

Per dissipare questo sospetto, dicono, sarebbe necessario usare tecniche di rilevazione assai più raffinate. Dato il concetto A, se ne dovrebbero scegliere degli indicatori tali da poter essere registrati con definizioni operative quanto più possibile diverse l'una dall'altra (quanto a fonti delle informazioni, forma delle domande, tipo di risposte previste, strumenti e situazioni in cui registrare le risposte, etc.). Lo stesso si dovrebbe fare anche con gli altri concetti generali B e C. Il sospetto sarebbe dissipato se le correlazioni fra gli indicatori dello stesso concetto, registrati con definizioni operative diverse, fossero considerevolmente più alte delle correlazioni fra gli indicatori di concetti diversi, registrati con definizioni operative simili. In tal caso si avrebbe — dicono gli autori — una *validazione convergente* (in quanto derivante da congruenza fra indicatori formalmente assai diversi) e una *validazione discriminante* (in quanto indicatori formalmente simili ma legati semanticamente a concetti diversi risulterebbero scarsamente congruenti fra loro). In effetti — osserviamo — non si avrebbe una vera e propria prova della validità degli indicatori (cioè del loro legame semantico proprio con il concetto che intende il ricercatore), ma certo un indizio confortante.

Tuttavia, per quanto la necessità che la congruenza fra indicatori derivi dalla loro vicinanza semantica, anziché da somiglianze fra le relative definizioni operative, possa sembrare un requisito abbastanza ovvio e minimale, Campbell e Fiske presentano numerosissimi esempi, tratti da ricerche pubblicate nei campi più diversi, in cui tale requisito è lungi dall'essere soddisfatto, e concludono che ne emerge « un quadro piuttosto tragico della validità degli indicatori implicati... quadro che — assicurano — non è assolutamente frutto di uno sforzo deliberato di selezionare tutti gli esempi più vistosamente negativi » (1959, 93 e 95).

In effetti, al di là della validità di certi indicatori, è la stessa capacità delle scienze sociali di conoscere il proprio oggetto senza alterarlo che così è messa in questione. La definizione operativa appare non più una maniera di registrare gli stati degli oggetti su una certa proprietà, ma un meccanismo che contribuisce a creare gli stati che registra. Da notare che, un decennio dopo Campbell e Fiske, il politologo Philip E. Converse è giunto a conclusioni altrettanto allarmanti in materia del tutto indipendente, cioè dopo aver constatato il modo, in molti casi del tutto casuale, in cui i membri di un *panel* (vedi sezione 2.4) cambiavano risposte da un'intervista all'altra (1970).

Non si direbbe che questi dubbi epistemologici abbiano imbarazzato la grande maggioranza dei ricercatori nelle scienze sociali. Gli studi tecnici sul miglior modo di calcolare la validità convergente e discriminante (per es. Jackson 1969; Althausen e Heberlein 1970; Schmitt, Coyle e Saari 1977; Hubert e Baker 1978) non sono mancati — essi permettono infatti eleganti elaborazioni matematiche — ma ben poche ricerche si sono adeguate alla richiesta di dare definizioni operative quanto più differenti possibile degli indicatori dello stesso concetto. Si deve peraltro riconoscere che tale adeguamento richiede in molti casi una notevolissima creatività metodologica, e pone inoltre problemi pratici spesso insormontabili nelle ricerche di ambito geografico più che locale.

Le considerazioni di Campbell e Fiske, nonché di Converse, dovrebbero quanto meno indurci a un sano scetticismo verso misure di attendibilità e validità basate sulla congruenza interna di batterie o altri gruppi di indicatori operativamente definiti in maniera troppo simile.

4.0. Classificazione, misurazione e conteggio

Nella sezione 2.2 si è affermato che il compito della definizione operativa è registrare lo stato di ciascun caso sulla proprietà in questione, assegnando tale stato a una delle categorie di un elenco già predisposto. Questa formulazione generica ci ha permesso di affrontare altri problemi, che si consideravano logicamente prioritari, rinviando la trattazione dei criteri con cui tali elenchi devono essere costituiti, e i casi assegnati. Infatti, come il lettore avrà probabilmente notato, si è evitato per quanto possibile di usare i termini 'classificazione' e 'misurazione' prima di averne precisato con cura il significato, come si cercherà di fare in questa e nelle successive sezioni.

Precisazione tanto più importante in quanto una malintesa visione positivista della scienza ha portato all'equiparazione implicita, e talvolta anche esplicita, di 'misurazione' con 'scienza' e di 'classificazione' con 'filosofia' (o comunque con 'non-scienza'). Di conseguenza si è cercato di radiare dal linguaggio scientifico lo stesso termine 'classificazione', ri-etichettando i tre principali concetti che designa con i termini 'misurazione' o 'scala'. Questa mistificazione merita tutte le critiche rivoltele, fra gli altri, da Sartori (1971, 22-28). Non si ignorano certo i vantaggi che presentano le proprietà misurabili dal punto di vista del trattamento matematico dei dati. Ma è fuorviante definire misurabili proprietà che non lo sono, e ancora peggio applicare tecniche matematiche a delle etichette che dei numeri hanno solo la veste, ma nessuna qualità essenziale.

C'è di più. Nelle sezioni che seguono la guerra sarà, per così dire, portata in casa del nemico, sostenendo che la misurazione è possibile solo in quanto è preceduta e seguita da atti di natura classificatoria. Tale argomentazione parte dalla premessa che — a seconda della natura della proprietà in questione — la *definizione operativa ha tre modi alternativi per assegnare un valore allo stato di un oggetto sulla proprietà stessa*.

A) Alcune proprietà consistono nel possesso di, o nella relazione con, un certo numero di oggetti (un individuo ha n figli; uno stato ha n province). Per queste proprietà la definizione operativa ordinerà di contare, direttamente o indirettamente, gli oggetti posseduti da, o relativi a, ogni caso, e di registrare il risultato del conteggio come valore di quel caso sulla relativa variabile.

B) Altre proprietà possono esser pensate come un *continuum*, sul quale ogni caso occupa una data posizione, corrispondente a un numero reale (termine con cui i matematici designano i numeri con infinite cifre). Poiché non è possibile registrare infinite cifre nella matrice dei dati, la definizione operativa dovrà dividere il *continuum* in categorie discrete, e assegnare a ciascun caso, sulla variabile che corrisponde alla proprietà in questione, il valore che corrisponde alla categoria in cui chi compie le operazioni concrete giudica che il caso vada a cadere.

C) Per tutte le altre proprietà, la definizione operativa dovrà dividere l'estensione del relativo concetto in categorie (ad es., dividere l'estensione del concetto 'tipo di occupazione' in un certo numero di gruppi di occupazioni) e assegnare a ciascun caso il valore che corrisponde alla categoria alla quale chi compie le operazioni concrete giudica che il caso appartenga.

Il modo A si può denominare conteggio. Il modo B è noto come misurazione, ma a rigore consiste in una classificazione, più una misurazione più una classificazione (vedi sezione 4.2). Il modo C è noto come classificazione, ma a rigore consiste in una classificazione, seguita da una classificazione. Il perché lo vediamo subito nella sezione che segue.

4.1. La classificazione

Il termine 'classificazione' è usato per designare numerosi concetti, di cui i tre principali sono tanto importanti per la presente trattazione che riteniamo opportuno distinguerli anche a livello terminologico, per evitare ambiguità. Piuttosto che proporre dei neologismi, ho preferito provvedere con le lettere a, b e c sottoscritte, come spesso si fa nei testi di logica.

Intenderemo per *classificazione*, l'operazione intellettuale con cui l'estensione di un concetto è divisa, rispettando alcune regole che vedremo, in un certo numero di classi o categorie. Intenderemo per *classificazione*, l'elenco di tali classi, cioè il risultato della classificazione. Intenderemo per *classificazione*, il procedimento con cui ogni singolo oggetto di un insieme è assegnato ad una delle classi della classificazione.

Per etichettare il concetto che qui viene designato con 'classificazione', lo psicometrico Stevens ha proposto (1946) i termini 'scala nominale', che hanno incontrato una fortuna universale, pur comportando una *contradictio in adjectu*. Infatti, secondo il proponente, e l'uso generale, si ha una scala nominale quando tra i 'nomi' (cioè le classi) intercorrono differenze soltanto qualitative, non quantitative. Ma questa condizione contraddice il concetto di scala, nel quale è implicito un dislivello quantitativo fra i gradini. E poiché le scelte terminologiche sono raramente prive di conseguenze sostanziali, molti ricercatori si sono sentiti autorizzati ad usare anche sulle 'scale nominali' strumenti statistici adatti alle scale vere e proprie, con risultati tragicomici³.

In seguito, Stevens ha portato avanti l'opera, ricomprendendo sotto il termine 'misurazione' anche il concetto qui designato con 'classificazione': « misurazione è l'assegnazione di numeri ad oggetti o eventi secondo regole; ogni genere di regole » (1951, 22). Coombs gli ha fatto eco: « La misurazione nelle sue forme più semplici consiste nel sostituire simboli o nomi agli oggetti concreti » (1953, 473). Né miglior sorte ha avuto il concetto che qui designiamo con 'classificazione': « un linguaggio di misurazione definisce classi di fenomeni prevedendo specifici criteri per decidere se un'osservazione può essere assegnata a una certa classe » (Przeworski e Teune 1970, 93).

