

COESIONE  
ITALIA



*A Scuola di*  
**OPENCOESIONE**

*Conoscere e preparare  
un'indagine di statistica ufficiale*

# INDICE

---

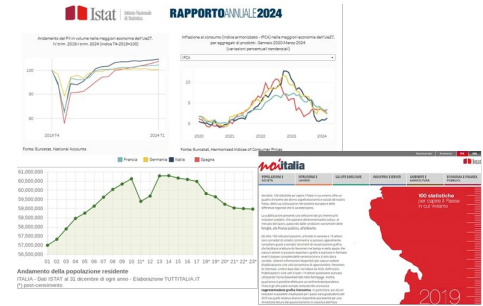
## ***Contenuti:***

- Introduzione alla statistica
- Le fasi di un'indagine statistica (approfondimento sulla progettazione)
- Esempi di indagine statistica
- Introduzione all'elaborazione dei dati
- La qualità dei dati (raccolti o utilizzati)

## ***Competenze acquisite:***

- Sapere come e dove cercare vari tipi di dati
- Discutere i vantaggi e svantaggi dell'utilizzo di un tipo di dato rispetto ad un altro
- Comprendere come progettare un'indagine statistica utile e concreta per la partecipazione e il controllo civico
- Esplorare le applicazioni pratiche dei dati nel monitoraggio civico

- La **conoscenza** dei dati consente ai cittadini di comprendere la realtà delle politiche o interventi, superando opinioni non fondate.
- Una corretta **lettura e interpretazione** dei dati permette di individuare criticità, successi e aree di miglioramento.
- La **diffusione** trasparente rende possibile un controllo partecipato, una responsabilizzazione delle istituzioni e stimola azioni di miglioramento condiviso.



I **DATI** (ossia i **NUMERI**) forniscono **INFORMAZIONI** su un problema o su una certa realtà in studio, ma non rappresentano di per sé conoscenza (**STATISTICHE**).

COMPITO DELLA STATISTICA “**dare forma ai dati**” (**CONOSCENZA**) mediante strumenti di modellazione del dato rigorosi che seguano regole (**FASI**) precise.

- **Statistica Descrittiva** **RIASSUME I DATI**. Esempio: l'abitante più anziano di una certa città ha 92 anni.
- **Statistica Inferenziale (o induttiva)** fa generalizzazioni e previsioni (**STIME**). Esempio: Se nel team ASOC del Lazio (campione) l'altezza media è di 172 cm, è possibile asserire che essa corrisponde all'altezza media di tutti i team ASOC di Italia (popolazione) ? Se sì, sono sicura al 100% della validità di questa asserzione?

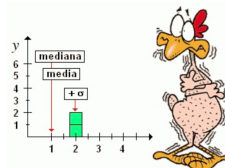
Per **INDAGINE STATISTICA** si intende un insieme di attività finalizzate ad approfondire la conoscenza di un fenomeno mediante alcuni passaggi (**FASI**):

- ❑ **TRADUZIONE** un concetto teorico (**carattere di interesse**) in un concetto operativo (**carattere misurabile** tramite variabili e categorie di misura) che è il **dato statistico**
- ❑ **RACCOLTA** informazioni sul comportamento del singolo componente (**unità statistica**) dell'insieme che si vuole indagare (**popolazione/fenomeno collettivo**)
- ❑ **TRATTAMENTO** delle informazioni raccolte:
  - **ELABORAZIONE** dei risultati dell'**osservazione** di uno o più **caratteri** del **fenomeno collettivo**
  - **ANALISI** dei dati raccolti (**quantitativi** e **qualitativi**) e **DIFFUSIONE** dei risultati ottenuti

Esempio

<b>Fenomeno collettivo</b>	Situazione occupazionale in Sardegna
<b>Popolazione</b>	Residenti in Sardegna in un certo momento
<b>Unità statistiche</b>	Singoli individui
<b>Caratteri</b>	Sesso, età, stato civile, condizione professionale, ...
<b>Osservazione</b>	Registrazione delle risposte ad un questionario
<b>Aspetti del fenomeno</b>	<ul style="list-style-type: none"><li>• Quota disoccupati</li><li>• Quota disoccupazione giovanile</li><li>• Durata della disoccupazione</li><li>• Relazione tra occupazione, disoccupazione e sesso</li><li>• ...</li></ul>

*La metafora del Pollo di Trilussa e gli inganni della Statistica*



Attività	Fase	Descrizione fase
<i>Decidere di monitorare la realizzazione di un'opera pubblica come una scuola o una pista ciclabile, stabilendo quali aspetti osservare (ad esempio, tempi di costruzione, rispetto del budget).</i>	<i>Definizione degli obiettivi/Ipotesi ricerca</i>	<i>Identificare cosa si vuole monitorare, quali progetti o interventi pubblici, e quali risultati valutare.</i>
<i>Effettuare sopralluoghi periodici, intervistare cittadini beneficiari, raccogliere dati da amministrazioni su stato di avanzamento lavori o spese sostenute.</i>	<i>Rilevazione dei dati</i>	<i>Raccolta delle informazioni sul campo tramite visite, interviste, questionari e documenti ufficiali.</i>
<i>Inserire le osservazioni e le risposte ricevute in tabelle strutturate, ad esempio in un foglio di calcolo dove si riportano date, riscontri e valutazioni raccolte.</i>	<i>Spoglio e tabulazione</i>	<i>Organizzazione e sistematizzazione dei dati raccolti per facilitarne l'analisi.</i>
<i>Calcolo di indicatori come percentuale di completamento lavori, confronto tra spese previste e sostenute, analisi qualitativa delle segnalazioni di criticità dagli utenti.</i>	<i>Elaborazione e analisi</i>	<i>Analisi dei dati per verificare coerenza, andamento e risultati rispetto agli obiettivi.</i>
<i>Redigere un report di monitoraggio civico con dati, foto, tabelle e raccomandazioni da presentare ai cittadini, media, e amministrazioni coinvolte.</i>	<i>Presentazione dei risultati</i>	<i>Creazione di report, grafici e comunicazioni chiare da condividere con comunità e istituzioni.</i>
<i>Confrontare i dati raccolti con le promesse iniziali, segnalare ritardi o sprechi, promuovere incontri pubblici per discutere criticità e proporre soluzioni insieme alle istituzioni.</i>	<i>Interpretazione e utilizzo</i>	<i>Utilizzare i risultati per agevolare processi decisionali, promuovere interventi correttivi o migliorativi.</i>

Nella progettazione e conduzione di una indagine statistica ci sono **RISORSE E VINCOLI** di diverso tipo che condizionano le varie fasi.

- ❖ **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 finanziarie sono disponibili solo in un certo intervallo di tempo)



## 1. I DATI GIÀ RACCOLTI DA ALTRI

- ❑ **I DATI SECONDARI** Sono i dati ottenuti dai ricercatori per uno scopo specifico ma che possono essere usati da altri ricercatori per scopi differenti o simili. Esempio le **statistiche ufficiali** ovvero prodotti da enti (prevalentemente) pubblici secondo canoni standard di qualità per soddisfare esigenze informative di pubblica utilità.
- ❑ **I DATI D'ARCHIVIO** Sono i dati raccolti e conservati da enti pubblici per gestire **servizi o procedimenti amministrativi** (es. registri, censimenti, archivi fiscali), utili perché spesso aggiornati sistematicamente e relativi alla popolazione totale o larga parte. Sono dati non raccolti per specifici scopi di ricerca. Esempio le **fonti amministrative** ovvero registrazioni di uno stato di fatto riguardanti persone fisiche o giuridiche del territorio di competenza per finalità fiscali, pensionistiche, giuridiche o anagrafiche.

## Le Banche Dati Tematiche della PA

### PERSONE



Anagrafe Nazionale della Popolazione Residente (ANPR)



Anagrafe tributaria gestita dall'Agenzia delle Entrate



Sistema Informativo del Casellario



Banca dati anagrafica dell'immigrazione

Fonte: FormezPa

### TERRITORIO



Banca dati catastale e ipotecaria

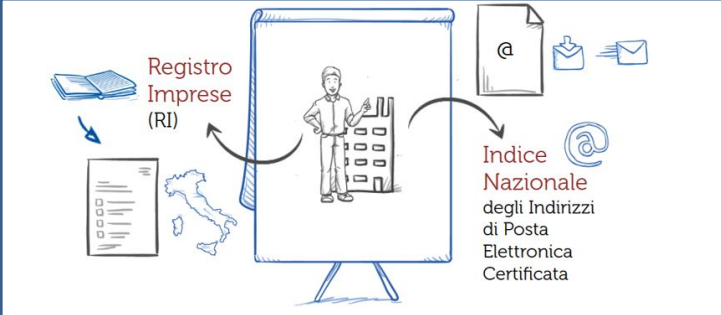


Archivio nazionale dei numeri civici delle strade urbane



SINFI, Sistema informativo nazionale federato delle infrastrutture

### IMPRESE



Registro Imprese (RI)

Indice Nazionale degli Indirizzi di Posta Elettronica Certificata

## 2. LE STATISTICHE RACCOLTE IN PRIMA PERSONA



### I DATI PRIMARI

Sono informazioni non esistenti, generate da un **processo di ricerca** per una indagine ad hoc: sono quelli raccolti attraverso attività in prima persona - come osservazione, registrazione, misurazione, monitoraggio di persone (e loro comportamenti), controllo di oggetti o eventi - per raggiungere obiettivi di ricerca specifici e generare nuove conoscenze.

