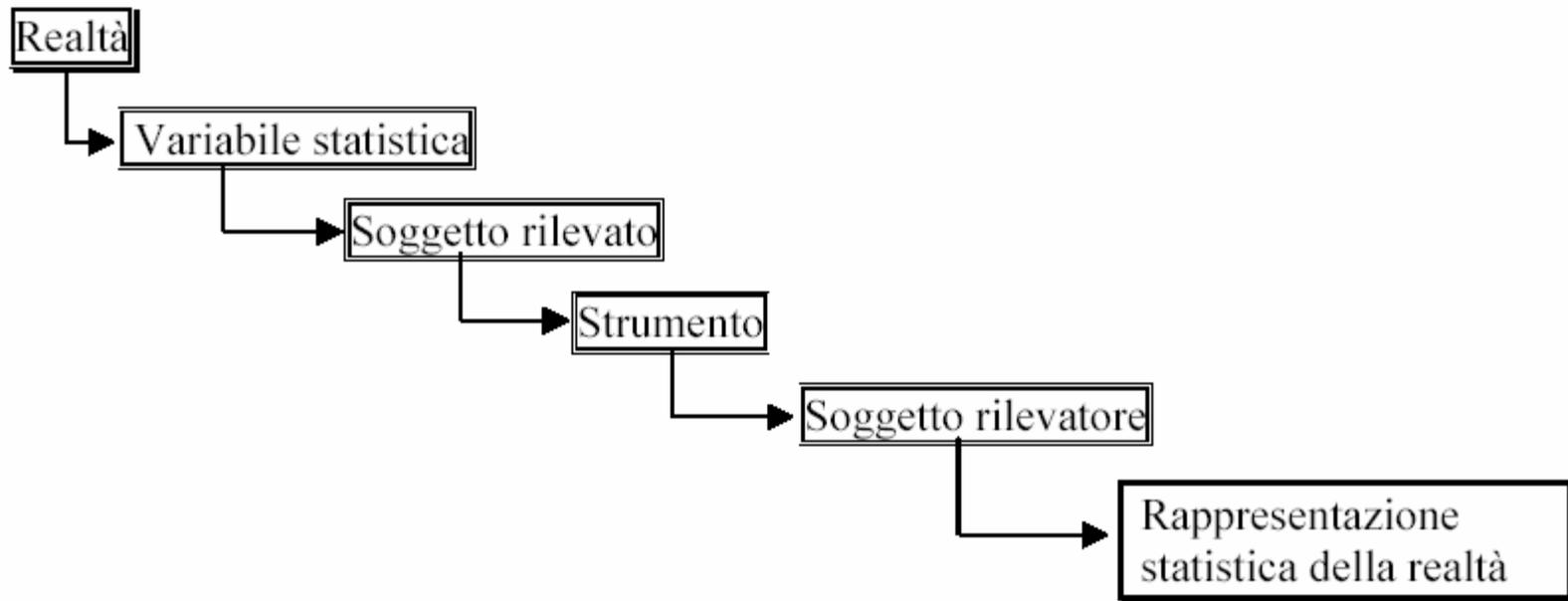
Impostare una strategia di campionamento (piano di campionamento - sampling design) implica la definizione delle **modalità di estrazione** del campione, delle sue **dimensioni** e delle **funzioni dei dati** utilizzate per ottenere, dal campione stime riferite alla popolazione di interesse.

La tecnica di estrazione del campione è scelta anche **tenendo conto delle caratteristiche degli archivi o liste di base disponibili** pertanto occorre valutare attentamente le caratteristiche di tali archivi.

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.

