

La valutazione e monitoraggio dei programmi  
finanziati dai fondi strutturali dell'Unione  
Europea

**Metodologia e tecnica dell'indagine  
campionaria**

**Giuseppe Storti**

*Dipartimento di Scienze Economiche e Statistiche*

*Università di Salerno*

**Email: [storti@unisa.it](mailto:storti@unisa.it)**

Roma, INEA, 17 gennaio 2003

## Schema della presentazione

- Le indagini campionarie
- Il piano dell'indagine e sue fasi
- Fonti di errore nelle indagini campionarie
- Casi studio

## Le indagini statistiche

➤ Oggetto di ogni indagine statistica è la conoscenza di una *popolazione* intesa come insieme di *unità elementari* in cui il fenomeno allo studio si manifesta (es. indagine sulla spesa delle famiglie)

➤ Le informazioni sul comportamento della popolazione relativamente alle variabili di interesse possono essere desunte da:

- ***rilevazioni totali o censuarie***: si prendono in considerazione tutte le unità della popolazione

- ***rilevazioni campionarie***: si limita l'analisi ad una parte delle unità, ad un *campione*, avendo come obiettivo lo studio dell'intera popolazione

## **Indagini campionarie e censimenti: un'analisi comparata I**

➤ Ogni indagine campionaria produce un'informazione affetta da errore (*errore campionario*) ma presenta, rispetto alle indagini totalitarie, alcuni vantaggi:

- il campione consente di ottenere informazioni in tempi più brevi
- l'indagine campionaria ha un costo ridotto
- in alcuni casi un'indagine totalitaria non è possibile (ad es. se si vuole misurare la durata media delle lampadine, la rilevazione comporta la distruzione dell'unità osservata)
- l'indagine campionaria consente un più accurato controllo del processo di acquisizione dell'informazione

# Indagini campionarie e censimenti: un'analisi comparata

## II

➤ Indagini campionarie su larga scala e censimenti spesso non si escludono ma si complementano:

- nella progettazione delle prime si ricorre spesso ad informazioni censuarie (informazione *a priori*)
- i censimenti sono spesso seguiti da indagini campionarie finalizzate all'accertamento della qualità dei dati.

(Ritorna a [sommario](#))

## Il piano dell'indagine

Ogni indagine campionaria richiede la redazione di un piano di lavoro che definisca gli aspetti fondamentali e le fasi dell'indagine:

- formulazione degli [obiettivi dell'indagine](#)
- [periodo di svolgimento e periodo di riferimento](#)
- determinazione della [lista di campionamento](#)
- Progettazione del campione e definizione del [piano \(o disegno\) di campionamento](#)
- Scelta dei parametri di interesse
- Predisposizione del [questionario](#) (molto importante!)
- metodo di [raccolta dei dati](#)
- lavoro sul campo ed addestramento dei rilevatori
- elaborazione ed analisi dei dati
- preparazione della [relazione finale](#)

(Ritorna a [sommario](#))

## **Nella definizione degli obiettivi dell'indagine rientrano:**

➤ L'indicazione del grado di [precisione](#) voluto per i risultati dell'indagine. Alcune considerazioni:

- esiste generalmente un *trade-off* fra costo finanziario dell'indagine e precisione
- il problema, a seconda delle esigenze del caso, si può affrontare in due modi diversi:
  - i) si fissa il livello di precisione desiderato come obiettivo prioritario minimizzando il costo in corrispondenza del valore prescelto
  - ii) si fissa il costo come obiettivo prioritario massimizzando la precisione in corrispondenza del valore prescelto

➤ Definizione della popolazione [obiettivo](#)

➤ Definizione dell'[unità statistica](#) e dell'[unità di rilevazione](#)

(Ritorna al [piano dell'indagine](#))

## Cosa si intende per precisione?

La precisione viene generalmente valutata in termini della distribuzione campionaria degli <b>stimatori</b>
---

Uno **stimatore** è una funzione delle osservazioni campionarie il cui fine è quello di produrre una **stima** ovvero una valutazione approssimata di un parametro caratteristico di una data popolazione

La **distribuzione campionaria** dello stimatore altro non è che la distribuzione delle diverse stime ottenibili al variare del campione estratto.



## La distribuzione campionaria

➤ Per n (=dimensione del campione) prefissato, le caratteristiche che ci interessa valutare sono:

la **distorsione** ovvero la differenza  $\mathbf{d}_n$  fra la media della distribuzione ed il valore incognito del parametro della popolazione.

la **variabilità** della distribuzione ovvero la dispersione delle stime intorno alla loro media, misurata dalla varianza  $\mathbf{s}_n^2$ .

➤ Un criterio sintetico di valutazione consiste nel calcolare l'**errore quadratico medio (MSE)**:

$$\mathbf{MSE} = \mathbf{d}_n^2 + \mathbf{s}_n^2$$

il cui valore è legato al concetto di **efficienza** dello stimatore

## Il concetto di consistenza

Al crescere di  $n$ , riveste invece una notevole importanza la verifica della proprietà di **consistenza** la quale risulta verificata quando il MSE tende ad annullarsi al crescere di  $n$ :

$$\lim_{n \rightarrow \infty} MSE_n = 0$$

### Nota bene

Tale proprietà ci dice quanto efficacemente lo stimatore utilizza l'informazione addizionale ottenuta a mezzo di un allargamento del campione.

(Ritorna a [obiettivi dell'indagine](#))

**Popolazione obiettivo o universo:** insieme delle unità elementari, o *unità statistiche*, nelle quali si manifesta il fenomeno oggetto di analisi. Può essere di dimensione sia finita che infinita. Nelle indagini socio-economiche ci si trova generalmente nel primo caso.

