

La metodologia relazionale

Ulisse Di Corpo e Antonella Vannini¹

La scoperta degli attrattori ha portato ad allargare l'indagine scientifica ai fenomeni non causabili, fenomeni che possiamo osservare, ma che non possiamo riprodurre in laboratorio. Finora si era ammesso, senza nemmeno postularlo chiaramente, che tutti i fenomeni fossero causabili e quindi studiabili per mezzo del metodo sperimentale; oggi sappiamo che accanto ai fenomeni causabili (entropici) esistono anche fenomeni non causabili (sintropici).

Attualmente si conoscono due metodologie che consentono di studiare relazioni:

1. la metodologia delle differenze, da cui nasce il metodo sperimentale e che consente di studiare unicamente relazioni causali (entropiche).
2. la metodologia delle variazioni concomitanti, inizialmente proposta da Stuart Mill nel 1843, da cui nasce il metodo relazionale e che consente di studiare qualsiasi tipo di relazione, causale e non causale, e che allarga la scienza anche agli aspetti sintropici della realtà.

La metodologia relazionale è stata finora poco utilizzata poiché:

- si pensava che le qualità proprie dei sistemi viventi implicassero l'esistenza di entità vitalistiche o metafisiche non indagabili con il metodo scientifico;
- mancava il potere di calcolo necessario per utilizzarla.

¹ syntropy@gmail.com

1. La metodologia relazionale

Per comprendere la differenza tra metodologia delle differenze e metodologia delle variazioni concomitanti, iniziamo dalla descrizione del metodo sperimentale applicato allo studio dei sistemi viventi.

Il metodo sperimentale studia le differenze tra gruppi, il procedimento è semplice:

1. si formano due gruppi simili;
2. si varia solo un elemento (trattamento);
3. le differenze che si osservano tra i due gruppi possono essere imputate unicamente al trattamento.

Ad esempio, per studiare gli effetti di un farmaco si utilizzano due gruppi simili (stessi valori medi, stessa varianza), si controllano tutte le variabili di disturbo e si somministra il farmaco solo ad un gruppo. Le differenze che si osservano tra i due gruppi possono essere attribuite solo al farmaco: si produce in questo modo conoscenza di tipo causa-effetto (la causa è il farmaco, gli effetti sono le differenze tra i gruppi). Il metodo sperimentale è stato utilizzato con successo in fisica e chimica, ma in medicina, biologia e psicologia ha portato a produrre una serie infinita di relazioni causa-effetto scollegate tra loro. L'esempio della farmacologia è lampante: una quantità enorme di relazioni causa-effetto, prive però di una visione d'insieme.

La metodologia relazionale invece di studiare le differenze studia le concomitanze (per esempio presenza-presenza, assenza-assenza). Come si procede? Basta progettare una scheda di rilevazione, un questionario, una griglia di osservazione, rilevare i dati secondo alcuni semplici criteri ed analizzare i risultati.

Ad esempio, nella tabella riportata di seguito si studia la relazione tra le variabili sesso e *incidenti stradali*. Si vede che il 20% degli uomini fa pochi incidenti contro il 70% delle donne, l'80% degli uomini fa molti incidenti contro il 30% delle donne:

Incidenti	Maschi	Femmine	Totale
Pochi	50 20%	105 70%	155 39%
Molti	200 80%	45 30%	245 61%
Totale	250 100%	150 100%	400 100%

Volendo interpretare questa relazione (*maschi-incidenti*) si potrebbero avanzare le ipotesi più disparate se solo non fosse possibile controllare la relazione inserendo una terza variabile. Ad esempio, nel momento in cui la popolazione viene divisa in due gruppi, quello delle persone che hanno percorso un basso numero di chilometri durante l'anno e quello di coloro che hanno invece percorso molti chilometri, si trova che la relazione tra sesso e incidenti si annulla:

Incidenti	pochi km percorso		Molti km percorsi	
	Maschi	Femmine	Maschi	Femmine
Pochi	70%	70%	20%	20%
Molti	30%	30%	80%	80%
Totale	100%	100%	100%	100%

Nell'esempio si vede che non vi è differenza nella quantità di incidenti tra maschi e femmine nel momento in cui si effettua il controllo con la variabile *numero di km percorsi*. Ciò significa che la relazione *sesso-incidenti*, osservata nella prima tabella, non è diretta, ma passa attraverso una terza variabile, *km percorsi* (i maschi percorrono più chilometri delle femmine), detta anche variabile interveniente. Si costruisce così la catena:

Sesso -> Km percorsi -> Incidenti

dove il numero di incidenti è spiegato dal numero di chilometri percorsi e non dal sesso.

Allo stesso modo si possono studiare gli ambiti entro i quali una relazione è valida. Poniamo il caso di aver trovato una relazione tra tipo di terapia e guarigione. Se, nel momento in cui si controlla questa relazione con la variabile *sesso* si osserva che la

terapia risulta molto efficace per le femmine e poco efficace per i maschi si può affermare che la relazione *terapia-guarigione* è specificata dalla variabile sesso e che non può essere generalizzata a tutta la popolazione.

Da questi esempi si evince che la metodologia relazionale permette di:

- studiare molte informazioni assieme e di utilizzare informazioni di tipo quantitativo e qualitativo, oggettive e soggettive;
- grazie ai controlli a posteriori, è possibile produrre informazione differenziata in grado di restituire la complessità dei fenomeni naturali;
- non richiede condizioni controllate a priori, né gruppi omogenei e/o randomizzati;
- permette di paragonare gruppi provenienti da realtà diverse e di attuare controlli avvalendosi di informazioni prese precedentemente (per esempio vecchie cartelle cliniche o altre rilevazioni).

