Indice

Prefazione

La scienza è fatta di dati come una casa è fatta di pietre.
Ma un ammasso di dati non è scienza più di quanto
un mucchio di pietre sia una vera casa.
Henri Poincarè

L'insieme degli argomenti trattati nel presente volume si propone come un essenziale ed agevole manuale operativo di *metodologia e tecnica della ricerca sociale*, affrontando con metodi assai pragmatici e concreti l'insieme degli argomenti di una attività di ricerca essenzialmente multidisciplinare.

È difficile coniugare gli aspetti metodologici caratteristici, diffusi, anche se condivisi con qualche distinguo, delle scienze della ricerca sociale, e gli strumenti che queste mettono a disposizione per la raccolta e l'elaborazione esplorativa e confermativa delle ipotesi formulate.

Gli argomenti attinenti le varie discipline vengono affrontati con un approccio semplificato, sufficiente a rendere autonomo il lettore, rinviando gli approfondimenti alla copiosa letteratura di settore, in parte citata in bibliografia.

Nel filo logico in cui la ricerca si esprime, si inseriscono ubiquitarie le nuove tecnologie, a prevalente derivazione informatica, sia hardware che software, con strumenti sempre più evoluti, che talvolta sopravvanzano lo sviluppo di un modello teorico di riferimento, introducendo implicazioni metodologiche non ancora comprese, o non comprese appieno.

Il ricercatore che ha maturato una qualche esperienza riesce a distinguere, a grandi linee, aspetti positivi e negativi delle nuove tecnologie, cosa che non riesce altrettanto bene a chi si avvicina solo marginalmente, o saltuariamente, a questi strumenti di ricerca sociale.

Il fervore riscontrato negli ultimi anni in questo settore della ricerca deve fare ora i conti sia con i rischi caratteristici di sempre, sia con gli errori introdotti da un uso superficiale dei nuovi prodotti, soprattutto quando questi promuovono indagini standardizzate, pronte per l'uso, con modelli o *template* preconfezionati.

L'esempio tipico del primo caso è l'indagine di opinione frettolosa, come ad esempio quella del giornalista che scende in strada ed intervista qualche passante, estrapolando dalle risposte ottenute l'opinione dell'intera popolazione in merito all'argomento del giorno.

Il secondo caso nasce paradossalmente dalla qualità e dalla quantità di strumenti software indirizzati specificatamente alla rilevazione ed alla analisi dei dati di indagine.

Proprio l'elevata qualità introdotta da parte di diversi applicativi, induce nel ricercatore una falsa sicurezza, portandolo a sviluppare l'attività di ricerca verso i paradigmi implicitamente inclusi nello strumento, senza porre accorta attenzione preliminare al disegno d'indagine ed alle diverse possibilità e metodi da utilizzare, con particolare riferimento agli obiettivi individuati nel momento della formulazione delle ipotesi iniziali.

Partendo da queste premesse, il presente lavoro vuole percorrere idealmente tutta la prassi oramai consolidata, almeno nelle sue linee più generali, di un percorso di ricerca sociale, con una particolare attenzione verso un suo specifico sottoinsieme, l'ambito socio – sanitario.

Idealmente il percorso si sviluppa attraverso le seguenti fasi:

– Scelta del problema e identificazione delle ipotesi d'indagine.
– Formulazione di un disegno della ricerca.
– Raccolta, codifica e validazione dei dati.
– Analisi esplorativa dei dati.
– Interpretazione dei risultati e analisi confermativa delle ipotesi iniziali.

È del tutto evidente che il compito che ci si prefigge è estremamente ampio, coinvolgendo nel suo percorso diverse discipline di derivazione sociologica, epidemiologica, informatica e statistica.

Per questo motivo gli argomenti vengono trattati ad un livello introduttivo, identificando come principale obiettivo il ruolo, non banale, di strumento multidisciplinare per la corretta implementazione ed attuazione di una indagine di ricerca.

