

1

1. Quali vantaggi e svantaggi presenta una rilevazione campionaria rispetto ad una rilevazione censuaria?


Prima parte progettazione ed esecuzione di una indagine campionaria

Indagine conoscitiva condotta con due approcci distinti:

- Indagine censuaria: il collettivo è esaminato in maniera completa
- Indagine campionaria: è presa in considerazione solo una parte del collettivo

	Indagine censuaria	Indagine campionaria
Risorse economiche	elevate	contenute
Durata delle operazioni	Lunga	Breve
Errore campionario	assente	presente
Altri tipi di errore	presenti	Presenti

Indagine campionaria è portatrice di un errore di campionamento, fornisce una stima

Si ha solo un valore approssimato dei parametri  al quale è associato un grado di fiducia quantificabile solo se il campione è probabilistico e sia rappresentativo

Campionamento

- inferenza statistica: si desumono caratteristiche della popolazione attraverso una parte di essa, il campione
- Popolazione: unità statistiche esaustiva
- Indagine campionaria

Il campionamento può essere

- ✗ Probabilistico: le unità statistiche scelte tramite estrazione casuale
Possibile inferenza e valutare errore campionario
- ✗ Non probabilistico: le unità statistiche sono scelte dalla soggettività di raccoglie le informazioni

Indagine campionarie

- Popolazione oggetto di indagine: universo
Si prendono in considerazione popolazioni finite di numerosità N
- Campione osservato
- Unità di rilevazione
- Unità di analisi

Problemi di selezione: se la popolazione selezionata non rappresenta quella obiettivo.

- Non c'è corrispondenza biunivoca fra la popolazione di selezione e quella obiettivo
- Non riposta

La popolazione sottoposta ad indagine si chiama popolazione d'indagine.

Vantaggi della rilevazione campionaria

- Flessibilità. Modalità organizzative più flessibili;
- Costi. Contenimento dei costi di indagine sul campo;
- Elaborazione dati. Maggiore rapidità nella raccolta e nella elaborazione dei dati;
- Qualità della rilevazione. Possibilità di approfondire l'analisi dei fenomeni oggetto di studio e di garantire una maggiore accuratezza in tutte le fasi dell'indagine, poiché la minore complessità organizzativa favorisce la possibilità di concentrare le risorse sul controllo della qualità della rilevazione.

Tecniche di campionamento non probabilistico

Trovano un grande impiego perché sono più facili da organizzare

Vantaggi:

- Semplicità di organizzazione
- Bassi costi
- Velocità di esecuzione

Limiti:

- ✗ Distorsione da selezione del campione per l'arbitrio di chi raccoglie i dati
- ✗ Non è possibile fare una stima della precisione dei risultati

Quali sono le principali modalità di selezione di un campione?

* Per giudicare dell'attendibilità dei risultati di un'indagine campionaria appaiono quindi decisive le modalità di selezione del campione, riconducibili alle fondamentali procedure:

- scelta di comodo (convenience sampling), esempio il televoto.
- scelta ragionata (judgmental sampling) nei campioni ragionati la scelta delle unità statistiche è in mano a chi dispone il piano di rilevazione, spesso questi ultimi contattano persone che per il ruolo occupato nel settore sono definite opinion leader o testimoni privilegiati;
- scelta casuale (random sampling), consiste in una procedura equivalente all'estrazione di palline numerate. Osservazione: solo quando il campione è scelto casualmente è possibile calcolare il rischio dell'errore a cui ci si espone nella stima delle caratteristiche oggetto di interesse (errore casuale di campionamento) ed estendere mediante procedimenti di inferenza induttiva inversa i risultati all'intera popolazione.

3. Di quali problemi specifici soffre il campionamento di comodo realizzato direttamente presso gli esercizi commerciali?

Soffre di tre potenziali fonti di distorsione, dovute alla scelta non casuale dei modi di rilevazione; sono rispettivamente:

- collocazione punto vendita, perché gli intervistati risiedono spesso nello stesso quartiere, i frequentatori più assidui hanno una maggiore opportunità di essere selezionati e il tipo di negozio può non rappresentare l'intero universo degli utenti del sistema distributivo;
- alla collocazione degli intervistatori, nel senso che il luogo fisico dove avviene l'intervista (all'entrata, all'uscita o all'interno del punto di vendita) influisce decisamente sulla disponibilità a collaborare e sulla qualità delle risposte;
- al tempo, perché il giorno della settimana e l'orario dell'intervista incidono anch'essi sull'andamento dell'indagine.

Si rischia di avere una serie di interviste che prendono in considerazione solo una tipologia di popolazione. Tutto questo porta al fatto che i **risultati sono difficilmente ritenuti rappresentativi**

4. Quali sono i limiti dal punto di vista statistico del televoto?

La scelta di partecipare all'indagine è lasciata all'adesione volontaria di chi è sintonizzato sulla trasmissione. I dati ottenuti non sono riferibili a tutti i radio/telespettatori, né tanto meno rappresentano l'opinione dell'intera popolazione, in quanto derivano da un processo di autoselezione spontanea e non dall'applicazione di un criterio di scelta di natura ragionata o probabilistica.

5. Che cosa si intende per campione ragionato?

Il campionamento ragionato: Nei campioni ragionati l'individuazione degli intervistati è affidata completamente a chi predispose il piano di rilevazione dell'indagine. A partire dalla disponibilità di dati e conoscenze a priori sulla popolazione obiettivo e sulle tematiche oggetto d'interesse vengono selezionate in maniera ragionata specifiche unità statistiche, in quanto ritenute depositarie delle informazioni più rilevanti ai fini del buon andamento della ricerca. Ad esempio, quando sono contattate persone che, per il ruolo occupato nel loro settore di attività o nella società. L'organizzazione della ricerca è snella e i tempi di realizzazione sono abbastanza rapidi.

Non è possibile fare inferenza statistica. L'attendibilità dei dati dipende dalla buona capacità di scelta del campione su cui la ricerca si è basata.

6. Che cosa si intende per campione per quote?

Chi intervista ha totale discrezionalità nel decidere chi coinvolgere nell'indagine, non è stabilito a priori da chi crea il piano di rilevazione (compromettendo la validità del campione).

Può essere interpretato come una variante del campionamento a scelta ragionata, realizzata però con le modalità organizzative tipiche del campionamento di comodo.

-Come per il campionamento di comodo (e a differenza del campione ragionato): Gli intervistatori non sono tenuti a seguire una lista (contenente i nominativi degli appartenenti alla popolazione scelti) e non vi è dunque

3

alcuna selezione campionaria che imponga ai rilevatori di contattare determinate unità statistiche (la scelta degli intervistati è casuale, ma sotto un vincolo).

- A differenza del campionamento di comodo: devono rispettare rigorosamente un vincolo prestabilito in fase di progettazione: il campione realizzato dovrà avere una composizione per specifiche caratteristiche (generalmente di natura sociodemografica) perfettamente identica a quella della popolazione di riferimento, cioè appunto dovrà riprodurre le quote.