Queste caratteristiche della metodologia relazionale permettono di lavorare direttamente sul campo, senza intervenire artificiosamente dall'esterno. L'attività di ricerca diventa semplice, accessibile a tutti, economica, e adatta ad affrontare qualsiasi tipo di problematica.

Un breve esempio in campo medico

In uno studio realizzato per la clinica pediatrica dell'Università di Roma sono state confrontate le terapie utilizzate nella cura delle trombocitopenie in età pediatrica utilizzando la metodologia relazionale. La Itp (trombocitopenia idiopatica o autoimmune) è una malattia abbastanza frequente nella sua forma acuta, ma fortunatamente molto rara nella sua forma cronica. Basti pensare che nella Clinica pediatrica dell'Università di Roma si sono registrati 15 casi di Itp cronica in dodici anni. Il numero esiguo di tali pazienti fa immediatamente capire quanto sia difficile un qualsiasi studio randomizzato o con gruppi di controllo. I pazienti, infatti, sono bambini diversi per età di insorgenza e per sesso, non riconoscibili prima che siano trascorsi almeno sei mesi dall'esordio della sintomatologia e che quindi possono aver intrapreso le più disparate terapie. A causa dei problemi di

campionamento dei bambini trombocitopenici è stata richiesta un'analisi multivariata (basata sulla metodologia relazionale). I risultati di questa analisi hanno evidenziato che: la terapia comunemente adottata si è rivelata, a lungo termine, meno efficace delle altre; il piano terapeutico utilizzato, per un disguido tecnico, in un ospedale del Nord Italia è risultato il più efficace; le risposte ai farmaci si sono differenziate in base al sesso e all'età. E' stata effettuata, parallelamente, anche un'elaborazione statistica con le metodiche classiche (analisi delle differenze, analisi della varianza). In questo caso è emersa solo la differenza per età e sesso; il resto delle informazioni si è persa.

Alcune considerazioni

Mentre, la metodologia delle differenze si avvale di test statistici parametrici, come il *t di Student* e l'*ANOVA* o analisi della varianza che richiedono condizioni particolari quali la distribuzione normale della popolazione - curva gaussiana - la randomizzazione dei gruppi, il controllo delle variabili di disturbo e misure effettuate almeno su scale a intervalli, condizioni che sono quasi impossibili da soddisfare nella ricerca applicata e nel campo clinico, la metodologia relazionale si avvale di test statistici non parametrici che permettono di analizzare qualsiasi tipo di informazione. Il *t di Student* e l'*ANOVA* possono essere utilizzate correttamente solo con gruppi randomizzati e controllando le fonti di disturbo, in caso contrario vengono evidenziate solo le differenze più marcate. In altre parole, si perde informazione e una tale perdita può condurre facilmente ad errori di interpretazione. Un'interpretazione restrittiva, dovuta a una carenza di informazioni, non permette un reale confronto delle diverse ipotesi e porta a mantenere ipotesi parziali che si contraddicono, senza possibilità di verifica.

L'esempio riportato mostra come le tecniche delle differenze, quando vengono utilizzate con gruppi non randomizzati e non omogenei (ciò è inevitabile nella ricerca applicata) evidenziano solo relazioni molto marcate e probabilmente scontate, per cui sebbene sia minima la probabilità di vedere qualcosa che non c'è si rischia di non vedere nulla e di perdere, così, il criterio fondamentale per la piena verifica delle ipotesi: la ripetibilità dei risultati. Nel campo scientifico non si arriva a conclusioni in base ad un singolo studio: le conoscenze sono scientifiche quando sono riproducibili. Solo le informazioni che si

ripresentano in più studi possono darci garanzie reali, poiché quelle dovute al caso hanno una scarsissima se non nulla probabilità di ripresentarsi. Con la metodologia classica è spesso impossibile verificare la riproducibilità dei risultati; è infatti impossibile confrontare dati che non si sono potuti ottenere.

2. Tecniche statistiche

Sembra una contraddizione, ma proprio utilizzando il livello più semplice di misura, cioè le variabili dicotomiche, si riesce a ricostruire la massima complessità. Un po' come capita con l'informatica. L'informatica si basa sulla traduzione di tutta l'informazione nel bit, il sì/no; dal bit si ricostruisce qualsiasi tipo di complessità. Tutti i tentativi di costruzione di computer analogici (basati su valori quantitativi) sono falliti. Lo stesso accade per l'analisi dei dati: perché si possa arrivare ad analisi veramente potenti è necessario tradurre tutta l'informazione in variabili dicotomiche, in bit di informazione.

Le variabili dicotomiche (*si/no*, *vero/falso*, *presente/assente*) sono il livello più semplice di misura in quanto presentano solo due alternative. Tutte le informazioni, anche quelle quantitative, possono essere tradotte in una o più variabili dicotomiche. Trasformando le variabili in variabili dicotomiche diventa possibile far interagire qualsiasi tipo di informazione: informazioni quantitative e qualitative, variabili nominali, ordinali, a intervalli e a rapporti.

E' diffuso il pregiudizio che il trattamento dell'informazione in base a variabili dicotomiche porti a perdere informazione, in quanto le variabili dicotomiche sono il livello più basso di misurazione. Ma, come vedremo, questo pregiudizio (in alcuni ambienti molto diffuso) è errato, infatti la gestione dell'informazione in base a variabili dicotomiche permette di specificare esattamente le ipotesi di lavoro, porta a risultati precisi e non ambigui e aumenta considerevolmente la qualità e la quantità dell'informazione che può essere trattata. La metodologia relazionale traduce tutte le informazioni in variabili dicotomiche e questo è forse l'elemento che più amplifica il potere informativo di questa metodologia rendendola incredibilmente potente e versatile.

Come si traducono le informazioni in variabili dicotomiche.

