

Percorso introduttivo alla statistica ufficiale

L'indagine statistica

Indice

1. Alcune definizioni
2. Rilevazioni dirette e indirette
3. Rilevazioni totali e campionarie
4. Fasi di un'indagine

Alcune definizioni

L'indagine statistica (1/2)

- Un'indagine statistica è un processo produttivo complesso che ha come obiettivo la produzione di informazione statistica
- Semplificando al massimo, potremmo dire che: essa consiste nell'assegnare a ciascuna un'entità (p.e. persona, impresa, etc.) la rispettiva modalità di una caratteristica (p.e. età, forma giuridica, etc.), ottenendo un insieme di dati statistici (serie statistica)



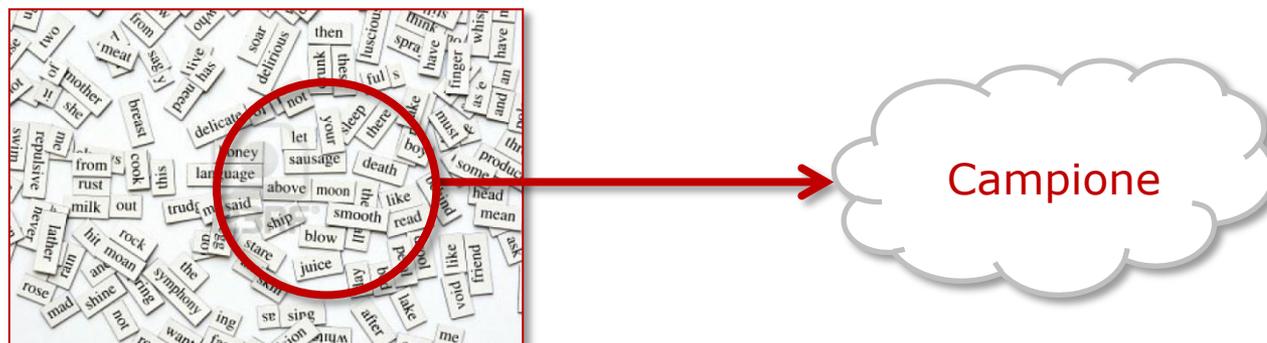
L'indagine statistica (1/2)

In alcuni casi la rilevazione fornisce risposte (opinioni, informazioni, scale di giudizi, etc.), mentre in altri casi occorre procedere a misure tramite strumenti (metro, bilancia, orologio etc.)



Campione

- **Campione:** sottoinsieme di una totalità di elementi che viene assunto a rappresentare la totalità stessa, costituita dalla popolazione oggetto di indagine
- Sostanzialmente il campione rappresenta la parte della popolazione che viene osservata
- I criteri usati per la formazione dei campioni possono essere diversi ma, affinché le informazioni ottenute dal campione possano essere estese alla popolazione di provenienza, è necessario che il campione sia **rappresentativo** ovvero riproduca le caratteristiche più rilevanti ai fini dell'indagine della popolazione di origine



Classificazione delle indagini

- Con riferimento alla **raccolta dei dati** un'indagine può essere:
 - **diretta**: i dati sono espressamente raccolti presso la popolazione oggetto di studio al fine di conoscere un determinato fenomeno
 - **indiretta**: ci si rivolge ad una risorsa che ha già questi dati raccolti e classificati
- Con riferimento **alla popolazione** da esaminare, un'indagine statistica può essere:
 - **totale** (rilevazioni censuarie o censimento): quando viene osservata l'intera popolazione oggetto di studio
 - **campionaria**: quando si osserva solo un sottoinsieme della popolazione tecnicamente definito campione

Rilevazioni dirette e indirette

Rilevazioni dirette

Definizione:

Rilevazioni dove l'informazione viene espressamente raccolta al fine di conoscere un determinato fenomeno

Esempi: i censimenti, le indagini campionarie ad hoc su settori specifici (indagini multiscopo, forze lavoro, consumi, sugli sbocchi professionali dei laureati etc.)



Vantaggi e svantaggi delle rilevazioni dirette

Vantaggi: l'informazione richiesta è proprio quella necessaria a descrivere un certo fenomeno.

