

PIANO NAZIONALE  
LAUREE SCIENTIFICHE

2016-2017



# Indagini campionarie

Prof. Francesco D. d'Ovidio



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

# Statistica

*Sai ched'è la statistica? E' 'na cosa che serve pe' fa' un conto in generale de la gente che nasce, che sta male, che more, che va in carcere e che sposa.*

*Ma pe' me la statistica curiosa è dove c'entra la percentuale, pe' via che, lì, la media è sempre eguale puro co' la persona bisognosa.*

*Me spiego: da li conti che se fanno secondo le statistiche d'adesso risurta che te tocca un pollo all'anno:*

*e, se nun entra ne le spese tue, t'entra ne la statistica lo stesso perché c'è un antro che ne magna due.*

*(Trilussa)*

# Approfondimento

**La statistica** non si occupa tanto (o soltanto) della media, quanto anche e soprattutto della variabilità, e tra le sue conquiste più importanti c'è quella di saper sintetizzare in pochi parametri e indicatori i fenomeni collettivi.

La risposta alla famosa storiella del pollo di Trilussa è che la statistica non vuole parlare soltanto della **media** (un pollo a testa) ma anche della **variabilità** (due polli a uno e nessuno all'altro).

Ad esempio, la statistica afferma che l'Italia oggi è mediamente più ricca che nel 2000 (maggiore PIL). Ma la gente comune non lo è, e ci sono molte aziende **in meno** (fallite) e qualcuna **molto** più solida.

## ***Seriamente: che cos'è la statistica?***

- **Scienza che ha per oggetto lo studio dei fenomeni collettivi suscettibili di misurazione e di descrizione quantitativa (specialmente quando il numero di unità interessate è tanto elevato da escludere la possibilità o la convenienza di seguire le vicende di ogni singola unità).**
- **Perviene alla formulazione di leggi fondamentali che governano tali fenomeni, dette leggi statistiche**
  - **basandosi sulla raccolta di un grande numero di dati inerenti ai fenomeni in esame, e partendo da ipotesi più o meno direttamente suggerite dall'esperienza o da analogie con altri fenomeni già noti**
  - **mediante l'applicazione di metodi matematici fondati sul calcolo delle probabilità**

La parola "statistica" ha due significati :

- analisi quantitativa dei fenomeni collettivi che hanno attitudine a variare, allo scopo di descriverli e di individuare le leggi o i modelli che permettono di spiegarli e di prevederli;
- raccolta sistematica e ordinata di dati: *la statistica dell'incremento delle nascite, la statistica dell'andamento dei prezzi, la statistica dei ricoveri ospedalieri, ecc.*

Di solito, ci si riferisce alla **"statistica"** nel primo caso e alle **"statistiche"** nel secondo.

- La **Statistica**, concepita inizialmente come *attività descrittiva* di fatti sociali (e in particolare come attività amministrativa dello Stato), ha via via ampliato i suoi confini, fino a diventare una vera e propria «**scienza del collettivo**», una disciplina con finalità non solo descrittive dei fenomeni, ma orientata anche a finalità di ricerca e di estrapolazione (o anche interpolazione).
- La sua base concettuale di partenza è generalmente una *indagine statistica*, che può essere di vari tipi.
- L'indagine statistica consente di raccogliere dati da cui trarre **statistiche** di vario genere: distribuzioni, medie, indici di variabilità ecc.

# Fasi dell'indagine statistica

1. Programmazione
2. Rilevazione
3. Elaborazione
4. Presentazione
5. Interpretazione



# 1. *La fase di programmazione*

L'attività di progettazione deve prendere in considerazione tutti gli aspetti coinvolti, da quelli riguardanti i fenomeni di interesse e quelli di carattere più operativo.

La progettazione e l'esecuzione di un'indagine sono generalmente il risultato dell'impegno multidisciplinare che coinvolge, necessariamente, diverse professionalità ed ha lo scopo di individuare una precisa *domanda di informazione*, domanda che l'indagine è chiamata e soddisfare.

La fase di programmazione riguarda la definizione del problema da analizzare in tutti i suoi aspetti (scientifici, statistici, organizzativi, economici).



I principali argomenti da prendere in considerazione sono:

- definizione del fenomeno in studio,
- definizione degli obiettivi,
- individuazione del collettivo e delle sue caratteristiche,
- ricerca e analisi delle fonti statistiche,
- indicazione dei mezzi tecnici da utilizzare (per raccogliere le informazioni sulle caratteristiche da studiare),
- fissaggio dell'estensione temporale e territoriale,
- individuazione delle fasi operative con relativa determinazione di tempi e costi,
- programmazione di un sistema di controllo della qualità,
- piano delle elaborazioni statistiche,
- piano di diffusione dei risultati.