#### Esempi di dati primari generati ad hoc

1. *Informazioni raccolte sul campo attraverso interviste dirette con cittadini, esperti, o stakeholder per valutare l'efficacia di un progetto pubblico, come nel caso della realizzazione di una pista ciclabile monitorata dall'inizio alla fine tramite osservazioni e valutazioni sul posto.*
2. *Dati quantitativi e qualitativi raccolti somministrando questionari per ottenere dalle comunità locali opinioni, impatti percepiti o risultati attesi di un intervento pubblico.*
3. *Informazioni su comportamenti e attività specifiche osservate in un territorio, come il traffico o l'uso di spazi pubblici, per verificare l'aderenza a quanto previsto da un progetto.*
4. *Dati geografici usati per creare mappe interattive che posizionano un fenomeno o un progetto in un contesto territoriale preciso, supportando la verifica locale dei risultati e l'analisi spaziale dell'impatto.*

### Dati Primari

Raccolti ad hoc per l'obiettivo specifico dell'indagine (esempio sondaggi, interviste, esperimenti). Controllo diretto su metodologia e contenuto.

**Specificità e Pertinenza:** rispondono esattamente alle domande di ricerca. Flessibilità e attualità delle informazioni.

**Costi e Tempi Elevati:** raccolta dispendiosa in termini di risorse e tempo. Rischio di errore (bias) del ricercatore nella fase di raccolta.

### Dati Secondari

Già esistenti, raccolti da terzi per scopi diversi (es. ISTAT, Eurostat, report di mercato, letteratura scientifica).

**Efficienza e Ampiezza:** rapidità di accesso, costi ridotti. Forniscono un quadro di riferimento ampio e validato.

**Pertinenza Limitata:** potrebbero non allinearsi perfettamente agli obiettivi specifici. Necessità di valutazione critica della fonte e della metodologia originale.

### Dati d'Archivio

Documenti storici, registri, verbali, corrispondenza (spesso non digitali). Offrono una prospettiva temporale e storica unica (monitoraggio di cambiamenti)

**Profondità Storica e Contesto:** permettono analisi longitudinali, tracciando l'evoluzione storica (andamenti/variazioni) dei fenomeni. Dettagli unici e non pubblicati altrove.

**Accessibilità e Qualità:** difficili da reperire e consultare (spesso manuali). Obsoleti, incompleti o difficile interpretabilità (es. scrittura antica) e accessibilità.



**I Big Data: i dati che potrebbero essere trasformati in opportunità future?**

Fase	Descrizione	Dati secondari/Archivio	Dati primari (indagine ad hoc)
<i>Definizione problema/Ipotesi ricerca</i>	<i>Identificazione del fenomeno e formulazione dell'ipotesi</i>	<i>attività comune*</i>	<i>attività comune*</i>
<i>Pianificazione</i>	<i>Scelta popolazione, variabili, strumenti</i>	<i>attività comune*</i>	<i>attività comune*</i>
<b>Raccolta dati</b>	<b>Acquisizione dei dati</b>	<b>Ricerca e verifica dati esistenti (fonti)</b>	<b>Progettazione campionamento e strumenti raccolta</b>
		<b>Estrazione e pulizia dati</b>	<b>Raccolta dati sul campo con questionari/interviste</b>
<i>Organizzazione dati</i>	<i>Codifica, classificazione, creazione dataset</i>	<i>attività comune*</i>	<i>attività comune*</i>
<i>Analisi statistica</i>	<i>Statistica descrittiva e inferenziale</i>	<i>attività comune*</i>	<i>attività comune*</i>
<i>Presentazione risultati</i>	<i>Report, grafici, conclusioni</i>	<i>attività comune*</i>	<i>attività comune*</i>

*\*= Fase che si svolge in ugual modo indipendentemente dalla fonte dei dati utilizzati per l'indagine statistica.*

L'integrazione dei dati storici offre un **grande valore** fornendo profondità storica e contesto documentale. Combinandoli con la specificità dei dati primari (i dettagli specifici) e l'ampiezza dei dati secondari (il contesto), si ottiene un'indagine **robusta e completa**= **Unire i dati, moltiplicare le conoscenze.**



**📌 Esempio: analizzare la soddisfazione della cittadinanza sui servizi pubblici locali**  
*Variabili: tipo di servizio usato, frequenza, grado di soddisfazione, suggerimenti per miglioramenti.*  
*Metodo: questionario somministrato a un campione casuale di cittadini.*  
*Analisi: confronto dati primari (questionari) con dati secondari (statistiche di utilizzo servizi) e dati d'archivio (storico cambiamento servizi).*

L'integrazione dei dati ha vantaggi e svantaggi/sfide

## VANTAGGI

- **Robustezza e Affidabilità** Triangolazione dei dati per validare i risultati tra fonti diverse.
- **Completezza** Combinazione tra attualità (primari), contesto ampio (secondari) e profondità storica (d'archivio).
- **Efficienza** Ottimizzazione delle risorse utilizzando dati esistenti prima di procedere alla raccolta primaria.
- **Insight Innovativi** Scoperta di correlazioni o tendenze emergenti dall'incrocio di prospettive diverse.

## SVANTAGGI/SFIDE

- **Compatibilità** Differenze nei formati, nelle definizioni e nelle metodologie di raccolta richiedono un lavoro intensivo di pulizia e armonizzazione dei dati.
- **Qualità Eterogenea** Mantenere un livello di affidabilità uniforme quando si combinano fonti con standard di qualità diversi.
- **Complessità Analitica** La gestione e l'analisi congiunta di set di dati eterogenei richiede competenze analitiche avanzate.

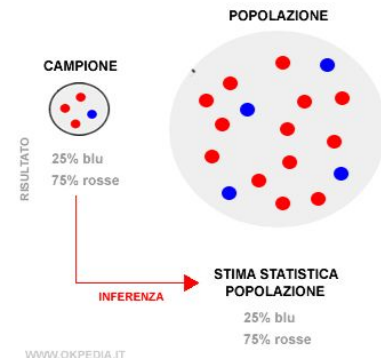
- ❑ Definizione del **disegno di indagine** ossia la specificazione della struttura metodologica complessiva del progetto di indagine (è il "come" dell'indagine).
- ❑ Individuazione della popolazione, progettazione della lista e estrazione del campione.

**Definire l'obiettivo:** cosa vuoi scoprire o monitorare? **Scegliere le variabili:** quali informazioni servono per raggiungere l'obiettivo? **Selezionare il campione:** chi o cosa sarà osservato? **Scegliere il metodo di raccolta dati:** questionari, interviste, osservazioni... **Pianificare tempistiche e risorse:** quanto tempo e quali strumenti servono? **Gestire la qualità nel processo di rilevazione:** precisione, validità, affidabilità dei dati...



La conoscenza e la valutazione statistica di un fenomeno può derivare da:

- ❖ **indagine totale o completa (censimento)** si ha la conoscenza esatta del fenomeno studiato: viene rilevata tutta la **popolazione** di interesse.
- ❖ **indagine parziale (campione)** si perviene a una **stima** (più o meno precisa) del fenomeno studiato: viene osservata una parte della popolazione di interesse.



## Tipo di indagine

## Vantaggi

## Svantaggi

### Indagine campionaria

Riduzione dei costi  
Tempi di rilevazione più brevi  
Minor carico organizzativo e minor errore di rilevazione  
Maggiore flessibilità organizzativa  
Possibilità di approfondire l'analisi in modo mirato

Minore precisione rispetto all'indagine totale  
Minore dettaglio nella disaggregazione territoriale  
Possibile perdita di rappresentatività del campione  
Richiede aggiornamento e attendibilità delle fonti  
Può portare a errori campionari e bias nella selezione

### Indagine totale (censimento)

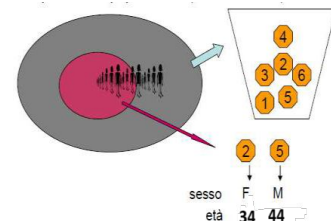
Elevato dettaglio di analisi  
Misure non affette da errore campionario  
Fornisce liste per indagini campionarie future

Costi elevati  
Tempi di realizzazione più lunghi  
Maggior carico organizzativo

**Indagine totale, approfondimento senza compromessi.  
Indagine campionaria, rapidità e costi contenuti.**



Il **piano di campionamento** rappresenta la parte operativa e metodologica del disegno con cui si definisce la strategia di estrazione del campione dalla popolazione target. Include le tecniche di selezione (casuale, stratificato, a probabilità differenziata, ecc.), la dimensione del campione, le modalità di estrazione e le probabilità di ogni unità di essere scelta.



