Tecniche di rilevazione statistica

- Il disegno di ricerca
- Indagini censuarie e campionarie
- Indagini basate su questionario
- Introduzione al campionamento

Istituto nazionale di statistica | Bandi di gara | Concorsi | Trasparenza valutazione e merito
A A A
It
En
Istat.it
Prodotti
Strumenti
Informazioni
Cerca...
Metodi e software
Definizioni e classificazioni
Qualita dei dati
Cartografia
Dossier
Quali sono le aree di interesse
CAMPIONE
- Disegno
- Selezione
- Calcolo dei coefficienti di espansione
- Calcolo degli errori campionari
ACQUISIZIONE DATI
- Codifica automatica
- Gestione acquisizione dati con più tecniche di rilevazione
CONTROLLO E CORREZIONE
- Controllo e correzione variabili qualitative
- Controllo e correzione variabili quantitative
- Controllo e correzione variabili unità complesse
DIFFUSIONE DEI DATI
TUTELA RISERVATEZZA
- Tutela dati elementari
- Tutela dati aggregati
ANALISI DEI DATI
Home / Strumenti / Metodi e software
Metodi e software
Le nozioni metodologiche di base per rilevazioni statistiche dirette o basate su dati amministrativi (tecniche di indagine, campionamento, acquisizione e codifica dei dati, ecc.) sono consultabili in questo documento.
In questa sezione si trovano software di tipo generalizzato per la produzione statistica. Lo scopo è di agevolare l'applicabilità delle tecniche e dei metodi, garantendo le funzionalità produttive delle indagini. I software sono dotati della documentazione di supporto che illustra ed esemplifica le potenzialità delle applicazioni.
Uno spazio è dedicato alla tecnica del concatenamento per la stima della crescita economica.
In una serie di pubblicazioni metodologiche sono disponibili le norme e le classificazioni adottate nello svolgimento delle indagini, oltre alla gestione degli aspetti metodologici e ai software utilizzati.
Strumenti
Metodi e software
Software
Definizioni e classificazioni
Qualita dei dati
Cartografia
Dossier
produzione
15° CENSIMENTO GENERALE DELLA POPOLAZIONE E DELLE ABITAZIONI 2011
datawarehouse I.Stat
SerieStoriche
Linee guida metodologiche per rilevazioni statistiche
Nozioni metodologiche di base e pratiche consigliate per rilevazioni statistiche dirette o basate su fonti amministrative
Marco Fortini
Istituto Nazionale di Statistica, 2000
http://www.istat.it/it/strumenti/metodi-e-software

Istat.it

Prodotti | Strumenti | Informazioni

Cerca...

Metodi e software | Definizioni e classificazioni | Qualita dei dati | Cartografia | Dossier

Home / Strumenti / Qualita dei dati

Qualità dei dati

Le **Linee guida per la qualità dei processi statistici** contengono i principi stabiliti dall'Istat per la progettazione, esecuzione e controllo di una rilevazione statistica e i requisiti di qualità che le statistiche devono soddisfare.

Le Linee guida, predisposte in una forma sintetica ma esaustiva, sono strutturate in due parti che trattano rispettivamente la Qualità di Processo e la Qualità di Prodotto. Le Linee guida sono indirizzate in particolare ai responsabili dei processi di produzione statistica dell'Istat, ma possono essere un riferimento anche per gli enti del Sistema statistico nazionale e per gli utilizzatori dei dati.

Presentazione (Enrico Giovannini, Presidente dell'Istat)

Il **Sistema Informativo sulla Qualità** (SIQual) contiene informazioni sulle modalità di esecuzione delle rilevazioni ed elaborazioni condotte dall'Istat e sulle attività svolte per garantire la qualità dell'informazione statistica prodotta.

Il sistema è dedicato alla navigazione dei metadati che descrivono il processo produttivo e le sue caratteristiche: contenuto informativo, scomposizione in fasi e operazioni, attività di prevenzione, controllo e valutazione dell'errore.

In SIQual le rilevazioni, le elaborazioni e i sistemi informativi statistici sono più genericamente indicati come processi statistici. Nell'area "Documenti" è disponibile la documentazione sia di carattere generale che di interesse specifico per un processo. Se il processo statistico rilascia dati on line, attraverso SIQual è possibile consultare tali informazioni.

Strumenti

- Metodi e software
  - Software
- Definizioni e classificazioni
- Qualita dei dati
- Cartografia
- Dossier

Istat

Linee guida

per la qualità dei processi statistici

Giugno 2011

A cura di:
Marina Signore, Giovanna Brancato, Riccardo Carbini, Marcello D'Orazio, Giorgia Simeoni

Istat — Sistema Informativo sulla Qualità dei Processi Statistici

SIQual

Scelta guidata

Ricerca multidimensionale

Elenco completo

Documenti

http://www.istat.it/it/strumenti/qualita-dei-dati

Esempio:

Istat — Sistema Informativo sulla Qualità dei Processi Statistici

Home | Glossario | Scelta guidata | Ricerca multidimensionale

English version | Suggerimenti | FAQ

Scelta Guidata | Ricerca Multidimensionale | **Elenco completo** | Stampa | > home > Elenco completo

Sono elencati tutti i processi produttivi statistici dell'Istat documentati nel sistema. È possibile effettuare una ricerca semplice per porzione di parola sul "Nome del caratteristiche del processo produttivo statistico.

Cerca

Stato Attivo Sospeso Cessato Tutti

Periodicità

Tipo Processo Rilevazione Elaborazione Sistema Informativo

Numero Processi: 232

[R] – Aggiornamento dell'archivio statistico delle unità economiche Asia - Unità locali
[R] – Aggiornamento delle basi territoriali di Census 2000
[E] – Aggiornamento, miglioramento ed ampliamento della banca dati indicatori comunali
[S] – Agricoltura e zootecnia
[R] – Agriturismo
[E] – Analisi della povertà assoluta

Presentazione Processo

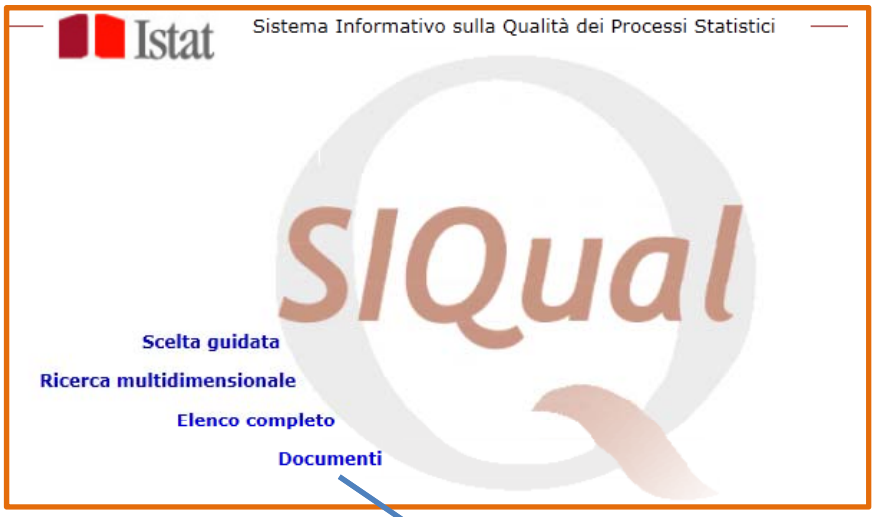
Approfondimenti
Dati Sintesi
Sorgenti normative
Fenomeni osservati
Fonti dell'elaborazione
Unità di analisi
Fasi del processo produttivo
Software

