

In “Sociologia e Ricerca Sociale” n. 57 (1998): 49-59.

Termometri con vincolo di ordinalità: il “gioco della torre” consente di aggirare la tendenza alla desiderabilità sociale?

di Alberto Marradi

1. Premessa

Opinioni, atteggiamenti e valori sono un settore di indagine di interesse cruciale per le scienze umane, e sicuramente inaccessibile alla misurazione in senso proprio. D'altra parte, è conveniente trattare le relative variabili come cardinali, perché ciò permette di sottoporle a potenti tecniche di analisi multivariata. Tuttavia, non tutte le tecniche con cui abitualmente si rilevano opinioni, atteggiamenti e valori generano variabili che è legittimo, o plausibile, trattare come cardinali.

Un'accettabile approssimazione alla cardinalità è raggiunta mediante le cosiddette *scale auto-ancoranti*, che ottengono questo risultato nell'unico modo possibile, cioè riducendo al minimo l'autonomia semantica delle categorie intermedie fra i due estremi^[1].

Altrove (1992, par. 5.2) ho sottolineato le profonde differenze che sussistono — sia dal punto di vista strutturale sia dal punto di vista semantico — fra queste ultime tecniche e la misurazione in senso stretto. Tuttavia, nel complesso si può ritenere che le caratteristiche delle variabili costruite con le tecniche auto-ancoranti giustifichino l'etichetta di “quasi-cardinali”, e che le (non gravi) distorsioni implicite nell'applicazione di tecniche di analisi che presuppongono una piena cardinalità siano ampiamente bilanciate dai vantaggi connessi all'impiego di tali tecniche.

Delle varie tecniche auto-ancoranti (ad es. il differenziale semantico, il termometro dei sentimenti, la scala auto-ancorante in senso stretto^[2]), trovo che il termometro — per vari motivi esposti nel luogo sopra citato — sia la preferibile, anche perché suscettibile di vari perfezionamenti. A questi perfezionamenti sto infatti lavorando da tempo, e in questo breve articolo esporrò alcuni risultati di questo lavoro.

2. Il termometro dei sentimenti nella versione canonica

Nella versione originaria^[3], e tuttora canonica, del termometro, all'intervistato viene mostrato un cartoncino con l'immagine di un termometro graduato da 0 a 100; lo si istruisce ad identificare con 0 il massimo di ostilità e con 100 il massimo di favore verso un oggetto cognitivo (persona, gruppo, istituzione, avvenimento, idea). Dopodiché l'intervistatore nomina uno di questi oggetti, e l'intervistato gli assegna *verbalmente* un punteggio fra 0 e 100 in base al suo sentimento verso di esso (le posizioni potenziali sono 101 — da 0 a 100 — ma in genere gli intervistati tendono a usare solo le cifre tonde: di solito i multipli di 10, talvolta anche i multipli di 5). Subito dopo l'intervistatore sottopone all'intervistato un altro oggetto da valutare, e così via. Non ci sono limiti al numero di oggetti che è possibile presentare all'intervistato.

Venni a contatto con questo strumento in occasione di una ricerca multinazionale su “Insoddisfazione, protesta e mutamento” diretta da Samuel Barnes e Max Kaase negli anni settanta, che includeva una batteria di 16 termometri. Il segmento italiano di questa ricerca era diretto da Sartori; quando Sartori si trasferì ad insegnare a Berkeley, ne ereditai la direzione.

Sfogliando i 2.500 questionari raccolti in quell'occasione mi risultò evidente un grave difetto

di questa versione dello strumento. Cioè il fatto che alcuni intervistati tendevano ad assegnare a tutti o quasi tutti gli oggetti punteggi alti (tra il 50 e il 100), altri intervistati tendevano ad assegnare punteggi bassi (tra 0 e 50), altri ancora preferivano i punteggi intermedi (fra 30 e 70), e qualcuno tendeva a polarizzare i suoi giudizi sui due estremi (molti punteggi vicini a 100 e molti altri vicini allo zero). Ricordo che, nel complesso, solo un 50% scarso degli intervistati in quella ricerca distribuiva i punteggi in modo equilibrato su tutto l'arco della scala, cioè usava correttamente lo strumento.