**Scelta del campione:** decidere chi/intervistare o quali aree/tempi osservare, garantendo che il campione sia rappresentativo della popolazione o dell'area da indagare.

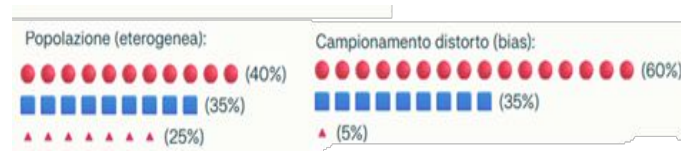
 **Esempio:** valutare come sono stati spesi i fondi pubblici per la riqualificazione dei parchi urbani in una città.

 **Obiettivo:** Verificare se gli interventi promessi sono stati realizzati e se i cittadini ne sono soddisfatti.

*Popolazione:* tutti i residenti della città.

*Campione:* 500 cittadini selezionati casualmente da diversi quartieri, con criteri di età, genere e zona di residenza per garantire la rappresentatività.

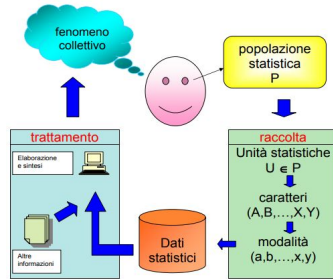
**Il disegno e la dimensione devono garantire livelli di accuratezza**



<b>Fattore</b>	<b>Descrizione</b>	<b>Esempio</b>	<b>Come influisce sulla dimensione del campione</b>
Popolazione	Numero totale di individui o unità della popolazione target.	Numero di cittadini residenti in un Comune oggetto di monitoraggio civico.	Popolazioni più grandi possono richiedere campioni più ampi, ma oltre una certa soglia la dimensione cresce poco.
Margine di errore	Precisione desiderata nelle stime, tipicamente espresso in percentuale.	Voler stimare con $\pm 3\%$ la percentuale di cittadini soddisfatti di un servizio pubblico.	Margini di errore più piccoli richiedono campioni più grandi per maggiore precisione.
Livello di confidenza	Grado di certezza che il campione rappresenti la popolazione (es. 90%, 95%, 99%).	Scelta di un livello di confidenza del 95% per dichiarare affidabili i risultati.	Livelli di confidenza più alti richiedono campioni più grandi.
Varianza o eterogeneità	Misura di quanto i valori nella popolazione variano tra loro (diversità delle risposte).	In un sondaggio civico con opinioni molto diverse fra i cittadini (alta varianza).	Popolazioni molto eterogenee richiedono campioni più grandi per rappresentare la variabilità.
Metodo di campionamento	Tecnica usata per selezionare il campione (casuale semplice, stratificato, cluster).	Stratificazione per zona geografica in città, per garantire rappresentanza di tutte le aree.	Metodi più complessi (es. cluster) possono ridurre o aumentare la dimensione necessaria.
Risorse e costi	Vincoli economici e temporali disponibili per l'indagine.	Budget limitato per piccole associazioni.	Risorse limitate possono imporre campioni più piccoli, accettando un margine di errore maggiore.

- ❖ Il campione è un sottoinsieme della popolazione.
- ❖ Viene utilizzato perché non è sempre possibile fare valutazioni sulla popolazione
- ❖ La tecnica di estrazione di un campione viene detta campionamento.
- ❖ Il tipo di campione viene scelto in base alle risorse disponibili e agli obiettivi della ricerca.

<b>Aspetto</b>	<b>Campionamento Probabilistico</b>	<b>Campionamento Non Probabilistico</b>
Definizione	<i>Metodo in cui ogni unità della popolazione ha una probabilità nota e diversa da zero di essere selezionata</i>	<i>Metodo non casuale in cui la selezione è basata su giudizio o convenienza</i>
Esempi	<i>Campionamento casuale semplice, stratificato, a grappolo, sistematico</i>	<i>Campionamento per quote, a convenienza, a scelta ragionata, a valanga</i>
Vantaggi	<i>Rappresentativo, consente inferenze statistiche valide, stime di errore</i>	<i>Facile, rapido, meno costoso, utile per ricerche esplorative o preliminari</i>
Svantaggi	<i>Più costoso e complesso, richiede lista completa della popolazione</i>	<i>Non rappresentativo, generalizzazioni limitate (non consente la stima dell'errore di campionamento)</i>
Quando usare	<i>Quando si desidera precisione e rappresentatività, e si hanno risorse adeguate</i>	<i>Quando si hanno vincoli di tempo e budget, o per studi esplorativi</i>



La **definizione operativa** trasforma concetti astratti in dati tramite la selezione di variabili appropriate, che classificano e misurano la realtà in modi coerenti con l'obiettivo della ricerca.

La **VARIABILE** (o **CARATTERE**) consente di classificare e raggruppare fatti o oggetti in categorie o valori numerici.

A seconda della natura del concetto e della realtà che si vuole descrivere, una variabile può essere:

- **QUALITATIVA** o **CATEGORIALE** quando le sue modalità sono espresse in forma verbale (sesso, livello di istruzione, gusto del gelato, ...).

A sua volta una variabile qualitativa può essere:

- **SCONNESSA** o **NOMINALE** se non esiste nessun ordinamento tra le modalità.

📌 *Esempio: la variabile gusto del gelato (cioccolato, vaniglia, fragola, .....)*

- **ORDINALE** se `e possibile individuare un ordinamento naturale delle modalità.

📌 *Esempio: la variabile giudizio (insufficiente, sufficiente, discreto, buono e ottimo).*



Oppure, una variabile può essere:

- **QUANTITATIVA** (o **NUMERICA**) quando le modalità sono espresse da numeri (età, peso, . . .) si identificano tramite numeri.

A sua volta una variabile quantitativa può essere:

- **DISCRETA** quando l'insieme delle modalità è finito o numerabile ed è ottenuto tramite un'operazione di conteggio (classe dei numeri naturali).

📌 *Esempio: la variabile numero di biglie (0,1,2.....).*



- **CONTINUA** quando l'insieme delle modalità è un intervallo (ossia un sottoinsieme) ottenuto tramite un'operazione di misurazione (classe dei numeri reali).

📌 *Esempio: la variabile altezza (in cm)*



La scelta di **variabili** richiede un'attenta valutazione della loro capacità di rappresentare correttamente il concetto (**validità**), di produrre risultati coerenti nel tempo (**affidabilità**), e di essere pertinenti e rappresentativi rispetto alla popolazione o al fenomeno studiato.

📌 *Esempio: la variabile attenzione e concentrazione misurata dal tempo in min che uno studente rimane concentrato senza distrarsi (esempio guarda altrove) durante una lezione.*

Tecnica	Vantaggi	Svantaggi	Ambito tipico di utilizzo
<b>Interviste faccia a faccia CAPI</b> (Computer Assisted Personal Interviewing)	Alto tasso di risposta, possibilità di chiarimenti, uso di supporti visivi, adatto a domande complesse o sensibili	Costoso, richiede tempo, personale formato, logistico complesso	Censimenti nazionali, ricerche etnografiche, popolazioni con bassa alfabetizzazione o accesso digitale
<b>Interviste telefoniche CATI</b> (Computer Assisted Telephone Interviewing)	Consente chiarimenti, meno costoso di faccia a faccia, più rapido, ampia copertura geografica	Nessun supporto visivo, possibile difficoltà nel creare rapporto, può avere tasso di risposta variabile	Sondaggi di opinione, ricerche di mercato, popolazioni ampiamente raggiungibili con telefono
<b>Sondaggi postali</b>	Copertura ampia, consente uso di supporti (seppur limitati)	Basso tasso di risposta, tempi lunghi per la raccolta dati	Rilevazioni con target geograficamente ampio ma poco accesso tecnologico
<b>Sondaggi online CAWI</b> (Computer Assisted Web Interviewing)	Molto veloce, costo contenuto, ampia scalabilità, possibilità di logica dinamica	Esclusione di popolazioni non connesse a internet, rischio di autoselezione, tasso di risposta variabile	Ricerche di mercato, feedback clienti, rilevazioni su popolazioni digitali
<b>Auto-compilazione assistita CASI</b> (Computer Assisted Self Interviewing)	Maggiore privacy, adatto a domande sensibili, risposte più sincere, riduce bias da intervistatore	Può necessitare supporto tecnico, esclude chi ha poca dimestichezza con il computer	Indagini su temi sensibili, salute, comportamenti personali
<b>Panel</b>	Disponibilità di panel reclutati, risposte rapide e continue	Può non essere rappresentativo, rischio di stanchezza rispondente	Monitoraggio continuo, ricerche commerciali, trend analysis

Il questionario è lo strumento (di misura) attraverso il quale vengono raccolti i dati (trasforma domande in conoscenza affidabile). La sua redazione rappresenta una delle fasi più delicate nella pianificazione dell'indagine: dalla sua formulazione dipende la validità delle informazioni raccolte e quindi la bontà dei risultati ottenuti.