Diffusione Dati
Disaggregazioni
Pubblicazioni
Dati on-line

Report di stampa

[E] - Aggiornamento, miglioramento ed ampliamento della banca dati indicatori comunali	
Descrizione	Database di dati comunali provenienti da fonti ufficiali che permette la consultazione, l'esportazione e la rappresentazione cartografica di informazioni relative a: Censimenti (dal 1971 per Popolazione e abitazioni e Industria e Servizi, dal 1990 per Agricoltura), Territorio, Popolazione, Sanità, Istruzione, Turismo, Cultura, Credito, Veicoli circolanti. I dati possono essere estratti e consultati secondo una serie molto ampia di partizioni territoriali predefinite (regioni, province, comuni, comunità montane, sistemi locali del lavoro, aziende sanitarie locali, ecc.), oppure attraverso selezioni personalizzate di comuni.
Classificazione Eurostat del tipo processo	Statistiche derivate in ambito diverso dalla Contabilità Nazionale
Primo anno di produzione	1971

http://siqual.istat.it/SIQual/welcome.do



Documenti

Documenti metodologici
Documentazione operativa
Documentazione Eurostat
Regolamenti europei
Questionari
Documentazione sulla qualità
Documenti approfonditi sulla qualità

Thesauri

Il disegno di ricerca

Con il termine “disegno di ricerca” ci si riferisce a tutti gli aspetti inerenti alla pianificazione e conduzione di una ricerca, comprese le fasi di analisi dei dati raccolti e di presentazione l’interpretazione finale dei risultati.

Le principali fasi di una indagine statistica sono:

- **Definizione degli scopi della ricerca**
- **Definizione della tipologia di studio**
- **Definizione della popolazione (e selezione eventuale del campione)**
- **Definizione delle variabili di interesse**
- **Definizione della tecnica di indagine (costruzione dello strumento e delle modalità di raccolta dat)**
- **Raccolta dati**
- **Registrazione e pulizia dei dati**
- **Analisi statistica**
- **Presentazione e interpretazione dei risultati**

Gruppo di progettazione

L'elevato grado di complessità della fase di progettazione e la multidisciplinarietà delle conoscenze richieste rende indispensabile la formazione di un gruppo di lavoro in cui tutti gli aspetti, da quelli concettuali a quelli operativi, sino affrontati alla presenza di tutti i rappresentanti delle diverse aree di esperienza coinvolte

E' necessario comprendere nel gruppo di lavoro in cui tutti gli aspetti, da quelli concettuali a quelli operativi, siano affrontati alla presenza di tutti i rappresentanti delle diverse aree di esperienza coinvolte.

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. Particolare attenzione nell'ambito del gruppo dovrà essere data all'integrazione fra i concetti e le procedure definiti, per assicurare il funzionamento dell'intero sistema e non solo la coerenza interna delle singole parti di esso. L'attività del gruppo di progettazione deve scaturire in un documento di progettazione nel quale vengono descritti nel dettaglio tutti gli aspetti relativi alla progettazione.

Risorse e vincoli nel disegno di ricerca

Nella progettazione e conduzione di una indagine statistica sono da considerare risorse e vincoli di diverso tipo che condizionano le varie fasi della ricerca, tra i quali:

- **Risorse e vincoli finanziari della ricerca:** riguardano le risorse necessarie per condurre ciascuna delle varie fasi di ricerca, da quelle iniziali (es. ricerca pilota), alla preparazione del materiale necessario (es. stampa questionari), alla scelta dell'ampiezza del campione, alla fase di analisi dei dati.

- **Vincoli temporali:** possono riguardare i contenuti scientifici della ricerca (ad esempio se i dati non vengono raccolti entro un certo intervallo di tempo, cambia il significato o la rilevanza dei risultati), o possono riguardare aspetti tecnici (ad esempio le risorse umane e finanziare sono disponibili solo in un certo intervallo di tempo)

Gli scopi della ricerca

Le indagini statistiche si distinguono in relazione alle finalità della ricerca che possono essere di tipo:

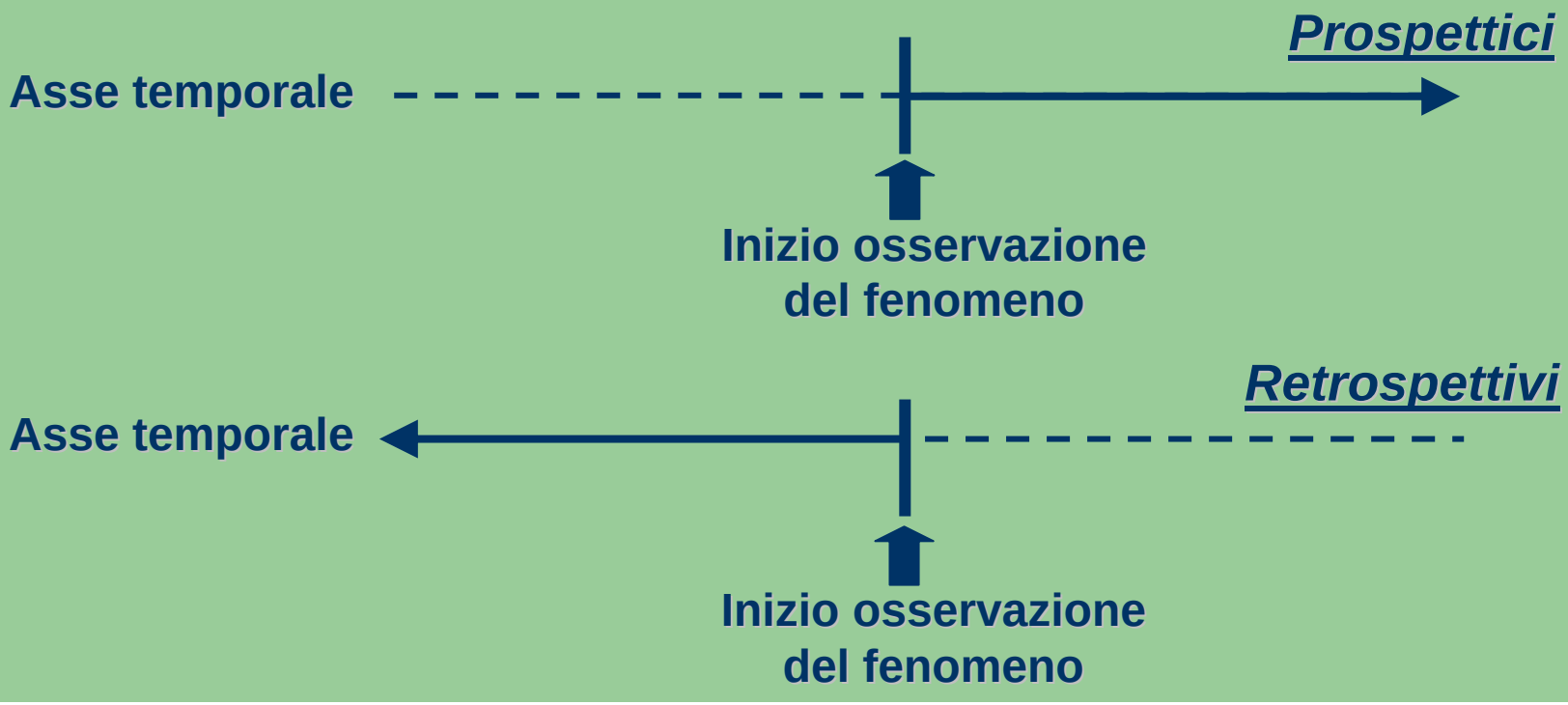
- **Descrittivo o esplorativo**: l'interesse del ricercatore si limita alla descrizione del fenomeno senza formulare e verificare ipotesi su quanto osservato (es. sondaggi, indagini pilota, …)