Spiegazione del vincolo: campione e popolazione devono condividere la medesima composizione relativa (rispetto a caratteri delle unità statistiche quali il sesso, l'età, la zona di residenza, il livello di istruzione, la condizione professionale, la posizione nella professione), il campione dovrà riprodurre le quote della popolazione.

Problematica del campionamento: grande importanza vengono ad avere in questo caso l'aggiornamento e l'attendibilità delle fonti statistiche a partire dalle quali sono tratte le informazioni necessarie per determinare le quote. L'errore consiste nel pensare che il campione che presenta la stessa presunta composizione relativa della popolazione possa essere rappresentativo di essa; tuttavia, questa rappresentatività non si verifica in concreto, in quanto spesso questo comporta una sottostima sistematica della variabilità.

7) Quali sono le fasi di progettazioni di una indagine campionaria

fasi di progettazione di una indagine campionaria

1. Identificazione della popolazione di riferimento
2. Definizione degli obiettivi della ricerca e delle informazioni da raccogliere
3. Scelta della procedura di rilevazione
4. Scelta dei criteri di selezione del campione
5. Scelta della metodologia di stima dei parametri di interesse della popolazione, se il campione è casuale
6. Determinazione della dimensione del campione
7. Messa a punto del questionario
8. Valutazione dei costi di progettazione della ricerca

8. Quali tipi di liste sono di possibili per campioni di famiglie, di individui o di operatori economici.

Liste e metodi di rilevazione

Le liste più utilizzate sono

- l'anagrafe della popolazione (consentito tuttavia solo agli enti pubblici e agli organismi privati con finalità pubbliche appartenenti al Sistema statistico nazionale),
- le liste elettorali,
- gli elenchi degli abbonati alla telefonia residenziale,
- gli elenchi delle utenze domestiche dell'ENEL, ecc
- specifici database di marketing, creati e gestiti ad hoc da imprese e/o società di consulenza.

Definire obiettivi di ricerca implica che committente e ricercatore si accordino sui risultati desiderati. La prima cosa è definire la popolazione oggetto di studio, identificare le unità statistiche che ne fanno parte e quelle che ne sono escluse.

È frequente non disporre di un elenco completo della popolazione di interesse, e la sua compilazione ex novo può comportare un significativo impiego di tempo e risorse. La progettazione di un campione probabilistico può risultare particolarmente complessa in relazione alla disponibilità e alla qualità delle liste campionarie utilizzabili. Le liste possono essere incomplete, presentare duplicazioni e difetti di aggiornamento, generando potenziali errori sistematici. Si potrebbero usare registri anagrafici che sono però consultabili solo da enti pubblici e organismi privati appartenenti al SISTAN.

Liste elettorali

Sono organizzate in sezioni e per elenchi generali in cui gli elettori sono in ordine alfabetico, distinti in maschi e femmine. L'esclusione dei minorenni limita la copertura completa della popolazione. Un problema nella costruzione di campioni familiari è la possibilità di selezionare due unità appartenenti allo stesso nucleo familiare, quando l'obiettivo è l'inclusione di una sola unità. Le liste elettorali sono utilizzate quando si vuole un contatto diretto con gli intervistati, con interventi e a domicilio

4

Metodo a due stadi: dalla lista elettorale viene sorteggiato casualmente un individuo, con l'idea di ricostruirne il suo nucleo familiare. Il secondo stadio comprende l'utilizzo del metodo del compleanno più recente: viene scelto per l'intervista l'individuo che ha compiuto gli anni più recentemente. Implicitamente stiamo dando più importanza ai nuclei familiari più grandi, in quanto è più facile intercettarli

Elenchi telefonici

lista Incompleta in quanti non tutti hanno una linea fissa o decidono di non apparire per questioni di privacy

Registro delle imprese

Esiste un archivio delle imprese chiamato ASIA Realizzato dall'ISTAT che viene aggiornato in maniera continuativa, ma non è consultabile e utilizzabile da parte dagli istituti di ricerca privati. Di conseguenza la principale lista di riferimento è costituita dal **registro delle imprese**, gestito su iniziativa delle Camere di Commercio nei quali sono elencate tutte le imprese della provincia di riferimento. Le informazioni disponibili sono numerose ma anche di qualità incerta perché non sono previste modalità di aggiornamento.

9. Quali sono le principali modalità di contatto degli intervistati in una indagine campionaria?

Tecniche di rilevazione

Per la raccolta delle informazioni esistono quattro tipologie di rilevazione:

- Interviste personali
- Interviste telefoniche
- Sondaggi postali
- Sondaggi via internet

La scelta della tipologia è strettamente legata al tema dell'indagine, all'ambiente, alla popolazione e ai vincoli di tempo e spesa da rispettare. Ogni tipologia presenta vantaggi e svantaggi, che vanno considerati in maniera attenta al momento della progettazione dell'indagine.

→Papi (paper and pencil interview): questionari stampati su carta

10. Che cosa si intende per metodologia CAPI?

Metodologia CAPI: (Computer Aided Personal Interview). Ci si siede davanti all'intervistato e man mano che l'intervista procede si compila il questionario in formato elettronico che appare sullo schema del dispositivo utilizzato.

Consente la progettazione di un questionario completamente assistito, garantendo un controllo formale di coerenza delle domande e fornendo uno schema di intervista lineare per l'intervistatore. Il percorso delle domande è controllato dal software, riducendo al minimo le possibilità di errore per il rilevatore.

Svantaggi sono i costi elevati, i tempi più lunghi di realizzazione delle interviste e il potenziale condizionamento dell'intervistatore, qualora nell'intervista si affrontino argomenti di natura personale e riservata con distorsione nelle risposte

11. Che cosa si intende per metodologia CATI?

CATI (Computer Aided Telephone Interview)

La rilevazione CATI si basa sull'interazione tra intervistatore e personal computer. Il questionario è memorizzato nel pc e nel corso della telefonata le domande scorrono a video. Le risposte sono digitate direttamente dall'intervistatore e poi salvate in un database abbinato al questionario. Il software controlla passo per passo lo svolgimento dell'intervista e compie in tempo reale verifiche di completezza e di coerenza tra le risposte ottenute. Questo tipo di rilevazione offre:

- vantaggi in termini di economicità e soprattutto di tempestività della rilevazione
- maggiore controllo sull'operato degli intervistatori durante lo svolgimento delle interviste
- ottenere un contatto più agevole con gli intervistati, con la possibilità di raggiungere anche coloro che negli orari delle interviste personali non sono reperibili

13. Che cosa si intende per metodologia CASI?

5

CASI (Computer Aided Mobile Interview)

Metodologie di autointervista attraverso procedure software dedicate (CASI, ovvero Computer Assisted Self Interview).

Questo metodo di rilevazione è un'evoluzione dei sondaggi postali, in quanto questo metodo in formato elettronico presenta il vantaggio di gestire la fase dei controlli durante l'immissione stessa dei dati.