- *Definizione obiettivi e concettualizzazione temi di interesse (fatti, conoscenze, opinioni, atteggiamenti, motivazioni)*
- *Preparazione lista di variabili da raccogliere rispetto ai temi identificati in precedenza*
  - *Stabilire successione logica dei temi trattati (sezioni del questionario)*
- *Definire sequenza di domande su uno stesso tema*
  - *Formulare quesiti*
  - *Predisporre domande filtro*
  - *Decidere organizzazione delle risposte*
- *Verificare questionario (prima della versione definitiva)*
  - *Preparazione piano provvisorio di analisi statistiche da compiere per verificare se tutti i contenuti dell'indagine sono espressi*

Le domande devono essere: - formulate in modo chiaro - orientate verso chi deve rispondere - impostate in modo sereno ed obiettivo, evitando ogni possibile fattore di distorsione delle risposte.

**La scelta della tecnica di somministrazione influenza la costruzione del questionario.**

Raccogliere dati precisi e affidabili che descrivano accuratamente le caratteristiche di interesse, consentendo così di comprendere il fenomeno studiato= **Misurare**

Assegnare valori (secondo scale di misura), che permettono di confrontare tra loro le osservazioni e di applicare metodi di analisi statistica appropriati= **Scalare**

A seconda del tipo di variabili osservate, sono possibili diverse analisi statistiche e strumenti statistici appositi per studiare tipi diversi di variabili=**VARIABILI QUALITATIVE vs QUANTITATIVE** (traduzione di espressioni verbali in numeri)

Le **scale di misura** sono gli strumenti per rilevare le modalità con cui una variabile rilevata da una domanda può manifestarsi. Tra i vari tipi di variabili è implicita una **gerarchia**.

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, etc.

		Operazioni		
		Uguaglianza/ Disuguaglianza	Ordinamento	Operazioni aritmetiche
<b>Caratteri</b>		=, ≠	>, <	+, -, *, /
<i>Qualitativi</i>	<i>sconnessi</i>	si	no	no
	<i>ordinabili</i>	si	si	no
<i>Quantitativi</i>		si	si	si

## Scala ordinale di Likert

<u>Modalità</u>	<u>Equivalente numerico</u>
Molto accordo	5
Abbastanza accordo	4
Né d'accordo né contrario	3
Abbastanza contrario	2
Molto contrario	1

I dati raccolti con le varie tecniche di rilevazione devono essere organizzati in una forma che ne permetta un'agevole analisi. A tal fine è pressoché indispensabile usare la matrice dei dati, definita anche matrice “**casi per variabili**” perché nella tabella ogni **riga** rappresenta un caso, ogni **colonna** una variabile (una proprietà) e ogni **cella** il valore rilevato per ogni caso su ogni variabile.

**Ogni riga un caso, ogni colonna una variabile: la matrice è il cuore dell'analisi!**

Uno, pochi, molti casi.....la matrice si complica quanto più grande è il numero  $n$  di osservazioni.

Qualitativa nominale

Qualitativa ordinale

ID	CORSO LAUREA	SESSO	MEDIA VOTI	CREDITI	RENDIMENTO
1	SAM	M	22	6	discreto
2	SAM	F	24	71	buono
3	SAM	M	21	19	discreto
4	SAM	F	26	27	buono
5	SAM	F	27	9	ottimo
6	SAM	M	26	10	buono
7	SAM	F	25	18	buono
8	SAM	M	24	27	buono
9	SAM	F	27	10	ottimo
10	SAM	F	24	17	buono
11	SAM	M	26	18	buono
12	SAM	M	30	18	ottimo
13	SAM	F	29	84	ottimo
14	SPO	M	27	27	ottimo
15	SPO	F	23	9	discreto
16	SPO	F	27	30	ottimo
17	SPO	M	28	33	ottimo
18	SPO	M	29	30	ottimo
19	SPO	F	28	48	ottimo
20	ORU	F	26	66	buono

Quantitativa continua  
(è una media!)

Quantitativa discreta (deriva  
da un conteggio)

– **L'analisi univariata** considera una sola variabile rilevata sulle unità.; **l'analisi bivariata** considera due variabili contemporaneamente; e così via. Lo studio congiunto di più variabili è detto **analisi multivariata** ed è per estensione, ogni forma di esplorazione e concettualizzazione dei dati raccolti.

Nella fase di **elaborazione** si applicano strumenti propri dell'analisi statistica al fine di ottenere **una sintesi e una descrizione dei dati** sotto forma di misure sintetiche (frequenze, indici di posizione, etc.) che permettono di trasformarli in informazioni utili.



La scelta del metodo da usare dipende dal tipo di fenomeno osservato.

## **Esempi: Numeri che parlano, decisioni che guidano**

<i>Indicatore</i>	<i>Misura</i>
<i>Percentuale di fasi del progetto completate</i>	<i>% di lavori eseguiti rispetto al cronoprogramma</i>
<i>Numero di incontri pubblici organizzati</i>	<i>Numero totale di incontri con cittadini durante i lavori</i>
<i>Percentuale di utenti soddisfatti</i>	<i>% di risposte positive a questionari di gradimento</i>
<i>Numero medio di visitatori settimanali</i>	<i>Conteggio tramite rilevazioni dirette</i>
<i>Percezione della sicurezza nel parco</i>	<i>Scala Likert (1-5) su sicurezza percepita dagli utenti</i>
<i>Numero di segnalazioni di danni o atti vandalici</i>	<i>Numero di segnalazioni ricevute in un mese</i>

**FREQUENZA ASSOLUTA ( $n_i$ )** numero di volte in cui si presenta una modalità della variabile.

$$0 \leq n_i \leq n$$
$$\sum_{i=1}^K n_i = n_1 + n_2 + \dots + n_K = n$$

$n$  = numero totale delle osservazioni  
 $K$  = numero dei valori/modalità/classi della variabile

**FREQUENZA RELATIVA: ( $p_i = n_i / n$ )** vengono calcolate come rapporto tra le frequenze assolute e la numerosità del collettivo. Vantaggio: confronto distribuzioni di frequenza basate su numeri differenti di unità statistiche.

$$0 \leq p_i \leq 1$$
$$\sum_{i=1}^K p_i = p_1 + p_2 + \dots + p_K = 1$$

**FREQUENZA RELATIVA PERCENTUALE: ( $p_i \% = n_i / n * 100$ )** vengono calcolate come prodotto delle frequenze relative per 100. Vantaggio: conto quanto volte un fenomeno si manifesta su 100 osservazioni.

$$0 \leq p_i \% \leq 100\%$$
$$\sum_{i=1}^K p_i \% = p_1\% + p_2\% + \dots + p_K\% = 100\%$$

**FREQUENZA CUMULATA ASSOLUTA E RELATIVA: ( $F_i$ );  $P_i = F_i / n$ ;  $P_i \% = F_i / n * 100\%$ )** numero di osservazioni il cui valore (o la cui %) è inferiore o uguale ad una data modalità o a un dato valore  $x_i$ .

**LA DISTRIBUZIONE DI FREQUENZA** (conteggio di quante volte ogni valore della variabile compare nei dati): alle modalità della variabile  $X$ , qualitativa o quantitativa, si associa il numero di volte  $n_i$  in cui ciascuna modalità si manifesta  $n_1, \dots, n_k$ . (Analisi Statistica Univariata).

$n$  = numero delle unità statistiche rilevate

$X$  = carattere oggetto di studio

$K$  = numero totale dei diversi valori assunti dalla variabile  $X$  (modalità)

$x_i$  = modalità  $i$ -esima del carattere  $X$   $i=1, \dots, k$

$n_i$  = frequenze assolute

$x_i$	$n_i$	$f_i$	$p_i$	$N_i$
$x_1$	$n_1$	$n_1/n=f_1$	$f_1*100$	$n_1$
$x_2$	$n_2$	$n_2/n=f_2$	$f_2*100$	$n_1+n_2$
$x_3$	$n_3$	$n_3/n=f_3$	$f_3*100$	$n_1+n_2+n_3=n$
	$n$	1	100	

$N_i = \sum_{j=1}^i n_j$   
 $\left\{ \begin{array}{l} N_1 = n_1 \\ N_k = n \\ N_i - N_{i-1} = n_i \end{array} \right.$

### Per caratteri quantitativi continui:

- Si raggruppano i valori in intervalli (classi);
- Gli intervalli non devono essere troppi né troppo pochi;
- L'ampiezza degli intervalli è preferibile che sia uguale per poter facilitare il confronto tra classi.