Le informazioni vengono raccolte sottoforma di variabili:

1. nominali (anche note come variabili qualitative sconnesse)
2. ordinali (anche note come variabili qualitative ordinali)
3. ad intervalli (anche note come variabili quantitative discrete)
4. a rapporti (anche note come variabili quantitative continue)

Queste 4 categorie di variabili si distinguono per le operazioni matematiche, sempre più complesse, che possono essere utilizzate:

1. nel caso di variabili *nominali* il valore è una modalità, ad esempio la provincia, il colore degli occhi, lo stato civile. Con le variabili nominali è possibile effettuare solamente conteggi (contare il numero di risposte per ogni singola modalità);
2. le variabili *ordinali* si differenziano dalle variabili nominali in quanto, oltre a consentire l'analisi dei conteggi (delle frequenze), consentono di effettuare analisi basate sull'ordine delle modalità. Ad esempio, con la variabile ordinale *istruzione* oltre a contare quante sono le persone per ogni modalità, (elementare, media, superiore e universitaria), è possibile contare qual è il livello di istruzione raggiunto da almeno (ad esempio) il 50% della popolazione, ecc... In altre parole esiste un ordine, una progressione. Si può utilizzare quest'ordine per creare nuove categorie (ad esempio: istruzione bassa e istruzione elevata) o per ordinare la popolazione (prima i laureati, poi i diplomati, quindi i possessori di licenza media, ecc...);
3. le variabili ad *intervalli* si differenziano dalle variabili ordinali in quanto, oltre a consentire l'analisi dei conteggi e l'ordinamento della popolazione, introducono il concetto di distanza: tra ogni modalità contigua esiste la stessa distanza. Ad esempio, la distanza tra 1 grado e 2 gradi è la stessa distanza che si trova tra 2 gradi e 3 gradi. Le osservazioni a intervalli possono essere analizzate oltre che con conteggi e ordinamenti, anche con operazioni quali l'addizione, la sottrazione ed è possibile calcolare i valori medi;
4. le variabili a *rapporti* si differenziano da quelle a intervalli per la coincidenza dello zero con lo zero assoluto. Ad esempio non è possibile dire che 2 gradi centigradi

siano il doppio di 1 grado centigrado. Infatti 0 gradi centigradi non corrispondono allo zero assoluto, ma alla temperatura in cui l'acqua si trasforma in ghiaccio. Lo zero assoluto è pari a -273 gradi centigradi (0 gradi Celsius). Quindi 1 grado centigrado nei fatti corrisponde a 274 gradi Celsius, 2 gradi corrisponde a 275 gradi Celsius: 275 gradi Celsius non è il doppio di 274 gradi.

La traduzione di queste variabili in variabili dicotomiche segue alcuni accorgimenti:

- nel caso di variabili nominali, la singola modalità (ad esempio la singola provincia, il singolo colore) può diventare una variabile dicotomica. Ad esempio la regione *Lazio* diventa la variabile dicotomica *Lazio* per la quale le risposte possono assumere solo i valori si/no. Durante la fase di progettazione della scheda con la quale si rilevano i dati è utile evitare di essere troppo analitici e puntuali. Ad esempio, se si effettua una ricerca raccogliendo 200 questionari e questa ricerca è distribuita su tutto il territorio nazionale, non ha senso chiedere il comune di residenza. Infatti, essendo i comuni italiani più di 8.000, la frequenza di risposta per i singoli comuni sarà eccessivamente bassa. E' invece necessario individuare che cosa interessa o è possibile sapere della variabile territorio, ad esempio Nord/Centro/Sud o dimensione del comune, ecc... ed inserire nel questionario direttamente queste domande. Risulta utile raccogliere l'informazione nello stesso modo in cui servirà per le elaborazioni. Le ricodifiche dei dati, la loro trasformazione, sono processi che possono risultare troppo onerosi e che possono portare al mancato utilizzo di variabili raccolte in forma troppo dettagliata.
- Le variabili ordinali sono tutte quelle in cui le modalità di risposta seguono una progressione, ad esempio il titolo di studio o variabili del tipo mai, poco, abbastanza, molto. Queste variabili possono essere trattate allo stesso modo delle variabili nominali traducendo ogni singola modalità di risposta in una variabile dicotomica. Ma, è anche possibile tradurre l'informazione nella forma alto/basso, poco/molto, ad esempio titolo di studio alto/basso. E' importante sottolineare che non esiste un criterio oggettivo per delimitare l'alto dal basso. In ogni studio le modalità che separeranno l'alto dal basso possono variare. Ad esempio se si effettua uno studio sui professori universitari il titolo di studio basso potrà essere la laurea, mentre il titolo di studio alto le specializzazioni. Se si effettua uno studio tra

la popolazione povera di un paese del terzo mondo, la separazione tra alto e basso corrisponderà a scolarizzato e non scolarizzato. Questo esempio mostra che la divisione di una variabile ordinale in una variabile dicotomica alto/basso, deve sempre tener conto del contesto e delle finalità dello studio. Nel caso in cui le informazioni che si chiedono sono tali da non consentire una suddivisione ragionata tra alto e basso, si suggerisce di equiripartire la popolazione tra le modalità alto/basso. Ad esempio, se si deve dicotomizzare la scala *mai, poco, abbastanza, molto*, il punto di separazione tra l'alto e il basso sarà quel punto che consente il massimizzare il bilanciamento dell'alto con il basso (valore mediano).

