

Statistica per il marketing

La Ricerca Economica e Sociale utilizza dati provenienti da fonti diverse e raccolti con strumenti particolari. Si possono individuare due classificazioni generali dei dati, da cui dipendono i modelli di ricerca utilizzati.

Dati primari: informazioni che vengono raccolte per la prima volta.

Essi riguardano:

- ✓ osservazioni raccolte in modo non strutturato (es. interviste o focus group)
- ✓ dati provenienti da esperimenti scientifici (dove esistono dei soggetti e un ambiente controllato o comunque con caratteristiche apposite)
- ✓ indagini campionarie svolte attraverso questionari

Dati secondari: informazioni già esistenti che possono essere utili per la ricerca, non devono essere costruiti.

Si dividono in due categorie:

- a. Dati secondari interni (prodotti dalla stessa organizzazione che intende studiarli, ad es. Università dati su studenti)
- b. Dati secondari esterni (es. ISTAT)

I dati secondari interni

Si tratta di informazioni che si trovano già in possesso del ricercatore (database aziendali/istituzionali, materiale già pubblicato in passato).

Vantaggi

- ✓ La fase preliminare di analisi di dati già disponibili è utile allo sviluppo dell'intera strategia di ricerca
- ✓ Bassi costi
- ✓ Facilità e velocità di ottenimento

Limiti

- ✓ Un problema molto diffuso è che molte aziende, specie quelle di dimensione medio/piccola, non sappiano gestire e analizzare con sufficiente competenza le informazioni che vengono raccolte di routine.

I dati secondari esterni

Comprendono le statistiche e i rapporti pubblicati da Enti Pubblici, Associazioni Commerciali di categoria, Istituti di Ricerca.

Tali dati si suddividono in quattro grandi categorie:

- ✓ dati ufficiali pubblicati (Fonti Statistiche Ufficiali)
- ✓ banche dati (possono essere sia pubbliche che private)
- ✓ dati da Internet (open data)
- ✓ dati d'agenzia (sondaggi, panel, audit su dettaglianti, studi sulla distribuzione)

Vantaggi

- ✓ bassi costi
- ✓ facilità e velocità di ottenimento

Limiti

- ✓ mancanza di rilevanza e accuratezza (c'entrano poco con quello che sto studiando)
- ✓ forma inappropriata
- ✓ intempestività

Le fonti statistiche

Fonte: principio da cui qualcosa emana o proviene direttamente (Devoto e Oli, 1971).

Le fonti statistiche sono fonti dalle quali attingiamo informazioni. Sapere la fonte dei dati ci dice molto sulla qualità dei dati, inoltre le fonti sono strettamente collegate al tipo di ricerca che stiamo conducendo.

Le fonti che forniscono informazione statistica sono anch'esse un principio. Un principio per l'analisi delle politiche aziendali, un supporto per il mondo dell'informazione e una base per le analisi dei fenomeni economici e sociali.

È possibile classificare le fonti statistiche in base a:

1. Natura dell'ente che le fornisce
2. Scopo della rilevazione
3. La territorialità

La natura dell'ente

Una prima classificazione è quella basata sulla natura giuridico-amministrativa della fonte. Considerando sia gli enti pubblici che quelli privati è possibile definire quattro grandi categorie di fonti:

- a) istituzionali;
- b) pubbliche o erogatrici di servizio pubblico;
- c) private;
- d) enti di natura scientifica.

La natura dell'ente: fonti istituzionali

Sono quelle fonti cui la legge ha relegato funzioni di raccolta, elaborazione e gestione di informazione statistica.

Esse producono sia dati istituzionali che statistici in senso proprio.

Le principali fonti istituzionali italiane sono l'ISTAT, e altri soggetti pubblici (Ministeri, Regioni, Province autonome, Province, Prefetture, Comuni, ecc).

