

La progettazione delle schede

Ulisse Di Corpo

I dati sono la materia prima delle analisi statistiche, ma non tutti i dati possono essere oggetto di elaborazione statistica; è infatti necessario che siano codificati e che siano stati raccolti in modo sistematico:

- *Codifica dei dati*: solo le informazioni quantitative e le informazioni che possono essere tradotte in numeri possono essere oggetto di analisi statistica. Tutte le altre informazioni non possono essere oggetto di analisi statistica.
- *Raccolta sistematica dei dati*: le analisi statistiche richiedono variabili che siano state rilevate in modo sistematico. Nella fase di progettazione dello strumento di rilevazione dei dati è quindi necessario valutare se tutte le variabili possano essere rilevate sistematicamente, ed eventualmente intervenire laddove vi siano problemi modificando il contenuto delle variabili, attivando risorse in grado di rilevare l'informazione o eliminando le informazioni che non possono essere raccolte sistematicamente.

Codifica dei dati

La caratteristica chiave che differenzia un archivio classico da un archivio statistico è la codifica dei dati. Non si usano più sigle alfabetiche, ma codici numerici che fanno riferimento a liste. E' importante prevedere la codifica dei dati già in fase di progettazione dello strumento di rilevazione dei dati. Ad esempio:

Scheda di un archivio non codificato

Nome e Cognome:

Indirizzo:

Comune:

Stato Civile: _____

Istruzione: _____

Scheda di un archivio codificato

Nome e Cognome:

Indirizzo:

Comune:

viene tradotto dal programma nel codice Istat

Stato civile:

1. *celibe/nubile*
2. *coniugato/a*
3. *separato/a*
4. *vedovo/a*

Istruzione:

1. *Analfabeta*
2. *Licenza Elementare*
3. *Licenza Media Inferiore*
4. *Licenza Media superiore*

In tutti i campi, al di fuori del Nome, Cognome e Indirizzo che non possono essere codificati, si ricorre al numero di codifica, ad esempio 2 a Coniugato, 4 a Licenza Media Superiore. Ogni campo ha una sua lista di modalità. Quando si utilizzano archivi cartacei la differenza sembra marginale, invece, quando si utilizzano archivi informatizzati i vantaggi sono immediati:

1. *Dimensioni*: si riduce al minimo lo spazio necessario alla memorizzazione delle schede, infatti mentre un numero (da 0 a 255) occupa un byte di memoria, una sigla alfabetica occupa tanti byte quanti sono i caratteri che la compongono, ad esempio CON (Coniugato) occupa 3 byte o tanti byte quanti sono gli spazi riservati per tale campo.
2. *Riduzione degli errori*: si riduce notevolmente la possibilità di errore. Infatti il programma accetta solo le risposte, le modalità, presenti nelle liste, non è possibile inserire una modalità non prevista o un codice al di fuori di quelli previsti;
3. *Facilità d'uso*: è possibile inserire le condizioni di ricerca in modo più semplice e immediato, semplicemente attivando i campi e le modalità che interessano direttamente sulla scheda e sulla lista delle modalità;
4. *Facilità di elaborazione*: quando si effettua una selezione si ottengono tutti i conteggi (cioè tutte le distribuzioni di frequenza). Ad esempio se si selezionano i coniugati che abbiano conseguito il diploma di media superiore, basta attivare la modalità coniugati, la modalità Diploma di Scuola Media Superiore ed effettuare una sola volta la lettura delle schede. Il programma fornirà automaticamente i conteggi su tutte le altre variabili.
5. *Elaborazioni territoriali*: le distribuzioni territoriali sono possibili in quanto ad ogni comune è associato un codice (numero d'ordine), il programma andrà a sommare le schede trovate per quel determinato comune nella sua posizione d'ordine.
6. *Tutte le informazioni a portata di mano*: lo stesso avviene per qualsiasi altra variabile, al termine di una selezione di schede, si ottengono automaticamente tutte le distribuzioni semplici su tutte le variabili presenti nella scheda. Nel caso di archivi tradizionali o di schede non codificate, è possibile effettuare solo un conteggio per volta in quanto non essendo presente la posizione d'ordine delle modalità il programma non sa dove andare a sommare le schede trovate. Inoltre, se l'informazione non è stata codificata può essere a volte inserita in modi diversi, ad esempio è possibile che l'informazione *Coniugato* sia stata registrata come CON, CONIUG, CON., MARITATA, ecc...rendendo quindi impossibile il conteggio dei coniugati.
7. *Ricerca scientifica*: oltre ad effettuare conteggi sulle modalità gli archivi statistici consentono di effettuare qualsiasi altra elaborazione statistica, ad esempio: tabelle doppie, tabelle delle connessioni, analisi fattoriale e individuazione di tipologie e di gruppi omogenei.