- Per le variabili continue (variabili ad intervalli e variabili a rapporti) si utilizzano in genere dei valori di taglio (cut-off points) che delimitano il passaggio da valori bassi a valori alti (da valori normali a valori patologici) o delle classi, degli intervalli. E' importante sottolineare che per queste variabili il taglio tra valori bassi e valori alti non segue necessariamente il criterio dell'equiripartizione. Infatti, queste variabili hanno un valore quantitativo preciso ed in genere a queste variabili si associano valori critici che indicano il passaggio da uno stato all'altro (ad esempio il passaggio dalla salute alla patologia). L'obiettivo del ricercatore e lo scopo dell'analisi dei dati è in genere quello di individuare questi valori critici. Capita quindi frequentemente che una stessa variabile continua venga tradotta in più variabili dicotomiche al fine di individuare quale valore di taglio consenta meglio di altri di individuare il passaggio da uno stato all'altro.

L'elaborazione di variabili dicotomiche.

L'operazione base che si applica alle variabili dicotomiche è il conteggio, il calcolo delle frequenze. Sulle frequenze si può utilizzare la somma, la divisione, la sottrazione e la moltiplicazione. In questo modo dalle frequenze si passa ad analisi più complesse, ad esempio l'analisi dei dati in percentuale, tabelle doppie e confronto di variabili, analisi delle connessioni e analisi fattoriali:

1. **distribuzioni semplici:** sono il livello più basso dell'analisi dei dati e corrispondono al conteggio delle risposte e al loro confronto grazie ai valori percentuali. Le

- distribuzioni semplici si utilizzano per fotografare la distribuzione dei dati;
2. **tabelle doppie:** sono il risultato dell'incrocio delle distribuzioni di due variabili e consentono di studiare le relazioni tra le modalità delle due variabili, grazie allo studio delle percentuali di riga e di colonne;
 3. **matrice delle connessioni:** è possibile calcolare per ogni coppia di variabili dicotomiche indici di connessione, ad esempio il chi quadrato e l'r-phi. L'analisi di questi valori consente di individuare le relazioni significative, evitando così di dover leggere tutte le tabelle doppie. In questo modo si riduce la quantità di materiale da analizzare: infatti le tabelle di connessione occupano pochissime pagine, e possono essere ordinate da quelle più significative a quelle meno significative in modo da essere guidati nella lettura;
 4. **analisi fattoriale:** le connessioni indicano la presenza di relazioni tra variabili prese a coppie. In uno studio medio le relazioni possono essere molte, basti solo dire che con 30 variabili dicotomiche le relazioni sono $30 \times 30 / 2 = 450$, con 100 variabili dicotomiche si arriva a 5.000. In uno studio medio che tratta varie migliaia di variabili dicotomiche le connessioni significative sono sicuramente nell'ordine di svariate migliaia se non di milioni. E' quindi chiaro che le connessioni, pur essendo molto più sintetiche, immediate e pratiche delle tabelle doppie possono risultare eccessivamente analitiche. Questo problema si risolve facendo analizzare la matrice delle connessioni direttamente dal programma grazie alla tecnica dell'analisi fattoriale. Questa tecnica individua le strutture (grazie alla loro rappresentazione su di uno spazio multidimensionale) e consente di ridurre l'intero problema ad un numero limitato di strutture, cioè di fattori.
 5. **analisi dei gruppi:** i profili fattoriali consentono di attribuire alle schede valori di similarità con ciascun profilo. In questo modo è possibile dividere le schede originarie in base all'appartenenza ai profili, ed è anche possibile studiare la sovrapposizione tra profili (schede che partecipano a più profili).

- distribuzioni semplici

Età	Soggetti
13 anni	2
14 anni	56
15 anni	161
16 anni	183
17 anni	194
18 anni	134
19 anni	72

Le distribuzioni semplici consentono di fotografare la distribuzione dei dati. Ogni riga della tabella corrisponde ad una variabile dicotomica.

Ad esempio la variabile dicotomica *13 anni* può assumere i valori si/no, lo stesso per la variabile dicotomica *14 anni*,...

Le distribuzioni semplici vengono in genere utilizzate per descrivere il campione: età, sesso, istruzione. Quando il campione è rappresentativo, le percentuali possono essere generalizzate a tutta la popolazione. Ad esempio, per stimare le dimensioni di un determinato fenomeno, per stimare le abitudini, le scelte, le opinioni.

- tabelle doppie

Incidenti	Maschi	Femmine	Totale
Pochi	50	105	155
	20%	70%	39%
Molti	200	45	245
	80%	30%	61%
Totale	250	150	400
	100%	100%	100%

Sono il risultato dell'incrocio di due variabili e consentono di studiare le relazioni tra le modalità delle due variabili, grazie allo studio delle percentuali di riga e di colonna. In questo esempio il 39% dei 400 soggetti intervistati ha

fatto pochi incidenti, mentre il 61% ne ha fatti molti (colonna a destra *Totale*). In assenza di relazione tra Sesso e Incidenti si osserverebbero le stesse percentuali della colonna "totale" nelle colonne *maschi* e *femmine*. Nell'esempio si legge, invece, che il 20% dei maschi ha fatto pochi incidenti, contro il 39% del totale (valore atteso in assenza di relazione) e il 70% delle femmine. Allo stesso modo l'80% dei maschi ha fatto molti incidenti contro il 30% delle femmine. E' quindi possibile affermare che esiste una relazione tra sesso e incidenti poiché i maschi fanno più incidenti delle femmine e le percentuali di risposta dei maschi e delle femmine si discostano sensibilmente dalle percentuali attese in assenza di relazione. Maggiore è lo scarto tra percentuali osservate e percentuali attese, maggiore è la forza della relazione. Per studiare la forza delle relazioni e la loro significatività statistica si possono utilizzare il test del chi quadrato, l'r phi, l'f di Fischer, o altri test statistici.

