

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR.DISPARI)

L'INDAGINE CAMPIONARIA

MANUELA SCIONI | 05 – 03 - 2021
A.A. 2020/21



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



L'INDAGINE CAMPIONARIA

È un modo di rilevare informazioni

1. Interrogando
2. Gli stessi individui oggetto della ricerca
3. Appartenenti ad un campione rappresentativo
4. Mediante una procedura standardizzata d'interrogazione
5. Allo scopo di studiare la consistenza e le relazioni fra più fenomeni

L'INDAGINE CAMPIONARIA (2)

1. Interrogando:

L'indagine campionaria comporta la formulazione di domande, con modalità di rilevazione definite (intervista faccia a faccia, intervista telefonica, questionario autocompilato)

2. Gli stessi individui oggetto della ricerca:

Le domande sono poste direttamente agli individui che costituiscono l'oggetto della ricerca, non a osservatori privilegiati, a esperti, ...

3. Appartenenti ad un campione rappresentativo:

poiché normalmente la popolazione oggetto di studio è troppo numerosa per interrogare tutte le unità, si rileva un campione "rappresentativo", ovvero che riproduce le caratteristiche dell'intera popolazione

L'INDAGINE CAMPIONARIA (3)

4. Mediante una procedura **standardizzata** d'interrogazione:

- a tutti i soggetti vengono poste le stesse domande nella stessa formulazione, per garantire la comparabilità delle risposte e la possibilità di analizzarle con metodi statistici. Anche le risposte sono standardizzate (tranne per domande aperte)

5. Allo scopo di studiare la consistenza dei fenomeni e le loro relazioni:

- Stimare il valore assunto da una statistica, ad esempio: consumo medio di sigarette al giorno, statura mediana della popolazione, ...
- Determinare il n° di unità che nella popolazione possiedono un certo attributo, ad es: N° di fumatori, % di popolazione oltre 1,70 mt,
- Stabilire entro quale intervallo cade, con probabilità determinabile, una statistica (stima intervallare)
- Verificare ipotesi concernenti il valore assunto da una statistica (media, mediana, ...)
- Esplorare se – ed eventualmente quali – relazioni esistono tra fenomeni d'interesse

CAMPIONE

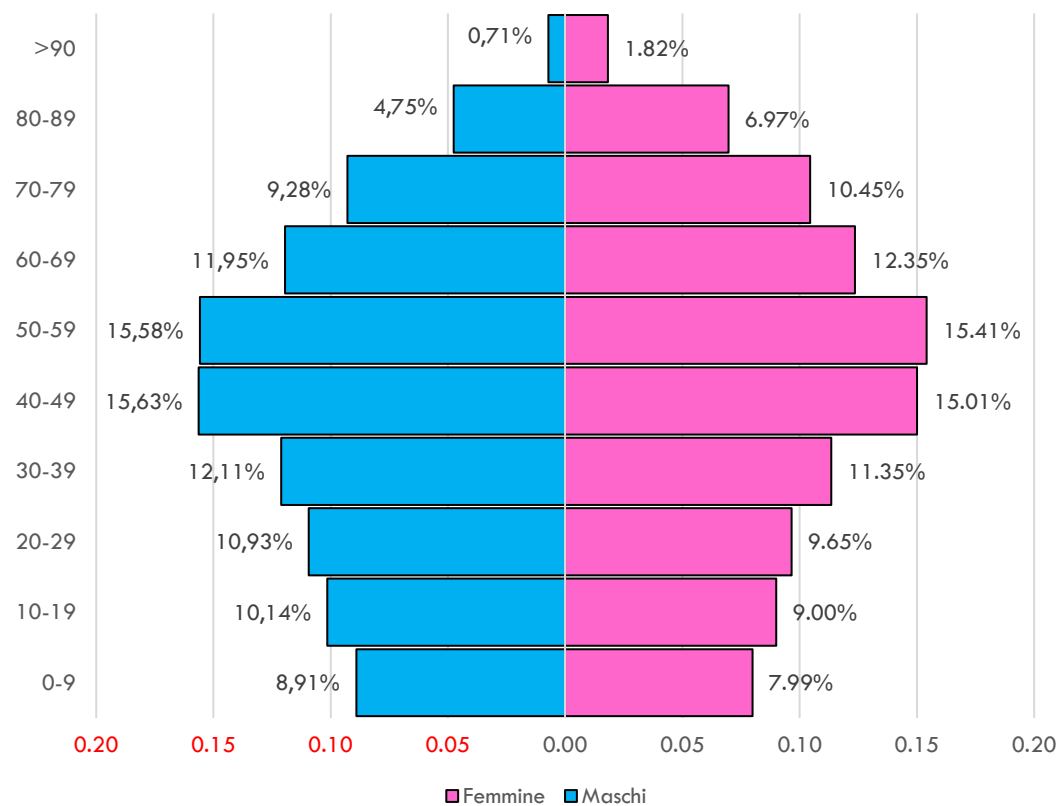
- È un **sottoinsieme** della popolazione, selezionato secondo un insieme di regole dette “disegno di campionamento”, che fanno sì che il campione sia rappresentativo della popolazione
- Le unità campionarie sono sottoposte a osservazione, al fine di ottenere informazioni su una *statistica* (*parametro*: misura di una caratteristica della popolazione; *statistica* o *stima*: misura corrispondente nel campione)