Tra le fonti statistiche istituzionali internazionali si ricordano:

- ✓ EUROSTAT
- ✓ OECD (Organisation for Economic Cooperation and Development)
- ✓ ONU (Organizzazione delle Nazioni Unite)
- ✓ OMS (Organizzazione Mondiale della Sanità)
- ✓ UNESCO (United Nations Educational Scientific and Cultural Organisation)
- ✓ FMI (Fondo Monetario Internazionale)

La natura dell'ente: fonti pubbliche

Tali fonti producono quei dati che sono utili al buon funzionamento dell'ente preposto alla loro raccolta (*process-produced data*).

Il ruolo informativo di questo tipo di dati sta crescendo costantemente, perché in alcuni casi essi costituiscono l'unica risorsa per la comprensione di determinati fenomeni di rilevante interesse sociologico.

Di seguito alcuni esempi di fonti italiane di questo tipo:

- ✓ Agea (Agenzia per le Erogazioni in Agricoltura),
- ✓ Ispra (Istituto superiore per la protezione e la ricerca ambientale),
- ✓ Inail (Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro),
- ✓ Inps (Istituto nazionale di previdenza sociale),
- ✓ Inran (Istituto nazionale di ricerca per gli alimenti e la nutrizione),
- ✓ Iss (Istituto Superiore di Sanità),
- ✓ Isvap (Istituto per la vigilanza sulle assicurazioni),
- ✓ Banca d'Italia.

La natura dell'ente: fonti private

Questa categoria di fonti è costituita da società ed enti di diritto privato.

Tra queste vi sono società a partecipazione statale quali:

- ✓ Aci (Automobile Club Italiano)
- ✓ Ferrovie dello Stato
- ✓ Isfort (Istituto Superiore di formazione e ricerca per i trasporti)
- ✓ Poste Italiane

consorzi, associazioni e organizzazioni sindacali di lavoratori dipendenti e dei datori di lavoro, tra cui figurano ad esempio:

- ✓ Confindustria, Confcommercio, Confartigianato, Confesercenti
- ✓ Caritas
- ✓ CIGL, CISL, UIL
- ✓ Inpdap
- ✓ INDIS (Istituto Nazionale Distribuzione e Servizi) – Unioncamere
- ✓ ABI (Associazione Bancaria Italiana)

oltre a innumerevoli uffici e studi di aziende private.

La natura dell'ente: enti scientifici

Questa categoria di fonti è costituita da enti ricerca che operano secondo criteri di scientificità quali, ad esempio:

- ✓ Università
- ✓ ISAE (Istituto Studi di Analisi Economica)
- ✓ CENSIS (Centro Studi Investimenti Sociali)
- ✓ CNR (Consiglio Nazionale delle Ricerche)
- ✓ Istituto Guglielmo Tagliacarne [fa parte del SISTAN dal 2002]
- ✓ Fondazione Gianni Agnelli
- ✓ Fondazione ENI Enrico Mattei

Lo scopo della rilevazione

Una seconda classificazione delle fonti è basata invece sullo scopo della rilevazione:

1. **fonti di origine statistica** (materiale che già in principio è stato raccolto con l'intento di conoscere alcuni aspetti della vita sociale, ad es. bankitalia raccoglie dati sulle famiglie per capire come si comportano e quale scelte attuano in relazione al loro bilancio familiare)
2. **fonti di tipo amministrativo** (documenti predisposti dalle amministrazioni pubbliche e private per scopi inerenti alla propria attività).

È pertanto possibile operare la seguente distinzione:

- a) Censimenti (fonte statistica, anche se un tempo serviva più per aggiustare i registri anagrafici e ripulire l'anagrafe)
- b) Indagini campionarie (fonte statistica)
- c) Rilevazioni amministrative (fonte amministrativa)

Il censimento

Il Censimento della Popolazione e delle Abitazioni è l'insieme delle operazioni di progettazione, raccolta, analisi e diffusione dei dati di carattere demografico, culturale, economico e sociale di tutti gli abitanti del Paese (ISTAT, 1993). IBM produceva macchinari per analizzare i dati anagrafici.

Il censimento è un'indagine esaustiva a cadenza decennale con cui si vuole determinare non solo l'entità numerica ma anche la struttura della popolazione di un Paese.