- *connessioni*

Molti test statistici consentono di studiare le relazioni tra variabili dicotomiche. Il test del Chi Quadrato (χ^2) è forse quello maggiormente utilizzato. Il χ^2 calcola la differenza tra le frequenze osservate e le frequenze attese. In assenza di relazione il χ^2 è uguale a 0, nel caso di relazione massima è uguale al numero di casi conteggiati. Il χ^2 consente anche di confrontare il risultato con distribuzioni probabilistiche note. In questo modo è possibile conoscere la significatività statistica della relazione. La significatività statistica indica la probabilità di errore che si accetta nel momento in cui si afferma l'esistenza della relazione. Convenzionalmente vengono considerate relazioni con probabilità di errore inferiore all'1%. Quando si incrociano variabili dicotomiche il χ^2 è significativo all'1% con valori uguali o superiori a 6,635.

La relazione può essere di due tipi: diretta o inversa. Se la relazione è diretta al presentarsi della prima variabile dicotomica si presenta anche la seconda, se la relazione è inversa al presentarsi della prima variabile dicotomica NON si presenta la seconda. Le relazioni inverse sono riportate con il segno negativo (-) mentre le relazioni dirette sono riportate senza segno.

Poiché il valore massimo del χ^2 varia a seconda del numero di casi conteggiati, risulta utile standardizzare i valori facendoli variare tra 0 e 1. Questa trasformazione è conosciuta come test r-Phi e si ottiene come radice quadrata del valore del χ^2 diviso per n (numero di casi conteggiati). Al valore così ottenuto si aggiunge il segno "-" nel caso di relazione inversa. Quando l'r-Phi viene utilizzato su variabili dicotomiche ottenute da variabili quantitative (tagliate utilizzando il valore mediano) si osserva che l'andamento di questo indice è simile all'indice di correlazione classico (r di Pearson).

Nella tabella riportata di seguito si nota, innanzitutto, che la variabile *depresso* è connessa perfettamente con se stessa (r-Phi pari ad 1,000, χ^2 pari al numero di schede conteggiate e frequenza osservata pari al 100%, contro il 53,29% di frequenza attesa), si tratta di un'informazione banale in quanto una variabile è sempre connessa perfettamente con se stessa.

Esempio di connessioni ottenute dalla variabile dicotomica *depresso* e le altre variabili di un questionario sul disagio giovanile, in uno studio effettuato su 974 studenti di scuole medie superiori nella provincia di Teramo
(elaborazione effettuata utilizzando Sintropia-DS)

Mi sento depresso:

ChiQ	rPhi	%Si	%No	
974.00	1.000	(100.00%/	0.00%)	Mi sento depresso
507.08	0.722	(85.42%/	12.96%)	Mi sento angosciato
229.19	0.485	(69.37%/	20.60%)	Mi sento inutile
209.18	0.463	(78.04%/	31.94%)	Mi sento solo
189.70	0.441	(72.14%/	27.78%)	Ho poca fiducia nella vita
189.15	0.441	(65.50%/	21.30%)	Scontento
179.18	0.429	(71.59%/	28.47%)	Insoddisfatto
173.57	0.422	(71.59%/	29.17%)	Infelice
169.88	0.418	(71.59%/	29.63%)	Ho poca fiducia in me stesso
.....				

Nell'esempio le connessioni vengono proposte organizzate dalle più forti per scendere progressivamente e bloccarsi quando la probabilità di errore supera l'1% di significatività statistica. Tutte le connessioni possono essere ritrovate sottoforma di tabelle doppie. E' però molto più immediato leggere una tabella delle connessioni che non alcune centinaia di pagine di tabelle doppie.

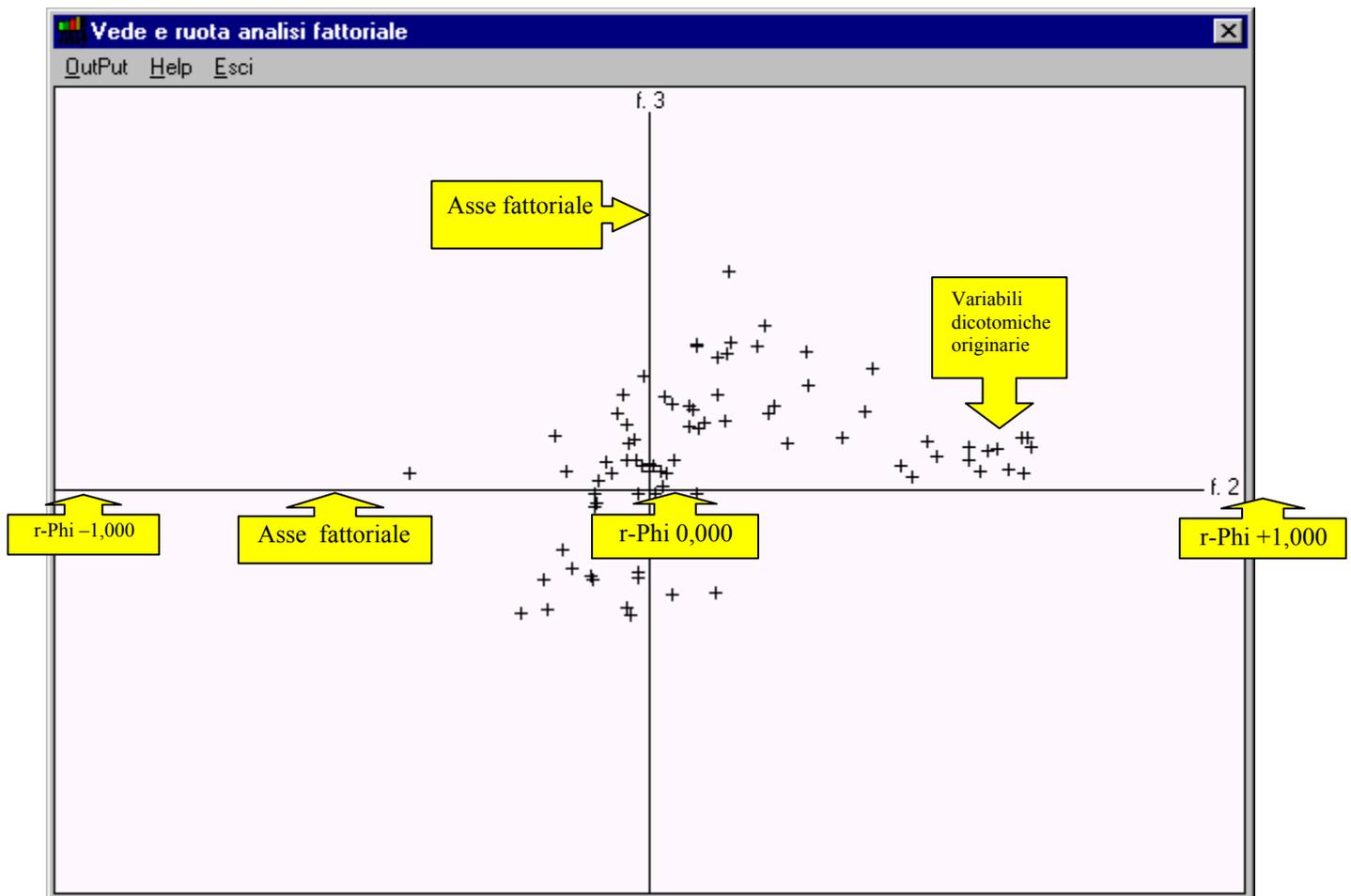
L'esempio mostra che la variabile dicotomica *Mi sento depresso* è fortemente connessa con *Mi sento angosciato*, *Mi sento inutile*, *Mi sento solo*, *Ho poca fiducia nella vita*, *Mi sento scontento*, *Mi sento insoddisfatto*, *Mi sento infelice*, *Ho poca fiducia in me stesso* ...