Campione casuale (o probabilistico)

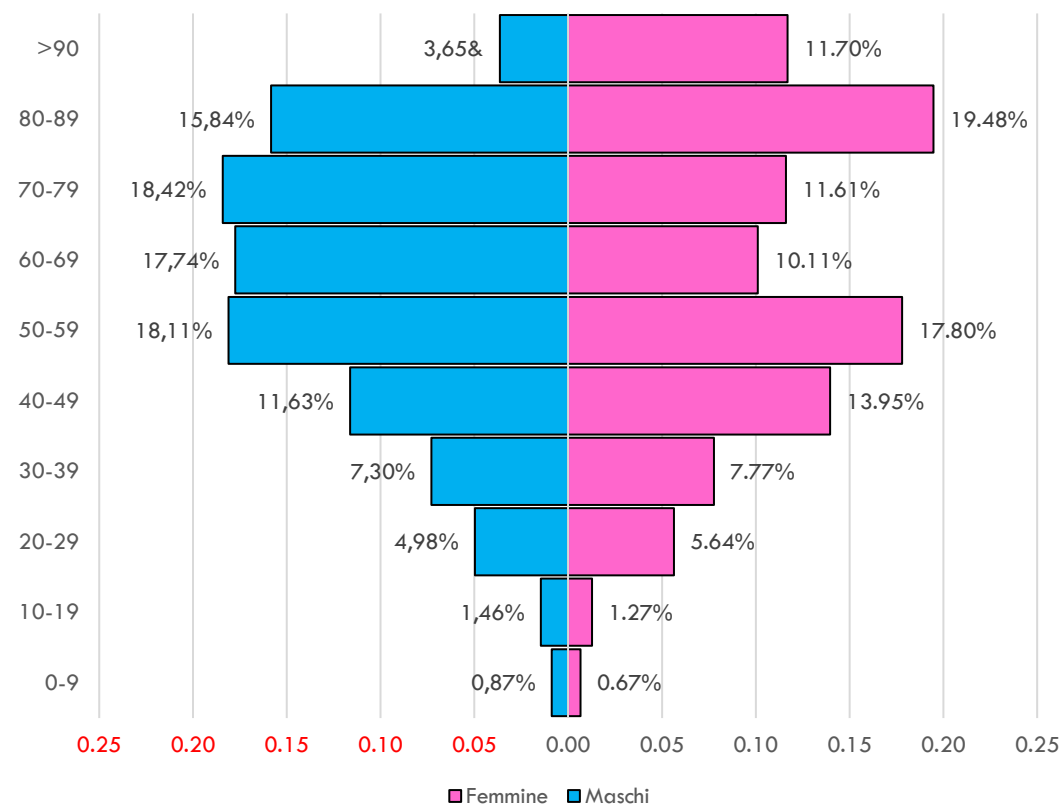
Ogni unità della popolazione ha probabilità non nulla di essere selezionata: $0 < p_i \leq 1, \sum p_i = 1$