(Ritorna a [obiettivi dell'indagine](#))

**Unità statistica:** insieme di unità elementari il cui aggregato costituisce l'universo ovvero la popolazione obiettivo.

**Unità di rilevazione:** unità di riferimento nella fase di acquisizione delle informazioni. Può coincidere con l'unità statistica o essere differente.

**Esempio:** nel caso dell'indagine ISTAT sulle forze di lavoro (complesso delle persone fra i 14 e i 70 anni occupate, disoccupate o in cerca di prima occupazione) l'unità statistica è il singolo individuo appartenente a detto aggregato mentre l'unità di rilevazione è la famiglia.

(Ritorna a [obiettivi dell'indagine](#))

**Periodo di svolgimento:** periodo in cui la raccolta dei dati andrà espletata

**Periodo di riferimento:** lasso di tempo a cui vanno riferite le informazioni concernenti le unità indagate.

**Esempio:** se si opera con fenomeni caratterizzati da forte stagionalità sarà generalmente opportuno assumere come periodo di riferimento l'anno ed effettuare l'indagine con periodicità infrannuale (ad es. mensile) come nel caso dell'indagine ISTAT sui consumi in cui la variabile rilevata è la spesa media mensile per consumi delle famiglie italiane con riferimento ad un determinato anno.

(Ritorna al [piano dell'indagine](#))

**Lista di campionamento (o *base campionaria*):** elenco dettagliato degli elementi (**unità di campionamento**) che costituiscono la popolazione e rappresenta la base per la scelta delle unità da inserire nel campione.

L'insieme degli elementi inclusi nella lista costituisce la **popolazione osservata** che è generalmente diversa dalla **popolazione obiettivo**.

**Alcuni esempi:** elenco degli abbonati al telefono, degli iscritti all'università, liste elettorali, elenco delle famiglie disponibile presso l'ufficio di stato civile dei comuni.

**Nota bene:**

La compilazione della lista avviene sulla base della popolazione osservata, diversa dalla popolazione obiettivo, la quale include le unità fra le quali viene effettivamente scelto il campione

**Esempio:** in un'indagine telefonica sulle preferenze dei consumatori di un dato prodotto:

- la *popolazione obiettivo* è data da tutti i potenziali consumatori
- la *popolazione osservata* comprende solo i consumatori possessori di un telefono.

(Ritorna al [piano dell'indagine](#))

## Definizione del piano (o disegno) di campionamento

È la fase in cui si determina il procedimento con cui viene formato il campione. I disegni di campionamento possono essere:

➤ casuali o non casuali, a seconda che l'inclusione delle singole unità del campione sia affidata al giudizio di un esperto o del rilevatore o ad una regola probabilistica.

➤ **semplici o stratificati.** La *stratificazione* comporta la suddivisione della popolazione in più sub-popolazioni omogenee all'interno delle quali viene selezionato, in modo indipendente, un sub-campione.

➤ **ad uno o a più stadi** a seconda che le unità finali oggetto di rilevazione siano selezionate direttamente o per tappe successive (vd. esempio).

(Ritorna al [piano dell'indagine](#))

**Esempio illustrativo.** Nelle indagini ISTAT sulla spesa delle famiglie e sulle forze di lavoro si segue una procedura a due stadi con stratificazione al primo stadio.

Le unità di primo stadio sono i comuni i quali vengono stratificati in base all'attività economica prevalente

Le unità di secondo stadio sono le famiglie residenti nei comuni selezionati al primo stadio.

(Ritorna al [piano dell'indagine](#))



## Campionamento casuale o probabilistico

Nei disegni di campionamento di tipo casuale o probabilistico le unità della popolazione hanno prefissate probabilità di essere incluse nel campione determinate dalla regola di selezione scelta.

- [campionamento casuale semplice](#) con reimmissione (bernoulliano) o senza reimmissione (in blocco)
- [campionamento stratificato](#) (proporzionale, uniforme, ottimale)
- [campionamento a grappoli](#) (*cluster sampling*): a uno (*single-stage*) o a più stadi (*multi-stage*)
- [campionamento sistematico](#)
- [campionamento ripetuto \(panel e panel ruotati\)](#)

(Ritorna a [Piano di Campionamento](#))

## **Campionamento Casuale Semplice (CCS)**

Si procede ad effettuare una estrazione (con o senza reimmissione) delle unità del campione dalla popolazione garantendo a ciascuna unità del collettivo la stessa probabilità di entrare a far parte del campione

### **Nota**

Viene spesso utilizzato come building block per la definizione di schemi di campionamento più complessi

➤ **Estrazione Bernoulliana o con reimmissione, ciascuna unità può figurare anche più di una volta nel campione.** Il numero di realizzazioni campionarie diverse fra loro per almeno un elemento che è possibile estrarre da una popolazione di  $N$  unità, è pari al numero delle disposizioni complete di  $N$  elementi in classe  $n$ :

$$D_{n,h}^c = N^n.$$

e la relativa probabilità di selezione è  $1/N^n$ .

**Estrazione in blocco.** ciascuna unità può figurare nella realizzazione estratta per una sola volta, il numero di possibili n-uple campionarie pari al numero delle combinazioni semplici di N elementi in classe n:

$$C_{N,n} = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

con probabilità di selezione  $\binom{N}{n}^{-1}$ .