## *Definizione degli obiettivi e concettualizzazione*

- Definire esattamente quali sono i temi che interessano l'indagine, escludendo quelli che non sono di 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 .

## Definizione degli obiettivi dell'indagine

In questa fase è necessario esplicitare gli obiettivi fissati dai committenti e specificare le definizioni e classificazioni delle unità di rilevazione e delle variabili da rilevare.

E' indispensabile coinvolgere nella progettazione uno o più esperti del settore specifico oggetto di indagine.

Vi sono alcuni aspetti definatori da considerare.

*Fenomeno di interesse:*

- delimitare precisamente cosa interessa e cosa non interessa.
- 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.

### *Classificazioni:*

- Insieme delle categorie assunte da una variabile qualitativa sconnessa o ordinale.
- E' opportuno ricorrere a classificazioni comunemente utilizzate.
- Per alcune variabili particolarmente complesse da definire (attività economiche, professioni, malattie) sono disponibili classificazioni standard riconosciute a livello internazionale.

## **Analisi delle fonti**

I canali informativi dai quali trarre informazioni si possono suddividere in quattro fonti:

- a) le esperienze di ricerca maturate sia in campo nazionale sia in quello internazionale;
- b) la documentazione bibliografica;
- c) la documentazione statistica (ad esempio documenti ISTAT o della rete del Sistan);
- d) i sondaggi presso utenti e/o operatori del settore (interviste agli opinion leader);
- e) indagini similari precedenti.

## 2. La rilevazione

Nella determinazione del piano di rilevazione si stabilisce come selezionare la popolazione di riferimento:

- **RILEVAZIONE TOTALE**: è estesa a tutte le unità dell'universo o popolazione statistica. Esempio: censimenti, rilevazione dei nati, dei decessi, dei ricoveri, delle IVG
- **RILEVAZIONE PARZIALE**: l'osservazione riguarda una parte delle unità dell'universo (detta "dominio"), definita a priori dal ricercatore, ma rileva **tutte** le unità di tale gruppo. Esempio: Rilevazione sulle grandi imprese, censimento intermedio dell'industria e dei servizi
- **RILEVAZIONE CAMPIONARIA**: l'osservazione riguarda una parte delle unità dell'universo scelte in modo casuale. Esempio: rilevazione sulle forze di lavoro, sui consumi ecc...

## Disegno di indagine

### *Indagine totale o indagine campionaria?*

Ricorrere ad un'indagine di tipo non opportuno può pregiudicare in tutto o in parte gli scopi della ricerca.

Raccogliere informazioni su *tutte* le unità statistiche (indagine censuaria) implica un aumento dei costi, ma il ricorso ad una indagine campionaria espone ad un più elevato rischio di errore.

Per una indagine campionaria occorre valutare i seguenti aspetti:

- a) determinare il metodo di selezione del campione in riferimento alla struttura degli archivi di base e alle informazioni in essi contenute, in modo da massimizzare l'efficienza delle stime prodotte, tenendo conto allo stesso tempo dei vincoli imposti;
- b) dimensionare il campione in modo da garantire stime della precisione desiderata, dati i vincoli di bilancio imposti.



## *Il grande dilemma*

*Nella rilevazione totale*  
si ha la *conoscenza*  
esatta del fenomeno  
studiato.

*Nella rilevazione campionaria*  
si perviene a una stima  
(più o meno precisa)  
del fenomeno studiato.

Rilevazioni  
totali



Rilevazioni parziali/  
campionarie

*Fattori che determinano la scelta*

- I **tempi** della rilevazione;
- I **costi** della rilevazione;
- La **ricchezza di dettagli** della rilevazione;
- Gli **errori** associati alla rilevazione.

## *Quindi...*

**Le informazioni relative alla popolazione o collettivo statistico, cioè alle variabili che la caratterizzano, possono derivare da una:**

### **Rilevazione censuaria o totale**

**Si ha la conoscenza esatta del fenomeno analizzato.**

**Si preferisce:**

**... per analisi a livello di micro - aree;**

**... quando le unità da analizzare sono rare;**

**... quando si vuole portare l'analisi ad un elevato livello di dettaglio.**

# *Invece...*

## Rilevazione campionaria / parziale

Si perviene ad una *stima* del fenomeno.