- **Esplicativo o analitico:** l'interesse del ricercatore va oltre la descrizione del fenomeno nel campione selezionato, e ricerca possibili relazioni tra variabili, generalizzabili all'intera popolazione.

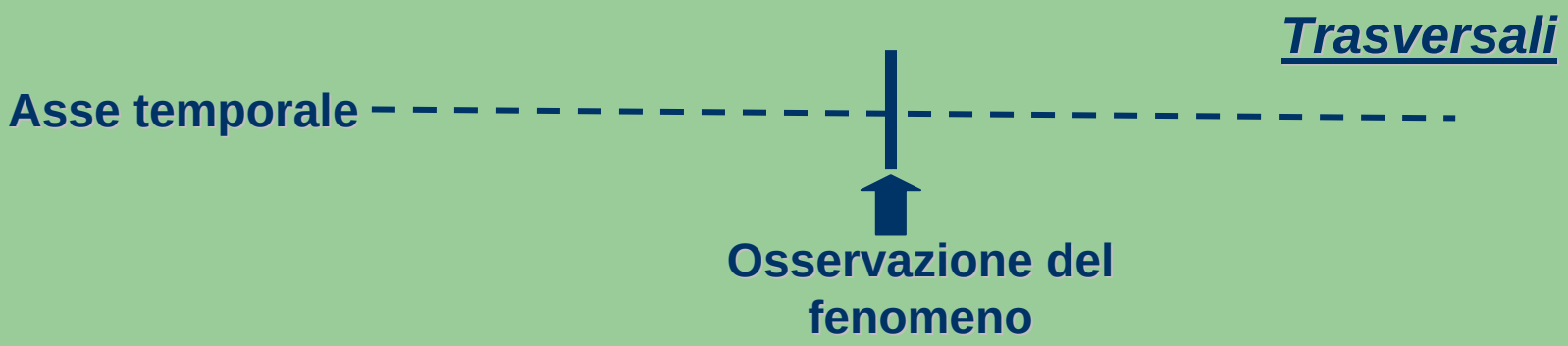
Nel definire il disegno di ricerca, il ricercatore deve valutare qual è la tipologia di studio più idonea rispetto allo scopo della ricerca e rispetto alle risorse e ai vincoli esistenti. Gli studi si distinguono in:

- **Studi longitudinali (es. indagini panel)**: si prevedono per ogni unità di rilevazione più misurazioni nel tempo (misure ripetute). Si misura ad esempio come variano nel tempo le preferenze dei consumatori o gli atteggiamenti rispetto ad una data tematica, …).

Gli studi longitudinali possono essere di due tipologie:



- **Studi trasversali:** viene effettuata una unica misurazione nel tempo. Rappresentano gli studi più frequenti.



Gli studi si distinguono anche in:

Studi osservazionali

Osservazione passiva di una realtà

Es. Si rilevano le opinioni di un campione di consumatori circa due diverse campagne pubblicitarie.

Ridotta formalizzazione metodologica

Problemi:
-problematicità dell'applicazione di procedure statistiche inferenziali

Vantaggi:
-possibilità di studiare problematiche non esaminabili tramite studi sperimentali.

Studi sperimentali

Intervento sperimentale del ricercatore

Es. Per valutare l'efficacia di due diverse campagne pubblicitarie, si selezionano casualmente due campioni di consumatori a ciascuno dei quali viene somministrata in modo random una campagna.

Approccio metodologico rigoroso

Vantaggi:
-diretta applicazione delle procedure statistiche inferenziali

Problemi:
-fattibilità dello studio / appropriatezza del metodo.

Indagini censuarie e campionarie

La conoscenza e la valutazione statistica di un fenomeno di interesse può essere derivata da:

- **Rilevazioni complete** (es. censimenti)

  Non sempre attuabili per problemi di costo, tempi non rapidi di ottenimento dati, distruzione delle unità osservate (es. controlli di qualità), …

- **Rilevazioni parziali** (indagini campionarie)

Sia le rilevazioni complete che quelle parziali possono fornire informazioni distorte della realtà a causa di:

- Distorsioni delle informazioni rilevate legate a:
  - ✓ Comportamento del rilevatore (es. quesiti proposti con orientamento alla risposta, …)
  - ✓ Comportamento del rispondente (es. ritrosia nelle risposte, scarso scrupolo nel fornire le risposte, …)
- Copertura non totale delle unità o presenza di dati mancanti
- Errori nella fase di registrazione, codifica, elaborazione dei dati

Per le rilevazioni parziali si aggiunge l'errore di campionamento.

Tale errore può essere stimato o meno a seconda che il campione sia **casuale** o **non casuale**.

B. Disegno di indagine

B.1. Carico statistico sui rispondenti

Principio B.1. Carico statistico sui rispondenti
È necessario rivolgere particolare attenzione alla minimizzazione del carico statistico sui rispondenti. Laddove possibile bisogna predisporre strumenti per la valutazione del suddetto carico.

B.2. Disegno di indagine per rilevazioni esaustive

Principio B.2. Disegno di indagine per le rilevazioni esaustive
Il ricorso ad una rilevazione esaustiva deve essere giustificato. In queste occasioni si deve valutare l'impatto di questa scelta rispetto ad una rilevazione campionaria in termini di costi, carico statistico sui rispondenti, accuratezza e dettaglio delle stime prodotte.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

Definizione della popolazione e delle variabili di interesse

Fenomeno di interesse. Delimitare precisamente cosa interessa da cosa non interessa ricordando che più ampio è l'arco degli argomenti trattati, maggiori divengono 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. Specificare quali ipotesi si intende sottoporre a verifica.