IL CAMPIONE RAPPRESENTATO NELLA PIRAMIDE A DESTRA NON È RAPPRESENTATIVO DELLA POPOLAZIONE ITALIANA

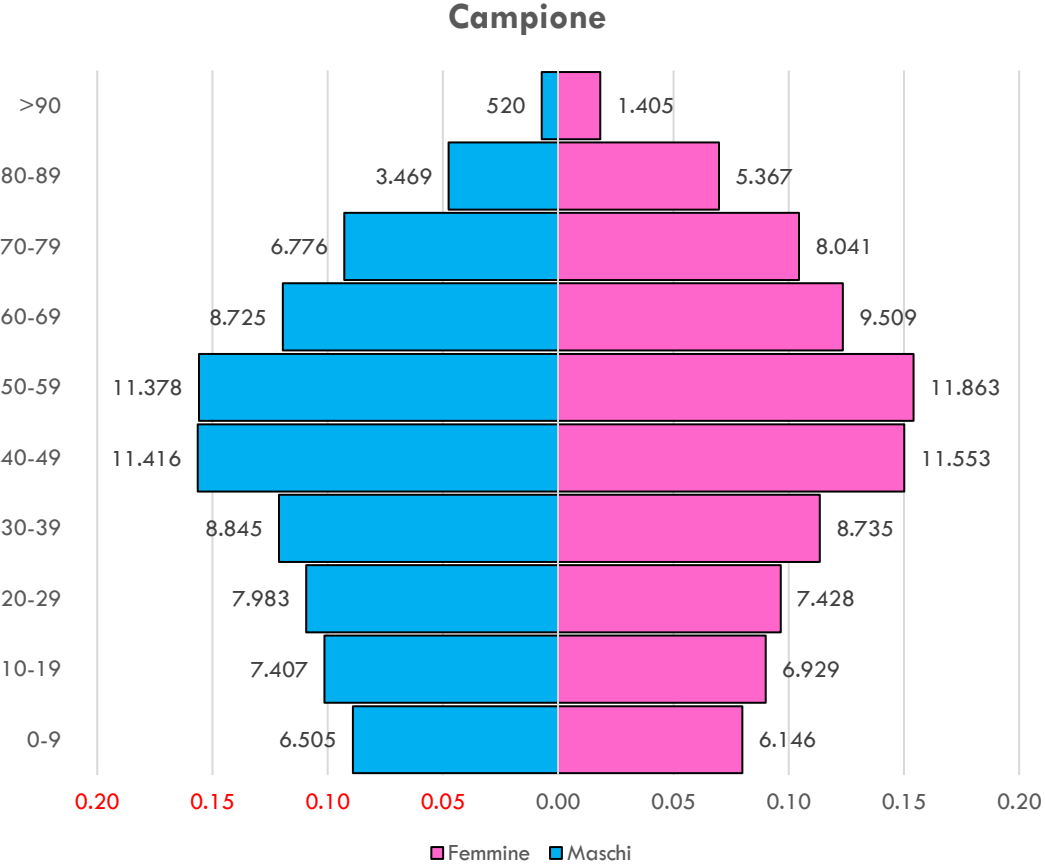
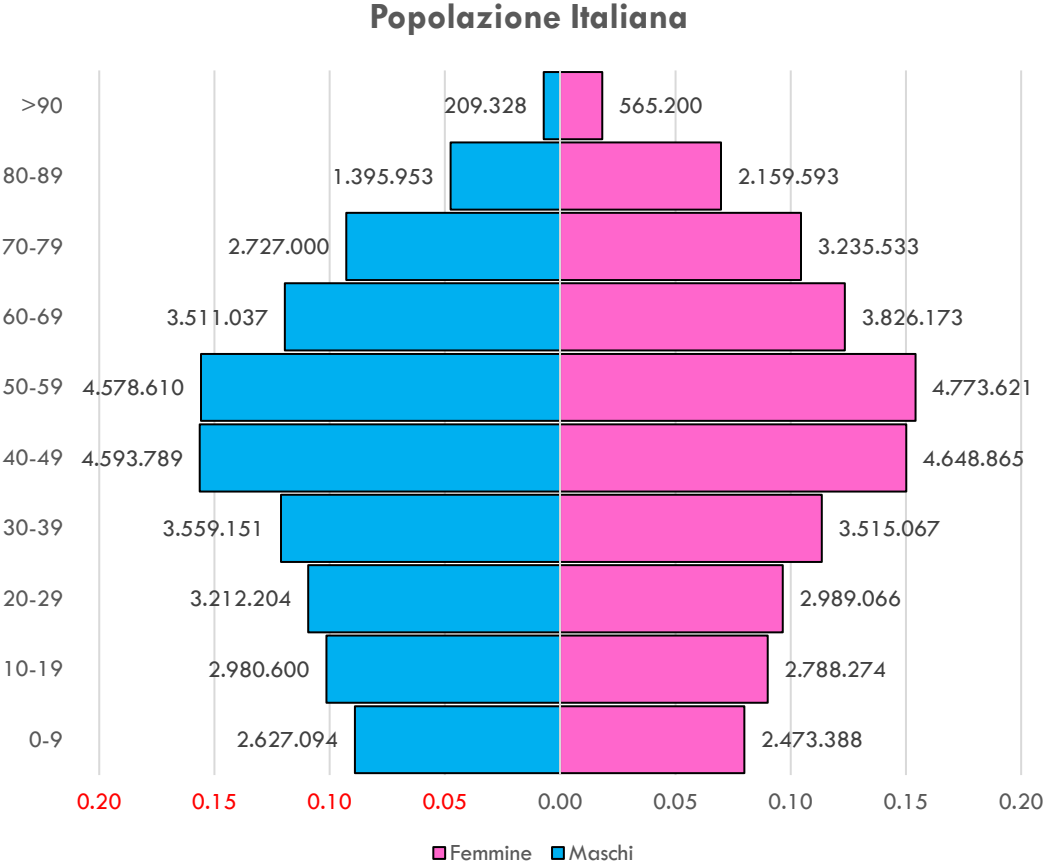
Popolazione Italiana



