

“Sapienza” Università di Roma
Dottorato in Metodologia delle Scienze Sociali
XXI Ciclo

Il contributo della logica fuzzy alla classificazione: oltre la cluster analysis

Simone Rambotti

L'intelligenza delle cose comincia con
la distinzione.

La grande intelligenza abbraccia, la
piccola discrimina.

(citazioni tratte da Campelli, *Tohu va-
vohu. Note non tecniche sul problema
della classificazione*)

[...] gli insiemi *fuzzy* sono reali. Gli insiemi *fuzzy* esistono fuori dalla matematica, esistono in ogni rete che gettiamo su ogni pezzo di universo. Ognuno di noi appartiene in una certa misura a ogni organizzazione, a ogni sistema ideologico, a ogni moda, o ogni mito, a ogni tendenza. Tutti aderiamo in una certa misura a ogni partito politico, a ogni orientamento sessuale, a ogni stile di vita, a ogni punto di vista in conflitto. Siamo tutti di sinistra, di destra, di centro, etero-, omo-, bi-sessuali, disinvolti, convenzionali, naturali, favorevoli, contrari e indifferenti. Possiamo anche non sapere in che misura siamo queste cose ma sappiamo che in una qualche misura siamo tutte queste cose. Il cinema, i libri, l'arte ci aiutano a vedere i vari lati di noi stessi. Cresciamo e impariamo con l'età; il che significa per l'appunto che cambiano quelle misure, che cambiano cioè i nostri *fits* (*fuzzy units*, o unità di misura *fuzzy*), allorché passiamo dall'essere una cosa a non essere quella cosa, dal credere al non credere e, per avventura, viceversa. I conflitti scoppiano quando una persona, un gruppo o un governo cerca di «arrotondarci» a suo modo, ossia cerca di renderci tutti A o tutti non-A, tenta di trasformare in nostri *fits* in *bits* e di negare le nostre sfumature di differenze.

Bart Kosko, *Il fuzzy-pensiero*

Do I contradict myself?
Very well then I contradict myself,
(I am large, I contain multitudes.)

Walt Whitman, *Song of Myself*

Indice

Introduzione	i
1. La classificazione: operazione e prodotto	1
1.1 Definizioni da vocabolario	1
1.2 La classificazione come struttura concettuale	2
1.2.1 I concetti ed il “concetto di concetto”	3
1.3 Classificazioni, tipologie e tassonomie	8
1.3.1 Classificazioni	8
1.3.2 Tipologie	12
1.3.3 Classificazioni estensionali	16
1.3.4 Tassonomie	19
1.3.5 Classificazione(c)	22
1.3.6 Altri contributi	23
1.4 La posizione epistemologica della classificazione	26
1.5 La valutazione di una classificazione	31
2. Le procedure di cluster analysis	39
2.1 La cluster analysis come arcipelago	41
2.2 Le operazioni preliminari	43
2.3 Le tecniche tradizionali	44
2.3.1 Le tecniche tradizionali gerarchiche	44
2.3.2 Le tecniche tradizionali non gerarchiche	46
2.3.3 Le strategie miste ed altre tecniche	48
2.4 Vecchi risultati e nuove proposte	49
2.4.1 Se cade l’esaustività: gli outliers	51
2.4.2 Se cade la mutua esclusività: l’overlapping clustering	51
2.4.3 Se cade l’appartenenza binaria: il fuzzy clustering	52
2.4.4 I possibili risultati di una cluster analysis	53
2.5 La valutazione di una classificazione automatica	53
2.5.1 L’efficacia	54
2.5.2 La stabilità	56
2.6 Conclusioni	57

3. Logica fuzzy e fuzzy clustering	58
3.1 L'irruzione della logica fuzzy: "uno schiaffo alla scienza moderna"	58
3.2 Gli antecedenti della logica fuzzy	60
3.3 Fuzzy sets	68
3.4 Le critiche alla logica fuzzy	70
3.5 Fuzzy clustering	74
4. La calibration	80
4.1 La Qualitative Comparative Analysis	80
4.2 La tecnica della calibration	88
5. Un'applicazione empirica	107
5.1 Provenienza e caratteristiche dei dati	107
5.2 La componente cognitiva del pregiudizio etnico: il contesto di Roma	114
5.2.1 Cluster analysis tradizionale e semisfocata	122
5.2.2 Calibration	134
5.3 La componente cognitiva del pregiudizio etnico: il contesto del Lazio	149
5.3.1 Cluster analysis tradizionale e semisfocata	152
5.3.2 Calibration	153
5.4 La componente affettiva del pregiudizio etnico	159
5.4.1 Cluster analysis tradizionale e semisfocata	164
5.4.2 Calibration	169
5.5 La componente attiva del pregiudizio etnico	175
5.5.1 Cluster analysis tradizionale e semisfocata	179
5.5.2 Calibration	182
5.6 Una breve analisi combinata	184
5.7 Conclusione	188
Riferimenti bibliografici	190

INTRODUZIONE

Ricordo esattamente il momento in cui sentii nominare per la prima volta la parola *fuzzy*: era una lezione di Metodologia e Tecnica della Ricerca Sociale. Una delle ultime lezioni, per l'esattezza, di un duro corso tenuto dalla professoressa Agnoli. Ero un *undergraduate* piuttosto stanco, così come i miei colleghi, e questo mi spinse a non fare domande per saperne di più, ma l'idea dell'esistenza di una logica (con un nome peraltro così "strano", informale, il cui significato mi sfuggiva completamente), che affondava le sue radici culturali nel pensiero di Buddha, colpì evidentemente con grande forza il mio immaginario di ragazzo occidentale. Misi da parte le domande ma anche la curiosità, che tirai fuori quando si presentò il momento opportuno e che ho portato con me fino a questa tesi di dottorato, nella quale il tema della *fuzziness* si è incontrato con quello della classificazione nelle scienze sociali.

Il lavoro inizia con un'approfondita analisi del concetto di classificazione, con particolare attenzione ed interesse per l'impostazione data da Marradi nel corso degli anni (cfr. Marradi 1984; 1993; 2007), nella quale si sottolinea con cura la distinzione tra gli aspetti intensionali e quelli estensionali insiti nel concetto e nell'attività classificatoria stessa.

Si prosegue con una rassegna delle principali tecniche di classificazione automatica (cfr. Tryon, 1939; Tryon e Bailey, 1970; Anderberg, 1973; Ricolfi, 1992; Biorcio, 1993), ossia quella *cluster analysis* che rappresenta l'applicazione su base algoritmica, con uso di matrice dei dati, della classificazione intesa in senso prettamente estensionale.

Viene poi introdotto il tema della logica *fuzzy*, di cui viene illustrato il significato, la portata “rivoluzionaria” ed il percorso storico (cfr. Zadeh, 1965; Kosko, 1993). Segue un’analisi delle principali tecniche di *fuzzy clustering*, attraverso le quali classificazione automatica e logica sfumata trovano un punto d’incontro (cfr. Bezdek, 1981; Aureli e Middei, 1997). È su queste basi che verrà effettuata parte della nostra applicazione empirica, il cui aspetto più innovativo riguarda l’unità d’analisi scelta: se, infatti, tranne rare eccezioni (cfr. Livolsi, a cura di, 1992; Erminio, 2001) la maggior parte delle applicazioni aveva base ecologica (cfr. Brunelli, 2000; 2001), nel nostro caso utilizzeremo una matrice dati costruita grazie all’intervista di un campione di studenti di scuola superiore secondaria del Lazio (cfr. Agnoli, a cura di, 2004).

L’ultimo passo è costituito dall’incontro con il metodo della *Qualitative Comparative Analysis* sviluppato da Ragin (cfr. Ragin 1987; 2000), da cui estrapoleremo una tecnica – la *calibration* – che, accoppiata all’uso delle cosiddette *truth tables*, ci offrirà l’opportunità di arrivare ad una classificazione dei nostri dati del tutto diversa ed indipendente da quella ottenuta tramite *cluster analysis*.

È nostra convinzione che questa applicazione, basata su *fuzzy-calibration*, rappresenti un’opzione utile ed interessante, soprattutto per la capacità di sottolineare e porre di nuovo al centro dell’attenzione quegli aspetti intensionali che devono sottostare a qualsiasi classificazione: le classificazioni estensionali, infatti, “diventano possibili – o dotate di senso, il che è, nel caso specifico, del tutto equivalente – solo *in relazione* all’eventualità che sia costruita, esplicitata ed interpretata una certa articolazione dell’intensione della proprietà (o delle proprietà) utilizzate” (Campelli, 2001, p. 14), altrimenti tutto ciò che viene prodotto sono solamente “aggregati muti, pure soluzioni statistiche di un «gioco» esclusivamente formale” (*ibid.*, p. 15).

CAPITOLO 1. LA CLASSIFICAZIONE: OPERAZIONE E PRODOTTO

1.1 DEFINIZIONI DA VOCABOLARIO

L'attività del classificare è, come si usa dire, vecchia quanto l'uomo ed il pensiero stesso, tanto che essa, oltre che utilizzata, è stata anche analizzata con frequenza innumerevole:

La stretta connessione che sussiste fra pensare e classificare nelle culture occidentali e nelle loro istituzioni – nella organizzazione e rappresentazione del mondo che esse producono e riproducono – fa sì che le classificazioni siano, per le scienze storico-sociali che studiano quelle culture, prima ancora che uno strumento, uno degli oggetti di indagine più affascinanti (Agodi, 2001, p. 168)

Moltissimo si è detto; è quindi necessario fare, per quanto ci è possibile, il punto della situazione, prima di tutto a livello terminologico. Le definizioni da vocabolario insistono su un aspetto in particolare del termine, per cui classificare significa *mettere insieme oggetti che possiedono caratteristiche comuni* (ritenute discriminanti):

operazione per la quale una moltitudine di oggetti affini viene inquadrata in un sistema gerarchico di raggruppamenti (Beer, 1949, p. 1782);

divisione in classi, ripartizione [di persone o oggetti] in cui sono raggruppati fra loro gli elementi affini per uno o più caratteri (Cecchini, 1967, p. 845);

ripartizione di vari oggetti, pratiche burocratiche, schedari e simili in gruppi sistematicamente predisposti secondo particolari criteri (AA.VV., 1987, p. 397);

principio che governa l'organizzazione degli oggetti in gruppi in accordo alle loro similarità e differenze o relazioni con gruppo di criteri (AA.VV., 1993, p. 356);

distribuire in classi, o in altre suddivisioni e categorie prestabilite e ordinate secondo determinati criteri (De Felice e Duro, 1993, p. 407).

Sembrerebbe tutto abbastanza chiaro ed efficacemente sintetizzato da Biorcio in apertura del suo lavoro del 1993, nel quale “l’operazione mentale di classificazione” viene definita “suddivisione di un insieme di oggetti in classi ritenute sufficientemente omogenee al loro interno, e sufficientemente distinte fra loro” (Biorcio, 1993, p. 7). Ma naturalmente non è tutto così semplice e, se Biorcio da un lato fa un passo avanti (che più avanti verrà ampiamente ripreso) con l’espressione “omogenee al loro interno e distinte fra loro” che suggerisce la parola *cluster*, dall’altro ci impone di farne uno indietro, con l’accenno all’operazione mentale.

Infatti, a fronte di una sostanziale concordanza sul significato della “classificazione” rinvenibile nei dizionari e nelle enciclopedie (la scienza dei manuali, direbbe Fleck – cfr. Fleck, 1983), una rapida analisi della letteratura scientifica (la scienza delle riviste) rompe questa unità ed impone alcune distinzioni, a cominciare da quella tra operazioni e prodotti: quando si parla di classificazione ci si riferisce ad entrambi e va fatta attenzione a non confondere le prime con i secondi. Ad esempio, riprendendo le parole di Biorcio, l’“operazione mentale” corrisponde alla procedura, mentre i prodotti sarebbero le classi. Ulteriori distinzioni sono possibili all’interno del panorama dei prodotti e delle operazioni, fermo restando che le distinzioni sono spesso utili a livello euristico ma altrettanto spesso confuse nella realtà.

1.2 LA CLASSIFICAZIONE COME STRUTTURA CONCETTUALE

Facciamo partire la nostra analisi dalla constatazione che la classificazione è una struttura concettuale (cfr. Marradi, 2007). Marradi la inserisce fra “gli

strumenti intellettuali con cui l'uomo affronta questi problemi [della conoscenza]" (Marradi, 2007, p. 47), classificati in tre livelli: 1) concetti e strutture concettuali; 2) asserti; 3) spiegazioni. La classificazione rientra nel primo livello; si tratta, dunque, di uno strumento più complesso di un semplice concetto, che si può formare per articolazione di quest'ultimo. Una breve trattazione del concetto in sé sarà dunque di giovamento per la comprensione dell'argomento di nostro di interesse, cioè la classificazione.

1.2.1 I CONCETTI ED IL "CONCETTO DI CONCETTO"

Non sono molti gli autori a trattare il tema del "concetto del concetto", tra questi Marradi cita e accetta le definizioni di chi vede nei concetti "le unità fondamentali del pensiero" (Sartori, 1984, p. 27), unità che sono "tutt'altro che rigide e tutt'altro che omogenee fra loro" (Selye, 1964, p. 268) e che "non sono unità ultime e indivisibili" (Gellner, 1964, p. 120).

Un concetto non è indivisibile perché può essere scomposto in vari aspetti, che sono a loro volta concetti. Aspetti che sovente si indicano associati al concetto di gatto, ad esempio, sono l'essere a quattro zampe, il miagolare, il mostrare indipendenza verso l'eventuale padrone. Aspetti più inusuali sono la sua parentela genetica col leone o la tigre. Ancora più inusuali: alcune particolarità dell'apparato digestivo o riproduttivo, utili agli zoologi per distinguerlo dagli altri felini.

"L'insieme di questi aspetti viene detto intensione del concetto" (Marradi, 2007, p. 48). L'autore ricorda che, probabilmente, la prima formulazione del concetto di "intensione" si deve a Locke (cfr. Locke, 1690), quando osservava che alcuni, pensando all'oro, si limitavano a considerare colore e peso, altri pensavano alla fusibilità, altri ancora alla duttilità, altri alla solubilità in acqua regia.

Marradi di nuovo concorda con Sartori quando afferma che "un concetto è la sua intensione", cioè l'insieme dei suoi aspetti, che, secondo Sartori, sono distinti in aspetti definitivi, o necessari, e aspetti contingenti (1984, p. 32-40): allo stesso modo alcuni cognitivisti (Osherson e Smith, 1981; Lakoff, 1987)

distinguono tra un nucleo (*core*) del concetto e le procedure di identificazione di tale nucleo. Nel concetto di donna, il nucleo sono le caratteristiche sessuali e riproduttive, le procedure di identificazione si servono della forma del corpo, il timbro della voce, gli abiti, ecc. L'identificazione è frutto di un bilanciamento tra assenza e presenza dei vari aspetti concettuali nello specifico referente, per cui una tigre a tre zampe viene riconosciuta unanimemente come tigre, anche se l'essere un quadrupede appartiene al nucleo degli aspetti definitori del concetto di tigre (cfr. Ziff, 1960).

“L'insieme dei referenti di un concetto è detto la sua estensione” (Marradi, 2007, p. 49), intendendo per referente “qualunque cosa alla quale il soggetto pensi in quel dato momento: non solo gli oggetti fisici, ma gli eventi, le altre persone e le loro caratteristiche, azioni e pensieri; non solo gli oggetti esistenti e esistiti, ma anche quelli solo immaginari” (Marradi, 2007, p. 29).

Questa dilatazione esistenziale pone alcune questioni. Ad esempio, raramente si delimita con precisione l'ambito spazio-temporale, per cui la maggior parte dei concetti che utilizziamo nella vita di tutti i giorni ha un'estensione dai confini incerti; la cosa non comporta problemi particolari.

Alcuni concetti, nel momento in cui si formano, hanno un rapporto stretto con i loro referenti, altri meno, altri ancora non hanno referenti: “concetti quali super-ego, libido, ritardo culturale, sviluppo rappresentano delle costruzioni piuttosto che delle astrazioni; anche se vogliamo considerarli delle astrazioni, si può dire che abbiano perso chiaramente ogni referente empirico” (Bruschi, 1990, p. 150). Più recisamente, per Sartori esistono “concetti che non hanno referente [...] ad esempio i concetti di funzione, di struttura, di equilibrio, di isomorfismo” (Sartori, 1979, p. 58).

Anche quando il rapporto tra un concetto ed i suoi referenti è stretto, non è detto che esso sia a-problematico. Benché, per Bruschi, “chiarendo l'intensione chiariamo anche l'estensione” (Bruschi, 1993, p. 66), va detto, come osservano i cognitivisti, che “nella realtà esistono casi ambigui, per i quali è controverso stabilire se possiedono tutti i requisiti per rientrare nell'estensione di un concetto” (Marradi, 2007, p. 49). Come vedremo più avanti, tale questione, sotto diversi punti di vista, sarà il filo conduttore di questo lavoro.

Sempre legata alla dimensione spazio-temporale dei referenti è la questione “se un concetto debba essere generale, cioè avere come referente un tipo, o possa avere anche un referente singolare, individuato nello spazio e nel tempo” (Marradi, 2007, p. 50). Opinione di Marradi è che entrambe le tesi – i concetti hanno solo referenti generali o solo referenti particolari – appaiono forzate. Sulla scorta del pensiero di Durkheim, che si chiedeva “se il pensiero concettuale può essere applicato al genere e alla specie [...] perché mai non potrebbe estendersi all’individuo, cioè al limite verso cui tende la rappresentazione col progressivo diminuire della sua estensione?” (Durkheim, 1963, p. 472), possiamo affermare, dunque, che formuliamo concetti generalissimi ma anche con referenti singolari, determinati spazio-temporalmente.

Riprendiamo da Marradi (2007) alcuni esempi che ci aiutino a chiarire ulteriori questioni. Si vedano i tre seguenti concetti: i laureati in sociologia, i laureati in sociologia a Firenze, i laureati in sociologia a Firenze nell’anno accademico 2006-2007. Passando dal primo al terzo concetto, aumenta l’intensione, in quanto si aggiunge un aspetto del concetto che non c’era in precedenza. Contestualmente, aumentando l’intensione, diminuisce l’estensione: alcuni referenti del primo concetto non saranno referenti del secondo ed alcuni del secondo non lo saranno del terzo:

Possiamo quindi immaginare questi concetti come situati su differenti gradini di una ipotetica **scala di generalità**. Poniamo il concetto A e il concetto B su due gradini diversi della stessa scala di generalità quando consideriamo che tutti i referenti di A sono anche referenti di B, mentre non tutti i referenti di B sono referenti di A. In tal caso diremo che B è un **genere** rispetto ad A, mentre A è una **specie** di B. Due concetti possono appartenere alla stessa scala di generalità solo se siamo disposti ad istituire fra loro questo rapporto di genere/specie¹ (Marradi, 2007, p. 50).

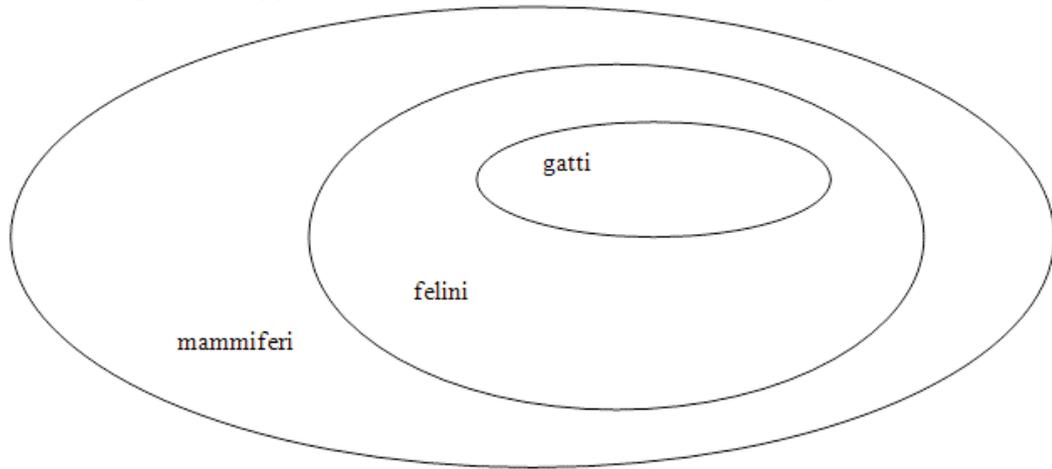
Vediamo una scala di generalità rappresentata nella figura 1.1 (cfr. Marradi, 2007). La rappresentazione è quella tipica degli insiemi: i gatti sono una specie di felini, che sono una specie di mammiferi.

La scala potrebbe contenere altri insiemi: insiemi più grandi (che fanno da genere) come i vertebrati o gli animali, o insiemi più piccoli (che fanno da specie)

¹ Grassetto e sottolineature sono presenti nell’originale.

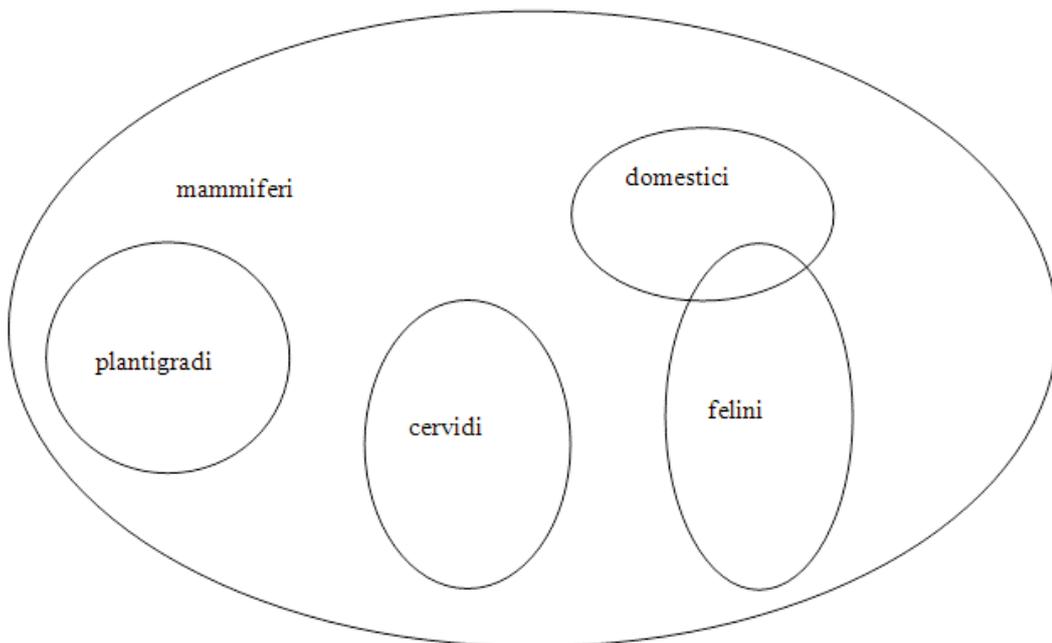
come i gatti siriani o i gatti siriani di Largo Argentina a Roma. Questi insiemi, è questo che ci interessa sottolineare, non avrebbero intersezioni: “se due cerchi hanno un’intersezione, cioè un’area comune, i relativi concetti non fanno parte della stessa scala di generalità” (Marradi, 2007, p. 51).

Fig. 1.1: Rappresentazione insiemistica di una scala di generalità



Consideriamo la figura 1.2 (cfr. Marradi, 2007). Non tutti i felini sono domestici e non tutti gli animali domestici sono felini: i due concetti hanno un’intersezione, dunque non sono posti sulla stessa scala di generalità.

Fig. 1.2: Rappresentazione insiemistica di quattro rapporti genere/specie



La figura ci fornisce ulteriori informazioni. Ognuna delle quattro specie che vediamo forma una diversa scala di generabilità col genere mammifero. Ovviamente se ne potrebbero creare di nuove e le esistenti potrebbero essere estese in un senso o nell'altro, aggiungendo o diminuendo l'intensione dei concetti genere/specie: "per ogni diverso modo di articolare l'intensione di un concetto si forma quindi una diversa scala di generalità" (Marradi, 2007, p. 52). Se, ad esempio, considero l'habitat, dal genere mammifero posso scendere a quello alla specie di mammifero marino; se considero la dieta, da mammifero a mammifero erbivoro; se considero la presenza attuale, da mammifero a mammifero estinto. Come vedremo meglio poco più avanti, il criterio che stiamo articolando si dice *fundamentum divisionis*.

Alcune brevi considerazioni finali (per quanto riguarda il tema del concetto in sé, anche se è facile rendersi conto come parlando del "concetto", si stia in realtà già affrontando il tema della classificazione) derivano dalle questioni che riguardano: se chi pensa, debba essere consapevole o meno dei concetti che ha in mente in un determinato momento; se i concetti debbano essere chiaramente delimitati; se i concetti siano prodotti della società o degli individui (cfr. Marradi, 2007, p. 53).

Marradi sostiene che queste tesi (cui molto sommariamente si può rispondere: no; no; in parte dell'una, in parte dell'altro) risentono della tendenza, tollerabile nella vita di tutti i giorni ma da evitare nell'attività scientifica, a "schiacciare i concetti sui termini che li designano" (Marradi, 2007, p. 54). Se non si presta attenzione a questo punto, si rischia di impostare la questione in modo riduttivo: "se riflettiamo sulla natura dei concetti senza lasciarci fuorviare dal loro rapporto con i termini, credo si debba giungere alla conclusione che ogniqualvolta qualcuno pensa a un referente, egli formula e usa (almeno) un concetto" (Marradi, 2007, p. 54). Siamo dunque caratterizzati da un dialogo continuo tra referenti ed aspetti di concetti mai del tutto ipostatizzati, una tensione continua tra intensione ed estensione: in poche parole, "siamo costantemente impegnati a formare e riformare concetti" (Fernandez, 1986, p. 12).

1.3 CLASSIFICAZIONI, TIPOLOGIE E TASSONOMIE

Vediamo ora i prodotti dell'attività classificatoria, in relazione alle diverse procedure attuabili per ottenerli.

1.3.1 CLASSIFICAZIONI

Una coppia di concetti possono stare fra loro in un rapporto di genere e specie: “la mera esistenza di questa struttura implica la possibilità di un'altra più complessa: la **classificazione**²” (Marradi, 2007, p. 55).

Riprendendo l'esempio precedente riguardante i gatti, se questi sono un concetto di specie rispetto ai felini (concetto di genere), ciò significa che i referenti del concetto felini sono divisibili in almeno due classi: i gatti e i non-gatti: “questa contrapposizione A/non-A è la forma più elementare di classificazione³” (Marradi, 2007, p. 55).

Il prodotto di base di un'attività classificatoria è detto **schema di classificazione**⁴. Marradi preferisce questa definizione a quella più diffusa di classificazione (cfr. Berger e Zelditch, 1968; Fox, 1982), in quanto quest'ultima “designa sia l'attività che il suo prodotto: è uno dei numerosi esempi di quello che in inglese si dice *process-product equivocation*” (Marradi, 2007, p. 56).

Nella dottrina classica⁵ i generi vengono divisi in specie prendendo in considerazione un solo aspetto dell'intensione del concetto di genere e

² Grassetto presente nell'originale.

³ La διαίρεσις (*diairesis*, scelta dicotomica) è una procedura concettualizzata da Platone, che dicotomizzava un genere in due specie, una caratterizzata dalla presenza di un certo attributo, l'altra dalla sua assenza. Proprio perché basata su una sola proprietà per volta fu contestata da Aristotele, che la giudicava inadatta alle scienze naturali.

⁴ Marradi (cfr. Marradi, 1993) ricorda che i linguisti sono soliti usare il termine ‘tipologia’ anche con un solo *fundamentum divisionis*, come in questo caso (cfr. Greenberg, 1957).

⁵ Sviluppata dai precetti Aristotelici, è rappresentata dalle opere di Hempel (Hempel e Oppenheim, 1936; Hempel, 1952, 1961) e, all'interno del panorama italiano, Sartori (Sartori, 1971).

articolandolo. I filosofi scolastici hanno chiamato questo aspetto *fundamentum divisionis*. Ad esempio, possiamo classificare il concetto di sistema politico articolando il principio di legittimazione dei governanti come *fundamentum divisionis*. Avremo, dunque, le seguenti classi: teocratico, autoritario, aristocratico, plutocratico, democratico: “in linea di principio tutti gli altri aspetti dell’intensione del concetto generale si ritrovano immutati nell’intensione di ciascuno dei concetti specifici” (Marradi, 2007, p. 56); le proprietà relative al concetto di sistema politico democratico saranno tutte le proprietà relative al concetto di sistema politico, eccezion fatta per quelle incompatibili con la democrazia.

Marradi definisce **classificazioni intensionali** le “operazioni con cui l’estensione di un concetto a un dato livello di generalità è divisa in due o più estensioni più ristrette, ciascuna corrispondente a un concetto a un minor livello di generalità” (Marradi, 1993, p. 22). Per lo stesso tipo di operazione sono stati proposti nominativi diversi, quali divisione (cfr. Cohen e Nagel, 1934) e categorizzazione (cfr. Scheffler, 1967).

Funzione della classificazione intensionale è, il più delle volte, quella di definire e chiarificare i concetti in uso: “nella sua forma strutturalmente più semplice, la classificazione intensionale può essere considerata un processo di esplicazione o chiarificazione concettuale” (Marradi, 1993, p. 22; cfr. anche vedi Sartori, 1970; Glaser, 1978). In parole molto povere, si tratta di capire di cosa stiamo parlando, e anche, o soprattutto, di cosa *non* stiamo parlando, instaurando un legame fra termine e concetto: “lo scopo dell’operazione è definire l’intensione di ciascuna delle classi che vengono costituite, cioè esplicitare la classe in quanto concetto, e denominarla con un termine o con un’espressione appropriati” (Marradi, 1993, p. 22; cfr. anche Saussure, 1916).

È proprio pensando a questo aspetto della classificazione, intesa scientificamente come articolazione intensionale di un concetto, che Hempel afferma: “la classificazione è una forma specifica di formazione dei concetti scientifici” (Hempel, 1965, p. 139). Chiarendo l’intensione di un concetto, si specificano le “condizioni necessarie e sufficienti dell’appartenenza alla classe, stabilendo certe caratteristiche che tutti e solo i membri della classe possiedono”;

ogni classe è dunque “definita come l’estensione del corrispondente concetto”. (Hempel, 1965, p. 137).

Altro precetto della dottrina classica è che i confini fra le classi siano rigidamente delimitati: “data una qualsiasi coppia di classi, nessun referente dev’essere attribuibile ad entrambe le classi della coppia” (Marradi, 2007, p. 56). Questo criterio è detto **mutua esclusività**.

In più, il complesso delle classi deve essere esaustivo, nel senso che ogni possibile stato di caso sulla proprietà (quindi ogni possibile referente) assunta come *fundamentum divisionis* deve essere assegnabile ad una classe. Questo criterio è detto **esaustività**.

La mutua esclusività è una proprietà di ciascuna coppia di classi, l’esaustività del complesso delle classi: considerando insieme questi criteri, ne consegue che ogni referente di un concetto di genere deve essere assegnato ad **una e una sola** delle classi derivanti dalla classificazione. A questo proposito, Marradi osserva:

I requisiti della dottrina classica sono ragionevoli, ed è abbastanza naturale rispettarli quando si classificano oggetti astratti, come per esempio i numeri. Altra cosa è rispettarli allorché si vuole classificare referenti empirici, sia percepibili con i sensi (come minerali, piante, animali) sia non percepibili con i sensi, come sentimenti, temperamenti, e così via. Come ha osservato Jevons, “le forme viventi non ammettono rigorose linee di demarcazione” [1874, 724] (Marradi, 2007, p. 57).

Dato che i requisiti logici valgono allo stesso modo per ogni prodotto classificatorio, apriamo ora una breve parentesi per affrontare qualche questione che li riguarda. Mutua esclusività ed esaustività sono requisiti che rappresentano i principi stessi della logica classica binaria su cui, da Aristotele in poi, si fonda l’intero pensiero occidentale (cfr. Calogero, 1968).

Vogliamo brevemente chiarire cosa s’intenda per logica classica. Essa si fonda sull’osservazione di tre principi: quello d’identità, quello di non contraddizione, e quello del terzo escluso.

Il principio d'identità ($A=A$) afferma che è impossibile che una cosa sia e non sia contemporaneamente, in modo che, qualora essa sia oggetto di studio nel tempo t , si abbia la garanzia che non muti le sue caratteristiche in detto tempo.

Il principio di non contraddizione (se $A=B$, non può essere $A=\text{non-B}$) impedisce di formulare affermazioni contraddittorie al tempo t sull'oggetto in questione.

Infine il principio del terzo escluso (A o non-A) afferma che non c'è via di mezzo tra due predicati contraddittori, se uno è vero l'altro è falso e viceversa.

La logica classica è quindi la logica della bivalenza, da non confondere con ambivalenza o ambiguità. Ambiguo è qualcosa che ha significato equivoco, poco chiaro; ambivalente è qualcosa che ha un duplice valore, un duplice significato contemporaneamente; bivalente è invece qualcosa che può assumere due significati, due valori, dei quali uno però esclude l'altro.

In questa logica una cosa esiste o non esiste (*bivalenza ontologica*), una proposizione è o vera o falsa (*bivalenza logica*⁶), ha un valore di verità 1 o 0. Applicata alla classificazione, è la logica dell'appartenenza o non appartenenza.

Quest'ultima considerazione ci riporta al nostro specifico campo d'indagine. Prima di andare avanti, possiamo sottolineare come, benché la classificazione scientifica dovrebbe distinguersi da quella spontanea “per una più precisa e impersonale definizione dei criteri su cui la classificazione stessa è basata” (Biorcio, 1993, p. 11), molto spesso “le classificazioni sono operate – non solo nella vita quotidiana, ma anche nelle scienze – in modo assai meno rigoroso di quanto prescrive la dottrina” (Marradi, 2007, p. 57), sia per queste tangibili difficoltà, sia per semplice trascuratezza⁷.

Più in generale, sottolinea Agodi, “l'autonomia tra classificazioni scientifiche e classificazioni profane è, tuttavia, il risultato di un processo mai definitivamente stabilizzatosi; non è mai definitivamente raggiunta ed è

⁶ Non è possibile affermare e negare nello stesso momento.

⁷ Come esempio, riguardo alle classificazioni non scientifiche, Marradi riporta l'esempio fornito dai *media* durante la recente guerra in Jugoslavia. All'epoca si parlò di serbi, croati e musulmani, confondendo due diversi *fundamenta divisionis*, l'etnia e la religione, e violando, di conseguenza, il criterio della mutua esclusività, in quanto i musulmani erano quasi tutti di etnia serba oppure croata.

continuamente ridefinita, così come lo è la distinguibilità tra cultura, struttura sociale o «scienza» nella definizione dei criteri cui quelle classificazioni rinviano” (Agodi, 2001, p. 170)⁸.

1.3.2 TIPOLOGIE

Aumentando il livello di complessità, incorriamo nel secondo tipo di prodotto classificatorio, la **tipologia**, che si forma attraverso l’articolazione di due o più *fundamenta divisionis*, presi simultaneamente. In questo modo, si viene a creare, anziché una serie unidimensionale di classi, un insieme *n*-dimensionale di tipi, dove *n* è appunto il numero degli aspetti dell’intensione del concetto di genere che vengono considerati⁹. Nella dottrina classica, i tipi, esattamente come le classi, devono essere esaustivi e mutuamente esclusivi.

Il termine tipologia deriva dal greco τυπος (impronta, conio, modello) combinato con la radice λογος (discorso, ragione). Come nel greco, così anche in italiano c’è stato un rapido slittamento dal significato originario (rintracciabile, ad esempio, nelle parole composte tipografia, tipografico) a quello metaforico, con passaggio da referenti tangibili a referenti non tangibili. Lo stesso discorso vale ad esempio per la parola carattere, per cui oggi in italiano si può parlare sia di carattere di stampa, che di carattere di una persona. Dunque etimologicamente tipologia significa dottrina dei tipi, riflessione sui tipi, discorso sui tipi.

⁸ Agodi, citando Latour (1987), riporta un interessante “esempio che potrebbe essere classificato come riferito tanto alla microbiologia, quanto alla storia sociale o alla sociologia della medicina” (Agodi, 2001, p. 170). A metà Ottocento, l’introduzione, da parte di Pasteur e gli igienisti, del concetto di “microbo” come causa essenziale delle malattie ridisegnò la classificazione sociale: se prima la divisione dei gruppi meglio definita era quella ricchi/poveri, capitalisti/proletari, essa divenne persone malate/persone portatrici sane/persone vaccinate/ecc: “un uomo ricco poteva morire contagiato dalla giovane cameriera portatrice di tifo. Nacque allora una nuova forma di solidarietà sociale” (Latour, 1987; tr. it. 1998, p. 155-6).

⁹ Questa caratteristica delle tipologie, per cui il numero dei tipi è una funzione moltiplicativa del numero delle classi in ciascuno dei *fundamenta* considerati, viene a volte definita potenza.

Facciamo un esempio piuttosto banale, ma esplicativo: operando una classificazione intensionale sui due *fundamenta* sesso (le cui classi sono naturalmente uomo e donna) e titolo di studio (dicotomizzato nelle classi laureato, non laureato), si avrà un insieme di quattro (due classi per due classi) tipi bi-dimensionali (una dimensione per ogni *fundamentum*): uomo laureato, uomo non laureato, donna laureata, donna non laureata. Il tipo “uomo laureato” è il prodotto logico di due classi (una per ogni *fundamentum*: la classe “uomo” per il *fundamentum* “sesso”, e quella “laureato” per il *fundamentum* “titolo di studio”) e la sua estensione è l’intersezione di quella delle classi che lo formano, l’insieme quindi di chi tra i laureati è uomo, e chi tra gli uomini è laureato, il che evidentemente è la stessa cosa, dato che, come abbiamo già visto, la divisione è simultanea: non è rilevante l’ordine in cui sono considerati i vari *fundamenta*: “il tipo delle persone di nazionalità italiana e medici di professione coincide con il tipo dei medici che hanno nazionalità italiana” (Marradi, 1993, p. 27).

Se consideriamo tre *fundamenta* (A, B e C) aventi rispettivamente cinque, tre e due classi ciascuno, il risultato della divisione delle relative intensioni sarà una gamma di trenta tipi tridimensionali, come possiamo vedere facilmente nella tabella che segue:

Tab. 1.1 – Tipi risultanti dai *fundamenta* A, B e C

		A				
B	C	A 1	A 2	A 3	A 4	A 5
B 1	C 1	1	2	3	4	5
	C 2	6	7	8	9	10
B 2	C 1	11	12	13	14	15
	C 2	16	17	18	19	20
B 3	C 1	21	22	23	24	25
	C 2	26	27	28	29	30

Questa tabella rappresenta lo spazio di attributi delle tre proprietà considerate. Lo spazio di attributi, prodotto logico concettualizzato da Hempel e Oppenheim (1936) e denominato *property space* da Lazarsfeld (1937), è la combinazione logica ed empirica di più proprietà. Quando si dice che una tipologia deriva dalla divisione di più *fundamenta* considerati

contemporaneamente, si sta dicendo in altre parole, che una tipologia si ottiene costruendo lo **spazio di attributi** di tali concetti.

Come dicevamo e come vediamo facilmente attraverso l'uso della tabella, il numero dei tipi in una tipologia è pari al prodotto del numero delle classi in ciascun *fundamenta*. In questo caso, abbiamo una potenza pari a trenta tipi.

Questo comporta tre problemi strettamente connessi: 1) il numero dei tipo è alto anche se si combinano pochi *fundamenta*, negli altri casi è altissimo; 2) alcuni tipi saranno probabilmente pure possibilità logiche prive di interesse concettuale; 3) alcuni tipi avranno estensione nulla o pochissimi referenti. Tutto ciò è riassumibile nella questione della *proliferazione dei tipi*.

I trenta tipi della tabella, ad esempio, sarebbero ingestibili in qualsiasi ricerca, ne deriverebbe una tipologia frammentata e dispersiva, in cui verrebbe a mancare l'indispensabile funzione di riduzione della complessità, tipica dell'attività classificatoria e di qualsivoglia impresa scientifica.

La soluzione alla proliferazione dei tipi sta nella riduzione dello spazio di attributi. Di pari passo con la diminuzione del numero dei tipi, diminuisce contestualmente anche la complessità intellettuale della tipologia. I tipi non possono essere eliminati, altrimenti sarebbe compromessa l'esaustività della tipologia; vanno aggregati in modo ragionato, tenendo in considerazione la prossimità semantica e la numerosità dei referenti. Non si possono mettere insieme due tipi concettualmente molto distanti solo perché hanno pochi casi assegnati, e non si possono mettere insieme due tipi concettualmente molto vicini se la distribuzione di frequenza ne risultasse decisamente sbilanciata.

Nella tabella 1.2 vediamo un esempio di possibile riduzione dello spazio di attributi, che contempererà i due requisiti:

Tab. 1.2 – Riduzione dello spazio di attributi dei fundamenta A, B e C

		A				
B	C	A 1	A 2	A 3	A 4	A 5
B 1	C 1	1	2	3	4	5
	C 2	6	7	8	9	10
B 2	C 1	11	12	13	14	15
	C 2	16	17	18	19	20
B 3	C 1	21	22	23	24	25
	C 2	26	27	28	29	30

Nella vita quotidiana è normale che si formino concetti tipologici senza preoccuparsi di considerare i criteri di divisione o tutte le possibili combinazioni. Sfortunatamente questo vale anche per molte tipologie scientifiche che risultano, esattamente come era per le classificazioni, imprecise per “insufficiente chiarezza del *fundamentum*” (Marradi, 1993, p. 23; come esempi di classificazioni difettose: Spencer, 1892; Freud, 1931; Malinowski, 1944; Gurvitch, 1950).

Tali errori possono però fornire lo spunto per ulteriori analisi, attraverso la ricostruzione inferenziale *ex-post* dei *fundamenta* impliciti in una tipologia già costruita da altri autori: è quella che Lazarsfeld ha chiamato **substruzione** e successivamente ricostruzione di uno spazio di attributi. Lazarsfeld stesso ha eseguito note ricostruzioni, ad esempio della tipologia delle relazioni di autorità proposta da Fromm, dei suicidi di Durkheim, delle norme sociali di Kingsley e Davis (cfr. Lazarsfeld e Barton, 1951). Altre note sono state fatte da Parsons (1951), sempre sui tipi di suicidio di Durkheim, o da Smelser (1976) sulla tipologia delle forme di autorità di Weber, o da Capecchi (1966) sulla tipologia delle forme di orientamento dell’azione sociale di Merton.

A detta di Lazarsfeld, l’importanza di questa operazione è nella possibilità di indagare le implicazioni logiche della tipologia proposta senza costruzione esplicita dello spazio di attributi. Questa operazione porta a molti vantaggi, “l’autore può rendersi conto di avere trascurato certi casi; può accertarsi che alcuni dei suoi tipi non invadano il campo degli altri; e la sua classificazione diverrebbe probabilmente più valida ai fini pratici della ricerca empirica. [...] La ricostruzione delle adeguate combinazioni d’attributi che sono alla base di un sistema dato di tipi, rende questi ultimi più atti alla ricerca pratica” (Lazarsfeld e Barton, 1951, p. 276).

Vediamo l’analisi fatta da Talcott Parsons sul *Suicidio* di Durkheim riportata in Lazarsfeld e Barton (1951, p. 275). L’autore francese individua tre tipi di suicidio (altruistico, egoistico e anomico), senza però specificare le dimensioni (attributi, proprietà, o se vogliamo *fundamenta*) dalle quali questi tipi derivano. Parsons ne indica due: la presenza o assenza di norme di gruppo e, ove presenti, il loro contenuto individualistico o collettivistico. Lo spazio ricostruito risultante è raffigurabile nella seguente tabella:

Tab. 1.3 – Ricostruzione dello spazio d'attributi per i tipi di suicidio di Durkheim

	Contenuto delle norme	
Esistenza delle norme	<i>Individualistico</i>	<i>Collettivistico</i>
<i>Norme presenti</i>	Suicidio egoistico	Suicidio altruistico
<i>Norme assenti</i>	Suicidio anomico	

Secondo Lazarsfeld: “tale formulazione non solo chiarisce il significato teorico della distinzione operata dal Durkheim fra i tre tipi di suicidio, ma introduce naturalmente ulteriori problemi. Quali alternative di comportamento esistono in ciascuna situazione, oltre al suicidio? Queste categorie esauriscono i possibili tipi di norme sociali?” (Lazarsfeld e Barton, 1951, p. 275).

Marradi, ad ogni modo, sottolinea come viene da taluni esaltata l'importanza della ricostruzione, giudicata “coessenziale alla stessa classificazione” (Marradi, 1993, p. 27), ma condividiamo la sua considerazione secondo cui questa importanza cruciale derivi semplicemente dalla scarsa sistematicità di molti ricercatori.

1.3.3 CLASSIFICAZIONI ESTENSIONALI

L'ultimo modo per costruire classificazioni che spesso sono piuttosto tipologie è attraverso l'utilizzo di tecniche di classificazione **estensionale**, e specificamente di classificazione estensionale automatica, cioè attraverso tecniche di *clustering*.

Marradi chiama classificazioni estensionali le “operazioni con cui gli *oggetti* o *eventi* di un dato insieme sono raggruppati in due o più sottoinsiemi a seconda di similarità percepite nei loro stati su una o (più frequentemente) più proprietà” (Marradi, 1993, p. 22). Come si vede, questa definizione rispecchia fedelmente quelle riportate ad inizio capitolo, riprese da dizionari ed enciclopedie, ed è inoltre molto più vicina, rispetto alla famiglia delle operazioni intensionali, a quello che il senso comune intende per classificazione. Questo perché, come hanno rilevato più autori, si ritiene che la classificazione estensionale sia più

vicina all'attività categorizzatrice degli individui nel processo d'apprendimento (cfr. Piaget e Inhelder, 1959; Biorcio, 1993; Marradi, 1993).

Biorcio (1993, p. 18) assegna a queste operazioni “un ruolo primario nelle prime fasi di sviluppo del pensiero infantile”, riservando all'adulto (ed alla scienza) “la capacità di elaborazione tipica della classificazione intensionale”. Marradi stesso (1992, p. 24) afferma che “i fanciulli sembrano acquisire il concetto di classificazione attraverso una serie di esperienze di natura estensionale (cioè basate sul raggruppamento di oggetti) piuttosto che intensionale, cioè basate sull'articolazione di concetti”. Considerato ciò e considerato che “una remota origine della procedura può forse trovarsi nelle critiche di Aristotele (Parti degli Animali, 642b, 21 - 644a, 10) alla $\delta\iota\alpha\rho\epsilon\sigma\iota\varsigma$ platonica”, l'autore nota che una “concezione estensionale della classificazione tardò almeno venti secoli ad attingere dignità scientifica” (Marradi, 1993, p. 24), trovando giustificazione di ciò in un motivo pragmatico di natura tecnico-operativa: “è ragionevole supporre che la classificazione estensionale, attività umana spontanea, per assurgere al rango di operazione scientifica abbia dovuto attendere che si facesse luce l'idea di registrare in modo ordinato gli stati di un insieme di oggetti su un insieme di proprietà, cioè il concetto di matrice dei dati” (Marradi, 1993, p. 24).

Perché una classificazione estensionale di un insieme ampio di oggetti possa ambire a qualche garanzia di intersoggettività, si devono determinare:

- gli oggetti sui quali operare la classificazione, cioè i casi, da trasformare in vettori-riga nella matrice;
- le proprietà sulle quali operare la classificazione, cioè le variabili, che diverranno i vettori-colonna della matrice, con lo scopo di minimizzare l'eterogeneità interna e massimizzare quella esterna;
- le procedure con cui gli stati sulle proprietà in questione verranno accertati e registrati;
- i criteri con cui saranno valutate, proprietà per proprietà, le differenze fra gli stati;
- una formula logico-matematica, che combini le differenze sulle varie proprietà considerate in un singolo valore della differenza fra due oggetti

(nella letteratura sulla classificazione estensionale questa formula viene detta “funzione distanza”; cfr. Sokal, 1961);

- un complesso di regole circa i criteri di formazione delle classi.

Data la mole delle informazioni da immettere e la complessità delle operazioni da svolgere, il ricorso a un'organizzazione matriciale e al conseguente calcolo elettronico appare inevitabile, soprattutto al crescere dei vettori riga e colonna, ragion per cui ci si spiega il ritardo nell'acquisita dignità scientifica di questa operazione.

Marradi sottolinea la contemporaneità (non casuale) tra gli studi dei proto-statistici dell'Università di Gottinga, che elaboravano una forma matriciale di organizzazione dei dati sulle risorse economiche e militari dei vari principati tedeschi, e quelli del botanico Michel Adanson, che richiedeva la considerazione di tutte le proprietà delle piante, nessuna esclusa, come requisito necessario per la loro classificazione, proposta sulla base di una proporzione di stati uguali, fra ogni coppia di piante, sul totale delle proprietà (cfr. Adanson, 1763). Adanson probabilmente, per questo scopo, si serviva dell'ausilio di una forma rudimentale di matrice dati e gli attuali specialisti ne riconoscono l'importante contributo nell'ideazione di una “funzione distanza”, tanto da chiamare “adansoniana” l'operazione qui denominata estensionale (cfr. Sneath, 1957; Sokal, 1961).