Questo inverosimile stiramento del termine 'misurazione' non può che provocare confusione concettuale, con le conseguenze esemplificate nella nota 5. Esso testimonia un penoso scimmiettamento delle scienze fisiche, ed è giustamente stigmatizzato dai maggiori filosofi di quelle scienze (Carnap 1950, 9; Hempel 1952, 58; vedi anche Ellis 1968, 39-41; Torgerson 1958, 14-19).

Torniamo alle regole da seguire in una classificazione, e quindi ai requisiti di una classificazione; secondo la logica classica, essi erano tre: esaustività dell'insieme delle categorie, mutua esclusività delle categorie, *fundamentum divisionis*.

La prima regola dice: ogni caso dev'essere attribuito a una classe; la seconda dice: nessun caso dev'essere attribuito a più di una classe; la terza dice: l'attribuzione deve basarsi su un unico criterio.

Per capire meglio la terza regola, consideriamo un esempio in cui essa è clamorosamente violata: questa classificazione, degli animali che Borges dice di aver trovato in un'enciclopedia cinese: appartenenti all'imperatore / imbalsamati / addomesticati / porcellini di latte / sirene / cani randagi / favolosi / inclusi nella presente classificazione / che si scuotono come pazzi / innumerevoli / disegnati con un pennello fine di pelo di cammello / etcetera / che hanno appena rotto l'anfora dell'acqua / che visti da lontano sembrano mosche. Sembra che il fine classificatore cinese (o magari lo stesso Borges) si sia divertito a cambiare criterio ad ogni voce della classificazione. Naturalmente, anche la seconda regola (mutua esclusività delle categorie) è violata ad ogni piè sospinto (un animale può essere addomesticato e appartenere all'imperatore, e così via), anzi programmaticamente, dato che una categoria (inclusi...) ricomprende tutte le altre. Invece, non è violata la prima regola (esaustività) perché tutti gli animali che non entrassero in alcun'altra classe sono ricompresi nella classe 'etcetera'.

In una ricerca, è la definizione operativa che garantisce la mutua esclusività delle categorie, vietando l'attribuzione a un caso di più di un valore sulla stessa variabile. Naturalmente, essa fissa anche dei criteri per la scelta del valore da assegnare in ogni situazione, ma non bisogna dimenticare che in pratica la scelta tocca a chi compie concretamente l'assegnazione, che assai spesso non è il ricercatore, ma un intervistatore o un codificatore. A seconda della tecnica con cui si raccolgono i dati, e a parte gli errori materiali, il margine di scelta può essere nullo (come quando si copia da una fonte ufficiale), ridotto (come quando le possibili risposte sono pre-codificate), o ampio (come quando si devono interpretare le risposte a domande 'aperte', le reazioni a figure, i significati di un discorso o documento ufficiale).

Più ampio è il margine di scelta, più è probabile che lo stato di un caso su una proprietà appaia assegnabile con pari fondatezza a due categorie. Ma una scelta dev'essere comunque compiuta; per questo, la norma « se sei proprio indeciso fra due categorie, fa' testa e croce » può fare legittimamente parte — per quanto ciò possa apparire paradossale — di una definizione operativa, funzionando come (nell'espressione dei giuristi) 'norma di chiusura dell'ordinamento'. Tra l'altro, i vantaggi di una soluzione del genere non sono solo formali (garantire che per ogni caso si arrivi a una decisione). Supponiamo infatti che le categorie di una classificazione siano 7; se l'intervistatore (o il codificatore) è incerto fra 2, vuol dire che a suo giudizio l'individuo considerato può appartenere a ciascuna delle due categorie, ma non alle altre 5. E pertanto preferibile che venga assegnato dalla sorte a una delle due cui può appartenere piuttosto che essere dato come 'inclassificabile' o 'altro', cioè egualmente estraneo a tutte e 7 le categorie previste.

Ad ogni modo, categorie residuali come 'inclassificabile' e 'altro' sono voci legittime di una classificazione; in loro assenza, il requisito di esaustività del complesso delle categorie potrebbe rendere oltremodo arduo il procedimento di classificazione, data la grande varietà dei fenomeni concreti, e provocherebbe molte distorsioni nella fase di classificazione. Altrettanto legittime sono le categorie che garantiscono l'esaustività anche quando non è rilevabile alcuno stato del caso sulla proprietà in questione, o perché mancano le informazioni nella fonte ('manca il dato') o perché il soggetto 'non sa', 'non risponde', non si sottopone al test, etc.

Di solito questi 'dati mancanti' sono poi esclusi dalle analisi e dimenticati.

³ Un caso clamoroso è la seguente equazione di regressione multipla, presentata dal noto 'metodologo' americano Rummel (1967, 203): stile sistemico = $2,15 + .36$ (status costituzionale) - $.18$ (carattere rappresentativo) - $.13$ (sistema elettorale) - $.12$ (libertà dell'opposizione di gruppo) - $1,83$ (regime non comunista) + $.31$ (leadership politica) - $.22$ (distribuzione orizzontale del potere) - $.31$ (tipo monarchico) + $.02$ (partecipazione militare) + X (errore statistico). Rummel non si preoccupa di spiegare come ha 'misurato' queste variabili; per la maggior parte esse sembrano assolutamente non-misurabili. Il lettore esperto scoprirà inoltre altre sciocchezze metodologiche in questa equazione, che sembra un compendio delle forme di nonsenso cui conduce l'ideologia della quantificazione.

Eppure, osserva Linz, anche « le mancate risposte... possono essere indizio di atteggiamenti ben definiti, e discernibili mettendole in relazione con altre variabili » (1969, 103). Se i dati mancanti risultano troppi, se ne può derivare la conclusione che qualche segmento della definizione operativa è difettoso (domanda mal redatta, o mal posta dagli intervistatori, fonte definita in modo impreciso, etc.), oppure che la proprietà che si sta registrando (ad esempio: l'opinione su un problema) è estranea a molti dei soggetti (che non hanno alcuna opinione: vedi sul punto le osservazioni di Converse 1964 e 1970). Invece, non è possibile in linea di principio attribuire alcun significato alla categoria residuale, che aggrega sotto la stessa etichetta gli stati più diversi, con l'unica caratteristica comune di non essere stati specificamente previsti in sede di classificazione. Se chi esegue la classificazione deve assegnare una proporzione troppo alta di casi alla categoria residuale, è inevitabile concludere che la classificazione è stata operata in base a un'insufficiente conoscenza della, o riflessione sulla, realtà.

Non prevedere categorie specifiche per degli stati molto diffusi nella realtà, e quindi destinare troppi casi alla categoria residuale, è quindi un modo di operare male una classificazione. Ce ne sono altri: per esempio, prevedere un numero troppo alto di classi, che recepiscono differenze di stati che sono praticamente irrilevanti per lo scopo della ricerca. Una classificazione inutilmente lunga appesantisce il lavoro di chi esegue la classificazione, accrescendo la possibilità di errori, materiali e non. È possibile anche cadere nel difetto opposto, cioè prevedere un numero troppo limitato di classi, in ciascuna delle quali vengano aggregati stati troppo disomogenei dal punto di vista che interessa la ricerca. Non bisogna dimenticare, infatti, che se a due casi è stato assegnato lo stesso valore su una variabile, agli effetti della ricerca essi risultano avere esattamente lo stesso stato sulla relativa proprietà. Nel costituire ciascuna categoria, chi opera la classificazione, deve pertanto preoccuparsi che i casi che ad essa verranno presumibilmente assegnati presentino, quanto al loro stato sulla proprietà in questione, la massima omogeneità fra loro, oltre che la massima differenza dai casi che verranno assegnati alle altre categorie.

Un requisito che può sembrare importante dal punto di vista logico, l'uguale livello di astrazione delle categorie, porrebbe invece gravi problemi; infatti è violato assai di frequente, come osserva Thurstone (1931a, 252), senza alcuna conseguenza negativa a condizione che la mutua esclusività sia rispettata (cioè che una delle categorie previste non ne includa un'altra). Vediamo meglio con un esempio. Supponiamo che unità sia l'elettore, ambito sia l'Italia, tecnica di raccolta sia l'intervista, e la proprietà che interessa sia l'intenzione di voto nelle prossime elezioni politiche. Scendendo nella scala di astrazione possiamo introdurre una prima distinzione fra quelli che hanno intenzione di votare e quelli che non hanno tale intenzione. Bisogna prevedere una terza possibilità: che l'intervistato non ci dica se ha o non ha intenzione. Con questa terza categoria, la classificazione è esaustiva, e non c'è bisogno di prevedere categorie residuali.

Ma è probabile che il ricercatore ritenga povera l'informazione che così otterrebbe, e scenda ancora nella scala di astrazione, suddividendo la prima categoria fra quelli che hanno intenzione di votare per un partito di governo e quelli che voteranno per un partito di opposizione. Se dovesse soddisfare il requisito del pari livello di astrazione, il ricercatore dovrebbe suddividere anche le altre due categorie ('non ha intenzione di votare' e 'non dice se ha intenzione o no'); ma la cosa si presenterebbe piuttosto problematica, oltre che scarsamente significativa. Ecco quindi un dislivello fra le categorie, che non crea alcun problema pratico, purché naturalmente

si rimuova dalla classificazione la categoria 'ha intenzione di votare', che è stata suddivisa. In questa veste, la classificazione presenterebbe però un altro problema: non sarebbe più esaustiva, perché non prevederebbe una categoria per quelli che, pur dichiarando l'intenzione di votare, non manifestassero la loro scelta fra governo e opposizione. Distinguere fra costoro e quelli che non dicono se voteranno o no può sembrare inutile, e le due categorie possono essere unificate ('non dice se voterà oppure non dice se voterà per il governo o l'opposizione') malgrado il fatto che, a rigore, appartengano a due diversi livelli di astrazione.