Il Censimento della Popolazione e delle Abitazioni non è l'unico censimento italiano:

- ✓ Censimento generale dell'Agricoltura,
- ✓ **Censimento generale dell'Industria e dei Servizi**
- ✓ Censimento generale delle istituzioni private ed imprese nonprofit
- ✓ ...

Il questionario sul censimento è molto povero a livello di informazioni, perché essendo un'indagine capillare devo avere elevati tassi di risposta per far rispondere più persone possibili.

I censimenti però, hanno un alto tasso di errore. Oggi si attua ogni anno prendendo dei campioni dalla popolazione e l'aggiustamento verrà fatto in corso d'opera. Prendendo un campione posso essere anche più accurato e approfondito.

Indagini campionarie

Sono caratterizzate dal fatto che **solo una parte** delle unità componenti la popolazione viene selezionata ed indagata.

Questo espediente, diminuendo l'onere della rilevazione, consente di destinare maggiore attenzione a tutte le attività connesse al miglioramento e al controllo della qualità dei dati raccolti.

Tuttavia, selezionare solo un campione implica ovviamente una minore attendibilità delle misure osservate in quanto stime dei parametri di interesse. Il problema riguarda sempre il passaggio dal campione alla popolazione.

È infatti chiaro che a seconda di quali unità vengono inserite nel campione prescelto, i risultati riferiti alla popolazione complessiva varieranno. Le indagini su campioni di popolazione rientrano nell'attività dell'ISTAT da circa un ventennio. Quella sulle **forze di lavoro**, realizzata nel settembre del 1952, è rimasta per molti anni l'unica indagine diretta ad individui con interviste personali. Dal 1983 con l'Indagine sulle strutture ed i comportamenti familiari, ora **Indagine Multiscopo** (copre diversi aspetti delle famiglie) **sulle Famiglie**, l'ISTAT indirizza sempre più la propria attività verso un uso delle indagini campionarie.

Le rilevazioni amministrative

Sono i dati prodotti dall'attività amministrativa (process-produced data) e consentono di acquisire informazioni regolarmente aggiornate su individui od enti.

Le più famose di queste rilevazioni sono sicuramente le anagrafi o registri della popolazione.

Esse sono tenute dalle amministrazioni comunali ed hanno la funzione di registrare nominativamente gli abitanti residenti nel comune.

Altri esempi di rilevazioni amministrative sono il registro delle imprese, il PRA (Pubblico registro automobilistico) e le liste elettorali.

La territorialità

I dati statistici possono essere ulteriormente suddivisi in base alla territorialità dell'ente che produce il dato. Su questa base è possibile definire almeno tre categorie di fonti.

- a) Internazionali/Sovranazionali
- b) Nazionali
- c) Locali

Il sistema statistico nazionale

Istituito con D.Lgs. 6 settembre 1989, n.322, il SISTAN connette in un'unica rete tutte le competenze disperse nelle varie amministrazioni, con l'obiettivo di acquisire nuove e maggiori conoscenze per influire sui processi decisionali della Pubblica Amministrazione e con quello ulteriore di ridurre la molestia verso il pubblico e le imprese che ha raggiunto livelli eccessivi. Una rete che connette tutti i produttori di dati.

In questo disegno, l'Istat diventa un centro di coordinamento, di ricerca, di consulenza, di formazione, di garanzia metodologica, anche se perde in un certo senso il vecchio privilegio di monopolio, trasformandosi in un'azienda in grado di fornire servizi al pubblico, puntando su schemi privatistici.

Secondo l'articolo 2 del D.L.vo 322/89, fanno parte del SISTAN:

l'ISTAT;

gli **uffici di statistica centrali e periferici** delle amministrazioni centrali e delle amministrazioni ed enti autonomi;

gli uffici di statistica delle **regioni** e delle **province autonome**;

gli uffici di statistica delle **province**;

gli uffici di statistica dei **comuni**, eventualmente consorziati;

gli uffici di statistica delle **ASL**;

gli uffici delle **Camere di Commercio**;

gli uffici di statistica comunque denominati di amministrazioni od enti pubblici;

gli uffici di statistica di **soggetti privati** che svolgono funzioni di interesse pubblico.