È interessante, dunque, notare come il dibattito sulla classificazione estensionale, operazione di cui è riconosciuta l'importanza nello sviluppo e nell'apprendimento, sia stato in un certo senso schiacciato da due punti di vista diametralmente opposti, per i quali tale classificazione estensionale rappresentava, prima, un'operazione non scientifica e, dopo, un'operazione semplicemente tecnica. Vale forse la pena, a questo punto, esortare ad una riflessione metodologica, ad un uso più “pensato” della *cluster analysis* (in altre parole, potremmo dire: ad un uso meno *automatico* della classificazione estensionale *automatica*), che metta al riparo dall'illusione quantofrenica e dalla tentazione di “immettere nel *computer shaker* quantità enormi di variabili indistinte, supponendo che per qualche magia il *computer* stesso ordini il tutto” (Statera, 1994, p. 138).

1.3.4 TASSONOMIE

L'ultimo prodotto classificatorio è la **tassonomia**, che è una “struttura concettuale ancora più complessa” (Marradi, 2007, p. 59). Si ottiene quando, dopo aver diviso l'estensione di un concetto applicando un *fundamentum divisionis*, l'estensione di alcune o di tutte le classi viene suddivisa applicando altri *fundamenta*, e così via per varie divisioni in successione. Insomma, come nella tipologia, si applicano due o più *fundamenta*, ma a differenza di questa non si applicano contemporaneamente, ma in successione; l'ordine in questo caso è decisivo: dividendo un concetto prima per il *fundamentum* A e poi per il *fundamentum* B si avrà una tassonomia, considerando prima B e poi A se ne avrà un'altra. La tipologia, infatti, è a differenza della tassonomia un prodotto logico, di conseguenza gode della proprietà commutativa di tutti i prodotti.

La tassonomia, letteralmente “la legge delle classi, degli ordini”, è il prodotto classificatorio che più è stato usato nelle scienze naturali, in quanto le sue caratteristiche ben si conciliavano con l'esigenza di dare un ordine sistematico al creato e renderlo quindi più intelligibile, quando in un clima di pieno essenzialismo, si pensava che le forme della natura fossero immutabili. In un secondo momento, la tassonomia si è dimostrata in grado di rendere conto dei percorsi e dei legami delle specie viventi, quando l'evoluzionismo aveva ormai sancito il suo trionfo ed era diventato la bandiera della nuova scienza.

La tassonomia ha tutte le altre caratteristiche di un qualsiasi schema di classificazione, per cui in ogni passaggio dal genere alla specie si assiste ad un aumento dell'intensione e a una diminuzione dell'estensione.

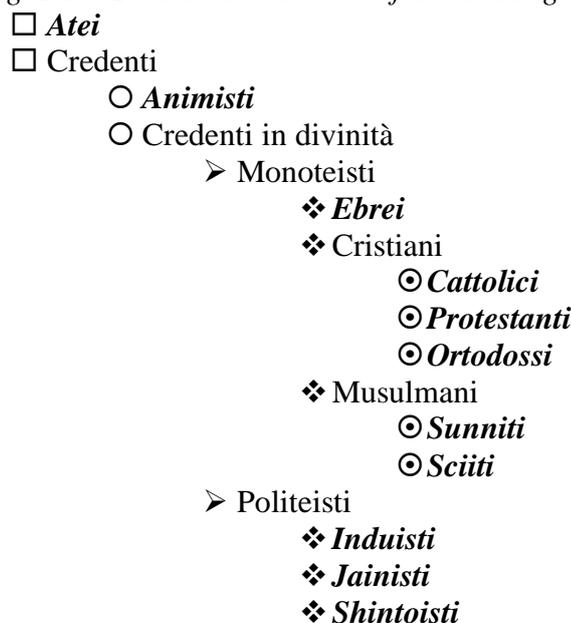
Va però ricordato che, se il singolo passaggio di una tassonomia è identico ad uno schema classificatorio, l'operazione non è isolata e va, quindi, giudicata nel suo complesso, nell'insieme della struttura gerarchica che ne viene fuori.

A proposito delle tipologie avevamo parlato di proliferazione dei tipi: in questo caso, abbiamo il problema della proliferazione delle specie. Per quanto riguarda le scienze sociali (poco possiamo dire, in questo caso, riguardo alle scienze naturali), il problema è risolvibile grazie ad un uso ragionato e utile al

nostro interesse cognitivo delle classi che abbiamo a disposizione. Una tassonomia è simile ad un albero genealogico, ed è nostra libertà “tagliare”, se così si può dire, questo o quel ramo.

Prendiamo come esempio una tassonomia sulle possibili confessioni religiose (Marradi, 1993, p. 26). La prima divisione è quella tra credenti e atei, i credenti possono essere divisi in credenti in divinità e animisti, i credenti in divinità in monoteisti e politeisti, i monoteisti in ebrei, cristiani e musulmani, i politeisti in induisti, jainisti, shintoisti, i cristiani in cattolici, protestanti e ortodossi, i musulmani in sunniti e sciiti, e così via. Se ci fermassimo qui, con cinque passaggi, ci troveremmo di fronte già ad undici classi, come si può facilmente vedere dalla rappresentazione grafica:

Fig. 1.3 – Tassonomia delle confessioni religiose



Le undici classi (in neretto e corsivo) sono quelle a cui non segue ulteriore divisione, sono soltanto specie, non genere. Undici è un numero già abbastanza consistente, ma un ricercatore esperto in materia potrebbe apportare ulteriori separazioni, ad esempio addentrandosi nell'intricato mondo degli animisti oppure approfondendo la divisione dei politeisti o dei musulmani, oppure ascoltando le varie voci presenti all'interno della chiesa cattolica, dall'Ordine Francescano ai neocatecumenali.

Ecco quindi la necessità di quell’ “abile impiego delle classi residuali” (Marradi, 1993, p. 28) di cui prima parlavamo. Così, se la ricerca si svolgesse in Italia potremmo addirittura scegliere, per bilanciare le distribuzioni di frequenza, di riunire in una sola classe i non cattolici, salvo poi recuperare le differenze interne ad esse in un secondo momento per un’analisi più specifica, ed allo stesso modo riuniremmo insieme i non protestanti se ci trovassimo in Inghilterra. In Irlanda la divisione decisiva sarebbe invece quella tra protestanti e cattolici, in Iraq tra sunniti e sciiti. In generale, sganciando la tassonomia da un contesto così specifico, potremmo mischiare *tàxa* di vario rango¹⁰, scegliendo ad esempio due *tàxa* del quinto rango (cattolici e protestanti), tre del quarto rango (ebrei, musulmani e induisti), nessuno del terzo, uno del secondo (animisti) e uno del primo (atei). Riunendo tutte le altre possibilità in una classe residuale, avremmo ora otto classi contro le undici iniziali. La classe residuale, però, in questo caso conterrebbe elementi molto differenti tra loro (ed anche piuttosto numerosi), come i cristiani ortodossi e gli shintoisti, ad esempio. Allora potremmo decidere di recuperare qualche classe, oppure di differenziare quella residuale al suo interno, ad esempio con le sottoclassi altri cristiani, altri politeisti.

Insomma, come si può vedere la tassonomia è uno strumento molto malleabile, che ci permette di aprire e chiudere canali a nostro piacimento, a seconda dell’interesse cognitivo, del contesto della ricerca; è probabilmente proprio a questa versatilità che essa deve il suo successo. Una versatilità che possiamo sfruttare appieno se seguiamo la buona norma di raccogliere informazioni attraverso classificazioni molto sensibili, le cui categorie possono eventualmente essere aggregate in un secondo momento, a seconda delle nostre necessità.

È interessante notare come la tassonomia, nonostante la sua complessità, sia una struttura concettuale molto utilizzata al di fuori della filosofia e delle

¹⁰ Si preferisce il concetto di rango (che si riferisce semplicemente al numero dei passaggi lungo un qualsiasi ramo della tassonomia, partendo dal concetto a livello più alto) in quanto “dato che non si può stabilire alcuna relazione di inclusione fra le estensioni dei *tàxa* appartenenti a rami diversi di una tassonomia, non si può confrontare il livello di generalità dei relativi concetti” (Marradi, 2007, p. 59).

scienze. Gli etnometodologi concordano sulla sofisticazione delle tassonomie rilevate in popoli detti primitivi (cfr. Durkheim e Mauss, 1901; Lévy-Bruhl, 1910; Douglas, 1966; Tyler, 1969). Alcuni schemi di classificazione di uso comune, come quelli abitualmente utilizzati nella comunicazione elettorale (voti per il partito X, per il partito Y, per il partito Z, nulle, bianche, astenuti) sono in realtà riduzioni di tassonomie (eseguite sui seguenti *fundamenta*: il soggetto è andato a votare; ha espresso un voto; ha espresso un voto valido).

1.3.5 CLASSIFICAZIONE(C)

Prima di passare a considerare il ruolo gnoseologico della classificazione e la sua valutazione, vorremmo fermarci su un'altra accezione della classificazione, in particolare sul procedimento che Marradi definisce **classificazione(c)** (cfr. 1984) o **assegnazione** (cfr. 1993) di oggetti a classi, tipi o *tàxa* già costituiti (a seconda che si consideri un solo *fundamentum*, due o più simultaneamente, o una serie in successione).

L'assegnazione può essere effettuata sia attraverso operazioni intensionali che estensionali. Le classificazioni intensionali sono precondizioni dell'assegnazione ed hanno proprio questo scopo: si formano le classi attraverso articolazione di un concetto per poter assegnare loro stati di specifici oggetti. Le classificazioni estensionali, che nel loro stesso svolgimento suddividono in gruppi un insieme già dato, rendono inutili le operazioni di assegnazione, a meno che non si presenti il problema di assegnare nuovi oggetti non presi in considerazione al momento della classificazione originaria.

Per quanto riguarda l'assegnazione di oggetti a classi o tipi già costituiti, nulla cambia nella natura dei prodotti finali. Per quanto riguarda i *tàxa*, se le classi di ordine superiore si formano quando la formazione di quelle di ordine inferiore è ancora in corso, avremo un prodotto ad albero con diramazioni irregolari, rappresentato graficamente da un dendogramma. Altrimenti, se il raggruppamento di secondo ordine avviene solo quando tutti i raggruppamenti di primo ordine sono finiti, il prodotto risulterà identico ad una tassonomia classica, ricordando

però che, in questo caso, avendo a che fare con una classificazione estensionale, i gruppi allo stesso rango sono tipi, mentre generalmente i gruppi allo stesso rango di una tassonomia sono classi in senso stretto. Questo secondo modo di produrre tassonomie è poco più che una possibilità astratta, visto che praticamente “non ne è rimasta traccia nella letteratura logica e metodologica” (Marradi, 1993, p. 29).

1.3.6 ALTRI CONTRIBUTI

In conclusione, vogliamo tracciare un veloce paragone tra le strutture classificatorie proposte da Marradi (ed ampiamente accettate) e quelle di altri autori. La bipartizione intensionale/estensionale che abbiamo potuto notare si rispecchia in altre bipartizioni, in termini e significati leggermente diversi, presenti nel dibattito sull'argomento.

Prima di tutto, la distinzione di Marradi può essere interpretata come risiedente nel fatto che “il processo di suddivisione dell'estensione di una proprietà può essere effettuato a priori oppure a posteriori” (Moriconi, 2000, p. 21): si può partire dai dati e arrivare alle categorie concettuali oppure, al contrario, partire dalle categorie concettuali e classificare gli stati dei casi. Accettando questo punto di vista, è lecito affermare che “questa distinzione [tra classificazione intensionale e classificazione estensionale] [...] è riconducibile a quella, rinvenuta in molti lavori, tra le tipologie formulate a priori e quelle induttive” (Moriconi, 2000, p. 21).

Assunta questa posizione, possiamo fare un parallelo tra la classificazione estensionale e, ad esempio, l'operazione che Bruschi (cfr. Bruschi, 1990) chiama segmentazione a posteriori (prima vengono formati i gruppi, e poi vengono interpretati sulla base delle proprietà che li denotano: quindi la tipologia è fatta a posteriori), così come tra la classificazione intensionale e la segmentazione a priori (il contrario: si formano i tipi, e poi si assegnano i casi). Allo stesso modo potremmo affermare che le tipologie euristiche (dedotte dalla teoria), proposte da Winch (cfr. Winch, 1947), sono un prodotto dell'operazione intensionale, mentre quelle empiriche (derivate dai dati) sono prodotte da quella estensionale.

Bailey (cfr. Bailey, 1973, 1984, 1994) mantiene una posizione controversa: individua tipologie teoriche ed empiriche, ma tenta di superare la frattura con il concetto di tipologia classica; propone un modello con tre differenti livelli d'analisi: quello concettuale, che esclude il riferimento ai dati, quello empirico, che non ha controparte teorica, e quello operativo, che combina i due livelli, ma non necessariamente. Se il pragmatismo ha legato indissolubilmente teoria e pratica, in Bailey il livello operativo, infatti, non cancella gli altri due: rimane possibile fare tipologie puramente concettuali senza controparte empirica, ed è altrettanto possibile farne altre puramente empiriche senza riferimenti teorici, ed è proprio in questo senso che Bailey legge la *cluster analysis*.

In quest'ultimo passaggio c'è a nostro avviso un punto critico: la separazione del momento teorico da quello empirico, una posizione ormai difficile (se non impossibile) da sostenere in qualsiasi ambito scientifico. L'ordine, dice Foucault, è "ciò che non esiste se non attraverso la griglia di uno sguardo, di un'attenzione, d'un linguaggio" (cfr. Foucault, 1966; tr. it., 1996, p. 10). L'illusione ingenua dei comportamentisti di un'osservazione pura è superata, così come l'idea che conoscere significhi svelare la realtà. Perché la realtà non è altro che quella "infinità priva di senso del divenire del mondo, alla quale è attribuito senso e significato dal punto di vista dell'uomo" (Weber, 1904; tr. it., 1958, p. 96). L'esperienza del mondo empirico può essere immediata solo a livello emotivo, la sua analisi è un'esperienza intellettuale che richiede una presa di distanza ed allo stesso tempo una presa di posizione, l'assunzione di un punto di vista, una prospettiva, un criterio che ci faccia da guida, come una bussola per un marinaio, nella scelta, nella selezione, nell'inquadramento e quindi, nella soluzione dei problemi.

Perfettamente in linea con ogni attività scientifica, dunque, anche quella del classificare consiste in un continuo rinvio fra dati e teoria, fra osservazione e concettualizzazione, potremmo dire fra pensiero e azione; in questo palleggio il riscontro empirico ha anche la fondamentale funzione di autocorrezione della classificazione. Esplicative le parole di Lazarsfeld, quando spiega come adattare uno schema ad una determinata situazione:

Si comincia col raccogliere osservazioni [...]. È necessario [...] compilare un'immagine o modello concreto di tutta la situazione cui si riferiscono le relazioni, e poi collocare le singole relazioni in questo 'schema strutturale'. [...] Poi si cerca di applicare sistematicamente questo schema ai dati, ci si rifà allo schema strutturale per perfezionarlo, si riapplica lo schema così riveduto ai dati, e così via. Si può perciò trovarsi, alla fine, con una classificazione piuttosto diversa da quella con cui si era partiti. (Lazarsfeld e Barton, 1951, tr. it. in Lazarsfeld, 1967, p. 240)

Questo significa dunque classificare: nessuna pura induzione e nessuna pura deduzione, bensì un'attività che presuppone "sia un'intima conoscenza della situazione concreta, sia la guida delle teorie sociali" (Lazarsfeld e Barton, 1951, tr. it. in Lazarsfeld, 1967, p. 245). Ogni ulteriore discorso epistemologico su induzioni e deduzioni è chiuso, dal punto di vista – direbbe Statera – della logica dell'indagine, da Campelli, per il quale "il problema se 'prima le osservazioni' o 'prima le ipotesi' è per la verità alquanto fittizio nella prassi concreta dell'indagine" (Campelli, 1991, p. 108).

In questa stessa prassi, sfuma quindi anche la distinzione tra intensionale ed estensionale, o meglio, le due fasi interagiscono continuamente. Riprendendo le parole citate poco sopra, con cui Biorcio (sulla scorta di Piaget) assegnava alla classificazione estensionale un ruolo basilare per lo sviluppo del pensiero infantile, riservando all'adulto ad alla scienza la capacità di elaborare operazioni intensionali, egli continua dicendo che "sebbene distinti, nel pensiero dell'adulto (e nella riflessione scientifica) i due processi coesistono e si influenzano e condizionano a vicenda" (Biorcio, 1993, p. 18).

Anche in un settore fortemente tecnicizzato ed automatizzato come la *cluster analysis*, ad esempio, si danno "situazioni di indecidibilità, che non possono essere risolte facendo ricorso unicamente a criteri statistico-matematici" (Biorcio, 1993, p. 18), ma che abbisognano di elaborazioni intensionali, oltre a "elementi conoscitivi e pragmatici di natura non quantitativa, legati alla 'conoscenza tacita'¹¹ del ricercatore" (Campelli, 1991, p. 42).

¹¹ Per conoscenza tacita intendiamo il "complesso di conoscenze, esperienze, convenzioni, abitudini concettuali e linguistiche che intervengono nell'osservazione (che in realtà è una costituzione) dei fatti" (Marradi, 1984, p. 166).

Ciò significa che in ogni fase della ricerca c'è spazio per la soggettività del ricercatore, *costretto* ad intervenire intensionalmente, teoricamente e con la sua conoscenza tacita, *prima*, scegliendo le variabili discriminanti (identificando così implicitamente i possibili *fundamenta divisionis* della tipologia finale), *durante* nelle scelte intermedie fra le diverse tecniche e le diverse opzioni, scelte che sono tutte portatrici di significati concettuali, e *dopo*, chiarendo i concetti legati ai gruppi formati e da interpretare.

Quindi, chiude Biorcio – e ci sentiamo di condividere in pieno – “il soggetto che opera la classificazione estensionale compie pertanto un processo di continua elaborazione-rielaborazione concettuale assimilabile alla classificazione intensionale” (Biorcio, 1993, pp. 19-20). Come avevamo detto in precedenza, siamo costantemente impegnati a formare e riformare concetti ed ogni volta che pensiamo, formuliamo od usiamo un concetto (la sua intensione), abbiamo in mente almeno un referente (la sua estensione).

1.4 LA POSIZIONE EPISTEMOLOGICA DELLA CLASSIFICAZIONE

Classificazione, tipologia e tassonomia, come abbiamo visto, sono strutture concettuali a crescente livello di complessità. Esse, come anche i concetti, una caratteristica ben precisa: “non sono né veri né falsi – anzi, non sono neppure pensabili come veri e falsi” (Marradi, 2007, p. 60).

Un concetto non afferma nulla, di conseguenza non c'è nessuna affermazione che possa essere verificata o smentita:

Se dico “il sole”, questa parola ha un significato, e lo suggerisce a chiunque mi stia ascoltando. Ma supponiamo di domandargli se è vero. Egli non può dare alcuna risposta, perché non c'è nulla che possa essere vero e falso [...] Fate invece che io dica “il sole esiste”. Di questa frase si può senz'altro dire se è vera o falsa. Questo perché la frase collega due distinti oggetti: uno è il sole, l'altro è l'esistenza. Non si dica che questo secondo concetto è implicito nel primo; infatti si può concepire il sole come non più esistente (Mill, 1843, I.I.2)

Questa semplice considerazione ha comportato due tipi di conseguenze: 1) per la sua natura non affermativa, la classificazione è stata da alcuni spinta ai margini della scienza o perfino oltre; 2) le classificazioni sono state, per qualche motivo, investite di significato ontologico ed ipostatizzate o reificate. Vediamo.

La principale funzione della classificazione è quella descrittiva. Essa ordina e semplifica l'infinità varietà del mondo sensibile, divide per criteri, individua categorie e dà loro un nome.

Proprio per questa sua peculiarità, la classificazione è stata spesso oggetto di due pesanti accuse (cfr. Marradi, 1993): la prima, più specifica, imputava ad essa il presunto ritardo delle scienze sociali nei confronti di quelle *forti*; la seconda, più generale e drastica, la indicava come forma di conoscenza elementare, inferiore, rifugio di scienziati incapaci.

La prima è una critica vecchia più nella concezione che nell'anagrafe, figlia del mal celato senso d'inferiorità delle scienze sociali, di cui si potrebbe fare tranquillamente a meno. Ancora nella seconda metà del secolo scorso possiamo leggere quanto segue:

Lo stesso successo di una classificazione tipologica può [...] congelare il livello della spiegazione [...]. Le classificazioni tipologiche nelle scienze sociali sono in parte responsabili del ritardo di una più potente spiegazione teoretica (Tiryakian, 1968, p. 179).

Una tale affermazione rivela un atteggiamento che vede nelle scienze forti ("esatte") il modello vincente da copiare per garantire a quelle sociali di raggiungere gli stessi successi. Successi che si ritiene siano dovuti al largo uso della misurazione nelle scienze fisiche. Dunque, finché la sociologia non abbandonerà la classificazione per la misurazione, resterà sempre al palo. Questa argomentazione ha sollevato varie obiezioni (cfr. Marradi, 1993) che si possono riassumere in quattro punti fondamentali:

- 1) nel caso le scienze sociali vogliano o possano raggiungere lo stesso prestigio di quelle fisiche, non è detto che debbano o possano farlo seguendo le stesse strade;

- 2) non solo la misurazione ma anche altri strumenti e numerosi fattori hanno contribuito al successo della fisica e affini;
- 3) la misurazione non sostituisce la classificazione, non è qualcos'altro, e qualcosa in più: “ogni livello di misurazione include, come requisito minimo, una procedura di classificazione” (Blalock, 1960, tr. it. 1970, p. 28¹²);
- 4) i fisici misurano proprietà concepite come *continuum* e per le quali hanno stabilito unità di misura, un tipo di proprietà per niente frequente nelle scienze sociali.

La seconda accusa parte dalla stessa premessa, l'incapacità dell'attività classificatoria di fornire spiegazioni, ma spinge la critica più in là, fino a portare la classificazione fuori dai confini della dignità scientifica. Chi sostiene questo punto di vista vede nella classificazione non solo l'incapacità, ma l'antitesi stessa della spiegazione.

Abbiamo già rapidamente osservato come la classificazione, facente parte del sapere pre-assertorio (come i concetti), sia tecnicamente esclusa dalla funzione esplicativa, riservata invece al sapere assertorio (teorie e, appunto, spiegazioni). È però erroneo, sulla base di ciò, ritenere il secondo tipo di sapere superiore al primo ed è ingenuo pensare che, al di là delle definizioni, quest'ultimo sia privo di utilità nel processo di comprensione e spiegazione.

Hempel (1952, tr. it. 1976), pur non risparmiando alcune critiche, annovera la classificazione fra i tre tipi fondamentali di formazione dei concetti scientifici. Gli altri due sono l'ordinamento non metrico e la misurazione ed Hempel li preferisce poiché essi sono più “idonei per la rappresentazione di caratteristiche graduabili” (Hempel, 1952, tr. it. 1976, p. 69): “gli oggetti e gli eventi del mondo nel quale viviamo non presentano le rigide delimitazioni richieste dagli schemi classificatori” (ibid., p. 71), i quali, confrontati ai concetti comparativi o quantitativi, sembrano “arbitrari e artificiali” (ibid., p. 71). Una posizione in apparenza (ed in buona sostanza) assai vicina alla critica da cui ci

¹² Vale la pena notare che Blalock non è un costruzionista o affini, ma l'autore di uno dei testi classici dell'approccio comportamentista alla ricerca sociale.

stiamo difendendo, ma che non giunge alle stesse frettolose conclusioni. Hempel, infatti, riconosce come talune classificazioni possano raggiungere una “certa portata sistematica” (ibid., p. 66), “elevata” perfino nel caso del “geniale raggruppamento” (ibid., p. 68) di Mendeleev, da egli riportato come esempio¹³.

Interessante, ed in un certo senso definitiva, è anche la posizione di Gil (cfr. 1981). La classificazione ha caratteristiche particolari che le consentono di passare dal piano descrittivo a quello esplicativo:

Mostrare classificando è una maniera embrionale di dimostrare, e le classificazioni sollecitano, per così dire, le teorie che le sosterranno compiutamente. Le classificazioni sono dunque costituenti a pieno titolo la conoscenza. [...] [Esse] congiungono analisi e sintesi [...] [e] si trovano al punto d'intersezione tra la rappresentazione non interpretata [...] di una fenomenologia [...] e una teoria propriamente detta: tra descrizione e spiegazione (Gil, 1981, p. 1042).

In questo senso vanno anche le parole di Marradi:

Se è vero che concetti e strutture concettuali (classificazioni, tipologie, tassonomie) non affermano nulla circa i loro referenti – e quindi non sono pensabili come veri o falsi – è anche vero che essi sono componenti necessarie di tutte le affermazioni possibili, e in un certo senso le predispongono; individuano infatti oggetti cognitivi a proposito dei quali si potranno formulare affermazioni pensabili come vere e false (Marradi, 2007, p. 61).

Riguardo al secondo problema, cioè il fatto che le classificazioni sono state spesso reificate, va detto che la struttura per organizzare i referenti è creata dal ricercatore; ovviamente questo non può implicare la pretesa che la realtà sia organizzata proprio in quel modo, non c'è “alcuna eccellenza esclusiva” (Jevons, 1874, p. 677) ed “è piuttosto grottesco” che, dopo aver costruito da noi classi e tipi, “cominciamo a domandarci quali sono le loro proprietà necessariamente vere in natura (Toraldo di Francia, 1986, p. 1986).

¹³ Vedi paragrafo successivo.

Marradi mostra come sia stato difficile per il pensiero occidentale¹⁴ raggiungere la “piena consapevolezza della natura stipulativa di classificazioni e tipologie” (Marradi, 2007, p. 60).

L’insidia maggiore viene dalla **fallacia essenzialista**: il mito della classificazione naturale, che ha attraversato lo spazio e il tempo, partendo da Platone e i suoi 5 *μεγιστα των γενων* (i più grandi di tutti i generi), passando dalle 10 categorie dei pitagorici; da Aristotele, per il quale l’osservazione empirica serviva ad eliminare l’accidentale ed isolare la sostanza immutabile, attraversando tutto il Medio Evo, per arrivare a Linneo, il cui motto era *tot numeramus species quot ab initio creavit supremus Ens*. Una fallacia che ha la sua origine nella confusione tra il piano ontologico e quello epistemologico, una confusione che induce a spacciare per oggettivo ciò che è creato dal soggetto pensante. Ciò porta a sostituire la parola *tipo* con la parola *struttura*, che ancora oggi, nonostante le dichiarazioni rassicuranti, suggerisce la reificazione delle tipologie come tentazione e trabocchetto ricorrente, che nasconde, sotto la versione metodologica della classificazione naturale come gruppo omogeneo¹⁵, il residuo tenace del vecchio essenzialismo.

Non meno grave appare la **fallacia scienziata**, evidente in Lamarck e Darwin, con il crollo della convinzione che le specie animali e vegetali fossero immutate dai giorni della creazione, un’ontologia statica abbattuta dai colpi dell’evoluzionismo. Evoluzionisti che, per via opposta, commisero lo stesso errore degli essenzialisti, confondendo il piano gnoseologico e quello ontologico: trovata colpevole di un’ontologia errata, la classificazione fu screditata come strumento cognitivo e fu ripudiata come anticaglia da conservatore. Come se questo non bastasse, un altro colpo lo diedero prima gli empiristi logici, poi i comportamentisti e gli operazionisti, amanti della misura, che presero a modello la fisica del diciannovesimo secolo.

¹⁴ Ma anche per quello orientale, vedi la teoria indiana delle categorie, esposta *Vaisesika sutra*, del primo secolo d.C., che mostra una visione essenzialista simile a quella di Aristotele.

¹⁵ Vedi prossimo paragrafo.

1.5 LA VALUTAZIONE DI UNA CLASSIFICAZIONE

Come sottolinea Marradi, “il fatto che i pre-asserti non siano pensabili come veri e falsi non comporta affatto la conseguenza che siano tutti equivalenti e fungibili (che uno valga l’altro). Essi possono essere, e sono, valutati in base ad altri criteri” (Marradi, 2007, p. 62).

Prima di tutto c’è la correttezza logica, le classificazioni devono cioè rispettare i requisiti di cui abbiamo parlato. La difficoltà in questo caso è fare i conti con il mondo reale, per cui “nella ricerca scientifica spesso si scopre che gli oggetti studiati resistono ai tentativi di sistemazione ordinata” (Hempel, 1965, p. 151).

Quando si effettua una classificazione è difficile, se non impossibile, affermare che le n classi individuate siano tutte le classi possibili per quello specifico genere. Prendendo ad esempio i tipi di governo sulla base del carattere e dell’indirizzo che li caratterizza, potremmo individuare: governo democratico, aristocratico, oligarchico. Ma chi ci garantisce, in realtà e al di sopra di ogni logico dubbio, che non ve ne siano altri che al momento ci sfuggono? Ovviamente, nessuno¹⁶. Vediamo quindi come l’esaustività sia un obiettivo raggiungibile solo con l’aggiunta di una classe residuale, che includa cioè tutti i casi esclusi dalle altre classi, uno stratagemma che è antico probabilmente quanto la classificazione stessa ed almeno quanto la *diairesis* platonica (cfr. Findlay, 1994; Reale, 2000), operazione in cui un genere viene dicotomizzato in due specie, una caratterizzata dalla presenza di un attributo, l’altra dalla sua assenza. Quando questa assenza è l’unica caratteristica comune agli oggetti della classe, la classe ha chiara natura residuale. La pratica del creare classi caratterizzate solo dalla comune assenza di un attributo, abbastanza ricorrente nelle scienze naturali,

¹⁶ Nessuno, da un punto di vista logico. Nella pratica di ricerca, si può godere di una garanzia provvisoria e contestuale fornita dalla *conoscenza della situazione*. È ovvio che questo discorso non vuole portare a derive relativistiche estreme che blocchino il corso dell’attività scientifica, ma vuole solo far riflettere sulle basi di questa attività per acquisire consapevolezza (sull’idea di verità come contesto, cfr. Kuhn, 1962).

è stata criticata da Henning (1950), che la definisce “parafiletica”, cioè caratterizzata da una scorretta applicazione delle procedure classificatorie

Detto delle critiche, va anche detto che un uso attento e sensato della classe residuale è non solo accettabile, ma consigliato e perfino necessario nelle scienze sociali e naturali. A questo scopo, è buona norma osservare delle piccole attenzioni (cfr. Marradi, 1993). La prima, nei casi che lo consentono, suggerisce di utilizzare non una sola, ma una pluralità di classi residuali, in modo di conservare un pur minimo contenuto semantico in ognuna di esse, recuperabile poi nelle successive analisi dei dati. Ad esempio, in una indagine elettorale si può sostituire la sola classe “altri partiti” con le classi “altri partiti di sinistra”, “altri partiti di destra”, “altri partiti confessionali”, “altri partiti locali”, ecc (Marradi, 1993, p. 26). In questo modo diventa possibile in un secondo momento ricostruire la totalità dei voti a sinistra, dei voti confessionali e così via.

Inoltre, anche se questo è un difetto spesso ineliminabile, si limita l’eterogeneità dei casi contenuti in classe residuale, che come abbiamo visto raggiunge il suo picco – altamente sconsigliabile – nella *diairesis*. Se questo è un difetto, per così dire, tollerato in termini *qualitativi*, esso non può essere tollerato in termini *quantitativi*. Vale a dire che una classe residuale mai deve contenere una percentuale troppo alta di casi. In questo caso, come in ogni distribuzione di frequenza eccessivamente sbilanciata, ogni analisi successiva viene vanificata e non resta altro da fare che cestinare la classificazione ed eventualmente ripeterla daccapo.

La situazione è più complicata per quanto riguarda la mutua esclusività. Il problema qui è sia sostanziale che formale. Da un punto di vista formale, l’errore gravissimo, che inficia la natura stessa della classificazione, in cui si può incorrere è quello di affiancare nella stessa classificazione due o più classi delle quali una sia specie dell’altra. Per chiarire, riprendiamo la tassonomia proposta in precedenza. Una riduzione che ponesse la classe *cattolico* affianco a quella *cristiano* sarebbe scorretta, perché, riproponendo i termini della definizione, i membri dell’estensione della prima sono anche membri dell’estensione della seconda. Fin qui i problemi sono relativi, riguardano il rapporto fra genere e specie e necessitano solo di una particolare attenzione da parte del classificatore.

Le cose vanno peggio dal punto di vista sostanziale, quando cioè alla rigida formalità della logica si oppone strenuamente la flessibile confusione del mondo esperito, e ci accorgiamo che “la mutua esclusività [...] è una proprietà degli schemi classificatori, ma raramente è anche una caratteristica degli oggetti da classificare” (Sandri, 1969, p. 84).

Il problema – se di problema si tratta – è che, come si accennava in precedenza, il mondo di cui facciamo esperienza ogni giorno, sia esso mondo sociale, fisico o psicologico, non è un universo in bianco e nero, è un mondo infinitamente e gradatamente grigio, più simile ad un fenomeno confuso e sfumato che ad uno chiaro e distinto (cfr. Kosko, 1991; 1993). Per assegnare un oggetto alla classe *a* invece che alla classe *b* serve spesso una buona dose di approssimazione, interpretazione, convenzione, tre elementi che, non si può negare, appartengono a buon diritto ad un lecito processo scientifico, ma che di tanto in tanto, nei casi più “sfortunati” o se preferiamo più incerti, possono sfociare in forzature che sarebbe preferibile evitare.

Pensiamo, ad esempio, all’indice di “povertà” usato dall’ISTAT, una linea netta sul *continuum* del reddito pro-capite nazionale. Sotto la linea (che, ad esempio, nel 2003 corrispondeva a 874,74 euro per una famiglia di due persone, nel 2004 a 919,98 e nel 2005 a 936,58 – cfr. Statistiche in breve, “La povertà relativa in Italia nel 2004” del 6 ottobre 2005; Statistiche in breve, “La povertà relativa in Italia nel 2005” del 11 ottobre 2006; reperibili *online*), si è ufficialmente “poveri”, sopra – fosse anche per pochi euro – non lo si è più.

Rimanderemo al prossimo capitolo una trattazione più sistematica dei problemi legati a questa impostazione; per ora ci limitiamo a dire che la mutua esclusività è sì un requisito logico fondamentale della classificazione, ma anche un criterio regolativo che, in quanto tale, impone alcuni vincoli e pone alcuni problemi ad essi relativi. Se, come osserva Marradi, non ci sono problemi a rispettare questo criterio logico quando si classificano oggetti astratti, altra cosa è classificare referenti come oggetti empirici. In questo caso la destrezza e la precisione normalmente richieste ad un ricercatore non sono sufficienti: “le forme viventi non ammettono rigorose linee di demarcazione” (Jevons, 1874, p. 724). Ne consegue che le soluzioni possibili ed utilizzabili di volta in volta possano, ed in

certa misura debbano, essere sottoposte ad un giudizio che sia capace di metterle totalmente in discussione e che lasci al ricercatore la possibilità di sondare altre strade, che vadano oltre non solo alla soluzione possibile, ma al criterio regolativo stesso, indipendentemente da quanto “costoso” possa essere questo superamento in termini di abbandono di pratiche e forme *mentis* tipiche della nostra cultura occidentale, per come essa si è sviluppata dagli insegnamenti aristotelici in poi. Questa, si vedrà più avanti, è la direzione nella quale il nostro lavoro si muoverà.

Logica a parte, “il pensiero occidentale è praticamente unanime sul criterio con cui vengono valutati i concetti: è l’utilità” (Marradi, 2007, p. 62). Parafasando ciò che Popper dice in un altro contesto, “la verità non è l’unico scopo della scienza. Vogliamo qualcosa di più della verità pura e semplice: cerchiamo la verità interessante” (Popper, 1969, p. 184).

Esempio classico a riguardo è “il geniale raggruppamento degli elementi chimici in base al sistema periodico, i cui principi fondamentali hanno consentito a Mendeleev (cfr. Mendeleev, 1994) di prevedere l’esistenza di parecchi elementi mancanti nel sistema e di anticiparne con grande accuratezza diverse proprietà fisico-chimiche” (Hempel, 1952, tr. it. 1976, p. 68).

Il sistema di Mendeleev si poggiava interamente sulla teoria dell’atomismo, per la quale si ipotizzava che le proprietà degli elementi, e di conseguenza anche quelle dei corpi semplici e composti, fossero in dipendenza periodica del loro peso atomico: disponendo gli elementi su una tabella in ordine crescente di peso atomico, ad intervalli regolari o, appunto, periodici, si trovavano elementi con proprietà simili.

Nel febbraio 1869, nel preparare un rapporto destinato alla Società Chimica Russa, Mendeleev scoprì che dentro la sua tabella, in seguito chiamata tavola periodica degli elementi, c’erano dei posti vuoti. Egli non cedette alla tentazione di rivedere la sua teoria in cerca di un errore, ma ipotizzò che i posti vuoti sarebbero stati occupati da elementi ancora ignoti, i cui pesi atomici dovevano essere pari a 2, 20 e 36. Ebbe ragione ed i tre posti furono in seguito occupati rispettivamente dall’elio (posizionato tra idrogeno e litio), dal neon (tra fluoro e sodio) e dall’argon (tra cloro e potassio).

Il cerchio si era chiuso perfettamente: partendo da una teoria si era costruito un sistema classificatorio, tanto dipendente da essa da risultare in sua assenza inimmaginabile. Dopodiché il sistema classificatorio – in seguito a successive scoperte – aveva finito per rinforzare clamorosamente la teoria, in quanto le previsioni implicate erano risultare confermate dagli eventi.

Difficile trovare una classificazione più riuscita e più feconda euristicamente. Vale inoltre la pena sottolineare come in questo esempio l'elemento contestuale non sia rappresentato unicamente dalla guida teorica, ma anche dall'aspetto specificatamente più tecnico. Dal tempo di Mendeleev, infatti, la tavola periodica ha subito notevoli modifiche, sono stati scoperti nuovi ulteriori elementi, e sono aumentate le conoscenze sulla struttura dell'atomo. Queste a loro volta hanno avuto una ricaduta teorica, per cui le proprietà degli elementi non vengono più relazionate al peso atomico, ma al numero atomico, cioè al numero di protoni contenuti nel nucleo di un atomo. Un esempio questo, fra gli innumerevoli che può fornire la scienza, di come vadano di pari passo progresso tecnologico e raffinamento teorico.

Secondo Marradi, “nel caso della classificazione e delle altre strutture concettuali, al criterio dell'utilità (che in questi casi riguarda principalmente la rilevanza dei *fundamenta divisionis* scelti ai nostri fini pratici e cognitivi) se ne aggiungono almeno altri due: una valutazione (in larga misura rimessa alla conoscenza tacita e assai difficilmente esplicitabile) del grado di corrispondenza fra le soglie divisorie fra le classi (o i tipi, o i *taxa*) e le zone di maggiore discontinuità nell'insieme dei referenti; e una valutazione di quanto bene siano state bilanciate la sensibilità (che induce a moltiplicare classi, tipi o *taxa* per accrescere la precisione con cui sono rappresentate le caratteristiche dei referenti) e la parsimonia (che invece suggerisce di limitare il numero di classi, tipi o *taxa* per mantenerne un miglior controllo concettuale)” (Marradi, 2007, p. 62).

Riguardo al secondo punto abbiamo ben poco da dire. La soluzione non può essere che essere pragmatica. Oltre al numero delle classi o tipi o *taxa*, un bilanciamento fra ricchezza di informazione e necessità di economizzare la ricerca può essere fatto anche per il numero di *fundamenta* da considerare o di proprietà da inserire in matrice. Buon senso suggerisce di raccogliere più informazioni

possibili all'inizio, visto che quelle non raccolte all'inizio sono perse definitivamente, e poi operare i tagli del caso, in termini di riduzione tipologica o selezione/elaborazione delle variabili.

Il primo punto, invece, con l'accento al grado di corrispondenza fra le soglie divisorie delle classi e le zone di maggiore discontinuità nell'insieme dei referenti ci fa venire in mente l'antica questione dell'adeguatezza della classificazione alla realtà o, in altri termini, delle classificazioni naturali.

Fatto salvo quanto detto a proposito della fallaccia essenzialista e scienziata, e facendo attenzione a non commettere "l'imprudenza di sostituire alla parola tipo la parola struttura" (Boudon e Bourricaud, 1982, tr. it. 1991, p. 544), possiamo ricanalizzare la distinzione naturale/artificiale in termini esclusivamente metodologici, sostituendo stavolta l'espressione "classificazione naturale" con quella "*cluster* naturale".

I *cluster naturali* sono "gruppi ben separati fra loro, fra i quali esistono soluzioni di continuità" (Di Franco, 1997, p. 82). In questo caso, "la 'naturalità' della classificazione viene infine ridotta ad una specifica 'forma' della 'struttura dei dati'" (Biorcio, 1993, p. 12).

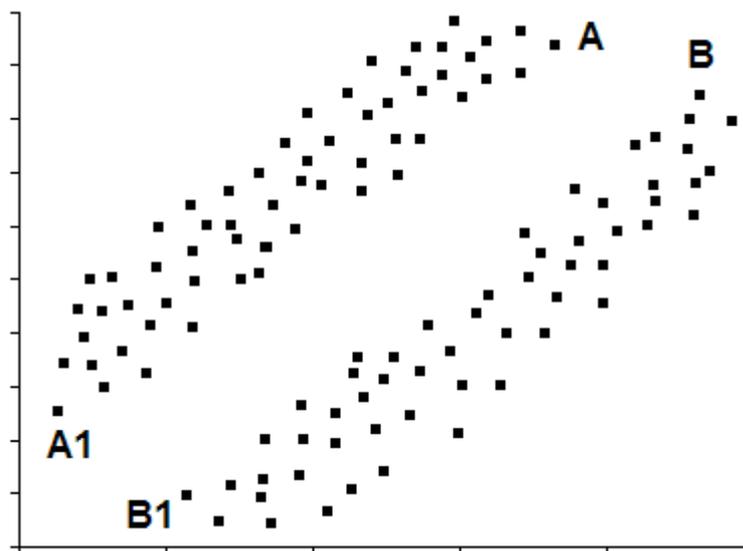
Se, insomma, si ha la fortuna di effettuare una classificazione automatica ed ottenere come risultato due gruppi ben distinti, si può anche parlare di *cluster naturali*, facendo attenzione a non spingere le nostre parole fino al piano ontologico. Di vera fortuna si tratta, tra l'altro, visto che nella concreta pratica di ricerca sono pochissimi i casi appena descritti, mentre di norma ciò che appare al ricercatore è una nuvola di punti distribuiti più o meno omogeneamente sugli assi.

Ciò non toglie che serva comunque cautela per non trasformare la fortuna in errore: si consideri il seguente esempio tratto da Biorcio (1993, pp. 26-27). Nella figura 1.4 vediamo la posizione spaziale di un insieme di oggetti in relazione a due variabili.

A prima vista saremo certi di aver individuato due *cluster naturali*, in quanto il nostro occhio tenderà a riunire i punti contigui nello stesso gruppo. Sarebbe un grave errore. Se la variabile in ascissa fosse il reddito e quella in ordinata il livello d'istruzione, è palese che l'individuo *A* non apparterebbe allo stesso tipo dell'individuo *AI*. *A* sarebbe molto più simile a *B*, che solo

apparentemente appartiene all'altro *cluster* “*naturale*” (virgolette necessarie, in questo caso), e *A1* a *B1*.

Fig. 1.4 – Esempio di *clusters* apparentemente “*naturali*”



Insomma, se la ricerca della classificazione *naturale* è intesa in senso *metodologico*, come individuazione di *gruppi omogenei* fra di loro, allora l'espressione è accettabile, ma rischia seriamente di essere equivoca e fuorviante. Se non meglio specificata il nostro consiglio è quello di abbandonarla, anche perché “la distinzione proposta [fra classificazione naturale e classificazione artificiale] non corrisponde alla gran parte delle situazioni di ricerca reali” (Biorcio, 1993, p. 26), per cui al ricercatore conviene “riconoscere la relativa artificialità della sua costruzione, funzionale ai propri obiettivi di ricerca” (ibid., p. 25), e mettersi così al riparo da ogni pericolo.

In fondo le parole chiave sono proprio queste: *funzionale agli obiettivi di ricerca*, una clausola contenuta anche nella versione metodologica della distinzione, cioè nell'individuazione di gruppi omogenei. Omogenei rispetto a cosa? Non esistono gruppi omogenei in assoluto, ma solo relativamente a date variabili, in dati contesti di ricerca, per dati fini conoscitivi. Vogliamo concludere questo discorso con un illuminante esempio ripreso da Dewey (1938, tr. it. 1974).

Qualità che sono estremamente dissimili l'una dall'altra nel loro presentarsi immediato (o 'sensibile'), vengono assimilate tra loro (ovvero vengono assegnate alla stessa specie), quando si trovi che lo stesso modo di risposta produce conseguenze analoghe; cioè, conseguenze soggette all'applicazione di una sola e medesima operazione successiva. Un bagliore di lampo è molto differente nella sua consistenza sensibile dalla scintilla elettrica quale era stata osservata prima di Franklin, come dall'attrazione esercitata dall'ambra strofinata, come dalla sensazione di formicolio che si percepisce quando in certe condizioni atmosferiche una persona che ha strascicato i piedi tocca un'altra persona. L'assimilazione fra loro di tali fenomeni e di un gran numero d'altri, come appartenenti alla specie *elettromagnetica*, non si è verificata attraverso la ricerca e il reperimento di qualità immediate 'comuni'. Essa venne effettuata mediante l'impiego di modi d'operazione e prendendo nota delle loro conseguenze. [...] 'Comune' designa non certe qualità, ma certi modi d'operazione. (Dewey, 1938, tr. it. 1974, pp. 316-17).

Tale classificazione di elettromagnetismo, ben poco "naturale" (o immediata e sensibile, come dice Dewey), è un esempio paradigmatico di come l'operazione del classificare sia conforme alle operazioni tipiche della logica dell'indagine e sia adatta a prestarsi al raggiungimento dello scopo che la stessa si prefigge: l'utilità euristica o, per dirla alla Popper, la verità interessante.

Il cerchio si è, dunque, chiuso: siamo all'interno della pura logica strumentale. L'unico criterio valido e ragionevole per valutare una classificazione è la sua *funzionalità* rispetto al contesto cognitivo in cui essa si colloca.

CAPITOLO 2. LE PROCEDURE DI CLUSTER ANALYSIS

Il nostro lavoro è iniziato dalla definizione del concetto di *classificazione*, si è parlato dei suoi diversi significati, dei requisiti, del ruolo epistemologico. Siamo a metà del percorso che ci porterà all'elaborazione empirica dei dati: questo capitolo parlerà delle *procedure di classificazione automatica*, l'applicazione su base algoritmica, con uso di matrice dei dati, di quella classificazione *estensionale* di cui abbiamo discusso nel primo capitolo.

Sono moltissime le strategie elaborate per la classificazione dei dati che possono essere utilizzate nella ricerca sociologica. Discuteremo, brevemente, di alcune tecniche di *cluster analysis*, per illustrarne caratteristiche, pregi e difetti; una parte di queste sarà poi utilizzata in un concreto caso empirico nel capitolo finale.

La *cluster analysis* è un insieme di tecniche che hanno lo scopo di individuare all'interno di un campione, in base ad alcune caratteristiche ritenute discriminanti, dei gruppi il più possibile omogenei all'interno ed eterogenei all'esterno.

Storicamente la nascita della *cluster analysis* viene fatta risalire al 1939, anno in cui Tryon pubblicò l'omonima monografia. In essa, però, a differenza della maggior parte dei lavori che seguiranno, la *cluster analysis* non opera raggruppamento automatico dei casi, o oggetti (*O-analysis*), ma delle variabili (*V-analysis*): obiettivo di Tryon era infatti quello di identificare gruppi di variabili simili con una tecnica più semplice e diretta dell'analisi fattoriale. Ad ogni modo,

all'epoca, al di là della prospettiva del raggruppamento, la tecnica ebbe scarso successo in quanto tutti i calcoli dovevano essere effettuati a mano.

Successivamente, dagli anni '60 in poi, l'attenzione verso l'analisi dei gruppi si rinnova, grazie anche, o forse soprattutto, alla possibilità di utilizzare i calcolatori elettronici, che iniziano a fare il loro ingresso nelle università americane intorno alla seconda metà degli anni '50. Aumentano dunque le applicazioni e le pubblicazioni a riguardo: fra le altre, per il grande contributo offerto, segnaliamo l'opera di due biologi, Sokal e Sneath, *Principles of Numerical Taxonomy*, del 1963. La sua importanza risiede soprattutto nell'aver formulato uno *schema di base*, tuttora valido, per l'analisi dei gruppi (cfr. anche Biorcio, 1993, p.15):

- a. la raccolta del maggior numero di informazioni rispetto ai diversi tipi di organismi studiati;
- b. l'uso di queste informazioni per stimare il grado di similarità;
- c. il ricorso a tecniche di *clustering* per classificare nello stesso gruppo gli organismi relativamente simili.