Per ottenere informazioni più interessanti, il ricercatore potrebbe scendere un ulteriore gradino nella scala di astrazione, e prevedere una categoria per ciascuno dei maggiori partiti, più le due categorie 'non ha intenzione...' e 'non dice se...' e una categoria 'altri partiti'. Si è detto sopra che tali categorie residuali sono prive di significato perché aggregano ogni genere di stati eterogenei. Ma si può ridurre l'eterogeneità, e quindi recuperare significato, prevedendo tre categorie residuali (altri partiti di sinistra / di centro / di destra) e specificando nella definizione operativa a quale di queste categorie dovrebbe essere assegnato ciascun gruppo politico o partito minore.

Oltre alla rilevanza delle divisioni operate per la ricerca in atto, nel costruire una classificazione, si deve tener conto anche del numero dei casi che prevedibilmente saranno assegnati a ciascuna classe. Supponiamo che unità sia l'individuo, ambito il mondo, tecnica di raccolta la consultazione di fonti ufficiali, e proprietà la confessione religiosa. Una classificazione potrebbe essere: cattolici / protestanti / ortodossi / musulmani / buddisti / brahmanisti / confucisti / scintoisti / animisti / altre confessioni / senza religione. Le prime tre categorie sono specificazioni di una categoria (cristiani) che sta allo stesso livello di astrazione della maggioranza delle altre. La categoria 'animisti' sta invece a un livello più alto, in quanto aggrega culti caratterizzati da differenze almeno altrettanto grandi di quelle fra cristiani, musulmani, buddisti, etc. Allo stesso livello può essere considerata la categoria residuale, e ancora più in alto la categoria 'senza religione', che si contrappone al complesso delle altre. Tuttavia, in un atlante o una fonte simile si troverà proprio una classificazione a più livelli come quella qui riportata, perché se si riducessero i dislivelli riaggregando tutte le confessioni cristiane e introducendo le opportune divisioni fra i culti animisti, si creerebbero categorie con centinaia di milioni di casi accanto ad altre con poche migliaia.

4.2. Conteggio e misurazione

L'espressione 'scala nominale', che è stata criticata nella sezione precedente, è parte di una classificazione dei livelli di misurazione proposta da Stevens nel 1946, e tuttora generalmente adottata dai testi di metodologia delle scienze sociali. Fatto a nostro avviso assai lamentabile, perché l'autore della classificazione era uno studioso di psicofisica (tempi di reazione agli stimoli visivi, e simili proprietà del sistema nervoso umano), largamente ignaro dei problemi che presenta la misurazione dei fenomeni sociali. Di conseguenza, i criteri scelti per dividere le categorie non rispecchiano i punti in cui le effettive operazioni di misura nelle scienze sociali mutano in modo significativo; alcune categorie presentano inoltre difetti logici. Le 'scale ordinali' in senso stretto non vengono praticamente mai utilizzate nelle scienze sociali; le 'scale cardinali' scavalcano la fondamentale distinzione fra proprietà discrete e proprietà continue, quindi fra misurazione e conteggio; peggio, le 'scale a

intervalli' . . . sono logicamente distinguibili dalle cardinali, e le scale nominali' sono una contraddizione in termini. Critiche analoghe se non più radicali valgono per tutte le varianti e complicazioni della classificazione di Stevens proposte da Coombs (1953), Coombs, Raiffa e Thrall (1954), Abell (1968), etc. Si può dire che tutta quella branca di studi che va sotto il nome di 'teoria della misura', con il suo culto per la logica simbolica e la formalizzazione (si vedano Pfanzagl 1968; Krantz *et al.* 1971; Conti 1972) è priva di rilevanza per i problemi pratici di misurazione dei fenomeni sociali.

Nel giustificare questo complesso di affermazioni abbastanza eterodosse, cominciamo con le scale cardinali (nella classificazione di Stevens, *ratio scales*, scale di rapporti). Esse hanno due requisiti: esistenza di un'unità di misura, e presenza di un punto zero, corrispondente all'assenza della proprietà studiata. Stevens, e molti altri con lui (la sola eccezione a me nota è Galtung 1967, 73-75) non distinguono fra unità di misura naturale e unità di misura convenzionale; ma questa differenza è essenziale dal punto di vista delle procedure operative, e corrisponde alla fondamentale distinzione dei matematici fra discreto e continuo.

Una proprietà è discreta quando consiste nel possesso (in senso lato) di un certo numero di oggetti: automobili, figli, abitanti, partiti al governo, etc. (vedi il punto A nella sezione 4.0). In tal caso, l'unità di misura non è definita convenzionalmente, ma esiste in natura, ed è appunto la stessa unità i cui multipli costituiscono la serie dei numeri interi, che non a caso si dicono 'naturali'. La definizione operativa, nella parte specifica alla variabile in questione, è molto semplice: 1) conta (o chiedi ad altri di contare) gli oggetti di quel tipo posseduti dal caso ennesimo, e 2) registra nella matrice il risultato del conteggio come il valore di quel caso sulla variabile. Il primo atto (contare) è molto diverso da un atto di misurazione (vedi oltre). Il secondo atto non ha assolutamente niente a che vedere con la misurazione; esso è un atto di classificazione, in quanto assegna un oggetto a una categoria prefissata (uno dei numeri interi). L'insieme delle categorie prefissate (cioè la classificazione) è la serie di numeri naturali; essendo naturale, tale classificazione non è stata stabilita attraverso una classificazione, del cui intervento non c'è, pertanto, bisogno ogniquale volta registriamo esattamente il risultato di un conteggio. A mio avviso, ci sono sufficienti motivi per considerare il conteggio una forma specifica di assegnazione di valori, distinguendolo dalla misurazione con cui ha in comune solo il fatto che la misurazione usa spesso la stessa classificazione (la serie dei numeri naturali). A rigore, si dovrebbe parlare di conteggio + classificazione; ma si può evitarlo perché la classificazione è l'inevitabile atto conclusivo di ogni processo di assegnazione di valori.

La classificazione, interviene ogniquale volta la definizione operativa prevede che i risultati del conteggio non siano registrati direttamente, ma subiscano un'elaborazione di qualche genere. Ad esempio, quando il numero di abitanti, o di voti dati a un partito, in una provincia, dopo essere stato prelevato da una fonte viene registrato, ad esempio, in migliaia. In tal caso la classificazione non è più 'naturale', perché la classificazione interviene a stabilire le nuove categorie (si registra in centinaia o migliaia, o magari in milioni?) e le regole di arrotondamento. Un'altra forma di elaborazione si ha allorché i risultati di un conteggio sono divisi per i risultati di un altro conteggio, come quando si registra la percentuale di voti dati a un partito sul totale dei voti validi. Il confronto fra due proprietà discrete produce una proprietà continua, che assume un numero infinito di cifre; la classificazione dovrà quindi intervenire a stabilire se e quante cifre decimali si registrano, e le regole di

arrotondamento. Certe proprietà possono essere costituite dai risultati di un conteggio divisi per quelli di una misurazione (abitanti per chilometro quadrato), o viceversa (metri quadrati di spazio abitabile pro capite); in tutti questi casi, la classificazione dei possibili valori della variabile non è naturale, ma è stabilita attraverso una classificazione.

Se la sfera di pertinenza del conteggio sono le proprietà discrete, la sfera della misurazione sono le proprietà continue, cioè quelle che variano non 'per quanti' (un figlio, due figli..., un voto, due voti...), ma per incrementi infinitesimali. In tal caso, l'unità di misura non esiste in natura, ma dev'essere definita convenzionalmente: per la lunghezza si usa il metro, ma si potrebbe anche usare un'altra unità (infatti altri popoli usano la yarda). Questa operazione di associare un'unità di misura a una proprietà continua è il necessario preliminare di ogni misurazione; la chiameremo pertanto misurazione. Le operazioni con cui l'unità di misura scelta per la proprietà X viene confrontata con l'ammontare della proprietà X posseduto dai vari casi sono atti che chiameremo di misurazione. L'elenco dei possibili esiti della misurazione — che sarebbe una forma di classificazione — non viene mai esplicitato, perché non ce n'è bisogno; tuttavia, il confine fra esiti ammessi ed esiti esclusi viene stabilito in via generale ad opera di una classificazione.

L'intervento della classificazione nei processi di misurazione è reso necessario dall'ovvia impossibilità di registrare tutte le infinite cifre di un numero reale in una matrice dei dati (o in qualsiasi altro luogo). Per questo motivo parlare di 'variabile continua', come si fa correntemente nelle scienze sociali (persino per variabili con soli 4 o 5 valori), è un'altra *contradictio in adjectu*. Anche quando le proprietà sono continue (cioè variano per incrementi infinitesimali), le variabili che le registrano non possono che essere discrete (perché variano 'per quanti', dove il quantum è la minor differenza possibile fra due numeri registrabili nella matrice dei dati).