Cercando di leggere e di interpretare queste connessioni, che intuitivamente sappiamo corrispondere a realtà, si percepisce subito la sensazione di sprofondare nella complessità. In uno studio medio le connessioni statisticamente significative possono essere migliaia. In uno studio di grandi dimensioni si arriva a diversi milioni. Pur trattandosi di uno strumento sintetico di lettura dei dati, le tabelle delle connessioni fanno subito intuire la necessità di un ulteriore strumento in grado di ridurre la complessità.

- *analisi fattoriale*

Continuando l'esempio precedente è possibile selezionare le variabili connesse con *Mi sento depresso* e su queste variabili effettuare un'analisi fattoriale.

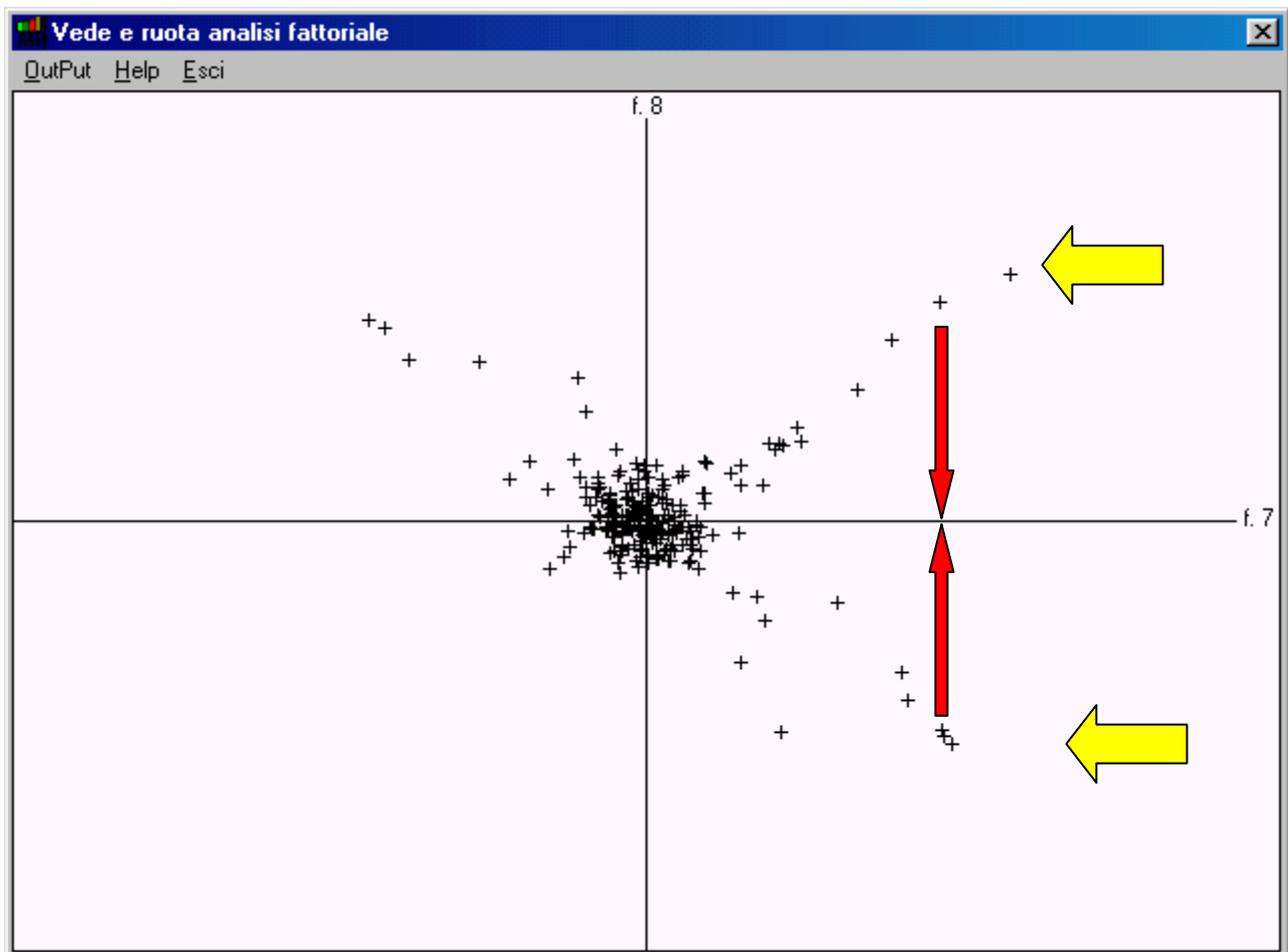
L'analisi fattoriale parte dalla matrice delle connessioni per costruire delle nuove variabili, non correlate tra di loro ($r\text{-Phi} = 0,000$). Per ognuna di queste nuove variabili si conosce il valore di connessione ($r\text{-Phi}$) con le variabili dicotomiche originarie. E' quindi possibile utilizzare queste variabili come assi cartesiani (assi fattoriali), in cui i valori di connessione diventano le coordinate con cui le variabili originarie vengono rappresentate nello spazio.