Un altro difetto non risultava evidente sulla base delle distribuzioni dei punteggi assegnati, ma era facilmente individuabile se si ricostruiva il processo di somministrazione dello strumento. Si trattava di questo: visto che gli oggetti da valutare erano sottoposti all'intervistato uno per volta e che lui/lei doveva pronunciarsi sul primo senza conoscere i successivi, poi sul secondo senza conoscere i successivi, e così via, ne conseguiva l'impossibilità per l'intervistato di calibrare le sue risposte iniziali (e, in misura decrescente, anche quelle via via successive) alla luce dell'intero paniere di oggetti che gli venivano sottoposti. Provai a mettermi nei panni di un intervistato di sinistra cui veniva sottoposto tra i primi oggetti la DC, alla quale lui dava punteggio 0. Se successivamente gli veniva sottoposto come oggetto da valutare anche il MSI, probabilmente questo intervistato avrebbe voluto dargli un punteggio inferiore a quello attribuito alla DC, ma questo gli era impossibile perché il termometro in quella forma non prevede punteggi negativi.

3. La deflazione

Al primo dei due difetti descritti cercai di porre rimedio analizzando proprio i dati della ricerca sopra menzionata, che furono raccolti nell'inverno 1974-75. Il successivo processo di codifica fu lento e faticoso perché a quel tempo ancora si perforavano le schede, e non c'era alcun modo di controllare la correttezza dei codici inseriti se non globalmente ed ex post. La perforazione fu affidata ad una ditta di Pisa, che lavorava per il CNUCE; per mesi e mesi i suoi incaricati continuarono ad andare e venire tra Pisa e Firenze con grossi pacchi di questionari e di schede^[4].

Passai l'estate '76 a controllare accuratamente la qualità di questi dati trovando errori di ogni genere; di conseguenza il *file* fu pulito e pronto per essere analizzato solo verso la fine di quell'anno. Provai a sottoporre ad analisi fattoriale la batteria dei 16 termometri di cui ho detto. Mi resi conto del fatto che — come verrà mostrato nell'articolo che segue — usando i punteggi “grezzi” (cioè quelli assegnati dagli intervistati) il fenomeno descritto nel par. 2, cioè la tendenza di molti intervistati a dare punteggi o prevalentemente alti o prevalentemente bassi, e così via, unito alla “desiderabilità sociale”^[5] di alcuni degli oggetti sottoposti a giudizio, produceva un annebbiamento delle dimensioni valoriali sottese all'assegnazione dei punteggi stessi: il *plot* dell'analisi fattoriale mostrava un indistinto nuvolone in cui a un'estremità si collocavano gli oggetti cognitivi cari ai progressisti (ad es. il PCI, i sindacati, gli immigrati extracomunitari) e all'altra gli oggetti cognitivi cari ai conservatori (ad es. la DC, la verginità, la grande industria), senza ulteriori articolazioni significative dello spazio.

Mentre non mi pareva possibile rimediare in sede di analisi, o comunque di trattamento dei dati, al problema della desiderabilità sociale di alcuni oggetti, pensai che si poteva rimediare alla tendenza degli intervistati a privilegiare i punteggi alti, o quelli bassi, o quelli centrali, o quelli estremi. Il rimedio si basava sull'idea che questa tendenza fosse un attributo costante del soggetto (almeno nel limitato arco di tempo di un'intervista) e che quindi — a certe condizioni — fosse possibile stimarla e neutralizzarla. La definizione operativa della tendenza di un individuo a dare punteggi alti o bassi era — ovviamente — la media dei punteggi da lui attribuiti a tutti gli oggetti che gli erano stati sottoposti. Sostituendo ai punteggi “grezzi” i loro scarti dalla media, la suddetta tendenza poteva essere neutralizzata. Per poter legittimamente applicare un rimedio piuttosto

grossolano come questo dovevano però essere tassativamente soddisfatte due condizioni:

1) il numero di oggetti sottoposti al giudizio dell'intervistato con questa tecnica doveva essere sufficientemente ampio (almeno quindici-venti), al fine di conferire robustezza alla media;

2) gli oggetti sottoposti dovevano inoltre occupare presumibilmente le posizioni più varie possibile su tutte le dimensioni concepibili (destra/sinistra politica, destra/sinistra economica, nuovo/vecchio, centro/periferia, pubblico/privato, secolare/religioso, e così via). In questa maniera si rendeva molto improbabile l'eventualità che un intervistato desse punteggi prevalentemente bassi semplicemente perché la gran parte degli oggetti sottoposti al suo giudizio gli erano effettivamente sgraditi, oppure prevalentemente alti perché la gran parte degli oggetti gli erano effettivamente graditi.

Se queste due condizioni erano soddisfatte, la trasformazione dei punteggi "grezzi" in scarti poteva ritenersi un accettabile rimedio alla tendenza a dare punteggi alti o bassi. La tendenza di un intervistato a dare prevalentemente punteggi intermedi o prevalentemente punteggi estremi poteva esser neutralizzata se era soddisfatta almeno la prima delle condizioni viste sopra, dividendo ogni punteggio trasformato in scarto per lo scarto-tipo della distribuzione dei punteggi dati da quell'intervistato a tutti gli oggetti della batteria di termometri.

Ho attribuito (1979) a questa procedura il nome 'deflazione', data la sua analogia con un procedimento usato dagli economisti per eliminare gli effetti del livello generale dei prezzi da varie altre grandezze fondamentali per la loro disciplina. La deflazione è praticamente analoga a una standardizzazione, salvo che è applicata sulle righe invece che sulle colonne della matrice; più esattamente essa è applicata al segmento di punteggi dati da un intervistato agli oggetti di una batteria di termometri.

4. Il termometro con il *gadget*

Per il secondo difetto descritto sopra (par. 2), un rimedio mi venne naturale in occasione di una ricerca (inedita) sull'immagine delle aziende a partecipazione statale nella città di Genova, da me diretta nel 1981 insieme a Roberto Cartocci.

Feci approntare dagli uffici tecnici della principale azienda committente, che era l'Italsider, un foglio di cartone di grande formato (50 * 70 cm.) e di colore vivace, sul quale veniva disegnato un termometro con 21 tacche corrispondenti alle cifre che crescevano, di 5 in 5, da 0 fino a 100. La grandezza del foglio faceva sì che fosse possibile collocare materialmente, su ciascuna delle righe che si estendevano ai due lati di ciascuna tacca, fino a sei cartellini con su scritti i nomi degli oggetti da valutare. Tali nomi venivano pertanto scritti su un certo numero di cartellini di colore bianco, in modo che risaltassero sul fondo colorato del foglio. Si insisteva molto sul fatto che prima di iniziare a collocare i cartellini l'intervistato si familiarizzasse con essi in modo da averne un pieno controllo intellettuale. Si incoraggiava, inoltre, l'intervistato a sentirsi libero di spostare i cartellini tutte le volte che voleva fino a che non fosse pienamente soddisfatto delle collocazioni che aveva attribuito.

Con la procedura e lo strumento ora descritti, le posizioni si riducono a 21 dalle 101 astrattamente possibili con la versione originaria del termometro. Ciò per ragioni puramente materiali (cioè poter disegnare il termometro su un foglio di dimensioni ragionevoli anche se ampie). Peraltro, questo espediente non altera sensibilmente il grado di approssimazione alla cardinalità dei punteggi prodotti con questa tecnica perché, come accennato nel par. 2, in realtà quasi tutti gli intervistati utilizzano punteggi multipli di 10, o al massimo di 5, anche nella versione originaria con risposta verbale.