Svantaggi:

- alti costi di rilevazione
- maggior "disturbo statistico"



Rilevazioni indirette

Definizione:

Si definisce rilevazione *indiretta* (o *indagine statistica basata su dati di fonte amministrativa*) quella indagine che utilizza i dati amministrativi con finalità di tipo statistico

La gran parte delle rilevazioni svolte dall'Istat ha base amministrativa

Caratteristiche:

- il dato statistico è generato dalla normale attività amministrativa
- nasce dalla registrazione di uno stato di fatto (ricovero ospedaliero, cambio di residenza, etc.)
- è un dato esaustivo

Esempi

- statistiche sul commercio con l'estero a partire dalle bolle doganali,
- statistiche dell'istruzione attraverso la registrazione degli iscritti e dei licenziati negli istituti scolastici
- statistiche sanitarie a seguito delle registrazioni fatte dai medici e dalle strutture ospedaliere etc.

Vantaggi delle rilevazioni indirette

- ❑ Riduzione del costo della rilevazione
- ❑ Riduzione del “disturbo statistico”

Svantaggi delle rilevazioni indirette

- la popolazione alla quale interessa riferire l'informazione può non coincidere con quella della raccolta dei dati amministrativi; in questo caso si parla di *errori di copertura*
- i concetti, le definizioni e le classificazioni adottate nella raccolta dei dati amministrativi possono non coincidere con quelli della ricerca statistica
- la qualità con cui le informazioni sono raccolte nei dati amministrativi può non essere sufficiente per gli obiettivi della ricerca; ciò accade soprattutto per le variabili che non sono fondamentali allo svolgimento dei processi amministrativi, ma che possono essere molto importanti ai fini statistici

Si consideri ad esempio le notizie riguardanti la professione del defunto, le quali sono ovviamente di difficile reperimento e non di immediato interesse al momento della registrazione del decesso, ma che sono invece fondamentali negli studi epidemiologici sulle malattie professionali

- ci sono fatti sociali che sfuggono a qualsiasi atto amministrativo (p.e. nuove forme di struttura familiare come le famiglie di fatto, reati non denunciati all'autorità giudiziaria, etc.)

Sintetizzando ...

	Dirette	Indirette
Caratteristica	L'informazione è espressamente raccolta al fine di conoscere un determinato fenomeno	Si utilizzano i dati amministrativi con finalità di tipo statistico
Pro	Rispondenza tra gli obiettivi dell'indagine e l'informazione raccolta	<ul style="list-style-type: none">• Minori costi di rilevazione• Minor "disturbo statistico"
Contro	Maggiori costi di rilevazione Maggior "disturbo statistico"	<ul style="list-style-type: none">• Il "collettivo amministrativo" può non coincidere con il collettivo statistico• I concetti, le definizioni e le classificazioni adottate possono non coincidere con quelli della ricerca statistica• Minore qualità dei dati amministrativi rispetto agli obiettivi della ricerca• Ci sono fatti sociali che sfuggono agli atti amministrativi

Rilevazioni totali e campionarie

Indagini totali e campionarie

Lo studio di una popolazione può essere effettuato scegliendo tra due metodi alternativi:

- ❑ conducendo **un'indagine totale (o censimento)** , che interessa tutte le unità statistiche della popolazione
- ❑ conducendo **un'indagine campionaria**: il fenomeno viene rilevato su una parte della popolazione (campione casuale), ma i risultati sono estesi, grazie a tecniche probabilistiche, all'intera popolazione

Il censimento (1/2)

Il censimento è una rilevazione:

Totale o universale

Viene osservata l'intera popolazione di riferimento

Generale

Riguarda i principali aspetti strutturali che caratterizzano le unità di rilevazione

Simultanea

Le notizie devono essere riferite allo stesso istante (la situazione rilevata deve essere quella del momento stabilito come data di riferimento del censimento)

Periodica

Viene ripetuta ad intervalli regolari (p.e. cadenza decennale)

Vantaggi e svantaggi delle indagini totali

Vantaggi

- ❑ Forniscono una misura non affetta da errore campionario del fenomeno
- ❑ Forniscono liste per indagini campionarie
- ❑ Consentono elevato dettaglio di analisi