Altri lavori da segnalare sono gli ormai classici *Cluster Analysis*, scritto dallo stesso Tryon a quattro mani con Bailey (1970), e *Cluster Analysis for application* di Anderberg (1973). A questo punto la situazione è già notevolmente più complessa: dagli anni '70 le risorse computazionali aumentano notevolmente, e le proposte di algoritmi si moltiplicano, così come le pubblicazioni (cfr. Edwards, Cavalli-Sforza, 1965; MacQueen, 1967; Hartigan, 1975; Blashfield, Aldenderfer, 1978; Blashfield, 1980; Jain, Dubes, 1988; Huang, 1998; Jain, Murty, Flynn, 1999). Aggiungiamo che lo sviluppo delle varie tecniche di analisi dei gruppi si frammenta in numerosi ambiti disciplinari scarsamente comunicanti, e ci ritroviamo alla situazione odierna, per cui parlare di *cluster analysis* significa muoversi in un territorio piuttosto ampio, un tema tanto ostico da meritare intere trattazioni a parte. Tra queste, in Italia, citiamo i classici lavori di Ricolfi (1992) e Biorcio (1993). Noi ci limiteremo ad un'analisi sommaria.

2.1 LA CLUSTER ANALYSIS COME ARCIPELAGO

Per descrivere la vastità dell'argomento *cluster analysis* si cita spesso la fortunata metafora di Ricolfi (1992), secondo il quale l'analisi dei gruppi più che una tecnica o un insieme di tecniche, sarebbe un vero e proprio *arcipelago*. Una stima *prudente* degli algoritmi proposti, infatti, suppone che essi siano qualche centinaio, una stima più audace, ma non inverosimile, si attesta quasi al migliaio. Un'enormità disarmante, soprattutto se paragonata alla sostanziale compattezza che si riscontra negli altri settori dell'analisi dei dati. Perché questa differenza? Se la capacità di calcolo è aumentata per tutti, perché solo nella *cluster analysis* essa ha portato alla proliferazione incontrollata di tecniche, metodi ed algoritmi?

Ricolfi propone due risposte, una legata all'altra: l'*ancoramento alle righe* della matrice dei dati, e la *manca di una teoria forte*.

L'analisi dei gruppi di norma opera sulle righe della matrice, sui casi, ciò che prima abbiamo definito *O-analysis*. In sociologia, nella maggior parte delle circostanze, il numero dei casi supera di gran lunga quello delle variabili. Dunque, una tecnica ancorata alle righe va incontro ad un *onere computazionale* di gran lunga superiore ad un'altra (ad esempio la riduzione fattoriale) ancorata alle colonne. Il fatto di operare su un numero di elementi dell'ordine delle centinaia o delle migliaia, anziché delle unità o delle decine, ha legato la *cluster analysis*, più di qualsiasi altra tecnica, allo sviluppo delle risorse di calcolo. Questo significa non solo un ritardo in termini di *tempo*, di anni, perché si potessero mettere in pratica o perfino concepire alcune idee, ma soprattutto che la stessa direzione in cui tali idee si è sviluppata è stata fortemente, necessariamente condizionata dall'enorme mole di calcoli cui bisognava andare incontro. Dunque, più di tutti, lo studioso di *cluster analysis* ha sviluppato una sensibilità, un'attenzione agli aspetti computazionali, che si è presto trasformata in *dipendenza*.

Arriviamo, così, al secondo punto. Ricolfi segnala come la *cluster analysis* sia l'unico settore dell'analisi dei gruppi in un cui manca una forte teoria, capace di inquadrare il problema in una prospettiva unitaria. La maggior parte degli algoritmi, infatti, sono soltanto varianti tecniche di un unico principio teorico: il

principio di *massima informazione*, che ordina di costruire gruppi minimizzandone la variabilità interna. Il costante tentativo di realizzare tale obiettivo ha portato alla tipica *confusione fra metodi ed algoritmi* che oggi caratterizza l'analisi dei gruppi. Anziché definire teoricamente un metodo, e poi realizzare un algoritmo in grado di metterlo in pratica, si trova un algoritmo che “funziona”, che classifica decentemente, verosimilmente, un insieme e lo si erge a metodo.

Per superare questa *impasse* Ricolfi propone un nuovo *software* (HELGA: cfr. Ricolfi, 1992), basato su un algoritmo (*hill-exploring*) capace di rompere tale monopolio teorico e di rendere visibili, oltre a strutture non standard di gruppo, anche due situazioni in genere ignorate nella ricerca empirica: quella in cui un problema di raggruppamento ammette *più di una* soluzione, e quella in cui non ne ammette *nessuna*. Senza entrare nello specifico della soluzione avanzata da Ricolfi, vogliamo soffermarci su questa ultima considerazione per segnalare un'importante caratteristica della *cluster analysis*, che non è stata ancora resa esplicita.

Riprendiamo la classica distinzione fra tecniche e modelli. Una tecnica d'analisi mira a riprodurre sinteticamente una informazione originaria, dunque dà sempre un risultato. Un modello sottopone a controllo empirico un'ipotesi di relazione fra variabili, dunque non dà sempre un risultato o almeno non sempre un risultato positivo: un modello sottoposto ad analisi può essere modificato o rigettato del tutto. L'analisi dei gruppi rientra fra le tecniche, e di conseguenza dà sempre un risultato. Ma questo non significa che tale risultato sia sempre buono o utile. Approfondiremo l'argomento quando parleremo di valutazione di una classificazione, ma già ora possiamo metterci in guardia dall'uso indiscriminato delle varie tecniche: di volta in volta, deve essere valutata la soluzione più adatta ad ogni situazione, altrimenti il rischio che si corre è che il proprio lavoro, prima ancora che erroneo, sia inutile.

Siamo pronti a questo punto ad una rapida analisi delle tecniche di *clustering* più usate, premettendo che essa sarà inevitabilmente incompleta e parziale. Il deficit di teoria di cui abbiamo prima parlato, infatti, incide anche su

un altro aspetto del problema: l'*incertezza dei confini*, sia interni che esterni, di questo settore di ricerca. Ricolfi, a tale riguardo, conclude ironicamente:

con una punta di malizia si potrebbe dire che la cluster analysis è forse l'unica disciplina autofalsificante: con sorprendente monotonia quasi tutti i manuali e le introduzioni esordiscono proponendo una classificazione delle tecniche di classificazione, per concludere mestamente che la classificazione proposta è largamente insoddisfacente. (Ricolfi, 1992, p. 13)

2.2 LE OPERAZIONI PRELIMINARI

Prima ancora di entrare nel vivo di un processo di classificazione automatica, il ricercatore deve effettuare dei passaggi necessari che influenzeranno il risultato finale. Questa è una delle fasi in cui, pur in ambito di analisi quantitativa e formalizzata, l'intervento soggettivo risulta, lo ripetiamo ancora una volta, decisivo. Vediamo brevemente quali sono tali operazioni preliminari¹:

- la scelta delle variabili in base alle quali classificare;
- il trattamento di tali variabili: possono essere standardizzate se riferite ad unità di misura diverse, si può attribuire loro un diverso peso in base alla diversa importanza che esse assumono, possono essere eliminate o sottoposte ad analisi fattoriale, se fortemente correlate fra loro, per evitare gravi distorsioni;
- la scelta della misura di similarità o distanza più idonea, necessaria per misurare la distanza fra caso e caso.

Le decisioni prese riguardo queste operazioni non sono di minore importanza rispetto alla decisione se utilizzare questa o quella tecnica: nella classificazione, che è un processo continuo e complesso, ogni scelta ha ripercussioni sul risultato finale. A questo punto, dovrà essere decisa quale tecnica

¹ Per maggiori approfondimenti rimandiamo a Biorcio (1993).

utilizzare e la nostra analisi continuerà con l'esposizione di quelle più utilizzate, quelle tradizionali.

2.3 LE TECNICHE TRADIZIONALI

La distinzione fondamentale all'interno delle tecniche tradizionali di *clustering* è fra quelle gerarchiche e quelle non gerarchiche.

Le tecniche gerarchiche si dividono in agglomerative e scissorie; quelle scissorie in politetiche e nomotetiche.

Le tecniche non gerarchiche si distinguono in metodi di suddivisione iterativa e metodi basati sulla programmazione matematica; tra i metodi di suddivisione iterativa si trovano le famiglie *K-Medie* e *Hill-Climbing*.

Questa è la suddivisione più largamente accettata: cominciamo a vedere le tecniche gerarchiche.

2.3.1 LE TECNICHE TRADIZIONALI GERARCHICHE

Il punto di partenza delle tecniche gerarchiche non è la matrice dati *case x variabili*, ma la matrice delle distanze o delle dissimilarità *case x case*. Prima di scegliere l'algoritmo, dunque, bisogna scegliere quale misura di similarità o distanza adottare.

Lo stadio iniziale delle tecniche gerarchiche agglomerative è rappresentato dalla presenza di n gruppi quanti sono gli n casi del campione; dopodiché cominciano una serie di unioni: nel primo passaggio si uniscono i due casi più vicini, poi si ricalcola la matrice delle distanze e si aggregano gli altri due gruppi (o casi) più simili della matrice. Si procede in questo modo fino alla formazione di un unico grande gruppo che contenga la totalità dei casi. Il numero delle iterazioni in cui si articola l'intero processo è pari ad $(n-2)$, dato che la prima (n gruppi per n casi) e l'ultima (un solo gruppo contenente tutti gli n casi) sono partizioni ridondanti.

Il principio degli algoritmi scissori è l'inverso: il punto di partenza è un unico grande gruppo a partire dal quale vengono effettuate divisioni fino al raggiungimento di n gruppi da un caso a testa. Gli algoritmi divisivi sono pochissimi, perché scarsamente applicabili: essi infatti prevedono una mole di calcoli proibitiva, e possono essere applicati solo a poche decine di casi. Sono dunque assenti nei *software* più diffusi e di scarso interesse per i ricercatori sociali, che lavorano sempre con campioni molto numerosi.

Gli algoritmi agglomerativi, invece, sono proliferati nel corso degli anni in maniera smisurata ed ormai se ne contano diverse centinaia. Qui elencheremo brevemente solo le tecniche presenti nei *software* più utilizzati nella ricerca sociale:

- *legame singolo*: la distanza fra due gruppi è uguale alla distanza fra i due casi più vicini;
- *legame completo*: il criterio è contrario al precedente, la distanza fra due gruppi è data dalla distanza fra i casi più lontani;
- *legame medio tra i gruppi*: la distanza fra i gruppi è data dalla media aritmetica di tutte le possibili distanze fra i casi dei due gruppi;
- *legame medio entro i gruppi*: la distanza fra i due gruppi è la media delle distanze tra tutte le possibili coppie di casi del nuovo gruppo;
- *tecnica di Ward*: si basa sul concetto di inerzia, aggregando ad ogni iterazione i gruppi o casi che provocano la minima perdita di varianza all'interno della nuova classe;
- *tecnica del centroide*: per ogni gruppo si calcola un baricentro, e la distanza fra due gruppi è quella fra i baricentri.

Il pregio maggiore delle tecniche gerarchiche è che può essere scelto a posteriori il numero dei gruppi che si preferisce, dopo aver analizzato il vasto numero di partizioni offerte dalla rappresentazione grafica del *dendrogramma*.

Il difetto che maggiormente si imputa a tali tecniche è che possono essere applicate soltanto a campioni non troppo numerosi. Questo è, però, vero solo in parte, visto che i nuovi *software* ed i nuovi elaboratori riescono a lavorare anche su molti casi.

Il vero difetto sta nel fatto che le iterazioni formano un gruppo *inscindibile*, a cui verranno semplicemente uniti altri casi o altri gruppi: in questo caso potrebbero dunque essere aggregati casi che, presi da soli, non sarebbero forse mai stati uniti.

2.3.2 LE TECNICHE TRADIZIONALI NON GERARCHICHE

Il principio sottostante alle tecniche non gerarchiche è quello di partire da una partizione qualsiasi e poi spostare i casi da un gruppo all'altro, fino all'individuazione della *classificazione migliore*.

Gli algoritmi proposti sono moltissimi ma seguono tutti lo stesso criterio; l'unica cosa che cambia è ciò che si intende per *classificazione migliore* ed il modo per perseguirla.

Il grande vantaggio delle tecniche non gerarchiche è che possono classificare un numero molto elevato di casi, per cui sono adeguate all'applicazione nella ricerca sociale.

La loro caratteristica peculiare è che il numero dei gruppi deve essere deciso a priori. Abbiamo detto caratteristica perché essa è a meta strada fra il pregio ed il difetto. Pregio se il ricercatore ha adeguata conoscenza, teoria di riferimento e ipotesi di lavoro sul fenomeno in esame; difetto se tali nozioni mancano. Spesso, in questo caso, si consiglia l'uso esplorativo di tecniche gerarchiche per avere un'idea del numero di gruppi più appropriato e della configurazione della partizione iniziale.

La dipendenza da tale partizione è, invece, la debolezza più grande di tali tecniche. Se la prima partizione non fosse buona, l'intero risultato sarebbe compromesso. Per questo si dice che le tipologie non gerarchiche siano poco stabili, perché formate da algoritmi che raggiungono solo un *equilibrio locale*, anziché globale. Si parla di equilibrio locale perché lo spostamento dei casi provocherebbe un peggioramento della funzione, ma si tratterebbe di un peggioramento temporaneo e destinato a raggiungere, con ulteriori spostamenti, un altro equilibrio, forse migliore del precedente: insomma i punti di equilibrio

sono molti, ma il procedimento iterativo trova solo quello più vicino. Quando parleremo di stabilità della classificazione vedremo come ovviare al problema.

Abbiamo detto che esistono, fra le tecniche non gerarchiche, tre famiglie: *k-medie*, *hill-climbing* e tecniche basate sulla programmazione matematica; cominciamo dalle prime.

Il procedimento inizia da una partizione qualsiasi, vengono calcolati i baricentri ed effettuato il riallocaimento (lo spostamento di un caso da un gruppo all'altro) dei casi più vicini al baricentro di un altro gruppo. Poi si calcolano i nuovi baricentri e si ripete la stessa operazione finché si raggiunge la convergenza, cioè finché nessun caso si sposta o non c'è variazione significativa.

I metodi *hill-climbing* raggiungono la soluzione finale allo stesso modo, partendo da una partizione qualsiasi, attraverso iterazioni intermedie ed effettuando spostamenti di casi. Il criterio che le guida è però l'ottimizzazione di una funzione (la *funzione obiettivo*). Alcuni autori, fra cui Anderberg (1973) e Biorcio (1993) ritengono forzata la distinzione fra *hill-climbing* e *k-medie*, in quanto anche quest'ultime seguirebbero una funzione obiettivo implicita: la distanza casi-baricentri.

Ricolfi (1992) esprime riserve su questi metodi sottolineando che le funzioni obiettivo maggiormente utilizzate sono semplicemente diverse varianti del medesimo principio: il *teorema della decomposizione della matrice delle varianze e covarianze*:

$$\underline{T} = \underline{B} + \underline{W}$$

dove \underline{T} è la matrice complessiva, \underline{W} la media ponderata di tutte le matrici delle varianze e covarianze locali e \underline{B} la matrice delle varianze e covarianze calcolata fra i baricentri dei gruppi.

In effetti, mentre le tecniche gerarchiche sono davvero alternative fra loro, quelle non gerarchiche (sia *k-medie* che *hill-climbing*) sembrano sottili varianti di uno stesso principio e risultano difficili da distinguere l'una dall'altra. Elenchiamo, ad ogni modo, le funzioni obiettivo più utilizzate:

- la minimizzazione della traccia² di \underline{W} , oppure, che è lo stesso, la massimizzazione della traccia di \underline{B} ;
- la minimizzazione del rapporto tra i determinanti delle matrici \underline{W} e \underline{T} :

$$|\underline{W}| / |\underline{T}|$$

- la massimizzazione della traccia di $(\underline{B} \cdot \underline{W}^{-1})^3$;
- la massimizzazione della traccia di $(\underline{B} \cdot \underline{T}^{-1})$.

Un brevissimo accenno, infine, ai metodi basati sulla programmazione matematica: come gli *hill-climbing* si propongono di ottimizzare una funzione obiettivo ma, a differenza di questi, gli spostamenti dei casi sono solo “virtuali”.

2.3.3 LE STRATEGIE MISTE ED ALTRE TECNICHE

Il ricorso ad una strategia mista si rivela utile quando gli algoritmi scelti di volta in volta non riescano a superare problemi di calcolo o non diano risultati soddisfacenti.

Esse si basano fondamentalmente sull'uso alternato e combinato di diverse tecniche, in modo di esaltare i pregi e limitare i difetti delle stesse. La strategia mista più comune è basata, semplicemente, sull'applicazione, in frequenza di algoritmi gerarchici e non gerarchici.

Un altro utilizzo di strategie miste può servire all'individuazione di *cluster* di natura diversa o, in altre parole, fondati su una diversa concezione di gruppo. Tutte le proposte finora analizzate sottendono una definizione di gruppo basata sulla *vicinanza* fisica, spaziale. Ci sono però proposte che collegano il gruppo ad un concetto di *densità*, per cui il gruppo è una parte dello spazio ad alta densità, circondato da zone a bassa densità. Sono numerosi, ma poco utilizzati, gli algoritmi proposti in tal senso: non ne parleremo in questa sede.

² Si può parlare di traccia solo se la matrice è quadrata: la traccia è la somma dei numeri collocati lungo la diagonale principale.

³ \underline{W}^{-1} significa l'inversa di \underline{W} .

2.4 VECCHI RISULTATI E NUOVE PROPOSTE

Tutte le tecniche finora analizzate, che sono le più utilizzate nella ricerca sociale, seguono un procedimento di classificazione di tipo *classico*. Tale procedimento individua gruppi che rispettino i seguenti criteri⁴:

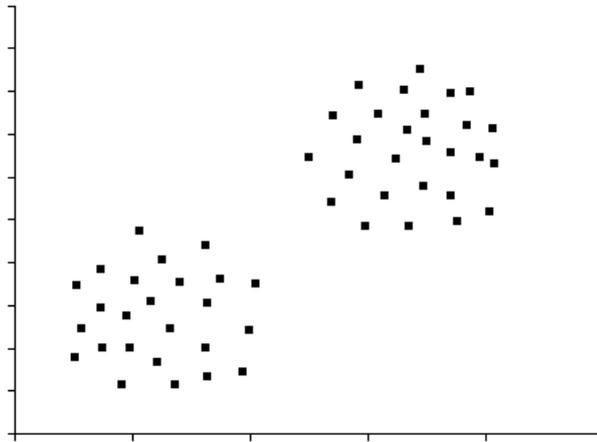
- *esaustività*: ogni caso appartiene almeno ad un gruppo, nessun caso è escluso dalla classificazione;
- *mutua esclusività*: un caso può appartenere soltanto ad un gruppo, non sono ammessi gruppi sovrapposti, i gruppi sono totalmente disgiunti fra loro;
- *appartenenza binaria*: seguendo i principi della logica classica, un caso può solamente appartenere o non appartenere ad un gruppo, non sono ammesse posizioni intermedie.

La *cluster analysis* classica, dunque, produce gruppi ben distinti e separati, effettuando una *sintesi* dell'informazione originaria molto efficace e maneggevole. Per questo motivo, finora, gli algoritmi classici sono quelli che hanno avuto più spazio nella discussione e nell'applicazione dell'analisi dei gruppi. Hanno però un limite di cui, col tempo, è stata presa sempre più coscienza.

Gli algoritmi classici partono da una situazione teorica di *clusters naturali*, come vediamo nella figura 2.1. In questa circostanza, con le proiezioni dei casi su piano bidimensionale così nettamente separate, le procedure tradizionali non avrebbero alcun problema ad individuare la partizione adeguata.

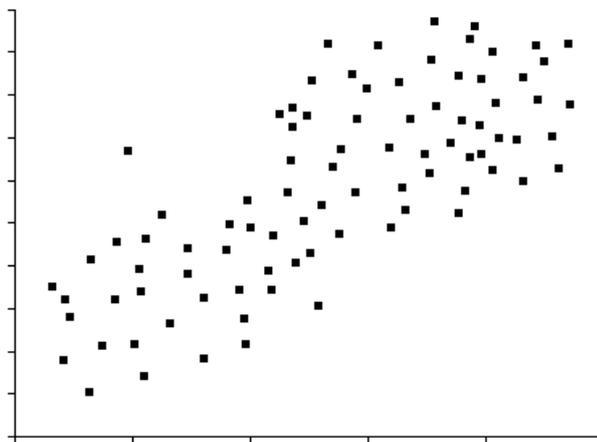
⁴ Dei requisiti formali di una classificazione avevamo già parlato nel capitolo precedente, ma lì si parlava di risultati provenienti sia da classificazioni intensionali che estensionali. Qui l'attenzione è rivolta in modo particolare ai risultati di una classificazione estensionale automatica.

Fig. 2.1 – Esempio di clusters naturali



Ma questa situazione è assai *rara* nelle concrete esperienze di ricerca empirica, dove spesso i casi si distribuiscono in modo uniforme e granulare e la nuvola dei punti assume l'aspetto caratteristico di un *blocco compatto*.

Fig. 2.2 – Esempio di granularità



È qui che emergono i limiti dell'analisi classica: la rappresentazione della realtà non sarebbe sufficientemente fedele, sarebbe anzi notevolmente *forzata*. Una rappresentazione migliore si potrebbe ottenere soltanto *allentando* il concetto stesso di *gruppo* (cfr. Ricolfi, 1992) ed accettando definizioni meno rigide.

Per ottenere tale risultato, devono essere messi in discussione i tre criteri prima elencati: se anche uno solo di essi salta, abbiamo nuovi risultati possibili e nuovi algoritmi per ottenerli.

2.4.1 SE CADE L'ESAUSTIVITA': GLI OUTLIERS

La prima delle violazioni possibili si ha quando cade il principio dell'eshaustività della classificazione. In tale condizione, non tutti i casi devono avere per forza una collocazione nei gruppi, per cui la somma dei relativi casi non necessariamente coincide col totale delle unità d'analisi. I casi non classificati vengono chiamati *outliers*, e la tipologia ottenuta è definita non *esaustiva*.

2.4.2 SE CADE LA MUTUA ESCLUSIVITA': L'OVERLAPPING CLUSTERING

La caduta della *mutua esclusività* genera gruppi che non sono completamente disgiunti fra loro, che hanno elementi in comune. Questo tipo di divisione di un insieme è definito *ricoprimento*, mentre la divisione che genera solo gruppi disgiunti è detta *partizione*. Le tecniche che danno questo risultato sono dette di *clump* (ricoprimento) o di *overlapping clustering*. Esse sono di grande interesse, in quanto costituiscono un primo tentativo di superare le difficoltà concettuali provocate dalla disgiunzione⁵, per cui avevamo visto che “la mutua esclusività [...] è una proprietà degli schemi classificatori, ma raramente è anche una caratteristica degli oggetti da classificare” (Sandri, 1969, p. 84).

Uno dei metodi utilizzati per l'individuazione dei *clumps* è l'algoritmo di Pichat. Il percorso che lo contraddistingue non è dei più semplici: si parte dalla matrice delle distanze fra i casi e si utilizza un valore soglia α per trasformarla in matrice booleana, ponendo uguali a 0 tutte le distanze maggiori della soglia, ed uguali a 1 quelle minori. Successivamente la matrice booleana viene scomposta in più sottomatrici, ognuna delle quali individuerà, alla fine, un *cluster*. Le

⁵ Vedi capitolo precedente.

sottomatrici sono composte solo da valori unitari, e ognuna di esse non è contenuta in nessuna delle altre. La matrice booleana originaria può essere rappresentata visivamente tramite un grafo, immaginando i casi come nodi che vengono uniti da segmenti solo quando la loro distanza è pari ad 1 (cfr. Miceli, 1995).

L'applicabilità di questa procedura è limitata dalla difficoltà di calcolo; un altro problema è che il valore di α deve essere deciso procedendo per tentativi: al suo aumentare, cresce la sovrapposizione fra i gruppi.

Ricolfi (cfr. 1989) ha sostenuto l'uso delle tipologie sovrapposte, soprattutto nelle analisi di tipo individuale, dove è frequente la situazione di *blocco compatto* prima vista, manca una vera struttura di *clusters* e la classificazione classica maggiormente evidenzia i suoi limiti, costringendo i dati ad una notevole forzatura. Il rovescio della medaglia di questa vicinanza al dato è in una rappresentazione più complessa, dove l'analisi rischia di compromettere la sintesi.

2.4.3 SE CADE L'APPARTENENZA BINARIA: IL FUZZY CLUSTERING

La violazione del principio di appartenenza binaria provoca le ripercussioni più pesanti, più feconde o controverse, a seconda del punto di vista assunto dallo studioso. Se, infatti, nei due casi appena visti eravamo ancora all'interno della classica logica aristotelica (o *booleana*, bivalente, binaria che dir si voglia), in questo caso il salto è maggiore, si fuoriesce da questa per entrare all'interno di una nuova cornice, quella della logica *fuzzy*.

Tratteremo questo argomento nel prossimo capitolo, nel quale, prima di entrare nello specifico degli algoritmi sfocati, necessariamente ci soffermeremo un momento sulla storia ed il significato della logica *fuzzy*.

2.4.4 I POSSIBILI RISULTATI DI UNA CLUSTER ANALYSIS

Abbiamo visto come la *cluster analysis* classica, benché nettamente la più utilizzata, non sia che un caso particolare all'interno di un panorama incredibilmente vasto.

Combinando le due alternative *partizione/ricoprimento*, e *appartenenza binaria/appartenenza fuzzy*, otteniamo un ventaglio non totale (mancano gli *outliers*) ma molto interessante di possibilità:

Tab. 2.1 – Possibili risultati di una cluster analysis

	Appartenenza classica $\mu : \{0,1\}$	Appartenenza fuzzy $\mu : [0,1]$
Disgiunzione $\Sigma \mu = 1$	(I) Partizioni classiche	(II) Partizioni sfocate
Sovrapposizione $\Sigma \mu > 1$	(III) Tipologie sovrapposte classiche	(IV) Tipologie sovrapposte sfocate

Il primo quadrante è il risultato più frequente in assoluto. Il secondo ed il terzo rappresentano le alternative più valide e più utilizzate. Il quarto quadrante, forse per eccesso di analisi, rappresenta per ora poco più che una possibilità teorica, non avendo stimolato discussioni, applicazioni o sviluppo di specifici algoritmi.

2.5 LA VALUTAZIONE DI UNA CLASSIFICAZIONE AUTOMATICA

Abbiamo appena visto come le molte tecniche di analisi dei gruppi possano produrre in generale una pluralità di soluzioni. Per scegliere fra le molte alternative, oltre all'irrinunciabile intervento soggettivo, guidato da considerazioni teoriche, pragmatiche ed euristiche, ci si può avvalere di diverse strategie.

Biorcio (1993, p. 88) individua tre piani su cui mettere a confronto le diverse possibilità:

- a. la funzionalità della soluzione rispetto ai fini della ricerca (semplicità e intelligibilità, capacità di descrivere efficacemente, di predire specifici fenomeni);
- b. la perdita di informazioni che si realizza rispetto ai dati di partenza;
- c. la stabilità della soluzione stessa.

Egli sottolinea anche come nella maggior parte dei casi pratici di ricerca, il secondo ed il terzo punto siano quasi sempre ignorati. Proprio su questi punti concentreremo l'attenzione.

2.5.1 L'EFFICACIA

Scopo di una classificazione è quello di effettuare una sintesi dei dati originari. I profili dei casi vengono riprodotti in parte dai profili dei tipi (cioè, i baricentri di ogni gruppo). Per giudicare la *bontà* di una classificazione bisogna valutare l'efficacia con cui essa riproduce l'informazione contenuta nella matrice di partenza. Si possono adottare tre diverse prospettive.

Ci si può chiedere quanto il profilo dei casi sia riprodotto nella classificazione, misurando la sua corrispondenza con il profilo dei tipi, per vedere quanto essi sono rappresentativi. Per fare ciò, ci si avvale dell'*analisi della varianza*, in particolare del test statistico dell'*eta quadro*. È la strategia più comune, che adotteremo nella nostra applicazione per valutare le tipologie. Trattiamo brevemente la questione: l'analisi della varianza, precisamente, esprime la relazione tra una variabile cardinale e una variabile categoriale. Secondo il teorema fondamentale di tale analisi, la varianza complessiva di una variabile cardinale può essere scomposta in due parti: la varianza interna alle categorie della variabile categoriale (cioè indipendente dall'associazione fra le due variabili) e la varianza esterna a tali categorie (dipendente dall'associazione fra le due variabili). Dunque ne consegue che:

$$\text{varianza totale} = \text{varianza interna} + \text{varianza esterna}$$

Se in una classificazione le variabili assunte come discriminanti sono, come spesso accade, variabili quasi-cardinali (come saranno, nel nostro caso, gli *items* di una scala Likert con punteggi assegnati ai vari gradi di accordo/disaccordo), allora le classi della tipologia possono essere interpretate come categorie di una variabile categoriale. Se la tipologia riuscisse a riprodurre completamente la varianza della variabile cardinale (l'inerzia della nuvola dei punti), allora la varianza interna ai gruppi scomparirebbe (massima omogeneità, cioè il profilo dei casi coinciderebbe col profilo dei tipi, cioè con il baricentro), mentre quella totale verrebbe a coincidere con la varianza esterna. Dunque tale classificazione riprodurrebbe alla perfezione il profilo dei casi.

L'indice *eta quadro* esprime la capacità di una variabile categoriale (come, ad esempio, una tipologia risultante da analisi dei gruppi) di riprodurre la varianza di una variabile cardinale: possiamo, quindi, sinteticamente affermare che tale indice ci dica, nel nostro caso, quanto una classificazione è *buona*. Esso è dato dal rapporto tra varianza esterna e varianza totale:

$$\eta^2 = \frac{\text{Varianza esterna}}{\text{Varianza totale}}$$

Il valore di *eta quadro* varia tra i casi limite di 0 (se la tipologia non riesce per nulla a riprodurre la varianza di una variabile cardinale) e 1 (se la tipologia la riproduce completamente). Esso può essere espresso anche in percentuale, per cui dire che il valore di *eta quadro* è pari a 0,4 significa affermare che la classificazione riproduce il 40% della varianza della variabile cardinale presa in considerazione, della nuvola dei punti e, per estensione del concetto, che una classificazione è *buona o efficace* al 40%.

La seconda prospettiva consiste nell'osservare quanto siano state riprodotte le relazioni fra i casi. Il procedimento è diverso a seconda che si operi con tecniche gerarchiche o non gerarchiche. In ogni caso si costruisce la matrice delle distanze riprodotta, da confrontare con la matrice delle distanze originaria.

L'ultima prospettiva cerca di giudicare il modo in cui vengono rappresentate le variabili. Il procedimento è simile al precedente, ma invece della matrice delle distanze riprodotta si costruisce la matrice delle correlazioni (fra variabili) riprodotta, e la si confronta con la matrice delle correlazioni originaria.

Essendo la *cluster analysis* una tecnica, si ha sempre un risultato. Sarebbe buona norma chiedersi se i gruppi individuati siano davvero presenti nella nuvola dei punti. È stato proposto un test, *Cubic Clustering Criterion* (CCC), utile a valutare se nella nuvola dei punti sia presente o meno una struttura di *cluster*. In realtà, basterebbe osservare con attenzione la nuvola dei punti – magari attraverso nuove soluzioni grafiche, campo questo largamente sottovalutato: “è opinione generale che nelle scienze sociali le tecniche di elaborazione matematico-statistica dei dati siano attualmente molto più sviluppate delle tecniche di raccolta delle informazioni o di presentazione grafica dei risultati” (Marradi, 2007, p. 22); si farà più avanti un tentativo in questa direzione –, tenendo sempre a mente come una struttura di *cluster* “naturali” è pressoché irrintracciabile nelle ricerche sociologiche a base individuale su un campione sufficientemente esteso.

2.5.2 LA STABILITA'

Al termine di una classificazione è importante determinare la stabilità della tipologia raggiunta. Una tipologia è *stabile* quando i profili dei tipi emersi da più tentativi siano sostanzialmente gli stessi.

La stabilità andrebbe controllata sia alla fine di un procedimento gerarchico che non gerarchico, perché in entrambe le situazioni i tipi finali dipendono dai passaggi iniziali, cioè il criterio d'aggregazione scelto nelle tecniche gerarchiche ed il principio che ha guidato il riallocaimento e la partizione iniziale in quelle non gerarchiche.

Per testare la stabilità di una classificazione si possono seguire due strategie. La prima è confrontare il profilo delle classi, calcolando la correlazione fra i tipi. Il profilo di una classe è rappresentato graficamente dal baricentro, che è

uguale alla media aritmetica delle variabili calcolata su tutti i casi che compongono il gruppo.

L'alternativa è il confronto fra la collocazione dei casi nei gruppi. Basta classificare più volte gli stessi casi ed effettuare un incrocio fra le diverse classificazioni tramite tabella di contingenza, vedendo così la percentuale dei casi collocati nello stesso tipo (ed il totale dei casi posti sulla diagonale principale).

2.6 CONCLUSIONI

Questa pur breve e parziale analisi delle procedure di classificazione automatica ci ha mostrato l'enormità dei mezzi e delle possibilità di questo settore di analisi dei dati. Per non perderci in questo *arcipelago*, bisogna usare in modo critico e consapevole le tecniche e le innumerevoli opzioni a disposizione. Si devono tener presenti pregi, difetti, predisposizioni di ogni strumento, cercando di mantenere il governo concettuale di ciò che si utilizza. Bisogna armarsi di pazienza se si vuole trovare un risultato sensato in un campo d'indagine spesso avaro di soddisfazioni, bisogna fare tentativi su tentativi e ricominciare spesso da capo. Se tutto questo non avviene, la pena da pagare non è tanto l'errore, quanto l'inutilità e, di conseguenza, la rapida caduta del dimenticatoio.

CAPITOLO 3. LOGICA FUZZY E FUZZY CLUSTERING

3.1 L'IRRUZIONE DELLA LOGICA FUZZY: "UNO SCHIAFFO ALLA SCIENZA MODERNA"

La logica *fuzzy* irrompe nella scienza nel 1965, quando Lofti Zadeh, allora direttore del Dipartimento di Ingegneria Elettrica dell'Università di Berkeley, pubblica l'articolo *Fuzzy Sets*. Nonostante Zadeh occupasse una delle più alte cariche in ambito accademico, il suo lavoro fu accolto con scetticismo, perplessità ed in certi casi aperta ostilità. I motivi di tali reazioni possono essere ricondotti a diversi livelli di profondità ed i più profondi di essi sono probabilmente riconducibili al fatto che la nuova logica, qualora si svolgano rigorosamente le sue implicazioni, tende, come vedremo, a mettere in discussione le stesse basi della *forma mentis* tipica del pensiero occidentale, forma che – secondo i seguaci della logica *fuzzy* – ricalca quella della divisione dicotomica:

gran parte della scienza, della matematica, della logica e della cultura muove dall'assunto di un mondo stabile di cose assolutamente bianche o nere: ogni asserzione è vera o falsa, ogni molecola del cosmo o appartiene al nostro dito o no, ogni legge, ogni norma, ogni regola di un club o si applica a una persona o no. (Kosko, 1993, p. 22).

Tale atteggiamento viene attribuito al successo di quello che viene definito come uno dei tanti possibili percorsi storici culturali del mondo occidentale, coincidente con l'affermazione del mondo culturale dell'antica Grecia, verificatosi più di 2000 anni fa:

La fede in questo presupposto dicotomico, questa *bivalenza*, in Occidente risale almeno agli antichi greci. Democrito riduce l'universo ad atomi e vuoto. Platone riempie il suo mondo di forme pure – del rosso, del giusto, della triangolarità e così via. Aristotele sottrasse un po' di tempo all'educazione del suo allievo Alessandro Magno per stendere quelle che ritenne fossero le leggi dicotomiche («o bianco o nero») della logica, leggi che scienziati e matematici impiegano tuttora per descrivere un universo «grigio» o in chiaroscuro e per discuterne. (Kosko, 1993, p. 22).

Procediamo, prima di tutto, con un chiarimento terminologico. *Fuzzy*, alla lettera, è traducibile in lanuginoso, coperto di peluria. Metaforicamente può essere reso con indistinto, sfumato, sfocato, ma nessuno di questi termini possiede la “valenza retorico-propagandistica (polemica, anticonformista, beffarda, innocentemente iconoclastica) che spinse Lofti Zadeh a scegliere questo aggettivo per ribattezzare e «rilanciare» un settore della logica che aveva dietro di sé una storia rispettabilissima, per nulla «anti-accademica» e tradizionalmente definiti dai sobri aggettivi di «vaga» o «polivalente»” (Kosko, 1993, p. 13 – N. d. T.). Stando a quanto riporta Kosko, (Kosko, 1993, p. 174) Zadeh scelse la parola *fuzzy* a causa dei suoi nessi col senso comune, gli pareva di aver trovato un termine che rendesse bene la nozione vaga di «un pezzo di realtà». Probabilmente non mancavano motivi di *marketing*, dato che né nei primi scritti di Zadeh né in quelli successivi vi erano riferimenti agli autori che avevano sviluppato il filone della logica polivalente. Zadeh cercava probabilmente “un nuovo inizio in un campo inesplorato, un mutamento di rotta e un tema nuovo rappresentato dalla matematica della teoria degli insiemi: meglio regnare all'inferno che servire in cielo”¹ (Kosko, 1993, p. 169). Zadeh scelse dunque la parola *fuzzy* come uno

¹ Kosko fa qui riferimento ad una diatriba intercorsa tra Zadeh e Rudolf Kalman, inventore dell'omonimo filtro. Kalman era uno studente della Columbia University quando Zadeh vi insegnava, tra i due c'era un'amicizia, anche se non troppo stretta, lavoravano negli stessi ambienti ed entrambi alla teoria degli insiemi: “l'opera di Zadeh sui «poli congelati» e sui «filtri modellanti» contribuì a gettare le fondamenta del filtro di Kalman. Ma fu Kalman, non Zadeh, ad avere la palma della vittoria” (Kosko, 1993, p. 167). Kalman fu poi un fiero oppositore della logica *fuzzy*, queste sono le sue parole pronunciate nel 1972 a Bordeaux, alla presenza di Zadeh, durante il convegno dedicato al tema «Man and Computer»: “non c'è dubbio che l'entusiasmo del professor Zadeh per la teoria *fuzzy* è stato rafforzato dal clima dal clima politico prevalente negli Stati Uniti – un clima di permissività senza precedenti. L'adozione della logica *fuzzy* è una sorta di

“schiaffo alla scienza moderna” (Kosko, 1993, p. 39), per suscitare l’ira della scienza che poi ottenne. Benché “il nuovo campo di studi fu costretto a crescere con tutti i problemi di «un ragazzo di nome Maria»” (Kosko, 1993, p. 39), il nome fece presa ed è per questo che in molti paesi, come l’Italia è stato deciso di mantenere il termine originale².

Ma cos’è la logica *fuzzy*? Ci sono due significati. Il primo è quello di logica *vaga*, un sistema logico la cui nascita è la naturale conseguenza degli studi sulla logica polivalente e sulla vaghezza. Il secondo è quello di *teoria degli insiemi fuzzy*, di ragionamento con insiemi *fuzzy*. La prima accezione è stata formalizzata ad inizio del secolo scorso, la seconda nasce con Zadeh.

Questi significati saranno chiarificati attraverso una sommaria ma necessaria analisi in chiave storica dei contributi culturali che hanno portato alla definizione di questa disciplina.

3.2 GLI ANTECEDENTI DELLA LOGICA FUZZY

Facciamo, dunque, un passo indietro ed analizziamo, seppur brevemente, gli antecedenti che hanno portato allo sviluppo della logica *fuzzy*. In estrema sintesi, schematizzando: Budda, Zenone, Russell, Heisenberg, Lukasiewicz, Black.

Budda visse in India due secoli prima di Aristotele. Narra la leggenda che a ventotto anni voltò le spalle alle ricchezze ed al potere del padre – che era un re locale dell’India nord orientale – per non tornarvi più. Da mendicante vagò per diffondere il suo messaggio di rinuncia³, non permise che i discepoli annotassero

permissività scientifica. Tende a risolversi in slogan di grande richiamo pubblico ma privi della disciplina propria del duro lavoro scientifico e dell’osservazione paziente” (Kosko, 1993, p. 167).

² A differenza, va detto, di quanto accade in altri paesi europei: ad esempio in Francia si usa *flou* ed in Spagna *borroso*.

³ Riassumibile nella formula delle “quattro verità”: a) la vita è dolore; b) il dolore deriva dal desiderio; c) eliminate il desiderio ed eliminerete il dolore; d) vivete una vita modesta, meditate e ciò vi aiuterà ad eliminare il desiderio. Come nota Kosko, ciò “è più una pillola intellettuale contro la sofferenza che una religione” (Kosko, 1993, p. 102).

le sue parole per iscritto (i suoi insegnamenti erano tutti orali), dipingessero, intagliassero ed in ogni modo mitizzassero o addirittura divinizzassero la sua immagine. Fu più simile ad un mendicante ateo o agnostico che diffondeva la sua personale filosofia di vita, nonostante ciò questa filosofia di vita divenne, nel corso dei secoli, la religione buddista. Buddha fu un anticipatore della dottrina *fuzzy* in senso specificatamente filosofico:

Budda non fu un teorico *fuzzy* in senso matematico, non scrisse alcuna pagina sugli insiemi o sui sistemi *fuzzy*, ma ebbe l'idea delle sfumature di grigio: ammise l'A E non-A. Accuratamente evitò la bivalenza artificiale che sorge nelle lingue naturali dal termine di negoziazione «non». Donde la sua frase famosa: «La non-mente non-pensa nessun-pensiero su nessuna-cosa». Budda sembra essere stato il primo grande pensatore a rigettare completamente il mondo dicotomico della bivalenza (Kosko, 1993, p. 102).

La cultura in lato senso buddista permea profondamente il mondo orientale: “il tema del «*fuzzy*», del «chiaroscuro» o «grigio» che dir si voglia, è riscontrabile nei sistemi dottrinari orientali vecchi e nuovi, dal taoismo di Lao Tze al moderno Zen del Giappone” (Kosko, 1993, p. 22).

L'Estremo Oriente in generale ed in Giappone in particolare hanno conosciuto un travolgente successo commerciale dei prodotti ad alta tecnologia a controllo *fuzzy*: il MITI (*Ministry for International Trade and Industry*) giapponese ha stimato che i prodotti *fuzzy*, il 70% dei quali appartiene all'elettronica di consumo (videocamere, aspirapolvere, televisori, ecc) hanno inciso per circa 1,5 miliardi di dollari nelle entrate del 1990 e per più di 2 miliardi nell'anno successivo.

Lo stato giapponese ha istituito due grandi laboratori, ciascuno dei quali promuove ad anni alterni un convegno sulla teoria *fuzzy*: il LIFE (Laboratory for International Fuzzy Engineering Research) a Yokohama, appena a sud di Tokyo, ed il FLSI (Fuzzy Logic System Institute) nel sud del Giappone, presso l'Istituto di Tecnologia di Kyushu. Membri del LIFE, fin dalla sua creazione, sono 48 tra le maggiori società giapponesi, tra cui, per fare qualche esempio: Canon, Hitachi, Honda, IBM Japan, Kawasaki, Mazda, Mitsubishi, Nissan, Sony, Toshiba,

Toyota. Secondo un memorandum dell'Ambasciata USA a Tokyo, inviato nel 1990 al Segretariato di Stato a Washington:

Il governo, le istituzioni industriali, commerciali e accademiche del Giappone stanno attivamente studiando la teoria della logica *fuzzy* nonché impiegando questa logica in una numerosa serie di applicazioni. [...] Alcune delle applicazioni a cui il LIFE sta lavorando sono un sistema di controllo per una centrale nucleare ed un prototipo di computer. [...] I ricercatori giapponesi impiegati nello studio dei sistemi *fuzzy* si aspettano che la logica *fuzzy* consentirà lo sviluppo di sistemi informatici che si adattino alla gente, al contrario di quanto accade ora (Kosko, 1993, p. 185).

Iniziative simili sono state prese in Corea del Sud, Singapore, Malesia, India, Taiwan, Hong Kong e Cina, dove – già nel 1989 – più di dieci mila studiosi e studenti *fuzzy*.

Kosko ritiene (a ragione, a nostro avviso) che “ci siano delle relazioni fra buddismo e ingegneria *fuzzy* nei paesi con economie ad alto sviluppo”. (Kosko, 1993, p. 96): se si prendono due mappamondi e si colorano i paesi che nella loro scienza, ingegneria e matematica studiano ed applicano la logica *fuzzy* nel primo, e si colorano i paese nei quali fiorisce il buddismo nel secondo, avremo due mappamondi colorati allo stesso modo. In questo senso, poco matematico e formale, ma molto sostanziale e profondo, si considera Budda come primo antesignano della filosofia *fuzzy*.

Zenone, in rappresentanza in un certo senso di tutti i sofisti, anticipa la logica *fuzzy* grazie ai paradossi – della forma del sorite –, con i quali già nella Grecia classica si mettevano in risalto le contraddizioni della logica bivalente. Togliendo un granello alla volta, si passava dal mucchio di sabbia al non-mucchio senza poter indicare quale fosse stato il granello che trasformava A in non-A (cfr. Kosko, 1993, p 23, 119-122). Il paradosso del sorite è presentabile in innumerevoli forme: se qualcuno tocca l'alluce della madre è incesto? Se tocca la caviglia, gli stinchi, il ginocchio? Russell lo presentava sotto la forma di un uomo non calvo che diventava calvo, subendo lo strappo di un capello uno alla volta. Quando l'uomo smetteva di essere non-calvo per diventare calvo?

Benché i paradossi ponessero questioni di difficile soluzione, visto che insinuavano dubbi al livello delle fondamenta del pensiero dicotomico, essi furono spesso liquidati come giochi di parole, non degni di attenzione in quanto i matematici non vedevano in essi una minaccia per la loro materia. Fino a quando Russell, all'inizio del ventesimo secolo, non propose uno di questi paradossi – un paradosso di autoriferimento, che si autoafferma e si autonega contemporaneamente, simile come forma a quello del mentitore di Creta (cfr. Kosko, 1993, p. 23, 32, 334) – dritto nel cuore della matematica, nello specifico della teoria degli insiemi.

Il paradosso del mentitore di Creta (che afferma: «tutti i Cretesi mentono») porta ad una doppia contraddizione: se sta mentendo, allora sta dicendo la verità (contraddicendo coi fatti la frase); se non sta mentendo, allora la frase è vera, per cui mente. Russell trovò nella teoria degli insiemi un paradosso della stessa forma che “pose termine alle certezze della matematica” (cfr. Kosko, 1993, p. 123). Gli insiemi sono contenitori, scatole di oggetti materiali ma anche immateriali. Secondo la logica bivalente, un elemento o appartiene ad un insieme oppure no, l'appartenenza non è questione di misura. Russell trovò, però un insieme che sembrava contenere e non contenere i suoi elementi: *l'insieme di tutti gli insiemi che non sono elementi di se stessi* (cfr. Kosko, 1993, p. 122-128). Gli insiemi che non appartengono a se stessi sono la maggior parte degli insiemi che possiamo osservare coi nostri occhi: ad esempio, l'insieme delle mele non appartiene a se stesso, in quanto i suoi elementi sono mele, non insiemi. Allo stesso modo, l'insieme delle persone, delle automobili, delle stelle o di innumerevoli altre cose, sono tutti insiemi che non appartengono a se stessi. Ora la domanda è: l'insieme di tutti gli insiemi che non sono elementi di se stessi appartiene a se stesso oppure no? Se appartiene, allora contraddice la regola (non appartenere a se stesso) e dunque non appartiene; se non appartiene, allora la soddisfa ed appartiene.

Questo paradosso insinuatosi nel cuore della teoria degli insiemi segnò – indipendentemente dai vari tentativi di soluzione o semplicemente minimizzazione che lo seguirono – una vera svolta epocale nella comunità scientifica:

I paradossi di Russell posero termine a migliaia d'anni di fede cieca nella certezza della matematica, della matematica bivalente, tanto che alcuni matematici, per descriverne l'effetto, parlano di «Paradiso perduto» (cfr. Kosko, 1993, p. 118).