Campione



UN CAMPIONE È RAPPRESENTATIVO QUANDO LA DISTRIBUZIONE DELLE VARIABILI CORRELATE CON IL FENOMENO CHE SI VUOLE STUDIARE È UGUALE A QUELLA NELLA POPOLAZIONE.




VARIABILE E PARAMETRO DI INTERESSE

Variabile d'interesse: $Y = \text{reddito}$

Parametro d'interesse: reddito medio

CI SARANNO TANTE VARIABILI IN
UN QUESTIONARIO
MA POCHE SONO QUELLE DI
INTERESSE



Variabile d'interesse: $Y = \text{fumare (si-no)}$

Parametro d'interesse: % fumatori

Ogni indagine prevede molte variabili, e quindi parametri, d'interesse

VARIABILI E PARAMETRI DI INTERESSE

9. ASSICURAZIONE

(PER TUTTI)

9.1 Lei ha (da solo o insieme ad altri componenti della famiglia):

(una risposta per ogni riga)

NO SÌ

Un'assicurazione sulla salute o
sugli infortuni1 ☐ 2 ☐

Un'assicurazione sulla vita o
per pensione integrativa.....3 ☐ 4 ☐

1. PESO E STATURA

(PER TUTTI)

1.1 Può indicare il suo peso e la sua statura attuali?

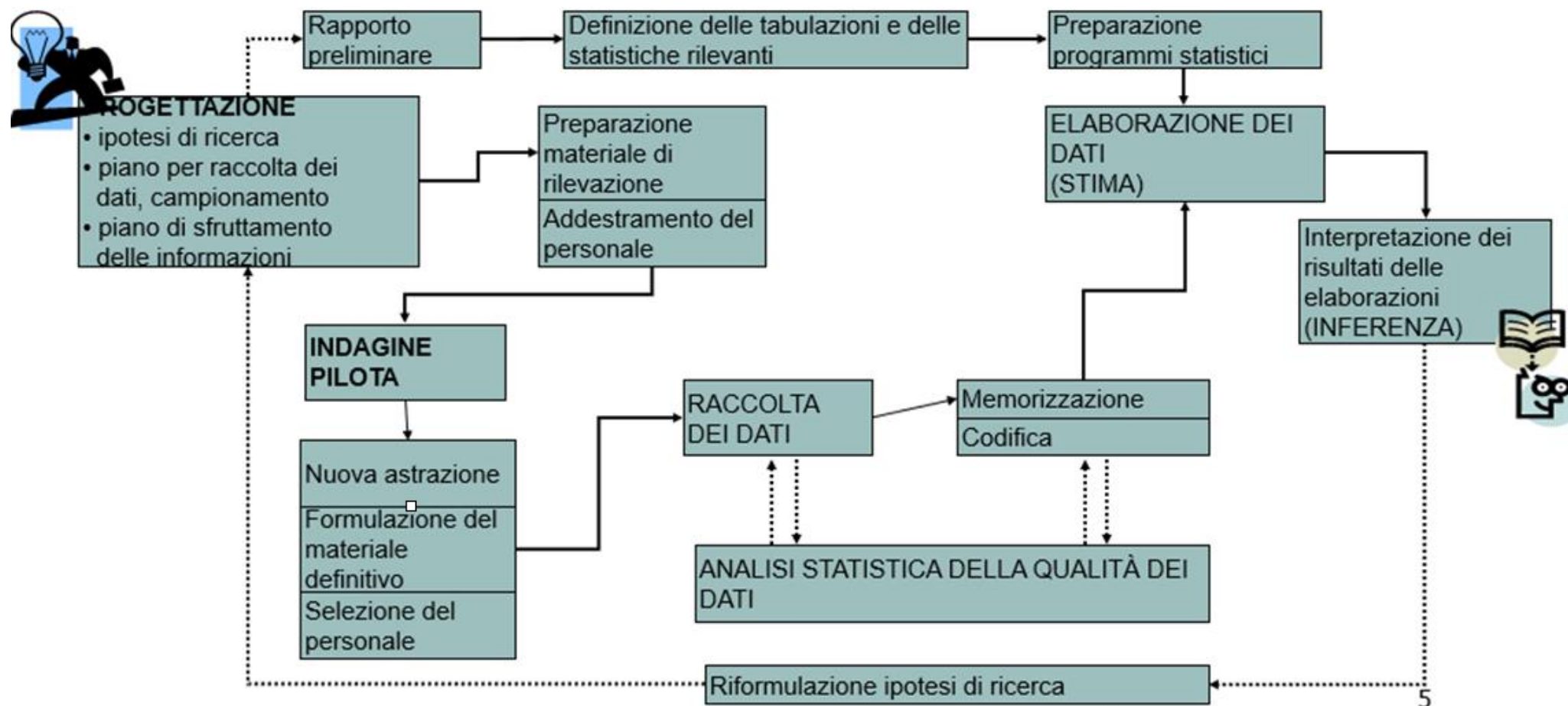
Le donne in attesa di un bambino indichino il peso prima della gravidanza.

Peso kg

Statura cm

DUE VARIABILI PER
UNA DOMANDA

FASI DI REALIZZAZIONE INDAGINE STATISTICA



PROGETTAZIONE DELL'INDAGINE

1. Specificazione degli obiettivi dell'indagine
2. Specificazione della popolazione obiettivo, delle variabili e delle informazioni già note su di essa e dei parametri di popolazione di interesse