Noi ci concentreremo sul secondo caso

**CCS: stima dei parametri della popolazione con riferimento ad un carattere X**

$$\bar{x} = \frac{\sum_{i=1}^N X_i}{N}$$

Media ( $\mathbf{m}_x$ )

$$\hat{T} = N\bar{x}$$

Totale ( $T$ )

$$s_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{x})^2}{n-1}$$

Varianza ( $\mathbf{s}_x^2$ )

$$p_x = \frac{n_x}{n}$$

Proporzioni ( $P_x$ )

## CCS: Stima della varianza campionaria delle stime – I

### Media

#### Estrazione in blocco

$$\text{var}(\bar{x}) = \frac{s_x^2}{n} \frac{N-n}{N} \qquad \hat{\text{var}}(\bar{x}) = \frac{s_x^2}{n} \frac{N-n}{N}$$

$(N-n)/N =$  *fattore di esaustività* o *fpc* (*finite population correction*)  $< 1$ .

#### Estrazione Bernoulliana

$$\text{var}(\bar{x}) = \frac{s_x^2}{n} \qquad \hat{\text{var}}(\bar{x}) = \frac{s_x^2}{n}$$

## CCS: Stima della varianza campionaria delle stime - II

### Stima di proporzioni

Estrazione in blocco

$$\text{var}(p_x) = \frac{P_x(1-P_x)}{n} \frac{N-n}{N}$$

$$\hat{\text{var}}(p_x) = \frac{p_x(1-p_x)}{n-1} \frac{N-n}{N}$$

Estrazione Bernoulliana. come sopra trascurando il fattore  $fpc$ .

(Ritorna a [Campionamento Probabilistico](#))

## Campionamento stratificato: definizioni generali

La **stratificazione** consiste nella costruzione di  $L$  subpopolazioni (strati) il più possibile omogenei al loro interno ed eterogenei tra loro, da cui si estraggono  $L$  campioni indipendenti:  $n = \sum_{h=1}^L n_h$

Se l'estrazione avviene con CCS si parla di *Campionamento Casuale Semplice Stratificato*.

**N.B.** Gli strati devono essere tali da escludersi reciprocamente e da non sovrapporsi, ovvero ogni unità della lista deve appartenere ad uno e ad uno solo degli  $L$  strati.

- [Stratificazione](#)
- [Tecniche di allocazione](#)
- [Considerazioni conclusive](#)

(Ritorna a [Campionamento Probabilistico](#))



➤ Come effettuare la stratificazione?

La soluzione ottimale è, in generale, quella di ripartire le unità della popolazione fra i diversi strati in base alla distribuzione della variabile oggetto di studio  $X$  nell'ambito della popolazione obiettivo.

➤ E se non si possiede alcuna informazione sulla distribuzione di  $X$ ?

E' necessario ricercare una o più variabili altamente correlate con  $X$  da utilizzarsi come ***variabili di stratificazione***.

**Nota bene:**

In conclusione, la procedura di stratificazione richiede l'utilizzo di informazioni *a priori* sulla popolazione oggetto di studio.

(Ritorna a [Campionamento Stratificato](#))

## **Ancora sul campionamento stratificato: l'allocazione del campione**

Determinare l'*allocazione del campione* fra gli strati significa decidere quante unità, fra le  $n$  appartenenti al campione, debbano essere selezionate all'interno di ciascuno strato.

In altre parole, una volta stabilita l'ampiezza campionaria, si decide in che misura ciascuno strato debba contribuire alla formazione del campione.

## Tecniche di allocazione

➤ Diversi tipi di allocazione sono possibili:

- [uniforme](#)
- [proporzionale](#)
- [ottimale](#)

### **Nota bene:**

Il tipo di allocazione scelto influisce pesantemente sui costi e sulla precisione delle stime ma non esiste una regola ottimale a priori: bisogna valutare caso per caso.

(Ritorna a [Campionamento Stratificato](#))

## Allocazione c.d. uguale o uniforme

All'interno di ciascuno degli  $L$  strati verrà campionato lo stesso numero  $n_h$  di unità dando luogo alla seguente situazione:

$$n_h = \frac{n}{L}$$

**Quando si usa?** quando la variabilità è pressoché uguale nei singoli strati. Se gli strati sono di uguale dimensione coincide con quella proporzionale.

(Ritorna a [Tecniche di Allocazione](#))

## **Allocazione proporzionale (o a *frazione di sondaggio costante o autoponderante*)**

Vi si ricorre quando le deviazioni standard degli strati sono pressoché costanti; in tal caso. La frazione di campionamento,  $f_h = (n_h / N_h)$ , è la stessa in ogni strato

$$f_h = f = \frac{n}{N}.$$

di modo che *il numero di unità campionate all'interno del generico strato è direttamente proporzionale all'ampiezza dello strato stesso.*

### **➤ Nota:**

L'adozione della *stratificazione proporzionale* consente di semplificare in maniera considerevole le procedure di calcolo.

(Ritorna a [Tecniche di Allocazione](#))

<b>media</b>
$\bar{X}_{st} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} X_{hi}}{n}$
N.B. Si noti che quest'ultima espressione coincide con la media campionaria.
<b>ammontare globale</b>
$\hat{X}_{st} = N \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} X_{hi}}{n}$
<b>stima di proporzioni</b>
$p_{x,st} = f \sum_{h=1}^L p_{hx}$

**Tabella 1.** Espressioni semplificate per la stima dei parametri della popolazione nel caso di campionamento stratificato con allocazione proporzionale.