Modi di risposta dell'intervistato:

- il proprio pc con spedizione e ritorno di un questionario all'istituto di ricerca che lo gestisce via e-mail;
- facendo uso della tastiera di un telefono fisso multifrequenza o di uno smartphone;
- o fornendo a voce le proprie valutazioni dopo aver contattato un numero verde telefonico dedicato

14. Che cosa si intende per metodologia CAWI?

CAWI (computer aided web interview)

attraverso questo metodo gli intervistati sono raggiunti da una mail inviata per posta elettronica contenente il link per collegarsi al sondaggio via web (sfruttando le potenzialità della rete internet)

I vantaggi e gli svantaggi di questa metodologia derivano dal mezzo utilizzato.

Gli aspetti positivi derivano dall'abbassamento dei costi, dalla velocità di spedizione e ricezione da parte degli intervistati. Gli accessi per la compilazione del questionario online sono personalizzati mediante l'assegnazione di personali codici di identificazione che permettono ai ricercatori di seguire la dinamica temporale delle risposte, in questo modo l'invio dei solleciti avviene con grande meticolosità

Gli aspetti negativi sono dovuti al fatto che bisogna far riferimento solo agli utenti della rete. Quanto più le caratteristiche sociodemografiche di questi differiscono con quelle della popolazione tanto più i risultati rischiano di dare una immagine distorta della realtà, non dando risultati significativi. L'utilizzo di Internet non può essere considerato un sottoinsieme casuale dell'intera popolazione italiana.

Mancano anche liste o elenchi che possono rendere agevole il lavoro di chi deve progettare la ricerca.

Indagine web

Ha come vantaggio i costi contenuti, non ci sono spese di stampa, di spedizione di materiali, costi connessi al reclutamento e il compenso dei rilevatori.

Come per tutte le tecniche Computer Assisted

- La registrazione dei dati è intrinseca all'acquisizione
- È prevista una trasmissione elettronica dei dati
- I percorsi di compilazione sono automatizzati
- I questionari presentano una elevata flessibilità
- È possibile prevedere controlli di coerenza in fase di acquisizione dei dati
- È possibile creare pop-up con le istruzioni per la compilazione corretta, anche per singoli quesiti
- La struttura può essere più complessa di un questionario cartaceo

15. Secondo quali schemi può avvenire l'estrazione di campione casuale semplice

Le unità della lista di campionamento hanno tutte stessa probabilità di essere selezionate pari alla frazione di campionamento $f = \frac{n}{N}$

La semplicità è il punto di forza, ha come svantaggi la possibilità che il campione di trovi sparso sul territorio con conseguenti costi elevati per l'organizzazione. Inoltre, poiché tutti i campioni hanno la stessa probabilità di essere estratti è possibile estrarre un campione non rappresentativo.

- Con remissione. L'unità selezionata viene reintrodotta nella popolazione prima dell'estrazione successiva, in modo che la probabilità di selezione rimanga costante x ogni estrazione.
- Senza remissione. L'unità selezionata non viene reintrodotta, riducendo di una unità la popolazione ogni estrazione successiva.

16. come può essere utilizzato un campionamento casuale con passo sistematico

Campionamento sistematico

6

È una variante del CCS, in un contesto in cui viene semplificata la procedura di selezione casuale: si seleziona una unità per ogni K presenti nella lista, dove k è detto passo campionamento, $k = \frac{N}{n}$.

Ha come vantaggio che, per formare un campione, è sufficiente una sola estrazione casuale; il rischio è una scarsa rappresentatività nel caso la lista da cui si estraggono le unità statistiche presenti un qualche ordinamento.

In sostanza il campionamento sistematico utilizza un passo ($k = N/n$) per l'estrazione delle unità statistiche, invece che un'estrazione randomica. Se si vuole estrarre da una popolazione costituita da N unità un campione di dimensione pari a n , posto: $N/n = k$, si potrà selezionare casualmente da un'urna contenente i primi k numeri naturali uno di questi (che risulterà minore o uguale a k), indicato con il simbolo r . Questo numero contrassegna il posto d'ordine occupato dalla prima unità da estrarre dalla lista di tutte le unità che compongono la popolazione di base. Dopo l'unità che occupa il posto r si contano k posizioni successive e si seleziona quella che occupa il posto $(r + k)$, e poi quelle che occupano via via i posti $(r + 2k)$, $(r + 3k)$, fino a quella che occupa il posto $[r + (n - 1)k]$; dopodiché il campione di n casi sarà stato selezionato.

17. Quali sono i vantaggi di un campione stratificato

Campionamento stratificato

L'obiettivo è quello di estrarre un campione più efficiente rispetto al CCS, avere stime quindi più precise o una numerosità campionaria inferiore. Sfrutta alcune conoscenze relative alla popolazione e possedute a priori dal ricercatore. Poiché esiste una relazione diretta tra l'ampiezza di un campione e la variabilità della caratteristica d'interesse, può risultare vantaggioso in termini di efficienza degli stimatori suddividere la popolazione in strati quanto più possibile omogenei al loro interno ed eterogenei tra loro.

Da ogni strato viene estratto un CCS, l'aggregazione di tali campioni produce un campione stratificato. È necessario disporre informazioni aggiuntive per ogni unità di popolazione

Si hanno delle stime più efficienti rispetto al CCS, la possibilità di stimare le variabili in sottopopolazioni di particolare interesse per gli scopi di ricerca e la riduzione della probabilità di estrarre campioni poco rappresentativi. Se le variabili ausiliarie non sono sufficientemente correlate con il fenomeno oggetto di interesse, ne può derivare una perdita di efficienza

18 cosa significa affermare che un campione stratificato di tipo proporzionale è auto ponderante?

Campione stratificato di tipo proporzionale

Riproduce la stessa composizione della popolazione in termini di dimensione degli strati. Ogni unità ha una probabilità di inclusione nel campione pari alla frazione di campionamento dello strato di appartenenza. Così si campiona di più gli strati più numerosi della popolazione. Il campione è così autoponderante, nel senso che per costruire gli stimatori di popolazione non è necessaria una ponderazione a posteriori dei dati.

19 che cosa si intende per campione stratificato di tipo ottimale

Mentre nel campionamento stratificato la numerosità degli strati nel campione è scelta in base alla numerosità degli strati della popolazione (più uno strato in popolazione è numeroso, maggiore sarà il numero di unità campionate); Nel campionamento stratificato ottimale si sfrutta un'informazione (la varianza all'interno degli strati), che se conosciuta a priori permette di campionare maggiormente negli strati in cui c'è una maggiore variabilità rispetto alla variabile oggetto di interesse. (In termini puramente numerici alcuni strati risulteranno sovrarappresentati ed altri sottorappresentati)

20 che cosa si intende per campione stratificato con criterio di allocazione non proporzionale e non ottimale

Questo si utilizza quando si ha interesse a fornire stime attendibili non solo a livello dell'intera popolazione ma anche per i singoli strati, che in questo caso svolgono il ruolo di sotto domini di riferimento. Se la variabile di stratificazione è di tipo territoriale nell'ambito di una indagine nazionale, si può volere che le stime dei parametri di interesse siano statisticamente affidabili anche a livello di regione. Potrebbe essere necessario fissare una numerosità campionaria minima necessaria da selezionare in ciascun strato, sovra campionando

7

da alcuni strati ogni qualvolta un criterio di allocazione puramente proporzionale non consenta il rispetto del vincolo indicato.