3. Analisi delle risorse **finanziarie**, umane, tecniche, materiali disponibili
4. Reperimento (o costruzione) della lista
5. Scelta se effettuare indagine esaustiva o campionaria
6. Specificazione dei tempi
7. Specificazione dei metodi e degli strumenti per la raccolta dati (es. costruzione questionari)
8. Specificazione del disegno di campionamento e dei suoi aspetti tecnici (selezione delle unità e determinazione della dimensione del campione)
9. Specificazione dei metodi di elaborazione (incluse le fasi di editing e imputazione)
10. Specificazione degli stimatori dei parametri e della loro precisione
11. Training del personale e organizzazione del lavoro sul campo
12. Allocazione delle risorse alle differenti operazioni legate all'indagine (compresa la fase di controllo e valutazione)

LE PRINCIPALI SCELTE DA FARE

- Indagine esaustiva o campionaria
- Tecnica di rilevazione
- Ampiezza del questionario
- Piano di campionamento (se indagine campionaria) e numerosità campionaria
- Modalità di controllo e correzione dei dati
- Modalità diffusione dei risultati
- Vincoli da rispettare:
 - Risorse umane
 - Budget
 - Tempi

INDAGINE SUGLI SBOCCHI PROFESSIONALI DEI LAUREATI

Definizione popolazione obiettivo e
popolazione statistica

Quali possibili variabili (e relativi parametri)
di interesse?

Quali variabili di stratificazione?



TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR.DISPARI)

INDAGINE ESAUSTIVA VS CAMPIONARIA

MANUELA SCIONI | Lezione 2- 11/03/2020
A.A. 2019/20



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



MATERIALE PARZIALMENTE TRATTO DA MATERIALE PRECEDENTE PROF.SSA BOCCUZZO

INDAGINI ESAUSTIVE

È rilevata tutta la popolazione oggetto d'indagine

Censimenti (oggi permanenti)

Della popolazione e delle abitazioni (fornisce la popolazione legale)

Dell'agricoltura

Dell'industria, commercio, artigianato e servizi

Vantaggi:

- Dati individuali collegabili nel tempo
- Dati su popolazioni rare
- Base per selezionare campioni e per calcolare tassi
- Collegare i dati provenienti da altre fonti
- Esente da errore campionario

Svantaggi:

- Estremamente complessi dal punto di vista organizzativo
- Costosi
- Risultati disponibili dopo molto tempo
- Considerevoli errori non campionari

IL CENSIMENTO DELLA POPOLAZIONE E DELLE ABITAZIONI

- Il primo Censimento della popolazione italiana risale al 1861
 - Caratteristiche: universalità, simultaneità, periodicità (cadenza tipicamente decennale)
- In realtà, il censimento è “una macchina complessa e costosa”
 - Il costo per la realizzazione del 15° Censimento della popolazione e delle abitazioni è stato di poco inferiore ai 500 milioni di euro
 - I dati sono stati diffusi nel dicembre 2013 e giugno 2014



I Censimenti diventano permanenti

IL CENSIMENTO PERMANENTE

- L'Istat ha avviato la stagione dei **censimenti permanenti** che vede la realizzazione di rilevazioni continue a cadenza annuale, biennale e triennale.
- La strategia dei censimenti permanenti è estesa a tutte le aree tematiche: popolazione e abitazioni, imprese, istituzioni non profit e istituzioni pubbliche, agricoltura.
- **DIFFERENZA** con i censimenti precedenti: non coinvolgono tutti i cittadini, le imprese e le istituzioni, ma di volta in volta dei campioni **rappresentativi**. Tuttavia, la restituzione al Paese dei dati ottenuti è paragonabile al livello censuario
- Questo è possibile grazie all'integrazione tra **fonti amministrative** e **rilevazioni campionarie**.

IL CENSIMENTO PERMANENTE: STRATEGIA DI RILEVAZIONE

CENSIMENTO PERMANENTE: DATI AMMINISTRATIVI + DATI DA INDAGINI

Flusso continuo di acquisizione di dati amministrativi

Sistema dei Registri (SIR)
presso ISTAT

Dati da Indagini
Campionarie

- **SIR:** sistema di registri, costruiti ed alimentati principalmente da fonti amministrative, relativi a individui (e famiglie), imprese, attività e luoghi. È un sistema in evoluzione, alcuni registri sono in produzione, altri in fase di progettazione, altri ancora da programmare
- **Dati da Indagini Campionarie:** il campione delle famiglie censuarie, definito master, diventa la base delle indagini censuarie

IL CENSIMENTO PERMANENTE COME VIENE REALIZZATO?

- Campionamento a due stadi:
 - 1 stadio: Territorio amministrativo (Comuni)
 - 2 stadio: Sezioni e Famiglie
- Le prime due rilevazioni del 2018 e 2019 hanno visto il coinvolgimento per anno di un campione di circa 1 milione e 400 mila famiglie in oltre 2.800 comuni. L'ultima rilevazione si è chiusa il 20 dicembre 2019
- Due indagini con due obiettivi diversi:
 - Rilevazione areale: per valutare la copertura del Registro degli Individui e Famiglie
 - Rilevazione da lista: arricchimento informativo del SIR

IL CENSIMENTO PERMANENTE — COVID 19

- Nel 2020, a fronte dell'emergenza Covid-19, l'Istat ha modificato parzialmente il disegno censuario
 - Sono state sospese tutte le attività di raccolta dati sul territorio e presso le famiglie, tradizionalmente svolte da ottobre a dicembre
 - Sono state avviate, a livello organizzativo, sin dal mese di marzo, e intensificando l'utilizzo dei dati amministrativi.
- Nel 2021 è prevista la ripresa delle rilevazioni sul campo e il recupero di tutte le attività sospese nel 2020.