Russell cercò una soluzione che rimanesse nell'ambito della bivalenza e la offrì con la «teoria dei tipi», secondo la quale si può parlare solo di un livello alla volta in una scala gerarchica o matematica:

Si può dire, ad esempio, «le mele sono rosse», poiché la rossezza è una proprietà delle mele e di altri oggetti; essa li precede immediatamente nella gerarchia logica. Per la medesima ragione si può anche dire «il rosso è un colore»; ma non può dire «le mele sono un colore» poiché così si salta un livello (Kosko, 1993, p. 126).

La coerenza di questa posizione è evidente, ma in un certo senso essa “costava troppa matematica” (Kosko, 1993, p. 126) e troppa logica, si pensi ad espressioni del linguaggio comune come il “color albicocca”. Ad ogni modo, il dubbio era stato insinuato al livello – ritenuto auto-evidente – della matematica:

Agli inizi degli anni Venti Russell gettò i fondamenti logici della logica *fuzzy* (vaga) senza peraltro dedicarsi mai a svilupparla. Aveva tuttavia scoperto gli altarini grigi della logica dicotomica (Kosko, 1993, p. 118).

La soluzione che offre la logica *fuzzy* ai paradossi, una volta che si decide di lasciar cadere il vicolo del terzo escluso è la seguente: i paradossi, che non sono eccezioni ma regole, sono letteralmente *mezze verità*. Non tutte le contraddizioni sono uguali, ci sono sfumature anche in esse ed i paradossi di autoriferimento si pongono esattamente a metà strada tra il vero ed il falso.

Questo è provabile anche matematicamente con una semplice dimostrazione. I paradossi di autoriferimento hanno tutti la stessa formula, per cui: $A \text{ implica } non-A \text{ e } non-A \text{ implica } A$ (ad esempio, il mentitore che “mente” implica che il mentitore “non mente”, e viceversa; l'insieme che “appartiene” all'insieme degli insiemi che non contengono se stessi implica che l'insieme “non appartiene”, e viceversa). Dunque, A e $non-A$ sono logicamente equivalenti: $A = non-A$; hanno anche gli stessi valori di verità: $t(A) = t(non-A)$. A questo punto, ci

troveremmo di fronte ad una contraddizione del tipo $0 = 1$ o $1 = 0$, ma se decidiamo di non proseguire con la logica binaria, sappiamo che il valore di verità di non-A è uguale a 1 meno il valore di verità di A, dunque: $t(\text{non-A}) = 1 - t(A)$. Facendo una sostituzione nella prima parte dell'equazione: $t(A) = 1 - t(A)$. Con una semplice operazione algebrica, abbiamo: $2t(A) = 1$; di conseguenza: $t(A) = \frac{1}{2}$. La verità, in un paradosso di autoriferimento, sta dunque nel mezzo. Il mentitore di Creta, nello stesso momento, mente e non mente nella stessa misura; l'insieme degli insiemi che non sono elementi di se stessi appartiene e non appartiene a se stesso nella stessa misura (Kosko, 1993, p. 126-128).

I paradossi, dunque, seguendo un'impostazione *fuzzy*, ci mostrano prima di tutto che il loro nome è improprio. Paradosso suggerisce l'idea di eccezione, mentre l'eccezione sembra essere nella dicotomia, nei rari casi di bianco e nero che si pongono alle estremità di un *continuum* di toni di grigio.

La seconda lezione dei paradossi è che ci mostrano i costi della bivalenza e ci suggeriscono come *al crescere dell'informazione, aumenti anche l'incertezza*. Il paradosso del sorite, tipico di Zenone, è formato da una catena logica di enunciati dalla forma «se A, allora B; se B, allora C; se C... allora Z». Se ipotizziamo una perfetta catena bivalente, moltiplicando enunciati che siano tutti perfettamente certi, abbiamo un prodotto logico che moltiplica ed ottiene certezze:

$$1 \times 1 \times 1 \times \dots \times 1 = 1.$$

Se invece i nostri enunciati sono incerti (come sono quelli che riguardano il mondo reale), otteniamo una catena di incertezza che aumenta indefinitamente:

$$1 \times 0,99 \times 0,98 \times 0,97 \times \dots \times 0,0001 = n \rightarrow 0.$$

Il prodotto della logica *fuzzy* «tende a zero» se i fattori di incertezza aumentano all'infinito. Il prodotto di un mucchio di incertezza dà incertezza composta: dunque, aumentando l'informazione, aumenta l'incertezza o, secondo l'espressione di Kosko, viste da vicino, le cose sono *fuzzy*:

paghiamo in punti di certezza ogni inferenza che facciamo. Il ragionamento non è privo di costi. Più passi facciamo nel nostro ragionamento, più questo diventa *fuzzy* (Kosko, 1993, p. 121).

Questa conclusione d'altronde è in linea con una visione moderna della scienza, per cui assunto fondamentale di essa è l'inaccuratezza e la riduzione di questa ultima ne è la meta. Ben venga, in questo senso, l'arrotondamento ma facendo attenzione a non eccedere:

L'arrotondamento condensa le informazioni, semplifica le questioni e riduce i molti a pochi, il complesso al maneggevole. Noi arrotondiamo per cavarcela e per avere una rapida presa sulle idee e sulle porzioni della nostra mutevole visione del mondo. Dobbiamo semplificare per far andare le cose, almeno all'inizio. In questo non c'è niente di male.

Ma le nostre azioni hanno dei costi; la bivalenza o l'arrotondamento barattano l'accuratezza con la semplicità. Quando si arrotonda, si paga in verità accuratezza e sincerità quello che si guadagna in semplicità, precisione e conformità. Il non ammettere i costi non li elimina. Un piccolo arrotondamento, come un piccolo debito, non ha mai fatto male a nessuno. Ma anche un piccolo arrotondamento, come una piccola iniziale gravidanza, può riservare delle sorprese. Se si arrotonda troppo si paga il fio dell'autocontraddizione bivalente e si cade nel paradosso (Kosko, 1993, p. 116).

Solo qualche anno dopo che Russell aveva messo in allarme il mondo della matematica con il suo paradosso degli insiemi, Werner Heisenberg pose una seria dose di incertezza all'interno della fisica con il suo principio di indeterminazione della meccanica quantistica. Esso afferma che nel momento in cui si sta misurando la velocità di una molecola, non se ne può calcolare la posizione, e viceversa. Velocità e posizione sono due «variabili coniugate», per cui o l'una o l'altra. Questo vale anche per altre coppie di variabili coniugate, come l'energia e il tempo o le forze del campo magnetico e quelle del campo elettrico. Nel momento in cui una si misura, l'altra varia. Se Russell aveva mostrato che la logica delle nostre menti è incerta, “ora Heisenberg mostrava che gli atomi dei nostri cervelli sono incerti” (Kosko, 1993, p. 128). Heisenberg dimostrò che nella meccanica quantistica alcune cose non possono mai essere conosciute, sono anzi “inconoscibili in linea di principio” (Kosko, 1993, p. 129). Egli rese scientifico il dubbio e di nuovo indicò come, in accordo col principio *fuzzy*, anche disponendo dell'informazione totale, non si poteva comunque asserire alcunché col 100% di certezza.

La strada che portava alla logica *fuzzy* era ormai spianata e non restava che passare per i lavori di Lukasiewicz e Black, che la elaborano sostanzialmente nello stesso modo in cui Zadeh la presentò alla comunità scientifica nel 1965.

Jan Lukasiewicz, logico polacco, fu il primo, già negli anni Venti, a tagliare la via di mezzo rappresentata dall'«indeterminato», dapprima ammettendo posizioni intermedie – 0,5 – tra la verità e la falsità e ideando così la logica trivalente e successivamente inserendo altri valori intermedi fino ad arrivare alla supposizione che l'indeterminazione definisse un *continuum* tra lo 0 e l'1, approdando così ad una logica che chiamò appunto a più valori o polivalente (cfr. Kosko, 1993, p. 39, 174-175, 332-333).

Nel 1937 il filosofo quantistico Max Black pubblicò un saggio intitolato *Vagueness: an exercise in logical analysis* (cfr. Black, 1937). Nei suoi studi sull'imprecisione del linguaggio, egli sostenne l'esistenza di alcuni individui per i quali sembrava impossibile sia applicare che non applicare un termine. Estese in questo modo la logica polivalente agli insiemi, delineando quella che oggi chiamiamo curva di appartenenza. Nella sostanza, l'opera di Black anticipò integralmente quello che avrebbe presentato Zadeh 30 anni dopo. Ma il suo lavoro cadde nell'oblio:

Max Black morì nell'agosto del 1989, l'anno in cui la logica *fuzzy* o vaga che egli aveva studiato faceva la sua prima comparsa alla televisione giapponese in comunicati pubblicitari di lavatrici «intelligenti». Egli morì in silenzio, come era morta la sua teoria degli insiemi vaghi. Il mondo della filosofia non riconobbe alcun interesse nella sua opera sulla vaghezza. Né gliene riconobbe il mondo della ricerca *fuzzy*. [...] Ci voleva un uomo dotato di abilità politica non meno che scientifica per far accettare la vaghezza alla scienza. (Kosko, 1993, p. 165)

Quell'uomo fu Lofti Zadeh, che scelse il nome *fuzzy* per i motivi di cui abbiamo parlato in precedenza. Zadeh era nato a Baku, sul Mar Caspio, nel 1921, proprio dove era nato Black, nel 1909: “questa era la seconda cosa che i due studiosi ebbero in comune” (Kosko, 1993, p. 166).

3.3 FUZZY SETS

Fatte le dovute premesse, possiamo ora ritornare ai due significati che ha l'espressione logica *fuzzy*: logica vaga e teoria degli insiemi *fuzzy*. Analizzando le origini della logica *fuzzy* come derivazione degli studi di logica polivalente e vaga, abbiamo implicitamente chiarito il primo significato, per il quale la logica *fuzzy*, secondo una delle molte belle immagini fornite da Kosko, "è come la differenza fra il giorno e la notte all'imbrunire" (Kosko, 1993, p. 105). Non ci resta che esplicitare il secondo, secondo la formalizzazione che ne ha dato Loft Zadeh, della cui famosa opera riportiamo l'*incipit*:

Un insieme *fuzzy* è una classe di oggetti con un continuum di gradi di appartenenza. Tale insieme è caratterizzato da una funzione di appartenenza (caratteristica) che assegna ad ogni oggetto un grado di appartenenza compreso tra zero e uno (Zadeh, 1965, p.338).

Un insieme classico, invece, prevede soltanto la totale o nulla appartenenza dei suoi elementi. Volendo formalizzare⁴, abbiamo queste due situazioni, intendendo per $f_I(x)$ la funzione di appartenenza di un qualsiasi elemento x ad un qualsiasi insieme I :

$$\begin{aligned} f_I(x) = \{0,1\} &\rightarrow \text{ nella teoria degli insiemi classici} \\ f_I(x) = [0,1] &\rightarrow \text{ nella teoria degli insiemi fuzzy} \end{aligned}$$

La funzione di appartenenza può essere interpretata anche come grado di verità dell'affermazione " x appartiene ad I ". Dunque nella teoria degli insiemi classici sono possibili soltanto due affermazioni:

$$\begin{aligned} x \text{ appartiene ad } I &\rightarrow \text{ valore di verità} = 1 \\ x \text{ non appartiene ad } I &\rightarrow \text{ valore di verità} = 0 \end{aligned}$$

⁴ In linguaggio logico, le parentesi graffe significano che sono validi solo i valori indicati in parentesi, le quadre intendono che valgono tutti i valori compresi, estremi inclusi.

La teoria degli insiemi *fuzzy* invece ha valori di verità che possono assumere tutti gli infiniti valori tra 0 e 1, estremi inclusi. È in questo senso che si interpreta l'affermazione per cui la logica *fuzzy* è una generalizzazione di quella classica, nello stesso senso in cui si può dire che il bianco ed il nero sono due casi speciali di toni grigio.

Per cogliere la diversità fra gli insiemi classici e gli insiemi sfocati, proponiamo due esempi tipici: l'insieme dei *numeri pari* e l'insieme dei *giovani*. Per discriminare il primo, la logica classica funziona: 2 è pari al 100%, 3 è pari allo 0%. Ci sarebbe il caso ambiguo dello 0, è un numero pari? È un numero? Qualcuno – come osservò Alfred North Whitehead – va forse al mercato per comprare zero pesci? Restiamo ad un livello più semplice ed ignoriamo questo problema, constatando dunque come la logica bivalente sia abile a dividere i numeri pari dai dispari.

Ma per discriminare un giovane da un non giovane? Si potrebbe dire tranquillamente che un neonato sia giovane al 100%, ed un centenario lo sia allo 0%, senza rischio di essere accusati di imprecisione o di forzare i dati. Ma un 28enne? Un 35enne? Con la logica classica avremmo dei problemi, che quella *fuzzy* supera facilmente: se pensassimo ad un neonato come alla persona più giovane pensabile e ad un centenario come la più anziana (ovviamente non è così, si tratta solo di un esempio; si potrebbe prendere come punto di riferimento l'età della persona più anziana del mondo di cui si ha conoscenza certa oppure l'età dell'aspettativa di vita in un determinato contesto spaziale e temporale) e ponessimo questi due limiti pari rispettivamente al 100% ed allo 0% dei valori di "gioventù", sarebbe facile calcolare il grado di appartenenza dei nostri soggetti all'insieme dei giovani (sottraendo a 100 la loro effettiva età): il 28enne sarebbe, dunque, giovane al 72% (grado di appartenenza = 0,72), il 35enne al 65% (grado di appartenenza = 0,65), e così via.

Gli insiemi *fuzzy* sono ovunque davanti ai nostri occhi: "numeri piccoli, aria calda, uomini alti: la maggior parte delle nostre parole significa insiemi *fuzzy*" (Kosko, 1993, p. 145): il nome casa sta per l'insieme *fuzzy* di case, a cui appartengono in diversa misura "castelli, roulotte, case trasportabili, villette bifamiliari, condomini in multiproprietà, tende pellerossa o mongole, tettoie,

caverne, padiglioni, scatoloni di cartone nei vicoli” (Kosko, 1993, p. 145); e quello che abbiamo appena visto, giova ribadirlo, non è soltanto un modo di giocare coi numeri, è un modo di risolvere i paradossi sopra osservati, di trovare soluzioni rapide ed efficaci a problemi spinosi, di prendere decisioni semplici in contesti difficili. Il fulcro della logica *fuzzy* è esattamente questo, diversamente dai sistemi logici classici, come abbiamo potuto vedere, essa mira a modellare un modo impreciso di ragionare, e quest’ultimo gioca un ruolo essenziale nell’abilità umana di *prendere decisioni razionali in un ambiente di incertezze e imprecisioni*. La logica *fuzzy* si avvicina al ragionamento umano e si adatta alle situazioni comuni:

i nostri cervelli sono pieni di insiemi *fuzzy*. Noi pensiamo in termini di insiemi *fuzzy* e ciascuno di noi definisce a modo suo i confini *fuzzy* in maniera differente e con esempi differenti. Noi ammucciamo il mobilio dell’universo in insiemi *fuzzy*. Raggruppiamo le cose in insiemi flessibili che sono *fuzzy* e poi giochiamo con questi mucchi e ne cerchiamo le connessioni: *il pensiero non è che un gioco con insiemi fuzzy*. E questo è appunto quel che è la logica *fuzzy*: ragionamento con insiemi *fuzzy* (Kosko, 1993, p. 149; corsivo originale).

Prima di passare all’analisi dei principali algoritmi *fuzzy* presenti nella letteratura sulla classificazione automatica, vediamo brevemente le principali critiche rivolte alla logica *fuzzy*.

3.4 LE CRITICHE ALLA LOGICA FUZZY

Kosko (cfr. Kosko, 1993, p. 175-182) individua tre critiche principali: la prima richiedeva applicazioni concrete che dimostrassero l’utilità del nuovo approccio; la seconda vedeva nella logica *fuzzy* una semplice imitazione della probabilità; la terza era una semplice, a volte irrazionale, difesa della bivalenza.

Le applicazioni di nuove idee richiedono sempre un po’ di tempo. Zadeh, in particolare, non ne fece mai, era un teorico matematico e tale rimase sempre. Negli anni Settanta cominciarono ad apparire le prime applicazioni sotto forma di simulazione al computer di idee matematiche. Apripista per quanto riguarda i

sistemi fuzzy di controllo fu l'opera di Ebrahim Mamdani (cfr. Mamdani, 1977). Nel decennio successivo in Giappone fu grande lo sforzo della ricerca di questi sistemi di controllo. Nel 1990 c'erano già più di cento applicazioni realmente operanti in prodotti commerciali.

La seconda critica veniva dalla scuola probabilistica: "Zadeh usava numeri fra 0 e 1 per indicare la vaghezza o le misure. Ai teorici della probabilità pareva di fare la stessa cosa, sicché il conflitto era inevitabile" (Kosko, 1993, p. 175). C'è da dire che Zadeh non si sforzasse neanche troppo di rispondere alle critiche in generale, quelle probabilistiche in particolare («tutta fatica sprecata»), piuttosto "si concentrava sul potere espressivo degli insiemi *fuzzy* e sulla loro adeguatezza alle parole" (Kosko, 1993, p. 176).

Ad ogni modo, l'argomentazione in risposta alle critiche probabilistiche è che la probabilità si fonda comunque sulla logica bivalente. La probabilità è un numero assegnato ad un evento, più grande è il numero, più è probabile (verosimile) che l'evento abbia luogo. Ma in termini matematici, tutti i valori di probabilità devono assommare a 1, la probabilità che accada un evento e che non accada è pari al 100%. Se si tira una moneta in aria, c'è una probabilità su due (50% o 0,5) che esca testa e una probabilità su due (di nuovo, 50% o 0,5) che esca croce. La probabilità che esca testa o croce è dunque una su una (100%). C'è di più, vediamo un esempio:

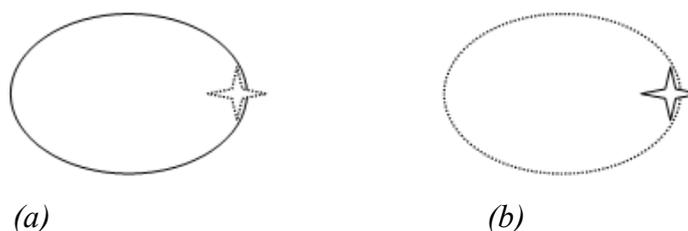
se nascondo un pedone bianco in una mano dietro la schiena e chiedo di indovinare in che mano lo tengo, io so in che modo va l'esperimento ma gli altri devono indovinare, fare supposizioni, calcolare probabilità. [...] La probabilità svanisce con l'aumento dell'informazione. Più informazione, meno probabilità (Kosko, 1993, p. 20).

Con la logica *fuzzy*, come abbiamo avuto modo di vedere, succede esattamente il contrario, essa aumenta di pari passo con l'aumento della precisione, dell'informazione. In un certo senso infatti, *fuzziness* e probabilità sono una l'inverso dell'altra, usano entrambe i numeri, ma per significare cose diverse:

mentre il concetto fuzzy di grado di appartenenza esprime in termini numerici il livello di somiglianza di un oggetto a una proprietà non definita con esattezza, la probabilità di appartenenza esprime numericamente l'aspettativa che un certo evento si verifichi, ossia l'appartenenza di un certo oggetto all'insieme degli eventi favorevoli (Brunelli, 2001, p. 38).

Dunque, nel caso della probabilità, sono incerte le informazioni sull'oggetto ma sono certi i confini del gruppo (l'insieme degli eventi favorevoli). Nel caso della logica *fuzzy*, le informazioni sull'oggetto sono accurate, ma sono imprecisi i confini del gruppo. Possiamo vedere questa differenza rappresentata graficamente:

Fig. 3.1 – Differenza tra probabilità di appartenenza (a) e grado di appartenenza (b)



L'ultima critica “era la più rilevante. Era la pura collera della bivalenza” (Kosko, 1993, p. 177-178). Oggi la logica *fuzzy* non desta particolare scalpore e, forse per i successi sul mercato, è diventata quasi di moda. Appena trent'anni fa, invece, rappresentava per molti scienziati una vera e propria provocazione:

secondo Zadeh la correttezza o anche la plausibilità [della matematica polivalente] implicavano che Aristotele si fosse sbagliato. Ciò significava che le cose non dovevano essere necessariamente o bianche o nere e che le verità bivalenti erano soltanto riflessi condizionati. Tremila anni di storia occidentale erano solo uno dei percorsi possibili attraverso lo spazio-tempo. poteva essere tutto sbagliato o, peggio, un limitato capriccio culturale: la scienza come precisione, la matematica e la logica come le linee di demarcazione fra «bianco e nero». Era troppo perché una sola conferenza o un solo articolo fossero sufficienti a farlo accettare. Zadeh sembrava dire che o aveva ragione lui o aveva ragione la gran parte della scienza o della cultura occidentale. Delle due cose o l'una o l'altra: e la scelta sembrava ovvia (Kosko, 1993, p. 178).

Le critiche erano di due tipi. Il primo tipo sottolineava come la bivalenza funzionasse, facesse lavorare i computer e ci avesse ben servito per un paio di millenni: può comportare vantaggi, ma è semplice e funziona. Si ammetteva l'artificialità e la forzatura di alcune sue applicazioni, ma si concludeva che una simile convenzione era innocente, visto che tutta la scienza era convenzionale (cfr. Quine, 1981).

Il secondo tipo era semplicemente “un grido di sdegno” (Kosko, 1993, p. 179) e, nonostante la mancanza di contenuti logici e fattuali, era vincente:

ci piace pensare che cose del genere non abbiano cittadinanza in una comunità civile e specialmente nella scienza; e in effetti è così, anche se la società e la scienza ne sono piene (Kosko, 1993, p. 181).

Per fare un solo esempio, ma significativo, citiamo il pezzo riportato da Kosko in cui un giornalista, nel 1975, riassume la posizione di Zadeh fra i suoi colleghi dell'Università di Berkeley in California:

«La teoria fuzzy è errata e perniciosa», dichiara William Kahan, un professore di informatica e matematico presso la Cal [l'Università di Berkeley], il cui studio presso la Evans Hall è a poche porte da quello di Zadeh. «Io non riesco a pensare a nessun problema che non possa essere risolto meglio con la logica ordinaria. ... Quello che va dicendo Zadeh appartiene allo stesso genere di affermazioni del tipo “La tecnologia ci ha cacciato in un grosso pasticcio e ora non è capace di tirarcene fuori”. Orbene, non è vero che è la tecnologia ad averci messo nei pasticci. L'avidità, la fiacchezza e l'ambiguità ci hanno messo nei pasticci. Abbiamo bisogno di più pensiero logico, non di meno. Il pericolo della logica *fuzzy* è che incoraggio quel genere di pensiero impreciso che ci ha creato tanti problemi».

Nel corso degli anni il professor Kahan non ha cambiato la sua posizione. Nel 1990, un giornalista chiese a Kosko di commentare la recente affermazione di Kahan secondo cui la logica *fuzzy* è una licenza per il pensiero in libertà ed è la “cocaina della scienza” (Kosko, 1993, p. 182). La risposta di Kosko fu: grazie.

3.5 FUZZY CLUSTERING

Ora, torniamo a noi, o meglio, torniamo alla *cluster analysis*. Vogliamo porre di nuovo attenzione alla parte più moderata della critica bivalente. Tale posizione, largamente condivisibile e certamente ragionevole, suggerisce la consapevole scelta di un paradigma come soluzione dirimente dei dubbi. Tali paradigmi si caratterizzano per le seguenti dicotomie:

Logica classica Vs Logica fuzzy

Sintesi Vs Analisi

Semplicità Vs Precisione

Se, dunque, questa posizione è condivisibile e lo è, è lecito domandarsi perché dovremmo provare a cambiare prospettiva anche nell'analisi dei gruppi? A nostro avviso, il tentativo è auspicabile per due ragioni: primo, il tipo di proprietà con cui lavora la nostra disciplina è più simile alla forma di un *fuzzy set* che a quella degli insiemi classici; secondo, i nostri campioni si distribuiscono in modo granulare sugli assi fattoriali, costringendo l'analisi classica ad evidenti forzature. Vediamo ora gli algoritmi *fuzzy* più usati.

Le tecniche di classificazione sfocata si distinguono in due tipi: quelle che ottimizzano una funzione obiettivo e quelle basate su una funzione caratteristica (Aureli e Middei, 1997).

La nostra attenzione si concentrerà sulle prime, che hanno maggiori risvolti pratici. Esse funzionano in modo analogo ai procedimenti non gerarchici visti in precedenza, il numero dei gruppi deve essere deciso a priori e c'è un principio (funzione obiettivo) in base al quale si classifica. Questi metodi hanno il difetto di convergere spesso ad un ottimo locale anziché globale⁵. Partono dalla matrice dati e sono in grande difficoltà quando essa è incompleta.

Le tecniche con funzione caratteristica sono assimilabili alle tecniche gerarchiche, la scelta del gruppo può essere fatta a posteriori, partono dalla matrice della similarità, per cui possono operare anche con dati mancanti. Danno

⁵ Vedi paragrafo 2.3.2.

però luogo ad una mole proibitiva di calcoli, che le rende utilizzabili solo per campioni poco numerosi (circa 100 casi). Perdono così gran parte della loro utilità.

Fra le tecniche sfocate con funzione obiettivo, una della più interessanti è la *Fuzzy C-Means* (FCM). Introdotto da Dunn per la prima volta nel 1973, la sua versione definitiva si deve a Bezdek (1981), questo metodo può essere considerato come la generalizzazione dell'algoritmo delle *k-medie* classico. La sua funzione obiettivo (da minimizzare) è la seguente:

$$\sum_{k=1}^c \sum_{i=1}^n (\mu_{ik})^m (d_{ik})^2$$

Dove c è il numero dei gruppi scelto, n è il numero dei casi del campione, μ_{ik} è il grado d'appartenenza di un generico i -esimo caso ad un altrettanto generico k -esimo gruppo, m è un parametro da fissare, e d_{ik} è la distanza dell' i -esimo caso dal baricentro del k -esimo gruppo.

Il parametro m è di fondamentale importanza, in quanto incide sulla sfocatura finale. Tale valore può essere determinato solo procedendo per tentativi, non esistono indicazioni a riguardo buone per ogni occasione ed è, a nostro avviso, proprio in questa "scelta che assai spesso deve essere fatta in maniera arbitraria e soggettiva" (Aureli e Middei, 1997, p. 32) il maggiore limite di questa tecnica. Aureli e Middei propongono, come soluzione, l'uso di un indice di sfocatura, l'indice I di Gini, da calcolare alla fine del processo classificatorio. La formula per calcolare l'indice di Gini è la seguente:

$$I = c / c-1 \left[1 - 1/n \sum_{i=1}^n \sum_{k=1}^c \mu_{ik}^2 \right]$$

dove c è il numero dei *clusters* della tipologia (deciso a priori), n è il numero dei casi che compongono il campione, e μ_{ik} è il grado d'appartenenza di un generico i -esimo caso ad un altrettanto generico k -esimo gruppo. L'indice varia tra 0 ed 1: nel caso di una tipologia classica è pari 0, nel caso di massima sfocatura è uguale ad 1.

Se, dunque, tale indice risulta troppo elevato, per i fini conoscitivi ed a parere del ricercatore, si può ripetere la procedura con un m più basso e viceversa. I due autori affermano che una scelta ragionevole per m sia quella che porta a un indice di Gini compreso nell'intervallo 0,4-0,5 pur precisando che il livello di sfocatura dipende anche dal numero di unità del collettivo, dalla sua particolare struttura e dal numero di *cluster* scelti per la classificazione. Uno stesso valore di m dunque potrebbe portare a risultati diversi in situazioni diverse ed abbiamo il sospetto che tale fascia di valori di Gini pari a 0,4-0,5 sia forse più adatta a studi ecologici che individuali, la cui confusione intrinseca – legata, tanto per fare un esempio, alla elevata numerosità dei campioni – porterebbe, con quei livelli di sfocatura, a risultati troppo dispersivi. Ne consegue, ad ogni modo, che l'unica cosa da fare sia armarsi di pazienza e provare finché non si raggiunge il risultato ritenuto più idoneo.

Il parametro m varia tra 1 e infinito, se uguale ad 1 genera una partizione tradizionale, la stessa che si otterrebbe con il classico metodo delle *k-medie*. Al crescere di m , cresce anche la sfocatura, fino alla situazione limite di equidistribuzione dei casi nei *clusters*.

Il ricercatore può anche stabilire un parametro di convergenza δ , per fermare il processo iterativo nel caso non fosse possibile raggiungere l'ottimo. Il procedimento si arresta quando la differenza tra ultima e penultima iterazione raggiunge tale soglia.

Dopo la scelta di m , δ , c e della partizione iniziale, vi è il calcolo dei baricentri e da qui inizia il procedimento di suddivisione e l'individuazione della classificazione finale.

Nelle sue applicazioni pratiche (cfr. ad esempio, Iacovacci, 1995), l'FCM si è dimostrato *robusto*, riuscendo ad individuare la stessa classificazione finale anche partendo da partizioni iniziali relativamente differenti. Altro pregio dell'FCM è che raggiunge spesso il minimo globale, senza incastrarsi in equilibri locali.

Il vantaggio delle classificazioni sfocate è che conservano più informazioni sul singolo caso, indagano le sfumature e le incertezze. Esse possono poi essere facilmente trasformate in tipologie classiche, ponendo $m = 1$, o

trasformando in totale il grado di appartenenza maggiore. Permettono inoltre di individuare, all'interno dei *fuzzy sets*, dei nuclei forti e molto rappresentativi, caratterizzati da grado d'appartenenza prossimo all'unità.

I pregi degli algoritmi sfocati, così come quelli degli algoritmi tradizionali, possono però trasformarsi in difetti: nell'approccio classico, la sintesi può diventare forzata, nell'approccio *fuzzy*, l'analisi può diventare ingestibile. Il rischio, infatti, è che la sfocatura possa essere introdotta anche laddove non era necessaria; questo è dovuto al fatto che in tali classificazioni un caso è assegnato interamente ad un gruppo soltanto se esso coincide con il suo baricentro: una situazione rarissima. In una nostra applicazione precedente, utilizzando un parametro $m=1,1$ – appena superiore al valore di 1 che avrebbe restituito una classificazione identica a quella tradizionale –, ben il 40% del campione (oltre 1400 individui) sarebbe stato non assegnato a nessuno dei 4 gruppi individuati; tra essi, casi come i seguenti, nei quali, si può vedere, la sfocatura è perfettamente inutile:

Tab. 3.1 – Esempio di sfocatura eccessiva

Tolleranti	Tolleranti con riserve	Conflittuali Moderati	Conflittuali
0,999	0,001	0	0
0,999	0	0	0,001
0,999	0	0,001	0
0	0,999	0,001	0
0	0,999	0	0,001
0,001	0,998	0	0,001
0,001	0	0,999	0
0	0	0,999	0,001
0	0,002	0,998	0
0	0	0,001	0,999
0	0,001	0	0,999
0,002	0	0,001	0,997

L'*output* (un elenco di gradi di appartenenza), inoltre, è molto meno maneggevole di quello derivante da una *cluster analysis* tradizionale, in quanto si presenta sotto la forma di una tabella come la precedente, dove il numero delle righe è pari al numero dei casi, il numero delle colonne a quello dei gruppi e di

conseguenza il numero delle celle è pari al prodotto di righe e colonne. Con 1400 casi e 4 gruppi, avremmo 5600 celle, decisamente un'informazione poco utile quanto difficile da gestire.

Per superare questa *impasse*, per coniugare pregi ed eliminare i difetti dei due approcci, si può ricorrere alla *terza via*: l'utilizzo di *tecniche semi-sfocate*. Benché finora quasi ignorata negli studi individuali, essa cerca di combinare i due approcci salvandone i pregi e limitandone i difetti. La differenza tra questo metodo e quello completamente sfocato è che, se nel secondo procedimento un caso è assegnato ad un gruppo solo se coincide col baricentro, nel primo è assegnato non solo se non coincide, ma anche se la sua distanza è minore ad una certa soglia stabilita dal ricercatore. Tale soglia, che caratterizza l'*algoritmo delle k-medie semisfocate*, differenziandolo da quello delle k-medie sfocate (*Fuzzy c-means*), è espressa dal rapporto:

$$\frac{1}{\alpha} D_k$$

dove α è un parametro scelto dal ricercatore, e D_k è la distanza tra il centro del *cluster k* e il centro del *cluster* più vicino. In questo modo, un caso viene assegnato a un gruppo non solo se è abbastanza vicino ad esso, ma anche se è sufficientemente lontano dagli altri gruppi.

Dunque, nell'algoritmo delle *k-medie semisfocate*, che che più avanti vedremo all'opera, il ricercatore influisce decisamente sulla classificazione finale decidendo il valore dei due parametri m (come nel FCM) ed α , ricordando che:

$m : [1, \infty)$ → varia tra 1 e infinito; se uguale a 1 genera una classificazione tradizionale bivalente, ed il suo valore influisce sul grado di sfocatura della classificazione.

$\alpha : [1, \infty)$ → varia tra 1 e infinito; influisce sul valore della soglia minima di appartenenza, al suo crescere cresce la soglia.

Quanto detto per l'arbitrarietà del valore m nell'algoritmo FCM vale anche per il valore α in questo caso, così come le annotazioni riguardanti l'indice di Gini. Va solamente fatto presente che in questo caso i parametri decisivi su cui si va ad incidere arbitrariamente sono due, è dunque conveniente, per evitare di creare troppa confusione, manipolarne uno alla volta, lasciando l'altro inizialmente invariato.

Nei casi che abbiamo visto precedentemente in tabella, che arrivavano a gradi di appartenenza pari anche a 0,999 ma comunque inferiori a 1, l'introduzione di una soglia α , pur se minima, avrebbe sicuramente eliminato tale inutile sfocatura, alleggerendo sensibilmente il peso, se così si può dire, dell'informazione. Anche a livello di lettura dell'*output* la situazione migliora sensibilmente, vedendo affiancate regioni incerte caratterizzate da un *continuum* di appartenenza, a regioni in cui tale *continuum* lascia spazio ai classici gruppi. Aumentando o diminuendo livello di sfocatura e soglia di appartenenza, un ricercatore potrebbe poi scegliere se escludere solo i casi più incerti o isolare quelli più resistenti, approfondendo magari lo studio delle rispettive caratteristiche socio-demografiche, degli atteggiamenti e delle variabili significativamente associate. Insomma, l'approccio semi-sfocato è una scelta obbligata in alcuni casi (come gli studi a livello individuale su grandi numeri, non lo sarebbe affatto in uno studio a livello ecologico, dove il numero delle righe non è elevato), ma è non è certo una scelta di ripiego: è un obbligo pieno di possibilità che vale la pena esplorare.

CAPITOLO 4. LA CALIBRATION

Prima di muoverci oltre, analizzeremo brevemente un contributo proveniente da un altro settore di indagine, quello degli studi comparativi. Il motivo di questo interesse è dovuto al fatto che estrapoleremo una tecnica sviluppata in questo contesto (la *calibration*) per applicarla ai nostri dati e tentare così un approccio del tutto inusuale alla classificazione. Più avanti le cose saranno più chiare, per ora, prima di entrare nel merito della *calibration* come tecnica, analizziamo brevemente il contesto in cui essa è stata formalizzata, ossia la QCA (*Qualitative comparative analysis*).

4.1 LA QUALITATIVE COMPARATIVE ANALYSIS

Fino a poco più di venti anni fa, gli studiosi di politica comparata avevano a disposizione sostanzialmente due metodi tra loro molto divergenti (cfr. Wagemann, 2007): un metodo di comparazione classica basato su campioni estremamente poco numerosi, che si riallacciava ai più importanti studi del passato, come quello proposto da John Stuart Mill (cfr. Mahoney, 2003; Morlino, 2005; Skocpol, 1984), oppure un metodo che lavorava su sondaggi di massa, recuperando informazioni a livello micro che venivano poi ricomposte per rispondere ad esigenze di ricerca che si ponevano quasi esclusivamente a livello macro. Questa era la situazione fino al 1987, anno in cui lo studioso americano Charles Ragin propose una terza alternativa presentata nel libro *The Comparative Method* (Ragin, 1987), che conteneva la sopraccitata tecnica QCA.

In poche parole, il metodo di Ragin è un metodo di analisi comparativa basato sulla logica formale, che si pone lo scopo di testare l'efficacia di modelli guidati teoricamente. Poste determinate variabili indipendenti e determinate variabili dipendenti e trasformate le variabili in funzioni di appartenenza, si procede all'incrocio dell'appartenenza dei casi alle variabili indipendenti con l'appartenenza degli stessi a quelle dipendenti, alla ricerca di condizioni o combinazioni di condizioni che risultino essere o meno necessarie e/o sufficienti, con l'ambizione ultima, previo il riscontro di un'adeguata e ripetuta mole di evidenze empiriche, di ipotizzare l'esistenza di relazioni causali, pur con tutta l'attenzione e la cautela che questo tipo di affermazioni possano comportare:

As always, claims of this type [di tipo causale] cannot be based simply on the demonstration of the subset relation. Researchers should marshal as much corroborating evidence as possible when making any type of causal claim (Ragin, 2007, p. 19).

È il caso, prima di tutto, di fare una piccola premessa relativa alla questione su quale sia la tradizione metodologica su cui si innesta la QCA rispetto all'asse qualitativo/quantitativo. Come abbiamo visto in precedenza, la Q della sigla è abbreviazione di *Qualitative*, constatazione questa che, da sola, darebbe risposta alla nostra domanda. Il sottotitolo del libro del 1987 è, però, *Moving beyond qualitative and quantitative strategies*. C'è da dire, intanto, che l'etichetta QCA non è introdotta nell'opera del 1987 ed è venuta fuori nel dibattito che è seguito a tale pubblicazione nel corso degli anni. Ragin in alcuni casi afferma esplicitamente che la QCA appartiene alla tradizione dell'indagine qualitativa (Ragin, 2000, p. 13), altrove e spesso preferisce l'espressione *case-oriented* (Ragin, 2000, p. 23) o più raramente *diversity-oriented* (Ragin, 2000, p. 19). Nell'esposizione concreta del metodo si sottolinea come esso sia un percorso intermedio, "*a middle path*" (cfr. Ragin, 2007), al classico dualismo, che cerca di coniugare il meglio di entrambi gli approcci.

Insomma, se da una parte resta il dubbio su che tipo di metodo sia la QCA (qualitativo o alternativo?), dall'altra possiamo "tagliare corto" affermando che essa sia un metodo che poggia le sue radici nell'analisi qualitativa per poi proporsi come ponte tra i due mondi (per una trattazione più esauriente, cfr. Wagemann,

2007; Ragin, 2004; Brady e Collier, 2004; Seawright e Collier, 2004; DeMeur e Rihoux, 2002).

Vediamo ora, brevemente, la QCA nei suoi aspetti più importanti, premettendo che, quando si usa l'abbreviazione QCA ci si riferisce normalmente alla prima versione presentata nel libro del 1987; sarebbe più corretto utilizzare la sigla csQCA (*Crisp Set Qualitative Comparative Analysis*) per differenziarla dalla sua versione *fuzzy* (fsQCA, *Fuzzy Set Qualitative Comparative Analysis*), presentata nel secondo famoso libro di Ragin, *Fuzzy Set Social Science* (Ragin, 2000). Entrambe queste versioni condividono, comunque, i medesimi obiettivi e lo stesso procedimento formale che – attraverso la mediazione della teoria degli insiemi – opera per implicazione logica (cfr. Ragin, 1987, 2000; DeMeur e Rihoux, 2002; Schneider e Wagemann, 2007; Wagemann, 2007); l'esposizione può, quindi, procedere di pari passo e per questo motivo, nel corso delle pagine successive, utilizzeremo la sola sigla QCA. Esiste anche versione a multi-valori (mvQCA, *Multi value Qualitative Comparative Analysis*) (cfr. Cronqvist, 2005), che per motivi tecnici e teorici differisce parzialmente dalle altre due ed ha trovato poche applicazioni concrete di ricerca (per una revisione critica, cfr. Vink e van Vliet, 2007).

L'obiettivo della QCA è, in sintesi, quello di testare se una determinata condizione (o, come si direbbe nei termini tipici della canonica analisi statistica a logica covariazionale: una variabile indipendente) sia sufficiente, oppure necessaria, oppure sufficiente e necessaria, oppure né sufficiente né necessaria per il realizzarsi di un determinato esito (in termini di analisi quantitativa: variabile dipendente). Questioni di questo tipo non sono inusuali nelle scienze sociali (cfr. Seawright, 2002) e sono affrontate dal punto di vista della QCA come tentativo di risposta ai seguenti quesiti (cfr. Wagemann, 2007):

1. la condizione ha ruolo causale¹ per l'esito oppure no?

¹ Per ogni volta che utilizzeremo il concetto di causalità, rimandiamo alla citazione riportata in precedenza, in modo da rammentare costantemente l'attenzione, la cautela e la mole e regolarità di evidenze empiriche necessarie per affermazioni che vanno in questa direzione. Una delle principali critiche che vengono mosse alla QCA è, effettivamente, quella di determinismo; va però

2. se sì, è una condizione sufficiente, una condizione necessaria, o è allo stesso tempo sufficiente e necessaria?
3. se la condizione ha ruolo causale per l'esito, ma non è né sufficiente né necessaria, allora come può essere specificato il suo ruolo causale?

Tipicamente, nell'analisi del mondo sociale, la QCA incorre nella terza situazione. Prima di tutto si costruisce una tabella che incroci i casi con le condizioni e gli esiti, in modo che sia possibile segnare per ogni caso la presenza o l'assenza di esiti e condizioni, per estrapolare alla fine una sorta di norma di comportamento, seguendo le normali regole dell'algebra booleana (nel caso della csQCA) e dell'algebra *fuzzy* (nel caso della fsQCA): tali tavole sono dette *truth table*. Un esempio di relazione che scaturisce da tale analisi può essere il seguente (cfr. Wagemann, 2007):

$$AB + aC \rightarrow Y$$

Date tre condizioni (A, B e C), ci sono due possibili modi, alternativi e non combinati, per arrivare all'esito Y: la presenza simultanea delle condizioni A e B, oppure l'assenza della condizione A e la presenza della condizione C. Si noti come la presenza di diversi percorsi e l'utilizzo di espressioni come "presenza simultanea" e "implicazione logica" attenuano, a nostro avviso, la tensione verso la relazione causale, come notavamo in precedenza.

Nello specifico di questa relazione, notiamo come nessuna condizione sia sufficiente (se la condizione A fosse sufficiente, avremmo avuto una soluzione del tipo: $A \rightarrow Y$) e nessuna sia necessaria (altrimenti almeno una delle 3 condizioni sarebbe stata presente in ogni possibile combinazione: ad es., $AB + Ac \rightarrow Y$):

considerato come, ad affermazioni del tipo "A è causa di B", si preferiscano affermazioni come "A implica logicamente B" oppure "A è sottoinsieme di B". La logica di questa analisi è, infatti, quella della relazione fra insiemi e sottoinsiemi ed in questa ottica l'implicazione logica, che di fatto sostituisce l'attribuzione causale, ci sembra essere pratica legittima e, per così dire, indolore (cfr. Goldthorpe, 1997; Mahoney, 2000; Goertz, 2005; Waldner, 2005).

sono le classiche condizioni INUS (*Insufficient but Necessary part of a condition which is itself Unnecessary but Sufficient for the result*) (cfr. Goertz, 2003, p. 68; Mackie, 1974, p. 62), caso tipico nelle scienze sociali e motivo per cui la QCA si interessa tipicamente non di condizioni ma di combinazioni di esse e dell'analisi puntuale e rigorosa di tali combinazioni.

Vediamo qualche altra caratteristica della QCA, prima di entrare nel merito di ciò che maggiormente ci interessa, in quanto sarà utilizzato nella nostra applicazione concreta, ossia la tecnica della *calibration*.

La prima riguarda il concetto di causalità, a cui abbiamo brevemente accennato. Posto, ribadiamo, che si preferisce utilizzare l'espressione meno forte "condizione", vediamo come il concetto in questione abbia in questo caso, sempre per la particolare strutturazione del metodo della QCA che ruota intorno alla teoria degli insiemi (classici e sfocati), caratteristiche specifiche che rendono piuttosto interessante questo metodo: l'equifinalità, la causalità congiunturale, e la causalità asimmetrica (cfr. Wagemann, 2007).

Equifinalità significa che è possibile trovare più di una condizione (o combinazione di condizioni) sufficiente ad implicare l'esito:

$$Ab + C \rightarrow Y$$

In questo caso funzionano come condizioni sia la simultanea presenza di A ed assenza di B che la sola presenza di C. Nei metodi statistici tradizionali, specialmente nella regressione, il contributo di ogni variabile alla spiegazione della variazione nella variabile dipendente è singolo ed in caso di alternative, grazie al controllo dei valori dei coefficienti sul test di significatività, solo una variabile resta valida. In questo senso, la QCA si dimostra più adatta allo studio delle combinazioni di condizioni.

Causalità congiunturale significa che alcune condizioni sono sufficienti solo ed esclusivamente se si presentano insieme:

$$Ab + C \rightarrow Y$$

L'equazione in esempio è la stessa di prima, l'attenzione ora va posta sul fatto che A e B, prese singolarmente, come presenza od assenza, non rappresentano condizione sufficiente allo sviluppo dell'evento; prese simultaneamente, sono invece condizione sufficiente. Mentre, statisticamente, la presenza di due variabili molto correlate è vista negativamente come multicollinearità (cfr. Wagemann, 2007).

La causalità asimmetrica significa che c'è un percorso logico nel processo di condizione ed esito:

$$A + B + C \rightarrow Y$$

In questo caso, l'accento va posto sulla freccia. Questa caratteristica differenzia questo tipo di causalità dalla covariazione. Non stiamo dicendo che C e Y siano altamente e positivamente correlati, ma che "se C, allora Y", il che non implica la relazione inversa "se Y, allora C". Anche questa caratteristica deriva dalla struttura a "relazioni fra insiemi e sottoinsiemi" propria del metodo della QCA, struttura che circoscrive tale metodo, ma lo rende probabilmente più funzionale ad alcuni ambiti di ricerca e più malleabile teoricamente e ancorabile empiricamente.

Altra peculiarità della QCA riguarda la popolazione ed i casi. In statistica il campionamento è un momento fondamentale dell'analisi, tanto che la statistica inferenziale si può pensare come una scienza che analizza i rapporti tra un campione ed una popolazione. Spesso, tuttavia, il momento centrale e precedente della definizione della popolazione, momento che in realtà dovrebbe presentarsi contestualmente alla stessa formulazione della domanda di ricerca, viene trattato sommariamente. Si studia il bipartitismo perché si vuole fare un'analisi dell'evoluzione politica italiana dalla metà degli anni '90 in poi o perché si vuole fare un confronto tra sistema elettorale inglese ed americano? E la portata delle nostre generalizzazioni a quale popolazione è applicabile nella dimensione spazio-temporale? Tutti i paesi del mondo oggi, le democrazie occidentali oggi, i paesi comunisti durante la guerra fredda (cfr. Wagemann, 2007)? Si tratta del problema

del *casing* (cfr. Ragin e Becker, 2001), ossia la risposta alla domanda *what is my case a case of* (cfr. Ragin, 2000, p. 53 e seguenti).

La struttura della QCA, con la costruzione delle tavole di verità, impone questo tipo di riflessione. Essendo un metodo *diversity-oriented*, nelle tavole devono essere presenti tutte le combinazioni possibili, in termini di assenza e presenza. Se si studiano sviluppo economico e democrazia, dobbiamo avere casi in cui lo sviluppo economico è presente e casi in cui è assente, casi in cui la democrazia è presente e casi in cui la democrazia è assente. Questo rende chiaro come, in realtà, costruire una tavola di verità su alcune condizioni equivale a costruire uno spazio di attributi su alcune proprietà (cfr. Lazarsfeld, 1937): nei record di riga della tavola troviamo gli stati (presenza/assenza – espressi rispettivamente con i valori di 1 e 0 – per la csQCA; gradi di appartenenza – espressi in valori che variano tra gli estremi di 0 ed 1 – per la fsQCA) dei casi sulle condizioni (che vanno in colonna). I casi, in riga, corrispondono dunque a delle vere e proprie configurazioni di casi, a dei tipi formati intensionalmente (vale la pena notare come nella versione fsQCA, dove 0 ed 1 sono casi estremi di un'appartenenza sfocata, si possa affermare che tali casi tendenti alla bivarianza possono essere interpretati come estremizzazioni empiriche proto-tipiche – non ideal-tipiche, in quanto l'ideal-tipo per definizione ha estensione pari a zero, appartiene ad una sfera esclusivamente concettuale – delle proprietà).