Fatturato mensile	Frequenze assolute
$[x_i - x_{i+1}[$	$n_i$
0 - 10	30
10 - 20	15
20 - 40	25
40 - 70	20
$\Sigma$	90

Partecipazione a:	Frequenza Assoluta (numero di persone)	Frequenza Relativa (quota del totale)	Frequenza Percentuale
nessuna partecipazione	15	$15/60 = 0,25$	25%
1 evento	30	$30/60 = 0,50$	50%
2 eventi	10	$10/60 \approx 0,17$	16,67%
3 o più eventi	5	$5/60 \approx 0,08$	8,33%
<b>Totale</b>	<b>60</b>	<b>1,00</b>	<b>100%</b>

### **Risultati e Interpretazione:**

*La maggior parte della popolazione monitorata (il 50%) ha partecipato ad un solo evento di quartiere, indicando un livello di coinvolgimento moderato.*

*Un quarto degli intervistati non ha partecipato a nessuna attività, mentre una piccola percentuale (8,33%) ha partecipato a tre o più eventi, segnalando un coinvolgimento più attivo.*

*Questi dati aiutano a capire meglio le abitudini di partecipazione e a pianificare future iniziative di coinvolgimento civico.*

Ore di volontariato	Frequenza Assoluta	Frequenza Cumulata	Frequenza Percentuale	Frequenza Percentuale Cumulata
0-5	10	10	16,70%	16,70%
6-10	20	30 (10+20)	33,30%	50,00%
11-15	15	45 (30+15)	25,00%	75,00%
16-20	10	55 (45+10)	16,70%	91,70%
21+	5	60 (55+5)	8,30%	100,00%
<b>Totale</b>	<b>60</b>		<b>100%</b>	

### **Risultati e Interpretazione:**

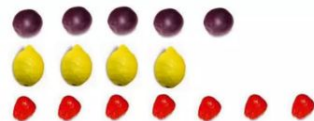
*La frequenza cumulata indica quanti cittadini dedicano fino a un certo numero di ore (ad es. fino a 10 ore sono 30 persone, il 50% del totale).*

*Le frequenze percentuali e cumulative aiutano a capire la distribuzione della disponibilità al volontariato nella popolazione osservata, mostrando ad esempio che il 75% della popolazione dedica fino a 15 ore.*

*Le frequenze cumulate sono uno strumento utile per sintetizzare e interpretare dati.*

## □ Moda o valore modale (variabile qualitativa sconnessa o ordinabile)

**DEF.** è la modalità della variabile alla quale è associata la **maggior frequenza**, cioè quella che si manifesta più volte in sede di rilevazione. Può essere calcolata per qualsiasi tipo di variabile. In una distribuzione è possibile individuare un solo valore modale (unimodale); possono esistere due valori che compaiono entrambi con la frequenza massima (distribuzione bimodale); ecc.

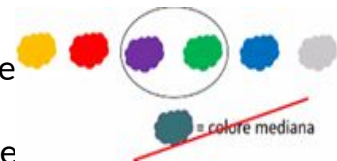


## □ Mediana (variabile quantitativa discreta o continua)

**DEF.** è la modalità (valore) che occupa la **posizione centrale** nella distribuzione ordinata della variabile (preceduto e seguito dallo stesso numero di osservazioni). Può essere calcolata per variabili almeno su scala ordinale (non su variabili sconnesse).

Per calcolare la mediana di  $n$  dati:

- si ordinano gli  $n$  di dati in ordine crescente o decrescente;
- se il numero di dati è dispari la mediana corrisponde al valore centrale, ovvero al valore che occupa la posizione  $(n + 1) / 2$ .
- se il numero  $n$  di dati è pari, la mediana è stimata utilizzando i due valori che occupano le posizioni  $(n / 2)$  e  $(n / 2 + 1)$  (generalmente si sceglie la semisomma dei due valori mediani).



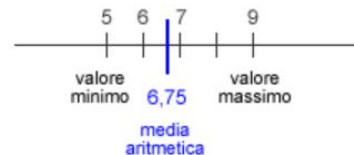
## Media (variabile quantitativa discreta o continua)

**DEF.** è la somma dei valori osservati divisi per il numero complessivo di valori (solo per variabili quantitative).

Se le modalità  $x_i$  si presentano ciascuna con frequenze assolute  $n_i$ , si calcola come il rapporto tra la somma dei prodotti di ciascun valore della variabile  $X$  per la rispettiva frequenza e la somma totale delle frequenze.

$$M(X) = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i}$$

- ❑ E' compreso tra il più piccolo e il più grande valore dei dati osservati.
- ❑ Non sempre coincide con un valore effettivo nella distribuzione



### POPOLARE O SAGGIA? La metafora sugli indici di posizione

- **La Media Democratica:** come una vera democrazia, dà voce a tutti i dati, sommando e dividendo per il numero totale, per un giudizio collettivo sull'insieme.
- **La Moda Trendy:** è il valore più popolare, quello che domina la scena come la tendenza del momento nel gruppo dei dati.
- **La Mediana Saggia:** divide i dati ordinati in due metà uguali, rimanendo imparziale e saggia, non influenzata dagli estremi o dai valori anomali.

**OBIETTIVO:** Chiarire cosa si vuole misurare (es. uso del parco, soddisfazione utenti, stato dei lavori) e individuare la popolazione coinvolta (cittadini residenti, beneficiari, operatori, tecnici).



**RACCOLTA E ORGANIZZAZIONE DEI DATI:** Raccogliere dati da fonti diverse e organizzarli in modo sistematico, possibilmente in un database o foglio di calcolo, per facilitare l'analisi.

- **Strumenti di raccolta dati:**
  - Questionari
  - Osservazioni dirette o visite sul campo per valutare lo stato reale delle cose.
- **Integrazione dati ufficiali** (es. dati open data del Comune) con dati raccolti sul campo tramite intervista o questionari civici.

## ANALISI QUALITATIVA E QUANTITATIVA DEI DATI - PIANO DI ANALISI

- **Qualitativa:** codificare e sintetizzare le risposte aperte per individuare temi ricorrenti, problemi segnalati e suggerimenti
- **Quantitativa:** calcolare indicatori (frequenze, medie, percentuali) per le risposte chiuse (es. scala di valutazione, frequenza di uso del servizio).

<b>Tipi di dati</b>	<b>Esempi</b>	<b>Riferimenti bibliografici / fonti</b>
<b>Secondari</b>	<ul style="list-style-type: none"><li>-Piani e progetti di riqualificazione urbana a Roma, es. "100 Parchi per Roma" (piani, finanziamenti, tempistiche)</li><li>-Statistiche demografiche e socioeconomiche del quartiere</li><li>-Censimenti alberature e verde urbano</li><li>-Rapporto Ecosistema Urbano 2024 con indicatori ambientali e urbani</li></ul>	<ul style="list-style-type: none"><li>-Comune di Roma, Piano "100 Parchi per Roma" (2024)</li><li>-Rapporto Ecosistema Urbano 2024 di Legambiente</li><li>-.....</li></ul>
<b>D'archivio</b>	<ul style="list-style-type: none"><li>-Documenti storici urbanistici e mappe catastali</li><li>-Relazioni tecniche e amministrative dei precedenti interventi</li><li>-Schede tecniche di parchi e aree verdi comunali</li><li>-Dati statistici storici sull'utilizzo e manutenzione del parco</li></ul>	<ul style="list-style-type: none"><li>-Archivio urbanistico Comune di Roma</li><li>-Documentazione storica rigenerazione urbana</li><li>-.....</li></ul>
<b>Primari</b>	<ul style="list-style-type: none"><li>-Rilevazioni sul campo: conteggio e tipologia utenti, attività svolte</li><li>-Questionari/interviste a residenti e frequentatori sul grado di soddisfazione e percezione di sicurezza</li><li>-Osservazioni dirette condizioni ambientali, stato delle attrezzature</li><li>-Misure di qualità dell'aria e ambientale nel parco</li></ul>	<ul style="list-style-type: none"><li>-Indagine statistica ad hoc</li></ul>

- **Quanto sei soddisfatto/a delle nuove aree attrezzate del parco?**
- **Hai notato miglioramenti nella pulizia e manutenzione rispetto a prima?**
- **Con quale frequenza frequenti il parco?**
- **Ti senti sicuro/a durante la visita al parco?**

Queste domande permettono di raccogliere dati quantitativi e qualitativi utili a valutare la fruizione, la sicurezza e la manutenzione del parco, oltre a raccogliere feedback diretto dai cittadini per miglioramenti futuri.