Popolazione di riferimento. Individua con precisione l'insieme di unità statistiche alle quali si intende estendere i risultati dell'indagine. Specificare esattamente le condizioni di eleggibilità, ovvero le caratteristiche che determinano l'inclusione (o l'esclusione) delle unità statistiche della popolazione. Discutere la possibilità reale di verificare le condizioni di eleggibilità sulle unità statistiche e giungere a condizioni che rappresentino un compromesso tra correttezza teorica ed effettiva praticabilità.

Variabili studiate. Misure di caratteristiche, solitamente elementari, riferite alle unità statistiche. Si raggruppano concettualmente in quattro grandi classi:

- Qualitative sconnesse. Assumono un insieme finito di categorie mutuamente esclusive tali che, per due differenti unità statistiche, si può definire soltanto se queste assumono la stessa o differenti categorie (sesso, stato civile)
- Qualitative ordinali. Assumono un insieme finito di categorie mutuamente esclusive tali da poter ordinare due unità statistiche secondo il possesso di caratteristiche possedute (grado di istruzione, grado di soddisfazione)
- Quantitative discrete. La caratteristica può essere descritta mediante un numero finito o infinito numerabile di valori numerici fra i quali abbia senso calcolare una differenza e/o un rapporto (numero di figli)
- Quantitative continue. La caratteristica può essere descritta mediante un'infinità non numerabile di valori fra i quali abbia senso calcolare una differenza e/o un rapporto (fatturato d'impresa)

La definizione delle variabili dovrebbe procedere attraverso una progressiva identificazione e raffinamento del fenomeno di interesse nelle sue componenti fino ad identificare gli aspetti salienti. L'obiettivo di tale procedimento dall'alto verso il basso serve a definire delle caratteristiche immediatamente utili all'obiettivo della ricerca. D'altro canto è necessario predisporre un analogo meccanismo dal basso verso l'alto considerando che le caratteristiche che si vogliono conoscere siano effettivamente misurabili sulle unità statistiche da indagare.

Classificazioni. Insieme delle categorie assunte da una variabile qualitativa sconnessa o ordinale. Definire una classificazione è un momento particolarmente critico. Ad esempio misurare il gradimento di uno spettacolo ricorrendo a quattro anziché a cinque categorie (ma anche denominando in modo appena diverso le stesse cinque categorie) può fornire risultati addirittura opposti. E' quindi opportuno, soprattutto se si desidera confrontare i risultati dell'indagine con altre fonti di informazione disponibili, ricorrere a classificazioni comunemente utilizzate. Per alcune variabili particolarmente complesse da definire (attività economiche, professioni, malattie) sono disponibili classificazioni standard riconosciute a livello internazionale.

http://www.istat.it/it/strumenti/definizioni-e-classificazioni

A. Obiettivi dell'indagine

A.1. Esigenze informative, utenti e usi

Principio A.1. Esigenze informative, utenti e usi
Le esigenze informative che si vogliono soddisfare devono essere ben definite. Contestualmente bisogna identificare i possibili utenti e usi dei risultati dell'indagine. La rilevazione deve essere quindi progettata in modo da soddisfare le esigenze degli utenti principali.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

Tecniche di indagine

Con il termine tecnica di indagine si intende l'insieme delle modalità di contatto delle unità statistiche interessate dalla rilevazione e di reperimento delle informazioni oggetto di interesse. La scelta della tecnica di indagine più idonea a raccogliere le informazioni oggetto della ricerca è uno degli aspetti di maggiore importanza nella pianificazione e nell'esecuzione di una indagine ed è strettamente connessa ad altre caratteristiche quali il fenomeno indagato, gli archivi di base, il strategia di campionamento, l'organizzazione del personale sul campo, i costi e i tempi attesi.

Inoltre non sono da sottovalutare le implicazioni della tecnica di indagine prescelta sulla qualità dei dati, in termini di mancate risposte e di errori di misura.

- La complessità delle scelte e le relazioni sopra menzionate possono essere facilmente illustrate mediante qualche esempio:
- Il contatto postale è difficilmente eseguibile se non si dispone di una lista di indirizzi affidabile. In questo caso è meglio ricorrere ad una indagine areale;
- Se si vogliono ottenere alti tassi di risposta è meglio ricorrere ad interviste personali condotte da rilevatori esperti;
- Domande su argomenti delicati (es. reddito, comportamenti sessuali, reati contro la persona) sono sottoposte a minore reticenza se condotte per telefono o mediante un questionario autocompilato;

Nel seguito elenchiamo le principali tecniche di indagine in uso per condurre una rilevazione, considerandone i più importanti vantaggi ed aspetti critici:

- Intervista diretta (o faccia a faccia);
- Intervista telefonica;
- Questionario postale autocompilato;
- Diario;
- Dati amministrativi;
- Osservazione diretta;
- Tecniche miste;
- Nuove tecnologie.

Intervista diretta (o faccia a faccia)

l'intervista viene condotta da un rilevatore che legge le domande e le opzioni di risposta nell'esatto ordine e con lo stesso linguaggio adottati nel questionario riportandovi quindi le risposte cosi come sono fornite dal rispondente.

Vantaggi

- Si presta meglio ad alcuni disegni di indagine (es.: censimenti e campionamento areale)
- Maggiore possibilità di contattare e convincere il rispondente a collaborare
- Si identifica esattamente il rispondente
- Possibilità di istruire il rispondente sul significato delle domande e sul modo corretto di fornire le risposte
- Flessibilità negli strumenti utilizzabili (audiovisivi, sezioni autocompilate, tecniche di probing,...)
- Interviste di maggiore durata

Svantaggi

- Costosa da implementare
- Necessita di una organizzazione capillare sul territorio
- Richiede tempi più lunghi di altri metodi per la raccolta dei dati
- Maggiori rischi di condizionamento

Intervista telefonica

L'intervista viene condotta al telefono da un intervistatore che legge le domande e le opzioni di risposta nell'esatto ordine e con lo stesso linguaggio adottati nel questionario riportandovi quindi le risposte così come sono fornite dal rispondente.

Vantaggi

- Costi minori rispetto all'intervista faccia a faccia
- Tempestività della raccolta dati
- Non è richiesta un'organizzazione sul territorio
- Maggiore possibilità di controllo dell'operato dei rilevatori
- Possibilità di contatto anche per le persone che non si trovano in casa in orari "canonici"
- Bassi rischi di condizionamento e maggiore possibilità di porre quesiti delicati

Svantaggi

- Impossibilità di contattare le famiglie senza telefono
- Il rispondente non è identificato con certezza
- Limitazioni nella lunghezza del questionario e nell'aiuto fornito ai rispondenti

Questionario postale autocompilato

Il rispondente riceve il questionario a mezzo posta o corriere e provvede a compilarlo nelle parti ad esso spettanti e a rispedirlo indietro o eventualmente a riconsegnarlo ad un addetto che lo ritira a domicilio.