Ogni confronto fra un'unità di misura e l'ammontare della relativa proprietà posseduto da un caso concreto, cioè ogni atto di misurazione, ha come esito un quoziente che è un numero reale: uno non ha mai 18 anni esatti, ma 18 anni, tot mesi, tot giorni, tot ore, tot minuti, tot secondi, tot decimi, all'infinito. Nella vita di ogni giorno, come nella scienza, usiamo per brevità solo le cifre più rilevanti del numero reale. Nel dire la propria età, il diciottenne ha la scelta fra « 18 anni », « 18 anni e 5 mesi », o magari « 18 anni e mezzo »; nell'attività scientifica, invece, il numero di cifre da registrare, e le regole di arrotondamento, non possono mutare da un caso all'altro della stessa ricerca. Fissarle è un compito della classificazione. Stabilire che lo spazio abitativo di cui dispone una famiglia si misura in metri quadrati, con arrotondamento al metro più vicino, comporta che — per esempio — ogni volta che la misurazione dà un esito intermedio fra 85,5000...1 e 86,4999...9, verrà registrato il valore 86. In altre parole, il continuum della proprietà 'spazio abitativo' è suddiviso in classi di ampiezza uguale. Il complesso di tali classi, a ognuna delle quali corrisponde un numero registrabile, è una classificazione. La misurazione, confrontando lo spazio abitabile a disposizione della famiglia Rossi con l'unità di misura 'metro quadrato', dà come esito un numero reale: ad esempio 79,832... Infine, la classificazione cerca quale dei numeri registrabili corrisponde al numero reale 79,832..., e registra 80 come valore del caso 'famiglia Rossi' sulla variabile 'spazio abitativo'.

Questa analisi ci ha permesso di specificare meglio quanto si era anticipato alla fine della sezione 4.0, e cioè che la 'misurazione' si scompone in quattro distinte operazioni: si stabilisce un'unità di misura (misurazione), si decide quante cifre

registrare come arrotondare (classificazione), si confronta l'unità di misura con l'ammontare della proprietà nel singolo caso (misurazione), e si trasforma l'esito della precedente operazione nel numero registrabile corrispondente (classificazione).

La prima operazione è di gran lunga la più ardua: a stabilire una plausibile unità di misura di certe proprietà si è arrivati, anche nelle scienze fisiche, dopo decenni, magari secoli di tentativi. Talvolta la difficoltà maggiore stava nel passo intellettuale preliminare, cioè nel concepire una data proprietà come qualcosa che assume stati quantitativamente diversi. La cosa è ovvia se la proprietà consiste nel possesso di oggetti distinti, oppure è assoggettabile a operazioni additive, come la lunghezza, il peso. Ma per altre proprietà, è stato necessario un lungo sviluppo concettuale: un pittore medioevale avrebbe sicuramente escluso che il colore variasse in modo quantitativo; perché ciò fosse concepibile, si dovettero attendere gli studi sulla lunghezza d'onda (Kemeny 1972, 204).

Nelle scienze sociali siamo appena agli inizi di questo sviluppo concettuale, nel duplice senso che percepiamo variazioni non-quantitative anche laddove forse le variazioni sono quantitative, e soprattutto che, anche se percepiamo variazioni quantitative, quasi mai riusciamo a metterci d'accordo su un'unità di misura. Abbiamo numerose 'scale cardinali', come le definisce Stevens; ma si tratta in massima parte di a) risultati di conteggi, b) variabili misurate con unità prese a prestito dalle scienze fisiche (età, anzianità sul lavoro, territorio di una provincia, estensione della rete ferroviaria di uno stato, etc.), c) variabili ottenute attraverso manipolazioni matematiche di uno o più risultati di conteggi e/o una o più variabili del tipo b (alcuni esempi sono stati presentati qualche paragrafo sopra). Per il tipo a), non si può parlare di misurazione; per i tipi b) e c), si può parlare di misurazione *derivata*, cioè che usa unità di misura proprie di altre variabili (Campbell 1928).

In psicologia, peraltro, da oltre un secolo è iniziata la ricerca di forme di misurazione non-derivata. Utilizzando in parte ricerche del fisiologo Ernst Weber sul potere di risoluzione (la capacità di discriminare fra stimoli sensoriali di diversa intensità), Gustav Theodor Fechner giunse alla conclusione che la 'differenza appena percettibile' fra due stimoli era l'unità di misura appropriata per alcune proprietà psicofisiche. Altri psicologi, come Titchener e E. L. Thorndike, tentarono in seguito di estendere i risultati di Fechner ad altre proprietà psicofisiche (vedi Torgerson 1958, 132-153; Boring 1961, 112-15; Frudà 1975, 57-8).

Solo dopo la metà degli anni venti si fece strada l'idea che anche proprietà meramente psichiche come i valori e gli atteggiamenti potevano essere concepite come dei *continua* sui quali ogni individuo aveva una posizione, rilevabile attraverso una serie di stimoli (domande) collocati anch'essi lungo quel *continuum*. Il problema era stabilire in modo non arbitrario la posizione degli stimoli. Lo psicologo Thurstone cominciò col proporre una tecnica per determinare tale posizione chiedendo ad ogni soggetto di compiere tante scelte quanto erano le possibili coppie di stimoli ed elaborando tali scelte con criteri derivati dal principio fechneriano della 'differenza appena percettibile' (1927a; 1927b). Ma questa tecnica era troppo laboriosa per essere applicata su vasta scala. Thurstone abbandonò l'idea di determinare a posteriori la posizione degli stimoli, e propose un'altra tecnica che divideva il *continuum* di un atteggiamento (pro o contro un dato oggetto) in dieci intervalli supposti uguali, usando poi la mediana dei giudizi di un folto gruppo di esperti per collocare un certo numero di stimoli su ciascuna delle undici posizioni che delimitavano gli intervalli. Una volta collocati, gli stimoli potevano essere usati in ricerche su larga scala; giustamente Thurstone annunciò questo sviluppo con un articolo dal

titolo trionfale: *Attitudes Can Be Measured* (1928a). Questa tecnica, nota come E.A.I., *Equal Appearing Intervals*^{*}, sarà illustrata nella sezione 5.2.

Molti rilutteranno all'idea di unità di misura costruite con l'intervento di fattori così 'esterni' e mutevoli come i giudizi di esperti. Ma — è stato fatto notare — anche in biologia varie proprietà di una vitamina o di un enzima sono misurate attraverso le reazioni di cavie, topi e altri esseri viventi (Scates 1937, 29). Quanto alla mutevolezza, se si allude alle differenze di giudizio fra esperti nella stessa ricerca, la tecnica E.A.I. riduce al minimo gli effetti di tali differenze in vari modi che vedremo; se invece si allude a differenze culturali fra esperti di luoghi e tempi diversi, riteniamo che sia abbastanza utopico, e forse anche epistemologicamente sbagliato, cercare un'unità di misura immutabile per concetti che cambiano di contenuto nel tempo e nello spazio. La proposta di Thurstone è una soluzione adeguata alla situazione epistemologica delle scienze sociali, perché mostra come costruire strumenti di misura assai attendibili per una vasta gamma di proprietà degli individui di una qualsiasi cultura, proprio utilizzando la conoscenza che di quella cultura hanno esperti tratti dal suo seno. Per questi motivi, riteniamo pienamente giustificato, se non addirittura riduttivo, il giudizio di Allport: « l'evento più importante nella storia della misurazione degli atteggiamenti è stato l'applicazione dei metodi psicofisici da parte di Thurstone » (1967, 10).

Nelle successive sezioni illustreremo altre tecniche per misurare gli atteggiamenti proposte in seguito, e che segnano comunque un regresso, dal punto di vista epistemologico, rispetto alle proposte di Thurstone. Tale opinione è condivisa, malgrado gli orientamenti epistemologici assai differenti, da un'autorità in fatto di misurazione: « Non è stato immaginato alcun metodo migliore per costruire una scala a intervalli che misuri gli atteggiamenti » (Coombs 1953, 521).

4.3. Sulle cosiddette 'scale a intervalli'

Coombs colloca quindi le tecniche di Thurstone nella categoria 'scale a intervalli', proposta da Stevens (1946); una categoria di cui non c'era traccia nel precedente trattato classico in fatto di misurazione (Campbell 1928). Sembra più che probabile che Stevens sia stato indotto a introdurla proprio dal fatto che Thurstone e la sua scuola avevano cercato un'unità di misura degli atteggiamenti che definisse intervalli uguali su un *continuum*, mentre il problema di trovare un punto zero aveva ricevuto minore attenzione (Thurstone 1928b). La nuova categoria è quindi caratterizzata dall'uguaglianza degli intervalli (e di conseguenza da un'unità di misura che permetta di stabilire tale uguaglianza) e dall'assenza di un punto zero — assenza che la distinguerebbe da una 'scala cardinale'. Ma « la differenza fra scala a intervalli e scala cardinale è puramente teorica, in quanto è molto difficile trovare una scala a intervalli che non sia anche una scala cardinale... una volta stabilita l'unità di misura, è possibile praticamente in ogni caso immaginare un soggetto la cui misura sia di zero unità » (Blalock 1970, 32-33). La tesi di Blalock risulta confermata in modo lampante se analizziamo gli esempi di 'vera scala a inter-

^{*} La traduzione corrente (intervalli apparentemente uguali) è inesatta, perché fa pensare « apparentemente, ma non realmente », significato estraneo al termine inglese *appearing*. In questo caso va meglio la traduzione letterale: 'intervalli che appaiono uguali' (si ignora se lo sono effettivamente).

valli' rit (ente citati nei testi di metodologia, e cioè le scale) misurare la temperatura, nelle quali esiste un'unità di misura (il grado) ma il punto zero è arbitrario, in quanto non corrisponde affatto all'assenza della proprietà. Tuttavia, proprio come dice Blalock, è possibile immaginare (e pur con grandi difficoltà tecniche realizzare con buona approssimazione) un punto zero assoluto ($-273,15^\circ$ centigradi).