In più, va considerato che i confini dello spazio di attributi identificano i confini di generalizzazione dei risultati. Con un campione così costruito, è chiaro che nessun caso finirà “fuori” dallo spazio. Se pensiamo ad uno spazio di attributi a due proprietà, quindi ad una tavola di verità a due condizioni, nessun caso può finire al di fuori del quadrato che geometricamente corrisponde a tale spazio. Nel caso dell'analisi classica i casi andranno in uno dei quattro quadranti che si formano, nel caso dell'analisi *fuzzy* i casi possono disporsi sia sui confini del quadrato, sia al suo interno, ed i casi che si disporranno sui quattro angoli saranno casi speciali, estremi di quella possibile combinazione empirica di proprietà (cfr. Ragin, 2000). Se lo spazio è a tre proprietà, sarà rappresentato da un cubo *fuzzy* con otto angoli bivalenti:

la linea numerica da zero a uno definisce un ipercubo a una dimensione. Un singolo quadrato definisce un ipercubo o un cubo *fuzzy* a due dimensioni (con quattro combinazioni o insiemi dicotomici di due oggetti). Un cubo solido come un cubetto di ghiaccio o un cubo di Rubik definisce un cubo *fuzzy* a tre dimensioni (con otto combinazioni dicotomiche di tre oggetti). In ciascun caso gli angoli definiscono gli unici risultati aristotelici – dicotomici – in cui è soddisfatto A O non-A. Il punto centrale del cubo definisce l'unico risultato buddista in A è uguale a non-A, ossia in cui l'equazione yin-yang è soddisfatta al 100%: A = non-A. Il punto centrale è l'unico punto del cubo che è ugualmente lontano da, e vicino a, ogni angolo (Kosko, 1993, p. 127).

Occorre sottolineare un problema legato a questa proprietà della QCA per cui i casi sono in realtà configurazioni di casi, ossia la questione della *limited diversity*. Dovendo ogni possibile combinazione delle proprietà essere rappresentata in almeno una riga, ne consegue che il numero delle condizioni influenza esponenzialmente il numero dei casi:

$$r = 2^k$$

dove r è il numero delle righe e k quelle delle condizioni. Con appena 5 condizioni avremmo, ad esempio, già 32 diverse configurazioni, per cui *almeno* 32 casi empirici (ragionevolmente molti di più). Se consideriamo che la QCA, in quanto metodo comparativo nato nelle scienze politiche, lavora su unità di analisi ecologiche, capiamo la difficoltà nel rintracciare un numero così elevato di configurazioni. Con circa 200 paesi esistenti al mondo, si potrebbe, ipoteticamente, lavorare con 7 condizioni al massimo (pari a 128 configurazioni; con 8 condizioni saremmo già a 256 configurazioni di casi necessarie), è però comprensibile come una costruzione del campione/popolazione del genere sia empiricamente irrealizzabile. Ne consegue che la QCA sia un metodo adatto a lavorare su n piccoli o medi, qualora, ad esempio, un n di 32 casi/configurazioni (5 condizioni; oppure un n di 64 configurazioni, pari a 6 condizioni) voglia essere considerato medio.

Benché il problema della diversità limitata a livello ecologico sia sostanzialmente insormontabile, è apprezzabile l'aiuto che la QCA fornisce nei termini di consapevolezza acquisita nel momento della definizione della popolazione. In secondo luogo, sono importanti, a nostro avviso, le conseguenze

implicite al riconoscimento dell'uguaglianza tra casi e configurazioni, da una parte, e tavole di verità e spazi di attributi, dall'altra, che abbiamo appena analizzato.

Soprattutto questo secondo punto potrebbe spingere la riflessione in direzione di un'interessante domanda: una classificazione costruita grazie all'uso congiunto di *truth table* e *calibration* è una classificazione che impropriamente potremmo definire "automatica intensionale"? È una definizione impropria perché contraddittoria in termini. Il punto che vogliamo sostenere è però il seguente: è possibile affermare che l'uso congiunto di *calibration* e *truth table* stia alla classificazione intensionale come le tecniche di *cluster analysis* stanno alla classificazione estensionale, in modo da esserne, per così dire, una sorta di "prosecuzione tecnica"? Tenteremo una risposta dopo le nostre applicazioni; ora andiamo avanti con l'analisi di questa procedura.

4.2 LA TECNICA DELLA CALIBRATION

Va prima di tutto specificato che la *calibration* si applica esclusivamente se si assume un'ottica *fuzzy*, altrimenti le variazioni dei casi vengono espresse in termini di presenza o assenza di determinate proprietà. Questo significa che tale tecnica compare solamente nella versione a *fuzzy set* della QCA ed il dibattito su di essa, con le relative applicazioni, si sviluppa soltanto dopo la pubblicazione del libro del 2000. È un dibattito molto giovane, quindi, come d'altronde è recente l'applicazione stessa della logica *fuzzy* alle scienze sociali, come sottolinea lo stesso Ragin:

Fuzzy sets are relatively new to social science. The first comprehensive introduction of fuzzy sets to the social sciences was offered by Michael Smithson (1987). However, applications were few and far between until the basic principles of fuzzy set analysis were elaborated through Qualitative Comparative Analysis (Ragin, 2007, p. 1).

Secondo Ragin, la chiave per un utilizzo proficuo dell'analisi a *fuzzy set* sta nella buona costruzione degli insiemi; emerge, di conseguenza, come

fondamentale il momento della calibrazione delle variabili originali. Egli considera le misure calibrate superiori, per molti aspetti, a quelle non calibrate e vede in esse uno strumento utile a colmare il divario tra strategie qualitative e quantitative che, a dispetto delle molteplici dichiarazioni d'intenti, ancora esiste nelle scienze sociali:

In essence, I argue that fuzzy sets offer a middle path between quantitative and qualitative measurement. However, this middle path is not a compromise between these two; rather, it transcends many of the limitations of both (Ragin, 2007, p. 1).

Ragin sottolinea come la *calibration* sia una procedura necessaria e di routine in molte scienze naturali come la chimica, la fisica e l'astronomia (cfr. Pawson, 1989). L'aggiustamento è fatto in relazione ad alcuni standard conosciuti. Ad esempio, la temperatura di 20 gradi Celsius è facilmente interpretabile perché situata tra due “*qualitative breakpoints*” (Ragin, 2007, p. 2): il momento in cui l'acqua ghiaccia (0° C) ed il momento in cui bolle (100° C). Questi momenti qualitativi, per indicare i quali la scala Celsius è appositamente calibrata, vengono anche definiti *phase shifts*, spostamenti di fase; è fondamentale, dunque, ancorare la *calibration* a tali punti, che affondano le loro radici alla conoscenza sostantiva, teorica ed empirica che è esterna alla calibrazione:

Knowledge of these phase shifts, which is external to the measurement of temperature per se, provides the basis for its calibration (Ragin, 2007, p. 2).

Tali criteri esterni di conoscenza sostantiva sono decisivi nell'applicazione della *calibration* e differenziano le variabili calibrate dalle variabili non calibrate, a vantaggio – secondo Ragin – delle prime, in quanto le seconde si limitano a mostrare la posizione di un caso rispetto ad un altro o rispetto alla media di un campione:

Uncalibrated measures, however, are clearly inferior to calibrated measures. With an uncalibrated measure of temperature, for example, it is possible to know that one object has a higher temperature than another or even that it has a higher temperature than average for a given

set of objects, but still not know whether it is hot or cold. Likewise, with an uncalibrated measure of democracy it is possible to know that one country is more democratic than another or more democratic than average, but still not know if it is more a democracy or an autocracy (Ragin, 2007, p. 2).

Così, ad esempio, se un paese appartiene all'insieme dei paesi democratici con una misura calibrata pari a 0,7 ed un altro con una misura calibrata pari a 0,3 sappiamo subito dire che il primo paese è più democratico che autoritario ed il secondo più autoritario che democratico, semplicemente perché essi sono rispettivamente sopra e sotto al valore centrale 0,5 che rappresenta il punto di *crossover*, il punto in cui un paese è nella stessa misura democratico e non democratico. Se invece si lavora con una misura non calibrata, fosse anche una scala di rapporti, e si sa che il primo paese ha un valore di democrazia pari a 18, mentre il secondo ne ha uno pari a 14, si può soltanto dire che il secondo è meno democratico del primo, ma potrebbero essere entrambi democratici od entrambi autoritari. In questo senso, la forza delle misure calibrate sta nell'aver non uno, ma ben tre punti di ancoramento:

While most textbooks assert that ratio scales are the highest form of measurement because they are anchored by a meaningful zero point, it is important to note that fuzzy sets have three numerical anchors: 1.0 (full membership), 0.0 (full nonmembership), and 0.5 (the cross-over point separating “more in” versus “more out” of the set in question). If it is accepted that such “anchoring” signals a higher level of measurement, then it follows that a fuzzy set is a higher level of measurement than a ratio-scale variable (Ragin, 2007, p. 4).

Si potrà obiettare che, conoscendo il valore massimo di una scala di rapporti, si può trovare un valore centrale in confronto al quale valutare gli altri. È certamente possibile, ma, di fatto, questo costituisce una forma implicita di calibrazione, una forma peraltro rudimentale, grezza, in quanto basata solo sui valori presenti in uno specifico insieme di casi:

Notice that the use of deviations from sample-specific measures of central tendency offers a crude but passive form of calibration. Its crudeness lies in the fact that the calibration standards (e.g., the mean and standard deviation) vary from one sample to the next and are inductively

derived. By contrast, the routine practice in the physical sciences is to base calibration on external, dependably known standards (e.g., the boiling point of water) (Ragin, 2007, p. 6).

Così se, ad esempio, avessimo 20 recipienti contenenti acqua, la cui temperatura vari da un minimo di 60 gradi Celsius ad un massimo di 95 e con una media di 78, potremmo essere tentati dall'affermare che l'acqua a 60 gradi è "fredda".

In più, il mancato riferimento a criteri esterni al campione si riflette anche in un'altra considerazione, cioè che in una scala di rapporti tutte le variazioni sono identiche, mentre non sono sempre identiche nella realtà. Se si misura la ricchezza di una nazione prendendo come variabile il reddito nazionale pro capite, la distanza tra i casi è sempre la stessa, indipendentemente dal fatto che sia la differenza tra 1.000 e 2.000 dollari o la differenza tra 21.000 e 22.000 dollari, ma:

Actually, there is a world of difference between living in a country with a GNP per capita of \$2,000 and living in one with a GNP per capita of \$1,000; however there is virtually no difference between living in one with a GNP per capita of \$22,000 and living in one with a GNP per capita of \$21,000. Such fine points are rarely addressed by researchers who use the conventional indicator approach, but they must be confronted directly in research that uses calibrated measures (e.g., fuzzy sets) (Ragin, 2007, p. 4).

Per quanto riguarda, invece, le "misurazioni" qualitative, il confronto tra esse e quelle non calibrate è su altri piani. Prendiamo come esempio uno studio sulle differenze tra nazioni che hanno avuto antica o recente formazione dello stato (cfr. Rokkan, 1975; Laitin, 1992). Come si determina qual è una formazione recente o antica? Ovviamente, la formazione può essere datata e tale variabile, in anni, può essere trattata come una scala di rapporti. Probabilmente, però, la variazione scaturita da una tale procedura è irrilevante ai fini che si pone in questo caso il ricercatore. Supponiamo, ad esempio, che uno stato si sia formato 500 anni fa ed un altro 250. Il primo è vecchio esattamente due volte il secondo, ma entrambi sono di formazione antica, se visti attraverso le lenti di una conoscenza esterna, teorica, sostantiva. Un tipo di ragionamento simile, che è frequente nelle ricerche qualitative, è una forma implicita di calibrazione:

“Age in years” must be adjusted on the basis of accumulated substantive knowledge in order to be able to interpret “early” versus “late” in a way that resonates appropriately with existing theory. Such calibrations are routine in qualitative work, even though they are rarely modeled or even stated explicitly (Ragin, 2007, p. 7).

Ritorna, dunque, l’idea per cui in una *calibration* le variazioni non sono sempre le stesse ed alcune scompaiono di fronte alla scelta dei criteri per calibrare.

Un’altra caratteristica tipica nei lavori qualitativi è che essi sono più orientati ai casi che alle variabili, a differenza di quanto accade nelle ricerche quantitative. Qui l’attenzione è posta su specifici insiemi di casi, sull’insieme dei paesi democratici, ad esempio, non sul livello di democrazia. La differenza è sottile ma importante in quanto spinge a ragionare in termini di appartenenza ed a cercare quali possano essere i criteri per determinare se un caso è fuori da un gruppo oppure no.

Queste caratteristiche avvicinano molto il lavoro qualitativo alla pratica di calibrazione, per questo motivo evidentemente Ragin, pur considerando la QCA un metodo ponte tra il mondo quantitativo e qualitativo, la inserisce comunque in filone che trae origine nel secondo:

the case-oriented view is more compatible with the idea that measures should be calibrated, for the focus is on the degree to which cases satisfy membership criteria, which in turn are usually externally determined, not inductively derived (e.g., using the sample mean). These membership criteria must reflect agreed-upon standards; otherwise, the constitution of a category or set will be contested (Ragin, 2007, p. 8).

Ad ogni modo, le “misurazioni” nelle ricerche qualitative mancano generalmente di precisione e spesso si rivelano non sistematiche e non scientifiche («haphazard and unscientific», Ragin, 2007, p. 8).

Il tentativo di Ragin si inserisce fra questi due universi, con l’ambizione di prendere il meglio di entrambi, in termini di precisione, da un lato, e riferimento a criteri sostantivi, dall’altro:

With fuzzy sets it is possible to have the best of both worlds, namely, the precision that is prized by quantitative researchers and the use of substantive knowledge to calibrate measures that is central to qualitative research. With fuzzy sets, precision comes in the form of quantitative assessments of degree of set membership, which can range from a score of 0.0 (full exclusion from a set) to 1.0 (full inclusion). For example, a country might have a membership score of 0.85 in the set of democracies, indicating that it is clearly more in this set than out, but still not fully in. Substantive knowledge provides the external criteria that make it possible to calibrate measures. This knowledge indicates what constitutes full membership, full nonmembership, and the point at which cases are more “in” a given set than “out” (Ragin, 2007, p. 9).

Secondo Ragin, idealmente, la calibrazione dei gradi di appartenenza dovrebbe essere basata interamente sulla conoscenza teorica e sostantiva del ricercatore, quindi sulla conoscenza collettiva di base degli scienziati sociali. Questo è non possibile in quanto le scienze sociali sono ancora nella fase della loro infanzia (cfr. Ragin, 2007, p. 11) e dunque questa base non esiste. A detta di Ragin, ciò è dovuto anche al fatto che il modello *variable-oriented* – la cui attenzione cade fundamentalmente sulla covariazione e sul confronto con valori medi centrali come chiave per valutare le relazioni fra i casi – sia quello dominante, e questo devia l’attenzione degli studiosi dalla ricerca di specifiche soglie che siano solidamente ancorate nella conoscenza sostantiva.

Ciononostante, è ancora possibile rintracciare alcuni criteri sulla cui base calibrare le variabili:

The external criteria that are used to calibrate measures and translate them into set membership scores may reflect standards based on social knowledge (e.g., the fact that twelve years of education constitutes an important educational threshold), collective social scientific knowledge (e.g., about variation in economic development and what it takes to be considered fully in the set of “developed” countries), or the researcher’s own accumulated knowledge, derived from the study of specific cases (Ragin, 2007, pp. 9-10).

Ci si può anche aiutare con criteri matematici oppure osservando la curva di distribuzione dei casi sulla proprietà, l’importante è che questi criteri vengano applicati in modo trasparente e sistematico, così che la calibrazione possa essere distinta da un generico ed implicito aggiustamento qualitativo.

Stabiliti i criteri, si può operare grazie ad una tecnica rodata che andiamo ora, finalmente, ad illustrare.

Cominciamo col dire che ci sono due diversi metodi per calibrare, uno diretto ed uno indiretto. Il metodo diretto si basa sulla determinazione di tre ancore qualitative (la piena appartenenza, il punto intermedio ed la piena non appartenenza) che determinano la struttura dell'insieme *fuzzy*. Il metodo indiretto utilizza invece tecniche di regressione per stimare le appartenenze dei casi, basandosi su uno schema iniziale di codifica a sei valori. Entrambi i metodi si avvalgono della metrica dei logaritmi naturali ed utilizzano, nello specifico, l'inverso della funzione *logit* per arrivare alla codifica finale, variante tra zero ed uno. Concentreremo l'analisi prevalentemente sul metodo diretto, visto che è quello che useremo nella pratica.

Entrambe le tecniche richiedono la soddisfazione di due requisiti preliminari all'applicazione:

1. il ricercatore deve avere a propria disposizione una convenzionale scala numerica, che sia a intervalli o di rapporto, come indicatrice del concetto su cui sta lavorando;
2. i concetti possono essere strutturati ed etichettati in termini di appartenenza ad un insieme assunto come insieme di riferimento.

Vediamo meglio questo secondo punto. Una delle differenze chiave tra un insieme *fuzzy* ed una variabile convenzionale sta nella loro concettualizzazione e nel loro etichettamento. Ad esempio, è possibile costruire una generica variabile “anni di istruzione”, ma è impossibile trasformare questa variabile direttamente in un *fuzzy set* senza prima designare e definire uno specifico «target set of cases» (Ragin, 2007, p. 10). Così, ad esempio, un ricercatore potrebbe essere interessato all'insieme degli “individui che hanno almeno un'istruzione superiore secondaria” oppure all'insieme degli “individui che hanno un'istruzione universitaria completa”. In questo caso, è chiaro come diverse designazioni degli insiemi di riferimento abbiano come conseguenza diverse calibrazioni. Una persona al primo anno di università avrà chiaramente appartenenza piena al primo gruppo, ma non

piena al secondo. Questo esempio, tra l'altro, evidenzia una delle principali caratteristiche della *calibration*, che abbiamo già avuto modo di vedere, cioè che va fatta distinzione tra le variazioni rilevanti e quelle irrilevanti: la differenza tra una matricola ed un laureato è ininfluenza nella costruzione del *set* degli individui con un'istruzione almeno secondaria.

Allo stesso modo, studiando l'economia di più nazioni ed utilizzando il reddito pro-capite nazionale, sarebbe necessario passare da un generico "livello di sviluppo economico" ad uno specifico *target set*, come ad esempio quelli dei "paesi sviluppati"; oppure studiando il pregiudizio etnico, dal "livello di pregiudizio" dovremmo individuare il gruppo degli "individui con pregiudizio" o quello degli "individui senza pregiudizio". Questo, prima di tutto incide direttamente sui valori calibrati: rispetto al primo gruppo, è lecito pensare che un neo-nazista abbia un valore di appartenenza pari a 1, rispetto al secondo pari a 0. In più, afferma Ragin, questa impostazione lega in modo semplice e diretto il discorso teorico all'analisi empirica, in quanto il primo assume generalmente la forma di legame tra specifici insiemi:

Notice that this requirement – that the researcher designate a target set – not only structures the calibration of the set, it also provides a direct connection between theoretical discourse and empirical analysis. After all, it is more common for theoretical discourse to be organized around designated sets of cases (e.g., the "developed countries") than it is for it to be organized around generic variables (e.g., "level of economic development") (Ragin, 2007, p 10).

Ancora nelle pagine successive si nota come la maggior parte delle teorie sociali siano formulate in termini verbali:

Calibrated measures [...] are especially useful when it comes to evaluating theory that is formulated in terms of set relations. While some social science theory is strictly mathematical, the vast majority of it is verbal. Verbal theory, in turn, is formulated almost entirely in terms of set relations (Ragin 2000; 2006) (Ragin, 2007, p. 18).

Consideriamo, ad esempio, l'affermazione "i paesi sviluppati sono democratici": essa, e molte altre ad essa simili, può essere interpretata come una

set relation, ossia i casi dell'insieme menzionato per primo (i paesi sviluppati) sono un *subset* dei casi dell'insieme menzionato per secondo (i paesi democratici).

Il ragionamento teorico è un “gioco” con insiemi, dunque, per riprendere le parole di Kosko, secondo cui il ragionamento è un gioco con insiemi *fuzzy* (Kosko, 1993, p. 149).

Soddisfatti dunque questi due requisiti, possiamo procedere all'applicazione. Il metodo diretto, dicevamo, utilizza l'inverso di una funzione *logit* e si serve, come passo intermedio, dei logaritmi naturali degli *odds*² di appartenenza. Consideriamo la tabella 4.1 (tratta – come le altre che mostreremo in questo capitolo – da Ragin, 2007), che mostra le diverse metriche utilizzate nell'applicazione del metodo diretto:

Tab. 4.1 – Trasformazione matematica delle etichette verbali

1. Etichetta verbale	2. Grado di appartenenza	3. Odd associato	4. Log odds sulla piena appartenenza
Piena appartenenza	0,993	148,41	5,0
Soglia della piena appartenenza	0,953	20,09	3,0
Prevalentemente dentro	0,881	7,39	2,0
Più dentro che fuori	0,622	1,65	0,5
Punto di crossover	0,500	1,00	0,0
Più fuori che dentro	0,378	0,61	-0,5
Prevalentemente fuori	0,119	0,14	-2,0
Soglia della piena non appartenenza	0,047	0,05	-3,0
Piena non appartenenza	0,007	0,01	-5,0

La prima colonna rappresenta le diverse etichette verbali che possono essere assegnate a vari gradi di appartenenza, da quella totale a quella nulla. La seconda colonna mostra alcuni gradi di appartenenza corrispondenti. Tre di questi sono fondamentali, in quanto sono le tre “ancore” scelte dal ricercatore: la soglia della piena appartenenza (0,95), il punto intermedio (0,5) e la soglia di appartenenza nulla (0,05). Posti questi valori, è possibile, come vedremo, eseguire il calcolo degli altri. La terza colonna mostra gli *odds* di appartenenza, calcolati secondo la seguente formula:

² Con il termine inglese *odds* in statistica si intende il rapporto tra la probabilità p di un evento e la probabilità $(1-p)$ dell'evento complementare, in formula: $p / 1-p$. Il logaritmo naturale degli *odds* è detto, appunto, *logit*. In formula: $\ln (p / 1-p)$.

$$\text{Odds di appartenenza} = (\text{grado di appartenenza}) / (1 - \text{grado di appartenenza})$$

L'ultima colonna riporta i logaritmi naturali degli *odds* così ottenuti. In effetti, possiamo affermare che i numeri riportati dalla colonna 2 alla colonna 4 siano diverse rappresentazioni, usando diverse metriche, dei medesimi valori. L'obiettivo è, dunque, trasformare la variabile numerica originale nella metrica dei *log odds* in un modo che rispetti l'etichettamento verbale che vediamo nella prima colonna.

Si noti come, nella seconda colonna, non compare nessun punteggio esattamente pari a 0 o 1. Con questa procedura è, infatti, matematicamente impossibile produrre tali gradi di appartenenza, in quanto essi corrisponderebbero ai valori, rispettivamente, di meno e più infinito per i *log odds*. Per questo motivo, i punteggi superiori a 0,95 sono interpretati come piena appartenenza al *target set* di riferimento e quelli inferiori a 0,05 come piena non appartenenza.

Lavorare con la metrica dei *log odd* è utile perché questa è completamente simmetrica rispetto allo 0 (che rappresenta un rapporto di 0,5/0,5) e non soffre né dell'effetto "pavimento" (*floor effect*) né di quello "soffitto" (*ceiling effect*) (Ragin, 2007, p. 13): in questo modo, se anche i valori degli *odds* presentassero numeri piuttosto grandi sia in positivo che in negativo, la loro traduzione in gradi di appartenenza sarebbe comunque compresa tra i confini 0 e 1, come una tipica appartenenza sfocata richiede. Tale trasformazione finale viene conseguita attraverso il calcolo dell'inverso di una funzione *logit*, secondo la seguente formula:

$$\text{Grado di appartenenza} = e^{\log \text{ odd}} / 1 + e^{\log \text{ odd}}$$

dove *e* è una costante matematica, base dei logaritmi naturali, che viene elevata utilizzando il valore di *log odd* come potenza; il risultato di questa operazione viene diviso per lo stesso valore aumentato di 1.

Vediamo ora un esempio che ci permetta, passo dopo passo, di illustrare la procedura. Il *target set* è quello dei "paesi sviluppati" e l'obiettivo è utilizzare il reddito nazionale pro-capite per calibrare i gradi di appartenenza a tale insieme. Il

campione incluso nella dimostrazione è di 136 paesi, si riporteranno i dati di 24 di questi, che sono stati scelti in quanto rappresentano un raggio molto ampio della variabile originale (cfr. Ragin, 2007). Il primo passo è scegliere i tre valori ai quali ancorare tutto il processo di calibrazione; in questo caso la scelta è caduta sui valori che riportiamo in tabella:

Tab. 4.2: Valori di ancoramento

Grado di appartenenza	Reddito pro-capite in USD
0,95	20.000
0,50	5.000
0,05	2.500

Una volta poste queste condizioni, abbiamo tutte le informazioni necessarie e sufficienti ad elaborare il calcolo. Il primo passo è calcolare gli *odds* di questi tre valori con la formula vista in precedenza, ottenendo così i valori riportati nella terza colonna della tabella 4.1; calcolando il logaritmo naturale di questi ultimi, abbiamo i tre valori di *log odds* corrispondenti alle tre ancore scelte dal ricercatore, che possiamo vedere nella quarta colonna della tabella 4.1 e che riportiamo in quella seguente. Come avevamo segnalato in precedenza, vediamo chiaramente come la metrica dei *log odds* ruoti intorno allo zero:

Tab. 4.3: Log odds dei valori di ancoramento

Grado di appartenenza	Log odds
0,95	3
0,50	0
0,05	-3

Mettiamo da parte questi valori e procediamo con l'analisi, riportando in tabella 4.4 tutti i valori corrispondenti alle 24 nazioni in esame. Il primo passo è calcolare la deviazione dal *crossover* di tutti i casi in esame. Per deviazione dal *crossover* si intende una semplice sottrazione operata sulla variabile originale: dal reddito pro-capite di ogni paese di sottrae il reddito pro-capite designato come punto intermedio tra i paesi sviluppati e quelli non sviluppati. Così, ad esempio, l'Argentina ha una deviazione pari a $(7470-5000) = 2470$ dollari americani; la

Turchia pari a (2980-5000) = -2020. Tutti i risultati sono riportati nella seconda colonna.

Tab. 4.4: Calibration dei gradi di appartenenza al set dei Paesi Sviluppati:
Metodo Diretto

Nazione	1. Reddito nazionale	2. Dev. dal crossover	3. Scalars	4. Prodotto di 2 x 3	5. Grado appart.
Svizzera	40110	35110	.0002	7.02	1.00
Stati Uniti	34400	29400	.0002	5.88	1.00
Paesi Bassi	25200	20200	.0002	4.04	.98
Finlandia	24920	19920	.0002	3.98	.98
Australia	20060	15060	.0002	3.01	.95
Israele	17090	12090	.0002	2.42	.92
Spagna	15320	10320	.0002	2.06	.89
Nuova Zelanda	13680	8680	.0002	1.74	.85
Cipro	11720	6720	.0002	1.34	.79
Grecia	11290	6290	.0002	1.26	.78
Portogallo	10940	5940	.0002	1.19	.77
Repubblica di Corea	9800	4800	.0002	.96	.72
Argentina	7470	2470	.0002	.49	.62
Ungheria	4670	-330	.0012	-.40	.40
Venezuela	4100	-900	.0012	-1.08	.25
Estonia	4070	-930	.0012	-1.12	.25
Panama	3740	-1260	.0012	-1.51	.18
Mauritius	3690	-1310	.0012	-1.57	.17
Brasile	3590	-1410	.0012	-1.69	.16
Turchia	2980	-2020	.0012	-2.42	.08
Bolivia	1000	-4000	.0012	-4.80	.01
Costa d'Avorio	650	-4350	.0012	-5.22	.01
Senegal	450	-4550	.0012	-5.46	.00
Burundi	110	-4890	.0012	-5.87	.00

Calcoliamo, dunque, i valori riportati in terza colonna, denominati *scalars*. Come è possibile vedere, ci sono due punteggi di *scalars*, quelli per i casi sopra il punto di *crossover* e quelli sotto. Per i primi, si deve recuperare il *log odd* relativo alla soglia di piena appartenenza (3) e si deve porre come numeratore. Il denominatore corrisponde alla differenza tra piena appartenenza e *crossover*, espressi nella forma della variabile originale. Il rapporto tra questi due numeri ci dà lo *scalar* per tutti i casi sopra il punto intermedio:

$$Scalar (appartenenza) = 3 / (20.000 - 5.000) = 3 / 15.000 = 0,0002$$

Per i casi sotto al punto medio, si mette al numeratore il *log odd* della soglia di non appartenenza (-3) ed al denominatore la differenza tra soglia di non

appartenenza e punto medio, sempre espressi secondo la metrica della variabile originale:

$$\text{Scalar (non appartenenza)} = -3 / (2.500 - 5.000) = -3 / -2.500 = 0,0012$$

A questo punto, come riportato nella quarta colonna della tabella 4.4, calcoliamo il semplice prodotto della deviazione dal *crossover* per gli *scalars*. Prendendo, ad esempio, i due valori più grandi, abbiamo: per la Svizzera ($35110 * 0,0002$) = 7,022; per il Burundi ($-4890 * 0,0012$) = -5,868. Arrotondati i decimali a due, abbiamo i valori riportati in tabella.

Questo risultato (visibile nella quarta colonna) corrisponde al *log odd* di tutti i casi. Più esattamente, come fa notare Ragin, questi valori corrispondono ai *log odds* di tutti i casi, *poste* le tre ancore scelte dal ricercatore come condizioni della procedura. Non è una semplice ricodifica, è una ricodifica fatta sulla base di quei tre punti fermi:

The values in column 4, in effect, are per capita national income values that have been rescaled into values reflecting the log odds of membership in the set of developed countries, in a manner that strictly conforms to the values attached to the three qualitative anchors--the threshold of full membership, the threshold of full nonmembership, and the cross-over point. Thus, the values in column 4 are not mere mechanistic rescalings of national income, for they reflect the imposition of external criteria via the three qualitative anchors. The use of such external criteria is the hallmark of measurement calibration (Ragin, 2007, p. 15).

L'ultimo passo per individuare i gradi di appartenenza consiste nell'applicazione della formula standard che converte i di *log odds* in valori che rientrano nell'intervallo [0,1]; questa formula è quella che abbiamo visto in precedenza, che si avvale della costante matematica *e*, base di tutti i logaritmi naturali. Otteniamo così i valori riportati nella quinta colonna, raggiungendo lo scopo dichiarato della *calibration*: la trasformazione dei valori di una variabile quantitativa in gradi di appartenenza *fuzzy* rispetto ad un designato *target set*, sulla base dei tre ancoramenti qualitativi scelti dal ricercatore.

Crediamo sia il caso, a scopo sintetico, di ricapitolare la procedura, elencandone punto per punto i passaggi fondamentali:

1. primo requisito: disponibilità di una scala numerica, ad intervallo o di rapporto, come variabile indicatrice del concetto che si vuole analizzare;
2. secondo requisito: determinazione di uno specifico *target set* di casi, in modo da riorganizzare la variabile originale come appartenenza ad tale insieme;
3. scelta delle tre ancore: soglia di piena appartenenza (0,95), punto di *crossover* (0,50), soglia di piena non appartenenza (0,05);
4. calcolo degli *odds* di questi tre valori (attraverso l'utilizzo della prima formula);
5. calcolo dei logaritmi naturali di tali *odds*;
6. calcolo dei due valori di *scalars* (*log odds* di piena appartenenza diviso deviazione della piena appartenenza dal *crossover*; *log odds* di piena non appartenenza diviso deviazione della piena non appartenenza dal *crossover*);
7. calcolo dei *log odds* di tutti i casi (deviazione dal *crossover* per gli *scalars*);
8. trasformazione dei *log odds* in gradi di appartenenza (attraverso l'applicazione della seconda formula).

Questa procedura, che può apparire proibitiva a livello di calcolo matematico, è implementata dal *software* “*fsQCA*”, scaricabile dal sito personale del professor Ragin (cfr. per l'uso di tale programma: Ragin, Drass, e Davies, 2006).

Riportiamo un altro breve esempio, per illustrare l'importante della scelta dei criteri esterni. Vediamo in questo caso una *calibration* che parte dagli stessi dati, ma che utilizza un diverso *target set* di riferimento, ossia quello dei “paesi almeno moderatamente sviluppati”, anziché quello dei “paesi sviluppati”. Dato

che cambia l'insieme di riferimento, devono cambiare anche i tre *breakpoints* qualitativi e la scelta, in questo caso, cade sui seguenti:

Tab. 4.5: Valori di ancoramento per i paesi almeno moderatamente sviluppati

Grado di appartenenza	Reddito pro-capite in USD
0,95	7.500
0,50	2.500
0,05	1.000

La procedura che ne consegue, ovviamente identica alla precedente, porta ai risultati espressi in tabella 4.6:

Tab. 4.6: Calibration dei gradi di appartenenza al set dei Paesi Almeno Moderatamente Sviluppato: Metodo Diretto

Nazione	1. Reddito nazionale	2. Dev. dal crossover	3. Scalars	4. Prodotto di 2 x 3	5. Grado appart.
Svizzera	40110	37610	.0006	22.57	1.00
Stati Uniti	34400	31900	.0006	19.14	1.00
Paesi Bassi	25200	22700	.0006	13.62	1.00
Finlandia	24920	22420	.0006	13.45	1.00
Australia	20060	17560	.0006	10.54	1.00
Israele	17090	14590	.0006	8.75	1.00
Spagna	15320	12820	.0006	7.69	1.00
Nuova Zelanda	13680	11180	.0006	6.71	1.00
Cipro	11720	9220	.0006	5.53	1.00
Grecia	11290	8790	.0006	5.27	.99
Portogallo	10940	8440	.0006	5.06	.99
Repubblica di Corea	9800	7300	.0006	4.38	.99
Argentina	7470	4970	.0006	2.98	.95
Ungheria	4670	2170	.0006	1.30	.79
Venezuela	4100	1600	.0006	.96	.72
Estonia	4070	1570	.0006	.94	.72
Panama	3740	1240	.0006	.74	.68
Mauritius	3690	1190	.0006	.71	.67
Brasile	3590	1090	.0006	.65	.66
Turchia	2980	480	.0006	.29	.57
Bolivia	1000	-1500	.0020	-3.00	.05
Costa d'Avorio	650	-1850	.0020	-3.70	.02
Senegal	450	-2050	.0020	-4.10	.02
Burundi	110	-2390	.0020	-4.78	.01

Così, ad esempio, possiamo notare come la Turchia, con un reddito pro-capite di 2.980 dollari americani abbia un grado di appartenenza di 0,08 all'insieme dei "paesi sviluppati", che la colloca quasi alla soglia di piena non appartenenza, e di 0,57 a quello dei "paesi moderatamente sviluppati", che la

piazza sopra il punto di *crossover*. In generale, possiamo vedere che nella tabella 4.6 ci sono molti più casi che si attestano su valori pari o prossimi a 1 rispetto alla tabella 4.4, coerentemente con la semplice evidenza che debbano esistere più paesi che hanno alta appartenenza al *set* dei paesi “almeno moderatamente” sviluppati, rispetto a quello dei paesi “completamente” sviluppati:

Il contrasto fra le due tabelle sottolinea sia la natura “knowledge-dependent” della *calibration*, sia l’impatto che ha l’applicazione di differenti criteri alle stesse misure:

Again, the key to understanding calibration is to grasp the importance of external criteria, which are based, in turn, on the substantive and theoretical knowledge that researchers bring to their research (Ragin, 2007, p. 16).

Riguardo al metodo indiretto, come detto in precedenza, faremo solo un breve accenno, in quanto non lo utilizzeremo poi nella nostra applicazione pratica. In poche parole, in questo metodo si inizia da un ordinamento qualitativo, operato dal ricercatore, dei casi in esame, assegnati a uno dei sei seguenti gruppi:

Tab. 4.7: Ordinamento qualitativo dei gradi di appartenenza: Metodo Indiretto

Etichetta verbale	Grado di appartenenza
Dentro il <i>target set</i>	1,0
Prevalentemente ma non completamente dentro il <i>target set</i>	0,8
Più dentro che fuori dal <i>target set</i>	0,6
Più fuori che dentro il <i>target set</i>	0,4
Prevalentemente ma non completamente fuori dal <i>target set</i>	0,2
Fuori dal <i>target set</i>	0,0

Si tratta di un “presunto” grado di appartenenza, scelto non in modo arbitrario ma sulla base della conoscenza a disposizione.

Come vediamo nella tabella 4.8, ci sono delle affinità fra questa e la 4.4, come ad esempio il fatto che un reddito di 20.000 USD o più è considerato piena appartenenza o che le nazioni con reddito maggiore di 5.000 USD sono considerate più dentro che fuori dal gruppo:

Tab. 4.8: Calibration dei gradi di appartenenza al set dei Paesi Sviluppati:
Metodo Indiretto

Nazione	1. Reddito nazionale	2. Codifica qualitative	3. Valore previsto
Svizzera	40110	1.00	1.000
Stati Uniti	34400	1.00	1.000
Paesi Bassi	25200	1.00	1.000
Finlandia	24920	1.00	1.000
Australia	20060	1.00	.999
Israele	17090	0.80	.991
Spagna	15320	0.80	.977
Nuova Zelanda	13680	0.80	.991
Cipro	11720	0.80	.887
Grecia	11290	0.80	.868
Portogallo	10940	0.80	.852
Repubblica di Corea	9800	0.60	.793
Argentina	7470	0.60	.653
Ungheria	4670	0.40	.495
Venezuela	4100	0.40	.465
Estonia	4070	0.40	.463
Panama	3740	0.20	.445
Mauritius	3690	0.20	.442
Brasile	3590	0.20	.436
Turchia	2980	0.20	.397
Bolivia	1000	0.00	.053
Costa d'Avorio	650	0.00	.002
Senegal	450	0.00	.000
Burundi	110	0.00	.000

Il passo successivo è usare le colonne 1 (come variabile indipendente) e 2 (come variabile dipendente) per il calcolo del grado di appartenenza di ogni caso riportato nella colonna 3. La tecnica migliore risulta essere l'applicazione di un modello *logit* frazionale, che è implementato nel *software* STATA. Analizzando la tabella, si scorgono delle somiglianze, ma anche delle diversità, ad esempio il notevole distacco tra il grado di appartenenza della Turchia (0,397) ed il paese successivo (la Bolivia: 0,053). Queste differenze, che emergono a dispetto dell'uso degli stessi criteri generali, sono dovuti alla natura indiretta del metodo ed alla sua necessaria dipendenza dalle tecniche di regressione.

Possiamo quindi affermare che, in presenza di criteri esterni a cui confrontarsi direttamente, il metodo diretto si lascia preferire. Nella nostra situazione ancora maggiormente, dato l'elevato numero di casi e la scarsa informazione sul singolo, tipiche dell'analisi individuale.

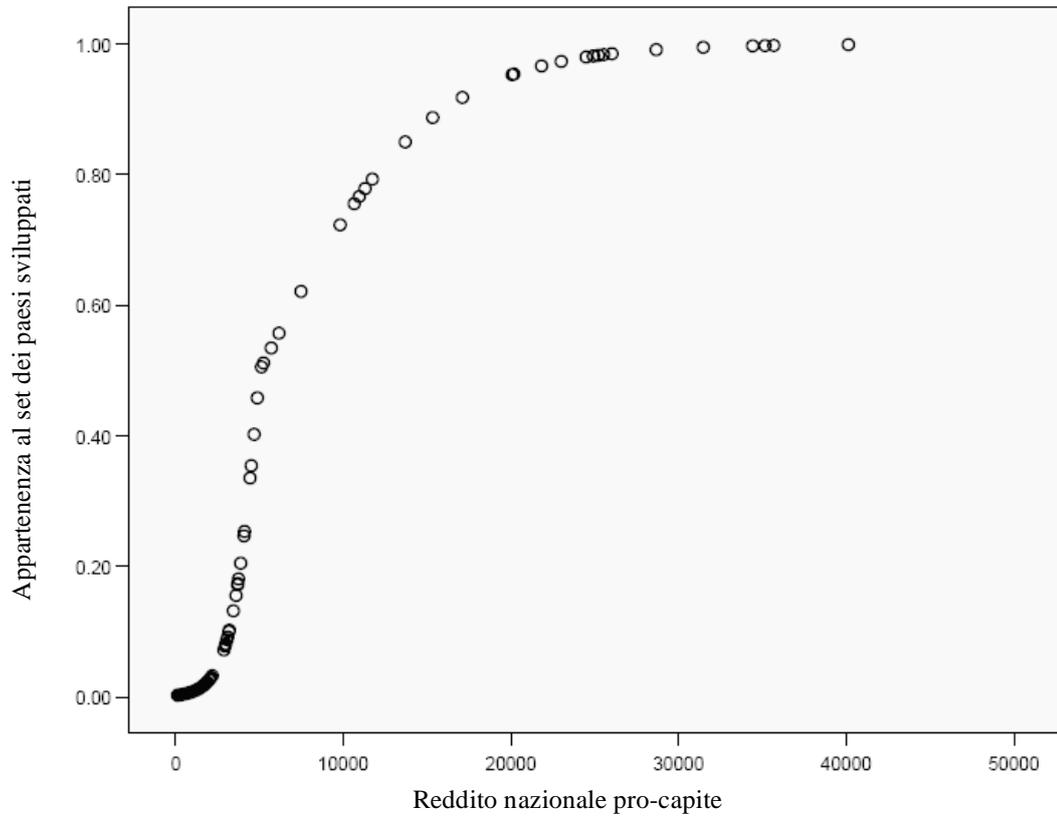
In un'ottica fsQCA, gli esempi che abbiamo visto rappresentano soltanto l'inizio dell'analisi. Più precisamente, rappresentano la trasformazione da variabile quantitativa ad appartenenza *fuzzy* ad un insieme designato, eseguita su una delle possibili condizioni da prendere in esame. Dopodiché si ripete la procedura per un'altra condizione, come ad esempio il livello di democraticità di una nazione, trasformato in appartenenza all'insieme dei paesi democratici, come vediamo nella tabella seguente:

*Tab. 4.9: Calibration dei gradi di appartenenza al set dei Paesi Democratici:
Metodo Diretto*

Nazione	1. Polity score	2. Deviations dal crossover	3. Scalars	4. Prodotto 2*3	5. Grado di appartenenza
Norvegia	10	8.00	0.43	3.43	0.97
Stati Uniti	10	8.00	0.43	3.43	0.97
Francia	9	7.00	0.43	3.00	0.95
Repubblica di Corea	8	6.00	0.43	2.57	0.93
Colombia	7	5.00	0.43	2.14	0.89
Croazia	7	5.00	0.43	2.14	0.89
Bangladesh	6	4.00	0.43	1.71	0.85
Ecuador	6	4.00	0.43	1.71	0.85
Albania	5	3.00	0.43	1.29	0.78
Armenia	5	3.00	0.43	1.29	0.78
Nigeria	4	2.00	0.43	0.86	0.70
Malesia	3	1.00	0.43	0.43	0.61
Cambogia	2	0.00	0.60	0.00	0.50
Tanzania	2	0.00	0.60	0.00	0.50
Zambia	1	-1.00	0.60	-0.60	0.35
Liberia	0	-2.00	0.60	-1.20	0.23
Tagikistan	-1	-3.00	0.60	-1.80	0.14
Giordania	-2	-4.00	0.60	-2.40	0.08
Algeria	-3	-5.00	0.60	-3.00	0.05
Ruanda	-4	-6.00	0.60	-3.60	0.03
Gambia	-5	-7.00	0.60	-4.20	0.01
Egitto	-6	-8.00	0.60	-4.80	0.01
Azerbaijan	-7	-9.00	0.60	-5.40	0.00
Bhutan	-8	-10.00	0.60	-6.00	0.00

Le varie appartenenze dei casi alle rispettive condizioni sono poi esaminate alla ricerca di condizioni che siano o meno necessarie e/o sufficienti. Nella figura 4.1 possiamo, ad esempio, vedere la proiezione del grado di appartenenza all'insieme dei "paesi democratici" dei 136 casi presi in esame sull'asse del reddito nazionale pro-capite in ascissa:

Fig. 4.1: grado di appartenenza al set dei “paesi democratici” vs reddito nazionale pro-capite



Il nostro obiettivo di ricerca si concentra invece sulla possibilità di classificare casi individuali grazie all’uso della logica *fuzzy*. Useremo, dunque, la *calibration* per trasformare le nostre variabili numeriche, dopodichè, grazie all’uso delle *truth table*, inseriremo i casi nei vari gruppi – ogni angolo (ossia ogni cella nello spazio d’attributi) rappresenta un gruppo ed ogni caso sarà assegnato all’angolo a cui si avvicina di più – che scaturiscono dallo spazio d’attributi corrispondente a tali tavole. A questo punto, considereremo il raggruppamento ottenuto come una classificazione a tutti gli effetti (pur se formata con una procedura e con delle tecniche molto diverse) ed in quanto tale la valuteremo.

CAPITOLO 5. UN'APPLICAZIONE EMPIRICA

5.1 PROVENIENZA E CARATTERISTICHE DEI DATI

Il nostro lavoro si conclude con un'applicazione delle tecniche esaminate finora su dati tratti dalla ricerca del 2004 "Lo straniero in immagine", a cura di Maria Stella Agnoli. Si tratta di "una ricerca-intervento sulla figura sociale dello straniero e sulla percezione che di essa manifestano gli studenti di scuola superiore di Roma e delle altre province del Lazio" (Campelli, 2004, p. 9).

L'indagine si presenta come prosecuzione ed ampliamento di quella precedentemente condotta da Roberta Cipollini nel 2002 sugli studenti dell'Istituto Magellano, nel XIII municipio del comune di Roma. La linea di continuità con questo studio è evidenziata soprattutto dal recupero del quadro teorico di riferimento. Il volume di Cipollini, infatti, "con precisione quasi filologica [...] non solo esplora il contributo di autori «classici» come appunto Simmel o Elias, ma introduce il lettore italiano alla conoscenza di analisi relativamente meno note, come quella di Wood o affatto rappresentate nel dibattito, come quella di Znaniecki" (Campelli, 2002, p. X).

Il tema dello straniero ha "impegnato la riflessione dei sociologi fin dagli inizi del '900. A partire da G. Simmel, molti autori come W. Sombart, R. Michels, R.E. Park, F. Znaniecki, M.M. Wood, A. Schutz, N. Elias, R.K. Merton, Z. Bauman, hanno contribuito a definire la rappresentazione sociologica dello straniero" (Cipollini, 2002, p. 3), al punto che tale tema "può considerarsi addirittura l'oggetto di una sorta di specializzazione disciplinare, la *sociologia dello straniero*" (Agnoli, 2004, p. 11), i cui contributi sono caratterizzati da uno sviluppo progressivo in grado di mantenere una certa identità teorica e concettuale di base (cfr. Tabboni, 1986; Colombo, 1999; Cipollini, 2000, 2002).

Nell'indagine coordinata da Agnoli, l'ottica che ha guidato il lavoro di ricerca si identifica con la "specificazione determinata degli «stranieri immigrati», un *oggetto collettivo*" (Agnoli, 2004, p. 11) caratterizzato da forte definizione contestuale. La scelta di tale espressione (stranieri immigrati) è stata fatta allo scopo di "arginare i rischi di eccessiva genericità ed astrattezza insiti nell'uso del solo termine *straniero*, così esteso da poter essere riferito all'Altro nella sua generalità indeterminata" (Agnoli, 2004, p. 18), che comprende turisti, stranieri in transito, sportivi, diplomatici e via dicendo. Si è voluto, quindi fornire un "riferimento non equivoco all'oggetto di valutazione, specificandolo in coloro che da altre parti del mondo si sono trasferiti nel nostro paese in cerca di migliori condizioni di vita" (Agnoli, 2004, p. 18). Per lo stesso motivo si è ritenuto di non poter sostituire il termine *immigrato*, "pur consapevoli dei connotati negativi progressivamente assunti" (Agnoli, 2004, p. 18), con quello *migrante*, considerato più efficace nel dar conto dello spostamento transnazionale di considerevoli quote di popolazione da un territorio ad un altro (cfr. Ferrarotti, 2000).

Oltre al quadro teorico, la continuità tra le ricerche del 2002 e del 2004 è rintracciabile anche a livello operativo, con la parziale sovrapposizione dello strumento di rilevazione, che nel secondo lavoro è stato affinato per complessità ed articolazione. Ne è scaturito "un questionario ad elevato grado di strutturazione [...], ampiamente collaudato, nelle parti omologhe, [...] e invece sottoposto a rigoroso *pretesting* nelle parti originali" (Agnoli, 2004, p. 23).

Tale collaudo, che ha investito, tra le altre, le parti relative alle componenti affettiva e attiva dell'atteggiamento, diretto da Sergio Mauceri, si è basato sulla procedura dell'*intervista sull'intervista* (cfr. Mauceri, 2003) ed è stato realizzato su un campione di 108 studenti di scuola secondaria superiore.