**media**

$$\text{var}(\bar{x}_{st}) = \left( \frac{N-n}{N^2} \right) \sum_{h=1}^L \left( \frac{N_h^2}{N_h-1} \right) \left( \frac{s_{hx}^2}{n} \right)$$

N.B. Quest'ultima, per valori sufficientemente elevati degli  $N_h$ , può essere approssimata in maniera soddisfacente da:

$$\text{var}(\bar{x}_{st}) \approx \left( \frac{\sum_{h=1}^L N_h s_{hx}^2}{N} \right) \left( \frac{N-n}{Nn} \right)$$

**ammontare globale**

$$\text{var}[\hat{X}_{st}] = N^2 \text{var}(\bar{x}_{st})$$

**stima di proporzioni**

$$\text{var}(p_{x,st}) = \left( \frac{N-n}{N^2} \right) \sum_{h=1}^L \left[ \frac{P_{hx}(1-P_{hx})}{n-1} \right]$$

**Tabella 2.** Varianza campionaria delle stime nel caso di campionamento casuale stratificato con allocazione proporzionale.

## **Allocazione ottima (o a frazione *allocazione ottima* di sondaggio variabile)**

Si definisce delle unità del campione quella che, in corrispondenza di un livello di costo prefissato, consenta di minimizzare la varianza campionaria delle stime o che, dato un certo valore della varianza campionaria, ci permetta di minimizzare il costo globale dell'indagine.



## Quando si usa l'allocazione ottimale?

Vi si ricorre quando gli strati hanno varianze notevolmente diverse tra loro; in particolare, un maggior numero di unità verrà campionato da quegli strati che:

1) hanno un'ampiezza maggiore

2) hanno una varianza interna maggiore

3) presentano un costo di campionamento per unità minore che negli altri strati.

➤ È necessario assumere che la struttura dei costi sia spiegata da una opportunamente definita [funzione di costo](#).

## Funzione lineare di costo

➤ L'andamento del costo globale dell'indagine può essere facilmente approssimato attraverso la funzione lineare:

$$C = c_0 + \sum_{h=1}^L c_h n_h$$

dove

$C$  = costo globale dell'indagine

$c_0$  = componente fissa di  $C$

$c_h$  = il costo di campionamento per singola unità, variabile da strato a strato.

$n_h$  = numerosità campionaria nello strato  $h$ -esimo

## Ancora sulle funzioni lineari di costo

➤ Se i costi di viaggio incidono pesantemente su  $C$ , sarà opportuno esprimerli separatamente come:

$$\sum_{h=1}^L t_h \sqrt{n_h}$$

dove  $t_h$  è il costo di viaggio sostenuto per singola unità.

➤ Quando il costo di campionamento è costante nei diversi strati ( $c_h=c$ , per ogni  $h=1,\dots,L$ ) si ha:

$$C = c_0 + cn$$

(c.d. allocazione di Neyman).

(Ritorna a [Tecniche di Allocazione](#))

## Confronto fra CCS e CCS stratificato

La stratificazione, se opportunamente attuata, consente di ottenere forti guadagni in efficienza rispetto al c.c.s. (vd. [esempio](#)).

➤ Non si può tuttavia affermare, che il campionamento casuale stratificato risulti in ogni caso più efficiente del c.c.s.:

*l'adozione di criteri di stratificazione inadeguati, insieme ad un'allocazione del campione molto lontana da quella ottima, possono far lievitare fortemente la varianza delle stime ottenute da campioni casuali stratificati.*

➤ L'informazione ottenuta con il campionamento stratificato presenta un elevato grado di analiticità. È possibile, infatti, ricavare stime separate dei parametri per ognuno degli strati formati

....

➤ In molte situazioni, poi, l'estrazione di un campione stratificato non presenta difficoltà maggiori di quelle legate all'estrazione di un campione casuale semplice.

In sintesi, rispetto al CCS, il campionamento casuale stratificato spesso consente di ottenere campioni maggiormente rappresentativi senza rinunciare alla semplicità concettuale del campionamento casuale semplice e senza condurre ad un eccessivo incremento dei costi dell'indagine.

(Ritorna a [Campionamento Stratificato](#))

## Esempio illustrativo: la stima della media

➤ Si considerino le seguenti quantità:

$V_{ccs}$  = varianza della media campionaria ottenuta nel caso del campionamento casuale semplice con estrazione in blocco;

$V_{prop}$  = varianza della stima della media ottenuta da un campione casuale stratificato con allocazione proporzionale;

$V_{opt}$  = varianza della stima della media ottenuta da un campione casuale stratificato con allocazione ottima, per  $n$  prefissato (*allocazione di Neyman*).

➤ Risulterà in ogni caso verificata la seguente relazione:

$$\boxed{V_{opt} \leq V_{prop} \leq V_{ccs}}$$

(Ritorna a [Ancora sul C. Strat](#))

## **Campionamento a grappoli**

**Ad uno stadio (single stage).** suddivisione della popolazione in sottoinsiemi (grappoli) e selezione casuale di tali grappoli; rilevazione su tutte le unità statistiche di ogni grappolo selezionato

**A due o più stadi (multi stage).** suddivisione della popolazione in sottoinsiemi (unità di I stadio), selezione casuale di tali unità, quindi selezione casuale delle osservazioni (unità di II stadio) dai sottoinsiemi precedentemente estratti e così via a catena.

## **Alcune considerazioni sul campionamento a grappoli**

- Generalmente è preferibile creare i grappoli (clusters) il più possibile omogenei fra loro ed eterogenei al loro interno in maniera tale da minimizzare la variabilità *esterna* fra grappoli e massimizzare quella *interna* fra le unità elementari.
- Sono possibili schemi misti come, ad esempio, il campionamento a grappoli stratificato.
- La dimensione del campione non è spesso nota ma va trattata come una variabile casuale

Ma passiamo ora ad esaminare i vantaggi che comporta l'adozione del campionamento a grappoli.