Come scegliere i campi da inserire in una scheda

La progettazione di uno strumento di rilevazione dei dati è il momento più delicato di tutto il lavoro. Infatti se si raccolgono informazioni irrilevanti o non si dà spazio ad informazioni che potrebbero invece essere importanti, si produce un lavoro sterile, poco interessante, poco utile e di scarso valore. Prima di avviare la progettazione dello strumento di rilevazione dei dati è perciò bene tener presente che:

1. E' necessario distinguere tra variabili chiave, variabili esplicative e variabili di struttura:
 1. le *variabili chiave* sono quelle che descrivono l'argomento trattato, ad esempio se la scheda è relativa ai tumori, le variabili chiave saranno quelle relative al tipo e progressione del tumore;
 2. le *variabili esplicative* sono tutte quelle variabili che si pensa possano essere collegate alle variabili chiave e che si vogliono porre in relazione con queste, ad esempio per i tumori potrebbero essere variabili che descrivono il tipo di alimentazione, situazioni di stress, l'ambiente in cui si vive;
 3. le *variabili di struttura*, invece, sono l'età, il sesso, l'istruzione, la professione, la provenienza, il reddito; variabili che in genere vengono utilizzate per descrivere la popolazione utilizzata.
2. Nella prima fase di progettazione della scheda è necessario farsi guidare e suggerire da coloro che hanno una buona *esperienza* nel settore, ad esempio esperti, testimoni privilegiati o ricorrendo a documentazione di vario tipo.
3. E' sempre necessario tenere presente che una scheda è utile e ricca di informazioni nel momento in cui *confronta ipotesi di diversa provenienza*. La raccolta dei dati non deve essere finalizzata a confermare le proprie ipotesi, altrimenti sarebbe del tutto inutile. L'interazione dei dati e la progettazione della scheda è un percorso di crescita, di cambiamento in cui si deve sempre essere disposti a rivisitare ed abbandonare le proprie certezze.

Per mettere a punto una scheda:

1. si inizia esplicitando gli obiettivi (variabili chiave);
2. si continua esplicitando tutte quelle variabili che potrebbero essere collegate all'argomento trattato (variabili esplicative). E' molto importante scrivere quale informazione dimostra che cosa, in modo da facilitare, in seguito, l'analisi dei dati. Se si lascia tutto nel vago, quando si leggeranno i risultati non si potrà dire se le ipotesi iniziali siano state o meno confermate e si rischia di soffermare l'attenzione su elementi secondari e irrilevanti, producendo così un lavoro noioso, arido e di scarso valore pratico e scientifico;
3. si termina mettendo a punto lo strumento di rilevazione dei dati, ad esempio testandolo per vedere come funziona, se incontra resistenze, se è possibile rilevare tutta l'informazione prevista. Questa fase di test deve essere ripetuta finché lo strumento di rilevazione non raggiunge quelle caratteristiche di efficienza e di qualità che sono poi alla base della qualità dei dati;
4. quando i dati raccolti sono relativi all'informazione gestita all'interno di un'organizzazione o di un servizio, diventa necessario coinvolgere le diverse figure professionali e i livelli centrali e periferici, per assicurarsi che lo strumento risponda alle varie esigenze e che non crei tensioni che ne minino il funzionamento.

Variabili oggettive e soggettive

Mettiamo il caso che l'interesse sia quello di studiare che cosa sia collegato ai vissuti di depressione, per capire come organizzare le attività di prevenzione. Si tratta di un campo molto vasto e complesso nel quale si confrontano impostazioni e teorie diverse. Si inizia perciò studiando queste diverse impostazioni e si cercano testimoni privilegiati che possano aiutare ad allargare il ventaglio di ipotesi che saranno inserite nello strumento di rilevazione dei dati. Maggiore è il numero di ipotesi a cui lo strumento di rilevazione dei dati risponde, maggiore sarà la ricchezza dei risultati. Messe assieme tutte queste ipotesi si inizia a tradurle in uno strumento che consenta di rilevare l'informazione.