21. In quali casi può essere opportuno fare ricorso ad un campionamento casuale a grappoli

L'utilizzo di questo tipo di campionamento dipende dalle caratteristiche della popolazione o è imposto da esigenze operative

Ogni grappolo deve riprodurre il più possibile la variabilità rispetto alla caratteristica d'interesse che si ritrova nell'intera popolazione. Per definizione i grappoli, all'opposto degli strati devono risultare fortemente eterogenei al loro interno ed omogenei tra loro. Si procede alla selezione casuale di un numero di grappoli e si includono nel campione tutti gli elementi ad essi appartenenti. Uno dei vantaggi del campionamento a grappoli risiede nel fatto che i grappoli possono costituire una naturale aggregazione delle unità per le quali non si possiede una lista. Poi effettuare la rilevazione solo su alcuni grappolo è meno dispendioso rispetto al CCS. Ma dall'altra parte è meno efficiente: i raggruppamenti naturali tendono essere omogenei all'interno ed eterogenei all'esterno. (il contrario da quello che viene richiesto)

22. Che cosa si intende per campionamento a più stadi

Nella popolazione vengono individuati raggruppamenti di unità, dette unità di primo stadio e all'interno di questi raggruppamenti delle unità elementari chiamate unità di secondo stadio. Si seleziona casualmente un campione di unità di primo stadio e da tali unità si seleziona casualmente un campione di unità di secondo stadio

L'adozione di quest'asta tecnica consente di ridurre i costi dell'indagine, a sfavore c'è la complessità dei metodi di stima dei parametri e in alcuni casi dal fatto che non tutti gli stimatori godono delle proprietà desiderabili. Si adatta alle grandi rilevazioni.

23. Nell'ambito del processo di determinazione della dimensione campionaria nei piani di campionamento probabilistici, cosa significa indicare la precisione delle stime?

* Indicare la precisione delle stime vuole dire :

- fissare il margine di errore che si è disposti a tollerare (epsilon)
- fissare un grado di fiducia o di confidenza, ovvero la probabilità che la stima si collochi all'interno di un intervallo definito appunto dai margini di errore che si ritengono tollerabili (intervallo di confidenza)

24. Come si determina la dimensione campionaria nel campionamento casuale semplice per attributi? (stima di una proporzione o di una percentuale di popolazione)

Campionamento casuale semplice x attributi

le variabili che è necessario conoscere

- la dimensione della popolazione di riferimento N
- Θ cioè il margine di errore massimo ammesso
- il grado di fiducia α che accompagna il margine d'errore
- Proporzione: P potrebbe essere stimato in base a un campione ragionato di piccole dimensioni, che un procedimento che non eviterebbe però di pervenire a risultati distorti e poco affidabili. Una soluzione consiste nel sostituire al prodotto $P(1-P)$ il massimo valore che può assumere

$$n = \frac{z_{\alpha/2}^2 N}{\frac{(N-1)\theta^2}{P(1-P)} + z_{\alpha/2}^2}$$

$$n = \frac{z_{\alpha/2}^2 N}{4(N-1)\theta^2 + z_{\alpha/2}^2}$$

Si è fatta l'ipotesi peggiore che $P=0,5$

Campionamento casuale semplice x variabili

- La dimensione della popolazione N
 - Margine di errore massimo ammesso
 - Il grado di fiducia che accompagna il margine di errore
 - La varianza in popolazione: potrebbe essere stimata in modo preliminare avendo delle conoscenze sulla popolazione in questione
- Una terza soluzione può sfruttare

$$n = \frac{z_{\alpha/2}^2 N \sigma^2}{\frac{(N-1)\theta^2}{N} + z_{\alpha/2}^2 \sigma^2}$$

26. Quali sono gli aspetti principali che devono essere presi in considerazione ai fini della predisposizione di un questionario? ***

- Obiettivo del sondaggio. Scelta della popolazione obiettivo dei rispondenti;
- Modalità Contatto. Modalità di contatto dei destinatari del sondaggio (intervista personale, intervista telefonica, oppure auto compilazione di un questionario di solito accessibile via Internet);
- Contenuto. il contenuto dei quesiti (Condizione indispensabile è l'impiego di una terminologia chiara e facilmente comprensibile da parte di tutti i rispondenti);
- Wording. La terminologia da utilizzare nella formulazione dei quesiti a seconda della popolazione a cui ci stiamo riferendo;
- Modalità Risposte. La forma e la scala di misura delle risposte;
- Sequenza dei quesiti (solitamente si pongono all'inizio i quesiti di maggiore importanza e che richiedono più sforzo mentale, mentre alla fine vengono posti quesiti più complessi che possono causare turbamento emotivo)
- Layout. presentazione in termini grafici e editoriali (layout deve essere semplice ed intuitivo in modo che il questionario sia comprensibile per tutti);
- Pre-test e l'eventuale revisione.

27. Quali sono le principali tipologie di domande inseribili in un questionario.

Progettazione del questionario

Il questionario è un insieme di domande volte ad acquisire delle informazioni.

Bisogna fissare bene le informazioni che dovranno essere rilevate,

Formulazione delle domande

- ✖ Domande chiuse: costringono l'intervistato a scegliere tra modalità di risposta predefinite, facili da processare ma limitanti per l'espressione del pensiero dell'intervistato. Presentano il vantaggio dell'omogeneità delle risposte e comportano procedure di elaborazione più semplice. Dall'altra parte non sempre le alternative proposte includono al loro interno che quelle che corrispondono alla posizione personale dell'intervistato.
- ✖ Domande aperte: permettono all'intervistato di rispondere liberamente, richiedono lunghi tempi di elaborazione
- ✖ Domande filtro: domande chiuse che hanno la funzione di indirizzare verso una sessione o l'altra del questionario. Vengono utilizzate quando le unità di analisi non sono omogenee. Vengono, quindi, inserite perché si presuppone che non tutti gli individui del campione hanno le stesse abitudini. Solitamente hanno come risposta sì o no.
- ✖ Domande fattuali/ non fattuali: le domande fattuali vanno a ricostruire situazioni vissute in prima persona dall'intervistato. Le domande non fattuali di solito riguardano atteggiamenti, opinioni.

9

- ✗ Gerarchizzate: si chiede all'intervistato di mettere in ordine diverse alternative
- ✗ Retrospective: generalmente fattuali. Si va a sollecitare la memoria e si possono creare delle distorsioni dei dati. Effetto telescopio: gli intervistati tendono a sbagliare avvicinando nel tempo i fatti più che allontanarli.
- ✗ di controllo: sembrano duplicazioni in realtà sono volute per controllare la coerenza tra le informazioni ricevute.