Per convertire le domande in numeri utili all'elaborazione e alle misure statistiche, si procede come segue:

- **Quanto sei soddisfatto/a delle nuove aree attrezzate del parco?** Assegnare un valore numerico da 1 a 5 corrispondente direttamente alla scala di soddisfazione (1 = per nulla soddisfatto, 5 = molto soddisfatto).
- **Hai notato miglioramenti nella pulizia e manutenzione rispetto a prima?** Codificare Sì = 1, No = 0.
- **Con quale frequenza frequenti il parco?** Codificare Mai = 0, Raramente = 1, Spesso = 2, Molto spesso = 3.
- **Ti senti sicuro/a durante la visita al parco?** Assegnare un valore numerico da 1 a 5 come nella prima domanda.

Ecco una possibile categorizzazione numerica per le risposte aperte del questionario:

### Domanda: Vuoi segnalare problemi o suggerimenti?

1. Problemi di manutenzione
2. Problemi di pulizia
3. Problemi di sicurezza
4. Suggerimenti per nuove attrezzature o servizi
5. Feedback positivo generale
6. Richieste di maggiore frequenza di pulizia/manutenzione
7. Segnalazioni riguardo alla fruibilità (es. accessibilità, orari)
8. Altro / Non classificabile

#### ***L'utilità di codificare le risposte aperte in categorie numeriche***

- *Trasformare input testuali in dati quantitativi da analizzare con frequenze, incroci e tendenze.*
- *Ogni risposta aperta viene letta e assegnata a una di queste categorie, poi si codifica con il numero corrispondente. Se emergono nuovi temi dai rispondenti, si possono aggiungere altre categorie numeriche alla codifica.*

- **Ogni domanda** ha un **codice univoco** per facilitarne il trattamento.
- Le **risposte (mancanti)** non date devono avere un **codice speciale** (ad esempio 99) per distinguerle dalle risposte date e valide.
- Le **risposte aperte** vanno lette attentamente e **raggruppate in categorie significative**, poi codificate numericamente.
- Le **scale da 1 a 5** rappresentano un **ordine naturale**, quindi è corretto calcolare indici numerici sintetici (medie).
- Le **domande Sì/No** sono codificate come **1 e 0**, utili per conteggi e percentuali.

ID Rispondente	Soddisfazione_are (1-5)	Miglioramenti_pulizia (Si=1, No=0)	Frequenza_parco (0-3)	Sicurezza_visita (1-5)	Categoria_risposte_aperte
1	4	1	3	5	5
2	3	1	2	4	1
3	5	0	1	3	2
4	2	1	0	2	4
5	4	0	3	4	3
...	...	...	...	...	...

I codici corrispondono a:

- Soddisfazione\_aree, Sicurezza\_visita: scala da 1 a 5;
- Miglioramenti\_pulizia: Si=1, No=0
- Frequenza\_parco: Mai=0, Raramente=1, Spesso=2, Molto spesso=3;
- Categorie\_risposte\_aperte: 1=Problemi di manutenzione, 2=Problemi di pulizia, 3=Problemi di sicurezza, 4=Suggerimenti, 5=Feedback positivo

### Preparazione dati

- ❖ Importa i dati in un foglio Excel o software statistico come SPSS.
- ❖ Controlla che ogni risposta rispetti la codifica (es. Soddisfazione\_aree da 1 a 5, Miglioramenti\_pulizia solo 0 o 1).
- ❖ Individua valori mancanti codificati come 99 e segnalali come "valore mancante" nel software.
- ❖ Verifica che non ci siano dati fuori scala o errori di inserimento.

Dati campione (10 risposte)

ID	Soddisfazione_aree (1-5)	Miglioramenti_pulizia (1=Si,0=No)	Frequenza_parco (0-3)	Sicurezza_visita (1-5)
1	4	1	3	5
2	3	1	2	4
3	5	0	1	3
4	2	1	0	2
5	4	0	3	4
6	5	1	2	5
7	3	0	1	3
8	4	1	3	4
9	2	0	0	2
10	3	1	2	3

### Codifica domande ed elaborazione dati

Hai notato miglioramenti nella pulizia? Si=1 No=0

**FREQ. ASSOLUTA (ni)**

Si(1):6

No(0):4

**FREQ. RELATIVA E FREQ. %**

Si(1): $6/10=0.6$  (60%)

No(0): $4/10=0.4$  (40%)

Frequenza con cui frequenti il parco (0=Mai, 3 Molto spesso)

**MODA (modalità più frequente):** 3 (3 risposte)

**MEDIA:**  $17/10=1.7$

**MEDIANA:** dati ordinati (0,0,1,1,2,2,2,3,3,3)

La mediana è la media tra il 5° e il 6° valore  $(2+2)/2=2$

### INTERPRETAZIONE

- **La moda** = 3 indica che il **valore più frequente** nella frequenza con cui si frequenta il parco è "3", cioè "molto spesso". Questo significa che la maggior parte delle persone risponde indicando una frequenza alta di visita al parco.
- **La media** = 1.7, invece, è un **valore medio calcolato su tutta la distribuzione** dei dati raccolti, che risulta più vicina a 2, quindi tra "mai" (1) e una frequenza moderata (2). Questo suggerisce che, sebbene il valore più comune sia "3" (molto spesso), ci sono molte risposte con frequenze basse che abbassano la media.
- **La mediana** = 2 è il **valore centrale** dei dati, indicando che la metà delle persone frequenta il parco con una frequenza pari o inferiore a 2 (una frequenza moderata), e l'altra metà con frequenze maggiori o uguali a 2.

*In sintesi, la maggioranza risponde "molto spesso" (moda=3), ma la distribuzione complessiva delle risposte è spostata verso frequenze più basse (media=1.7 e mediana=2), probabilmente a causa di un numero significativo di persone che frequentano poco o mai il parco, evidenziando quindi una variabilità nelle abitudini di visita al parco.*

### IMPORTANZA

- ❖ Agevolano il confronto tra obiettivi pianificati e risultati ottenuti.
- ❖ Supportano proposte di miglioramento e azioni di advocacy.
- ❖ Permettono di costruire report accessibili e visualizzazioni efficaci per i cittadini e le istituzioni.
- ❖ Favoriscono la trasparenza e la partecipazione informata.

**Analisi bivariata:** Tabelle che incrociano due variabili, es. Frequenza\_parco vs Miglioramenti\_pulizia

Frequenza_parco	Miglioramenti_pulizia = 0	Miglioramenti_pulizia = 1	Totale
0	2	0	2
1	1	2	3
2	1	3	4
3	1	0	1
<b>Totale</b>	<b>5</b>	<b>5</b>	<b>10</b>

La tabella che riassume la **distribuzione combinata** di queste due variabili nel campione, utile per analizzare se la frequenza di visita al parco è associata alla percezione di miglioramenti nella pulizia.

Ad esempio, i soggetti che rilevano miglioramenti nella pulizia frequentano il parco soprattutto con frequenze 1 o 2.

### Analisi bivariata (% di riga)

Frequenza_parco	Miglioramenti_pulizia = 0	Miglioramenti_pulizia = 1	Totale	% Miglioramenti_pulizia = 0	% Miglioramenti_pulizia = 1
0	2	0	2	$(2/2) \times 100 = 100\%$	0%
1	1	2	3	$(1/3) \times 100 = 33.3\%$	66.7%
2	1	3	4	$(1/4) \times 100 = 25\%$	75%
3	1	0	1	$(1/1) \times 100 = 100\%$	0%
<b>Totale</b>	<b>5</b>	<b>5</b>	<b>10</b>		

Le freq, % per riga= divido la frequenza assoluta di ciascuna cella per il totale della riga corrispondente.

Esempio, alla frequenza 2 del parco, il 75% ha risposto di sì ai miglioramenti nella pulizia, mentre alla frequenza 0 o 3, la totalità ha risposto no.

### Analisi bivariata (% di colonna)

Frequenza_parco	Miglioramenti_pulizia = 0	Miglioramenti_pulizia = 1	% Miglioramenti_pulizia = 0	% Miglioramenti_pulizia = 1
0	2	0	40%	0%
1	1	2	20%	40%
2	1	3	20%	60%
3	1	0	20%	0%
<b>Totale</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>

Le freq. % per colonna= divido ogni cella per il totale della colonna corrispondente.

Esempio, nel gruppo "Miglioramenti\_pulizia = 0", il 40% ( $=2/5 \times 100$ ) ha frequenza 0, il 20% ( $=1/5 \times 100$ ) frequenza 1, e così via.

Questa lettura evidenzia la distribuzione relativa della frequenza parco in ogni categoria di miglioramenti percepiti, utile per confronti verticali tra gruppi.