## Vantaggi del campionamento a grappoli

Sono legati essenzialmente alla riduzione dei costi e dei tempi dell'indagine:

- Non è necessaria una lista delle unità elementari di campionamento ma solo di quelle di primo stadio
- Si riducono i costi ed i tempi di viaggio perché le unità campionate all'interno di ogni cluster saranno generalmente legate da un vincolo di contiguità territoriale.

(Ritorna a [Campionamento Probabilistico](#))

## Campionamento sistematico

È caratterizzato dalla scelta casuale della prima unità del campione seguita dalla selezione di quelle successive con passo di campionamento costante  $k$ , con  $k=N/n$  ( $n$  = numerosità campionaria,  $N$  = numerosità della popolazione).

- Può essere definito come caso particolare sia del campionamento stratificato che del campionamento a grappoli.
  
- Viene utilizzato dall'ISTAT al secondo stadio delle indagini sulla spesa delle famiglie e sulle forze di lavoro per il campionamento delle famiglie dalle liste anagrafiche dei comuni selezionati al primo stadio.

(Ritorna a [Campionamento Probabilistico](#))

## **Campionamento ripetuto (*Panel*)**

Un **panel** è un campione continuativo (o permanente) di unità statistiche contattate in successivi periodi di tempo (a meno di sostituzioni dovute ad esigenze tecniche)

Scopo: studio delle dinamiche e delle determinanti dei comportamenti

Inconvenienti:

- [difficoltà di costituzione iniziale](#)
- [mortalità spontanea](#)
- [effetto panel](#)

(Ritorna a [Campionamento Probabilistico](#))

**Difficoltà di costituzione iniziale:** è difficile selezionare il campione in maniera tale da garantire al tempo stesso la volontarietà dei partecipanti e la rappresentatività del campione.

**Mortalità iniziale:** nei primi mesi di vita il tasso di abbandono è molto elevato e bisogna programmare i rimpiazzi fin dall'avvio dell'indagine. La mortalità iniziale si fa sentire soprattutto in alcuni gruppi sociali con differenze fra i diversi paesi (es. donne giovani, anziani, tecnici ed operai con basse qualifiche professionali)

**Mortalità spontanea:** legata all'abbandono dell'indagine in fasi successive a quella iniziale. Le cause principali sono stanchezza, diminuito interesse, impressione di non aver più nulla da dire.

**Effetto panel:** si tratta di una modifica qualitativa e/o quantitativa dei comportamenti di acquisto legata alla sensibilizzazione dei partecipanti riguardo ad alcuni della rilevazione (prezzi, pubblicità, fedeltà ad una marca, razionalizzazione degli acquisti) ed alla consapevolezza di poter influire sui risultati dell'indagine.

In più bisogna considerare tutti gli effetti derivanti dalla stanchezza e dal logoramento indotti da una partecipazione prolungata all'indagine.

Un possibile correttivo consiste nel ricorrere alla tecnica del [Panel Ruotato](#)

## Il Panel Ruotato

Un **panel ruotato** campione continuativo con sostituzione a rotazione delle unità statistiche sotto osservazione (Es. [indagine ISTAT sui consumi delle famiglie](#) e sulle forze di lavoro, [indagine Banca d'Italia sui bilanci delle famiglie italiane](#) ).

### Vantaggi:

- migliora la rappresentatività del campione
- riduce il carico di lavoro dei rispondenti
- aumenta l'efficienza delle stime

(Ritorna a [Panel](#))

## Campioni non casuali

- Storicamente precedono quelli probabilistici
- Presentano lo svantaggio di non consentire o rendere difficoltosa la determinazione del grado di precisione delle stime

Fra gli schemi più diffusi e comunemente utilizzati rientrano:

- a) il [campionamento a scelta ragionata \(in senso stretto\)](#)
- b) il [campionamento per quote](#)
- c) [altri disegni](#) fra cui il campionamento bilanciato, i campioni di volontari, il campionamento accidentale e quello a valanga.

(Ritorna a [Piano di Campionamento](#))

## **Il campionamento a scelta ragionata**

➤ Le unità vengono selezionate non casualmente sulla base della valutazione, da parte di un pool di esperti, di informazioni preliminari concernenti la popolazione indagata.

➤ La rappresentatività dei campioni a scelta ragionata dipende:

a) dalla scelta e dalla qualità delle informazioni sulle quali la regola di selezione si basa.

b) dal punto di vista e dall'orientamento del selezionatore. È stato osservato che la scelta ragionata fornisce buone stime solo per caratteri legati da una relazione (lineare) positiva con le variabili che hanno guidato la scelta del campione (Neyman ed il censimento italiano del 1921)



## **Alcune considerazioni sui campioni a scelta ragionata**

➤ In alcune situazioni essi sono tuttavia preferibili a quelli casuali:

a) per piccoli campioni come, ad esempio, nel caso di ampie indagini che interessano relativamente poche unità territoriali (valutazione di una caratteristica da valutare su base sulla base di un campione la cui numerosità, per motivi economici, è limitata a poche unità)

b) per ricerche su beni industriali nelle quali considerazioni di carattere economico sulla rete produttiva e/o distributiva richiedono l'inclusione di alcuni nodi vitali.

(Ritorna a [Camp. Non Prob.](#))

## **Il campionamento per quote**

- La popolazione viene suddivisa in classi o sottogruppi omogenei secondo alcune caratteristiche. È opportuno che le classi differiscano il più possibile fra loro per quanto riguarda la variabile oggetto di studio:
- Il processo di selezione del campione avviene per passi successivi seguendo le fasi che vengono di seguito illustrate.