Vantaggi

- Bassi costi di realizzazione
- E' richiesta un'organizzazione minore
- Bassi rischi di condizionamento
- Adatta per porre quesiti delicati
- Disponibilità di tempo per reperire eventuale documentazione necessaria alla compilazione
- Possibile sottoporre più categorie di risposta

Svantaggi

- Tempi lunghi di raccolta
- Impossibilità di identificare con certezza il rispondente
- Autoselezione dei rispondenti
- Minore capacità di ottenere la partecipazione all'indagine (il tema deve essere coinvolgente)
- Più difficile aiutare i rispondenti nella comprensione delle domande e nella compilazione del questionario (importanza della grafica)

Diario

E' un particolare tipo di questionario strutturato appositamente per registrare eventi frequenti e di scarsa importanza quali spese di bassa entità o attività quotidiane. L'organizzazione di tale strumento è tale da permettere la registrazione degli eventi nel momento della giornata in cui essi avvengono in modo tale da non dover ricorrere ad uno sforzo di memoria, con una conseguente sottonotifica degli eventi, nello svolgimento di una intervista di tipo classico.

Vantaggi

- Non affetto da problemi di memoria per la rilevazione di eventi poco rilevanti e ad elevata frequenza (ad esempio: spese giornaliere, uso del tempo, visione di programmi TV)

Svantaggi

- Struttura del questionario complessa
- Sottonotifica degli eventi col passare del tempo di osservazione
- Rischi di condizionamento dei comportamenti da registrare
- Necessita di un rilevatore per la consegna, il ritiro e il supporto alla compilazione

Dati amministrativi

Dati, riferiti a soggetti individuali, raccolti allo scopo di intraprendere decisioni o azioni che riguardano gli individui medesimi (es. licenze, assicurazioni tributi, regolamenti, pagamenti,...).

Vantaggi

- Relativamente economici da utilizzare a fini statistici
- Nessun disturbo ai rispondenti
- Spesso riguardano la totalità della popolazione e sono utili per costituire archivi

Svantaggi

- Possibili distorsioni dovute alla non coincidenza fra le definizioni usate per i dati amministrativi e quelli interessanti ai fini statistici
- Le leggi che regolano la raccolta possono cambiare pregiudicando la confrontabilità dei dati nel tempo
- Lo statistico non è in grado di controllare la qualità della raccolta dei dati
- Le informazioni utili ai fini statistici sono spesso raccolte in modo inaccurato perché non di primaria importanza ai fini amministrativi

Osservazione diretta

L'informazione viene raccolta dal rilevatore per mezzo dei propri sensi o mediante strumenti di misurazione fisici (applicazioni in antropologia, psicologia, geologia, telerilevamento,...).

Vantaggi

- Preferibile qualora l'informazione fornita da un rispondente non sia considerata sufficientemente precisa (ambito sperimentale)

Svantaggi

- L'interazione fra osservatore e oggetto osservato riproduce gli stessi problemi di condizionamento che si possono riscontrare con l'uso di rilevatori

Tecniche miste

utilizzate quando una sola tecnica di rilevazione non si comporta bene in tutte le situazioni pratiche

Esempi di tecniche miste:

Indagine postale + indagine diretta sui non rispondenti all'indagine postale

Indagine telefonica + indagine diretta su coloro che non possiedono il telefono

Indagine diretta + questionario individuale

Diario + intervista finale

Prima intervista diretta e successive con modalità telefonica

Dati amministrativi + controllo campionario con questionario postale autocompilato

Nuove tecnologie a supporto delle tecniche di indagine

CATI (Computer Assisted Telephone Interviewing)

CAPI (Computer Assisted Personal Interviewing)

Il questionario è contenuto nel computer cosicché le domande vengono poste così come compaiono sullo schermo e le risposte sono registrate direttamente su supporto magnetico

Vantaggi

- Alcuni controlli di qualità sono eseguiti dal computer al momento dell'immissione con un conseguente risparmio nelle successive fasi di controllo di qualità
- Si gestiscono facilmente questionari molto articolati
- Possono essere predisposte formulazioni alternative delle domande
- Si accorciano i tempi di completamento dell'indagine (soprattutto nel CATI)

Svantaggi

- Occorre dotare i rilevatori di un Computer portatile (CAPI)
- E' necessario un maggiore addestramento dei rilevatori
- Problemi di hardware (CAPI - pesante, lento, batterie,...)

D. Acquisizione dei dati

D.2. Acquisizione diretta dei dati

Principio D.2.1. Scelta della tecnica di raccolta dei dati
Le tecniche e gli strumenti di raccolta dei dati devono essere scelti in funzione degli obiettivi dell'indagine e delle caratteristiche della popolazione oggetto di studio e in modo tale da massimizzare la qualità dei dati e, al tempo stesso, limitare il carico statistico sui rispondenti e i costi.

Principio D.2.2. Presenza di organi intermedi o società di rilevazione
Per ridurre la mancata risposta totale e ottenere un elevato livello di qualità delle informazioni acquisite, vanno intraprese misure che migliorino la cooperazione con organi intermedi alla rilevazione e vanno previste azioni di monitoraggio delle attività di tali organi.

Principio D.2.3. Rispondenti
Per garantire adeguati tassi di risposta ed un elevato livello di qualità delle informazioni acquisite, la partecipazione dei rispondenti dovrebbe essere incoraggiata, e misure oggettive sulla fase di rilevazione dovrebbero essere prodotte e valutate.

Principio D.2.4. Intervistatori
Gli intervistatori devono essere selezionati, formati e monitorati in maniera tale da garantire che i dati raccolti siano il più accurati possibile. È buona pratica che l'effetto intervistatore venga valutato attraverso studi ad hoc.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

Indagini basate su questionario

Molte ricerche in vari campi di indagine, si prefiggono di individuare, valutare e spiegare alcune caratteristiche degli individui o di altre unità statistiche di interesse, utilizzando il *questionario* come modalità di raccolta dei dati.

Si distinguono diverse tipologie di questionario:

- Questionario cartaceo
- Questionario informatizzato
  (Indagini CAPI – Computer Assisted Personal Interview)
  (Indagini CATI – Computer Assisted Telephone Interview)

Si distinguono diverse tipologie di compilazione del questionario:

- Autocompilazione o autosomministrazione (es. questionario postale)
- Compilazione mediante intervista diretta o telefonica
- Somministrazione mista mediante intervista e autocompilazione

Il questionario di indagine è lo strumento di misura designato a raccogliere le informazioni sulle variabili qualitative e quantitative oggetto di indagine. Il questionario deve essere visto come uno strumento di comunicazione finalizzato a facilitare l'interazione fra il ricercatore, il rilevatore e il rispondente. Affinché possa svolgere il suo ruolo occorre che il questionario sia uno strumento standardizzato; ovvero domande e comunicazione devono essere identiche per tutti i rispondenti al fine che le informazioni raccolte siano confrontabili fra loro.