# Disegno d'indagine

## Strategia di campionamento - 2



# Disegno d'indagine

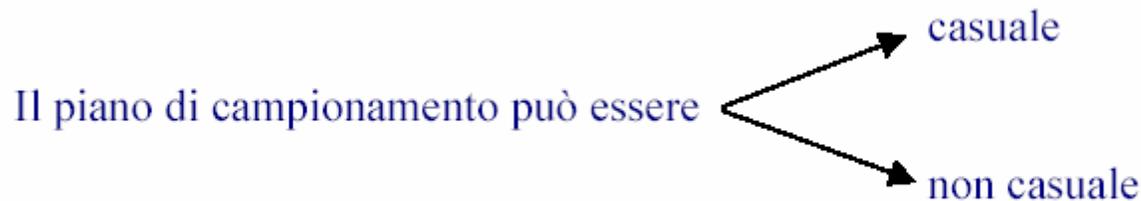
## Strategia di campionamento - 3

### Statistica campionaria

- qualsiasi funzione calcolata sui dati campionari, che non dipende da parametri ignoti
- sono statistiche campionarie, tra l'altro, tutti gli indici descrittivi (media, mediana, varianza, ecc.) quando siano calcolati su un campione piuttosto che sulla popolazione completa

# Disegno d'indagine

## Strategia di campionamento - 4



Un **campione casuale** affida al **caso** la scelta delle unità statistiche da campionare.

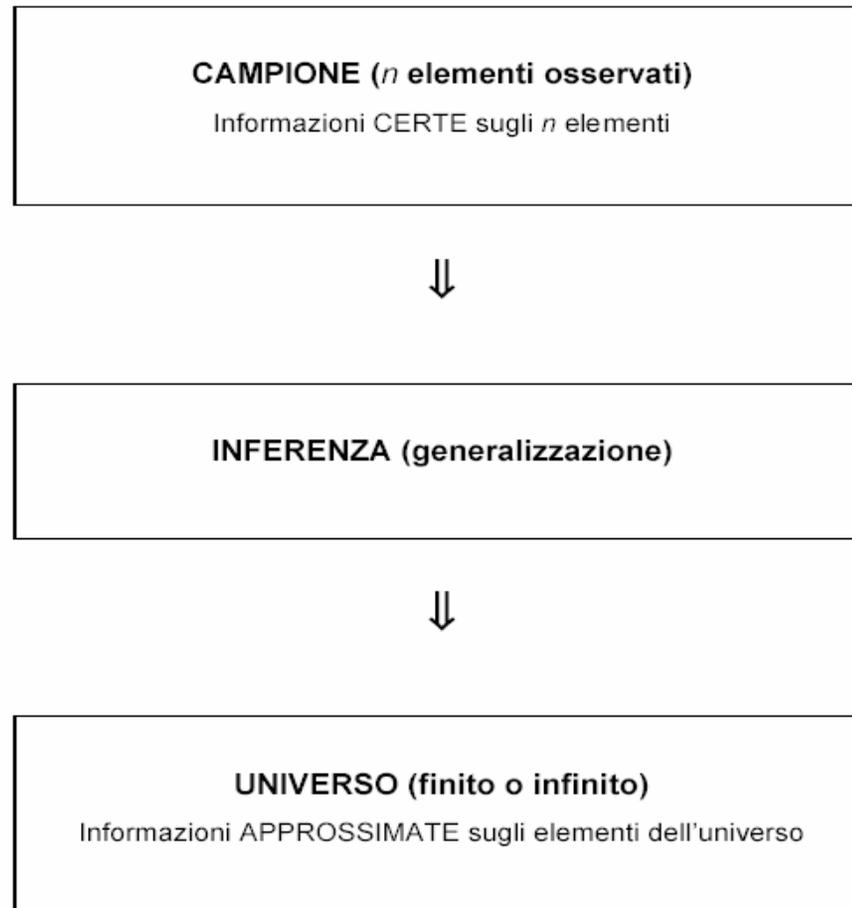
“A caso” però non vuol dire “a casaccio” (in sostanza il controllo della procedura di estrazione delle unità che vanno a formare il campione viene sottratto all’uomo e affidato al caso pertanto si eliminano elementi di soggettività).

Il concetto di caso è infatti strettamente connesso a quello di probabilità.

Il caso è un concetto intuitivo strettamente connesso all’idea di impossibilità di previsione, di individuare un ordine, un legame.