## Selezione del campione nel campionamento per quote

1) Dai dati censuari (o da altre fonti) si ricava il peso percentuale di ogni classe.

2) Il totale delle unità nel campione viene poi suddiviso fra le classi in modo da rispecchiare la composizione della popolazione pervenendo così alla definizione delle quote ovvero del totale di interviste da effettuare in ciascuna classe. In merito si distinguono due casi:

- a) quote marginali (assegnazioni indipendenti)
- b) quote associate (indicazioni a due o più dimensioni)

3) Nell'ambito delle quote assegnate, la scelta delle unità da intervistare è demandata al rilevatore introducendo così un margine di soggettività nell'analisi che può indurre alcune distorsioni nei risultati.

## **Pregi....**

➤ Consente una riduzione dei costi e dei tempi dell'indagine in quanto non si rende necessaria la compilazione di una lista di campionamento e, conseguentemente, l'identificazione nominativa degli intervistandi.

Ma quale è il prezzo da pagare per tutto questo?

## ....e difetti del campionamento per quote

➤ Un problema comune è che spesso la scelta delle persone da intervistare risponde a criteri utilitaristici (interviste "facili") ovvero: non tutti i componenti della popolazione hanno la stessa probabilità di entrare a far parte del campione.

**Ad esempio** gli abitanti ai piani superiori in zone periferiche mal collegate difficilmente vi rientreranno, gli intervistatori tenderanno ad avvicinare le persone più disponibili etc...

➤ il metodo non comporta il controllo dei rifiuti a collaborare rischiando così di indurre un minore impegno degli intervistatori.

➤ gli intervistatori spesso tendono ad avvicinare persone più o meno simili portando ad una sottostima della variabilità della popolazione.

## Possibili rimedi

- **Punto di partenza prefissato.** Il punto di partenza per ciascun intervistatore è fissato in modo casuale dall'ente che effettua la ricerca.
- **Sentiero prestabilito.** Il rilevatore è vincolato ed effettuare le interviste lungo un tragitto ben specificato
- **Piano delle interviste.** Le interviste debbono essere effettuate alternando un predeterminato insieme di modalità.
- **Orari prestabiliti.** Possono essere fissati degli intervalli orari all'interno dei quali effettuare le interviste.
- **Schemi misti.** Si può utilizzare una procedura a due stadi con stratificazione ed estrazione probabilistica al primo stadio e campionamento per quote al secondo (*campionamento probabilistico per quote*).

(Ritorna a [Camp. Non Prob.](#))

## Altri schemi non probabilistici I

➤ **Il campionamento bilanciato:** le unità del campione vengono selezionate in maniera tale che la media nel campione di una o più caratteristiche note (variabili di controllo) sia uguale, o comunque molto vicina, alla media nella popolazione. *Il legame fra la variabile di controllo e la variabile oggetto di studio ha un'importanza vitale.*

➤ **Il campionamento a valanga:** è molto utile per il campionamento da popolazioni rare (minoranze linguistiche o etniche disperse sul territorio, reduci). Composto da più fasi, dopo aver intervistato alcune persone dotate delle caratteristiche richieste, queste persone servono per identificare altri soggetti che possono essere intervistati in una fase successiva e che a loro volta producono informazioni per identificare altri soggetti con le caratteristiche per essere inclusi nel campione, creando così un effetto *a valanga*.

## Altri schemi non probabilistici II

➤ **Campioni di volontari.** Spesso usati negli esperimenti di tipo clinico o farmacologico.

➤ **Campionamento accidentale.** Si ha un campionamento accidentale quando il ricercatore sceglie come rispondenti alla sua indagine le prime persone che capitano, senza criteri definiti. Ciò che si perde in accuratezza del campione, lo si risparmia in tempo e denaro.

(Ritorna a [Camp. Non Prob.](#))



## **Alcune avvertenze per la compilazione dei questionari: chiarezza e “semplicità”**

### **➤ Evitare formulazioni generiche o ambigue:**

- Il significato di ciascuna domanda deve essere univocamente comprensibile
- Le domande devono essere coerenti con il background culturale dell'intervistato

### **➤ Le domande devono richiedere all'intervistato solo un limitato sforzo di memoria (errori di memorizzazione nella ricostruzione dei fatti o nella loro collocazione temporale)**

## **Alcune avvertenze per la compilazione dei questionari: neutralità delle domande**

- Evitare di condizionare od orientare la risposta:
  - Evitare domande "orientate" oppure che coinvolgano il prestigio o l'autovalutazione dell'intervistato.
  - Evitare di formulare le domande in maniera tale da creare tensione psicologica nell'intervistato.
- Risposte chiuse: bisogna arginare la tendenza a concentrarsi sui valori medi evitando però di orientare la risposta dell'intervistato.
- Domande di "controllo"

(Ritorna al [piano dell'indagine](#))

## Scelta del metodo di raccolta dei dati

Può influire pesantemente sia sulla precisione che sul costo dell'indagine.

➤ intervista diretta

Pregi: tassi di risposta più elevati che per le altre tecniche ed una maggiore accuratezza

Difetti: costi elevati

➤ sondaggio postale

Pregi: basso costo

Difetti: tassi di risposta bassi (solleciti), *autoselezione* dei rispondenti, tempi.

➤ sondaggio telefonico

Pregi: velocità, basso costo (CATI).