28. Quali sono i requisiti delle diverse scale di misura con le quali possono essere espresse le modalità di risposta nel caso di domande chiuse?

Le scale di misurazioni

- **Caratteri quantitativi derivano da misurazione o operazioni di conteggio**
- **Caratteri qualitativi** sono assenti concetti di quantità

Una delle difficoltà maggiori è quella di proporre quesiti in modo che le risposte siano traducibili in misure. Molto complessa è la misurazione di atteggiamenti e opinioni, per queste misurazioni sono state ideate scale ad hoc.

Per i caratteri qualitativi

- Scale nominali: assegnano un codice convenzionalmente numerico ad un oggetto, al fine di identificare e classificare l'unità statistica messa sotto osservazione. Le modalità non possono essere messe in ordine secondo una sequenza logica, l'unico confronto statistico è l'uguaglianza o meno tra due unità rispetto a un carattere
- Scale ordinali: associano le risposte degli intervistati a categorie diverse, tra le quali esiste un ordinamento naturale: si può classificare una unità statistica come superiore o inferiore rispetto a un'altra. Si possono fare dei confronti applicando i concetti di maggiore o minore.

Per i caratteri quantitativi

- ◇ Scala ad intervalli: non prevedono lo zero assoluto che indichi l'assenza di quantità. Il confronto è ammissibile x differenza tra i valori assunti dal carattere. La costruzione della scala ad intervalli permette di usare statistiche di sintesi: medie, valori di variabilità, coefficienti di correlazione lineare, e tecniche di analisi multivariata

30 Cosa si intende per scala a differenziale semantico

Scala a differenziale semantico: è caratterizzata da coppie di aggettivi o proposizioni di significato contrapposto utilizzate come polarità opposte. Si utilizzano un numero dispari di modalità, in modo da utilizzare la modalità centrale come quella neutra. Se le modalità fossero pari si avrebbe uno spostamento o verso il positivo o verso il negativo

In questo modo si possono effettuare misure statistiche

1	2	3	4	5
Molto insoddisfatto	Abbastanza insoddisfatto	Nè soddisfatto Nè insoddisfatto	Abbastanza soddisfatto	Molto soddisfatto

29 cosa si intende per scala Likert

Scala di Likert: caso particolare della scala a differenziale semantico ma a intervalli. Consiste nel porre qualche affermazione alla quale l'intervistato deve esprimere l'intensità del consenso o meno attraverso l'uso di modalità distinte.

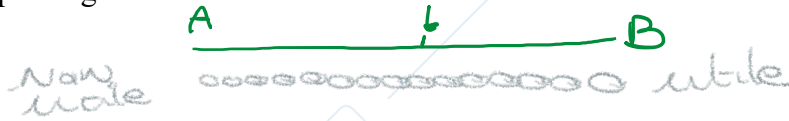
-2	-1	0	+1	+2
Del tutto disaccordo	in Abbastanza disaccordo	in Nè d'accordo Nè in disaccordo	Abbastanza d'accordo	Del tutto d'accordo

31. Quali sono possibili esempi di scale di valutazione a rapporti?

10

Scala a rapporti: assegna valori numerici alle risposte in modo tale che esista un valore zero, naturale e non arbitrario, e che i rapporti fra le posizioni nella scala abbiano un significato reale e non convenzionale. Il confronto statistico è ammissibile come rapporto tra valori assunti del carattere in due unità diverse

Segmento di retta/sequenza di pallini si supera l'idea di ancorarsi a un numero di posizioni all'interno della scala, si privilegia l'idea di continuo tra i due estremi



Scala con punto di ancoraggio: viene fissato in modo ragionato un punteggio a una marca si chiede all'intervistato di attribuire un punteggio di confronto basandosi sulla valutazione prefissata

Scala a somma costante: si inserisce un vincolo di somma, un totale da distribuire come punteggio alle varie alternative.

32. A quali regole deve sottostare l'ordine di presentazione delle domande in un questionario?

- Ordine delle domande. Secondo una scala decrescente di importanza, e riservare quindi quelle meno utili e meno rilevanti alla fine del questionario;
- Raggruppamento per argomento. Raggruppare le domande in funzione dell'argomento trattato e, nell'ambito di un determinato argomento, in funzione della loro struttura (chiuse o aperte), il che comporterà un minor sforzo cognitivo da parte del rispondente;
- fare seguire per ciascun argomento le domande che più probabilmente saranno soggette ad obiezioni e dubbi a quelle più innocue (prima domande semplici per mettere a suo agio l'intervistato, poi complesse o quelle che possono creare tensioni emotive);
- Garantire un senso di continuità al questionario. Ricorrendo ad opportune proposizioni di transizione, il cui scopo sarà proprio quello di guidare il rispondente da un argomento all'altro aiutandolo a stabilire un tono simile a quello di una conversazione.
- Creazione di un rapporto di fiducia con il rispondente, mettendo domande all'inizio che non causino tensione emotiva.

33. Che cosa si intende per pre-test di un questionario?

* Con pre-test si intende un questionario preliminare di verifica, il quale viene somministrato ad un campione (ristretto) di individui che sia il più vario possibile e quindi che rispecchi nel modo migliore la popolazione di riferimento. Lo scopo del pre-test è quello di verificare l'adeguatezza del questionario, valutandone i tempi di compilazione, se l'ordine di presentazione delle domande è corretto o confusionario e, soprattutto analizzando il tasso di partecipazione e completezza .

Crescendo la dimensione del campione cresce anche i costi dell'indagine.

È quindi importante fissare la numerosità del campione cercando di salvaguardare la massima precisione delle stime in modo compatibile con il limite di spesa prestabilito. Indicare la precisione delle stime vuol dire fissare il margine di errore che ai è disposti a tollerare in relazione con gli obiettivi dell'indagine e il grado di confidenza, ovvero la probabilità che la stima si collochi all'interno di un intervallo definito dai margini di errore che si ritengono tollerabili.

34. Che cosa si intende esattamente per errori non campionari?

Gli errori non campionari corrispondono alla differenza tra la media del campione effettivamente realizzato (c_y) e la media del campione originariamente programmato (cp_y).

Sono errori sistematici, riguardano problemi di corretta misurazione, solitamente sono:

- a) Unità statistiche mancanti. Problemi di mancata osservazione delle unità da contattare (errori di mancata risposta o di mancata copertura);
- b) Problemi di corretta misurazione (che dipendono da strumenti di rilevazione, tecniche di raccolta delle informazioni, intervistati ed intervistatori).

Tali errori, denominati non campionari si sommano all'errore campionario.