E' noto che quando si effettuano lavori di ricerca sociale, spesso le persone rispondono alle domande mascherandosi. Anche se la persona si sente insoddisfatta, sola, depressa, probabilmente cercherà di dare un'immagine di se ben diversa, un'immagine di persona realizzata e soddisfatta. Molti ricercatori si sono fatti bloccare dal problema della *maschera*. E' importante fare chiarezza a questo riguardo in quanto le tecniche utilizzate da Sintropia-DS si applicano bene anche su informazioni soggettive (mascherate), ad esempio è possibile chiedere direttamente: *Ti senti solo?*, *Ti senti depresso?*, *Ti senti angosciato?* La maschera impedisce di utilizzare le risposte a queste domande per effettuare stime quantitative, ma è quasi insignificante nel momento in cui si vogliono individuare strutture, relazioni tra eventi. Infatti, la maschera non si applica su una sola risposta, ma è coerente su tutte. Se una persona afferma di non essere depressa quando nei fatti è depressa, affermerà anche, di non essere angosciata, quando nei fatti è angosciata. La relazione tra depressione e angoscia rimane invariata.

Ad esempio, mettiamo il caso in cui le persone intervistate siano 200, e a tutte sia stato chiesto se si sentivano depresse e se si sentivano angosciate. Le persone hanno risposto in modo prevalente di non sentirsi depresse e di non sentirsi angosciate quindi angoscia e depressione risultano correlate (concomitanza delle risposte No). Correggendo l'effetto della maschera utilizzando un test psicologico e non una domanda diretta, si sarebbe ottenuta una

Risposte mascherate			
	Depressi	Non depressi	Totale
Angosciati	15	3	18
Non Angosciati	2	180	182
Totale	17	183	200

Risposte depurate dall'effetto maschera			
	Depressi	Non depressi	Totale
Angosciati	158	10	168
Non Angosciati	2	30	32
Totale	160	40	200

La relazione emerge in entrambe le tabelle

prevalenza delle risposte *Sì* ad angosciato e a depresso. Cioè la situazione effettiva è quella di persone prevalentemente depresse e angosciate. Poiché la relazione si studia come presenza di concomitanza la depressione e l'angoscia risultano ancora fortemente connesse (concomitanza delle risposte *Sì*). L'esempio precedente mostra che se esiste relazione tra le variabili, questa relazione emerge anche quando le risposte vengono mascherate. Questo è un punto fondamentale in quanto il problema della maschera è un problema onnipresente quando si lavora in ambito sociale. Il problema della maschera limita la possibilità di realizzare studi quantitativi e analisi delle differenze (t di student e analisi della varianza) non la possibilità di individuare concomitanze, relazioni e strutture fattoriali.

Ridondanza dell'informazione

Sintropia-DS consente di lavorare con fenomeni complessi dove la stessa informazione può avere più significati. Ad esempio è possibile avere il dubbio se sia meglio chiedere *ti senti solo?* o *sto spesso da solo*, o ... E' sempre bene prevedere una certa ridondanza nelle informazioni. Se l'argomento è la solitudine è utile inserire nel questionario/scheda più domande relative alla solitudine, non una sola.

Univocità dell'informazione

E' importante, inoltre, che ogni domanda chieda un'unica informazione. Sarà poi l'analisi dei dati a dire se le informazioni sono legate, collegate tra di loro. Ad esempio è errato formulare una domanda del tipo: *La solitudine ti provoca angoscia?* Infatti viene inserita nella domanda la relazione tra solitudine e angoscia. L'analisi dei dati non potrà quindi dire se questa relazione esiste effettivamente. Per studiare la relazione tra angoscia e solitudine è necessario formulare due domande separate: *Ti senti solo? Ti senti angosciato?* Oltre ad utilizzare un unico concetto per domanda e oltre ad evitare domande nelle quali si ipotizza un'eventuale relazione è importante che l'informazione che si chiede sia chiara. Chiedere ad una persona in modo indiretto se si sente sola, può non dare nessuna informazione.

Univocità della variabile

E' importante che ad un campo (ad una domanda) corrisponda un'unica variabile. Ad esempio una domanda così formulata è errata:

Il padre ha fruito del congedo parentale?