REGISTRI AMMINISTRATIVI

Sono frequentemente utilizzati a fini statistici. Esempi: registri delle nascite, delle morti, della popolazione residente, degli iscritti a scuola. NB: l'unità di rilevazione è quasi sempre un evento.

Vantaggi:

- Disponibili a costo praticamente nullo

Svantaggi:

- Generalmente contengono poche informazioni, raccolte a fini amministrativi e non statistici
- Scarsa qualità
- Difficile capire se sono davvero stati registrati tutti gli eventi
- Non si coglie ciò che non è ufficiale (es. aborti clandestini)
- Difficile mantenere la serie storica: dipendono da modifiche normative

INDAGINI CAMPIONARIE O INDAGINI ESAUSTIVE?

Vantaggi dell'indagine campionaria

- Ricchezza e profondità dei contenuti
- Rapidità di esecuzione e tempestività dei risultati
- Più accurata la rilevazione
- Flessibilità del disegno
- Minor costo
- Minor numero di rispondenti coinvolti
- Indispensabile quando le unità campionarie sono destinate a distruzione (es. durata lampadina)
- Fattibile in paesi in via di sviluppo

Svantaggi dell'indagine campionaria

- Non idonea per analisi dettagliate, stime non significative per aggregati minori
- La copertura è spesso incompleta
- Invasiva per gli intervistati
- Errore campionario

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR.DISPARI)

GLI ERRORI NELL'INDAGINE STATISTICA

MANUELA SCIONI | Lezione 2- 11/03/2020
A.A. 2019/20



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



MATERIALE PARZIALMENTE TRATTO DA MATERIALE PRECEDENTE PROF.SSA BOCCUZZO

GLI ERRORI NELL'INDAGINE STATISTICA



ERRORE NON CAMPIONARIO: SELECTION BIAS

Un rigoroso campionamento probabilistico riduce l'errore non campionario dovuto a distorsione da selezione (**selection bias**): errore che emerge se la popolazione obiettivo è diversa dalla popolazione campionata.

Ad esempio si può indurre selection bias se:

- Si selezionano le unità più facili da raggiungere
- Si selezionano le unità che paiono "rappresentative" a giudizio dell'intervistatore o del ricercatore
- Si specifica male la popolazione obiettivo
- Si usano cattivi frame (es. sottocopertura)
- Si sostituiscono le unità del campione (per assicurarsi la numerosità del campione desiderata)
- Si utilizzano nel campione unità che si selezionano sulla base della volontarietà
- Si riscontrano mancate risposte (per irreperibilità o rifiuto)

DATI MANCANTI

- **Missing Completely At Random (MCAR)**

I dati mancanti non dipendono né dalla distribuzione della variabile di interesse né da altre variabili rilevate;

- **Missing At Random (MAR)**

I valori mancanti non dipendono dalla variabile di interesse, ma da altre variabili rilevate;

- **Not Missing At Random (NMAR)**

I valori mancanti dipendono dalla variabile di interesse.

ERRORE NON CAMPIONARIO: MEASUREMENT BIAS

Distorsione da misura: errore dovuto al fatto che la variabile di interesse è misurata nel modo errato

Possibili cause:

- Le persone talvolta non rispondono il vero, spesso in caso di domande su temi sensibili
- Le persone potrebbero non capire la domanda, magari perché espressa in modo complicato (es. doppia negazione)
- Le persone dimenticano gli eventi o tendono a dichiarare eventi accaduti prima del periodo di riferimento (telescoping)
- Le persone possono rispondere diversamente a intervistatori diversi, talvolta per accondiscendenza
- L'intervistatore può codificare la risposta in maniera errata
- L'ordine delle domande e il modo di porle può portare a risposte differenti

IL DISEGNO DELL'INDAGINE TOTALE

Ricordiamoci sempre:

“Garbage in – Garbage out”

Non basta un buon disegno campionario, è necessario un disegno dell'indagine in tutte le sue fasi, che tenga sotto controllo anche tutte le fonti di errore non campionario

IL DISEGNO DELL'INDAGINE TOTALE (2)

AZIONI

- Scomporre l'indagine negli aspetti che possono minare le qualità delle stime
- Prevenire i rischi di scarsa qualità
- Inserire metodi-sistemi di controllo della qualità dei dati

PRINCIPI

- Minimizzare i costi del rispondente
- Massimizzare le ricompense attese
- Trasmettere il senso di serietà dell'indagine