Difetti: rischio di costituire campioni non rappresentativi, tassi di risposta bassi

➤ soluzioni miste (es.: sond. telefonico + intervista diretta)

(Ritorna al [piano dell'indagine](#))

## Relazione finale

➤ Nella redazione della relazione o rapporto finale è opportuno seguire alcune raccomandazioni. In particolare la relazione deve contenere delle informazioni specifiche relative a:

- finalità dell'indagine
- popolazione obiettivo e popolazione osservata
- periodo di riferimento e di svolgimento
- piano di campionamento
- metodo di raccolta dei dati
- metodi di stima adottati
- presentazione dei risultati
- accuratezza delle stime e struttura dei costi
- evidenziare le potenziali fonti di [errore](#) che devono indurre cautela nel valutare i risultati dell'indagine.

(Ritorna al [piano dell'indagine](#))

## **Errore campionario e non campionario**

- L'errore di natura campionaria è “necessario” in quanto è legato alla natura stessa dell'indagine campionaria che cerca di ricostruire il comportamento della popolazione da un campione di dimensione ristretta.
- L'errore di natura non campionaria è invece spesso “evitabile” ed è spesso dovuto ad una non accurata progettazione dell'indagine

(Ritorna a [sommario](#))

## **Gli errori "non campionari" nelle rilevazioni campionarie**

**Errori di specificazione:** specificazione difettosa dei fatti o specificazione non rispondente agli scopi della rilevazione; omissioni o duplicazioni dovute a difetti di definizione delle unità della popolazione, o a errori materiali di numerazione.

**Errori di accertamento:** mancanza di rilevatori addestrati ed esperti; errori di misurazione da strumenti o da difficoltà inerenti alla raccolta dei dati, dovuti ad errori di memoria, a mancate risposte e ad altri fattori; carenza di controlli e di supervisione da parte degli organi direttivi.

**Errori di tabulazione:** difettoso esame dei dati di base; errori nelle operazioni di elaborazione dei dati (codificazione, verifica, tabulazione, ecc); errori commessi nella presentazione e stampa di tavole, diagrammi ecc...

Fra le fattispecie considerate, alcuni casi particolari sono rappresentati dagli errori dovuti alle [liste](#) ed a [mancate risposte](#)

(Ritorna a [errori](#))

## Errori dovuti alle liste

➤ Sono riconducibili a:

- **incompletezza** delle liste
- **grappoli di unità compresi in uno stesso nominativo** (es. alloggi in cui vivono più famiglie; come correttivo si può aggiungere 1 stadio)
- **nominativi inesistenti o estranei** (se previsto, integrazione delle liste)
- **lista con nominativi ripetuti** (es. selezione della famiglie da una lista elettorale con probabilità proporzionale al numero di elettori nella famiglia)

(Ritorna a [altri errori](#))



## **Errori dovuti a mancate risposte**

➤ Introducono una componente di distorsione negli stimatori (non response bias). Dipendono da:

- cause accidentali (es. assenza)
- rifiuto cosciente o impossibilità di rispondere

## **Approcci al trattamento delle mancate risposte -I**

➤ **Call-backs** (si specifica in anticipo quanti tentativi verranno fatti prima di rinunciare definitivamente)

➤ **Sottocampionamento dei non rispondenti** : usato per i sondaggi postali; è riconducibile ad un particolare tipo di *campionamento stratificato*:

1. il campione viene diviso in due strati: *rispondenti* e *non rispondenti*.
2. un sottocampione di questi ultimi viene quindi visitato da intervistatori qualificati

## Approcci al trattamento delle mancate risposte -II

➤ **Metodo delle sostituzioni:** non tiene conto delle differenze sostanziali che esistono fra rispondenti e non rispondenti; si rimpiazzano le mancate risposte con nominativi di rispondenti scelti a caso.

(*Ad es.* in indagini periodiche, Kish e Hess, 1959, come correttivo, suggeriscono di sceglierli fra quelli che nelle indagini precedenti non avevano risposto)

➤ **In fase di analisi** si utilizzano poi una serie di **correttivi**. Alcuni di essi prevedono la ricostruzione delle mancate risposte attraverso una serie di variabili ausiliarie e l'impiego di tecniche di regressione.

(Ritorna a [altri errori](#))

# **ALCUNI CASI STUDIO**

**Indagine ISTAT**

**Indagine Banca D'Italia**

## **L'indagine ISTAT sui consumi delle famiglie**

- L'indagine ISTAT sui consumi delle famiglie viene condotta con cadenza mensile dal 1968.
- L'indagine si basa sulla rilevazione di tutti i beni e servizi acquistati nel periodo di rilevazione dalle famiglie campionate espressi in termini di spesa e quantità, per quanto riguarda i beni alimentari, e solo in termini di spesa, per quanto riguarda i beni non alimentari ed i servizi.

## **Riforma dell'Indagine ISTAT sui consumi**

➤ L'indagine è stata radicalmente rinnovata a partire dal gennaio 1997. Le modifiche hanno riguardato tutte le fasi del processo di produzione dei dati ed essenzialmente i seguenti punti:

- disegno di campionamento
- questionari
- procedure di revisione, acquisizione e correzione dei dati
- procedure di controllo delle operazioni sul campo

➤ Fra gli obiettivi della riforma: migliorare la qualità dell'informazione prodotta, armonizzazione di definizioni e metodologie con quelli concordati in sede comunitaria

## Il disegno di campionamento

➤ Si estrae un campione casuale stratificato a due stadi. Il campione comprende ogni anno 488 comuni e 24.000 famiglie.

- Le unità di I stadio sono i comuni
- Le unità di II stadio sono le famiglie. Il territorio italiano è stato suddiviso in **234 strati** in base alla tipologia del comune, alla sua dimensione demografica ed alla regione di appartenenza.