Si preferisce:

- ... quando è impossibile effettuare una rilevazione totale;
- ... quando la rilevazione del carattere comporta la distruzione delle unità osservate;
- ... quando si vogliono ridurre i costi e/o i tempi di un'indagine.

# Caratteristiche dell'indagine campionaria

La **ricchezza di dettagli** della rilevazione

- Le indagini campionarie si distinguono dalle indagini esaustive per la possibilità di andare in profondità nella ricerca dell'informazione.

La **rapidità:**

- nel raccogliere e trattare i dati;
- nel pubblicare i risultati delle analisi.
- Analisi di eventi stagionali o periodici che richiedano interventi immediati. a (Attività produttive, Occupazione, Malattie diffuse, Migrazioni, ...).

- ✓ La **precisione** di una stima è sempre direttamente proporzionale alla dimensione del campione.
- ✓ L'**accuratezza** è invece legata al passaggio dei dati su un supporto adeguato per l'elaborazione automatica. Gli errori di rilevazione e di trattamento dei dati sono un *rischio maggiore nelle indagini di vaste dimensioni*. Il problema non va sottovalutato: l'*inaccuratezza* a volte supera l'*imprecisione* dovuta al campionamento.
- ✓ Il concetto che riassume in sé sia la precisione che l'accuratezza è l'**attendibilità** di un'indagine.

***Precisione, Accuratezza e Attendibilità*** sono assolute nelle indagini esaustive e decrescono in funzione della numerosità del campione.

# La fase di rilevazione

- Raccolta dei dati:
  - **RILEVAZIONI DIRETTE:** sono eseguite dall'Istat o da altri Enti del SISTAN mediante un'intervista diretta, mirata a raccogliere le informazioni, rivolta alle famiglie, alle imprese, alle istituzioni
  - **RILEVAZIONI INDIRETTE:** le informazioni sulle unità statistiche sono desunte da documenti e/o registri compilati a fini amministrativi, oppure mediante osservazione del fenomeno senza contatto diretto con eventuali attori dello stesso.

## *Modalità di rilevazione diretta*

- **Intervista diretta “faccia a faccia”**  
Libera (senza questionario): “*biografica*” oppure “*focalizzata*”  
Con questionario cartaceo (*PAPI*=Paper Aided Personal Interview)  
Con Tablet o PC (*CAPI*=Computer Aided Personal Interview)
- **Intervista postale (con questionario cartaceo)**
- **Intervista con diario**
- **Intervista telefonica**  
Con questionario cartaceo (*PATI*=Paper Aided Telephone Interview)  
Con Tablet o PC (*CATI*=Computer Aided Telephone Interview)
- **Indagine web (con questionario e-mail o CAWI)**  
Con questionario e-mail (*MAWI*=Mail Aided Web Interview)  
Con Tablet o PC (*CAWI*=Computer Aided Web Interview)



# Intervista diretta

## VANTAGGI

- Certezza dell'identità del rispondente
- Possibilità di fornire chiarimenti
- Possibilità di usare tecniche di *probing*
- Possibilità di stimolare e motivare gli intervistati
- Controlli strutturati di coerenza delle risposte

## SVANTAGGI

- Organizzazione complessa
- Costi elevati
- Effetto rilevatore
- Diversa formulazione delle domande ⇒ ridotta omogeneità
- Modifiche nell'ordine delle domande
- Eliminazione di domande
- Reticenza

# *Indagine postale con questionario autocompilato*

## **VANTAGGI**

- Costi minori dell'intervista diretta
- Organizzazione meno complessa
- Nessun effetto intervistatore
- Meno reticenza
- Flessibilità nei tempi di compilazione del questionario

## **SVANTAGGI**

- Necessità di disporre degli indirizzi
- Mancate risposte totali e parziali
- Incertezza sull'identità dei rispondenti
- Errori dei rispondenti (risposte non pertinenti)
- Autoselezione dei rispondenti

# *Diario*

## *con questionario autocompilato*

### **VANTAGGI**

- Contatto con l'intervistatore solo iniziale
- Dati più accurati
- Niente "effetto memoria"

### **SVANTAGGI**

- Contatto iniziale con l'intervistatore (se poco esperto)
- Richiede un buon grado di collaborazione da parte degli intervistati
- Uso limitato a brevi periodi d'indagine
- Distorsione del campione (elevata autoselezione dei rispondenti)