Il programma statistico nazionale

Il Programma statistico nazionale (PSN), secondo quanto indicato dall'art. 13 del D.L.vo 322/89 e successive integrazioni stabilisce le rilevazioni statistiche di interesse pubblico affidate al SISTAN ed i relativi obiettivi.

Il PSN parte dal presupposto di fornire un **servizio pubblico** per la **circolazione delle informazioni**, con la tutela del **segreto statistico**; il SISTAN, creando una rete di collegamento, vuole riportare la statistica là dove si formano le decisioni, ma il **pluralismo** delle **fonti** deve essere **coordinato** con gli atti di indirizzo e coordinamento.

Il PSN afferma che la **produzione** deve essere **orientata all'utente**, il quale deve avervi ampio accesso; i dati devono essere usati in maniera coordinata, organizzati in serie temporali e forniti **tempestivamente**, debbono essere **imparziali**, coordinati con EUROSTAT, **standardizzati** ed **integrati** per l'interconnessione dei sistemi informativi.

Qualità delle fonti statistiche

La qualità delle fonti informative è un problema che si affronta di frequente nelle analisi statistiche, poiché la qualità dei risultati è strettamente connessa alla qualità dei dati rispetto al livello di complessità delle metodologie e tecniche di analisi statistica. Ci sono tutta una serie di problemi legati alla qualità del dato che non è possibile correggere a posteriori.

Storicamente la qualità dell'informazione statistica si identificava quasi unicamente col concetto di accuratezza, ovvero vicinanza tra la stima prodotta di un parametro (caratteristica della popolazione studiata) e il vero valore (ignoto) dello stesso.

Oggi i requisiti per un dato di qualità sono:

- a) Pertinenza
- b) Accuratezza
- c) Tempestività
- d) Accessibilità
- e) Comparabilità
- f) Completezza
- g) Coerenza

Pertinenza

È la capacità dell'informazione statistica di soddisfare le esigenze degli utenti.

Viene valutata all'interno del programma statistico.

È opportuno conoscere i bisogni informativi degli utenti finali dell'informazione.

Accuratezza

È la vicinanza tra la stima prodotta per un parametro (caratteristica della popolazione studiata) e il vero valore (ignoto) dello stesso.

La stima differisce dal vero (ignoto) valore del parametro a causa di una serie di errori. L'errore è ineliminabile, ma comunque si può tenere sotto controllo entro certi limiti.

Gli errori possono essere presenti in qualsiasi fase dell'indagine (es. non completezza delle liste di riferimento, irreperibilità delle unità di rilevazione, autoselezione del campione quando lanciamo un sondaggio ma non tutti rispondono, mancate risposte, ...).

Tempestività

Si riferisce al tempo che intercorre tra l'istante o il periodo cui si riferisce l'informazione statistica e il momento in cui le informazioni sono disponibili per l'utenza.

La tempestività è strettamente legata alla pertinenza.

La maggiore accuratezza di solito va a discapito della tempestività

Informazioni poco tempestive possono diventare poco rilevanti

Accessibilità

Si riferisce alla facilità con cui gli utenti possono:

- ✓ Apprendere dell'esistenza dell'informazione
- ✓ Localizzarla
- ✓ Trasferirla nel proprio contesto lavorativo

Risultano pertanto determinanti:

- ✓ Il sistema di annuncio del rilascio dell'informazione
- ✓ Il sistema di distribuzione dell'informazione
- ✓ Il costo
- ✓ La disponibilità di informazioni aggiuntive per una corretta interpretazione dell'informazione

Comparabilità

Indica in che misura l'informazione è confrontabile nello spazio e nel tempo.

Il problema della comparabilità dei dati ha un'importanza cruciale sia perché si presenta con una frequenza elevata nella ricerca applicata (sul territorio e non) sia perché costituisce un fattore fortemente distorsivo nell'analisi dei fenomeni.