35. Che cosa si intende per Total Survey Error (TSE)?

L'errore complessivo (Total Survey Error) è la somma degli errori non campionari (sistematici) e degli errori campionari (dovuti alla scelta casuale degli elementi che compongono il campione). Se i primi non possono essere modificati scegliendo un campione più ampio, i secondi sì, in quanto aumentando le unità statistiche nel campione, riduco la casualità. L'errore complessivo può essere scomposto in :

- Errore campionario : distanza della media del campione programmato in origine dal valore medio incognito della popolazione
- Errore non campionario: corrispondenti alla differenza tra la media del campione effettivamente realizzato e la media del campione originariamente programmato

Total survey error

$$TSE = f(\bar{y} - \bar{y}_{CE}) = f(\bar{y} - \bar{y}_{CP}) + f(\bar{y}_{CP} - \bar{y}_{CE})$$

Dove

CP rappresenta il campione programmato

CE rappresenta il campione effettivamente osservato

L'indagine perfetta sarebbe quella affetta esclusivamente da errore campionario.

36. Da quali fattori dipende l'intensità dell'errore di mancata copertura?

Errori di mancata copertura. Sono errori dovuti dalla scelta di una lista non rappresentativa (che non copre) tutta la popolazione, può essere una lista incompleta, non aggiornata o contenente informazioni che si ripetono (ad esempio l'elenco telefonico può essere una lista incompleta).

I fattori che influenzano l'intensità dell'errore dovuto di mancata copertura sono:

- lista non completa;
- lista non aggiornata;
- lista non rappresentativa.
- lista contenere duplicazioni (nel caso non sia stata sottoposta a revisioni accurate e sistematiche). In sostanza la lista delle unità statistiche da cui viene estratto il campione è incompleta, non pulita (presenta elementi uguali più volte).

È dovuta all'incompletezza, alle duplicazioni, al fatto che il campione non è aggiornato. Dipende dall'oggetto dell'indagine e non dall'indagine in sé

N è la dimensione della popolazione di riferimento

N_{NC} È quella parte di popolazione non inclusa nel campione

Quindi si ha che

$$N = N_C + N_{NC}$$

L'interesse è sapere se in media per la variabile di interesse il fatto di guardare una lista che non contiene certe unità genera qualche tipo di differenza

La formula

$$\bar{y}_c - \bar{y} = \frac{N_{NC}}{N} (\bar{y}_C - \bar{y}_{NC})$$

Consente di qualificare consente di identificare da che cosa dipende maggiormente l'errore di mancata copertura

$\frac{N_{NC}}{N}$ Rappresenta il tasso di mancata copertura della lista, ovvero la proporzione di

$(\bar{y}_C - \bar{y}_{NC})$ rappresenta la differenza fra le medie riferite agli individui presenti e a quelli esclusi della lista. Se è pari a zero posso trascurare l'errore di mancata copertura, perché ho la sicurezza che non ci sia una differenza significativa fra le due parti di popolazione.

37. Quali metodi possono consentire di attenuare l'impatto dell'errore di mancata copertura sulle stime?

12

* 1. Un metodo che può essere utilizzato è la post stratificazione, che consente di riordinare gli elementi del campione in gruppi pesati proporzionali alla struttura della popolazione obiettivo.

2. Un'alternativa per affrontare il problema della mancata copertura, che ha riflessi anche sui costi di rilevazione, consiste nell'uso simultaneo di una doppia base di campionamento e di una duplice modalità di raccolta dei dati (dual frame mixed mode surveys)

Per ottenere da un campione informazioni di buona qualità è necessario che si adotti un procedimento probabilistico di selezione delle unità. Il rispetto di questa condizione esige ovviamente la disponibilità di una lista che consenta di stabilire una corrispondenza biunivoca perfetta fra gli elementi in essa contenuti e le unità della popolazione obiettivo, lista che non di rado risulta incompleta o contiene informazioni non aggiornate oppure duplicazioni, dando origine appunto al cosiddetto errore di mancata copertura.

38. Che cosa si intende per post-stratificazione?

È una tecnica statistica volta a migliorare la precisione delle stime, viene utilizzata nel momento in cui la stratificazione della popolazione non è stata possibile o non stata sufficiente; interviene sul campione suddividendo i dati campionari in base ad una variabile di interesse.

L'obiettivo è quello di rendere i gruppi omogenei al loro interno, mantenendo la composizione relativa della popolazione. Le stime vengono quindi pesate in base alla proporzione di ciascuno strato nella popolazione. La post-stratificazione ha come obiettivo la correzione delle distorsioni campionarie potenzialmente derivanti dalla mancata copertura di una lista.

Tecnicamente la poststratificazione non è altro che un'operazione di riclassificazione dei dati campionari, allo scopo di riproporcionarli alla struttura della popolazione obiettivo. In altre parole, ci si propone di adattare l'informazione campionaria già rilevata alla composizione (relativa) della popolazione rispetto alle modalità di una o più variabili (di post-stratificazione) connesse sia con quelle oggetto di interesse ai fini dell'indagine campionaria sia con la motivazione che determina l'incompletezza della lista in uso. Condizioni di applicabilità

a) sia noto il peso relativo di ciascun post-strato in popolazione;

b) i post-strati risultino il più possibile omogenei al loro interno ed eterogenei tra di loro, così da rendere minime le differenze, entro ciascuno di essi, fra i profili delle unità comprese e non comprese nella lista in uso.

39. Che cosa si intende per doppia base di campionamento e duplice modalità di raccolta dei dati? *

Per affrontare il problema della mancata copertura, che ha riflessi anche sui costi di rilevazione, consiste nell'utilizzo simultaneo di una doppia base di campionamento e di una duplice modalità di raccolta dati. Si ricorre a una lista anagrafica (lista elettorale) sia ad un elenco di telefonia, assieme alle relative modalità di contatto. Si combina la maggiore copertura della prima con i costi più contenuti tipici della selezione campionaria.

$$N = n_{a1} + n_{a2} + n_b \text{ Dove}$$

n_{a1} Corrisponde a unità interpellate direttamente dalla lista elettorale, che non sono coperte dalle liste elettorali

n_{a2} Corrisponde a unità campionarie partendo dalla lista elettorale che sono presenti anche nella lista telefonica

n_b Corrisponde alle unità presenti nella lista telefonica e intervistate telefonicamente

$$\bar{y}_{DB} = p\bar{y}_{a1} + (1 - p)[\gamma\bar{y}_{a2} + (1 - \gamma)\bar{y}_b]$$

Stimatore della media aritmetica

P è una stima iniziale della proporzione di unità statistiche
 γ è un parametro compreso tra 1 e 0 che definisce la proporzione di unità statistiche coperte dalla lista fissa ma selezionate dalla lista elettorale r intervistate direttamente

Lo stimatore è ottenuto come media ponderata dei valori medi campionari relativi ai due gruppi delle unità coperte e non coperte dalla lista telefonica.

Fra le situazioni di mancata copertura va considerata anche quella di persone o famiglie che pur disponendo di una linea telefonica fissa, non figurano nell'elenco sia per motivi di riservatezza, sia perché il loro numero

13

è associato ad un operatore telefonico diverso da quello scelto. Esistono alcune tecniche che consistono di ridurre l'effetto di questa mancata copertura:

- ✓ Generazione di numeri telefoni in modo casuale. È l'approccio più immediato ma non si utilizza per i troppi numeri non esistenti
- ✓ Utilizzare liste diverse da quelle tradizionali di elenchi telefonici.