### Codifica domande aperte e produzione report

#### Analisi delle risposte aperte

- Raccogli tutte le risposte testuali.
- Leggi e assegna manualmente o con strumenti software category codes (es. 1=Problemi manutenzione).
- Conta quante risposte rientrano in ciascuna categoria.
- Riporta alcuni esempi significativi di commenti per ogni categoria nel report.

#### Sintesi e reportistica

- Organizza i risultati in un documento chiaro, con sezioni per ogni analisi.
- Inserisci grafici e tabelle per facilitare la lettura.
- Evidenzia i punti chiave, come il grado di soddisfazione medio o la percentuale che ha notato miglioramenti.
- Formula raccomandazioni basate su dati oggettivi, ad esempio interventi per migliorare sicurezza o manutenzione.

Elemento di criticità è l'aumento di costi nella produzione di informazione. Vi è il rischio di non produrre statistiche a causa dell'onerosità o della difficoltà connessa alla garanzia di informazioni di qualità.


Non sempre più soldi significano più qualità, ma la qualità ha sempre un valore.

## Vincoli di costo? La qualità si costruisce con scelte intelligenti!

- Personale impiegato nell'indagine (anni/persona)
  - per livello e tipologia di contratto (di ruolo/non di ruolo)
- Raccolta e registrazione
  - stampa modelli
  - spedizione modelli
  - raccolta dati
  - registrazione in service
- Budget annuale (a preventivo)

Per indagini con periodicità inferiore all'anno, costo per replicazione (ottenuto come media sul costo annuale).





**QUALITÀ= OBIETTIVI PROGRAMMATI E RISULTATI RAGGIUNTI**

- Conformità a requisiti dati
- Assenza di difetti
- Corretto funzionamento
- Adeguatezza rispetto agli scopi

Criteri di qualità	Descrizione
<b>Pertinenza</b>	Il grado in cui le statistiche rispondono alle esigenze attuali e potenziali degli utilizzatori.
<b>Accuratezza</b>	Il grado di corrispondenza fra le stime e i valori veri ignoti.
<b>Comparabilità</b>	La misurazione dell'impatto delle differenze tra i concetti statistici applicati, gli strumenti e le procedure di misurazione quando le statistiche si comparano per aree geografiche, ambiti settoriali o periodi di tempo.
<b>Coerenza</b>	L'idoneità dei dati a essere attendibilmente combinati in modi diversi e a vari scopi.
<b>Tempestività</b>	L'intervallo di tempo intercorrente fra il momento della diffusione dell'informazione e l'evento o il fenomeno da essa descritto.
<b>Puntualità</b>	L'intervallo di tempo intercorrente fra la data di diffusione dei dati e il termine previsto per la loro diffusione.
<b>Accessibilità e chiarezza</b>	Le modalità e le condizioni alle quali gli utilizzatori possono acquisire, utilizzare e interpretare i dati.

Fonte: Corte dei conti europea, sulla base dell'articolo 12 del regolamento (CE) n. 223/2009.

## IL SISTEMA DI CONTROLLO DELLA QUALITÀ

<b>Azioni preventive</b>	<i>Prevenire i problemi prima che si presentino</i>	<b>Analisi dei rischi, definizione di standard</b>	<b>Valutazione dei rischi criteri di input</b>
<b>Azioni di monitoraggio</b>	<i>Tenere sotto controllo i processi e i dati</i>	<b>Monitoraggio continuo, uso di metriche</b>	<b>Controllo qualità dati, validazione dati</b>
<b>Azioni di controllo e correzione (CAPA)</b>	<i>Correggere i problemi esistenti e prevenirne il ritorno</i>	<b>Azioni correttive e preventive</b>	<b>Identificazione cause errori, piani CAPA</b>

$$\text{Valore osservato (variabile)} = \text{Valore vero (concetto)} + \text{Errore}$$



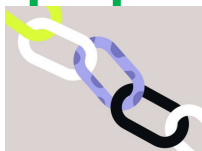
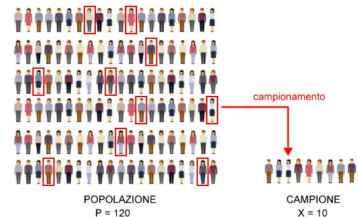
Per creare **VALORE PUBBLICO** e benefici per la collettività è indispensabile garantire la **QUALITÀ** costante dei dati statistici prodotti e diffusi.

## Errore non campionario

- \* errori di lista delle unità da osservare
- \* errori nel questionario
- \* errori dovuti alla tecnica di indagine
- \* errori dei rilevatori
- \* errori dovuti ai rispondenti
- \* mancate risposte (totali e parziali)



## Errore campionario

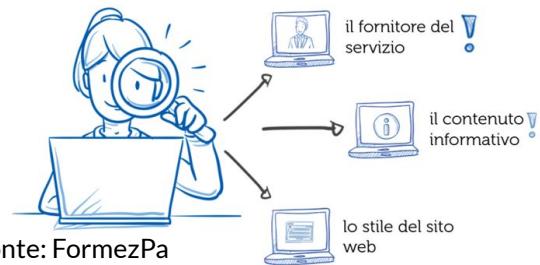


Qual è l'aspetto caratterizzante?

Il rispetto

- di regole metodologiche condivise,
- di principi etico-professionali,
- dell'autonomia scientifica.

- ❑ **WHO=** Chi è il produttore? È riconosciuto ed autorevole a livello pubblico?
- ❑ **WHY=** Per quali obiettivi viene prodotta l'informazione statistica: per servire il paese e i decisori politici o per obiettivi commerciali?
- ❑ **WHAT=** La produzione statistica è riportata in programmi statistici normati? Quali fenomeni sono investigati? Le statistiche sono confrontabili?
- ❑ **HOW=** Si adottano metodologie solide, classificazioni standard, procedure e strumenti standardizzati, sistemi di controllo della qualità?



- ❖ *Evitare di basarsi su dati errati o fuorvianti: comprometterebbero la qualità dell'indagine.*
- ❖ *Se i dati sono inaccurati, incompleti o obsoleti, possono portare a decisioni sbagliate, costose e poco efficaci.*



Home / Comunicato stampa / Tentata truffa alle imprese...

Condividi:    

## Contatti

PER INTERVISTE, DICHIARAZIONI E  
CHIEDIMENTI AI MEDIA  
Ufficio stampa  
Orario 9-19  
tel. 06 4673 2243-4  
Contact centre per i media  
ufficio stampa@istat.it

PER FORNITURA DATI E  
ELABORAZIONI  
Contact Centre  
tel. 06 4673 3102  
contact centre

## Tentata truffa alle imprese con false credenziali Istat

Diverse aziende hanno ricevuto una lettera con intestazione Istat in cui si chiedevano informazioni per uno studio riguardante le relazioni economiche nazionali e internazionali.

La missiva, allegata in copia a tale lettera, è un falso che Istat ha già denunciato alle autorità competenti.

Tutte le rilevazioni in corso presso le imprese o le famiglie seguono una procedura certificata che consente agli interpellati di riconoscere in maniera incontrovertibile l'identità Istat. L'Istituto, nel rendere nota questa truffa in atto, invita coloro che avessero ricevuto la falsa lettera a non rispondere. In caso di dubbio sulla veridicità della comunicazione [contattare il numero unico Istat 1810](#).

L'importanza di riconoscere possibili errori, effettuare controlli accurati e correggere i dati in modo appropriato può garantire la validità e l'affidabilità delle analisi basate sui dati primari.



**evitare errori, controllare con rigore, correggere con metodo.**

### Problemi di riporto

- ❖ I dati possono essere incompleti, errati o falsamente riportati.
- ❖ Cause comuni: errore umano, strumenti difettosi, intenzioni distorte.
- ❖ Impatto: dati sbagliati portano a conclusioni errate.

### Controllo dei dati

- ❖ È importante verificare e validare i dati.
- ❖ Tecniche: Controllo incrociato con altre fonti; Revisione da parte di esperti; Semplici controlli di coerenza (ad esempio, valori impossibili).

### Correzione e imputazione

- ❖ Quando mancano dati o sono sbagliati, si può provare a correggerli.
- ❖ Imputazione = stimare i valori mancanti basandosi su dati esistenti o modelli statistici.
- ❖ Attenzione: imputare bene richiede metodi statistici appropriati per evitare di introdurre bias.