# Intervista telefonica

## VANTAGGI

- Tempestività
- Costi ridotti
- Controlli strutturati di coerenza delle risposte
- Meno diffidenza, più disponibilità
- Minor effetto intervistatore
- Se *CATI*, struttura guidata dell'intervista e controlli automatici

## SVANTAGGI

- Tasso di copertura dipendente da fattori esterni
- Incertezza sull'identità del rispondente
- Impossibilità ad usare tecniche di *probing*
- Se *CATI*, costo iniziale del sistema di rilevazione

***CATI* = Computer Aided Telephone Interview**

# Intervista web

## VANTAGGI

- Tempestività
- Velocità di rilevazione
- Costi ridottissimi
- Meno diffidenza, più disponibilità
- Nessun effetto intervistatore
- Se *CAWI*, struttura guidata dell'intervista e controlli automatici

## **SVANTAGGI**

- Tasso di copertura dipendente da fattori esterni
- Mancate risposte totali e parziali
- Incertezza sull'identità dei rispondenti
- Errori dei rispondenti
- Autoselezione dei rispondenti
- Necessità di una certa competenza informatica

***CAWI* = Computer Aided Web Interview**

## *Modalità di rilevazione indiretta*

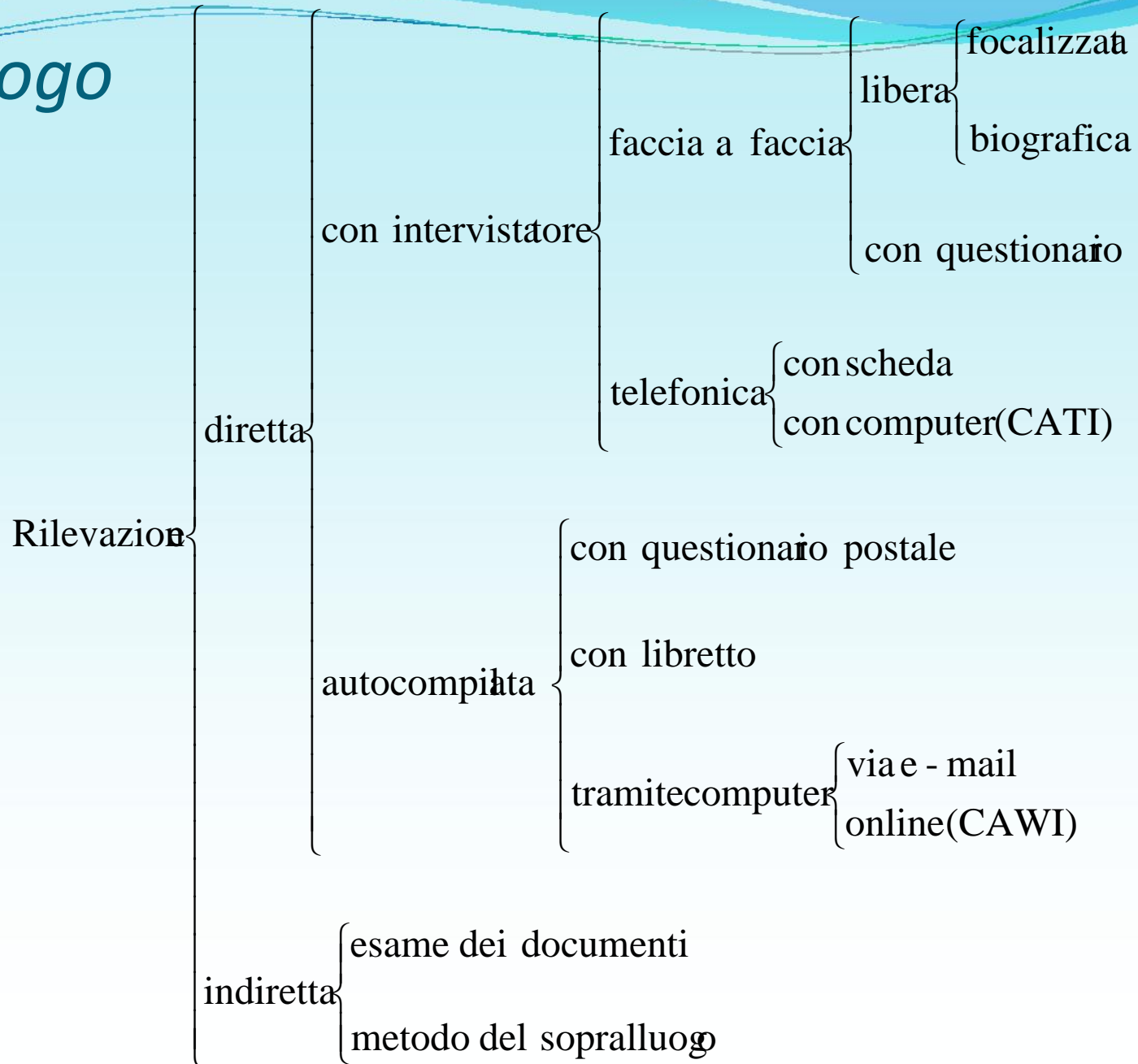
- Indagine “con sopralluogo”