Oltre che nella ricerca sopra menzionata, ho usato il termometro in questa forma (che chiamerò ‘con il *gadget*’)^[6] in varie tesi e in qualche altra ricerca rimasta inedita.

5. Il termometro come strumento di *rank ordering*

Nel 1988 mi capitò di descrivere questo strumento in un seminario che tenni nell’Istituto di sociologia dell’Università di Perugia. In quell’occasione Claudio Bezzi, ricercatore dell’IRRES (istituto di ricerca economica e sociale della Regione Umbria), intervenne sostenendo l’illegittimità di fare un uso cardinale dei punteggi assegnati agli oggetti mediante questo strumento. Capii in seguito che tener conto dell’osservazione di Bezzi significava in pratica riesumare le classiche tecniche di *rank-ordering* degli psicometrici (vedi ad es. Thurstone, 1931), con l’aggiunta di due dettagli importanti: l’uso di un supporto grafico come il cartone descritto sopra (che dava l’idea della scala, e soprattutto permetteva all’intervistato di meditare e di calibrare le sue risposte), e il fatto che gli intervistati erano incoraggiati a rivedere e ri-calibrare i loro giudizi spostando materialmente i cartellini lungo la scala.

L’occasione di sottoporre a prova pratica il funzionamento di quest’ultima versione del termometro mi fu offerta da una tesi cui Francesca Giovani e Carla Pellizzari^[7] lavorarono dal 1987 al 1990. Le interviste furono svolte tutte in Toscana, in centri di dimensioni diverse.

Per la parte che qui ci interessa (perché la ricerca era più ricca) si trattava di sottoporre agli intervistati, a distanza di tempo in quanto intercalate con batterie di scale Likert, tre versioni del termometro applicate agli stessi 21 oggetti: quella originaria, quella con il *gadget*, e quella che avevo ideato per tener presente l’obiezione di Bezzi appena menzionata. In quest’ultima versione si istruiva l’intervistato a fare un uso meramente ordinale della scala, cioè a sistemare non più di un cartellino su ogni gradino. Quindi lo si induceva inevitabilmente a operare una serie di confronti a coppie o a terne tra oggetti diversi (“tra questi tre cartellini che mi piacciono molto, quale metto al livello 100 della scala?” e analogamente per tutti i gradini successivi della scala).

Il fatto che non si potesse sistemare più di un cartellino su ognuno dei 21 gradini della scala comportava la necessità di limitare comunque a 21 il numero degli oggetti sottoposti a giudizio anche quando essi venivano somministrati con le altre versioni del termometro, che in sé non avrebbero posto questo limite.

6. Il “gioco della torre”

Una caratteristica importante di questa ricerca era che si chiedeva preliminarmente agli intervistati di commentare liberamente sia le affermazioni che sarebbero poi state sottoposte loro mediante le scale Likert^[8] sia gli oggetti cui avrebbero poi dovuto assegnare un punteggio mediante il termometro.

Ascoltando i nastri con questi giudizi, constatai che alcuni intervistati si esibivano in grandi declamazioni retoriche a proposito di alcuni oggetti “socialmente desiderabili”^[9]. Ma quando poi si trattava di confrontare questi oggetti con altri, la desiderabilità sociale finiva spesso per essere ignorata, e gli oggetti in questione venivano collocati molto più in basso (a) di quello che le dichiarazioni spontanee facevano pensare, (b) della posizione in cui erano collocati con tecniche che non comportavano un confronto diretto.

INSERIRE FIGURA 1 più o meno QUI

Per sottoporre a controllo empirico questa impressione, immaginai un particolare diagramma a bandiera, che serviva a confrontare — oggetto per oggetto — i punteggi medi assegnati con una delle altre due versioni del termometro e quelli assegnati con la versione ordinale. Ovviamente, in questo caso si confrontavano i punteggi grezzi (non deflazionati) relativi a ciascun oggetto. La figura 1 mostra appunto le differenze fra le medie dei punteggi assegnati agli stessi oggetti con i termometri nella versione originaria e con quelli nella versione ordinale.