# Disegno d'indagine

## Strategia di campionamento - 5



## Strategia di campionamento - 6

### Il campionamento casuale

**Definizione:** Le unità sono selezionate con meccanismo casuale e hanno tutte una probabilità nota, o calcolabile, e non nulla di essere selezionate. (Cochran)

- è possibile definire l'insieme dei campioni distinti che possono essere estratti dalla popolazione;
- a ciascuno dei campioni sappiamo associare una probabilità di selezione;
- esiste un meccanismo di selezione casuale che garantisce la selezione di ciascun campione secondo la probabilità teorica;
- la scelta della numerosità campionaria sulla base di un criterio probabilistico.

#### **Il campionamento casuale: caratteristiche delle stime**

- Garantisce le migliori proprietà statistiche degli stimatori
- Garantisce la possibilità di stimare la “bontà” dei risultati ottenuti
- A parità di numerosità campionaria (e di costo) fornisce risultati più attendibili
- Semplifica la costruzioni di modelli statistici, che spesso assumono che i dati siano stati raccolti in modo casuale

# Disegno d'indagine

## Strategia di campionamento - 8

Campionamento casuale semplice: tutti i soggetti hanno la stessa probabilità di essere estratti

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, tante biglie quante sono le unità che si vogliono campionare.

Può effettuarsi con reinserimento e senza reinserimento.

Affinché si possa applicare tale metodo è necessario disporre di una lista che elenchi tutte le unità statistiche della popolazione e che tutti i soggetti siano ugualmente reperibili.

Uno dei metodi impiegati per ottenere un campione casuale era la tavola dei numeri casuali. OGGI USIAMO degli applicativi informatici (ad es.Excel ha un generatore di numeri casuali).

### Strategia di campionamento - 9

#### **PREGI**

1. E' privo di errori di selezione: nessuna unità o gruppo di unità è favorito a priori
2. E' molto semplice ed ha quindi un basso costo

#### **DIFETTI**

1. E' necessario disporre di una lista completa delle unità
2. Può non essere "rappresentativo": alcune particolari sezioni della popolazione potrebbero non essere rappresentate
3. Non sfrutta informazioni ausiliarie eventualmente disponibili sulla popolazione
4. La rilevazione sul campo può essere molto costosa se l'intervista viene effettuata attraverso un rilevatore e le unità selezionate sono fra loro lontane (alti costi di spostamento)

# Disegno d'indagine

## Strategia di campionamento - 10

Stratificazione del campione: è necessario che le caratteristiche usate nella formazione degli strati siano disponibili sulla lista per ogni unità della popolazione

Se riteniamo che un campione casuale semplice possa non rappresentare nelle giuste proporzioni sezioni fra loro eterogenee della popolazione, possiamo procedere alla stratificazione.

### **Il procedimento di stratificazione consiste:**

1. nel raggruppare le unità della popolazione in strati il più possibile omogenei al loro interno rispetto al carattere (od ai caratteri) investigati. Gli strati costituiscono una partizione: sono a due a due disgiunti e complessivamente esaustivi;
2. nell'estrarre casualmente un certo numero di unità campione da ciascuno strato.

La stratificazione deve essere effettuata sulla base di uno o più caratteri noti su tutti gli elementi della popolazione e correlati con il carattere indagato: quanto più stretto è il legame fra le variabili di stratificazione e la variabile oggetto di indagine, tanto più efficace sarà la stratificazione.

### Strategia di campionamento - 11

#### **PREGI**

1. Garantisce una migliore rappresentatività, a parità di numerosità del campione, rispetto al campionamento casuale semplice, genera quindi stime più efficienti
2. Sfrutta informazioni disponibili sulla popolazione

#### **DIFETTI**

1. E' necessario disporre di una lista completa delle unità della popolazione che riporti le informazioni utilizzate nella formazione degli strati
2. Richiede un lavoro preliminare alla rilevazione: scelta delle variabili di stratificazione, realizzazione di un'indagine pilota per la verifica dell'efficacia della stratificazione e, se l'indagine è multi-obiettivo, identificazione della stratificazione compromesso

# Disegno d'indagine

## Strategia di campionamento - 12

Campionamento a più stadi: non sia disponibile una lista complessiva delle unità della popolazione

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.