Osservazione del comportamento delle unità statistiche, **con partecipazione** (quando i soggetti vengono stimolati esplicitamente o implicitamente dall'osservatore) o **senza partecipazione** (quando l'osservatore resta in incognito e si mantiene distaccato dai soggetti osservati)

- Indagine documentale

Disamina di documentazione (cartacea o informatica) in cui siano contenute le informazioni ricercate: tipica è l'analisi di registri storici (ad es., i registri delle nascite e delle morti tenute dalle parrocchie nei secoli scorsi).

# Riepilogo





# Ancora sulle rilevazioni

- Le rilevazioni possono essere:
  - ✓ Occasionali
  - ✓ Periodiche
  - ✓ Continue
- I fenomeni che si concretizzano solo in riferimento a dati istanti di tempo si dicono *fenomeni di stato*
- I fenomeni soggetti ad un rinnovamento continuo la cui misurazione non può aversi all'istante, si dicono *fenomeni di flusso* o *di movimento*

### 3. *La fase dell'Elaborazione*

- I dati vengono presentati e sintetizzati attraverso *tabelle di spoglio*.
- Questa operazione consiste nel raggruppare ed elaborare le unità statistiche omogenee secondo le modalità di uno o più caratteri .
- Se il raggruppamento viene effettuato in classi, bisogna fare attenzione che le classi siano disgiunte, cioè che si escludano a vicenda (l'estremo superiore dell'una *non può* coincidere con l'estremo inferiore di quella successiva, altrimenti non si saprebbe a quale delle due classi assegnare una unità che presenti tale valore.)

# Lo spoglio dei dati

## SPOGLIO MANUALE:

Se i dati sono pochi oppure non si sa (o non si vuole) usare un computer si può procedere come segue:

- a) ORDINAMENTO: i dati vengono disposti in ordine crescente (di **grandezza**, se quantitativi; di **rango**, se ordinali; **concettuale** o anche **alfabetico** se nominali);
- b) TABELLE DI SPOGLIO: tabelle inizialmente vuote (anche a doppia entrata), le cui celle identificano le categorie di risposta delle variabili di interesse; man mano che viene esaminato un dato, viene inserito un segno di spunta nella cella corrispondente alle sue modalità.
- 3) TABELLE DI CONTEGGIO: nelle cui celle vengono posti i conteggi dei segni spuntati nelle tabelle di spoglio.

## **SPOGLIO AUTOMATIZZATO:**

Lo spoglio dei dati avviene invece con il computer se essi sono molto numerosi oppure se l'utilizzo di tale strumento risulta agevole; sono necessarie le seguenti attività:

- 1) CODIFICA: definizione di una corrispondenza biunivoca tra cifre alfanumeriche e/o le denominazioni, con un insieme di codici che ne faciliti l'inserimento (o "inputazione"), riduca lo spazio occupato e ne faciliti controlli ed elaborazioni;
- 2) INSERIMENTO: fase di trasferimento dei dati (monotona ma delicata) dal supporto su cui sono già registrati ai programmi di elaborazione;
- 3) TABULAZIONE: l'elaborazione, in questo caso effettuata direttamente dal computer, di tabelle di conteggio che evidenzino la distribuzione dei dati nelle diverse categorie.

## *4. La fase di presentazione*

- Consiste nell'esposizione dei dati statistici in forma compatta (distribuzioni statistiche, medie, variabilità, grafici, ecc.). Prevede anche, nel caso di dati codificati, la **decodifica** o **descrizione in chiaro** delle modalità (se non già operata nella fase di tabulazione automatica).

## *5. La fase di interpretazione*

- Nella fase di interpretazione vengono spiegati i risultati dell'indagine alla luce delle teorie e delle precedenti conoscenze del fenomeno studiato.