I metodi di ricerca trattati sono essenzialmente quantitativi, ma non per questo vengono trascurati alcuni strumenti di tipo qualitativo che da sempre sono elementi fondanti della analisi sociologica, e che qui vengono visti come utili strumenti preliminari ed integrativi nella definizione degli obiettivi di indagine, dei temi da approfondire e delle modalità più idonee per farlo.

Tra questi vengono accennati:

– Intervista discorsiva.
– Osservazione e osservazione partecipata.
– Focus Group.
– Intervista con esperti, il metodo Delphi.

Grande risalto viene dato al momento della raccolta del dato, sia questo proveniente da fonti ufficiali o di tipo documentale, o costruito attraverso tecniche di intervista e di questionario.

Vengono inoltre definite caratteristiche, criticità ed opportunità delle varie modalità con cui viene somministrato il questionario:

– Indagine postale o email.
– Indagine faccia a faccia.
– Indagine via web.
– Indagine telefonica.
– Indagine con questionario autocompilato.

In tutte queste modalità di somministrazione del questionario esiste, più o meno diffusa, una versione *Computer Assisted*, ovvero con una integrazione di strumenti hardware e software che vanno dalla semplice opportunità di raccolta diretta del dato, alla somministrazione di questionari strutturati e personalizzati a seconda dell'utente-paziente-cliente a cui ci si rivolge.

Il mondo anglosassone, da sempre noto per la sua pragmaticità, non si smentisce neanche in questo settore, ed ha coniato per i diversi tipi di indagine i seguenti acronimi:

– CAPI *Computer Assisted Personal Interview.*
– PAPI *Paper Assisted Personal Interview.*
– CATI *Computer Assisted Telephone Interview.*
– CAWI *Computer Assisted Web Interview.*
– CASI *Computer Assisted Self Interview.*

In queste tipologie di indagine il questionario assume un ruolo determinante, non solo come strumento operativo per rilevare sul campione il valore o la modalità assunte dalle diverse variabili, ovvero le risposte registrate alle domande che vengono poste, ma come vero e proprio *framework* ove interpretare concettualmente le ipotesi iniziali del progetto di ricerca.

Nella redazione del questionario si deve tenere conto di una notevole quantità di elementi, come la sequenza delle domande, il modo di chiedere informazioni, la cautela nelle domande che causano imbarazzo, la congruenza semantica nelle possibilità di risposta, la unidimensionalità delle richieste, la comprensibilità del linguaggio utilizzato, l'aspetto grafico, ed altro ancora.

A tutti questi primari aspetti atti a tradurre nel concreto e nel migliore dei modi il disegno della ricerca, si aggiunge l'esigenza di particolari strutture di questionario che sono richieste a seconda degli strumenti di analisi statistica preventivamente ipotizzati.

In ambito epidemiologico, ad esempio, vi deve essere una domanda a risposta dicotomica, o dicotomizzabile, che permetta di distinguere tra casi e controlli, o tra esposti e non esposti, o tra malati e non malati, a seconda del tipo di analisi del rischio.

Oppure nella definizione di variabili latenti comuni a determinati tipologie di utenti-pazienti, nella ricerca di atteggiamenti attraverso la misura di opinioni espresse, si devono costruire delle domande multiple atte a rilevare delle scale, come ad esempio le scale di Likert, o le scale di Guttman o, delle scale aventi

una accezione probabilistica, come quelle identificate dal modello di Rasch e di Bradley-Terry.

La raccolta delle informazioni viene seguita da una fase di validazione e trasferimento su supporto informatico, operazione che richiedono una certa accortezza ed una puntigliosa coerenza se si vuole pervenire ad una *Matrice di dati* di buona qualità e di facile fruibilità.

Fanno parte di questo momento l'analisi della coerenza dei dati raccolti, la ricerca di dati palesementi errati, ed il difficile passo ove vengono considerati i dati mancanti, compresa una loro eventuale imputazione.

L'accuratezza dei dati vorrebbe che, in presenza di una mancata risposta venga rimossa l'intera unità statistica, ma questo pesa inevitabilmente quando si hanno numerosità non molto elevate, o quando la rilevazione ha richiesto molte risorse, o quando si raccolgono dati su casi rari, come capita sovente in ambito sanitario in riferimento alle patologie poco frequenti nella popolazione.