L'assenza di comparabilità può portare a risultati discordanti. Le discordanze tra fonti diverse sono imputabili a:

- ✓ classificazione dei fenomeni
- ✓ modalità di rilevazione
- ✓ procedure di controllo dei dati
- ✓ non veridicità della fonte intesa come non «ufficialità» della fonte

Soluzioni

Si trovano caso per caso mediante un approccio critico da parte dell'utilizzatore dei dati.

Coerenza, completezza

Coerenza

statistiche derivate da fonti diverse sono **coerenti** se basate su definizioni, classificazioni e metodologie standard ossia sono integrabili

Completezza

la capacità dei dati di fornire informazioni sugli aspetti indagati **senza lacune** o omissioni non giustificate, ma anche la **necessaria copertura geografica** delle stesse informazioni che non sempre sono disponibili per tutto il territorio nazionale

Uso delle fonti

Le fonti informative forniscono i dati necessari al calcolo di indicatori per l'analisi dei fenomeni economici e sociali.

La raccolta dei dati produce:

serie territoriali (dati longitudinali o cross-section): sequenza dei valori assunti da una variabile nello stesso momento in diversi aggregati territoriali (es. grado di invecchiamento della popolazione per area, livello di sviluppo economico regionale)

serie storiche (time series): sequenza dei valori assunti da una variabile nello stesso aggregato territoriale in momenti diversi (es. andamento del tasso di natalità).

dati panel (serie storiche di dati longitudinali): sequenza dei valori assunti da una variabile in diversi aggregati territoriali e in momenti diversi (es. *Indagine sui bilanci delle famiglie italiane* promossa dalla Banca d'Italia inquadrata nell'indagine della Banca Centrale Europea: *Household Finance and Consumption Survey - HFCS*). Viene selezionato un campione di individui seguito nel tempo e rilevato più volte, i soggetti poi vengono sostituiti per un breve periodo e poi ritornano.

Metadati

Sono il **curriculum vitae** dei dati e contengono l'informazione che descrive un insieme di dati: **dove, quando, come** e **da chi** i dati sono stati ottenuti. Sono un elemento indispensabile per rendere i dati fruibili correttamente da chiunque e a distanza di tempo e di spazio.

La funzione principale di un sistema di metadati è quella di consentire il raggiungimento dei seguenti obiettivi: **ricerca, localizzazione, selezione, gestione** del dato grazie all'intermediazione di banche dati e cataloghi e **disponibilità**.

I campi di una collezione di metadati sono costituiti da informazioni che descrivono le risorse informative a cui si applicano, con lo scopo di migliorarne la visibilità e facilitarne l'accesso. I metadati infatti permettono il recupero di documenti primari indicizzati attraverso le stringhe descrittive contenute in record: schede in cui vengono rappresentate le caratteristiche più significative o le proprietà peculiari dei dati in questione, così che la loro essenza possa essere catturata da un'unica concisa descrizione, che, in modo sintetico e standardizzato, fornisce a sua volta una via di recupero dei dati stessi.

Le banche dati

Gli elaboratori elettronici consentono la raccolta, la elaborazione e la disseminazione delle informazioni. I vantaggi delle banche dati sono riconducibili all'aggiornamento continuo delle informazioni e alla possibilità di accedervi selettivamente, anche a distanza.

Le banche dati possono essere on-line o off-line (solitamente su CD ROM). Alcuni esempi:

CERVED (Società Nazionale di Informatica delle Camere di Commercio Italiane) rende disponibili informazioni raccolte nelle diverse Camere (anagrafe ditte, bilanci, ecc.).

NEW CRONOS di Eurostat contiene serie temporali macroeconomiche dei vari paesi

COMEXT di Eurostat contiene le statistiche degli scambi commerciali tra i paesi della comunità e circa 200 paesi terzi.

SEAT PG (CONSODATA) gestisce la base informativa *cd superstat* che contiene dati comunali tratti dai censimenti della popolazione e dell'industria, indicatori di provenienze diverse (Telecom, Enel, Aci) e elaborazioni delle informazioni desumibili dagli elenchi telefonici

GFK Eurisko fornisce servizi basati su centinaia di importanti banche dati internazionali, relative a quasi tutti i Paesi del mondo al fine di effettuare ricerche di mercato su nuovi prodotti, tecnologie innovative, individuazione di fornitori o partner, ecc.