Svantaggi

- ❑ La realizzazione e il rilascio delle informazioni richiede tempi molto lunghi
- ❑ Costi elevati sia in termini di risorse che economici

Indagine campionaria

L'indagine campionaria si rende necessaria quando:

- l'ampiezza della popolazione è infinita
- quando la rilevazione comporta la distruzione delle unità osservate (si pensi di voler verificare il funzionamento di una partita di fuochi d'artificio)

Vantaggi della rilevazione campionaria:

- riduzione dei costi
- riduzione dei tempi
- riduzione del carico organizzativo (e quindi dell'errore di rilevazione)

Limiti:

- minore dettaglio nella disaggregazione territoriale (servono campioni molto grandi per ottenere stime attendibili a livello comunale/provinciale)

Indagini totali e campionarie

	Totali	Campionarie
Caratteristica	<ul style="list-style-type: none">• Viene osservata l'intera popolazione oggetto di studio	<ul style="list-style-type: none">• Viene osservata solo un sottoinsieme della popolazione tecnicamente definito campione
Pro	<ul style="list-style-type: none">• Elevato dettaglio di analisi	<ul style="list-style-type: none">• Riduzione dei costi• Riduzione dei tempi• Riduzione del carico organizzativo
Contro	<ul style="list-style-type: none">• Costi elevati• Minore accuratezza dei dati• Ottenimento delle informazioni in tempi più lunghi	<ul style="list-style-type: none">• Minore dettaglio nella disaggregazione territoriale

Totalità o campione?

Nella scelta fra un'indagine campionaria o esaustiva si deve bilanciare l'attendibilità, cioè la precisione e l'accuratezza di una stima, con il costo da sostenere per ottenerla

Precisione: riguarda l'errore campionario

- La precisione delle stime è direttamente proporzionale alla dimensione del campione: è assoluta nelle indagini esaustive e decresce in funzione della numerosità del campione sondato

Accuratezza: riguarda gli errori di rilevazione e di trattamento dei dati

- Sono un rischio maggiore nelle indagini di vaste dimensioni, che richiedono l'impiego massivo di personale e sono gravate da enormi volumi di dati con scarse possibilità di controllo sulla veridicità

Fasi di un'indagine

Fasi di un'indagine statistica

Un'indagine, sia essa totale o parziale, nasce seguendo un preciso ciclo di fasi ognuna orientata alla qualità del prodotto finale

Fasi:

1. definizione degli obiettivi dell'indagine
2. definizione del disegno di indagine
3. rilevazione dei dati (parziale o totale)
4. registrazione
5. revisione e validazione
6. elaborazione metodologica
7. presentazione ed utilizzazione dei risultati
8. diffusione

1. Definizione degli obiettivi dell'indagine

Gli **obiettivi** devono essere particolareggiati, al fine di:

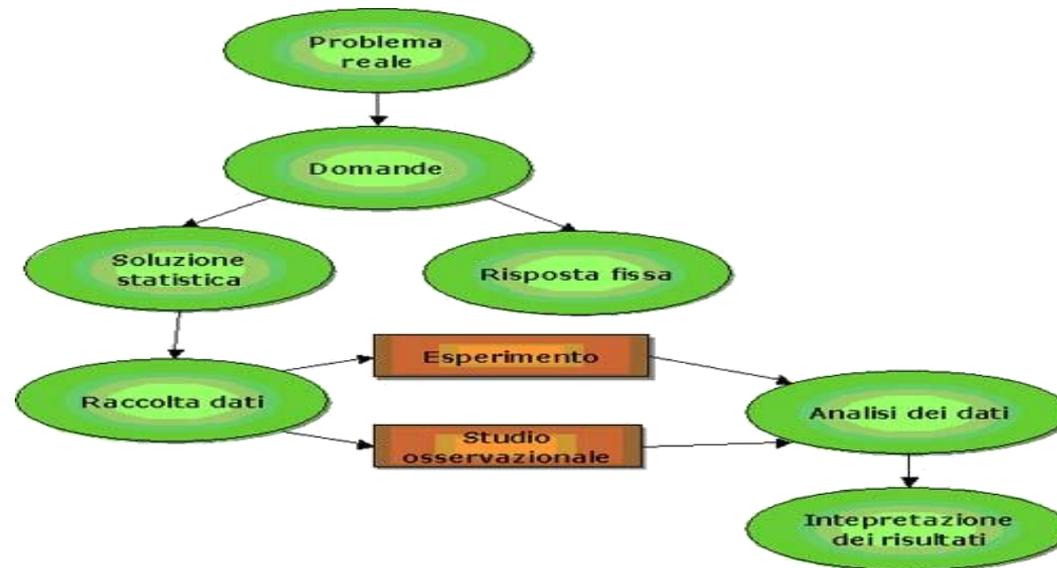
- individuare con precisione le informazioni da ricercare
- evitare equivoci definatori (definizione di popolazione obiettivo, definizione dell'unità statistica)
- circoscrivere con esattezza il territorio e il periodo di riferimento dell'indagine