In questo esempio è riportato l'incrocio tra l'asse fattoriale 2 e l'asse fattoriale 3. Come si vede i due fattori si incrociano a 90°, in quanto non sono correlati tra loro. Le variabili originarie sono rappresentate come delle piccole croci. L'obiettivo dell'analisi fattoriale è di

selezionare gruppi di variabili tra loro correlate. L'idea è che variabili tra loro correlate esprimano un comune denominatore, un fattore (da qui il termine analisi fattoriale). L'analisi fattoriale può individuare un numero elevato di assi fattoriali (ma in numero sempre inferiore alle variabili inserite nell'analisi). In questo esempio sono state inserite 85 variabili, il numero di assi individuato è di poco superiore a 20.

L'analisi fattoriale porta ad una rappresentazione multidimensionale (ogni asse fattoriale corrisponde ad una dimensione) e che la rappresentazione delle variabili su di un piano (incrocio di due assi fattoriali) è una riduzione della reale complessità della rappresentazione fattoriale. In altre parole, due variabili dicotomiche che appaiono vicine su di un piano, potrebbero invece essere distanti su di un altro piano.



Prima di leggere i risultati dell'analisi fattoriale è quindi necessario ridurre il rischio di leggere assieme variabili che non sono tra loro correlate. A tal fine si ruotano gli assi fattoriali in modo da far coincidere le strutture fattoriali con gli assi. Nell'esempio appena

riportato si notano due strutture (indicate dalle frecce gialle). Queste due strutture sono poco correlate tra loro in quanto formano tra loro un angolo (come si vede dalle rette blu) prossimo ai 90°. Ma, le coordinate fattoriali (i valori r-Phi) di queste due strutture sono, sul fattore 1, coincidenti. Leggendo il fattore 1 si è portati a confondere assieme queste due strutture. Per ovviare alla sovrapposizione delle strutture, si opera ruotando le coppie di assi in modo da far coincidere (o avvicinare) le due strutture con gli assi fattoriali (una struttura su ogni asse).

Effettuata la rotazione di tutte le coppie di assi fattoriali si può passare alla lettura dei risultati. Ogni asse presenterà la lista delle variabili dicotomiche connesse. Si opera quindi cercando di dare un nome, di associare un concetto generale all'asse. Nell'esempio che stiamo seguendo, troviamo che le variabili connesse alla depressione si raggruppano in:

Fattore 1 – <i>fisicità</i>	Fattore 2 – <i>giudizio altrui</i>	Fattore 3 - <i>ambiente</i>
Mi sento brutto Non mi piace il mio corpo Mi sento impacciato sessualmente	Mi sento molto impacciato Ho poca fiducia in me stesso E' difficile avere dei veri amici Ho molta paura del giudizio altrui Vengo spesso criticato	Conosco molta gente sola Conosco molti tossici

Ovviamente l'analisi fattoriale deve essere presa come un suggerimento. Abbiamo appena visto i passaggi che stanno dietro a questa analisi. E' chiaro che nella lista delle variabili collegate ad un fattore possono capitare variabili non coerenti, sporcizia, come conseguenza di una non adeguata rotazione degli assi fattoriali. E' quindi sempre necessario prendere i risultati fattoriali in modo critico, accettando solo ciò che ha una logica ed una sua coerenza.

- *analisi dei gruppi*

L'analisi fattoriale porta alla produzione di profili, cioè elenchi di variabili tra loro correlate che descrivono le caratteristiche di un fattore. E' possibile attribuire ad ogni scheda un valore di similarità con ciascun profilo. I valori di similarità variano da 0% a 100%, dove 0 indica assenza di similarità e 100 indica identità con il profilo. Questi valori diventano delle nuove variabili che è possibile utilizzare per selezionare le schede, creando così gruppi omogenei relativamente a ciascun fattore, ad esempio è possibile selezionare tutte le schede con un valore di similarità maggiore a 75% con il fattore 1, ecc... Ovviamente una stessa scheda può ottenere valori di similarità elevati con più profili e, allo stesso modo, si possono avere delle schede che ottengono valori di similarità bassi con tutti i profili. I valori di similarità possono essere trattati come delle nuove variabili e gestiti alla stregua di una qualsiasi variabile producendo distribuzioni di frequenze, tabelle doppie, tabelle delle connessioni ed eventualmente utilizzandole come macro-variabili per nuove analisi fattoriali.