Le operazioni che devono essere curate per la realizzazione di un questionario possono essere schematizzate come segue:

Definizione degli obiettivi e concettualizzazione

Definire esattamente quali sono i temi che interessano l'indagine escludendo quelli che non sono un interesse primario

Preparare la lista delle variabili (e non direttamente le domande) da raccogliere rispetto ai temi di interesse identificati in precedenza

Preparare un piano provvisorio delle analisi statistiche da compiere per accertarsi che i contenuti necessari allo studio siano tutti espressi

Redazione del questionario

Stabilire la successione logica dei temi trattati (le sezioni del questionario)

Predisporre le domande filtro

Definire la sequenza di domande su uno stesso tema

Formulare i quesiti

Decidere l'organizzazione delle risposte

Verifica del questionario

Prima di rilasciare la versione definitiva del questionario occorre valutare se:

1. risponde alle esigenze conoscitive dell'indagine;
2. sono state omesse domande;
3. i riferimenti spaziali e temporali dei quesiti sono sufficienti;
4. linguaggio e struttura delle domande sono adeguati;
5. è facilmente comprensibile per gli intervistati e semplice da gestire per gli intervistatori.

Occorre mettere in atto una serie di controlli

1. Revisione estesa da parte di esperti del fenomeno;
2. pre-test: rilevatori esperti intervistano un campione ragionato di individui per raccogliere elementi utili a valutare completezza, chiarezza e gestibilità del questionario;
3. test di alternative: si sperimentano versioni alternative del questionario su piccoli campioni indipendenti di unità statistiche;
4. indagine pilota: versione completa dell'indagine su scala ridotta per verificare il gradi di integrazione tra le fasi dell'indagine ed effettuare eventuali ultimi ritocchi anche sul questionario.

Contenuti di un questionario:

- **Fatti**

  (es. possesso di specifiche caratteristiche del rispondente, situazioni, avvenimenti, …)
- **Conoscenze**

  (informazioni che il rispondente possiede su un dato argomento)
- **Opinioni**

  (modo di pensare del rispondente su un dato argomento)
- **Atteggiamenti**

  (es. stili di vita)
- **Motivazioni**

  (informazioni che spiegano il comportamento e l'atteggiamento del rispondente)

La difficoltà di reperire informazioni aumenta generalmente passando dai fatti alle motivazioni

Fonti di variazione o errore nella qualità dei dati raccolti

Per il buon esito della rilevazione, la formulazione del questionario deve essere tale da dissipare resistenze, ambiguità, timori, incertezze interpretative e condizionamenti di ogni genere.

Le principali fonti di variazione o errore sono riconducibili in generale a:

- **Caratteristiche del questionario e del contesto in cui avviene la raccolta dati** (inadeguatezza delle domande, mancanza di privacy, presenza di rumore, …)
- **Caratteristiche del rispondente** (livello di istruzione, personalità, grado di motivazione, …)
- **Caratteristiche del rilevatore o dell'intervistatore** (professionalità, caratteristiche individuali, …)

Fonti di variazione o errore nella qualità dei dati raccolti

I principali tipi di errore o distorsioni nelle risposte sono:

- **Errori motivazionali**: il rispondente non è sufficientemente motivato ad assumere il ruolo di "informatore", per cui tende a dare risposte incomplete o non accurate, rispondendo ad esempio "*non so"* anche quando egli possiede di fatto la risposta
- **Errori sociali**: il rispondente fornisce risposte alla scopo di fornire anche inconsapevolmente una certa immagine di sé
- **Errori di memoria**: il rispondente ha difficoltà a recuperare in memoria le informazioni richieste
- **Errori di comunicazione**: il rispondente ad esempio non ha compreso appieno la domanda o la tematica oggetto di indagine, o il rilevatore non ha compreso adeguatamente la risposta
- **Errori strutturali**: il questionario o l'intervista è troppo lunga o troppo breve, le categorie previste per la risposta sono inadeguate, l'ordine delle domande è inadeguato, …

La classificazione delle domande può essere intesa secondo diversi criteri:

1)
- **Dirette o personali** (chiamano in causa direttamente il rispondente)
- **Indirette o impersonali** (informazioni che il rispondente fornisce su terzi)

2)
- **Semplici** (domande che prevedono una sola risposta)
- **Multiple** (domande che prevedono più risposte)

3)
- **Primarie o domande filtro** (domande che non dipendono da altre domande)
- **Secondarie** (domande che dipendono da altre domande)

4)
- **Aperte** (domande a risposta libera, non precodificata)
- **Chiuse** (domande a risposta preformulata)
- **Semichiuse** (prevedono la modalità “altro: specificare”)

5)
- **Di controllo** (domande poste per verificare l’attendibilità di risposte fornite in precedenza)

B. Disegno di indagine

B.4. Sviluppo e test del questionario

Principio B.4.1. Sviluppo del questionario
Il questionario deve essere disegnato in modo da raccogliere efficacemente le informazioni di interesse, contenere i soli quesiti necessari, non comportare un eccessivo carico statistico per i rispondenti, ed essere strutturato in modo da garantire che l'intervista sia scorrevole. Laddove possibile, dovrebbero essere utilizzate definizioni e classificazioni standard o ad esse riconducibili. Dovrebbe essere disegnato in modo da facilitare il successivo trattamento dei dati, per esempio la registrazione e la codifica.
La grafica dei questionari cartacei dovrebbe indurre nel suo utilizzatore una percezione positiva e aiutarlo nell'orientarsi tra le diverse sezioni. Le soluzioni grafiche adottate devono essere utilizzate in modo coerente all'interno del questionario.

Principio B.4.2. Sviluppo e test del questionario elettronico
Il questionario elettronico deve essere sviluppato in modo da sfruttare le potenzialità tecnologiche per la gestione degli svincoli e per i controlli di qualità on-line, senza appesantire eccessivamente lo svolgimento dell'intervista.

Principio B.4.3. Strategia di test del questionario
Questionari nuovi o nuovi quesiti/sezioni di questionari pre-esistenti devono essere valutati attraverso una strategia di test. Questionari già in uso in indagini correnti andrebbero periodicamente sottoposti a valutazione. La strategia di test o valutazione dovrebbe comprendere sia test informali e/o di laboratorio che test sul campo.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

Le scale di misura sono gli strumenti per rilevare le modalità con le quali una variabile rilevata da una domanda può manifestarsi:

- **Scala nominale** (la variabile si manifesta secondo due o più modalità qualitative non ordinabili)
- **Scala ordinale** (la variabile si manifesta secondo due o più modalità qualitative ordinabili)
- **Scala numerica** (la variabile si manifesta secondo modalità quantitative)

Le modalità scelte come categorie di risposta ad una data domanda, devono essere mutuamente escludentesi ed esaustive

Le scale di valutazione sono gli strumenti per indagare su orientamenti ideologici, opinioni, atteggiamenti, su un dato argomento.

Sono costituite da una batterie di asserzioni o domande (item) nei confronti delle quali gli intervistati devono precisare il proprio punto di vista in termini di accordo/disaccordo, soddisfazione/insoddisfazione, accettazione/rifiuto, etc.

Tra le più comuni:

Scala ordinale di Likert che prevede che per ogni item siano rilevati diversi livelli di accordo/disaccordo secondo ad esempio il seguente schema:

Modalità	Equivalente numerico
Molto accordo	**5**
Abbastanza accordo	**4**
Né d'accordo né contrario	**3**
Abbastanza contrario	**2**
Molto contrario	**1**

Tali scale possono prevedere un numero diverso di modalità, in genere variabile da quattro a sette.