Rientra nella definizione degli obiettivi l'indicazione del grado di precisione voluto per i risultati dell'indagine

2. Definizione del disegno dell'indagine

La definizione del **disegno di indagine** mira a rispondere alle seguenti necessità:

- definire qual è il tipo di indagine più consono a produrre le statistiche che si desiderano
- decidere tra indagine totale e campionaria e, in tal caso, disegnare ed estrarre il campione (strategia di campionamento)



3. Rilevazione dei dati

Lo strumento di osservazione scelto per la raccolta dei dati può considerarsi come un filtro posto tra l'osservatore e l'oggetto, cambiando il mezzo di osservazione dallo stesso oggetto possono ottenersi dati differenti, ed è per questo che la loro scelta influenza i dati dell'indagine

I dati possono essere raccolti attraverso:

- misurazione di entità oggettive (rilevazione delle qualità delle acque dei fiumi italiani)
- dichiarazioni (questionario)

In quest'ultimo caso rivestono particolare importanza ai fini della qualità dei dati (mancate risposte, errori di misura):

- il questionario (strumento di osservazione)
- le tecniche di intervista (le modalità di somministrazione del questionario)
- il rilevatore

4. Registrazione

La fase di registrazione su supporto informatico consiste nel convertire le informazioni raccolte presso i rispondenti, e disponibili su questionario cartaceo, su supporto di formato interpretabile dalle procedure informatiche predisposte dall'indagine

In alcuni casi la fase di registrazione può essere assente, come quando la rilevazione viene effettuata in modalità assistita dal computer (CATI, CAPI), in altri casi l'operazione può essere sostituita dalla lettura ottica dei questionari

5. Revisione e validazione

- Individuazione delle fonti di errore
- Adozione di metodi di correzione in fase di analisi dei dati
- Validazione dei dati corretti



7. Presentazione ed utilizzazione dei risultati

Presentazione dei risultati

Lo statistico deve porre particolare cura nella presentazione dei risultati sotto forma di:

- tabelle
- grafici
- rapporti sintetici



Utilizzazione dei risultati

Lo statistico contribuisce ad una corretta utilizzazione dei risultati di una indagine:

- circoscrivendo l'ambito applicativo
- richiamando i limiti entro cui essa assume validità

8. Diffusione



Diffusione

La diffusione rappresenta la fase conclusiva di una rilevazione attraverso la quale si concretizza la disponibilità dei dati raccolti

L'utilizzazione dei dati è resa possibile da un'adeguata offerta produttiva e da efficienti forme di diffusione

Diffusione e qualità

La fase della diffusione ha notevoli ripercussioni sulla *rilevanza* dei dati

La rilevanza è una componente della qualità dei dati, essa si traduce in termini di:

- *Accessibilità*: possibilità per gli utilizzatori di entrare in possesso dei dati
- *Confrontabilità*: possibilità di paragonare nel tempo e nello spazio le statistiche riguardanti il fenomeno di interesse

...e adesso...
buon lavoro!

Rete per la promozione
della cultura statistica

Per ulteriori moduli didattici relativi alle scuole secondarie di secondo grado si prega di consultare anche la piattaforma Scuola di statistica – Lab (accessibile dal link <http://scuoladistatistica-lab.istat.it/>)