40. Che cosa si intende per Random Digit Dialing (RDD)?

Con RDD, si fa riferimento ad una tecnica volta a ridurre un particolare effetto di mancata copertura (mancanza di unità statistiche nella lista, ad esempio le persone che per privacy decidono di non figurare nell'elenco telefonico). L'RDD consiste nella generazione pseudocasuale di numeri telefonici dall'insieme di tutti i numeri possibili, tenendo conto sia di quelli compresi negli elenchi sia di quelli esclusi. Generando così una lista di contatti, tra cui alcuni non presenti nell'elenco telefonico. In sintesi, permette di raggiungere le persone che non fanno parte dell'elenco telefonico, attraverso la generazione di numeri casuali.

41. A quali problemi danno origine le mancate risposte totali nel corso di un'indagine campionaria?

Le mancate risposte nell'indagine campionaria, ossia la mancata compilazione del questionario da parte di unità statistiche nella lista, causa un errore di tipo non campionario e quindi una distorsione delle stime. L'errore sarà dato dalla differenza tra la media del campione realizzato e la media del campione originariamente programmato. (generano quindi una discrepanza fra il campione programmato e quello realizzato che può produrre conseguenze importanti).

Errore di mancata risposta totale

Errori dovuti all'impossibilità di osservare una parte delle unità campionarie, sia per l'impossibilità di contattare alcune unità selezionate sia a causa del rifiuto di parte di queste a partecipare all'indagine,

La potenziale distorsione delle stime è una delle conseguenze più gravi delle mancate risposte:

- ❖ Se il meccanismo che genera le mancate risposte non è casuale, le statistiche campionarie basate sui soli dati dei rispondenti risulteranno distorte
- ❖ La riduzione della numerosità campionaria rispetto a quella prevista determina come conseguenza immediata un incremento della varianza degli stimatori

$$N = N_R + N_{RC}$$

$$\bar{y}_R - \bar{y} = \frac{N_{NR}}{N} (\bar{y}_R - \bar{y}_{NR})$$

L'intensità dell'errore di mancata risposta è esprimibile come funzione del tasso di mancata risposta e della differenza di comportamento in media rispetto alla variabile di interesse fra le due sottopopolazioni dei rispondenti e non rispondenti.

Sul tasso di mancato contatto si può pensare di intervenire

- Sul numero di tentativi di intervista
- Sul momento della giornata in cui viene effettuata l'intervista: la sera è più produttivo mentre non ci sono differenze significative per i giorni della settimana

Un altro tipo di mancata risposta è il rifiuto: può dipendere dal momento in cui viene proposta l'intervista, da come si pone l'intervistatore.

42. Da quali elementi dipende l'intensità dell'errore di mancata risposta totale?

L'intensità dell'errore di mancata risposta totale dipenderà dai modi per attenuarlo a priori e posteriori:

- Tentativo con un'altra tecnica. selezionato un sub-campione di non rispondenti, si ritenta un'intervista mediante un'altra tecnica;
- Utilizzare indagini precedenti. Utilizzazione di informazioni di indagini precedenti, condotte sulla stessa popolazione;
- Usi di intervistati simili. dall'utilizzazione di dati di unità simili che hanno collaborato.

43. Quali sono i principali interventi preliminari utili a prevenire le mancate risposte?

Per quanto concerne il tasso di mancato contatto, sono soprattutto tre le variabili ad esso correlate sulle quali si può pensare di intervenire:

- N. Tentativi di contatto. Agire sul numero di tentativi di contatto;

14

- Cambiare orario di contatto. Agire sul momento della giornata dei tentativi di contatto: ore serali più produttive (Determinante è il momento in cui viene effettuato il contatto, perché non basta sapere che la persona da intervistare è reperibile per assicurarsene la collaborazione. Una persona può, infatti, non essere disponibile per un'intervista, e ciò avviene generalmente durante le ore dedicate al riposo) Tentativi successivi di intervista per convincere coloro che rifiutano hanno difficilmente successo, al contrario di quanto accade per i mancati contatti. L'esperienza dimostra che far cambiare idea a chi rifiuta fin dall'inizio un'intervista è un evento possibile, ma assai poco probabile.
- Analizzare il numero di squilli (contatto telefonico) prima della risposta: creazione di algoritmi di telefonata automatici.

44. A quali limiti si espone il metodo della sostituzione delle unità statistiche non rispondenti con unità di riserva?

1. Similarità a chi ha già risposto (campione poco eterogeneo, con unità simili causa sottostime). Le unità aggiuntive potrebbero essere maggiormente simili a chi ha risposto ed è già presente nel campione, piuttosto che a chi non ha risposto. Questa evenienza determinerebbe di fatto una compressione artificiale della variabilità delle stime, pur a fronte dell'aumento dei costi di rilevazione dovuto alla gestione delle sostituzioni

2. Diminuzione dell'impegno negli intervistatori. Può causare una diminuzione dell'impegno da parte dell'intervistatore, il quale sa che saranno predisposte unità statistiche di riserva. Generando in definitiva un aumento delle mancate risposte, con conseguente perdita di rappresentatività statistica del campione effettivamente realizzato.

45. A quali limiti si espone il metodo del sovra campionamento?

Il metodo del sovra campionamento viene utilizzato per ovviare al problema della mancanza di risposte ai questionari, in questo modo il numero delle unità contattate sarà superiore fin dall'inizio alla dimensione campionaria preventivata. I limiti del metodo rimangono gli stessi dell'utilizzo di unità di riserva:

1. problemi di autoselezione (le unità statistiche aggiuntive non vengono scelte in modo casuale);
2. perdita di rappresentatività (compressione artificiale delle stime).

46. Che cosa si intende per metodo della domanda cruciale?

Si tratta di un metodo più rapido ed economico per ovviare al problema del rifiuto dell'intervista personale o telefonica (CAPI o CATI). E consisterebbe nell'affrettarsi a porre una sola domanda in sostituzione sia pure parziale o imperfetta del nucleo centrale del questionario, durante i contatti preliminari e nell'imminenza dell'interruzione degli stessi. Questa consentirebbe il confronto delle risposte date alla domanda cruciale dai rispondenti all'intero questionario e dai soli rispondenti a questa domanda, essa deve essere formulata in circostanze il più possibile simili.

47. Quali interventi correttivi ex post possono essere attuati per attenuare l'impatto delle mancate risposte totali sulle stime?

1. Procedure di riponderazione, ad esempio la post-stratificazione (riproporzionare l'informazione campionaria, nel tentativo di attenuare il potenziale impatto distorsivo sulle stime provocato dalle mancate risposte di alcune delle unità selezionate originariamente e non sostituite dalle corrispondenti unità di riserva);
2. Sovracampionamento (consiste nel prelevare un numero di unità statistiche maggiori rispetto alla numerosità campionaria preposta dall'indagine)
3. Procedure di estrapolazione. In pratica si potrà ipotizzare che coloro che partecipano solo dopo k solleciti siano, al crescere di k , sempre più simili nel loro comportamento di risposta ai non rispondenti. Dunque, un'extrapolazione degli eventuali trend temporali emersi nei valori assunti da ciascuna variabile d'interesse dovrebbe riflettere in maniera appropriata il comportamento dei non rispondenti.