Ciascuna provincia del Lazio è stata sottoposta ad una preliminare analisi contestuale, volta a stimare la consistenza della popolazione studentesca, della presenza degli stranieri immigrati, degli alunni stranieri nelle scuole, l'andamento del fenomeno migratorio, le caratteristiche del tessuto lavorativo e la penetrazione di lavoratori stranieri in tale tessuto. Al momento dell'indagine, la Regione Lazio è la prima in Italia per numero di residenti stranieri immigrati in rapporto alla popolazione (quasi 5 stranieri ogni 100 italiani) e la seconda regione

italiana per numero totale (dietro alla Lombardia), con oltre 150 nazionalità rappresentate (cfr. Caritas, 2003). Dall'analisi contestuale emerge, ovviamente, la peculiarità della città di Roma: su 10 stranieri soggiornanti nel Lazio, 9 vivono a Roma, così come a Roma vive il 73% della popolazione del Lazio; il 6% della popolazione romana è dato da stranieri residenti, con picchi del 12% nel XVIII municipio; 170 diverse nazionalità sono rappresentate, di cui 144 nelle scuole. Queste considerazioni hanno spinto ad una trattazione centrata sul solo capoluogo di provincia (come è stato anche per le altre provincie del Lazio) e su un'esposizione separata dei risultati. Tutto ciò è in linea con l'impostazione teorica per cui "si assume che alla varietà dei contesti socio-economici-demografici si associno specifici sistemi di relazione, influenti sulle stesse strategie/regole di contatto con lo straniero" (Agnoli, 2004, p. 24).

Per il campionamento è stata adottata una strategia ragiona a multi-stadio. Per quanto riguarda il comune di Roma, nel primo stadio si è proceduto ad una selezione di 8 dei 19 municipi che compongono la città, che fossero fra loro diversi per estensione e per presenza di immigrati, sulla base di elaborazioni di dati forniti dall'Ufficio Statistico del comune di Roma; nel secondo stadio, sono state selezionate, per ogni municipio, due scuole di diverso orientamento formativo, un istituto liceale e un istituto tecnico-professionale; infine, nel terzo stadio, è stato selezionato, in ognuna delle 16 scuole, un intero ciclo di istruzione superiore, costituito da cinque classi appartenenti a diverse sezioni. In questo modo la rilevazione è stata condotta su un totale di 80 classi e 1443 intervistati.

Per quanto riguarda gli altri capoluoghi laziali, sono state scelte due scuole di formazione liceale e due di formazione tecnico-professionale per ogni capoluogo. Lo stadio successivo, identico al terzo romano, ha previsto la scelta, lasciata in mano ai presidi, di un intero ciclo formativo, possibilmente di sezioni diverse. Così, anche nel contesto non metropolitano, sono state sottoposte ad analisi 80 classi, corrispondenti in questo caso a 1520 studenti.

La somministrazione del questionario è stata effettuata tramite rivelazione di gruppo, mediante intervista personale indiretta. Prima di distribuire il questionario, "gli intervistatori-ricercatori presenti in aula hanno illustrato il complessivo quadro della ricerca e le modalità tecniche di compilazione di alcune

domande particolarmente complesse; hanno poi assistito alla somministrazione e raccolto i questionari a rilevazione conclusa” (Agnoli, 2004, p. 23). Tale modalità di somministrazione ha permesso di ridurre i limiti classici dei questionari auto-compilati, cioè la caduta delle risposte e l’incompletezza della compilazione. La caduta delle risposte, in particolare, è risultata molto limitata.

Le variabili su cui lavoreremo sono le risposte date a tre scale Likert (cfr. Likert, 1932), costruite per indagare il pregiudizio etnico, “inteso come costruito avente una struttura complessa, articolata in dimensioni analiticamente distinguibili in cognitiva, affettiva e conativo/attiva, (cfr. Allport, 1954), ancorché non sempre e non necessariamente coerente” (Agnoli, 2004, p. 21).

In questa ricerca, come nella precedente di Cipollini e “contrariamente alla tendenza prevalente a sovrapporre indifferenziatamente i due momenti di analisi” (Campelli, 2002, p. X), si è dunque scelto di separare la componente cognitiva – relativa a credenze, conoscenze e stereotipi, “tratti della rappresentazione sociale dello straniero definiti dalla teoria sociologica, integrati con quelli più diffusi nell’opinione pubblica” (Agnoli, 2004, p. 21) – da quella attiva, relativa all’orientamento all’azione normativa verso gli stranieri; si è deciso anche di aggiungere la componente affettiva, assente nella ricerca precedente, del pregiudizio: dunque, “*sentimenti e reazioni emotive*, unitamente all’*orientamento all’azione* nei loro confronti [nei confronti degli stranieri immigrati] sono stati specificatamente riferiti a dimensioni dell’atteggiamento concernenti il piano dell’incontro interetnico” (Agnoli, 2004, p. 21).

Il supposto *continuum* delle proprietà indagate dalle tre scale Likert considerate è stato diviso in sei intervalli, onde evitare l’attrazione del medio. La scala relativa alla componente cognitiva (domanda 16 del questionario, che più avanti vedremo) si presenta come una classica batteria di items, 25 in questo caso, verso i quali si può manifestare un accordo o disaccordo sfumato nelle tre categorie: “del tutto”, “prevalentemente”, “parzialmente”.

Le scale relative alle componenti affettiva ed attiva sono invece costruite sulla forma del differenziale semantico (cfr. Osgood, 1952). Agli estremi, però, anziché aggettivi opposti, troviamo le seguenti espressioni: “mi darebbe fastidio/non mi darebbe fastidio”, “mi interesserebbe/non mi interesserebbe”, “mi

farebbe piacere/non mi farebbe piacere” per quanto riguarda la componente affettiva; “approverei/disapproverei” per la componente attiva. All’interno, la proprietà è stata scomposta in sei caselle, sempre per scoraggiare l’attrazione del medio.

In tutti e tre i casi, la polarità degli items è stata invertita e sono stati alternati items positivi ed items negativi in modo non sistematico, come tentativo di evitare il fenomeno del *response set*. Una volta riorientati nella direzione di un aumento del pregiudizio, gli items saranno sottoposti a riduzione fattoriale attraverso la tecnica dell’analisi delle componenti principali. Le variabili ottenute saranno così considerate e trattate convenzionalmente come variabili *quasi-cardinali*, senza dimenticare però che è illusorio pensare che gli intervalli in cui è scomposta la proprietà siano equidistanti.

È, infatti, necessario ricordare i problemi a cui l’uso della scala Likert va incontro, benché una problematizzazione adeguata di questi esuli dall’interesse cognitivo di questo lavoro – che concentrandosi sulle procedure e tecniche di classificazione, in un certo senso ha *soltanto* bisogno di una stessa base empirica su cui lavorare – e per certi aspetti dalle nostre possibilità, dato che interveniamo a ricerca ampiamente conclusa. Esiste una vasta letteratura in merito, a cui rimandiamo per ulteriori approfondimenti, ci limitiamo a citare alcuni capisaldi: Pitrone, 1984; Marradi, 2002; Marradi e Gasperoni, a c. di, 2002; Pavsic e Pitrone, 2003.

In questa sede, oltre ai già menzionati *response set* e non equidistanza fra gli intervalli, citiamo, prima di tutto, la curvilinearità e la reazione all’oggetto. La prima capita quando un’affermazione è approvata o disapprovata per motivi molto diversi, così che si assegnano medesimi punteggi a opinioni collocate su posizioni molto distanti: in questo modo, il supposto *continuum*, identificato geometricamente da un segmento, si piega metaforicamente su se stesso.

La reazione all’oggetto è invece un fenomeno che capita quando alcuni intervistati non reagiscono all’intero significato della frase, ma ad un singolo termine, più precisamente all’oggetto/soggetto designato da quel termine. Ad esempio, uno studente di psicologia si dichiarò completamente in disaccordo con la frase “I politici si interessano solo al voto, non ai bisogni degli elettori”,

commentando ad alta voce: “Dovrebbero interessarsi anche ai bisogni, non solo al voto” (cfr. Marradi, 2007, p. 148). Egli pertanto avrebbe dovuto esprimere completo accordo, ma concentrandosi solo sull’oggetto della frase e non sull’intera frase stessa, reagì nel modo opposto a quello che avrebbe dovuto. Marradi ha stimato in un 5-10% l’incidenza della reazione all’oggetto sui risultati di una scala Likert.

Un altro noto problema è quello della distorsione delle distribuzioni di frequenza, secondo cui le approvazioni superano sempre le disapprovazioni, in quanto “dichiararsi d’accordo con qualunque affermazione richiede un impegno psichico molto minore che dichiararsi in disaccordo” (Marradi, 2007, p. 147; cfr. Jordan, 1965; Galtung, 1967; Amisano, Rinaldi e Pampanin, 2002). Vale la pena notare comunque che, con i dati che abbiamo a disposizione, una semplice analisi delle distribuzioni di frequenza degli items non ancora ricodificati sembra sfuggire a quest’assunto. Se consideriamo la domanda 16 (componente cognitiva), notiamo come, per il contesto romano, in 14 casi su 25, la percentuale cumulata delle prime tre opzioni “del tutto d’accordo”, “prevalentemente d’accordo” e “parzialmente d’accordo” risulta essere inferiore al 50%. Nel contesto laziale, la stessa cosa si verifica 13 volte su 25. Quello che speriamo e crediamo è che in questo caso il ruolo degli intervistatori, che coincideva con quello dei ricercatori – con tutto ciò che questo comporta in termini di preparazione ed interesse – abbia esercitato una attiva influenza benefica sulla qualità dei dati, andando ad interrompere quel “gioco a somma positiva” di interessi privati, in cui l’unica vittima è la fedeltà dei dati (Marradi, 2007, p. 151).

L’ultima precisazione che dobbiamo fare prima di illustrare i risultati della nostra applicazione riguarda la fase di controllo delle scale Likert, nella quale si accertano coerenza interna ed unidimensionalità della scala (cfr. Corbetta, 1999). La coerenza, così come era successo nella ricerca di Cipollini del 2002, è stata analizzata attraverso la procedura della correlazione «corretta» elemento-scala e attraverso il calcolo dell’alfa di Cronbach. Le scale sono risultate molto coerenti (cfr. Nota metodologica in Cipollini, a c. di, 2002; Nota metodologica in Agnoli, a c. di, 2004), queste analisi non verranno pertanto ripetute in questa sede. L’unidimensionalità è stata, invece, testata attraverso l’analisi delle componenti

principali. Le scale sono risultate unidimensionali, in quanto tutti gli items concorrono alla costruzione della prima componente in modo coerente e con punteggi significativi. Sono anche emerse alcune dimensioni latenti che vedremo di volta in volta.

Possiamo ora procedere con l'analisi. Prenderemo in considerazione una scala alla volta, a partire dalla componente cognitiva (domanda 16 del questionario), passando per quella affettiva (domanda 19), per finire con quella attiva (domanda 20). Analizzeremo il contesto romano per primo e laziale per secondo, in questo modo avremo a che fare, di fatto, con due matrici.

Il primo passo sarà l'analisi delle componenti principali, dopodiché costruiremo diverse tipologie attraverso l'uso della *cluster analysis* tradizionale, della *cluster analysis* semi-sfocata e della *calibration*. Queste verranno valutate in base alla loro stabilità, attraverso semplici tabelle di contingenza, ed alla perdita di informazioni rispetto ai dati di partenza, attraverso l'indice dell'eta quadro¹. Questi, come abbiamo riportato nel secondo capitolo, sono rispettivamente il secondo ed il terzo criterio di valutazione che Biorcio (1993) individua, segnalandone al contempo la sistematica omissione nella maggior parte dei casi pratici di ricerca. Ove possibile, non mancheranno considerazioni di tipo non statistico. A causa della considerevole quantità di dati (abbiamo costruito in totale 36 diverse tipologie), prediligeremo l'uso di tavole sintetiche, in modo da non appesantire esageratamente la lettura.

Per la nostra analisi abbiamo utilizzato diversi *software*: SPSS per l'analisi tradizionale; un programma sviluppato da Francesco Middei e che ci è stato fornito dalla ricercatrice Chiara Brunelli per l'analisi semi-sfocata; il programma fsQCA, sviluppato da Charles Ragin e scaricabile dal suo sito personale, per la *calibration*. Per l'esportazione e l'importazione dei dati, ci siamo serviti del *software* Microsoft Excel. Per alcune elaborazioni grafiche tridimensionali, abbiamo utilizzato l'applicazione Grapher, di proprietà Apple, fornita negli accessori del sistema operativo Mac OS Tiger e Leopard.

¹ Vedi paragrafo 2.5.1.

**5.2 LA COMPONENTE COGNITIVA DEL PREGIUDIZIO ETNICO: IL
CONTESTO DI ROMA**

Come detto in precedenza, la componente cognitiva del pregiudizio etnico è stata operativizzata nella forma di una scala Likert a 6 intervalli, corrispondente alla sedicesima domanda del questionario, che riportiamo a seguire:

Tab. 5.1: scala Likert – componente cognitiva del pregiudizio etnico

16. Ti sottoponiamo una serie di affermazioni comuni a proposito degli stranieri immigrati. Per ciascuna di esse indica il tuo grado di accordo o di disaccordo.		ACCORDO			DISACCORDO		
		Del tutto Prevalentemente	Parzialmente		Del tutto Prevalentemente	Parzialmente	
a.	Gli stranieri immigrati apportano nuovi stimoli culturali	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b.	Gli stranieri immigrati hanno spesso comportamenti invadenti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c.	Gli stranieri immigrati contribuiscono allo sviluppo culturale del paese	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d.	Gli stranieri immigrati sono troppo legati alle proprie tradizioni	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e.	Gli stranieri immigrati badano solo ai propri interessi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f.	Gli stranieri immigrati sono spesso persone originali	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g.	Molto spesso gli stranieri immigrati non sono persone pulite	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h.	Gli stranieri immigrati introducono nel paese ideologie pericolose	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i.	Generalmente gli stranieri immigrati hanno poca voglia di lavorare	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j.	Gli stranieri immigrati introducono comportamenti sessuali troppo liberi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
k.	Gli stranieri immigrati sottraggono lavoro agli italiani	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
l.	Gli stranieri immigrati hanno troppo figli	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
m.	Gli stranieri immigrati spesso disturbano con il loro comportamento	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
n.	Gli stranieri immigrati generalmente non rispettano le regole della convivenza civile	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
o.	Gli stranieri immigrati introducono malattie prima inesistenti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
p.	Gli stranieri immigrati usufruiscono abusivamente del nostro sistema di assistenza	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
q.	Gli stranieri immigrati aggravano i problemi di ordine pubblico	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
r.	Gli stranieri immigrati non sanno stare al loro posto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
s.	Gli stranieri immigrati contribuiscono allo sviluppo economico del paese	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
t.	Con la loro presenza, gli stranieri immigrati degradano il quartiere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
u.	Gli stranieri immigrati introducono nuove mode	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
v.	Gli stranieri immigrati diffondono comportamenti in contrasto con le nostre tradizioni	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
w.	Gli stranieri immigrati rendono più difficile la ricerca di abitazioni in affitto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
x.	Accettando lavori non regolamentati, gli stranieri immigrati, di fatto, contribuiscono alla diffusione del lavoro non regolamentato anche per gli italiani	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
y.	Gli stranieri immigrati spesso sono troppo critici nei confronti della società italiana	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

La codifica per tutti gli items, indipendente dalla loro polarità, è espressa sotto forma di un crescente disaccordo, quindi:

- 1) Del tutto d'accordo
- 2) Prevalentemente d'accordo
- 3) Parzialmente d'accordo
- 4) Parzialmente in disaccordo
- 5) Prevalentemente in disaccordo
- 6) Del tutto in disaccordo
- 7) Parzialmente in disaccordo + parzialmente d'accordo

Ciò evidentemente significa che nel caso di item “negativo” (items *b, d, e, g, h, i, j, k, l, m, n, o, p, q, r, t, v, w, x, y*; es. mancanza di igiene) il totale accordo, quindi il punteggio 1, esprime l’adesione ad uno stereotipo negativo dello straniero, e quindi la presenza di pregiudizio contro di esso. Il punteggio 6, cioè il totale disaccordo, esprime mancanza di pregiudizio verso l’immigrato.

Negli items “positivi” (items *a, c, f, s, u*; es. contributo allo sviluppo economico del paese) il discorso è naturalmente inverso: il punteggio 1 (totale accordo) esprime stereotipo positivo, il punteggio 6 (totale disaccordo) esprime stereotipo negativo.

Il valore “7” è stato assegnato a quei pochissimi intervistati che hanno aggirato l’*escamotage* anti-attrazione del medio indicando contemporaneamente il parziale accordo ed il parziale disaccordo.

La prima cosa da fare, dunque, era modificare la codifica degli items e ciò è stato fatto, seguendo l’esempio della ricerca, riorientando quelli negativi:

da	1	2	3	4	5	6
a	6	5	4	3	2	1

Abbiamo ricodificato le risposte “7” come “mancanti di sistema”, considerando la loro esiguità, il fatto che gli intervistati si erano adoperati per non propendere per nessuna delle due polarità e per mantenere un *range* fisso da un minimo di 1 a un massimo di 6 per l’intera batteria.

Dunque, come nella ricerca, avevamo 25 items varianti tra i punteggi 1 e 6, dove 1 rappresentava il minimo pregiudizio etnico (o stereotipo positivo dello straniero) e 6 il massimo (o stereotipo negativo).

A questo punto, abbiamo trattato le variabili qualitative della scala Likert come variabili quasi-cardinali, ed abbiamo effettuato, con il *software* SPSS, analisi delle componenti principali. I valori mancanti sono stati sostituiti con la media del campione, visto che lavoreremo con l’algoritmo delle k-medie semisfocato, che, come abbiamo spiegato precedentemente, essendo un algoritmo che ottimizza una funzione obiettivo e parte dalla matrice dei dati, si trova in difficoltà quando questa è incompleta.

Questa è la lista delle componenti estratte e della relativa varianza spiegata:

Tab. 5.2: Componenti estratte e varianza spiegata – Roma, d16

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9,98	39,94	39,94	9,98	39,94	39,94
2	1,61	6,44	46,38	1,61	6,44	46,38
3	1,09	4,37	50,75	1,09	4,37	50,75
4	0,98	3,92	54,67			
5	0,84	3,36	58,03			
6	0,82	3,28	61,31			
7	0,77	3,10	64,41			
8	0,72	2,86	67,27			
9	0,69	2,76	70,02			
10	0,64	2,57	72,59			
11	0,62	2,47	75,06			
12	0,59	2,38	77,44			
13	0,55	2,22	79,66			
14	0,54	2,14	81,80			
15	0,52	2,06	83,86			
16	0,50	2,02	85,88			
17	0,49	1,96	87,84			
18	0,47	1,89	89,73			
19	0,44	1,74	91,47			
20	0,42	1,67	93,14			
21	0,39	1,58	94,72			
22	0,38	1,53	96,25			
23	0,34	1,35	97,60			
24	0,32	1,29	98,89			
25	0,28	1,11	100,00			

Seguendo il metodo di estrazione per autovalore > 1 (cfr. Kaiser, 1960), sono state estratte le prime tre componenti, analogamente a quanto effettuato nella ricerca del 2004 a cura di Agnoli.

È il momento di fare una precisazione. I dati a nostra disposizione sono tratti dalla ricerca di Agnoli, ma non sono esattamente quelli elaborati nel capitolo redatto da Cipolli, Cuppone e Giannone, relativo alla componente cognitiva del pregiudizio etnico, che è stato pubblicato. O meglio, i dati sono gli stessi, ma è diversa la matrice. Quella in nostro possesso è una matrice in SPSS grezza, i cui dati sono privi di qualsiasi trattamento. Per questo motivo la nostra analisi parte fin dall'orientamento degli items. Durante l'originario lavoro di ricerca, l'elaborazione è stata eseguita con il programma SPAD. Questo, come abbiamo anche visto in nostre precedenti analisi, porta ad alcune differenze, ad esempio nelle distribuzioni di frequenza delle analisi dei gruppi, dove si sono registrate variazioni anche considerevoli, mentre l'analisi fattoriale è risultata sempre molto stabile. Tutto ciò ha anche comportato il vantaggio di avere il controllo totale sull'intero processo di elaborazione statistica. Ad ogni modo, questo ci impone una limitazione: non avendo i dati originari, non potremo fare alcun confronto in merito alla stabilità delle nostre tipologie con quella pubblicata o in merito all'efficacia, visto che l'età quadro della tipologia pubblicata non è di nostra conoscenza. Per questo motivo, il nostro lavoro sarà dipendente da quello originale per quanto riguarda l'interpretazione dei risultati, dei fattori, l'etichettamento dei gruppi e così via, ma sarà indipendente a livello statistico. Ultima annotazione: nella ricerca sono presenti 1443 casi sul campione romano, mentre nella nostra matrice i casi sono 1442. Evidentemente, nel corso dei vari passaggi di matrice in matrice, un caso è stato erroneamente o volontariamente eliminato, non ci è dato saperlo.

Il primo fattore spiega una quota di varianza pari al 39,94% con autovalore di 9,98; il secondo spiega il 6,44% con autovalore pari a 1,61; il terzo spiega il 4,37% della varianza totale ed ha un autovalore di 1,09. La varianza cumulata spiegata dalle tre componenti è pari al 50,75%. A conferma di quanto diceva sulla stabilità di questa tecnica, i risultati ottenuti con SPAD e pubblicati sono assolutamente identici. Vediamo ora la correlazione tra componenti e variabili.

Nella ricerca sono stati considerate correlazioni significative solo quelle uguali o superiori al valore assoluto di 0,30; noi abbiamo adottato la stessa scelta.

Tab. 5.3: Correlazione componenti-variabili – Roma, d16

Items	Correlazioni variabili-fattori		
	1	2	3
Apporto di nuovi stimoli culturali	0,63	0,41	
Comportamenti invadenti	0,66		
Contributo allo sviluppo culturale del paese	0,64	0,46	
Legame eccessivo alle tradizioni	0,33		0,73
Interessi personali	0,65		0,33
Originalità	0,34	0,48	
Mancanza di igiene	0,68		
Introduzione ideologie pericolose	0,73		
Scarsa voglia di lavorare	0,60		
Comportamenti sessuali troppo liberi	0,60		
Sottrazione lavoro agli italiani	0,64		
Numero eccessivo figli	0,58		
Comportamenti di disturbo	0,77		
Mancato rispetto regole di convivenza civile	0,77		
Introduzione nuove malattie	0,62		
Fruizione abusiva sistema di assistenza	0,74		
Problemi di ordine pubblico	0,74		
Non saper stare al proprio posto	0,79		
Contributo allo sviluppo economico del paese	0,59	0,42	
Degrado quartiere	0,74		
Introduzione nuove mode	0,43	0,59	
Diffusione comportamenti contrari tradizione	0,55		
Aumento difficoltà in ricerca abitazioni in affitto	0,57	-0,31	
Diffusione lavoro irregolare	0,52		
Atteggiamento troppo critico verso società italiana	0,58		

I valori di correlazione relativi alla prima componente confermano quanto detto riguardo all'unidimensionalità della scala. Tutti gli items concorrono positivamente alla formazione della prima componente, con valori in certi casi anche piuttosto elevati: “emerge quindi una conferma dell'evidenza empirica riscontrata nella precedente utilizzazione della scala tra gli adolescenti del XIII municipio di Roma, che a sua volta, conferma l'ipotesi alla base della sua costruzione: la struttura prevalente del pregiudizio nella sua componente cognitiva è costituita da un insieme coordinato e coerente di immagini dello straniero in cui convivono tutte le dimensioni della sua rappresentazione sociale” (Cipollini, 2004, p. 107).

Oltre alla dimensione principale, emergono dalla riduzione fattoriale anche due dimensioni latenti del pregiudizio. La terza componente è correlata positivamente a due soli items, la seconda è correlata incoerentemente a sei items: positivamente alle cinque variabili che indicano mutamento socio-economico-culturale di cui sarebbe investito il ruolo dello straniero, negativamente l'item riguardante l'aumento della difficoltà nella ricerca di abitazioni in affitto.

Tab. 5.4: Interpretazione della prima componente – Roma, d16

ROMA	Orientamento	Semiassse neg	Semiassse pos
Apporto di nuovi stimoli culturali	+	Accordo	Disaccordo
Comportamenti invadenti	-	Disaccordo	Accordo
Contributo allo sviluppo culturale del paese	+	Accordo	Disaccordo
Legame eccessivo alle tradizioni	-	Disaccordo	Accordo
Interessi personali	-	Disaccordo	Accordo
Originalità	+	Accordo	Disaccordo
Mancanza di igiene	-	Disaccordo	Accordo
Introduzione ideologie pericolose	-	Disaccordo	Accordo
Scarsa voglia di lavorare	-	Disaccordo	Accordo
Comportamenti sessuali troppo liberi	-	Disaccordo	Accordo
Sottrazione lavoro agli italiani	-	Disaccordo	Accordo
Numero eccessivo figli	-	Disaccordo	Accordo
Comportamenti di disturbo	-	Disaccordo	Accordo
Mancato rispetto regole di convivenza civile	-	Disaccordo	Accordo
Introduzione nuove malattie	-	Disaccordo	Accordo
Fruizione abusiva sistema di assistenza	-	Disaccordo	Accordo
Problemi di ordine pubblico	-	Disaccordo	Accordo
Non saper stare al proprio posto	-	Disaccordo	Accordo
Contributo allo sviluppo economico del paese	+	Accordo	Disaccordo
Degrado quartiere	-	Disaccordo	Accordo
Introduzione nuove mode	+	Accordo	Disaccordo
Diffusione comportamenti contrari tradizione	-	Disaccordo	Accordo
Aumento difficoltà in ricerca abitazioni in affitto	-	Disaccordo	Accordo
Diffusione lavoro irregolare	-	Disaccordo	Accordo
Atteggiamento troppo critico verso società italiana	-	Disaccordo	Accordo

Per aiutarci nella comprensione delle componenti (e per avere ulteriore conferma dell'unidimensionalità della prima) è conveniente sostituire ai valori di correlazione l'espressione "accordo/disaccordo", ricordando che gli items hanno polarità diversa e quelli negativi sono stati orientati².

² In caso di correlazione positiva: per gli items positivi, all'aumento dei valori corrisponde aumento del disaccordo; per gli items negativi, all'aumento dei valori corrisponde aumento

L'osservazione della tabella 5.4 conferma l'interpretazione data da Cipollini. All'aumento del punteggio della componente, aumentano l'accordo verso gli items che esprimono i tradizionali pregiudizi negativi di cui è investita la figura dello straniero immigrato; contemporaneamente aumenta il disaccordo verso gli items che investono lo straniero di caratteristiche positive, come l'originalità o il contributo economico e culturale al paese. Dunque, "la rappresentazione sembra strutturarsi in riferimento alle caratteristiche ambivalenti della figura sociale dello straniero, ritenuto soggetto minaccioso, perturbante, deviante da coloro che registrano un'alta adesione al pensiero per stereotipi e, in alternativa, come soggetto in grado di stimolare il mutamento sociale tra coloro che si distaccando da tale immagine ancorata ad una rappresentazione negativa" (Cipollini, 2004, p. 107). La categorizzazione segue la forma descritta da Moscovici, per cui nelle classificazioni di cose non familiari, quello che entra maggiormente in gioco è il bisogno di definirle come normali, conformi o divergenti e aberranti (Moscovici, 1984). Visto che la componente verte sulle caratteristiche ambivalenti della figura dello straniero, è stata essa stessa definita "**Ambivalenza**". Vediamo ora la seconda componente, nella tabella 5.5:

Tab. 5.5: Interpretazione della seconda componente – Roma, d16

ROMA	Orientamento	Semiassse neg	Semiassse pos
Apporto di nuovi stimoli culturali	+	Accordo	Disaccordo
Contributo allo sviluppo culturale del paese	+	Accordo	Disaccordo
Originalità	+	Accordo	Disaccordo
Contributo allo sviluppo economico del paese	+	Accordo	Disaccordo
Introduzione nuove mode	+	Accordo	Disaccordo
Aumento difficoltà in ricerca abitazioni in affitto	-	Accordo	Disaccordo

All'aumento della componente, corrisponde un aumento del disaccordo sia nei confronti degli items positivi che indicano mutamento socio-culturale, sia di quello negativo che imputa agli stranieri immigrati una maggiore difficoltà nella ricerca di case in affitto. Questa dimensione sembra essere in un certo senso tipica del contesto metropolitano, nel momento in cui si sottostima il cambiamento

dell'accordo. In caso di correlazione negativa: per gli items positivi, all'aumento dei valori corrisponde aumento dell'accordo; per gli items negativi, all'aumento dei valori corrisponde aumento del disaccordo.

apportato dallo straniero, se ne segnala contemporaneamente la non percezione “come soggetto conflittuale, almeno sul piano della loro possibile contesa di un bene prezioso e tipico del vivere in città: la disponibilità di abitazioni” (Cipollini, 2004, p. 108). Tale componente, denominata “**Distanza cognitiva**”, fa riferimento ad “un tratto tipico dell’abitante della metropoli che vive una vita di «stranieri fra stranieri»” (Cipollini, 2004, p. 109; cfr. Wood, 1934; Lofland, 1973; Bauman, 1999). Non vengono messe a fuoco le caratteristiche dei soggetti che agiscono sullo sfondo della metropoli, si tratta di “presenze constatate, forse tollerate”, rispetto alle quali mancano riflessione e coinvolgimento e prevale “un senso di distacco, se non di indifferenza” (Cipollini, 2004, p. 109). Il semiasse negativo è segnato dall’atteggiamento opposto, nel quale convivono curiosità, riconoscimento del ruolo innovativo ma anche timore e competizione, in linea con le teorizzazioni di Simmel (cfr. Simmel, 1908), per cui la figura dello straniero è sinonimo di vicinanza e lontananza, inserimento ed emarginazione, che genera sentimenti di coinvolgimento capaci di suscitare allo stesso tempo di interesse e rifiuto.

L’ultima componente è stata denominata “**Diversità con orientamento utilitaristico**”.

Tab. 5.6: Interpretazione della terza componente – Roma, d16

ROMA	Orientamento	C3, semiasse neg	C3, semiasse pos
Legame eccessivo alle tradizioni	–	Disaccordo	Accordo
Interessi personali	–	Disaccordo	Accordo

Essa è correlata con soli due items, al crescere della componente cresce l’accordo verso una visione che associa un legame eccessivo alle tradizioni d’origine dello straniero immigrato ed una concezione dello stesso come troppo legati ai suoi particolari interessi personali. Lo straniero non è solo percepito come diverso, ma come diverso caratterizzato da interessi competitivi rispetto a quelli della comunità. Sulla sponda opposta, questa percezione scompare.

5.2.1 CLUSTER ANALYSIS TRADIZIONALE E SEMISFOCATA

Alle tre componenti estratte è stata applicata una *cluster analysis*, con metodo delle k-medie. È stata imposta una divisione a 4 gruppi, così come 4 erano quelli della ricerca³. Il punto di equilibrio è stato raggiunto dopo 12 iterazioni, vediamo i centri finali dei gruppi:

Tab. 5.7: Centri finali dei gruppi – Roma, k-medie, d16

	1	2	3	4
Ambivalenza	0,21	1,24	-0,98	-0,14
Distanza cognitiva	0,76	0,04	0,35	-1,03
Diversità con orientamento utilitaristico	0,89	-0,77	-0,66	0,36

La disposizione dei centri sulle 3 componenti (specialmente sulla prima) e il calcolo dell'Anova sugli items hanno permesso di etichettare i gruppi come "Tolleranti" (3), "Tolleranti con riserve" (4), "Conflittuali moderati" (1) e "Conflittuali" (2). Vediamo brevemente l'analisi della varianza sui singoli gruppi.

Nei **tolleranti** (tabella 5.8) vediamo come i valori medi del gruppo siano costantemente inferiori a quelli del campione, anche se con delle differenze. Di pari passo seguono i valori della deviazione standard. Gli items relativi al cambiamento culturale sono quelli su cui il gruppo è più eterogeneo, arrivando, di fatto, a pareggiare i valori medi del campione. Per quanto riguarda la definizione sociologica di questo gruppo, Cipollini segnala come siano risultate discriminanti otto variabili illustrative: "si tratta di adolescenti che prevalentemente si collocano nell'orientamento politico di «sinistra, centro-sinistra», impegnati sul terreno politico e sociale, il cui sistema di appartenenza è allargato al mondo e all'Europa e quindi lontani da confini di tipo nazionalista e localista, appartenenti ai livelli più alti della stratificazione sociale e le cui famiglie presentano anche alti livelli di *status* socio-culturale" (Cipollini, 2004, p. 112). In più, sono presenti: sottostima della propria posizione di status, ateismo o disinteresse rispetto alla sfera religiosa e fruizione di molteplici fonti di informazione e comunicazione.

³ La partizione ottimale a 4 gruppi era scaturita da osservazione del dendrogramma, era stata quindi utilizzata una tecnica gerarchica.

Il secondo gruppo, quello dei **tolleranti con riserve**, si distacca in modo meno significativi dai punteggi del campione, come ci è possibile vedere anche attraverso il solo valore medio sull'intera batteria, di quasi un punto superiore al gruppo precedente e appena inferiore alla media del campione (tabella 5.9). Anche i valori della deviazione standard crescono a conferma di una maggiore variabilità intra-gruppo e maggiore ambivalenza nei confronti della figura dello straniero immigrato.

Tab. 5.8: Tolleranti – Roma, k-medie, d16

Items	Gruppo		Campione	
	M	σ	M	σ
Apporto di nuovi stimoli culturali	2,73	1,07	3,45	1,44
Comportamenti invadenti	2,63	1,33	3,75	1,57
Contributo allo sviluppo culturale del paese	3,03	1,32	3,91	1,56
Legame eccessivo alle tradizioni	3,03	1,30	4,36	1,43
Interessi personali	2,39	1,25	3,70	1,62
Originalità	3,60	1,51	3,68	1,55
Mancanza di igiene	2,70	1,46	3,95	1,71
Introduzione ideologie pericolose	1,92	1,15	3,09	1,64
Scarsa voglia di lavorare	1,77	1,09	2,67	1,67
Comportamenti sessuali troppo liberi	1,70	1,08	2,64	1,67
Sottrazione lavoro agli italiani	2,13	1,43	3,20	1,85
Numero eccessivo figli	2,03	1,31	3,35	1,79
Comportamenti di disturbo	2,30	1,32	3,62	1,70
Mancato rispetto regole di convivenza civile	2,39	1,23	3,73	1,58
Introduzione nuove malattie	1,91	1,20	2,78	1,64
Fruizione abusiva sistema di assistenza	1,95	1,20	3,11	1,68
Problemi di ordine pubblico	2,53	1,28	3,68	1,58
Non saper stare al proprio posto	1,78	1,03	3,05	1,72
Contributo allo sviluppo economico del paese	3,35	1,42	3,95	1,60
Degrado quartiere	2,14	1,26	3,28	1,70
Introduzione nuove mode	4,14	1,59	4,47	1,61
Diffusione comportamenti contrari tradizione	2,63	1,31	3,62	1,56
Aumento difficoltà in ricerca abitazioni in affitto	2,21	1,29	3,05	1,66
Diffusione lavoro irregolare	2,76	1,60	3,74	1,75
Atteggiamento troppo critico verso società italiana	2,59	1,46	3,75	1,66
Media	2,49	1,30	3,50	1,64

Il fatto che ambivalenza e eterogeneità siano gli elementi che fanno da padrone in questo gruppo è confermato anche dalla sua definizione sociologica. Solo due variabili illustrative risultano significative: indifferenza rispetto alla collocazione politica e la condizione di “credente/praticante” per quanto riguarda

la sfera religiosa. È interessante quello che fa notare Cipollini, secondo cui “la presenza di un numero così esiguo di modalità-variabili caratteristiche può essere messo in relazione alla presenza diffusa [...] di una «dimensione latente» e condivisa della rappresentazione sociale dello straniero e per questo non suscettibile di specifica caratterizzazione in termini di appartenenza” (Cipollini, 2004, p. 115).

Tab. 5.9: Tolleranti con riserve – Roma, k-medie, d16

Items	Gruppo		Campione	
	M	σ	M	σ
Apporto di nuovi stimoli culturali	2,76	1,11	3,45	1,44
Comportamenti invadenti	3,61	1,36	3,75	1,57
Contributo allo sviluppo culturale del paese	3,11	1,26	3,91	1,56
Legame eccessivo alle tradizioni	4,99	1,00	4,36	1,43
Interessi personali	3,65	1,44	3,70	1,62
Originalità	2,73	1,25	3,68	1,55
Mancanza di igiene	3,73	1,55	3,95	1,71
Introduzione ideologie pericolose	3,07	1,41	3,09	1,64
Scarsa voglia di lavorare	2,49	1,50	2,67	1,67
Comportamenti sessuali troppo liberi	2,74	1,57	2,64	1,67
Sottrazione lavoro agli italiani	3,25	1,69	3,20	1,85
Numero eccessivo figli	3,70	1,63	3,35	1,79
Comportamenti di disturbo	3,41	1,50	3,62	1,70
Mancato rispetto regole di convivenza civile	3,68	1,35	3,73	1,58
Introduzione nuove malattie	2,81	1,52	2,78	1,64
Fruizione abusiva sistema di assistenza	3,07	1,45	3,11	1,68
Problemi di ordine pubblico	3,60	1,43	3,68	1,58
Non saper stare al proprio posto	2,87	1,48	3,05	1,72
Contributo allo sviluppo economico del paese	3,04	1,38	3,95	1,60
Degrado quartiere	2,98	1,45	3,28	1,70
Introduzione nuove mode	3,32	1,48	4,47	1,61
Diffusione comportamenti contrari tradizione	3,81	1,34	3,62	1,56
Aumento difficoltà in ricerca abitazioni in affitto	3,35	1,51	3,05	1,66
Diffusione lavoro irregolare	4,00	1,60	3,74	1,75
Atteggiamento troppo critico verso società italiana	4,00	1,48	3,75	1,66
Media	3,35	1,43	3,50	1,64

Ed effettivamente, osservando i centri dei gruppi, notiamo un forte sbilanciamento di questo gruppo sull’asse negativo della seconda componente, cioè della distanza cognitiva, che nei tolleranti con riserve non viene percepita: gli stranieri, pur se tendenzialmente tollerati, sono percepiti come vicini e in competizione col gruppo.

Speculare ai tolleranti con riserve è il gruppo dei **conflittuali moderati**. I punteggi sono leggermente superiori alla media del campione e la deviazione standard leggermente inferiore. Anche in questo caso, prevalgono eterogeneità e ambivalenza. Osservando i centri dei gruppi, vediamo che questo è caratterizzato da distanza cognitiva e percezione della diversità utilitaristica. Insomma, prevale leggermente il pregiudizio, pur se con leggera intensità ed attenuato dalla percezione dello straniero come di un soggetto distante e perciò non competitivo.

Tab. 5.10: Conflittuali moderati – Roma, k-medie, d16

Items	Gruppo		Campione	
	M	σ	M	σ
Apporto di nuovi stimoli culturali	4,12	1,28	3,45	1,44
Comportamenti invadenti	4,81	1,21	3,75	1,57
Contributo allo sviluppo culturale del paese	4,03	1,43	3,91	1,56
Legame eccessivo alle tradizioni	4,60	1,39	4,36	1,43
Interessi personali	5,35	1,15	3,70	1,62
Originalità	4,14	1,37	3,68	1,55
Mancanza di igiene	5,21	0,87	3,95	1,71
Introduzione ideologie pericolose	4,44	1,32	3,09	1,64
Scarsa voglia di lavorare	4,64	1,39	2,67	1,67
Comportamenti sessuali troppo liberi	3,07	1,47	2,64	1,67
Sottrazione lavoro agli italiani	2,55	1,56	3,20	1,85
Numero eccessivo figli	2,34	1,54	3,35	1,79
Comportamenti di disturbo	2,98	1,72	3,62	1,70
Mancato rispetto regole di convivenza civile	3,49	1,72	3,73	1,58
Introduzione nuove malattie	3,99	1,49	2,78	1,64
Fruizione abusiva sistema di assistenza	4,10	1,34	3,11	1,68
Problemi di ordine pubblico	2,31	1,35	3,68	1,58
Non saper stare al proprio posto	3,00	1,53	3,05	1,72
Contributo allo sviluppo economico del paese	3,82	1,41	3,95	1,60
Degrado quartiere	3,20	1,59	3,28	1,70
Introduzione nuove mode	3,52	1,53	4,47	1,61
Diffusione comportamenti contrari tradizione	3,67	1,60	3,62	1,56
Aumento difficoltà in ricerca abitazioni in affitto	2,44	1,46	3,05	1,66
Diffusione lavoro irregolare	3,59	1,73	3,74	1,75
Atteggiamento troppo critico verso società italiana	3,94	1,51	3,75	1,66
Media	3,73	1,44	3,50	1,64

Il gruppo è caratterizzato sociologicamente da mancata auto collocazione politica e distacco dalle forme di partecipazione ad essa, per quanto ci sia un interesse molto generico. Il sistema di appartenenza è spiccatamente locale (città o quartiere), l'informazione familiare è assente o molto limitata.

L'ultimo gruppo, quello dei **conflittuali**, è speculare al primo. Crescono notevolmente i valori della media e diminuiscono quelli della deviazione standard. Diminuisce quindi l'ambivalenza, a vantaggio di una visione piuttosto omogenea dello straniero come portatore di tutti i tipici pregiudizi rintracciabili nell'analisi della teoria dello straniero e nell'opinione pubblica.

Tab. 5.11: Conflittuali – Roma, k-medie, d16

Items	Gruppo		Campione	
	M	σ	M	σ
Apporto di nuovi stimoli culturali	4,52	1,42	3,45	1,44
Comportamenti invadenti	5,04	1,22	3,75	1,57
Contributo allo sviluppo culturale del paese	4,68	1,34	3,91	1,56
Legame eccessivo alle tradizioni	5,20	1,16	4,36	1,43
Interessi personali	5,42	1,00	3,70	1,62
Originalità	4,96	1,26	3,68	1,55
Mancanza di igiene	4,18	1,35	3,95	1,71
Introduzione ideologie pericolose	4,59	1,50	3,09	1,64
Scarsa voglia di lavorare	5,05	1,41	2,67	1,67
Comportamenti sessuali troppo liberi	4,69	1,32	2,64	1,67
Sottrazione lavoro agli italiani	4,27	1,57	3,20	1,85
Numero eccessivo figli	4,14	1,58	3,35	1,79
Comportamenti di disturbo	4,84	1,53	3,62	1,70
Mancato rispetto regole di convivenza civile	4,42	1,60	3,73	1,58
Introduzione nuove malattie	5,18	1,11	2,78	1,64
Fruizione abusiva sistema di assistenza	5,11	1,10	3,11	1,68
Problemi di ordine pubblico	4,44	1,41	3,68	1,58
Non saper stare al proprio posto	4,85	1,20	3,05	1,72
Contributo allo sviluppo economico del paese	5,12	1,06	3,95	1,60
Degrado quartiere	4,82	1,34	3,28	1,70
Introduzione nuove mode	4,92	1,32	4,47	1,61
Diffusione comportamenti contrari tradizione	4,61	1,34	3,62	1,56
Aumento difficoltà in ricerca abitazioni in affitto	4,49	1,45	3,05	1,66
Diffusione lavoro irregolare	4,87	1,38	3,74	1,75
Atteggiamento troppo critico verso società italiana	4,69	1,49	3,75	1,66
Media	4,76	1,34	3,50	1,64

Anche la definizione sociologica è chiara e inversa rispetto ai tolleranti. Questi giovani aderiscono ad ideologie di destra o centro-destra, hanno un forte senso di appartenenza a contesti locali come quartiere o città, l'informazione familiare è molto bassa, così come sono bassi gli status socio-economico e socio-culturale. È confermata l'interpretazione relativa alla componente dell'Ambivalenza che già si era data a commento del gruppo dei tolleranti: il

pregiudizio appare, in questo contesto, “un insieme coordinato di immagini dello straniero in cui convivono tutte le dimensioni della sua rappresentazione sociale definite dalla teoria sociologica” (Cipollini, 2004, p. 117).

Prima di analizzare le distribuzioni di frequenza di questa *cluster analysis*, sottolineiamo come sia arrivato il momento in cui si fanno sentire le esigenze di sintesi di cui prima parlavamo. Nella prossima tavola sintetica, dunque, oltre alle distribuzioni della tipologia appena vista, inseriremo anche i valori medi di M e σ su tutta la batteria, la varianza media riprodotta espressa in termini percentuali (ossia il valore di eta quadro medio su tutta la batteria, moltiplicato per cento) e la stabilità della tipologia appena vista con tre tipologie semi-sfocate che abbiamo costruito proprio a partire da questa. Un breve accenno sulla costruzione delle 3 tipologie con logica *fuzzy*.

Il punto di partenza è la tipologia sfocata. Si esportano i dati nel programma di cui parlavamo, sviluppato da Middei. Si devono esportare le variabili (tre, cioè le componenti principali) e l'appartenenza al gruppo. Questa deve essere però espressa in forma binaria, cioè come 1 o 0. In parole povere, se in SPSS abbiamo una sola variabile, una sola colonna, con appartenenza ai gruppi espressa alternativamente dai valori 1, 2, 3 o 4, nel programma per le classificazioni sfocate, dobbiamo inserire quattro colonne ed ogni caso deve avere un valore di 1 nella colonna del gruppo a cui appartiene ed un valore di 0 in tutte le altre. Il caso che quindi appartiene al secondo gruppo, si presenterà in questa forma: (0, 1, 0, 0). Tale codifica è definita *codifica sotto forma disgiuntiva completa*.

A questo punto è possibile manipolare i parametri m e α che caratterizzano l'algoritmo delle k-medie semisfocato⁴. Ricordiamo che la sfocatura cresce al crescere dei parametri e viceversa, e ad m uguale ad 1, la classificazione sfocata si trasforma in classificazione binaria. Avendo due parametri da modificare e nessun criterio preliminare sull'entità della sfocatura da applicare, è buona norma stabilire un parametro, previo qualche tentativo, e bloccarlo, salvo poi modificare l'altro. In nostre precedenti analisi, avevano trovato un interessante punto di

⁴ Di tale algoritmo e dei parametri m ed α abbiamo parlato diffusamente nel paragrafo 3.5, a cui rimandiamo.

equilibrio, in termini di perdita delle informazioni iniziali, con i valori ($m = 1,9$) e ($\alpha = 1,3$). Da questi, bloccato m , ci eravamo mossi in senso di una maggiore e minore sfocatura, identificando 5 tipologie ($m = 1,9$; $\alpha = 1,1$; 1,2; 1,3; 1,4; 1,5). Questa volta, per ridurre il lavoro, abbiamo saltato la seconda e la quarta impostazione del parametro α . Inoltre, visto che nella sfocatura maggiore c'era stato un crollo della varianza riprodotta (i mancanti di sistema erano diventati maggiori dei casi presi in esame), abbiamo deciso di abbassare leggermente m , impostandolo sul valore di 1,8.

Dopo aver modificato i parametri, si manda il programma in esecuzione, poi si possono esportare i dati di nuovo in SPSS e si può continuare con l'analisi. Si può teoricamente procedere in più modi, quello da noi prescelto consiste nella semplice non assegnazione dei casi ad appartenenza sfumata. Avremo così una tipologia caratterizzata da zone di piena appartenenza (e piena non appartenenza) a zone in cui l'elemento appartiene e non appartiene. L'esclusione di questi ultimi elementi è da intendersi come scelta consapevole di eliminare le assegnazioni forzate dall'analisi, in linea di continuità con gli insegnamenti della logica *fuzzy*.

I gruppi manterranno le stesse etichette avute attraverso l'analisi tradizionale. In questo caso la logica *fuzzy* interviene, infatti, solo in un secondo momento, come affinamento di un lavoro impostato in modo tradizionale. È una miglioria tecnica, non un cambiamento di orizzonti vero e proprio. Vediamo comunque i valori di media e deviazione standard sull'intera batteria:

Tab. 5.12: Media e deviazione standard delle tipologie sulla batteria – Roma, d16

Roma – d16	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	M	σ	M	σ	M	σ	M	σ
Tolleranti	2,49	1,30	2,50	1,22	2,56	1,20	2,64	1,18
Tolleranti con riserve	3,35	1,43	3,53	1,34	3,50	1,26	3,50	1,22
Conflittuali moderati	3,73	1,44	3,66	1,41	3,71	1,36	3,76	1,32
Conflittuali	4,76	1,34	4,76	1,31	4,82	1,27	4,77	1,27
Totale	3,50	1,64	3,57	1,58	3,63	1,54	3,65	1,48

Come ci si poteva aspettare, i valori di M aumentano e quelli di σ diminuiscono al crescere della sfocatura, in termini non del tutto trascurabili, pensiamo ad esempio al valore di σ dei tolleranti con riserve che già alla seconda sfocatura ($\alpha = 1,3$) passa da 1,43 a 1,26. Questo vale sia per i valori dei gruppi, sia

per quelli del campione. I valori del campione della terza tipologia semi-sfocata tendono, ad esempio, ad assomigliare a quelli dei conflittuali moderati nell'analisi tradizionale. Ad ogni modo, in linea con quanto ci aspettavamo, i gruppi migliorano in termini di omogeneità col crescere della sfocatura, che comporta l'eliminazione dei (tanti) casi di dubbia assegnazione.