1. Sì
2. No - Non ne ha fruito
3. No - Non ne ha diritto
4. Famiglia monoparentale solo padre
5. Famiglia monoparentale solo madre

Si tratta infatti di due variabili e sono quindi necessarie due domande e due campi di risposta. La versione corretta è perciò la seguente:

Il padre ha fruito del congedo parentale?

1. Sì
2. No - Non ne ha fruito
3. No - Non ne ha diritto

Tipologia della famiglia:

1. Famiglia monoparentale solo padre
2. Famiglia monoparentale solo madre

E' fondamentale tenere distinte le variabili. Ad ogni variabile deve coincidere una domanda e almeno un campo di inserimento dati.

Raccolta dei dati ed organizzazione: i sistemi informativi

Una azienda, una organizzazione, un servizio sociale, un servizio sanitario, presentano sempre momenti in cui le informazioni vengono raccolte, ad esempio quando si effettua un colloquio di ingresso, un trattamento. Esplicitare l'informazione che viene raccolta in una organizzazione significa progettare una scheda. Ad esempio una scheda di primo ingresso, una scheda sul trattamento, una scheda sull'utente, una scheda sulle famiglie. Elaborare l'informazione delle schede significa progettare gli interventi, valutarli, definire i ruoli all'interno del servizio: chi prende quale informazione, come circola l'informazione, chi è responsabile di quale informazione.

Esplicitare una scheda è un momento spesso divertente, un gioco, un gioco che può toccare profondamente l'organizzazione. Un gioco che a volte può scatenare dinamiche forti. Esistono infatti sovrapposizioni tra ruoli: il medico a volte non riconosce il ruolo dello psicologo o viceversa, l'infermiere effettua colloqui che dovrebbero essere effettuati dall'assistente sociale, e così via. Lavorare sulla scheda consente di esplicitare le competenze, i confini di ogni operatore, l'informazione che deve circolare tra gli operatori e quella che è propria di ogni figura professionale. In altre parole la scheda indica le regole del gioco, cioè l'organizzazione.

Meglio si esplicitano le schede e migliore sarà l'organizzazione e la qualità del servizio, l'efficienza, l'efficacia e la valorizzazione delle risorse. In altre parole, lavorare sulla scheda significa lavorare sull'organizzazione. Se il lavoro di progettazione delle schede non viene svolto, manca un tassello fondamentale di tutta l'organizzazione e l'intero sistema tende a bloccarsi e a lavorare solo secondo una logica delle emergenze. Un'organizzazione funziona bene nel momento in cui ci si mette d'accordo sulle regole del gioco. Le schede sono le regole del gioco, definiscono gli spazi di ciascuno.

La progettazione delle schede è un momento molto delicato. Infatti se le informazioni non vengono inserite nella scheda non potranno di certo essere raccolte ed analizzate. L'esito dell'intero lavoro dipende dalla progettazione della scheda. Inoltre, se la scheda tratta il flusso dell'informazione interno ad un servizio o ad un'organizzazione, la progettazione della scheda implicherà la riprogettazione del servizio e dell'organizzazione. In altre parole la progettazione della scheda comporta un lavoro delicato di analisi organizzativa.

Archivi relazionali e archivi statistici

Perché i dati possano essere elaborati è necessario che questi siano organizzati in un archivio statistico. Un archivio non è altro che una sequenza di schede (cartacee o informatizzate) sulle quali sono presenti gli stessi campi.

Gli archivi relazionali consentono di archiviare e recuperare l'informazione e di effettuare semplici conteggi, ma non si prestano bene per le analisi statistiche. La differenza fondamentale tra archivi relazionali e archivi statistici sta nel fatto che mentre gli archivi relazionali scompongono il record in tabelle e sotto archivi, fisicamente distinti e collegati tra di loro solo tramite dei puntatori, gli archivi statistici mantengono (nei limiti del possibile) l'unitarietà degli archivi (dei record).

Esistono vari software di archiviazione dei dati, ma il mercato attualmente offre, quasi unicamente, programmi per l'archiviazione relazionale dei dati. Si pone quindi spesso un problema di dialogo tra archivi relazionali e programmi di analisi statistica dei dati. Sintropia-DS offre procedure di acquisizione dei dati da archivi esterni e organizza i dati acquisiti in modo da ottimizzare le analisi statistiche dei dati.