La differenza^[10] tra queste medie di punteggi assegnati ad ogni oggetto con le due versioni è leggibile il fondo al relativo segmento, cioè sul frastagliato bordo della bandiera.

Per facilitare l'interpretazione della figura farò un esempio: ai Verdi era stato assegnato un punteggio medio di 52,9 con il termometro in versione originaria e di 54,9 con il termometro in versione ordinale: per questo il relativo segmento è lungo 2 unità e compare a destra dell'asse della bandiera. A destra infatti si trovano i segmenti relativi a tutti gli oggetti che hanno ricevuto una valutazione media più alta con la versione ordinale che con la versione originaria del termometro.

Osservando con attenzione la figura 1 si nota innanzitutto un fatto: gli intervistati avevano manifestato la tendenza a dare punteggi tendenzialmente bassi quando non erano costretti a coprire tutta la scala, come in effetti erano costretti a fare con il termometro ordinale^[11]. A questa conclusione si arriva constatando che per la maggior parte degli oggetti il punteggio medio assegnato era più alto quando l'intervistato era costretto a usare tutta la scala.

La differenza in questione appare particolarmente sensibile per gli oggetti collocati in fondo alla bandiera nella figura 1. Questi oggetti sono tre oggetti-simbolo tradizionali della sinistra (sindacati, PCI, cooperative). Essi evidentemente non fruivano di alcun appello emotivo superficiale di tipo "socialmente desiderabile", che induceva molti intervistati a dare punteggi alti quando reagiva verbalmente. Ma al momento di scegliere direttamente fra questi oggetti e altri magari più *à la page*, gli intervistati toscani (prevalentemente di sinistra) sentivano il richiamo dei simboli tradizionali della sinistra e tendevano a farli salire qualche gradino più in alto.

All'estremo opposto, cioè nella parte alta della bandiera, a sinistra dell'asta, stanno gli oggetti — pochi in questa ricerca^[12] — che hanno ricevuto una valutazione media più alta con la versione originaria che con la versione ordinale (nella fig. 1: gli extracomunitari, un uomo che lava i piatti e gli omosessuali). Questi oggetti si trovano sulla scala in una posizione diametralmente opposta rispetto agli oggetti-simbolo della sinistra tradizionale, posizione che rispecchia perfettamente la loro situazione psicologicamente speculare. Essi infatti — essendo *à la page*, e quindi socialmente desiderabili — hanno ottenuto valutazioni alte quando l'intervistato non era tenuto ad operare confronti espliciti e diretti, potendo collocare più cartellini sullo stesso gradino della scala. Tuttavia, al momento di confrontarli direttamente con altri oggetti, e in particolare con gli oggetti-simbolo della sinistra tradizionale, molti intervistati finivano per preferire questi ultimi, con la conseguenza di far scendere qualche gradino più in basso gli oggetti socialmente desiderabili.

INSERIRE FIGURA 2 più o meno QUI

Anche la figura 2, dedicata a un confronto tra i punteggi dati agli stessi oggetti con il *gadget* e con la versione ordinale del termometro, conferma queste osservazioni. Gli oggetti socialmente desiderabili (extracomunitari, uomo che lava i piatti e omosessuali) sono a un estremo della scala; gli oggetti-simbolo tradizionali della sinistra (sindacati, PCI, cooperative) sono all'altro estremo.^[13]

Riflettendo su questo risultati, che considero importanti, mi resi conto che soddisfacendo l'obiezione di Bezzi, ispirata a lodevoli criteri di correttezza metodologica, avevo inopinatamente

conseguito un obiettivo non immaginato dall'autore, e al quale io stesso non avevo pensato prima di constatare i relativi risultati empirici. L'obiettivo conseguito consisteva nel fatto che, inducendo l'intervistato a fare una specie di "gioco della torre" tra oggetti disparati, si finiva per aggirare in molti casi il condizionamento della desiderabilità sociale. In altre parole, una volta che erano obbligati a scegliere quale fra due oggetti collocare su un dato gradino e quale fare scendere a gradini sottostanti, gli intervistati tendevano a manifestare le loro preferenze di fondo, cioè manifestavano più spesso giudizi presumibilmente autentici anziché giudizi ispirati dalla preoccupazione di "presentare il sé"^[14].