ASSAP-NIELSEN registra le componenti essenziali degli annunci pubblicitari, apparsi su quotidiani, periodici, affissioni stradali, radio e televisione.

Dati di agenzia

Informazioni vendute da società di ricerca specializzate a gruppi di clienti che condividono bisogni informativi comuni.

Vantaggi

- ✓ bassi costi e velocità
- ✓ Accuratezza

Svantaggi

- ✓ i dati non sono specifici per le esigenze del singolo committente

Ambiti analizzati

- ✓ comportamenti dei consumatori
- ✓ segmentazione del mercato
- ✓ evoluzione del mercato
- ✓ monitoraggio di campagne promozionali

Dati di agenzia da famiglie

sondaggi (interviste su campione con domande predefinite)

panel (campioni di soggetti che forniscono informazioni a intervalli regolari)

sondaggi omnibus (sondaggi commissionati da più utilizzatori/imprese)

Dati di agenzia da operatori

audit sui dettaglianti inventari periodici (solitamente bimestrali) condotti presso i negozi sulla merce in vendita e in magazzino; si raccolgono informazioni su vendite, prezzi, promozioni, ecc.

studi sulla distribuzione al dettaglio forniscono informazioni complete sui punti vendita che hanno a carico, in quel momento, i prodotti delle imprese.

informazioni sui clienti industriali vengono raccolte direttamente dalle agenzie o attraverso dati pubblicati presso gli acquirenti di prodotti utilizzati come beni intermedi.

Principali fornitori di dati d'agenzia in Italia

Doxa Demoskopea Abacus A.C. Nielsen Eurisko

Il censimento dell'industria e dei servizi

Ogni dieci anni i censimenti economici forniscono un quadro esaustivo della dimensione e delle caratteristiche del sistema economico nazionale con un dettaglio territoriale fine (area e microarea).

La base dati ottenuta consente la predisposizione e l'aggiornamento di archivi delle unità produttive (aziende, imprese, istituzioni) da utilizzare anche come base per l'esecuzione delle indagini campionarie.

Il Censimento generale dell'industria e dei servizi 2011 ha l'obiettivo di rappresentare in maniera puntuale e dinamica il mondo delle *imprese*, del *non profit* e delle *istituzioni pubbliche*, offrendo un contributo fondamentale alle decisioni di politica economica e alla governance di tre importanti settori della vita del Paese.

Rispetto ai passati censimenti economici presenta forti novità:

- ✓ si articola in tre diverse rilevazioni che coinvolgono un **campione di 260mila imprese** (sia grandi gruppi industriali sia piccole e medie realtà), oltre **470mila istituzioni non profit** e **13mila istituzioni pubbliche**;
- ✓ prevede una restituzione multicanale del questionario compilato;
- ✓ rileva nuove tematiche per ciascuno dei tre campi d'indagine.

La rilevazione sulle imprese

A differenza dei censimenti svolti fino ad oggi, la rilevazione non coinvolge direttamente tutte le realtà imprenditoriali – circa 4,5 milioni – ma un campione pari a circa 260mila unità di cui fanno parte tutti i grandi gruppi industriali, le imprese di grandi dimensioni, e circa 190mila piccole e medie imprese. La restituzione dei dati ottenuti sarà tuttavia di tipo **censuario**: le informazioni strutturali delle imprese saranno tratte dai registri statistici e dalle fonti amministrative, mentre la rilevazione dei dati tramite il questionario offre approfondimenti inediti sulla competitività e la capacità innovativa delle imprese.

La rilevazione sulle istituzioni non profit

L'Italia è tra i pochi Paesi in Europa a censire periodicamente il variegato mondo del non profit. Il Censimento vede coinvolte **474.765 istituzioni** che sono state individuate conformandosi alla definizione internazionale del System of National Accounts (SNA) che considera principalmente il criterio del "divieto di distribuzione di