La graduazione dei livelli può essere realizzata mediante opportuna verbalizzazione, o tramite simboli o in maniera grafica tramite una scala graduata sulla quale il rispondente riporterà la propria posizione e che sarà in seguito convertita in termini numerici:

D'accordo **contrario**

Esempio: Scala di Likert per la valutazione dell'atteggiamento della popolazione italiana nei confronti degli extracomunitari

Items	*Approvo decisamente*	*Approvo*	*Sono indeciso*	*Disapprovo*	*Disapprovo decisamente*
Gli extracomunitari tolgono lavoro agli italiani	5	4	3	2	1
Gli extracomunitari contribuiscono a peggiorare la qualità di vita degli italiani	5	4	3	2	1
Gli extracomunitari dovrebbero abitare in zone separate dagli italiani	5	4	3	2	1

. . .

N.B. Barrare la casella che corrisponde al proprio atteggiamento

Introduzione al campionamento

Campioni casuali o probabilistici

Consentono la stima dell'errore di campionamento e la "bontà" dei risultati.

Le unità dello studio sono selezionate con procedimenti equivalenti alla scelta di elementi da un insieme omogeneo, secondo un criterio che non comporti la selezione preferenziale di un particolare insieme di unità.

La selezione casuale può essere effettuata in "blocco", ossia senza ripetizione delle unità che comporranno il campione, oppure con ripetizione.

Per applicare un procedimento di estrazione casuale occorre che sia disponibile una lista completa di tutte le unità della popolazione o collettivo da cui si seleziona il campione. Ogni unità del collettivo ha quindi una probabilità nota di essere inclusa nel campione.

Alcune tipologie di campionamento casuale:

- **Campionamento casuale semplice**
- **Campionamento casuale stratificato**
- **Campionamento a cluster o a grappoli**
- **Campionamento sistematico**

Campione casuale semplice

Un campione casuale semplice di ampiezza *n* da una popolazione finita di ampiezza *N* è un campione selezionato in modo tale che ciascuna unità ha la stessa probabilità di essere estratta.

Dalla lista degli elementi della popolazione la procedura di selezione si basa sull'uso dei numeri casuali, ad esempio mediante le tavole dei numeri casuali.

Tali tavole contengono sequenze di numeri ottenuti con una selezione casuale ad esempio mediante elaboratore elettronico.

Campione casuale stratificato

La popolazione è inizialmente divisa in *H* gruppi detti strati. Da ciascuno strato *h* è quindi selezionato un campione casuale di ampiezza n_h.

Se la variabilità all'interno degli strati è minore della variabilità tra gli strati, il campionamento stratificato può condurre ad una precisione maggiore delle stime.

Ad esempio, una università vuole condurre una indagine sul salario iniziale dei propri laureati. Si vuol tener conto del fatto che sono attivi cinque indirizzi di studi: Accounting, Finance, Information system, Marketing e Operations management. La popolazione dei laureati sarà quindi divisa in cinque strati in corrispondenza dei diversi indirizzi e da ciascuno strato verrà selezionato un campione casuale di laureati.

Campionamento a cluster o grappoli

La popolazione è inizialmente divisa in *N* gruppi detti cluster che rappresentano delle unità ausiliarie che contengono le unità da rilevare (es. la popolazione di elettori viene divisa in base alle province di residenza che costituiscono i cluster).

Viene quindi selezionato un campione casuale di *n* cluster tra gli *N*. A questo punto si può procedere come segue:

Campionamento ad uno stadio: Tutte le unità di rilevazione contenute nel campione casuale di cluster costituiranno il campione finale.

Campionamento a due stadi: si selezionerà un campione casuale di unità di rilevazione da ogni cluster.

Campionamento sistematico

Il campionamento sistematico consiste nello scegliere casualmente le unità dalla popolazione tramite una selezione che segue un modulo fisso, predeterminato a priori.

In genere, scelto casualmente il termine iniziale r nell'intervallo $(1,N/n)$, si selezioneranno le successive unità applicando la procedura $r+ik$, con $i=0,1,\ldots,n-1$.

Ad esempio dovendo selezionare un campione di $n=50$ unità da una popolazione di $N=5000$ unità, verrà scelta casualmente la prima unità tra le prime $k=5000/50=100$ della lista, ad esempio la decima per cui $r=10$, si procederà a selezionare la seconda unità che sarà la $r+k=10+100=110^a$, la terza unità che sarà la $r+2xk=10+200=210^a$, e così via.

Campioni non casuali o non probabilistici

Non consentono la stima dell'errore di campionamento.
Tra gli schemi più frequenti:

- Campioni ragionati

  Sono utilizzati ampiamente nelle ricerche di mercato finalizzate ad esempio alla valutazione delle potenzialità di un nuovo prodotto o di una campagna pubblicitaria in settori particolari della popolazione o in aree territoriali di particolare interesse.

- Campioni per quote

  Si seleziona la popolazione oggetto dell'indagine secondo alcune variabili strutturali (es. sesso, classi di età, professione, ecc.) rispettando delle "quote" proporzionali.

B. Disegno di indagine

B.3. Disegno di campionamento

Principio B.3. Disegno di campionamento
La selezione del campione deve avvenire in accordo a criteri probabilistici ben definiti. Il ricorso a criteri di selezione non probabilistici deve essere giustificato. Il disegno e la dimensione del campione devono essere tali da garantire il livello di accuratezza prefissato per le variabili chiave in corrispondenza dei principali domini di studio.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

Organizzazione dei dati raccolti

La corretta organizzazione dei dati in banche dati (database)

Esempio: sono stati raccolti i dati relativi alla performance (1Yr$Ret=rendimento percentuale a un anno) di un campione di 194 fondi di investimento, suddivisi in 59 a capitalizzazione integrale (Object=1) e 135 misti (Object=2). Per una corretta ed efficace analisi statistica dei dati, essi devono essere strutturati secondo il seguente schema:

N	Fund	1Yr$Ret	Object
1	Alliance Capital A GrowInc	30.8	2
2	Berger SmCoGrow	29.9	1
3	Jurika & Voyles Kaufmann	28.9	1
4	Baron Funds BanRosSC	35.5	2
...	....	...	...
192	MainStay Inst MainPwrGr	36.1	2
193	Vanguard Index Inst	30.9	2
194	Vanguard Index 500	30.8	2

Nome Variabili

Unità statistica

Non devono esserci né righe né colonne completamente vuote. Se ci sono dei dati mancati essi vanno codificati in maniera appropriata (es. cella vuota).