La differenza fra scale cardinali e scale a intervalli si rivela solo apparente anche se analizziamo da vicino un altro aspetto spesso richiamato dai testi di metodologia, e cioè il fatto che è corretto calcolare la differenza fra due valori di una scala a intervalli, ma è scorretto calcolare il loro rapporto: non si può dire che « 20° centigradi è un caldo doppio di 10° » (Galtung 1967, 74). Il rapporto è calcolabile solo fra i valori di una scala cardinale, che non a caso Stevens ha chiamato *ratio scale* (scala di rapporti). Tuttavia, è legittimo calcolare il rapporto fra due intervalli su una scala a intervalli: ha senso dire: « la differenza fra 30° e 10° è il doppio della differenza fra 30° e 20° ». Se è così, anche quando calcoliamo il rapporto fra due valori di una scala cardinale, quelli che confrontiamo sono in realtà gli intervalli fra due coppie di punti; quando diciamo « 10 metri è il doppio di 5 metri » usiamo una formula abbreviata per dire « l'intervallo fra il punto misurato da 0 metri e il punto misurato da 10 metri è il doppio dell'intervallo fra il punto misurato da 0 metri e il punto misurato da 5 metri ». Quando il punto zero non c'è o è fittizio, non si può usare la formula abbreviata perché l'origine degli intervalli non è determinata in modo univoco, e quindi dev'essere menzionata. Ma al di là delle abbreviazioni verbali, le due situazioni sono logicamente identiche.

Dal punto di vista dell'analisi dei dati, poi, la presenza o meno di un vero punto zero è del tutto irrilevante: tutte le tecniche statistiche ignorano l'origine del *continuum*, sia essa assoluta o convenzionale, e fanno coincidere il punto zero con la media della distribuzione empirica dei punteggi, trasformando il punteggio attribuito a ciascun soggetto in uno scostamento dalla media.

In conclusione, Stevens e i suoi seguaci, di fronte al problema di differenziare le tecniche di misurazione proposte da Fechner e Thurstone dalle tecniche abituali nelle scienze fisiche, fermano la loro attenzione su un aspetto formale (esistenza o meno di un punto zero) anziché su un aspetto sostanziale, e cioè sul fatto che le prime, a differenza delle seconde, necessitano della collaborazione da parte degli oggetti misurati. Non solo gli individui studiati devono collaborare alla misurazione dei loro stati sulle varie proprietà (misurazione), ma — quel che è più importante — alcuni individui devono aver collaborato alla costruzione dell'unità di misura (misurazione_a): i pazienti che segnalano a Fechner il momento in cui percepiscono variazioni di intensità nello stimolo tattile, visivo o auditivo, gli esperti che posizionano gli stimoli di Thurstone sul *continuum*.

In certe tecniche di misurazione proposte negli anni sessanta, sono gli stessi individui studiati che stabiliscono sia l'unità di misura, sia il punto zero. Con le scale 'auto-ancorate' (*self-anchoring*), si istruisce l'interrogato a identificare con 0 il peggiore stato possibile e con 10 il miglior stato possibile su una certa proprietà (soddisfazione della propria vita, della situazione politica, etc.), e poi ad auto-assegnare una posizione al proprio stato (Cantril e Free 1962). Con i 'termometri dei sentimenti' (*feeling thermometers*) si istruisce l'intervistato a identificare con 0 il massimo di ostilità e con 100 il massimo di favore verso un oggetto (gruppo, istituzione, idea, etc.), e poi ad auto-assegnare una posizione al proprio grado di favore/sfavore per quell'oggetto. In entrambi i casi, il punto 0 definisce l'assenza della proprietà (soddisfazione, favore), e quindi si può dire che un favore auto-misurato con

60 è doppio di un favore auto-misurato con 30. Tuttavia, queste tecniche sono comunemente presentate come scale a intervalli nella letteratura. Ciò significa che il criterio discriminante al quale si fa implicitamente riferimento non è la presenza di un punto zero, ma appunto il fatto che la misurazione, sia o no possibile senza la collaborazione dell'oggetto studiato, e quindi che l'unità di misura sia fissata una volta per tutte, o possa cambiare da ricerca a ricerca (nelle ultime tecniche presentate, addirittura da caso a caso).

Trasformare la questione sostanziale del rapporto fra ricercatore e oggetto studiato in una questione formale, logicamente infondata, di presenza o meno del punto zero, è quindi uno dei modi in cui si manifesta il rifiuto, più o meno conscio, di prendere atto della diversità epistemologica fra scienze fisiche e scienze sociali, per timore che tale riconoscimento metta in forse lo status scientifico delle seconde. Questo stesso motivo spiega anche l'immediata diffusione e la grande popolarità goduta dal concetto di 'scala a intervalli' nelle scienze sociali, per l'implicita autorizzazione che esso dà ad usare statistiche metriche su variabili non misurabili con una scala cardinale, senza curarsi dell'arbitrarietà dell'operazione. Naturalmente, le tecniche faticosamente messe a punto da Thurstone per ridurre al minimo tale arbitrarietà sono state largamente sostituite da tecniche più semplici e grossolane; ma l'etichetta 'scala a intervalli' con la sua rassicurante vernice scientifica è stata usata anche quando non c'era, né si cercava, alcuna garanzia che gli intervalli fossero effettivamente uguali. A giudizio di Blalock, « la maggior parte delle tecniche sino ad oggi elaborate per la misura degli atteggiamenti danno origine a scale che solo in via di estrema approssimazione possono esser considerate a intervalli: in molti casi il risultato che si ottiene non può neppure considerarsi corretto se usato come scala ordinale » (1970, 34).

4.4. Scale ordinali o scale con categorie ordinate?

Per il classico trattato di Campbell (1928), si ha un livello cardinale di misurazione quando l'ammontare della proprietà X posseduto da un oggetto è confrontato con l'unità di misura della stessa proprietà X, e l'esito del confronto è un numero reale. Si ha invece un livello ordinale quando, in assenza di unità di misura, l'ammontare della proprietà X posseduto dall'oggetto A è confrontato con l'ammontare della proprietà X posseduto dall'oggetto B, e l'esito del confronto è un giudizio di maggiore/uguale/minore. Molti testi di metodologia definiscono 'scala ordinale' nello stesso modo di Campbell (per esempio Coleman 1964, 59; Galtung 1967, 73; Kemeny 1972, 209).

Ma, come osserva Wilson (1971), le scale ordinali così definite sono uno strumento praticamente ignoto alle scienze sociali; solo in aree di ricerca molto marginali capita infatti che i casi siano confrontati fra loro. E questo non tanto perché « le proposizioni da esse [variabili misurate a livello ordinale] derivabili... sono proposizioni sull'ordine relativo tra coppie di casi » e quindi scarsamente utili « in modelli che, come quelli attualmente usati nelle scienze sociali, tendono principalmente a concettualizzare i processi interni ai singoli casi o sistemi » (Perrone 1977, 142), quanto per due ragioni che stanno a monte:

A) Confronti fra tutte le coppie di casi sono praticamente impossibili se i casi sono più di qualche decina. Con 50 casi e 10 variabili, i confronti sarebbero $10 \cdot 50 \cdot 49/2 = 12.250$. E ben pochi ricercatori si accontentano di 50 casi e 10 variabili.

B) Se la definizione operativa di una proprietà prevede un confronto fra coppie di casi, la corrispondente variabile non occupa più una sola colonna della matrice dei dati (vedi sezione 2.3), ma tante colonne quanti sono i casi. Matrici del genere sono perfettamente legittime, ma sono appunto utilizzate solo in settori relativamente marginali, come le ricerche 'sociometriche' sui piccoli gruppi.

La definizione adottata nella gran parte dei testi presenta quindi un netto divario rispetto alla pratica della ricerca, che usa come 'ordinali' scale in cui sono le categorie, non i casi, ad essere confrontate in termini di maggiore o minore possesso della proprietà misurata. E per di più, si tratta di un confronto che soddisfa per definizione la condizione di transitività (se $a > b$ e $b > c$, allora $a > c$), che non è affatto garantita, invece, quando il giudizio di maggiore/minore viene portato sui singoli casi.

Giustamente, allora, Tufte parla (1970, 438) di « categorie ordinate », cioè disposte in ordine lungo un *continuum*. Ma se le cose stanno così, siamo di fronte al solito problema di dividere un *continuum* in un certo numero di segmenti adiacenti, ognuno corrispondente a una categoria di una classificazione. Quando abbiamo un'unità di misura, possiamo fare in modo che questi segmenti abbiano ampiezza uguale, se lo desideriamo (possiamo anche servirci dell'unità di misura per definire segmenti — per esempio classi di età — di ampiezza diversa). Se non abbiamo un'unità di misura, possiamo soltanto dare un giudizio sull'ampiezza del segmento di *continuum* corrispondente a ciascuna delle categorie ordinate, e cercare di assegnare alla serie delle categorie una serie di etichette numeriche tale che le due serie siano in rapporto monotonicamente, e per quanto possibile anche lineare.