Si preferisce talvolta imputare manualmente il dato mancante, sostituendolo con un valore neutro come potrebbe essere la mediana per dati almeno su scala ordinale, purchè i valori mancanti siano pochi punti percentuali rispetto al totale.

Nei casi di variabili qualitative si deve costruire una distribuzione di frequenza e valutare opportunamente caso per caso assegnando, ad esempio, la modalità più frequente, oppure alternando le due modalità più frequenti.

Il trasferimento su supporto informatico è quasi sempre necessario, anche per indagini svolte con interviste assistite dal computer, dato che si utilizzano sovente strutture di dati complesse, soprattutto in presenza di questionari strutturati.

La tecnologia ha fatto passi avanti notevoli nel settore dei sistemi informativi ma, in generale, i dati raccolti vengono trasferiti in un foglio elettronico per rilevazioni sino a circa 100 variabili e sino a circa 300 interviste, oltre si preferisce usufruire delle innegabili opportunità offerte da un gestore di base di dati, tipicamente una base di dati relazionale, utilizzando un *Relational Data Base Managment System (RDBMS).*

La fase di trasferimento delle informazioni è anche il momento più opportuno per definire il *dizionario dei dati (data dictionary)* o *libro dei codici (codebook)*, documento che raccoglie le definizioni dei nomi mnemonici dati alle variabili, come traduzione delle forme discorsive che queste assumono all'interno del questionario e le relative codifiche assegnate ad ognuna delle modalità possibili per la singola variabile.

Nel dizionario dei dati vengono annotate anche tutte le variazioni e le osservazioni di un qualche rilievo verificatesi nella somministrazione del questionario, e le ricodifiche o imputazioni di dati mancanti intervenute successivamente.

La successiva analisi dei dati può essere idealmente suddivisa in due momenti distinti, una prima fase comune di *analisi esplorativa*, valida per ogni tipo

di indagine, che utilizza strumenti statistici di base, ove si fanno classificazioni univariate del campione indagato, come distribuzioni per età, o territorio di provenienza, ed altro, utilizzando strumenti di statistica descrittiva, integrati da grafici ad istogramma, o di altro tipo.

Il secondo passo è la ricerca di associazione e di dipendenza tra due variabili, confrontando le variabili oggettive raccolte come età, genere, titolo di studio, ecc., con la o le variabili obiettivo dell'indagine che meglio identificano le ipotesi iniziali.

Lo strumento principale qui è la tabella bivariata, ove si confrontano le distribuzioni di frequenza assolute e relative percentuali per variabili categoriali o continue ridotte in classi, sino a giungere eventualmente alla sintesi massima delle tabelle tetracoriche.

Le ipotesi associative che sorgono vengono usualmente verificate, sia nella intensità puntuale, che nella significatività statistica attraverso il test chi – quadro e l'analisi del rischio utilizzando l'odds ratio quando risulti opportuno.

Anche in questo caso sono utili all'interpretazione i grafici ad istogrammi appaiati, oppure anche una verifica distributiva con boxplot sempre appaiati.

Il secondo momento di analisi dei dati, *l'analisi confermativa delle ipotesi*, si esplica attraverso una serie di test statistici, oppure attraverso la costruzione di veri e propri modelli statistici multivariati, i quali sono diversi da caso a caso, applicati a seconda del contesto, e sono qui solamente accennati.

L'aspetto operativo viene approfondito attraverso esempi pratici, tratti da lavori di indagine effettivamente svolti, secondo le diverse tipologie ed i diversi modelli proposti.

L'archiviazione dei dati e le elaborazioni più semplici vengono svolte utilizzando Calc di OpenOffice[1] o Excel di Microsoft Office, mentre le elaborazioni statistiche più complesse vengono attuate attraverso il software R[2].

[1] http://www.openoffice.org; http://it.openoffice.org
[2] http://www.r-project.org