Per aumentare l'efficienza del campionamento a più stadi conviene spesso procedere ad una stratificazione delle unità di primo stadio, come avviene in molte indagini su larga scala (indagine ISTAT sui consumi, indagine ISTAT sulle forze di lavoro, ecc.).

# Disegno d'indagine

## Strategia di campionamento - 13

### 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

## Disegno d'indagine

### Strategia di campionamento - 14

#### Raccomandazioni

- 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 ritenerla difficilmente applicabile per evitare che in fase di selezione del campione siano introdotti errori di cui sia difficile valutare gli effetti sulle stime

### Strategia di campionamento - 15

## Il campionamento non casuale

Si ha un campionamento non probabilistico quando le unità della popolazione non hanno una probabilità predefinita e non nulla di entrare a far parte del campione.

Precede storicamente il campionamento probabilistico.

Non consente il calcolo dell'errore campionario e quindi la valutazione della bontà delle stime.

### Principali piani

1. Campionamento ragionato
2. Campionamento per quote
3. Snowball sampling (campionamento a valanga)
4. Campionamento accidentale

## Disegno d'indagine

### Strategia di campionamento - 16

Campionamento ragionato: La scelta delle unità campionarie viene effettuata sulla base di informazioni riguardanti la popolazione indagata.

E' appropriato quando le indagini sono ampie ma interessano poche unità territoriali (test market o provincia di prova). Se, ad esempio, per motivi economici un'indagine deve essere limitata a due sole città, sarà opportuno scegliere le città con criteri mirati, avendo cura che esse appartengano alla tipologia più rilevante ai fini della ricerca.

E' anche una scelta obbligata quando non è possibile accedere alla lista delle unità della popolazione. La rappresentatività del campione ragionato dipende dalla competenza del gruppo di esperti e quindi del loro livello di conoscenza della popolazione.

Non permette la valutazione dell'errore campionario.

## Disegno d'indagine

### Strategia di campionamento - 17

Campionamento per quote: si seleziona la popolazione oggetto di studio secondo alcune variabili strutturali in base alle quali si definiscono le quote del campionamento. Le unità statistiche da inserire nell'indagine sono scelte dai rilevatori in base alle quote loro indicate dal ricercatore

Si ricorre a questo piano di campionamento quando si hanno a disposizione dati analitici e vi sia ragione di ritenere che l'oggetto dello studio non possa essere influenzato dall'interazione rilevatore – rispondente.

Si eliminano i vincoli imposti dall'identificazione nominativa degli intervistandi ma si riduce la probabilità di inserire nel campione unità statistiche “svantaggiate” (abitanti dei piani superiori di uno stabile o dei quartieri periferici, lavori non canonici).

## Disegno d'indagine

### Strategia di campionamento - 18

Campionamento a valanga: si utilizza per studiare caratteristiche rare nella popolazione

1. Si seleziona un piccolo gruppo iniziale in modo casuale
2. Si effettua l'intervista e si chiede al rispondente di identificare amici o conoscenti con la caratteristica da analizzare
3. Si intervistano le nuove unità identificate e si continua a chiedere di identificare amici o conoscenti con la caratteristica da analizzare sino ad ottenere la dimensione del campione desiderata

## Disegno d'indagine

### Strategia di campionamento - 19

Campionamento accidentale: si intervistano unità statistiche facili da raggiungere

Possono essere usati solo a fini esplorativi o per indagini pilota. La selezione è affidata al rilevatore e pertanto i rispondenti sono inseriti nel campione solo perchè si trovano nel posto giusto al momento giusto.

#### Esempi

- ◆ Interviste per strada
- ◆ Questionari inclusi in riviste o allegati a prodotti in vendita