Si ha rapporto monotonicamente quando l'ordine delle etichette numeriche non contraddice l'ordine delle categorie (supponiamo che quest'ultimo sia rappresentato dall'ordine alfabetico; la seguente assegnazione di etichette numeriche crea un rapporto monotonicamente: A=2, B=3, C=5, D=8, E=9, mentre quest'altra assegnazione ne crea uno non monotonicamente: A=1, B=4, C=3, D=8, E=6). Se l'ampiezza dei segmenti corrispondenti alle categorie è uguale, assegnando la serie dei numeri naturali (1, 2, 3, 4...) si crea un rapporto lineare oltre che monotonicamente. Ma se, per esempio, il segmento corrispondente alla categoria C è più ampio, allora l'assegnazione del primo esempio qui sopra si avvicina di più alla creazione di un rapporto lineare.

Idealmente, il giudizio sull'ampiezza dei segmenti dovrebbe essere frutto di un accordo intersoggettivo fra ricercatori esperti di quel determinato campo di studi. Ma quanto meno, esso « dovrebbe incorporare la conoscenza sostanziale che il ricercatore ha della proprietà misurata... l'assegnazione della serie dei numeri naturali alle categorie è di solito inappropriata. Essa non è, comunque, una scelta più saggia o più prudente dell'assegnazione di una qualsiasi altra serie di numeri, anzi è una scelta chiaramente peggiore se urta contro le nostre conoscenze sostanziali » (Tufte 1970, 440-1).

Le differenze fra scale cardinali, a intervalli e ordinali si riducono in fondo a un grado decrescente di sicurezza con cui conosciamo l'effettiva ampiezza dei segmenti su un *continuum*, cui corrisponde una crescente importanza delle conoscenze sostanziali del ricercatore nello stimare tale ampiezza. Quando l'unità di misura può essere stabilita senza l'intervento degli oggetti studiati, non abbiamo dubbi sull'ampiezza dei segmenti (almeno da un punto di vista formale: i 5 anni che intercorrono fra i 15 e i 20 non hanno, dal punto di vista della personalità umana, la stessa ampiezza dei 5 anni fra i 70 e i 75). Quando l'unità di misura è stabilita dagli og-

getti stessi, oppure con la collaborazione di esperti, il peso delle conoscenze sostanziali del ricercatore resta limitato, senza che per questo egli possa dirsi certo che l'effettiva ampiezza dei segmenti in cui è stato diviso il *continuum* sia correttamente riprodotta dalle etichette numeriche delle sue categorie. Quando non c'è stato alcun tentativo di stabilire un'unità di misura, tutto il peso del giudizio sull'ampiezza delle categorie ricade sulle conoscenze sostanziali di chi fa ricerca.

Se tali conoscenze sono sufficienti, e bene utilizzate, l'impiego di strumenti statistici metrici (cioè adatti a variabili misurate) è epistemologicamente meno scorretto di quanto sia stato finora sostenuto dai metodologi (del nostro avviso Abelson e Tukey 1970; Tufte 1970, 438-441). In ogni caso, purché le etichette numeriche siano assegnate creando un rapporto monotonicamente con l'ordine delle categorie, la scelta di questa o quella serie di etichette non ha effetti molto sensibili sul livello dei coefficienti che misurano le relazioni fra variabili. Questo fatto è stato provato da Labovitz, che assegnò varie serie alternative di etichette numeriche alle categorie ordinate di una variabile, riscontrando altissime correlazioni fra le serie, e correlazioni molto simili di ciascuna serie con variabili esterne (1970). In seguito, Doreian ha confermato questo risultato, mostrando però che vale finché il numero delle categorie è sufficientemente ampio; altrimenti il livello dei coefficienti statistici decresce sensibilmente, il che è un ovvio indizio di distorsione (1972, 256-71). Osservazione che ci introduce all'argomento delle sezioni seguenti.

5.0. Distorsioni introdotte classificando e misurando

Nella sezione 3.2 l'attendibilità è stata definita come la capacità della definizione operativa di registrare fedelmente gli stati degli oggetti sulla proprietà che si studia. Dopo aver valutato i problemi connessi a classificazione e misurazione, siamo in grado di esaminare in modo più approfondito la questione dell'attendibilità, e di renderci conto di come « registrare fedelmente » sia un'espressione sintetica, che sta per « registrare con il minor grado possibile di distorsione ». Poiché l'attendibilità come da noi definita è una proprietà della definizione operativa, essa non dipende dalle distorsioni causate da errori materiali commessi da chi esegue le operazioni previste, a meno che tali errori non siano prodotti o favoriti da qualche aspetto della definizione stessa.

Il grado di distorsione non può essere misurato, ma soltanto stimato, in base a considerazioni che si presentano notevolmente diverse a seconda del tipo di proprietà in questione. Se la proprietà consiste nel possesso di oggetti (tipo A nella sezione 4.0), la possibilità di distorsione può essere introdotta nel modo in cui si richiedono le informazioni, e soprattutto nella fonte dalla quale si assumono (ad esempio, se unità è l'azienda, e la proprietà che interessa è il numero di dipendenti, proprietari e dirigenti possono avere motivi per ridurre oppure gonfiare artificialmente tale numero).

Se fra gli stati della proprietà esistono differenze qualitative ma non quantitative (tipo C nella sezione 4.0), la principale fonte di distorsione sono errori commessi in fase di classificazione, cui si è già accennato nella sezione 4.1. Se non è stata prevista alcuna categoria che corrisponda a uno stato della proprietà nettamente distinto dagli altri stati, in fase di classificazione, a tutti i casi che presentano quello stato verrà attribuito uno degli stati previsti dalla classificazione (un

altro stato (anziale, oppure 'inclassificabile'). D'altra parte, la classificazione non può prevedere un numero troppo alto di categorie, perché questo aumenta le possibilità di distorsioni in fase di classificazione.

Se la tecnica di raccolta dei dati è l'intervista, le voci della classificazione vengono lette al soggetto, che manifesta il suo stato sulla proprietà scegliendo una voce come la propria risposta. Tale scelta è distorta se il soggetto non è in grado di valutare con uguale attenzione tutte le alternative. Alcuni esperimenti sembrano indicare che il numero massimo di alternative che l'individuo medio riesce a valutare contemporaneamente è sette (Miller 1956; Luce 1959, 5 ss.). Considerando però che durante un'intervista scelte del genere devono essere operate parecchie decine, talvolta centinaia, di volte nell'arco di un'ora, e che il soggetto è mediamente assai meno attento e motivato che in una situazione sperimentale, è ragionevole collocare tale limite massimo ancora più in basso, attorno alle 4 o 5 alternative, come infatti sostiene Galtung alla luce delle sue esperienze di ricerca (1967, 108). Il limite di sette è invece abbastanza prudente se la scelta avviene per scritto (e quindi il soggetto ha più tempo per riflettere), oppure se essa non è operata dal soggetto stesso, ma dal ricercatore o da collaboratori esperti sulla base dell'osservazione degli stati dei soggetti.

È sorprendente che così poca attenzione sia stata dedicata a questa limitazione nel numero di alternative contemporaneamente valutabili, e alle sue conseguenze per l'attendibilità delle risposte a un'intervista. Un metodologo polacco ha recentemente proposto una tecnica intesa ad aggirare la limitazione, dividendo le voci di una classificazione in due elenchi, e sottoponendoli separatamente all'intervistato, come possibili risposte alla stessa domanda. Il soggetto sceglie le risposte preferite in ciascun elenco, poi compie una scelta definitiva fra le alternative che hanno superato questo primo vaglio. Successive interviste in profondità hanno permesso di accertare che la scelta operata in due riprese corrispondeva assai meglio al reale stato dell'intervistato sulla proprietà di una scelta unica operata sull'intero elenco delle alternative (Kistelski 1978).

A parte queste pregevoli, ma ancora per niente diffuse, innovazioni tecniche, in fase di classificazione, il ricercatore si trova comunque di fronte al problema di ridurre una varietà potenzialmente amplissima di stati su una determinata proprietà ad un piccolo numero di categorie. Se la proprietà non interessa in sé ma come indicatore di una proprietà più generale, il compito del ricercatore può essere notevolmente facilitato per la possibilità di concentrarsi su quelle differenze fra stati che sono rilevanti per il concetto generale, e per il suo ruolo nel disegno complessivo della ricerca. Supponiamo che si intenda approfondire il classico tema weberiano dei rapporti fra confessione religiosa e motivazione al successo, svolgendo un sondaggio in un paese multiconfessionale, come l'Olanda, l'India o gli Stati Uniti. Nella classificazione, una categoria dovrà esser riservata per ognuna delle confessioni più diffuse in quel paese; ma le confessioni meno diffuse potranno esser raggruppate in un minor numero di categorie tenendo conto delle loro caratteristiche potenzialmente rilevanti per la motivazione al successo. Se le valutazioni del ricercatore su quelle caratteristiche sono fondate, la distorsione insita nell'aggregare più confessioni diverse nella stessa categoria può avere limitatissimi effetti negativi per la ricerca in questione.