Problematica degli interventi ex post : si fondano implicitamente sull'ipotesi, debole e comunque non verificata, che gli intervistati possano utilmente rappresentare gli assenti dalla rilevazione, a parità di certe caratteristiche correlate con le variabili di interesse.

48. Quali sono le principali tecniche di imputazione a fronte di mancate risposte parziali? L'integrazione dei dati mancanti avverrà grazie all'applicazione di metodi di imputazione, a partire da un'attenta analisi delle informazioni contenute nei questionari completi e nelle parti compilate dei questionari incompleti. Le principali tecniche di imputazione di mancate risposte parziali sono tre, note come:

- criteri da donatore. Hot deck : una volta che ad una unità, la quale presentava omissione di risposta, è stata donata la risposta, anche essa potrà donare la risposta; Coldk deck: quando una unità non presenta omissioni di risposta può donare le sue risposte alle unità simili. Si basa sull'ipotesi che una risposta mancante possa essere stimata attraverso la risposta data da un'unità statistica simile a quella che non ha risposto. (utilizzato per domande associate alla stima di parametri di variabili quantitative o qualitative)
- imputazione da modello. Utilizza la regressione lineare (esclusivamente per stima di caratteri quantitativi). Una volta suddivisi i rispondenti per classi di imputazione (chi non ha risposte mancanti "cold dekc" e chi ha risposte mancanti), viene utilizzato un modello che pone la variabile dipendente (domanda mancante) in funzione di variabili esplicative note moltiplicate per coefficienti considerati "invarianti", per cui prendendo i coefficienti stimati e attraverso procedure algebriche si otterrà un valore per la variabile dipendente relativa alla risposta mancante
- imputazione stocastica (esclusivamente per stima di caratteri quantitativi). Consiste nel modificare il valore ottenuto per l'imputazione attraverso uno dei due criteri precedenti aggiungendovi (o sottraendovi) un valore aleatorio. In questo modo si riesce ad ottenere una diminuzione della compressione artificiale.

Post stratificazione

Operazione di riclassificazione dei dati del campione che sono stati raccolti attraverso l'utilizzo di pesi w . La post stratificazioni

La post stratificazione non è altro che una operazione di riclassificazione dei dati campionari allo scopo di di riproporcionarli alla struttura della popolazione obiettivo.

Ci si propone di adottare l'informazione campionaria già rilevata alla composizione della popolazione rispetto alle modalità di una o più variabili, dette di post stratificazione, sia connesse con quelle oggetto di interesse ai fini dell'indagine campionaria sia con la motivazione che determina l'incompletezza della lista in uso.

Per poter applicare questa strategia è necessario che:

- sia noto il peso relativo di ciascun post-strato in popolazione
- I post strati risultino il più possibili omogenei all'interno ed eterogenei tra di loro così da rendere minime le differenze entro ciascuno di essi Sia H il numero dei post strati nei quali è suddivisa la popolazione

$$n_1, n_2, \dots, n_H \quad \text{dove } \sum_n^H n_h = n \text{ numerosità di ogni sottogruppo}$$

$$\bar{y}_1, \bar{y}_2, \dots, \bar{y}_H \quad \text{dove } \bar{y} = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ik} \text{ valori medi calcolato all'interno di ciascun sottogruppo del campione}$$

$$\bar{y}_{PS} = \sum_{h=1}^H w_h \bar{y}_h \quad w_h = \frac{N_h}{N} \quad \forall h$$

Si tratta di uno stimatore corretto ed efficace

I pesi di popolazione w_h rappresentano la consistenza relativa dei diversi gruppi per il criterio di stratificazione. Dovrebbero tendere ad attenuare l'effetto di sbilanciamento, a patto che i valori medi rappresentino in maniera adeguata la situazione che caratterizza la situazione di ciascuna stratificazione

49. A quali cause principali possono essere ricondotti gli errori di misura?

Le cause di errori di misura sono attribuibili :

- al questionario,
- al rispondente,
- al procedimento di raccolta delle informazioni,
- all'intervistatore.

50. Come possono essere classificate le diverse componenti di costo di un'indagine campionaria? a) Costi fissi. corrispondenti a quelli che devono essere comunque sempre sostenuti, indipendentemente dalla dimensione campionaria e qualunque sia il tipo di campionamento adottato (probabilistico o meno). Si tratta

16

di costi di progettazione dell'indagine, di reperimento delle informazioni preliminari necessarie (ad esempio 102 liste o database contenenti informazioni sulla popolazione di riferimento), di predisposizione del questionario e di redazione del rapporto finale;

b) Costi specifici. Nel caso di campioni probabilistici, costi specifici dipendenti dalla strategia di campionamento prescelta (ad esempio quelli relativi alla preparazione delle liste delle unità elementari per ciascun strato in caso di stratificazione, oppure quelli connessi alla messa a punto degli elenchi delle unità elementari per i diversi stadi di rilevazione, qualora il campione sia a due o più stadi di selezione);

c) Costi variabili. Dipendono essenzialmente dalla strategia di campionamento adottata, dalla dimensione del campione prefissata di conseguenza e dalla modalità di rilevazione dei dati prescelta (ad esempio costi legati ai materiali di consumo, ai servizi utilizzati, al lavoro degli intervistatori e all'analisi dei dati).

51. Nella stesura del report di presentazione dei risultati di un'indagine campionaria quali informazioni dovrebbero sempre essere indicate per poterne valutare l'attendibilità?

1. gli obiettivi principali dell'indagine ed il nome del committente, specificando se si tratta di un privato o di un ente pubblico, un media di informazione, un organismo politico, ecc.;
2. chi ha eseguito l'indagine (rilevazione dei dati, elaborazioni, ecc.), ossia il nome della società di ricerca, agenzia di informazione o altro;
3. la popolazione di riferimento, ossia il collettivo rispetto a cui si vogliono generalizzare i risultati dell'indagine campionaria;
4. quando si utilizzano campioni non casuali questo deve essere espressamente indicato in sede di pubblicazione dei risultati con la dicitura «campione non casuale»;
5. la numerosità campionaria, ossia il numero di persone intervistate, la loro età ed il sesso, e con quale procedura sono state selezionate per costituire il campione statistico;
6. il numero di persone che non hanno risposto all'intervista, distinte per età e sesso;
7. il numero di mancate risposte ad ogni domanda;
8. il periodo dell'anno in cui sono state fatte le interviste;
9. il metodo dell'intervista, cioè se personale, telefonica, con questionario auto compilato o di altro tipo;
10. nelle tabelle di presentazione dei risultati le domande del questionario a cui le stesse si riferiscono; 11. per i risultati campionari di maggiore importanza rispetto agli obiettivi dell'indagine i margini di errore e i livelli di confidenza dei parametri statistici (proporzioni, percentuali, valori medi, ecc.)