Regole di costruzione dei database: la codifica delle variabili

Codice_cliente	Età	Sesso
1	85	Maschio
2	72 anni	m
3	81	Femmina
4	72	maschio
5	ANNI 77	F
6	78	2
7	70 e 5 mesi	M
8	74	Femmina
…	…	…

Codice_cliente	Età	Sesso
1	85	Maschio
2	72	Maschio
3	81	Femmina
4	72	Maschio
5	77	Femmina
6	78	Femmina
7	70	Maschio
8	74	Femmina
…	…	…

E. Trattamento dei dati

E.1. Codifica delle variabili testuali

Principio E.1. Codifica delle variabili testuali
La procedura di codifica, sia essa manuale, automatica o assistita da computer, deve garantire un elevato livello di qualità delle informazioni codificate. Misure oggettive della qualità della codifica andrebbero prodotte e valutate.

E.2. Registrazione

Principio E.2. Registrazione
La procedura di registrazione, sia essa da operatore o mediante lettura ottica, deve garantire un elevato livello di qualità delle informazioni registrate. Misure oggettive della qualità della registrazione andrebbero prodotte e valutate.

E.3. Identificazione e trattamento degli errori

Linee guida

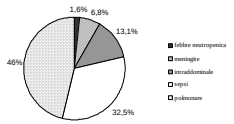
…

http://www.istat.it/it/strumenti/qualita-dei-dati

PERCORSO DI ANALISI STATISTICA

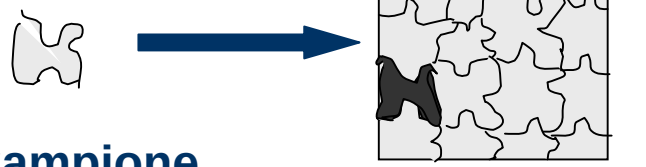
Analisi descrittiva del campione

N. fattori predisponenti	N. pazienti
1	273
2	188
3	106
4	56
Totale	623



Calcolo di statistiche descrittive: Frequenze, valori medi, indicatori di variabilità, di forma, …

Analisi inferenziale



Campione

Popolazione

Problemi di stima dei parametri, verifica di ipotesi sui parametri, modellistica, …

E. Trattamento dei dati

E.5. Processo di stima

Principio E.5. Processo di stima
Al fine della produzione delle stime, i dati raccolti e opportunamente trattati (codifica, editing, …) devono essere elaborati in accordo a metodologie consolidate e condivise a livello nazionale o internazionale. L'utilizzo in fase di elaborazione di informazioni ausiliarie attraverso modelli statistici deve essere motivata e le assunzioni alla base dei modelli devono essere esplicitate e ne deve essere valutata la effettiva validità. Le stime prodotte devono essere accompagnate da stime dell'errore che ne permettano un corretto uso e interpretazione.

E.8. Validazione dei risultati

Principio E.6. Validazione dei risultati
I risultati delle analisi, prima della pubblicazione, dovrebbero essere valutati insieme ad esperti del settore per verificare se vi siano o meno delle anomalie.
Laddove possibile, i risultati devono essere confrontati con gli stessi risultati ottenuti in precedenti replicazioni del processo o con risultati simili ottenuti al termine di altri processi nell'ambito dello stesso ente o provenienti da enti esterni.
Inoltre dovrebbero essere calcolati ed analizzati gli indicatori di qualità di processo e, nel caso di indagine campionaria, il livello di precisione delle stime diffuse, anche per valutare l'eventuale adozione di azioni di miglioramento nelle successive edizioni della rilevazione.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

F. Archiviazione, diffusione e documentazione

F.1. Archiviazione, diffusione dei dati e documentazione

Principio F.1. Archiviazione, diffusione dei dati e documentazione
I microdati validati devono essere archiviati secondo gli standard di Istituto prima della loro diffusione all'esterno.
I macrodati e i microdati diffusi devono essere preventivamente trattati per garantire una adeguata tutela della riservatezza.
Il calendario di diffusione dei risultati statistici deve essere reso pubblico.
Tutte le fasi del processo devono essere adeguatamente documentate.

Linee guida

…

http://www.istat.it/it/strumenti/qualita-dei-dati

PARTE I: QUALITÀ DI PROCESSO

...

Istat

Linee guida

per la qualità dei processi statistici

Giugno 2011

A cura di:
Marina Signore, Giovanna Brancato, Riccardo Carbini, Marcello D'Orazio, Giorgia Simeoni

PARTE II: QUALITÀ DI PRODOTTO

Definizione G.1. Rilevanza
*La **rilevanza** è definita come il grado in cui l'informazione statistica soddisfa le esigenze attuali e potenziali degli utenti. Essa comprende la completezza dell'informazione prodotta (tutte le statistiche necessarie agli utenti devono essere prodotte) e il livello in cui i concetti utilizzati (definizioni, classificazioni…) riflettono le esigenze degli utenti.*

Definizione G.2. Accuratezza
*L'**accuratezza** dei risultati viene definita dal punto di vista statistico come il grado di vicinanza tra le stime e i corrispondenti valori veri.*

Definizione G.3. Tempestività e puntualità
*La **tempestività** dei risultati è definita come il periodo di tempo che intercorre tra l'evento o il fenomeno che i risultati descrivono e il momento in cui gli stessi vengono resi disponibili.*
*La **puntualità** è definita come il periodo di tempo che intercorre tra la data di rilascio dei dati e la data di rilascio programmata, quest'ultima può essere annunciata dal calendario ufficiale di diffusione, stabilita da un Regolamento oppure frutto di un accordo preventivo tra partner.*

http://www.istat.it/it/strumenti/qualita-dei-dati

Definizione G.4. Coerenza e comparabilità

*La **coerenza** tra due o più statistiche si riferisce a quanto i processi statistici che le hanno prodotte hanno utilizzato i medesimi concetti – classificazioni, definizioni e popolazioni obiettivo – e metodi armonizzati. Statistiche coerenti possono essere correttamente combinate e usate congiuntamente. Esempi di uso congiunto si hanno quando le statistiche fanno riferimento alla stessa popolazione, periodo di riferimento e regione, ma comprendono differenti gruppi di variabili (es. dati sull'occupazione e dati sulla produzione) o quando comprendono le stesse variabili (es. dati sull'occupazione) ma per diversi periodi, regioni o altri domini.*

*La **comparabilità** è definita come un caso particolare della coerenza e si riferisce all'ultimo esempio citato, dove le statistiche fanno riferimento alle stesse variabili e vengono combinate per fare confronti nel tempo, tra regioni o tra altri tipi di domini.*

Definizione G.5. Accessibilità e chiarezza

*L'**accessibilità** delle statistiche è la facilità con cui gli utenti possono ottenere i dati. Essa è determinata dalle condizioni attraverso cui gli utenti ottengono i dati: dove recarsi, come richiederli, tempi di consegna, politica dei prezzi, politica di diffusione, disponibilità di micro o macrodati, formati disponibili (carta, files, CD-ROM, Internet…).*

*La **chiarezza** delle statistiche è la facilità con cui gli utenti vengono messi in grado di capire i dati. Essa è determinata dal contesto informativo in cui vengono presentati i dati, se sono accompagnati da metadati appropriati, se vengono utilizzate illustrazioni quali grafici o mappe, se sono disponibili informazioni sull'accuratezza dei dati (incluse eventuali limitazioni d'uso) e fino a che punto viene fornita assistenza aggiuntiva dal produttore del dato.*