Naturalmente, quella che è una buona classificazione per un ricercatore può non andare altrettanto bene per un altro. Supponiamo che quest'altro ricercatore voglia collegare la confessione religiosa con la preferenza politica. Egli non dovrebbe

adottare la classificazione, stabilita dal collega senza prima controllare se alcune categorie aggregano confessioni che invece è meglio tenere distinte allorché ci interessano i rapporti con le preferenze politiche, e se viceversa sono operabili altre aggregazioni. In altre parole, la distorsione che una classificazione può introdurre — e quindi l'attendibilità della definizione operativa — non solo non è misurabile, ma non può neppure esser valutata in assoluto, senza fare riferimento agli scopi della ricerca che la utilizza. Ciò è vero in particolare quando le categorie hanno autonomia semantica, cioè sono comprensibili anche senza far riferimento al concetto più generale, la proprietà, di cui sono specificazioni (le categorie 'cattolico', 'battista' sono semanticamente autonome dalla proprietà; le categorie 'sì', 'no', 'molto', 'poco' non lo sono, perché acquistano senso compiuto solo in riferimento a una domanda). Tornerò sul concetto di autonomia semantica nella sezione che segue.

Se la proprietà che si studia può esser considerata un *continuum* che la classificazione divide in segmenti adiacenti (tipo B nella sezione 4.0), le possibilità di distorsione sono molto diverse a seconda del modo in cui tale divisione è operata. Se ci si può valere di un'unità di misura che non presuppone la collaborazione degli oggetti misurati, problemi di distorsione possono sorgere solo se le categorie stabilite sono troppo ampie, e assegnano lo stesso valore a oggetti le cui differenze nell'ammontare della proprietà posseduta sono tali da avere un significato qualitativo. Se il *continuum* della proprietà 'età' viene diviso in segmenti di ampiezza annuale, è impensabile che le differenze fra soggetti collocati nella stessa categoria (cioè divisi da un anno al massimo, a meno di errori di arrotondamento o altri errori materiali) siano talmente significative da introdurre distorsione. Le cose stanno diversamente se l'ampiezza delle classi è quinquennale, o addirittura decennale: assegnare lo stesso valore sulla variabile 'età' a un quindicenne e a un ventenne annulla differenze che quasi certamente sono significative.

Un problema assai più grave dell'ampiezza delle categorie si pone allorché l'unità di misura è stabilita dagli stessi soggetti studiati, come in alcune tecniche ('termometri dei sentimenti', scale 'auto-ancorate') viste nella sezione 4.3. Altre tecniche del genere chiedono all'intervistato di collocare se stesso, oppure un gruppo o un personaggio pubblico, su una scala composta in genere di dieci o undici posizioni, che va da 'destra' a 'sinistra', da 'conservatore' a 'progressista', e simili. Un'altra tecnica molto usata in psicologia, il *differenziale semantico*, chiede ai soggetti di collocare un qualsiasi concetto su una serie di scale a sette posizioni che vanno da 'onesto' a 'disonesto', da 'gentile' a 'crudele', da 'caldo' a 'freddo', e così via (Osgood 1952; Heise 1969).

L'etichetta 'auto-ancorate', che Cantril ha dato alle scale da lui proposte, si presta bene a caratterizzare tutta questa famiglia di tecniche, che hanno in comune la caratteristica di stabilire il numero delle categorie, ma di affidare a ciascun soggetto il compito di definire il punto zero e l'altra estremità del *continuum*, e quindi implicitamente l'unità di misura (che sarà la distanza fra i due estremi divisa per il numero delle categorie). Di conseguenza, se due soggetti scelgono la stessa posizione (per se stessi o per l'oggetto da valutare), non abbiamo alcuna garanzia che il loro stato sulla proprietà che si vuole misurare sia lo stesso, o comunque sufficientemente simile. Traendo un esempio dalla situazione italiana, chi si colloca nella posizione 7 di una scala in cui lo 0 è etichettato 'destra' e il 10 'sinistra' può essere un socialdemocratico che si sente sufficientemente a sinistra del centro, oppure un

comunista che tiene conto della distanza fra le posizioni del suo partito e quelle di alcuni gruppi extraparlamentari.

Un'indicazione probante della diversità fra soggetto e soggetto nel modo di concepire il *continuum* si ha scorrendo dei questionari che contengono una batteria di termometri riferiti agli oggetti più diversi. Alcuni usano tutto l'arco dei punteggi, da 0 a 100, altri solo i punteggi centrali (quindi fissano gli estremi in una posizione che sembra loro 'troppo estrema'); alcuni usano di preferenza i punteggi alti, altri quelli bassi. Naturalmente, la differenza fra individui non è dovuta solo al modo in cui ciascuno fissa gli estremi, ma anche alle caratteristiche psicologiche: ottimismo/pessimismo, estroversione/introversione, sicurezza/insicurezza, etc. Qualunque sia l'oggetto da collocare sulla scala, il punteggio dato dall'individuo i può essere così scomposto: $a_{ij} = t_i + n_{ij} + e_{ij}$, dove t_i è la tendenza di i a dare punteggi alti o bassi indipendentemente dall'oggetto, n_{ij} è la valutazione che i dà dell'oggetto j , depurata di detta tendenza, e l'addendo e_{ij} è dovuto ad altri fattori contingenti.

Al ricercatore interessa n_{ij} (altrimenti non avrebbe posto quella domanda), ma è nota solo a_{ij} , la risposta manifesta. Se i termometri della batteria sono abbastanza numerosi, la media dei punteggi assegnati da i a tutti i vari termometri presenti nella batteria (che rappresenteremo con $\bar{a}_i$) può essere assunta come una ragionevole stima di t_i , cioè della tendenza di i a dare punteggi alti o bassi. Sottraendo tale stima al punteggio manifesto ($a_{ij} - \bar{a}_i = n_{ij} + e_{ij}$), ci si avvicina meglio alla vera valutazione che i dà dell'oggetto j , anche se la componente erratica e_{ij} non può essere stimata, e quindi eliminata⁷. Questa procedura potrebbe essere chiamata *deflazione*, in quanto presenta molte analogie con la neutralizzazione del livello generale dei prezzi che gli economisti operano sulle variabili monetarie. Essa si dimostra (Maradi 1979b) assai efficace nel ridurre le correlazioni spurie fra i termometri, facendo emergere soluzioni fattoriali assai meglio interpretabili (vedi sezione 6.1). Il che significa, naturalmente, che i termometri 'deflazionati' sono assai più validi come indicatori dei relativi concetti generali, perché ne è stata ridotta la parte estranea (vedi sezione 3.2).

Una procedura di deflazione è opportuna per ogni tipo di scala auto-ancorata. Tuttavia, essa è tanto meno legittima ed efficace quanto più ristretto è l'arco dei punteggi possibili (l'arco è 1-100 nei termometri, ma solo 1-10 nelle altre scale auto-ancorate più usate, e 1-7 nel differenziale semantico) e, soprattutto, quanto minore è il numero di scale dello stesso tipo riunite in batteria (perché la media dei punteggi è poco attendibile se calcolata su una ristretta base di elementi; al limite, se la scala è isolata, la deflazione non ha senso). D'altra parte, più alto è il numero delle domande in una batteria, più cresce il pericolo che le risposte siano date in maniera ripetitiva, senza prestare attenzione allo specifico contenuto della domanda. Questa forma di distorsione, la più grave e diffusa quando i dati sono raccolti tramite intervista, verrà trattata a fondo nella sezione che segue.

⁷ Si può anche tener conto non solo della tendenza a dare punteggi alti o bassi, ma anche della tendenza a usare un arco più o meno ampio di punteggi. Quest'ultima può essere rappresentata dalla deviazione standard dei punteggi dati da i (s_i), e usata poi come divisore nella formula $(a_{ij} - \bar{a}_i) / s_i = n_{ij} + e_{ij}$. In questo caso il processo è analogo alla standardizzazione, salvo il fatto che opera sulle righe della matrice dei dati (cioè sui punteggi dello stesso individuo su variabili diverse), mentre la standardizzazione opera sulle colonne (cioè sui punteggi diversi sulla stessa variabile).

5.1. Scale con categorie ordinate

Quando il *continuum* con cui si rappresenta la proprietà viene diviso in segmenti senza l'ausilio di un'unità di misura, ma semplicemente disponendo sul *continuum* stesso un certo numero di categorie, i pericoli di distorsione sono talmente acuti da meritare una trattazione a parte. Infatti, alle distorsioni implicite nell'aggregare sotto la stessa categoria stati individuali più o meno diversi, si aggiungono quelle comportate dal fatto che è il ricercatore stesso a collocare a suo giudizio la categoria in una posizione sul *continuum* nel momento in cui gli attribuisce una etichetta numerica, che poi verrà trattata dalle tecniche statistiche come se fosse un numero cardinale.

Vediamo innanzitutto che forme assume il primo tipo di distorsione in questa situazione, in cui immaginiamo la proprietà come un *continuum* sul quale ogni caso occupa una certa posizione, che però non è rilevabile direttamente perché non si dispone di un'unità di misura con cui confrontarla. Non resta quindi che stimare tale posizione indirettamente, chiedendo a ogni individuo di scegliere fra un certo numero di categorie che si suppongono disposte sullo stesso *continuum*, nell'assunto che egli sceglierà la categoria la cui posizione ritiene meno lontana dalla propria.