Il "gioco della torre", costringendo l'intervistato a operare confronti, sempre sgraditi e spesso dolorosi^[15], fra oggetti e simboli disparati, sembra una promettente arma in mano al ricercatore che voglia aggirare la desiderabilità sociale sperando di ottenere così dati più fedeli, cioè meglio corrispondenti al vero stato d'animo dell'intervistato. Una volta scoperta questa sua virtù, il principio del "gioco della torre" può naturalmente essere applicato a un supporto grafico diverso da un termometro: una rudimentale torre, o una scala a gradini. In entrambe queste vesti si è rivelato particolarmente efficace come supporto grafico di domande ricorrenti nei questionari dei sociologi, come quelle relative ai problemi più importanti, agli interventi di *policy* preferiti, e simili. Di questi risultati parlerò in altro articolo.

Il risultato che qui ho riportato è stato ottenuto con tecniche di analisi molto semplici: un confronto tra medie esposto con l'adeguata rappresentazione grafica. L'articolo che segue mostrerà che esso appare confermato anche ricorrendo a tecniche più sofisticate, come l'analisi delle componenti principali e la matrice multi-tratto / multi-tecnica.

Riferimenti bibliografici

H. Cantril, L. A. Free, 1962, «Hopes and Fears for Self and Country: The Self-Anchoring Striving Scale in Social Research», *American Behavioral Scientist*, VI, 2 (october), supplement.

G. Gasperoni, F. Giovani, 1992, *Come e perché non funzionano le scale Likert con items a polarità semantica invertita*, in A. Marradi, G. Gasperoni (a cura di), *Costruire il dato 2*, Milano, Angeli, pp. 60-99.

E. Goffman, 1959, *Presentation of Self in Everyday Life*, New York, Doubleday. Tr. it. *La vita quotidiana come rappresentazione*, Bologna, Il Mulino, 1969.

F. P. Kilpatrick, H. Cantril, 1960, «Self Anchoring Scaling. A Measure of Individuals' Unique Reality Worlds», *Journal of Individual Psychology*, XVI, 2 (november), pp. -73.

A. Marradi, 1979, «Dimensioni dello spazio politico in Italia», *Rivista Italiana di Scienza Politica*, IX, 2 (agosto), pp. 263-96.

A. Marradi, 1992, *L'analisi monovariata*, Milano, Franco Angeli.

L. L. Thurstone, 1931, «Rank Order as a Psychophysical Method», *Journal of Experimental*

[1]

[2]

[3]

[4]

[5]

[6][6] Gadget [6] Gadget è un termine generale che può essere tradotto con 'aggeggio'. Nella pratica delle agenzie di sondaggi. per gadget

[7]

[8]

[9]

[10]

[11]

Il fatto che la ricerca fosse condotta in Toscana spiega probabilmente il basso livello medio dei punteggi assegnati dagli intervistati quando non erano obbligati a coprire tutto l'arco della scala. E' noto infatti che una caratteristica subculturale dei toscani è la forte propensione alla critica.

[12]

[13][13] Gli oggetti collocati in posizione significativamente diversa nelle due figure sono solo tre: il matrimonio in chiesa, il movimento dei consumatori e i Verdi. Tutti e tre questi oggetti avevano ricevuto — per ragioni sulle quali posso solo avanzare delle deboli congetture — un punteggio medio di 3-4 punti più alto con la versione originaria rispetto a quello ricevuto nella versione con il gadget

[14]

[15]