Vediamo ora le distribuzioni, la stabilità, la varianza media riprodotta e le iterazioni necessarie delle quattro tipologie:

Tab. 5.13: Distribuzioni di frequenza, stabilità ed eta – Roma, d16

Roma – d16	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
Gruppi	N	% val	N	% val	N	% val	N	% val
Tolleranti	380	26,35	308	25,39	240	25,42	175	24,04
Tolleranti con riserve	410	28,43	308	25,39	233	24,68	194	26,65
Conflittuali moderati	365	25,31	334	27,54	250	26,48	194	26,65
Conflittuali	287	19,90	263	21,68	221	23,41	165	22,66
Totale valido	1442	100,00	1213	100,00	944	100,00	728	100,00
Mancante di sistema	0	-	229	15,88	498	34,54	714	49,51
Iterazioni	12		23		16		9	
Stabilità	-		89,69		95,34		97,80	
Eta quadro	28,52		29,36		30,76		28,52	

Prima di tutto, osserviamo le distribuzioni di frequenza. Con l'aumento della sfocatura si registra un assottigliamento della differenza in percentuale valida dei due gruppi estremi, i tolleranti e i conflittuali: la maggiore numerosità dei tolleranti che si registra nella tipologia classica e nella prima semi-sfocata, tende quasi a scomparire. Le tre tipologie semi-sfocate risultano tutte molto stabili se paragonate con quella tradizionale (il numero in parentesi indica la percentuale sul totale valido che cade nella diagonale principale). L'eta quadro, che indica in un certo senso la bontà della classificazione, cresce nelle prime due tipologie semi-sfocate e, nel terzo caso, ritorna curiosamente sul livello di quella tradizionale.

Con l'applicazione di algoritmi *fuzzy* abbiamo, quindi, gruppi più omogenei al loro interno, più eterogenei fra loro, valori medi sugli items più alti e valori di deviazione standard più bassi. Sotto molti punti di vista, sembrano classificazioni migliori. È però vero che c'è una caduta di casi, nel senso che molti vengono esclusi dall'analisi. Se però consideriamo che questi sono i casi in cui

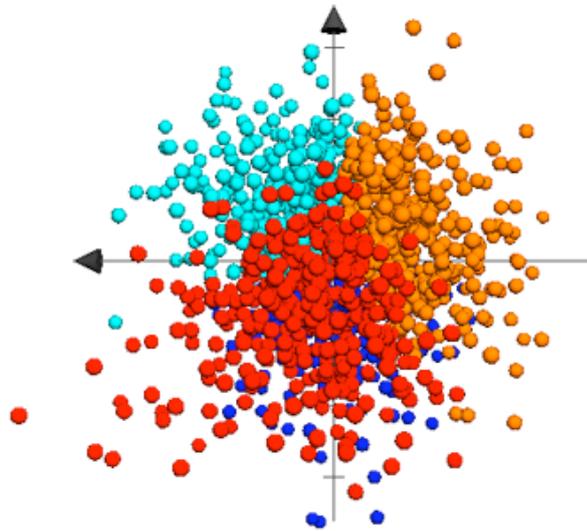
l'assegnazione risulta più forzata, questo “difetto” potrebbe ben essere interpretato come un pregio. Un'interessante scelta di compromesso potrebbe essere rappresentata dalla seconda tipologia semi-sfocata, dove poco più di un terzo di casi vengono esclusi e dove allo stesso tempo quasi 1000 giovani mantengono la loro assegnazione. Questa tipologia è anche molto stabile e presenta il valore di età quadro più alto, con un incremento di più del 2%, che non sembra eccezionale in termini assoluti, ma se consideriamo che stiamo lavorando su dati individuali, è tutt'altro che trascurabile.

Per controllare se nei dati è presente una sorta di tendenza alla *clusterizzazione*, spesso si dice che è utile guardare la distribuzione grafica dei dati sulle componenti. Il problema è, però, che di norma si osservano due componenti alla volta, si schiaccia quindi l'osservazione sul piano cartesiano e scompare la possibilità di un'utile osservazione.

Con il programma che abbiamo prima nominato, Grapher, è possibile sfuggire a questa *impasse*. Grapher è un programma che permette l'elaborazione di grafici statistici anche in tre dimensioni. Possiamo inserire i valori di ogni caso, colorare i casi in modo diverso a seconda del gruppo al quale appartengono e vedere questi casi proiettati nello spazio tridimensionale formato dalle nostre variabili. Questa proiezione è poi trasciabile con il *mouse* in modo che la distribuzione dei casi possa essere osservata da ogni angolazione possibile. Si possono estrarre filmati o fotografie.

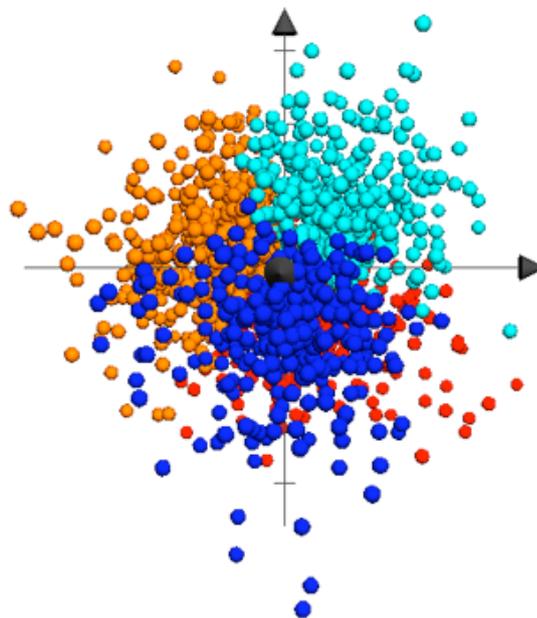
Sfortunatamente, su carta è impossibile rendere il 3D. Ci limiteremo all'inserimento di qualche *screenshot*. Cominciamo a vedere la tipologia tradizionale, guardando i casi dal lato dei tolleranti. Gli assi rappresentano le tre componenti principali estratte; abbiamo colorato i tolleranti (che erano sociologicamente schierati verso la sinistra politica) di rosso, in arancio i tolleranti con riserva, in celeste i conflittuali moderati ed in blu i conflittuali (schierati a destra).

Fig. 5.1: Rappresentazione in 3d, K-medie – Roma d16



Vediamo ora il grafico dal lato opposto:

Fig. 5.2: Rappresentazione in 3d, K-medie – Roma d16

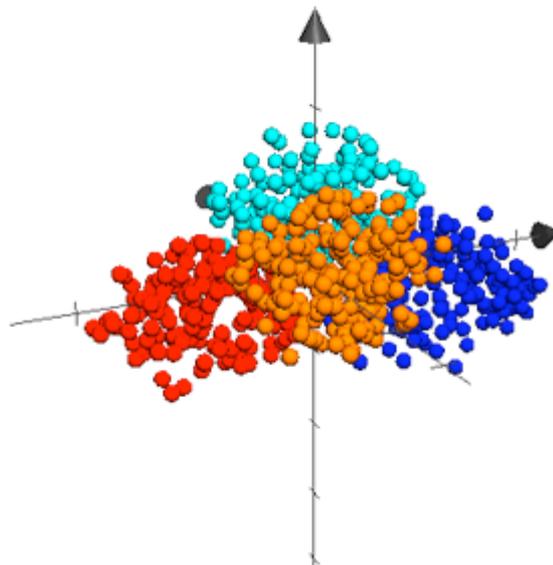


È possibile notare come una “struttura” tendente alla classificazione non ci sia. I discorsi fatti nei capitoli precedenti riguardo alle classificazioni naturali (anche se intese in senso puramente metodologico) tornano d’attualità. Immaginiamo di colorare di nero i quattro gruppi e ci ritroviamo di fronte al tipico

“blocco compatto” visto molte volte su assi cartesiani. La confusione (nel senso etimologico del termine) è l’aspetto prevalente. Ciò nondimeno, bisogna dire che la costruzione (ci sembra appropriato l’uso di questo termine) della tipologia ha portato ad esiti quantomeno decenti. I gruppi sembrano “fuggire” in direzioni diverse, rammentando che ogni direzione rappresenta una proprietà, o meglio una combinazione di più proprietà. In questo manifestano la loro diversità. La situazione più problematica è al centro, nei pressi del punto zero, dove le divisioni fra un gruppo e l’altro diventano un vero e proprio rebus.

Il secondo grafico che abbiamo scelto di mostrare riguarda la terza tipologia semi-sfocata, quella in cui la sfocatura è quindi maggiore. Abbiamo scelto questa perché ci aspettiamo che sia quella più lontana dalla precedente anche a livello d’immagine. Abbiamo scelto una prospettiva leggermente angolata, partendo dal lato favorevole ai tolleranti con riserva.

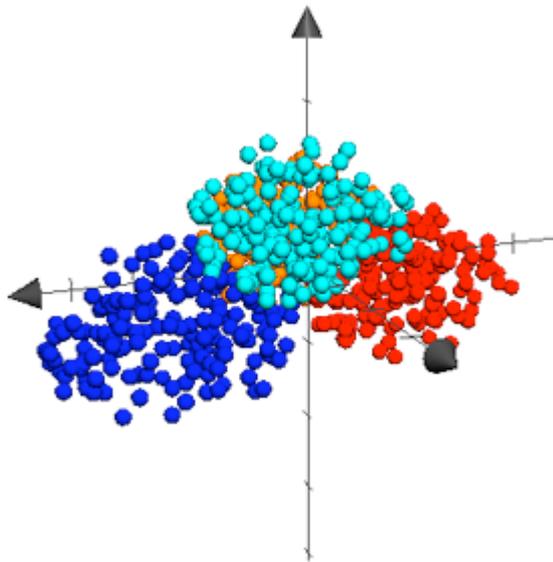
Fig. 5.3: Rappresentazione in 3d, Fuzzy III – Roma d16



È evidente come la situazione appaia decisamente più nitida. I gruppi sono sempre gli stessi, chiaramente, per questo prima avevamo parlato di miglioria tecnica e non di cambio di prospettiva. Ma le assegnazioni sembrano molto più ordinate, è in generale un grafico più “pulito” rispetto al precedente. Verso il

centro degli assi, in prossimità del punto zero, permane una certa, inevitabile confusione, ma in misura enormemente inferiore rispetto alla tipologia tradizionale. I *clusters*, benché ovviamente non naturali, si presentano alla vista molto più separati, distinti e distinguibili se così si può dire. Quello che non avevamo preso in considerazione, è l'eliminazione dei casi più lontani. Avevamo pensato che l'applicazione degli algoritmi *fuzzy* avrebbe inciso nella zona intermedia, eliminando i casi in cui era più arbitraria una distinzione fra, ad esempio, conflittuali moderati e tolleranti con riserva. Come invece mostrano chiaramente i grafici, escono dalla tipologia anche i casi più lontani dal centro, la spiegazione è ovvia, la distanza di questi casi dal baricentro del gruppo è necessariamente maggiore alla soglia introdotta dal parametro α . Vediamo ora l'immagine dal punto di vista opposto:

Fig. 5.4: Rappresentazione in 3d, Fuzzy III – Roma d16



C'è poco da aggiungere se non confermare quanto detto e sottolineare che l'aspetto grafico si pone come un punto di vista in più rispetto all'analisi e – in un certo qual modo e ovviamente dal nostro punto di vista – segna un piccolo punto a favore verso le tipologie *fuzzy* che va ad aggiungersi alle considerazioni già fatte a margine delle tabelle.

5.2.2 CALIBRATION

Facciamo ora il terzo *step* della nostra elaborazione, andando ad applicare la tecnica della *calibration*. Quest'applicazione si differenzia in modo notevole dalla precedente, perché introduce la visione *fuzzy* in un momento precedente dell'analisi. Se prima avevamo detto che eravamo di fronte solo ad una migliore tecnica, qua crediamo di trovarci di fronte ad un'assunzione di un diverso punto di vista. Questo consiste prima di tutto nella scelta di utilizzare misure calibrate, il che comporta un'assunzione di un *target set* di riferimento e, di conseguenza, l'adesione ad un discorso scientifico che segue la logica del ragionamento con insiemi. Come se questo non bastasse, ricordiamo come la formazione di gruppi non avviene estensionalmente, ma in base alla costruzione di una tavola di verità/spazio di attributi. In questo senso, possiamo dire che le tipologie sono definite *ex-ante*, intensionalmente, e solo dopo vengono riempite di materiale esistenziale. Riprendendo la distinzione di Marradi (1993) delle procedure classificatorie in: intensionali, estensionali e per assegnazione a tipi già formati; possiamo dire che la *cluster analysis* è una versione operativa su base matriciale della seconda procedura; l'uso di *truth table* e *calibration* è, a nostro avviso, una versione operativa dell'uso combinato della prima e della terza procedura. La formazione di classi per via intensionale è seguita da un'assegnazione automatizzata tecnicamente. Ne è prova il fatto che, pur utilizzando dati a livello individuale, riusciamo ad arrivare ad una tipologia *bypassando* la *cluster analysis*, che è il metodo universalmente adottato per raggiungere tale scopo.

Punto di partenza per l'applicazione della *calibration* è, come detto nel capitolo precedente, la disponibilità di una scala quantitativa e la possibilità di interpretare le variabili originali in termini di appartenenza ad un gruppo di riferimento. A questo punto basta scegliere le tre soglie (0,05; 0,50; 0,95) per procedere con la trasformazione della variabile in appartenenza *fuzzy*.

Le nostre scale quantitative sono rappresentate dalle nostre componenti. Queste possono essere pensate in termini di appartenenza, ad esempio il livello di ambivalenza può essere rivisto nei termini di appartenenza al gruppo dei conflittuali o dei tolleranti, allo stesso identico modo in cui il livello di reddito

nazionale pro-capite può essere pensato come appartenenza di un paese all'insieme dei paesi sviluppati, dei paesi non sviluppati, dei paesi almeno moderatamente sviluppati e via dicendo.

Abbiamo deciso di operare la calibrazione sempre nella direzione del gruppo più portatore di pregiudizio, declinato nelle sue varie forme. Le soglie le abbiamo stabilite avvalendoci di criteri matematici, osservando le distribuzioni di frequenza, in particolare avvalendoci dei percentili, il decimo ed il novantesimo. In parole povere, abbiamo scelto di considerare per intero appartenenti al gruppo di riferimento l'ultimo 10% dei casi osservati (nel caso in cui la soglia 0,95 sia espressa da un numero positivo; il primo 10% in caso di numero negativo) e per nulla appartenenti il primo 10% dei casi osservati (nel caso in cui la soglia 0,05 sia espressa da un numero negativo; l'ultimo 10% in caso di numero positivo). Tutto sarà più chiaro una volta osservato in tabella.

Cominciamo prima di tutto a determinare l'insieme di riferimento per ognuno dei tre fattori della componente cognitiva del pregiudizio, ricordando che abbiamo sempre scelto come riferimento i gruppi con manifestazione maggiore di pregiudizio:

1. Ambivalenza → Giovani conflittuali;
2. Distanza cognitiva → Giovani che percepiscono vicinanza e competizione;
3. Diversità con orientamento utilitaristico → Giovani che percepiscono diversità utilitaristica.

Facciamo presente che si tratta di una decisione, in alcuni casi, che porta con sé un grado di arbitrarietà. Ad esempio, la seconda componente potrebbe essere interpretata in modo opposto, visto che, come abbiamo potuto osservare, la vicinanza è percepita dai tolleranti con riserve, che, seppur lievemente, sono caratterizzati da pregiudizio in quantità minore rispetto ai conflittuali moderati (che, sulla seconda componente, sono sulla sponda della distanza cognitiva), come dimostrano i valori sulla prima componente, che è quella dominante del pregiudizio. C'è però quella connotazione di “vicinanza e competizione” che ci fa

propendere per questa interpretazione, coerentemente con l'idea per cui il conflitto dei moderati sia appunto smorzato da questa distanza, e le riserve dei tolleranti sono espresse proprio in termini competitivi.

In generale, dobbiamo ricordare che ci stiamo muovendo in un campo ben poco esplorato, soprattutto se consideriamo che la *calibration* nasce ed è utilizzata negli studi comparativi, che operano a livello ecologico. Sotto questo punto di vista, ben vengano eventuali errori se stimolano la riflessione e fanno muovere appunti critici.

A questo punto, per sottolineare il carattere intensionale dell'operazione, possiamo già costruire la nostra tipologia, sulla base di uno spazio di attributi, le cui combinazioni si articolano in assenza e presenza delle proprietà considerate. Il modello è il seguente:

Tab. 5.14: Spazio di attributi a 3 proprietà (presenza/assenza)

		Y ₀	Y ₁
X ₀	Z ₀	X ₀ Y ₀ Z ₀	X ₀ Y ₁ Z ₀
	Z ₁	X ₀ Y ₀ Z ₁	X ₀ Y ₁ Z ₁
X ₁	Z ₀	X ₁ Y ₀ Z ₀	X ₁ Y ₁ Z ₀
	Z ₁	X ₁ Y ₀ Z ₁	X ₁ Y ₁ Z ₁

Considerando la nostra prima componente come X, la seconda come Y e la terza come Z, i tipi nelle celle equivarranno ai seguenti:

- Tolleranti, non percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (000);
- Tolleranti, non percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (001);
- Tolleranti, percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (010);
- Tolleranti, percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (011);
- Conflittuali, non percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (100);

- Conflittuali, non percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (101);
- Conflittuali, percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (110);
- Conflittuali, percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (111).

È chiaro che l’etichettamento di questi gruppi potrebbe essere molto migliorato. Ciò potrebbe avvenire attraverso qualche lettura teorica, l’aiuto di qualche esperto proveniente da altri rami, quali la psicologia, o anche solo per intuizione o uso di metafore. Ci sembra però altrettanto chiaro il contributo che una tale tipizzazione potrebbe fornire ad un’analisi approfondita. Ad esempio, pensiamo alla visione alla Simmel, per cui lo straniero lontano e vicino è visto con curiosità e timore. Ma chi prova la curiosità, chi il timore? Come e dove la vicinanza va di pari passo col pregiudizio? Come e dove va di pari passo con la diminuzione del pregiudizio? Queste combinazioni sono presenti nella tipologia appena vista, è là che possiamo trovare i tolleranti che si sentono vicini, i tolleranti che si sentono lontani, i conflittuali che si sentono vicini ed i conflittuali che si sentono lontani.

Ora bisogna vedere come questa procedura si comporta in termini statistici. Prima di tutto, due parole sul programma, fsQCA, sviluppato sull’impostazione metodologica di Charles Ragin. Il funzionamento è molto semplice, si esportano i dati in formato CSV, si inserisce una prima riga per etichettare le variabili (nel nostro caso le tre componenti), poi si può calcolare la variabile calibrata, scegliendo il comando *compute* e dando una sintassi così composta:

```
calibrate(vecchiavariabile,0.95,0.5,0.05)
```

dove “vecchiavariabile” è l’etichetta della misura non calibrata e i tre valori numerici sono le rispettive soglie che avremo stabilito. Il risultato dell’operazione sarà una serie di colonne, tante quante erano le variabili iniziali,

con valori che variano tra 0 e 1, intesi come appartenenza *fuzzy* di ogni caso al gruppo di riferimento.

Nel nostro caso, come dicevamo, ci siamo avvalsi dei percentili per determinare le soglie (tabella 5.15). Come soglia 0,5 abbiamo scelto sempre il punto zero. Considerando che per la seconda componente, la vicinanza (associata all'insieme di riferimento) aumenta in corrispondenza del semiasse negativo, ne consegue che la soglia 0,95 sarà il numero negativo, corrispondente al decimo percentile.

Tab. 5.15: Statistiche descrittive e percentili delle tre componenti

		FAC1_16	FAC2_16	FAC3_16
N	Valid	1442	1442	1442
	Missing	0	0	0
Minimum		-2,29	-3,34	-4,06
Maximum		2,41	3,80	3,01
Percentiles	10	-1,34	-1,27	-1,26
	50	0,00	0,02	0,02
	90	1,36	1,24	1,26

Questo non ha alcun effetto nocivo sulla calibrazione. Infatti, se calibriamo due volte una variabile invertendo le soglie e sommiamo le due appartenenze sfocate che ne scaturiscono, il totale è sempre 1, sia che si lavori su numeri che hanno un *range* (0, n), sia che si lavori con numeri che hanno un *range* (-n, +n). Queste sono, quindi, le soglie della nostra calibrazione:

$c161 = \text{calibrate}(f161, 1.36, 0, -1.34)$

$c162 = \text{calibrate}(f162, -1.27, 0, 1.24)$

$c163 = \text{calibrate}(f163, 1.26, 0, -1.26)$

Dove $c161$ è la misura calibrata della prima componente (fattore $f161$) e via dicendo. A questo punto, si dà il comando *truth table* e si ottiene l'assegnazione dei casi ai gruppi/angoli più vicini. Un caso con appartenenze (0,12;0,74;0,27) sarà dunque assegnato al gruppo (0;1;0). Questo è l'output che ci dà il programma:

Tab. 5.16: Truth table

C161	C162	C163	number
1	0	1	216
0	1	1	192
0	0	0	190
0	1	0	170
1	1	0	170
1	1	1	166
1	0	0	159
0	0	1	153

Nel programma fsQCA non c'è una procedura che renda la variabile di assegnazione dei casi al relativo gruppo, o almeno noi non l'abbiamo trovata. Quindi, una volta importati i nuovi dati in SPSS, si deve procedere alla creazione e ricodifica di una nuova variabile. Ricordando che un caso che ha valore 0,5 non propende per nessuna delle due sponde del vettore e, quindi, non è soggetto all'assegnazione ad alcun gruppo, la ricodifica può essere espressa direttamente in termini di minore o maggiore del punto di *crossover*, per esempio attraverso la seguente sintassi di SPSS, che riguarda la ricodifica del solo primo gruppo:

```
DO IF (c161 > 0.5 and c162 < 0.5 and c163 > 0.5) .  
RECODE  
    cal16 (ELSE=1) .  
END IF .  
EXECUTE .
```

A seguire, vediamo la tabella 5.17 che mostra la distribuzione di frequenza associata ai tipi etichettati secondo la logica prima vista e la media della varianza riprodotta dalla tipologia sugli items della domanda 16.

Possiamo notare che i gruppi, evidentemente, non sono ordinati in termini di combinazioni logiche, come abbiamo fatto prima noi quando abbiamo esposto la tipologia, ma in ordine di numerosità decrescente. Abbiamo deciso di lasciarli in questo ordine.

Tab. 5.17: Distribuzioni di frequenza, eta – Roma, calibration, d16

	N	% Val
Conflittuali, non percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (101)	216	15,25
Tolleranti, percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (011)	192	13,56
Tolleranti, non percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (000)	190	13,42
Tolleranti, percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (010)	170	12,01
Conflittuali, percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (110)	170	12,01
Conflittuali, percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (111)	166	11,72
Conflittuali, non percepiscono vicinanza e competizione, non percepiscono diversità utilitaristica (100)	159	11,23
Tolleranti, non percepiscono vicinanza e competizione, percepiscono diversità utilitaristica (001)	153	10,81
Totale valido	1416	100,00
Missing	26	1,80
Totale	1442	-
Eta quadro	34,65	

I risultati ci appaiono subito decisamente interessanti. La distribuzione è molto ben bilanciata, tanto che i gruppi, nonostante siano ben otto, sono quasi perfettamente equidistribuiti. La differenza in percentuale valida tra il gruppo più numeroso e quello meno numeroso è, infatti, inferiore al 5%. Questo potrebbe consentirci, con una certa agilità, di accorpate qualche tipo, nel caso volessimo ridurre la tipologia, senza che ne derivi un forte squilibrio interno alla stessa. Sempre in direzione di un'ipotetica riduzione di tipologia, avremmo già chiaro davanti ai nostri occhi il significato semantico di ogni tipo. Non avremmo problemi ad unire il tipo (001) col tipo (000), qualora volessimo rinunciare alla terza dimensione, ad esempio, ma dovremmo essere veramente distratti per mettere insieme tipi opposti come quello (010) e quello (101), che non hanno nessuna dimensione in comune. Insomma, una tipologia costruita in questo modo ha dalla sua anche un non indifferente grado di “malleabilità”, per così dire.

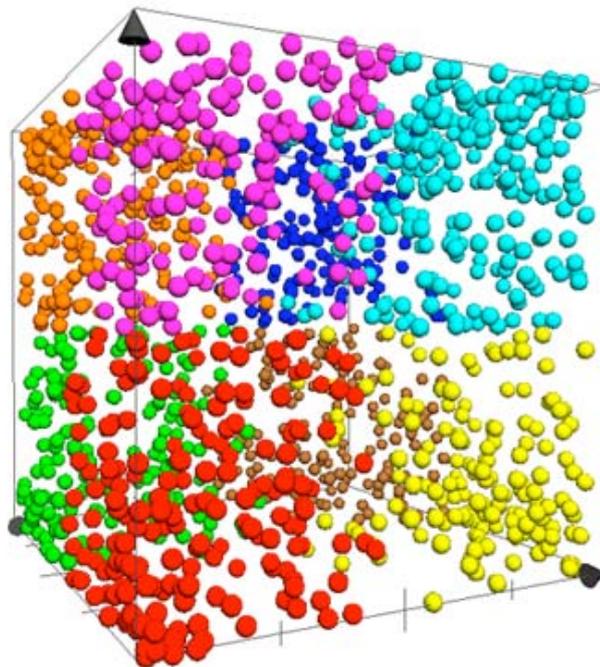
I casi non assegnati sono pochissimi, soltanto 26 su 1442. Escono dalla classificazione, infatti, solo quei casi che presentano un valore esattamente di 0,5 su almeno una delle variabili calibrate. È chiaramente una situazione molto rara, identica a quella di un caso che abbia un valore preciso di 0 su una delle componenti principali estratte.

L'ultimo dato riguarda la perdita di informazioni iniziali e, anche qui, il risultato è positivo. Ci sbilanciamo nell'affermare che è molto positivo. Il valore percentuale di varianza riprodotta è del 34,65%. Rispetto al 28,52% della *cluster analysis* tradizionale, registriamo un aumento assoluto del 6,13% che, se consideriamo l'enorme variabilità e la confusione tipiche dei dati individuali, ci sembra davvero un dato notevole. In termini relativi, facendo cento il valore di eta dell'analisi tradizionale, è un aumento superiore al 21% (21,49%, per la precisione).

Ora, esattamente come avevamo fatto per la *cluster analysis*, proiettiamo questi valori sul grafico a tre dimensioni. Se prima avevamo tre assi che si incrociavano allo zero, ora avremo tre vettori che partono da un punto zero in comune e si dirigono ognuno verso il limite massimo di uno. Quindi, geometricamente, vedremo un cubo, che ricorda l'ipercubo *fuzzy* a tre dimensioni di cui parlava Kosko.

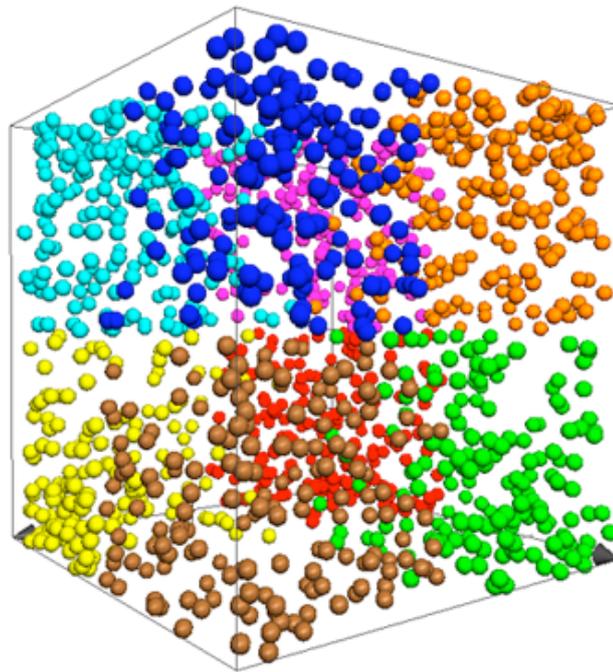
Vediamo un'immagine leggermente angolata, partendo dalla prospettiva dell'angolo (0, 0, 0), i cui casi, segnati da tendenza al non pregiudizio, sono colorati di rosso e sono contraddistinti dall'essere più fuori che dentro ai *target set* scelti per calibrare:

Fig. 5.5: Rappresentazione in 3d – calibration, Roma d16, angolo (0,0,0)



La differenza nella costruzione di questa tipologia, punto su cui tanto abbiamo insistito, balza subito agli occhi in tutta la sua evidenza grafica. Se prima vedevamo una nuvola di casi, tendente al blocco compatto e tendente al collasso sul punto zero, qui vediamo un cubo a tre vettori/proprietà, il cui spazio geometrico (e concettuale) è completamente e letteralmente utilizzato dai nostri casi. I punti non sono attratti specificamente né verso il centro né verso gli angoli, ma si distribuiscono in modo piuttosto armonico, ognuno mantenendo i suoi valori di appartenenza sfocata alle componenti e, allo stesso tempo, vedendosi assegnato all'angolo che più gli somiglia. Vediamo il grafico dal punto di vista opposto, quello dell'angolo (1, 1, 1) colorato di blu:

Fig. 5.6: Rappresentazione in 3d – calibration, Roma d16, angolo (1,1,1)

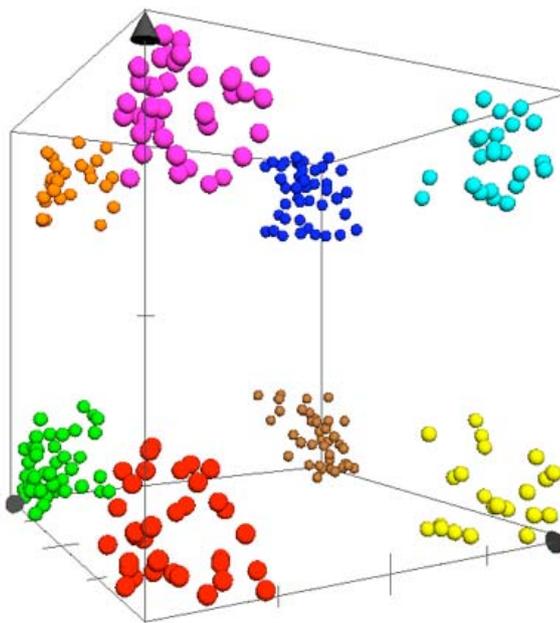


Con questa classificazione non perdiamo nulla, guadagniamo invece, in termini di varianza riprodotta, in termini di flessibilità concettuale ed in termini di mera quantità di informazioni, due per ogni caso (l'assegnazione e l'appartenenza), invece che una (la sola assegnazione). Queste informazioni possono essere recuperate ed utilizzate a piacimento a seconda dell'obiettivo

cognitivo che di volta in volta il ricercatore si pone. Se si volesse lavorare sul nucleo centrale di persone che non assomigliano troppo a nessun angolo in particolare, ci basterebbe selezionare tutti i casi che hanno appartenenze che variano tra 0,25 e 0,75. I casi selezionati sarebbero, così, una sorta di tipologia dei “poco tipici”, coloro che assomigliano al punto 0.5, dove, riprendendo le parole di Kosko, l’equazione yin-yang è al suo massimo ed A assomiglia a non-A.

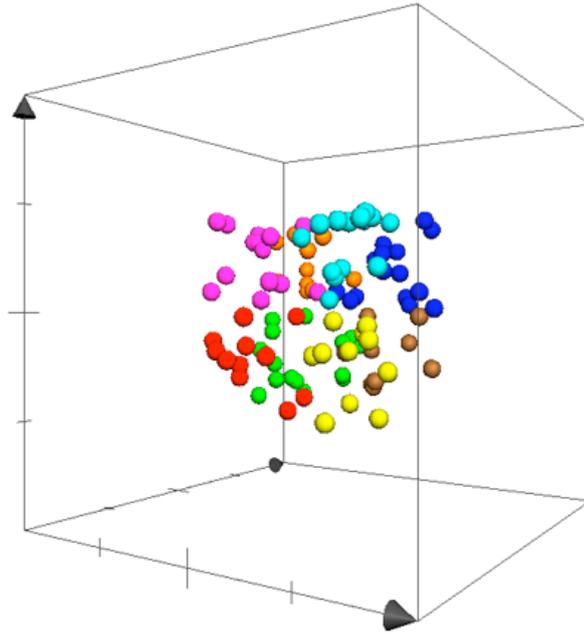
Al contrario, selezionando i casi che hanno appartenenze superiori a 0,75 o inferiori a 0,25, si potrebbe lavorare sui soli casi in cui la sfumatura cede gradualmente il passo alla bivarianza, i casi proto-tipici, come li avevamo prima definiti. Possiamo vedere le due soluzioni rappresentate nei grafici che seguono:

Fig. 5.7: Rappresentazione in 3d – calibration, Roma d16, gli angoli



Basta riempire i vettori dei significati più disparati che possono venirci in mente, guidati dall’esperienza acquisita sul campo, dai nostri interessi cognitivi, dalla lettura delle teorie, dalle possibilità di definizione operativa delle proprietà e così via, e avremo a portata di mano classificazioni che crediamo siano utili e interessanti.

Fig. 5.8: Rappresentazione in 3d – calibration, Roma d16, il centro



Per fare un esempio, immaginiamo che i vettori rappresentino la controparte empirica di proprietà che hanno a che vedere con la sfera della politica, quindi, ad esempio, il comportamento elettorale passato, le intenzioni di voto, l'informazione su temi politici, l'interesse e via dicendo. Immaginiamo che le polarizzazioni siano costruite nei termini dell'interesse/non interesse, dell'asse destra/sinistra, dell'informazione/non informazione, ecc. Costruita una simile classificazione, dovremmo trovare agli angoli dei gruppi di individui che corrispondono proto-tipicamente a queste polarizzazioni, mentre al centro individui che sono poco di destra, poco di sinistra, non molto informati né troppo disinformati, mediamente interessati. Non è forse questo il pubblico ideale a cui il *marketing* politico si potrebbe rivolgere in cerca di voti, cioè quel bacino di moderati eternamente indecisi? O forse, il giusto *target* è quello di chi ha 0,5 sull'asse di collocazione politica e assenza di tutte le altre proprietà? Chiaramente non abbiamo risposte, ma ci sembra che questo sia uno strumento che potrebbe essere utile di fronte a domande di questo tipo.

A questo punto è arrivato il momento di fare una piccola pausa. Ci spieghiamo meglio. Consideriamo “fatta” la gran parte della nostra applicazione, nel senso che abbiamo visto le tre tecniche diverse, ne abbiamo spiegato il

significato e l'utilizzo ed abbiamo esposto ed analizzato i risultati. Quello che ci aspetta è, sotto un certo punto di vista, una parte minore o comunque più sintetica. A livello concettuale ed interpretativo, il più è già stato detto, non è nostra intenzione annoiare il lettore ripetendo cose già scritte. Nelle applicazioni che seguiranno (mancano un paio di analisi sulla parte di Roma per la componente cognitiva, la ripetizione dell'analisi nel contesto laziale, e l'analisi per la componente affettiva ed attiva in entrambi i contesti) cercheremo di essere più brevi. Ci limiteremo a sottolineare diverse interpretazioni ove queste si presentino, ma per lo più ci limiteremo ad illustrare i risultati. Il motivo di una mole quantitativamente importante di dati come questa è, in effetti, avere una base empirica su cui ragionare un po' più solida, visto che stiamo facendo applicazioni non del tutto convenzionali. Ma crediamo che ripetere le stesse considerazioni sull'analisi della varianza relativa di volta in volta al *cluster* di turno, o ribadire ripetutamente cosa significhi assumere un *target set* non sia di alcun aiuto o interesse. Speriamo che il lettore abbia trovato spunti sufficienti nell'elaborazione fin qui realizzata ed apprezzi un'esposizione più agile da qui in poi.

Continuiamo l'analisi sempre su Roma e sulla componente cognitiva del pregiudizio, con alcune piccole integrazioni. Abbiamo ritenuto di non sottoporre alla prova della stabilità la tipologia formata per *calibration* appena vista, dato che presenta il doppio dei gruppi rispetto alla tipologia tradizionale. Abbiamo così deciso di usare le prime due componenti calibrate per formare una tipologia a quattro gruppi, rappresentata da un quadrato con quattro angoli. Questa è la tipologia scaturita dall'analisi:

Tab. 5.18: Distribuzioni di frequenza, eta – Roma, calibration, d16 (2 comp.)

	N	% Val
Conflittuali, non percepiscono vicinanza e competizione (10)	376	26,40
Tolleranti, percepiscono vicinanza e competizione (01)	366	25,70
Tolleranti, non percepiscono vicinanza e competizione (00)	345	24,23
Conflittuali, percepiscono vicinanza e competizione (11)	337	23,67
Totale valido	1424	100,00
Missing	18	1,25
Totale	1442	-
Eta quadro	30,98	

Quanto detto per la precedente *calibration* in termini di “malleabilità” vale anche per questa. L’eta quadro scende sensibilmente rispetto alla tipologia a otto gruppi, che evidentemente riproduce meglio il totale delle informazioni. D’altronde, quegli items, sottoposti ad ACP, davano tre componenti, di cui solo due sono entrati in gioco ora. È comunque superiore alla *classificazione* a k-medie.

Ora dovremmo analizzare la stabilità tra le due. C’è però da fare una considerazione. I gruppi della tipologia classica sono etichettati sostanzialmente sulla base della loro posizione sulla prima componente. Basta andare a rivedere i centri dei *cluster* e ripensare alle intuizioni di Cipollini, secondo cui nei tolleranti con riserva c’era probabilmente una forte influenza di una dimensione latente che però non emergeva abbastanza da influire sull’appartenenza. Ora, ci rendiamo conto che questo non accade con la *calibration*, dato che ogni dimensione ha pari dignità e pari influenza nella costruzione dei tipi. Ogni angolo, quindi ogni gruppo, è formato in parti uguali dalla convergenza di tre vettori (due in questo caso). È un pregio o difetto? Se si guarda alla completezza e profondità dell’analisi, diremmo un pregio (in più, l’eta quadro aumenta). Se si guarda al fatto che una componente che spiega il 40% dell’inerzia (dimensione dominante) “conti” quanto una componente che ne spiega il 7% (dimensione latente), si direbbe un difetto.

Anche in questo caso, non c’è una soluzione nettamente e definitivamente migliore di un’altra, come, di fatto, non c’è (quasi) mai. Il punto è trovare la soluzione che meglio funziona rispetto al determinato scopo che ci prefiggiamo. Di certo, ci sembra che questo pregio/difetto sia l’ennesimo indizio (o conferma) del fatto che calibrando si costruisce una tipologia intensionalmente. Vediamo dunque la stabilità tra le due tipologie, ma a scopo descrittivo, non statistico, anche perché in un certo senso decade l’idea stessa di una “diagonale principale”:

Tab. 5.19: *cal_2comp * k-medie – Roma, d16*

	Tolleranti	Tolleranti con riserva	Conflittuali moderati	Conflittuali	Totale
Conflittuali, non percepiscono vicinanza e competizione (10)	14	0	213	149	376
	3,72	0,00	56,65	39,63	100,00
Tolleranti, percepiscono vicinanza e competizione (01)	131	226	8	1	366
	35,79	61,75	2,19	0,27	100,00
Tolleranti, non percepiscono vicinanza e competizione (00)	231	0	114	0	345
	66,96	0,00	33,04	0,00	100,00
Conflittuali, percepiscono vicinanza e competizione (11)	0	174	29	134	337
	0,00	51,63	8,61	39,76	100,00
Totale	376	400	364	284	1424
	26,40	28,09	25,56	19,94	100,00

I tolleranti restano tolleranti e i conflittuali restano conflittuali: entrambi i gruppi si dividono fra chi, in aggiunta alla tolleranza/conflitto, percepisce anche la vicinanza/competizione, e chi no. Ci interessano di più i gruppi centrali. I tolleranti con riserva sono in larga misura tolleranti che percepiscono vicinanza e competizione, ma in misura quasi uguale, convergono nel gruppo che continua a percepire vicinanza e competizione, ma smette di essere tollerante (con riserva) per diventare conflittuale. Viceversa, ma con stesso andamento, i conflittuali moderati continuano a vedere lo straniero immigrato come un soggetto distante, a costo di entrare nel gruppo dei tolleranti della seconda tipologia. Insomma, l'idea di Cipollini è confermata, la differenza tra i gruppi intermedi non è tanto nella misura di pregiudizio che emerge sulla prima componente, ma nella grande divergenza che c'è sulla seconda, tanto che ci sarebbe da mettere in discussione l'etichettamento originale dei gruppi. D'altronde si vedeva anche nei grafici, soprattutto quelli ripuliti dai casi più ambigui (figure 5.3 e 5.4), dove i gruppi interni ed esterni si dispongono quasi "a croce" sulle prime due componenti. Non possiamo non sottolineare, tra l'altro, come questo aspetto sia stato colto brillantemente attraverso *calibration*, tanto che l'etichetta giusta era già là, scaturita da un incrocio di intensioni, prima ancora che il tipo si riempisse di casi.

Per finire, abbiamo deciso di provare ad usare soltanto la prima componente, sempre calibrata, per costruire una tipologia a quattro gruppi. Il *target set* designato è rimasto lo stesso, i giovani conflittuali, ma sono cambiate le regole di assegnazione e l'etichettamento dei gruppi, nelle modalità che vediamo elencate:

1. Giovani tolleranti ($0 \leq \text{appartenenza} \leq 0,24$);
2. Giovani prevalentemente tolleranti ($0,25 \leq \text{appartenenza} \leq 0,49$);
3. Giovani prevalentemente conflittuali ($0,51 \leq \text{appartenenza} \leq 0,75$);
4. Giovani conflittuali ($0,76 \leq \text{appartenenza} \leq 1$).

Vediamo ora le distribuzioni di frequenza e l'età quadro:

Tab. 5.20: Distribuzioni di frequenza, età – Roma, calibration, d16 (1 comp.)

	N	% Val
Giovani tolleranti	455	31,84
Giovani prevalentemente tolleranti	258	18,05
Giovani prevalentemente conflittuali	288	20,15
Giovani conflittuali	428	29,95
Totale valido	1429	100,00
Missing	13	-
Totale	1442	-
Età quadro	34,95	

La varianza riprodotta è sorprendente, se si considera che stiamo utilizzando soltanto la prima componente, e supera di 6,43 punti percentuali quella della tipologia tradizionale. L'aumento relativo è del 22,55%.

Vediamo ora quanto questa classificazione è stabile rispetto a quella formata con k-medie:

*Tab. 5.21: cal_1comp * k-medie – Roma, d16*

	Tolleranti	Tolleranti con riserva	Conflittuali moderati	Conflittuali	Totale
Giovani tolleranti	281	122	52	0	455
	61,76	26,81	11,43	0,00	100,00
Giovani prevalentemente tolleranti	82	105	70	1	258
	31,78	40,70	27,13	0,39	100,00
Giovani prevalentemente conflittuali	14	110	123	41	288
	4,86	38,19	42,71	14,24	100,00
Giovani conflittuali	0	64	119	245	428
	0,00	14,95	27,80	57,24	100,00
Totale	377	401	364	287	1429
	26,38	28,06	25,47	20,08	100,00

La stabilità è molto bassa, d'altronde c'era da aspettarselo viste le distribuzioni di frequenza molto diverse. Non sono molti i casi che fanno un salto

doppio, come i 64 giovani conflittuali che diventano tolleranti con riserva. La poca stabilità è comunque accettabile in funzione del fatto che sono due procedure molto diverse di per sé, per di più una è costruita sulle prime tre componenti e l'altra solo sulla prima. Quale scegliere? Di nuovo, la decisione è in mano al ricercatore, ci sono vari aspetti in ballo: l'intensione/estensione, la logica classica/fuzzy, le misure calibrate o meno. Come sfondo a tutto ciò, c'è il valore dell'età quadro, che spinge per la nuova tipologia.

L'analisi dei dati relativi alla matrice di Roma è finita, vediamo brevemente i dati relativi agli altri capoluoghi di provincia laziali.

5.3 LA COMPONENTE COGNITIVA DEL PREGIUDIZIO ETNICO: IL CONTESTO DEL LAZIO

Iniziamo dall'analisi delle componenti principali. Come nel caso di Roma, le componenti estratte (secondo il criterio dell'autovalore maggiore di uno) sono tre. Vediamo subito la correlazione tra queste e gli items, visto che ci sono delle piccole differenze che vogliamo analizzare, anche se relative alle sole dimensioni latenti, mentre il primo fattore è del tutto sovrapponibile a quello già visto e si chiamerà, di nuovo, **Ambivalenza** (tab. 5.22).

Nella seconda componente scompare il riferimento alla competizione relativa alle abitazioni, tratto che era stato giustamente interpretato come tipico metropolitano. Nella terza componente gli items che prima erano correlati positivamente (rispetto alla diversità utilitaristica) sono ora correlati negativamente. In più sono correlati positivamente due items relativi alla conflittualità sul piano lavorativo. Vediamo l'andamento di accordo e disaccordo su queste componenti nelle tabelle 5.23 e 5.24.

Tab. 5.22: Correlazione componenti-variabili – Lazio, d16

Items	Correlazioni variabili-fattori		
	1	2	3
Apporto di nuovi stimoli culturali	0,60	0,43	
Comportamenti invadenti	0,65		
Contributo allo sviluppo culturale del paese	0,61	0,48	
Legame eccessivo alle tradizioni	0,35		-0,62
Interessi personali	0,65		-0,30
Originalità	0,38	0,43	
Mancanza di igiene	0,61		-0,30
Introduzione ideologie pericolose	0,71		
Scarsa voglia di lavorare	0,57		
Comportamenti sessuali troppo liberi	0,58		
Sottrazione lavoro agli italiani	0,61		0,30
Numero eccessivo figli	0,55		
Comportamenti di disturbo	0,78		
Mancato rispetto regole di convivenza civile	0,78		
Introduzione nuove malattie	0,59		
Fruizione abusiva sistema di assistenza	0,73		
Problemi di ordine pubblico	0,73		
Non saper stare al proprio posto	0,80		
Contributo allo sviluppo economico del paese	0,50	0,49	
Degrado quartiere	0,74		
Introduzione nuove mode	0,39	0,60	
Diffusione comportamenti contrari tradizione	0,57		
Aumento difficoltà in ricerca abitazioni in affitto	0,58		
Diffusione lavoro irregolare	0,51		0,40
Atteggiamento troppo critico verso società italiana	0,53		

Caduto il riferimento alla competizione, la seconda componente in provincia si caratterizza come semplice **Coinvolgimento/indifferenza** rispetto alla figura sociale dello straniero. Manca, di fatto, una connotazione in termini di schietto pregiudizio, la divisione è piuttosto fra chi vede gli stranieri e li vede come portatori di cambiamento e chi invece non percepisce tutto ciò e si dichiara in disaccordo con tutti gli items a polarità positiva. Chi è sul semiasse positivo “dimostra di non percepire i mutamenti sociali generati dal fenomeno dell’immigrazione e non evidenza, d’altra parte, neanche eventuali aspetti negativi legati al fenomeno dell’immigrazione” (Cipollini, 2004, p. 143). Tale mancanza di aspetti negativi ci ha suggerito, in fase di calibrazione delle componenti, di assegnare l’appartenenza al *target set* a coloro che vedono il mutamento, che si collocano sul semiasse negativo. Questo non significa necessariamente avere pregiudizio, che è la regola che, ove possibile, abbiamo seguito per dare la piena

appartenenza, ma abbiamo scelto di privilegiare la parte, per così dire, “attiva” della dimensione (vedere il mutamento) rispetto a quella passiva (non vedere).

Tab. 5.23: Interpretazione della seconda componente – Lazio, d16

LAZIO	Orientamento	Semiasse neg	Semiasse pos
Apporto di nuovi stimoli culturali	+	Accordo	Disaccordo
Contributo allo sviluppo culturale del paese	+	Accordo	Disaccordo
Originalità	+	Accordo	Disaccordo
Contributo allo sviluppo economico del paese	+	Accordo	Disaccordo
Introduzione nuove mode	+	Accordo	Disaccordo

Il semiasse positivo della terza componente è caratterizzato da disaccordo verso una visione degli stranieri come troppo legati alle loro tradizioni ed ai loro interessi, al contempo c'è accordo verso l'opinione che gli stranieri immigrati sottraggano lavoro agli italiani e ne diffondano di irregolare. La componente è stata, dunque, denominata **Conflittualità sociale**, nel semiasse positivo cadranno i *paritari*, che si sentono in competizione alla pari con gli stranieri. Nel semiasse negativo andranno invece i *separati*, che non avvertono tale competizione alla pari.