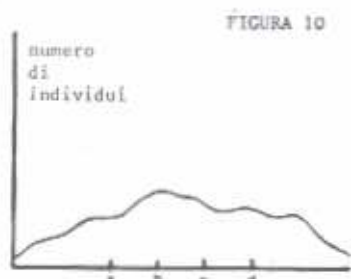
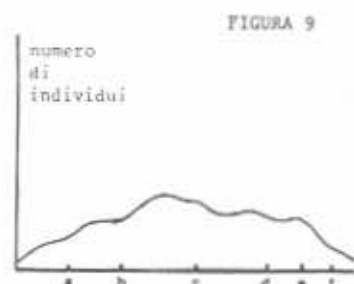
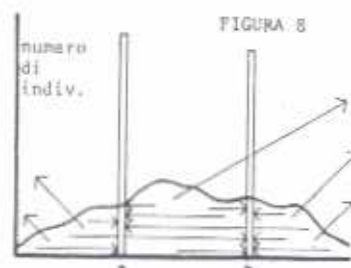
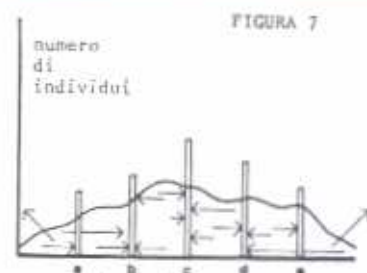
Il processo che si genera al momento in cui tale scelta si prospetta è descritto nella figura 7, in cui l'ascissa rappresenta la proprietà X come un *continuum*, e in ordinata stanno i casi la cui posizione su X si intende rilevare. Per ogni punto in ascissa, l'ordinata della curva corrisponde al numero di individui la cui 'reale' posizione su X corrisponde a quel punto. Si tratta naturalmente di una curva ipotetica, perché non possiamo effettivamente misurare la posizione degli individui. C'è inoltre da rilevare che in ascissa sta solo un'approssimazione a un *continuum*; se ci fosse un vero *continuum* le posizioni di due individui, avendo infinite cifre, non potrebbero coincidere, e quindi l'ordinata non potrebbe superare l'unità (questa osservazione vale naturalmente per tutte le curve che rappresentano distribuzioni di frequenza, anche di fenomeni fisici).

Nell'esempio della figura 7, le categorie sottoposte agli interrogati come risposte alternative fra cui scegliere sono cinque (a, b, c, d, e). Per il momento, non sono state loro assegnate delle etichette numeriche, per cui la posizione in cui sono rappresentate sul *continuum* è quella che ipoteticamente è attribuita loro dalla media degli stessi interrogati. Le frecce rappresentano qualche esempio delle distorsioni cui le posizioni degli individui vengono sottoposte nel momento in cui essi scelgono una delle cinque categorie. Non bisogna dimenticare, infatti, che l'infinita varietà di posizioni esistenti lungo il *continuum* viene condensata in cinque sole: lo spazio sotteso dalla curva si distribuisce in cinque istogrammi isolati.

Se tutti gli individui scegliessero la categoria a loro più vicina sul *continuum*, la distorsione totale sarebbe la somma delle distanze di ciascun individuo dalla categoria più vicina. In realtà, la distorsione è più alta, perché per vari motivi (cattiva interpretazione del testo delle risposte, disattenzione, altri stati momentanei) molti non scelgono la categoria più vicina. Nella figura vediamo infatti che qualcuno si trova più vicino a c ma sceglie b , un altro si trova più vicino a d ma sceglie c ; c'è anche chi, pur trovandosi oltre a , scavalca a e sceglie b , e chi, pur trovandosi oltre e , scavalca e e sceglie d . C'è infine chi, ritenendo che nessuna delle cinque risposte sia abbastanza vicina alla propria posizione, sceglie di non rispondere, e quindi risulta non avere alcuno stato sulla proprietà X (la sua freccia esce infatti dal *continuum*).

Sembra evidente che la distorsione è in linea di principio tanto maggiore quanto

più è ridotto il numero delle categorie. Il massimo di distorsione si ha allorché una proprietà rappresentabile su un *continuum* è ridotta a uno stato dicotomico, come quando si chiede agli intervistati se sono pro o contro un dato oggetto. In tal caso, le distanze fra la posizione reale di un individuo e quella attribuitagli in sede di classificazione, saranno mediamente più lunghe e, a parità di altre condizioni, un maggior numero di individui, non gradendo nessuna delle due alternative, rifiuterà di rispondere (i due fenomeni sono rappresentati dalle frecce della figura 8). Le dicotomie hanno inoltre il difetto di massimizzare le conseguenze di errori materiali commessi da intervistatori e codificatori: se i valori possibili sono solo due, un errore trasforma automaticamente lo stato di un individuo nello stato opposto.



Dal punto di vista dell'analisi dei dati, gli svantaggi delle dicotomie sono altrettanto gravi: se due proprietà continue sono registrate con due dicotomie, non è possibile distinguere se tra loro esiste una relazione curvilinea oppure lineare; in quest'ultimo caso il coefficiente di correlazione è artificialmente compresso (come si è accennato in chiusura della sezione 4.4), tanto più pesantemente quanto più la distribuzione dei casi fra le due categorie è diseguale (Doreian 1972, 257-9. Vedi anche Gangemi 1977 per i nefasti effetti di una distribuzione diseguale su tutti i coefficienti statistici relativi alle dicotomie).

Naturalmente, questi svantaggi tecnici non possono essere evitati se la proprietà è effettivamente dicotomica, come il sesso. Ma se le proprietà sono continue, registrarle con variabili dicotomiche è, a parte gli svantaggi tecnici, « la maniera più infame di censurare l'informazione » (Tuftes 1970, 441; con toni più moderati, Blalock 1963, 989). Tuttavia, si tratta di una pratica comunissima nelle scienze sociali, avallata com'è da metodologi rinomati come Davis (1971), Galtung (1967, 95-98 e 250-9) e soprattutto Lazarsfeld (1972), che ha esercitato grande influenza

con la sua decisione di lavorare esclusivamente con variabili dicotomiche nei suoi studi sul mutamento di opinione nei *panels* (1940; sui *panels*, vedi sopra, sezione 2.4) e nella tecnica statistica detta 'analisi della struttura latente' (1950).

Se riducendo il numero delle categorie si accresce la distanza media tra la posizione di un individuo e quella della categoria a lui più vicina, la naturale soluzione al problema delle distorsioni parrebbe consistere nell'adozione di un numero molto più ampio di categorie. Ma ciò urta contro la limitata capacità di valutare più alternative contemporaneamente, di cui si è detto nella sezione precedente. Se le alternative superano quel limite, l'interrogato non è effettivamente in grado di considerarle tutte quando sceglie; si produrrà quindi un maggior numero di scavalcamenti di posizione, del tipo esemplificato nella figura 7. Si potrebbe ovviare a questo con la tecnica del doppio elenco (Kistelski 1978), già presentata nella sezione precedente. Resta però il problema che, aumentando il numero delle categorie, diventa sempre più difficile — e arbitrario in assenza di una qualsiasi unità di misura — stabilire quale sia la loro giusta successione ordinale sul *continuum*.

Per inciso, la stessa difficoltà si manifesta anche quando vengono registrate le risposte spontanee degli intervistati anziché chiedere loro di scegliere fra alternative pre-codificate, e in seguito si raggruppano tali risposte in un certo numero di categorie, operando cioè una classificazione, ex post anziché preventivamente. In tal caso, accade spesso di costituire, stimolati dalla varietà delle risposte, troppe categorie, cui poi non siamo in grado di dare un'accettabile disposizione ordinale sul *continuum*. Beninteso, di fronte a una difficoltà del genere, si può sempre decidere di non ordinare le categorie; ma se si è convinti che la proprietà è correttamente rappresentabile come un *continuum*, ciò significa rinunciare a parte delle informazioni raccolte.

Non è semplice, quindi, trovare il numero ideale di categorie da ordinare, tanto più che la quantità di distorsione dipende non solo dal numero, ma anche dal grado di corrispondenza fra le posizioni delle categorie e la distribuzione delle posizioni individuali. Cosa ciò significhi sarà chiaro a chi osservi le figure 9 e 10. Nella prima, per le posizioni centrali del *continuum* è stata prevista la sola categoria *c*, che sarà quindi scelta da individui con posizioni effettive anche piuttosto distanti fra loro, e finirà per includere un numero di casi assai più alto delle altre. Nella figura 10, sono le posizioni estreme del *continuum* a non essere rappresentate, il che lascia a chi le occupa la scelta fra convergere nelle categorie *a* o *d*, oppure rifiutarsi di rispondere.

Veniamo ora alle ulteriori distorsioni introdotte al momento di attribuire alle categorie un'etichetta numerica. In questa fase la posizione individuale, che è già stata più o meno dislocata per ricondurla a una categoria, può subire un nuovo sensibile spostamento in quanto è tutta la categoria che viene spostata; tale spostamento può riavvicinarla alla sua situazione originaria, ma può anche allontanarla ulteriormente. Supponendo che la posizione reale di ciascuna categoria sul *continuum* sia definibile come la media delle posizioni reali di tutti gli individui che scelgono quella categoria, è impensabile che le categorie, per felice intuito del ricercatore nello sceglierle, risultino collocate proprio ad intervalli uguali sul *continuum*. Alcune categorie (di solito quelle centrali) risulteranno più vicine a loro di altre. Come osserva Tuftes (1970, 440, citato nella sezione 1.4.4), nell'assegnare le etichette numeriche si dovrebbe cercare di riprodurre la distanza effettiva fra le categorie, utilizzando le nostre conoscenze sostanziali e — se la proprietà in questione è usata come indicatore — tenendo principalmente presente il concetto generale.