- ❑ Promuovere la partecipazione alla rilevazione delle unità di rilevazione (famiglia, impresa,...)
- ❑ Scelta della **tecnica di indagine** (contatto) più idonea a raccogliere le informazioni. Spesso sono usati due o più modi di raccolta dei dati per: risparmiare, aumentare il tasso di risposta, ridurre gli errori di misura, ect.
- ❑ Formulazione del **questionario e pre - test**
  - Concettualizzazione (relazione e gerarchie), individuazione aree e sotto-aree, contenuti, formulazione e sequenza domande
  - Progettazione dello strumento di raccolta dati: preparare questionari con domande semplici e chiare
- ❑ Rilevazione su campo - Formazione dei **rilevatori**



Questionario a domande chiuse

Reti di rilevazione



**Sistema di controlli dell'errore affiancati alla fase:** invio di lettera di una lettera di preavviso ai rispondenti; istituzione di un numero verde per le richieste di chiarimento; selezione e formazione dei rilevatori; analisi della completezza e della ridondanza delle liste utilizzate; assegnazione delle unità a strati; monitoraggio mancate risposte, indagini ad hoc su non rispondenti, ect.

Fasi	Operazioni	Fonti errore	Tipo errore
Progettazione	scelta unità, variabili	modello concettuale	rilevanza teorica
	redazione questionario	struttura del vocabolario, quesiti, codifica	errori di misura
	diffusione dei dati		rilevanza effettiva trasparenza

- ❑ **Raccogliere**, senza influenzare il rispondente, le informazioni utilizzando le diverse modalità
- ❑ **Registrazione** i dati riportando le risposte dei questionari (strumento di osservazione) su supporto informatico:
  - *lettura ottica dei questionari*
  - *i quesiti aperti vanno ricodificati da linguaggio libero in codici standard (es. ATECO)*
  - *non è necessaria quando la rilevazione è effettuata in CATI, CAPI, CAWI*

Fasi	Operazioni	Fonti errore	Tipo errore
Rilevazione sul campo	formazione elenchi di unità	liste supervisor rilevatori	errori di misura (selez. da mancata copertura)
	raccolta dati	supervisor rilevatori rispondenti	errori di misura (mancate risposte, incongruenze)
	registrazione	operatori	errori di misura

**Sistema di controlli dell'errore affiancati alla fase:** lasciare una buona impressione per facilitare contatti futuri (indagini longitudinali, indagini di controllo), uso di programmi per la registrazione controllata dei dati; applicazione di tecniche di identificazione automatica di incoerenze nei dati (es. un professionista con la sola licenza elementare); ripristino dell'informazione con valori accettabili con una reintervista delle unità statistiche; calcolo di indicatore di qualità (tassi di risposta), ect.

La **revisione** consiste nel:

- valutare la congruità delle informazioni raccolte con le finalità dell'indagine;
- individuare le **fonti di errore** più rilevanti analizzando i dati e correggendo gli errori;
- predisporre modifiche al processo di produzione in modo da ridurre gli effetti degli errori in successive occasioni di indagine;
- validare i dati corretti (se la **qualità dei dati** è buona (es. tempestività ai fini della diffusione) dell'informazione agli utenti.

Fase	Operazioni	Fonti errore	Tipo errore
Revisione	registrazione	operatori	errori di misura
	revisione	revisori programmi automatici	errori di misura (possibile selezione)

**Sistema di controlli dell'errore affiancati alla fase:** uso di procedure manuali o automatiche e/o interattive per individuare violazioni di regole logico formali (regole di compatibilità) relative ai limiti imposti sul campo di variazione delle singole variabili, alle relazioni intercorrenti fra le variabili e alle relazioni formali stabilite dalle norme di compilazione dei questionari, ect.

Nella fase di **elaborazione** si applicano strumenti propri dell'analisi statistica (metodi inferenziali-induttivi) al fine di ottenere una sintesi e una descrizione dei dati sotto forma di:

- tabelle
- grafici
- calcolo di indici e rapporti sintetici

La scelta del metodo da usare dipende dal tipo di fenomeno osservato (quantitativo o qualitativo).

<b>Fase</b>	<b>Operazioni</b>	<b>Fonti errore</b>	<b>Tipo errore</b>
Elaborazione dati	elaborazione	programmi base statistica	errori di calcolo rilevanza effettiva
	diffusione		tempestività

**Sistema di controlli dell'errore affiancati alla fase:** analisi descrittive e esplorative sui dati per individuare incongruenze (valori anomali ed errori); ricerca di relazioni; analisi di tipo confermativo o verifica di ipotesi; calcolo di indicatori di qualità (varianza), ect.

### Diffondere dati affidabili per decisioni migliori




La diffusione è la fase conclusiva della rilevazione, grazie alla quale si rendono disponibili ([accessibili e comprensibili](#)) i dati raccolti ([le statistiche e l'analisi statistica](#)).

È l'occasione per far percepire all' «[esterno](#)» la specificità e l'utilità del lavoro svolto (promozione).

### Aspetti da valutare su contenuti e strumenti della diffusione

- adeguata offerta di informazioni di qualità
- facilità di accesso ai dati e ai metadati
- tempo ridotto tra raccolta, elaborazione e diffusione
- forme efficienti di diffusione e analisi del grado di utilizzo delle statistiche
  - caratteristiche degli utenti
  - canali utilizzati

La qualità dei dati è fondamentale non solo per l'analisi interna dei dati, ma anche per garantire che le informazioni diffuse siano utili, credibili e contribuiscano a decisioni consapevoli.

- ❑ L'informare come «dare forma ai dati e dare senso alla realtà» richiede quantomeno informazioni **CORRETTE**.  
 *Esempio: Monitoraggio dell'inquinamento ambientale con sensori precisissimi per riflettere la realtà e raccolta regolare dei dati.*
  
- ❑ I dati statistici - **ASSET STRATEGICO** per conoscere la realtà che si traducono in **INFORMAZIONE STATISTICA** quanto più questa si diffonde e viene utilizzata.  
 *Esempio: I dati ISTAT sulla disoccupazione giovanile in Italia diffusi e utilizzati da ministeri, enti di formazione, imprese e associazioni.*
  
- ❑ Un mondo di dati e di produttori di **STATISTICHE NON TUTTE UGUALI** per esigenze conoscitive sempre più diverse.  
 *Esempio: Nel settore sanitario, diversi enti producono statistiche sui dati di pazienti e malattie: ogni produttore adatta le statistiche alle proprie esigenze specifiche.*

- **AUTOREVOLEZZA DELLA FONTE** (credenziali e reputazione di imparzialità, indipendenza scientifica, impianto organizzativo pubblico)
- **QUALITÀ DEL DATO PRODOTTO** (affidabilità, pertinenza, trasparenza, riservatezza, tempestività etc.)

- I dati sono numeri o fatti raccolti
- Da soli non dicono nulla senza interpretazione
- La statistica organizza e analizza i dati per fornire conoscenza
- La conoscenza consente di capire il problema e prendere decisioni informate
- Cos'è la statistica: (definizione semplice e chiara) raccolta, analisi e interpretazione di dati riguardanti fenomeni reali.
- Perché è importante: aiuta a comprendere fenomeni collettivi, prendere decisioni consapevoli e prevenire disinformazione.

DATO



INFORMAZIONE



CONOSCENZA



### OGNI DATO CONTA, OGNI ERRORE PESA!

- Una buona progettazione è la base per dati di qualità e risultati affidabili
- Valutare sempre la qualità e la pertinenza dei dati disponibili (qualità “specifica per fonti”)
- Sfide: Unendo dati, potenziamo le decisioni pubbliche.

## BIBLIOGRAFIA ESSENZIALE

AA.VV., Manuale di tecniche di indagine, Vol. 1-6, ISTAT (1989)

[https://lipari.istat.it/digibib/Metodi%20e%20norme/Manuale e tecniche di indagine/](https://lipari.istat.it/digibib/Metodi%20e%20norme/Manuale_e_tecniche_di_indagine/)

H.M. Blalock H.M. (1969), Statistica per la ricerca sociale, Il Mulino (riedizioni) [ed. orig.: Social Statistics, McGraw Hill, New York, 1960].

G.Cicchitelli, P.D'Urso, M.Minozzo, Statistica. Principi e Metodi, Mylab (2022).

P.Corbetta, Metodologia e tecniche della ricerca sociale, Il Mulino (2014).

ISTAT (2020-2025). Linee guida e manuali metodologici sulle indagini statistiche. Roma.

G.Leti, Statistica Descrittiva, Il Mulino. (1983 e riedizioni)

D.Piccolo, Statistica, Il Mulino. (2010 e riedizioni).

L.Ricolfi, Manuale di analisi dei dati. Fondamenti, Laterza (2024).

Progetto finanziato con il sostegno di



Cofinanziato  
dall'Unione europea



Presidenza del Consiglio dei Ministri  
Dipartimento per le politiche di coesione e per il sud

In collaborazione con



**MIM**  
Ministero dell'Istruzione  
e del Merito



Con il supporto di

**EUTALIA**  
studiare sviluppo



COESIONE  
ITALIA



*A Scuola di*  
**OPENCOESIONE**

Rita Lima  
[lima@istat.it](mailto:lima@istat.it)  
[www.istat.it](http://www.istat.it)

*Grazie*



Istat