3. Esigenze conoscitive e informazioni da raccogliere

I dati sono la materia prima per le analisi statistiche, ma non tutti i dati possono essere oggetto di elaborazione statistica. E' infatti necessario che i dati siano codificati e che siano stati raccolti in modo sistematico:

- *Codifica dei dati*: solo le variabili dicotomiche, nominali, ordinali e continue 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.

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 inficiare le nostre ipotesi, si produce un lavoro sterile, poco interessante, poco utile e di scarso valore scientifico. Prima di avviare la progettazione dello strumento di rilevazione dei dati è perciò bene tener presente che è necessario distinguere tra variabili chiave, variabili esplicative e variabili di struttura:

- le **variabili chiave** sono quelle che descrivono l'argomento trattato, ad esempio se l'indagine è relativa ai tumori, le variabili chiave saranno quelle relative al tipo e progressione del tumore;
- 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;
- 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 il campione utilizzato nell'indagine.

Per mettere a punto lo strumento di rilevazione dei dati:

1. si inizia esplicitando gli obiettivi dell'indagine (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 di ricerca 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 trova 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 assicurarci che lo strumento risponda alle varie esigenze e che non crei tensioni che ne minino il funzionamento.

4. Informazioni distorte dall'effetto maschera

L'effetto maschera è dovuto al fatto che le persone rispondono cercando di dare un'immagine di se stesse positiva o che soddisfi ciò che loro pensano sia l'aspettativa del ricercatore (effetto complicità). L'effetto maschera porta però all'impossibilità di utilizzare i dati a fini quantitativi, ad esempio se chiediamo alle persone di dirci quanto si sentono depresse, non potremo con questi dati dire quanto le persone siano in effetti depresse.

Questo punto è estremamente importante in quanto molti cadono nel tranello di pensare che l'obiettivo della ricerca sia quello di quantificare la realtà. E' importante sottolineare nuovamente che l'obiettivo della ricerca scientifica è quello di individuare relazioni. Tutti i fenomeni si definiscono in base alle relazioni che hanno con altri fenomeni. Ad esempio la depressione è associata ad un vissuto di inutilità, alla mancanza di significati e di scopi, alla percezione di una sensazione particolarmente fastidiosa, alla mancanza di energie,.... relazioni che permettono di definire il fenomeno depressione, ma anche di comprenderne i meccanismi sottostanti.

E' importante fare chiarezza a questo riguardo in quanto la metodologia relazionale, a differenza della metodologia sperimentale, non risente dell'effetto maschera. Ad esempio, con la metodologia relazionale è possibile chiedere direttamente: *Ti senti solo?*, *Ti senti depresso?*, *Ti senti angosciato?* L'effetto maschera impedirà di utilizzare i dati da un punto di vista quantitativo, ma non per lo studio di relazioni. Infatti, la maschera si applica in modo omogeneo su tutte le risposte. 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, si prenda 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*):

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

Correggendo l'effetto della maschera, ad esempio utilizzando un test psicologico e non una domanda diretta, si sarebbe ottenuta una 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*):

	Depressi	Non depressi	Totale
Angosciati	158	10	168
Non Angosciati	2	30	32
Totale	160	40	200

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 con dati qualitativi. Se non fosse possibile analizzare i dati che sono oggetto di maschera la metodologia relazionale non sarebbe, dopo tutto, una tale innovazione nel campo della ricerca scientifica. Il problema della maschera limita la possibilità di realizzare studi quantitativi e analisi delle differenze, non la possibilità di individuare concomitanze, e relazioni.

Ridondanza dell'informazione

La metodologia relazionale si applica su fenomeni complessi dove la stessa informazione può avere più significati. Ad esempio all'inizio del nostro lavoro possiamo avere il dubbio se è 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 più domande relative alla solitudine, non una sola. Infatti se l'unica domanda che è stata inserita non dovesse funzionare, salterebbe completamente l'intero lavoro di ricerca.

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, la relazione viene inserita nella domanda. 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.

Le connessioni tra variabili soggettive (mascherate) e variabili oggettive sono spesso basse, la maschera non si estende infatti alle informazioni oggettive. Le connessioni tra variabili soggettive e oggettive possono risultare più basse del dovuto a causa dell'uso della maschera solo sulle variabili soggettive, non è però possibile la creazione di connessioni inesistenti: se si osservano connessioni significa quindi che le connessioni esistono. Nel momento in cui l'interesse sia quello di correlare informazioni soggettive con informazioni oggettive è quindi necessario aumentare il campione utilizzato nell'indagine.

5. Il campione

E' importante notare che:

- Quando l'indagine ha finalità quantitativa si utilizzano informazioni oggettive e si utilizza un campione rappresentativo.
- Quando l'indagine ha finalità qualitativa, cioè studiare i meccanismi, le relazioni, si utilizzano informazioni oggettive/soggettive e campioni eterogenei.

Ad esempio, se il nostro obiettivo è quello di studiare che cosa caratterizza i tossicodipendenti utilizzeremo un campione in cui metà dei soggetti sono tossicodipendenti, mentre l'altra metà non lo sono. Il confronto tra questi due campioni consentirà di stabilire che cosa si collega, si correla, caratterizza l'essere tossicodipendente. Se lo studio venisse effettuato solo sui tossicodipendenti saremmo limitati a descrizioni quantitative (ad esempio distribuzioni di frequenze), e non potremmo in nessun modo conoscere che cosa caratterizza i tossicodipendenti.

In altre parole, per massimizzare le correlazioni, è necessario utilizzare campioni eterogenei. Maggiore è la variabilità del campione, maggiore è la possibilità di individuare correlazioni. Quando i dati non variano cessa anche la possibilità di studiare correlazioni. Questa caratteristica della metodologia relazionale rende la ricerca estremamente semplice, economica ed intuitiva.