## Campione stratificato di I STADIO

➤ In particolare 107 strati sono formati da un unico comune (**comuni autorappresentativi - AR**) e comprendono tutti i capoluoghi di provincia più altri 4 comuni non capoluogo. Partecipano ogni mese all'indagine.

➤ Gli altri comuni (**comuni non autorappresentativi - NAR**) vengono ripartiti, all'interno di ciascuna regione, in maniera tale da ottenere strati all'incirca della stessa dimensione demografica. In ciascuno dei 127 strati così ottenuti si estraggono 3 comuni campione che partecipano all'indagine rispettivamente il I, il II ed il III mese di ogni trimestre.



## Campione di II STADIO

- Le **famiglie da intervistare** sono estratte in modo casuale dalle liste anagrafiche del comune (famiglie dell'*elenco base* al quale viene affiancato un gruppo di unità di riserva incluse nell'*elenco suppletivo*).
- Al fine di eliminare l'influenza di **effetti di "calendario"**, per ogni mese e per ogni regione sono scelti casualmente due periodi di sette giorni (periodi di riferimento) nei quali effettuare le interviste.

## **Raccolta dei dati**

- La raccolta dei dati è affidata ai comuni che
  - selezionano le famiglie da intervistare
  - individuano, formano, supervisionano e danno assistenza ai rilevatori sotto il controllo dell'ISTAT

## **Coinvolgimento delle famiglie**

➤ Alle famiglie che partecipano alla rilevazione viene chiesto:

- di tenere nota quotidianamente, nei 7 giorni di riferimento, delle spese effettuate per generi di largo consumo (alimentari, tabacchi, giornali, ecc.) in un apposito questionario detto *libretto o diario degli acquisti*.
- di partecipare all'intervista conclusiva condotta all'inizio del mese successivo nella quale vengono rilevate in un secondo questionario (detto *riepilogo delle spese*) notizie socio-demografiche riguardanti i componenti della nonché ulteriori spese relative a mobili ed apparecchiature, ad abbigliamento e calzature, alla salute, ai trasporti e comunicazioni, al tempo libero, spettacoli ed istruzione, ad altri beni e servizi oltre a notizie su reddito e risparmio.

## **Trattamento preliminare dei dati**

- I questionari compilati, dopo un primo esame presso i comuni, vengono controllati da personale esperto presso l'ISTAT il quale verifica che le operazioni di raccolta dei dati siano state effettuate secondo le indicazioni fornite.
- I singoli questionari vengono quindi registrati dal personale dell'ISTAT su supporto informatico, mediante un software che consente il controllo delle coerenze logico-formali dei dati rilevati e degli eventuali valori "sospetti" delle spese annotate.
- Le incoerenze vengono eliminate mediante specifici software, basati su una metodologia probabilistica, i quali correggono gli eventuali valori errati. Il metodo usato si basa sulla ricerca di una famiglia che abbia caratteristiche simili a quella che abbia il dato errato e che sia in grado, quindi, di "donare" l'informazione colmando quella mancante o rimpiazzando quella errata.
- I dati "controllati" sono utilizzati per la stima

## **Informazione prodotta dall'indagine e pubblicizzazione dei dati**

➤ Sebbene la rilevazione avvenga mensilmente, il piano di diffusione dei risultati prevede la pubblicazione annuale della spesa media mensile familiare a differenti livelli di disaggregazione, secondo alcune principali variabili:

- ripartizione geografica e regione
- ampiezza e tipologia familiare
- caratteristiche socio-demografiche della persona di riferimento

ed altre ancora.

➤ Contestualmente alla pubblicazione viene messo a disposizione degli utilizzatori il file standard dei dati elementari conforme alla normativa per la tutela della riservatezza e del segreto statistico.

## **Altri fini “incidentali” dell’indagine**

- Oltre a quelle riferite all'intero anno di rilevazione ogni trimestre vengono calcolate stime per particolari capitoli di spesa, utilizzate per la valutazione dei conti economici nazionali trimestrali.
- I dati dell'indagine sui consumi delle famiglie vengono utilizzati anche per le analisi della distribuzione del reddito e della povertà.

## **L'indagine sui bilanci delle famiglie italiane della Banca d'Italia**

- L'indagine della Banca D'Italia è stata condotta annualmente dal 1964 al 1984. Svolta nel 1986 l'indagine è stata successivamente condotta con cadenza biennale dal 1987 al 1995. L'ultima indagine è riferita al 1998.
- La gamma di variabili rilevate dall'indagine è molto ampia (redditi e risparmi, aspettative di reddito e occupazione, ricchezza reale e attività finanziarie detenute dalle famiglie, diffusione dei mezzi di pagamento, dimensione dell'abitazione di residenza).

## **Il disegno di campionamento e modalità di rilevazione**

- Il disegno adottato nell'indagine sui bilanci delle famiglie tenuta dalla Banca D'Italia è il campionamento stratificato a due stadi.
- All'ultima indagine hanno partecipato 318 comuni e 7.147 famiglie (ca. 1000 in meno delle precedenti indagini).
- Dal 1989 è stato introdotto uno schema in base al quale viene assicurata la permanenza nel campione di un gruppo di famiglie già presenti nel campione relativo alle precedenti indagini (cosiddette famiglie *panel*). La natura *panel* del campione permette di effettuare analisi dinamiche come ad esempio quelle relative allo studio dell'evoluzione della distribuzione dei redditi.
- La rilevazione è affidata ad una società specializzata (CAPI e PAPI).

(Ritorna a [sommario](#))