Tab. 5.24: Interpretazione della terza componente – Lazio, d16

LAZIO	Orientamento	Semiasse neg	Semiasse pos
Legame eccessivo alle tradizioni	–	Accordo	Disaccordo
Interessi personali	–	Accordo	Disaccordo
Mancanza di igiene	–	Accordo	Disaccordo
Sottrazione lavoro agli italiani	–	Disaccordo	Accordo
Diffusione lavoro irregolare	–	Disaccordo	Accordo

In questo caso abbiamo calibrato la componente nel senso inverso rispetto a Roma. In quel caso, davamo la soglia di piena appartenenza a chi era d'accordo con gli items legati alla diversità utilitaristica. In questo caso, invece, tale soglia va a chi è in disaccordo con questi items. Questo perché, secondo la nostra interpretazione, in un certo senso il pregiudizio esce dalla porta (non hanno interessi personali, non sono troppo legati alle loro tradizioni) per rientrare dal portone (sottraggono lavoro agli italiani, diffondono lavoro irregolare). L'interpretazione potrebbe anche essere inversa, come a dire, si coglie la conflittualità, ma non si associa a tratti negativi. Ad ogni modo, abbiamo optato

per la prima versione e comunque sia la calibrazione, otterremo una tipologia a otto gruppi con tutte le combinazioni possibili.

5.3.1 CLUSTER ANALYSIS TRADIZIONALE E SEMISFOCATA

Procediamo con l'analisi dei gruppi. Tratteremo congiuntamente l'approccio binario e quello *fuzzy*.

I gruppi sono speculari a quelli visti a Roma, hanno valori simili ed identico etichettamento, eccezion fatta per l'ultimo gruppo, che Cipollini ha scelto di chiamare Ostili. Nella sostanza cambia ben poco. Vediamo subito i valori, oltre che per la tipologia tradizionale, per quelle semi-sfocate (i parametri di m ed α sono gli stessi già utilizzati sui dati relativi al contesto romano):

Tab. 5.25: Media e deviazione standard delle tipologie sulla batteria – Lazio, d16

Lazio – d16	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	M	σ	M	σ	M	σ	M	σ
Gruppi								
Tolleranti	2,59	1,30	2,51	1,22	2,55	1,18	2,61	1,14
Tolleranti con riserve	3,13	1,42	3,16	1,36	3,21	1,33	3,34	1,30
Conflittuali moderati	3,67	1,44	3,71	1,36	3,68	1,31	3,76	1,26
Ostili	4,58	1,38	4,57	1,37	4,43	1,36	4,22	1,33
Totale	3,46	1,62	3,48	1,57	3,44	1,50	3,43	1,42

È confermata la tendenza all'aumento dei valori medi e la diminuzione della deviazione standard, nell'ottica di quella opera "pulizia" dei gruppi che in precedenza abbiamo visto anche graficamente.

Passiamo ora alla successiva tavola sintetica, che ci mostrerà, per ognuna di queste 4 tipologie, distribuzioni di frequenza, percentuali valide, iterazioni, stabilità misurata sul confronto con la tipologia tradizionale e per ultimo, il solito valore della media della varianza riprodotta dalle classificazioni sull'informazione relativa a tutta la batteria:

Tab. 5.26: Distribuzioni di frequenza, stabilità ed eta – Lazio, d16

Lazio – d16	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	N	% val	N	% val	N	% val	N	% val
Gruppi								
Tolleranti	401	26,38	331	25,33	277	26,28	219	27,83
Tolleranti con riserve	392	25,79	328	25,10	265	25,14	197	25,03
Conflittuali moderati	368	24,21	323	24,71	269	25,52	204	25,92
Conflittuali	359	23,62	325	24,87	243	23,06	167	21,22
Totale valido	1520	100,00	1307	100,00	1054	100,00	787	100,00
Mancante di sistema	0	-	213	14,01	466	30,66	733	48,22
Iterazioni	12		7		8		20	
Stabilità	-		97,78		97,82		96,70	
Eta quadro	26,49		28,14		25,51		22,37	

Prima considerazione sulle distribuzioni, buone ed equilibrate. Mentre prima il divario tra tolleranti ed ostili andava ad assottigliarsi, qui cresce. Le tipologie sono davvero molto stabili, l'assegnazione a tipi diversi è irrilevante. Quello che balza agli occhi sono però i valori di varianza riprodotta. Per i valori di Roma c'era stato un aumento nella prima e seconda tipologia semi-sfocata ed una di munizione nella terza, che aveva segnato il ritorno sul valore di quella tradizionale. Anche in questo caso vediamo un discreto aumento sulla prima tipologia (1,65%). Il valore cade però drasticamente nella seconda, scendendo sotto il livello di quella tradizionale, per crollare poi nella terza. La semi-sfocatura, in termini di varianza riprodotta, qui non ha funzionato. Non sappiamo dare una spiegazione, forse una sfocatura perfino minore della prima avrebbe funzionato meglio. La "miglioria tecnica" che prima avevamo visto, ora non migliora affatto la situazione, almeno stando fermi sulla scelta di questi parametri. Scelta che, come abbiamo notato e ripetiamo, non ha però criteri di guida.

5.3.2 CALIBRATION

Vediamo come va con la *calibration*. Abbiamo già accennato alla direzione delle calibrations mentre interpretavamo le componenti, vediamo ora un riassunto dei *target set* di ogni componente:

1. Ambivalenza → Giovani ostili;
2. Coinvolgimento/indifferenza → Giovani coinvolti che percepiscono lo straniero come portatore di mutamento;

3. Conflittualità sociale → Giovani che percepiscono conflittualità sociale alla pari con gli stranieri.

Come fatto prima, decidiamo le soglie di appartenenza con l'osservazione dei percentili:

Tab. 5.27: Statistiche descrittive e percentili delle tre componenti

		FAC1_16	FAC2_16	FAC3_16
N	Valid	1520	1520	1520
	Missing	0	0	0
Minimum		-2,24	-4,85	-2,80
Maximum		2,54	3,16	3,81
Percentiles	10	-1,33	-1,27	-1,26
	50	0,00	0,00	0,00
	90	1,32	1,28	1,28

In base all'orientamento degli assi, ne deriva che la calibrazione segue il seguente schema, dove per c161 intendiamo l'appartenenza *fuzzy* al primo *target set* derivata da calibrazione della prima componente, e così via:

$$c161 = \text{calibrate}(f161, 1.32, 0, -1.33)$$

$$c162 = \text{calibrate}(f162, -1.27, 0, 1.28)$$

$$c163 = \text{calibrate}(f163, 1.28, 0, -1.26)$$

Le combinazioni logiche delle appartenenze e non appartenenze agli insiemi di riferimento considerati si traducono in 8 gruppi che hanno la numerosità mostrata nella seguente tavola di verità:

Tab. 5.28: Truth table

C161	C162	C163	number
0	1	1	195
1	0	0	194
1	0	1	194
1	1	0	188
0	1	0	185
0	0	1	183
0	0	0	180
1	1	1	164

Facciamo presente che i tipi, a rigor di lettera, sono già formati intensionalmente, come abbiamo mostrato prima. Per brevità ora li omettiamo o

meglio li mostriamo direttamente nella tabella 5.29, insieme alle relative statistiche, ma ci piace rammentare il carattere intensionale del procedimento.

Tab. 5.29: Distribuzioni di frequenza, eta – Lazio, calibration, d16

Tolleranti, percepiscono lo straniero come fattore di mutamento, percepiscono conflittualità alla pari (011)	N	% Val
Ostili, non percepiscono lo straniero come fattore di mutamento, non percepiscono conflittualità alla pari (100)	195	13,15
Ostili, non percepiscono lo straniero come fattore di mutamento, percepiscono conflittualità alla pari (101)	194	13,08
Ostili, percepiscono lo straniero come fattore di mutamento, non percepiscono conflittualità alla pari (110)	194	13,08
Tolleranti, percepiscono lo straniero come fattore di mutamento, non percepiscono conflittualità alla pari (010)	188	12,68
Tolleranti, non percepiscono lo straniero come fattore di mutamento, percepiscono conflittualità alla pari (001)	185	12,47
Tolleranti, non percepiscono lo straniero come fattore di mutamento, non percepiscono conflittualità alla pari (000)	183	12,34
Ostili, percepiscono lo straniero come fattore di mutamento, percepiscono conflittualità alla pari (111)	180	12,14
Tolleranti, percepiscono lo straniero come fattore di mutamento, percepiscono conflittualità alla pari (011)	164	11,06
Totale valido	1483	100,00
Missing	37	2,43
Totale	1520	-
Eta quadro	33,24	

Quanto detto in riferimento alla *calibration* sulle tre componenti nel contesto metropolitano è più che mai valido anche in questo ultimo caso. La caduta dei casi è bassissima: sono solo 37 i casi che hanno almeno un valore di appartenenza pari al punto di *crossover*. I gruppi sono molto equidistribuiti, la differenza in numerosità tra il primo e l'ottavo gruppo è pari a poco più del 2%. Anche il valore di varianza riprodotta è soddisfacente, segnando un aumento assoluto rispetto al valore registrato con la tecnica delle k-medie (26,49%) del 6,75%. Tradotto in termini relativi è un aumento del 25,48%, la riproduzione dell'informazione iniziale è stata aumentata di un quarto.

Come fatto in precedenza, allo scopo di fare un confronto fra la tipologia scaturita da *cluster analysis* e una tipologia scaturita da *calibration* che abbia lo stesso numero di gruppi, vediamo ora frequenze ed eta quadro medio delle combinazioni derivanti dall'incrocio delle sole prime due variabili calibrate. Eliminando, dunque, la terza dimensione, abbiamo la seguente tipologia:

Tab. 5.30: Distribuzioni di frequenza, eta – Lazio, calibration, d16 (2 comp.)

	N	% Val
Ostili, non percepiscono lo straniero come fattore di mutamento (10)	390	26,07
Tolleranti, percepiscono lo straniero come fattore di mutamento (01)	385	25,74
Tolleranti, non percepiscono lo straniero come fattore di mutamento (00)	365	24,40
Ostili, percepiscono lo straniero come fattore di mutamento (11)	356	23,80
Totale valido	1496	100,00
Missing	24	1,58
Totale	1520	-
Eta quadro	29,94	

Abbiamo dei gruppi di nuovo molto ben distribuiti, mai come in questo caso, dove raggiungono quasi la stessa numerosità. La caduta dei casi è ancora irrilevante. La riproduzione della varianza scende rispetto alla tipologia ad otto gruppi, ma rimane decisamente sopra il valore di quella tradizionale, con un aumento assoluto del 3,45%. Questa classificazione ci consente anche qualche riflessione non statistica. L'indecisione che prima avevamo manifestato sulla direzione da adottare nella calibrazione della seconda componente, cioè in quale senso era l'aumento del pregiudizio, trova risposta radicale: nessuna direzione giusta è giusta. Ci spieghiamo meglio: appare, attraverso questa costruzione, come la dimensione della percezione del mutamento sia del tutto indipendente da connotazioni stereotipate, siano esse in positivo o in negativo. D'altronde, è una dimensione a sé stante dell'ACP, ortogonale a quella dominante. Vediamo ora la stabilità con la tipologia tradizionale, intesa sempre in senso non statistico (tab. 5.31).

Accade qualcosa di molto simile a quanto avevamo visto nel contesto metropolitano. I casi estremi restano sostanzialmente quello che sono (tolleranti o ostili), ma semplicemente si colorano di una proprietà, di un dettaglio in più: l'attribuzione (o meno) allo straniero immigrato del ruolo di motore di cambiamento sociale, economico e culturale. I gruppi interni, invece, mantengono questa seconda connotazione, andando a scambiarsi in parte di posto sulla prima componente. I tolleranti con riserva della prima tipologia sono tolleranti o conflittuali (in parte tolleranti ed in parte conflittuali in se stessi, come suggeriscono i valori di M molto vicini fra loro ed alla media del campione) che, ad ogni modo, non percepiscono lo straniero come fattore di cambiamento. I conflittuali moderati della tipologia classica sono un po' conflittuali e un po'

tolleranti, di certo per loro l'ingresso degli stranieri in Italia è un fatto che cambia le cose.

Tab. 5.31: *cal_2comp * k-medie – Lazio, d16*

	Tolleranti	Tolleranti con riserva	Conflittuali moderati	Ostili	Totale
Ostili, non percepiscono lo straniero come fattore di mutamento (10)	11	109	0	270	390
	2,82	27,95	0,00	69,23	100,00
Tolleranti, percepiscono lo straniero come fattore di mutamento (01)	188	66	131	0	385
	48,83	17,14	34,03	0,00	100,00
Tolleranti, non percepiscono lo straniero come fattore di mutamento (00)	177	187	0	1	365
	48,49	51,23	0,00	0,27	100,00
Ostili, percepiscono lo straniero come fattore di mutamento (11)	22	15	234	85	356
	6,18	4,21	65,73	23,88	100,00
Totale	398	377	365	356	1496
	26,60	25,20	24,40	23,80	100,00

Spesso, leggendo la ricerca, ci eravamo chiesti come fosse visto questo mutamento, un fattore positivo o negativo? L'ultima tipologia ci aiuta a cogliere questa sfumatura, perché ci indica che alcuni lo vedono come un bene (i tolleranti che lo percepiscono, il gruppo 01) ed alcuni come un male (gli ostili che lo percepiscono, il gruppo 11). Anche questo tipo di riflessione, di analisi sui concetti, ci sembra utilmente stimolato dalla classificazione eseguita tramite *calibration*, che col passare del tempo ci convince sempre un po' di più.

A questo punto, per finire, abbiamo costruito una tipologia, sempre a quattro gruppi, sulla calibrazione della sola prima componente, secondo lo schema visto già per Roma:

1. Giovani tolleranti ($0 \leq \text{appartenenza} \leq 0,24$);
2. Giovani prevalentemente tolleranti ($0,25 \leq \text{appartenenza} \leq 0,49$);
3. Giovani prevalentemente ostili ($0,51 \leq \text{appartenenza} \leq 0,75$);
4. Giovani ostili ($0,76 \leq \text{appartenenza} \leq 1$).

Perdendo anche la connotazione relativa al mutamento, tipica della seconda componente, questa tipologia risulta semanticamente più simile a quella tradizionale che, se operativamente è costruita sui tre fattori estratti in ACP, è poi etichettata sostanzialmente sulla base della posizione dei gruppi sul primo fattore. Vediamo subito distribuzioni e varianza riprodotta:

Tab. 5.32: Distribuzioni di frequenza, eta – Lazio, calibration, d16 (1 comp.)

	N	% Val
Giovani tolleranti	490	32,56
Giovani prevalentemente tolleranti	266	17,67
Giovani prevalentemente ostili	284	18,87
Giovani ostili	465	30,90
Totale valido	1505	100,00
Missing	15	0,99
Totale	1520	-
Eta quadro	32,70	

Di nuovo abbiamo una caduta dei casi irrisoria ed una buona equidistribuzione, pur se con uno spiccato aumento dei casi “estremi”. Il valore di eta quadro percentuale è di nuovo molto buono (+6,21%): l’aumento relativo è ancora una volta nell’ordine di un quarto circa (23,44%). Vediamo l’incrocio fra le due tipologie, nella tabella 5.33.

Coerentemente coi dati romani, la stabilità è sostanzialmente bassa. Forse pretendere una buona stabilità da procedure così diverse (una usa *cluster analysis* e tre componenti, l’altra *calibration* e una componente) è volere troppo.

Tab. 5.33: cal_1comp * k-medie – Lazio, d16

	Tolleranti	Tolleranti con riserva	Conflittuali moderati	Ostili	Totale
Giovani tolleranti	276	158	56	0	490
	56,33	32,24	11,43	0,00	100,00
Giovani prevalentemente tolleranti	91	99	75	1	266
	34,21	37,22	28,20	0,38	100,00
Giovani prevalentemente ostili	32	74	114	64	284
	11,27	26,06	40,14	22,54	100,00
Giovani ostili	1	50	120	294	465
	0,22	10,75	25,81	63,23	100,00
Totale	400	381	365	359	1505
	26,58	25,32	24,25	23,85	100,00

Anzi, la poca stabilità potrebbe essere vista come un segno buono, d’altronde stiamo cercando un modo nuovo di classificare, se i risultati sono come quelli del vecchio, sarebbe un esercizio utile? L’analisi della componente cognitiva è, a questo punto, da ritenersi conclusa. Procediamo con quella affettiva.

5.4 LA COMPONENTE AFFETTIVA DEL PREGIUDIZIO ETNICO

Abbiamo visto come la componente affettiva del pregiudizio sia indagata attraverso l'utilizzo di una scala Likert a 8 items a polarità invertita, costruiti secondo il modello del differenziale semantico, con una coppia di espressioni verbali contrastanti agli estremi e nessuna espressione verbale associata alle caselle interne, che sono in totale sei, onde evitare l'attrazione del medio. Tale scala corrisponde alla domanda 19 del questionario, che riportiamo:

Tab. 5.34: scala Likert – componente affettiva del pregiudizio etnico

19. Immagina di trovarti nelle seguenti situazioni. Come ti sentiresti? (Per ciascuna frase indica il quadratino che meglio rappresenta la tua posizione personale rispetto all'una o all'altra risposta).								
a.	Il gruppo di amici che frequenti abitualmente ha deciso di partecipare ad una manifestazione pubblica contro gli immigrati.	<i>Mi darebbe fastidio</i>	<input type="checkbox"/>	<i>Non mi darebbe fastidio</i>				
b.	Nella tua scuola inizierà un corso tenuto da stranieri immigrati per conoscere culture diverse da quella italiana.	<i>Mi interesserebbe</i>	<input type="checkbox"/>	<i>Non mi interesserebbe</i>				
c.	Un tuo familiare sta per sposare uno straniero immigrato	<i>Mi darebbe fastidio</i>	<input type="checkbox"/>	<i>Non mi darebbe fastidio</i>				
d.	Un tuo amico ha invitato un giovane straniero immigrato a venire in vacanza con voi	<i>Mi darebbe fastidio</i>	<input type="checkbox"/>	<i>Non mi darebbe fastidio</i>				
e.	Il tuo locale preferito è sempre più frequentato da stranieri immigrati	<i>Mi darebbe fastidio</i>	<input type="checkbox"/>	<i>Non mi darebbe fastidio</i>				
f.	I tuoi amici vorrebbero far entrare dei giovani stranieri immigrati nel vostro gruppo	<i>Mi farebbe piacere</i>	<input type="checkbox"/>	<i>Non mi farebbe piacere</i>				
g.	Tu e la tua famiglia siete invitati a pranzo a casa di stranieri immigrati	<i>Mi farebbe piacere</i>	<input type="checkbox"/>	<i>Non mi farebbe piacere</i>				
h.	Gli stranieri immigrati che vivono nel tuo quartiere saranno presto allontanati	<i>Mi farebbe piacere</i>	<input type="checkbox"/>	<i>Non mi farebbe piacere</i>				

Come possiamo vedere, la polarità degli items è invertita in due sensi: prima di tutto, nel senso che alcune frasi esprimono sentimenti positivi/di accettazione e alcuni items esprimono sentimenti negativi/di rifiuto; inoltre, alcune risposte sono ordinate nel senso del crescente sentimento negativo (es. mi farebbe piacere/non mi farebbe piacere), altre nel senso inverso (es. mi darebbe fastidio/non mi darebbe fastidio). Qualche che sia l'orientamento, le risposte sono state codificate inizialmente tutte allo stesso modo, le prime caselle hanno avuto un valore 1 e via crescendo fino alle ultime, che hanno valore 6.

Sono stati riscontrati anche qui dei casi in cui si tornava al punto medio, barrando le due caselle centrali. Abbiamo scelto di eliminare dall'analisi questi

casi, codificati con un 7. I ricercatori, tra l'altro, hanno identificato degli errori di risposta, codificato col valore 8. Tali errori non hanno mai superato il 2%, quindi abbiamo escluso anche questi. C'è poi la mancata risposta, molto ridotta. In generale, la caduta delle risposte è sempre stata limitatissima.

Abbiamo quindi operato ricodifica degli items, che è andata nella direzione dell'aumento di pregiudizio, ovviamente identificato con sentimenti negativi e di rifiuto, disinteresse, fastidio. Gli items i cui punteggi sono stati invertiti (1/6, 2/5, 3/4, 4/3, 5/2, 6/1) sono quelli contrassegnati dalle lettere *c*, *d*, *e*, *h*. Ad esempio, l'items:

c. Un tuo familiare sta per sposare uno straniero immigrato *Mi darebbe fastidio* *Non mi darebbe fastidio*

ha, di fatto, assunto una forma come quella che segue:

c. Un tuo familiare sta per sposare uno straniero immigrato *Non mi darebbe fastidio* *Mi darebbe fastidio*

I valori degli items *a*, *b*, *f* e *g* sono rimasti intatti. Siamo stati, a questo punto, pronti ad eseguire l'analisi delle componenti principali sulle variabili così ottenute. In questo paragrafo e nel successivo porteremo avanti di pari passo l'illustrazione dei dati di Roma e degli altri capoluoghi del Lazio, in modo che, se ci saranno differenze, queste balzeranno subito agli occhi e saranno segnalate e commentate immediatamente.

Cominciamo a vedere quanti sono i fattori estratti per Roma ed in quale misura è spiegata la varianza riprodotta dalle 8 domande:

Tab. 5.35: Componenti estratte e varianza spiegata – Roma, d19

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,11	51,42	51,42	4,11	51,42	51,42
2	0,81	10,17	61,59	0,81	10,17	61,59
3	0,63	7,93	69,51			
4	0,62	7,76	77,27			
5	0,58	7,28	84,55			
6	0,46	5,70	90,25			
7	0,42	5,21	95,46			
8	0,36	4,54	100,00			

È evidente la differenza con la componente cognitiva, dovuta al numero molto minore di items, che nella domanda 16 erano più del triplo rispetto alla domanda 19. Le conseguenze sono due: la scala risulta fortemente unidimensionale, tanto che il primo fattore spiega più del 50% della varianza. Seconda conseguenza: solo il primo fattore ha autovalore maggiore di uno. Se seguissimo il criterio seguito per la componente cognitiva (cfr. Kaiser, 1960), dovremmo estrarre solo la prima componente. Questo però impoverirebbe l'analisi, scegliamo così di estrarne due, come d'altronde è stato fatto nella ricerca. Va comunque detto che questo potrebbe in qualche modo inficiare sulla qualità dei dati che andremo ad analizzare, in termini, ad esempio, di difficoltà di interpretazione della seconda componente che potrebbe sembrare contraddittoria o dal significato ambiguo. Non sappiamo se e come potrebbero esserci influenze sull'efficacia delle nostre tipologie. Ad ogni modo, andremo avanti lavorando sulle prime due componenti. Vediamo i fattori estratti nelle altre province:

Tab. 5.36: Componenti estratte e varianza spiegata – Lazio, d19

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,22	52,78	52,78	4,22	52,78	52,78
2	0,79	9,90	62,67	0,79	9,90	62,67
3	0,63	7,90	70,57			
4	0,57	7,07	77,64			
5	0,53	6,57	84,21			
6	0,46	5,75	89,96			
7	0,43	5,39	95,35			
8	0,37	4,65	100,00			

Autovalori e percentuali di varianza riprodotta sono simili, possiamo andare avanti analizzando la correlazione dei fattori con gli items, sempre cominciando dai dati romani, che vediamo esposti nella tabella 5.37. Si conferma la forte unidimensionalità della scala, dato che tutte le variabili sono correlate positivamente con il primo fattore, con un valore minimo di 0,62. Tre variabili sono correlate con la seconda componente, di cui una negativamente. L'unico item con una buona correlazione è comunque quello che riguarda l'ingresso di

uno straniero immigrato nella cerchia delle amicizie, gli altri due sono, infatti, appena sopra la soglia dello 0,3:

Tab. 5.37: Correlazione componenti-variabili – Roma, d19

Items	Correlazioni variabili-fattori	
	1	2
Partecipazione del gruppo ad una manifestazione contro gli stranieri	0,68	
Corso scolastico per conoscere culture diverse	0,69	
Matrimonio di un familiare con uno straniero	0,74	
Vacanza con uno straniero	0,73	
Locale preferito frequentato da stranieri	0,73	-0,32
Stranieri nel gruppo amicale	0,62	0,65
Pranzo a casa di stranieri	0,76	0,36
Allontanamento degli stranieri dal quartiere in cui vivi	0,77	

Per quanto riguarda il Lazio, la situazione è pressoché identica sulla prima componente, sulla seconda c'è un po' di confusione in più. Sono correlati sei items, anziché tre. Anche in questo caso però le correlazioni sono molto basse, addirittura tre valori sono pari a 0,30. L'item più correlato, l'unico che si stacca leggermente dagli altri, è di nuovo quello relativo al gruppo amicale, sarà dunque una buona idea privilegiare quest'ultimo in fase di interpretazione.

Tab. 5.38: Correlazione componenti-variabili – Lazio, d19

Items	Correlazioni variabili-fattori	
	1	2
Partecipazione del gruppo ad una manifestazione contro gli stranieri	0,69	
Corso scolastico per conoscere culture diverse	0,68	0,30
Matrimonio di un familiare con uno straniero	0,73	-0,38
Vacanza con uno straniero	0,74	-0,30
Locale preferito frequentato da stranieri	0,71	-0,36
Stranieri nel gruppo amicale	0,69	0,49
Pranzo a casa di stranieri	0,77	0,30
Allontanamento degli stranieri dal quartiere in cui vivi	0,80	

Riportiamo ora le tabelle che affiancano agli items gli andamenti dei punteggi ad essi relativi sui due semiassi. Se prima, trattandosi di una scala volta ad indagare la componente cognitiva, l'orientamento era stato fatto in termini di accordo/disaccordo, ora, avendo a che fare con reazioni emotive, l'orientamento è dato sull'asse del fastidio/piacere e, solo in un caso, del disinteresse/interesse. Come vediamo nella tabella 5.39, la prima componente è coerente, sul semiassi

negativo vanno tutti coloro che provano sentimenti positivi verso lo straniero e dichiarano predisposizione ad accettarlo:

Tab. 5.39: Interpretazione della prima componente – Roma, d19

ROMA	Asse neg	Asse pos
Partecipazione del gruppo ad una manifestazione contro gli stranieri	Fastidio	Piacere
Corso scolastico per conoscere culture diverse	Interesse	Disinteresse
Matrimonio di un familiare con uno straniero	Piacere	Fastidio
Vacanza con uno straniero	Piacere	Fastidio
Locale preferito frequentato da stranieri	Piacere	Fastidio
Stranieri nel gruppo amicale	Piacere	Fastidio
Pranzo a casa di stranieri	Piacere	Fastidio
Allontanamento degli stranieri dal quartiere in cui vivi	Fastidio	Piacere

Sull'asse positivo, in modo complementare, si trovano tutti i giovani che manifestano sentimenti negativi e rifiuto all'accettazione. Coerentemente con l'elaborazione emersa nella ricerca (il capitolo in questione, il terzo, è stato redatto da Maria Paola Faggiano), tale componente è stata denominata: **Accettazione dello straniero negli spazi materiali e simbolici dell'autoctono/non accettazione**. L'interpretazione è la medesima per l'area del Lazio, come possiamo vedere nella seguente tabella:

Tab. 5.40: Interpretazione della prima componente – Lazio, d19

LAZIO	Asse neg	Asse pos
Partecipazione del gruppo ad una manifestazione contro gli stranieri	Fastidio	Piacere
Corso scolastico per conoscere culture diverse	Interesse	Disinteresse
Matrimonio di un familiare con uno straniero	Piacere	Fastidio
Vacanza con uno straniero	Piacere	Fastidio
Locale preferito frequentato da stranieri	Piacere	Fastidio
Stranieri nel gruppo amicale	Piacere	Fastidio
Pranzo a casa di stranieri	Piacere	Fastidio
Allontanamento degli stranieri dal quartiere in cui vivi	Fastidio	Piacere

Procediamo ora con l'analisi della seconda componente. Per quanto riguarda Roma, nel semiasse negativo aumenta l'espressione di sentimenti positivi verso l'ingresso di stranieri immigrati nella cerchia di amicizie e verso l'idea di essere da questi ospitati a pranzo, insieme con la propria famiglia. Contemporaneamente si avverte un leggero fastidio di fronte ad un aumento della frequentazione straniera del proprio locale preferito. Sul semiasse positivo i

sentimenti sono opposti: è visto di buon occhio un aumento di stranieri nel locale preferito, mentre si preferisce evitare un contatto che abbia luogo nella cerchia delle amicizie o attraverso la famiglia.

Tab. 5.41: Interpretazione della seconda componente – Roma, d19

ROMA	Asse neg	Asse pos
Locale preferito frequentato da stranieri	Fastidio	Piacere
Stranieri nel gruppo amicale	Piacere	Fastidio
Pranzo a casa di stranieri	Piacere	Fastidio

I dati del Lazio si sovrappongono parzialmente alla dimensione vista su Roma, ma con delle aggiunte. Nello stesso semiasse dove emerge principalmente piacere verso l'ingresso di uno straniero nelle amicizie (ci riferiamo soprattutto a questa variabile in quanto era l'unica con una correlazione che si staccava un po' dalla soglia dello 0,30), si prova anche fastidio per una vacanza o per un matrimonio interetnico. Viceversa nell'altro semiasse.

Tab. 5.42: Interpretazione della seconda componente – Lazio, d19

LAZIO	Asse neg	Asse pos
Corso scolastico per conoscere culture diverse	Interesse	Disinteresse
Matrimonio di un familiare con uno straniero	Fastidio	Piacere
Vacanza con uno straniero	Fastidio	Piacere
Locale preferito frequentato da stranieri	Fastidio	Piacere
Stranieri nel gruppo amicale	Piacere	Fastidio
Pranzo a casa di stranieri	Piacere	Fastidio

Questi dati sembrano un po' contraddittori, e qua va ricordato che il fattore non raggiunge l'autovalore di uno e, in secondo luogo, che la maggior parte delle correlazioni è molto debole. A questo punto, privilegiando in particolar modo la correlazione con l'ingresso di stranieri nel gruppo amicale, accettiamo l'interpretazione che nella ricerca si dà alla componente: **Accettazione dello straniero nel sistema di relazioni dell'autoctono/non accettazione.**

5.4.1 CLUSTER ANALYSIS TRADIZIONALE E SEMISFOCATA

A questo punto possiamo applicare le tecniche di classificazione automatica, sia in versione tradizionale (k-medie) sia semi-sfocate. Sceglieremo,

come nella ricerca e come abbiamo fatto finora, la partizione a quattro gruppi. Vediamo prima di tutto dove cadono i centri finali dei *cluster* (siamo ancora nella fase della bivarianza). Abbiamo scelto di vedere in successione le tabelle di Roma e del Lazio per fare un rapido confronto in vista dell'etichettamento dei gruppi:

Tab. 5.43: Centri finali dei gruppi – Roma, k-medie, d19

	1	2	3	4
Accettazione dello straniero negli spazi materiali e simbolici dell'autoctono/non accettazione	-0,76	-0,09	0,52	1,18
Accettazione dello straniero nel sistema di relazioni dell'autoctono/non accettazione	-0,19	1,92	-1,30	0,21

Tab. 5.44: Centri finali dei gruppi – Lazio, k-medie, d19

	1	2	3	4
Accettazione dello straniero negli spazi materiali e simbolici dell'autoctono/non accettazione	-0,80	0,10	0,18	1,44
Accettazione dello straniero nel sistema di relazioni dell'autoctono/non accettazione	-0,06	1,50	-1,30	-0,03

Prima di tutto, ricordiamo che, in entrambe le componenti, il semiasse negativo è quello contraddistinto da maggiore accettazione. I gruppi “esterni” (numero 1 e numero 4) sono sostanzialmente segnati da rispettiva accettazione (gruppo 1: -0,76 su Roma; -0,80 sul Lazio) e non accettazione (gruppo 4: 1,18; 1,44) sulla prima componente. I gruppi “interni” (numero 2 e numero 3) tendono alla non accettazione nel primo caso (gruppo 2: 1,92; 1,50) e accettazione nel secondo (gruppo 3: -1,30; -1,30) sulla seconda componente. Coerentemente, i gruppi sono stati etichettati da Faggiano nel seguente modo:

1. Giovani con sentimenti di apertura/accettazione/accoglienza verso lo straniero;
2. Giovani con sentimenti di apertura rispetto agli spazi materiali e simbolici d'appartenenza e di chiusura rispetto al sistema di relazioni;
3. Giovani con sentimenti di apertura rispetto al sistema di relazioni e di chiusura rispetto agli spazi materiali e simbolici di appartenenza;
4. Giovani con sentimenti di chiusura/rifiuto/repulsione verso lo straniero.

Non siamo del tutto sicuri, però, sulle definizioni dei gruppi centrali. Ad esempio, nel secondo gruppo, mentre è chiara la chiusura sulla seconda

componente (sistema di relazioni), non è affatto chiara l'apertura rispetto alla prima (gli spazi materiali o simbolici). Il secondo gruppo del Lazio in particolare, tende, anche se molto debolmente, al rifiuto anche sulla prima componente.

Insomma, nell'etichettamento dei gruppi centrali, ci pare giustificata la parte che si riferisce al sistema di relazioni (i gruppi sono ben posizionati sulla seconda componente), meno giustificata quella relativa agli spazi materiali e simbolici (la posizione sulla prima componente è nei pressi del centro dell'asse, oltre per quanto riguarda il secondo gruppo laziale).

Il problema è, probabilmente, nella qualità della seconda componente, su cui la maggior parte delle variabili che incidono, lo fanno in modo davvero poco significativo, tanto che l'autovalore è, sia per Roma che per il Lazio, inferiore a uno.

In più, ci differenzia dalla ricerca il fatto che in essa i dati sono stati analizzati insieme, non è stata fatta distinzione di contesto, mentre noi separiamo le matrici. Per cui, forse, i valori delle correlazioni fra componenti e variabili, i centri finali dei gruppi e di conseguenza i punteggi medi sugli items potrebbero variare, anche se lievemente. Nella ricerca questi valori non ci sono, per cui non possiamo fare confronto.

Abbiamo deciso, dunque, di etichettare i gruppi in modo indipendente dalla ricerca e, sull'esempio di quanto visto nell'analisi della scala cognitiva, di farlo prendendo come punto di riferimento solamente la prima componente ed ordinando su di essa i gruppi, consapevoli che quelli centrali sono molto vicini fra loro, mentre sono invece fortemente influenzati da una dimensione latente, che è però espressa da una seconda componente assai poco significativa.

Facciamo presente che in sede di calibrazione, come visto in precedenza, i fattori avranno "pari dignità". Questo ci porta a credere che attribuire un peso specifico ad una componente debole e confusa potrebbe inficiare sulla qualità della classificazione derivante da *calibration*. Sarebbe un po' come costruire uno spazio d'attributi su due concetti, di cui uno a nostro avviso non funziona o è mal definito. Proveremo comunque e vedremo i risultati.

Vediamo ora i valori medi dei nostri gruppi sulla batteria, sia per Roma che per il Lazio. Vedremo contestualmente anche l'etichettamento dei gruppi.

Nelle tabelle provvederemo ad inserire immediatamente anche le tre tipologie semi-sfocate, che mantengono gli stessi parametri di m ed α visti in precedenza.

Tab. 5.45: Media e deviazione standard delle tipologie sulla batteria – Roma, d19

Roma – d19	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	M	σ	M	σ	M	σ	M	σ
Gruppi								
Giovani con sentimenti di accettazione	1,82	1,12	1,70	1,02	1,66	0,98	1,51	0,84
Giovani con sentimenti di parziale accettazione	2,74	1,46	2,74	1,46	2,65	1,42	2,79	1,37
Giovani con sentimenti di parziale non accettazione	3,47	1,54	2,98	1,44	2,92	1,43	2,94	1,37
Giovani con sentimenti di non accettazione	4,33	1,49	4,40	1,46	4,05	1,43	4,38	1,37
Totale	2,82	1,79	2,76	1,75	2,60	1,64	2,62	1,63

Ci sono due tendenze che spiccano in particolar modo. Prima di tutto, la deviazione standard del primo gruppo, i giovani con sentimenti di accettazione, cala drasticamente di pari passo con l'aumento della sfocatura. Secondo, sfumano quasi completamente le differenze nei due gruppi interni, tanto che il secondo gruppo va a finire con la media sopra al campione. Testimone è anche il valore della deviazione standard, che non scende sotto l'1,37. Analizzando le medie, sembrerebbe che i due gruppi interni formino un solo gruppo, sarebbe stata forse migliore una partizione a tre. Guardiamo i dati del Lazio:

Tab. 5.46: Media e deviazione standard delle tipologie sulla batteria – Lazio, d19

Lazio – d19	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	M	σ	M	σ	M	σ	M	σ
Gruppi								
Giovani con sentimenti di accettazione	1,70	1,00	1,56	0,88	1,56	0,87	1,54	0,85
Giovani con sentimenti di parziale accettazione	2,88	1,51	2,83	1,46	2,80	1,45	2,84	1,40
Giovani con sentimenti di parziale non accettazione	2,98	1,46	2,86	1,38	2,81	1,36	2,87	1,34
Giovani con sentimenti di non accettazione	4,60	1,48	4,75	1,44	4,74	1,43	4,76	1,42
Totale	2,74	1,77	2,71	1,74	2,69	1,73	2,69	1,72

L'interpretazione è la stessa. Qua la deviazione standard tocca ha un punteggio di 1,42 perfino nella tipologia a maggiore sfocatura, nel quarto gruppo. Questo gruppo ha, almeno, una media ben distinta ed isolata, quindi, pur nell'eterogeneità, prevale netta la non accettazione. I due gruppi vanno entrambi

sopra la media del campione, con grande variabilità interna. Di contro, il primo gruppo è molto omogeneo e predisposto positivamente verso lo straniero. Vediamo distribuzioni di frequenza, stabilità e varianza riprodotta:

Tab. 5.47: Distribuzioni di frequenza, stabilità ed eta – Roma, d19

Roma – d19	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
Gruppi	N	% val	N	% val	N	% val	N	% val
Giovani con sentimenti di accettazione	690	47,85	597	42,92	563	43,88	448	39,16
Giovani con sentimenti di parziale accettazione	174	12,07	178	12,80	174	13,56	198	17,31
Giovani con sentimenti di parziale non accettazione	213	14,77	283	20,35	253	19,72	281	24,56
Giovani con sentimenti di non accettazione	365	25,31	333	23,94	293	22,84	217	18,97
Totale valido	1442	100,00	1391	100,00	1283	100,00	1144	100,00
Mancante di sistema	0	-	51	3,54	159	11,03	298	20,67
Iterazioni	10		18		19		24	
Stabilità	-		90,94		89,40		76,05	
Eta quadro	43,18		44,44		40,27		46,14	

Balza subito agli occhi la distribuzione squilibrata a livello di numerosità (come risultava pure dalla ricerca). Confermiamo l'idea che fondere i due gruppi centrali in uno solo poteva essere una soluzione valida. La varianza riprodotta è, di contro, costantemente altissima. Crediamo che ciò sia dovuto alla relativa esiguità della varianza iniziale, causata da un numero non troppo elevato di domande. È relativamente bassa anche l'uscita dei casi dalla tipologia con l'aumento della sfocatura, che arriva ad un valore massimo di poco superiore al 20%. È interessantissimo il rapporto tra stabilità e varianza riprodotta: alla diminuzione della stabilità, infatti, aumenta l'eta quadro. Sembra quasi che la sfocatura abbia liberato la tipologia dai vincoli imposti dalla classificazione binaria, ed i casi, più che uscire dalla tipologia, sono stati riassegnati, in un modo che ha arrecato benefici in termini di riproduzione dell'informazione. L'andamento resta piuttosto strano, alla prima sfocatura la varianza riprodotta aumenta di poco, per poi crollare alla seconda sfocatura e risalire ad un livello ben oltre quello tradizionale nell'ultima tipologia semi-sfocata. C'è da dire, comunque, che tutti questi dati vanno presi con le pinze, a causa delle considerazioni che prima abbiamo fatto sul dubbio valore della seconda

componente. Vediamo come si presenta la medesima situazione sulla matrice dati riguardante gli altri capoluoghi di provincia della regione Lazio:

Tab. 5.48: Distribuzioni di frequenza, stabilità ed eta – Lazio, d19

Lazio – d19	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
Gruppi	N	% val	N	% val	N	% val	N	% val
Giovani con sentimenti di accettazione	657	43,22	561	38,45	556	39,07	543	40,43
Giovani con sentimenti di parziale accettazione	275	18,09	313	21,45	299	21,01	277	20,63
Giovani con sentimenti di parziale non accettazione	280	18,42	314	21,52	300	21,08	268	19,96
Giovani con sentimenti di non accettazione	308	20,26	271	18,57	268	18,83	255	18,99
Totale valido	1520	100,00	1459	100,00	1423	100,00	1343	100,00
Mancante di sistema	0	-	61	4,01	97	6,38	177	11,64
Iterazioni	13		18		13		14	
Stabilità	-		90,68		90,51		90,02	
Eta quadro	45,25		48,16		48,61		50,47	

L'andamento è leggermente diverso rispetto a quanto visto sui dati romani. Ad esempio, là c'era stato un rigonfiamento del terzo gruppo, sotto il punto di vista della percentuale valida, che qui manca. Si ritorna ad una soluzione che sembra coerente ed è allo stesso tempo ottimale, visto che la stabilità è sempre nei dintorni del 90% e l'eta quadro va ad aumentare, fortemente tra la tipologia tradizionale e la prima sfocata, trascurabilmente tra questa e la seconda, di nuovo grandemente tra questa e la terza, che arriva ad un aumento superiore al 5% sulla *cluster analysis* tradizionale. Tutto sommato, la *clusterizzazione* semi-sfocata, tra alti e bassi, ha funzionato o quanto meno ha dato risultati interessanti.

5.4.2 CALIBRATION

Cominciamo ora con la calibrazione delle componenti. Sono solo due, quindi in questo caso non avremo uno spazio concettuale rappresentato da un cubo, ma da un quadrato o, se vogliamo, da uno spazio di attributi con due proprietà declinate in termini di presenza ed assenza.

Il primo passaggio è, come al solito, l'individuazione degli insiemi di riferimento e delle soglie di appartenenza. I *target set* per le componenti sono i seguenti:

1. Accettazione dello straniero negli spazi materiali e simbolici dell'autoctono/non accettazione → Giovani che non accettano lo straniero negli spazi materiali/simbolici;
2. Accettazione dello straniero nel sistema di relazioni dell'autoctono/non accettazione → Giovani che non accettano lo straniero nel sistema delle relazioni.

Dunque, secondo la linea che abbiamo scelto, l'appartenenza è sempre designata nei termini dell'aumento del pregiudizio, espresso sulla componente affettiva come espressione di sentimenti negativi, di rifiuto, di non accettazione. Avendo interpretato le componenti allo stesso modo, sia per Roma che per il Lazio, è superfluo dire che i gruppi di riferimento sono gli stessi.

Dato che in entrambe le elaborazioni la non accettazione cade sul semiasse positivo, la soglia di piena appartenenza corrisponderà ad un numero positivo e viceversa. Di nuovo ci serviremo dei percentili. Vediamo le statistiche descrittive delle componenti, per entrambe le matrici:

Tab. 5.49: Statistiche descrittive e percentili delle due componenti

ROMA		FAC1_19	FAC2_19
N	Valid	1442	1442
	Missing	0	0
Minimum		-1,44	-3,47
Maximum		2,52	3,66
Percentiles	10	-1,24	-1,09
	50	-0,10	-0,13
	90	1,46	1,19

Facciamo notare come il secondo quartile, in questo caso, non corrisponda esattamente al punto zero, che però abbiamo scelto di porre comunque come punto di *crossover*. Sulla base di queste statistiche, possiamo vedere le impostazioni della calibrazione per i dati di Roma:

c191=calibrate(f191,1.46,0,-1.24)

c192=calibrate(f192,1.19,0,-1.09)

dove c191 corrisponde alla variabile calibrata di f191, primo fattore estratto dalla componente 19. A seguire le statistiche descrittive delle componenti sulla matrice delle province laziali:

Tab. 5.50: Statistiche descrittive e percentili delle due componenti

LAZIO		FAC1_19	FAC2_19
N	Valid	1520	1520
	Missing	0	0
Minimum		-1,36	-3,79
Maximum		2,56	3,78
Percentiles	10	-1,18	-1,18
	50	-0,17	-0,08
	90	1,50	1,32

I valori per la *calibration* sono ragionevolmente molto simili. Queste sono le impostazioni, dunque:

c191=calibrate(f191,1.5,0,-1.18)

c192=calibrate(f192,1.32,0,-1.18)

Non ci soffermiamo sull'esposizione della *truth table*, il cui significato dovrebbe ormai essere chiaro. Vediamo subito i tipi formati nella tabella che segue, corredati del relativo valore medio di eta quadro percentuale:

Tab. 5.51: Distribuzioni di frequenza, eta – Roma, calibration, d19

	N	% Val
Giovani che accettano lo straniero negli spazi materiali/simbolici e nel sistema di relazioni (00)	506	35,84
Giovani che non accettano lo straniero negli spazi materiali/simbolici né nel sistema di relazioni (11)	333	23,58
Giovani che non accettano lo straniero negli spazi materiali/simbolici ma lo accettano nel sistema di relazioni (10)	313	22,17
Giovani che accettano lo straniero negli spazi materiali/simbolici ma non lo accettano nel sistema di relazioni (01)	260	18,41
Totale valido	1412	100,00
Missing	30	2,08
Totale	1442	-
Eta quadro	41,98	

La prima considerazione è sulle distribuzioni, benché non perfettamente bilanciate, lo sono molto di più rispetto alle tipologie tradizionali che abbiamo visto ed alle relative sfocature. Se consideriamo il peso relativo del secondo fattore, è ragionevole aspettarci di pagare questa distribuzione omogenea in termini di varianza riprodotta, come, di fatto, accade, anche se in termini minori rispetto a quanto ci aspettassimo: a fronte del 43,18% delle k-medie, il disavanzo è solo dell'1,20%. In compenso, la tipologia ci sembra chiara e intellegibile come ci sono sembrate tutte quelle basate su calibrazione delle componenti e tavola della verità. In sede di ricerca, quello che più ci premerebbe in una situazione del genere, lo ripetiamo, sarebbe il miglioramento dei dati relativi al secondo fattore. Vediamo i dati relativi al Lazio:

Tab. 5.52: Distribuzioni di frequenza, eta – Lazio, calibration, d19

	N	% Val
Giovani che accettano lo straniero negli spazi materiali/simbolici e nel sistema di relazioni (00)	488	32,62
Giovani che accettano lo straniero negli spazi materiali/simbolici ma non lo accettano nel sistema di relazioni (01)	337	22,53
Giovani che non accettano lo straniero negli spazi materiali/simbolici ma lo accettano nel sistema di relazioni (10)	337	22,53
Giovani che non accettano lo straniero negli spazi materiali/simbolici né nel sistema di relazioni (11)	334	22,33
Totale valido	1496	100,00
Missing	24	1,58
Totale	1520	-
Eta quadro	42,34	

La situazione è simile a quella vista sui dati romani, anzi la tendenza all'equidistribuzione è più spinta, tanto che i gruppi corrispondenti a quelli intermedi (cioè i gruppi 01 e 10) sorpassano il gruppo caratterizzato da maggiore pregiudizio e si attestano, alla pari, a quota 337 casi. Penseremmo di pagare questa situazione in termini di riproduzione della varianza esterna, ma questa è superiore a quella registrata nell'elaborazione relativa a Roma. La paghiamo però se confrontiamo il dato con quello della *cluster analysis* bivariata sul Lazio: il calo è di quasi il 3%.

Non pensiamo sia un dramma, lo sarebbe se pensassimo che il valore dell'eta quadro fosse l'unica chiave di interpretazione dei risultati. È una chiave, importante, ma non l'unica. C'è una sua diminuzione, ma su valori che rimangono

comunque superiori al 40%, su una scala particolare, con solo 8 items e con l'incidenza di una componente poco chiara e poco significativa, che in un'analisi classica, vincolata al criterio dell'autovalore maggiore di uno, sarebbe stata eliminata. Per contro, rimane la costruzione della tipologia su base intensionale, che è la cosa che più ci interessa, e l'impostazione di fondo che rimanda direttamente alla teoria degli insiemi (sfocati) che apprezziamo molto, con le conseguenti determinazioni dei *target set* ed il dialogo che ne scaturisce in termini di *set relations*.

Vediamo ora la stabilità di queste tipologie calibrate con quelle tradizionali.

Tab. 5.53: *calibration * k-medie – Roma, d19*

	A.	P. A.	P. N. A.	N. A.	Totale
Sì spazi materiali/simbolici	462	0	44	0	506
Sì sistema di relazioni (00)	91,30	0,00	8,70	0,00	100,00
No spazi materiali/simbolici	21	76	0	236	333
No sistema di relazioni (11)	6,31	22,82	0,00	70,87	100,00
No spazi materiali/simbolici	21	0	169	123	313
Sì sistema di relazioni (10)	6,71	0,00	53,99	39,30	100,00
Sì spazi materiali/simbolici	165	95	0	0	260
No sistema di relazioni (01)	63,46	36,54	0,00	0,00	100,00
Totale	669	171	213	359	1412
	47,38	12,11	15,08	25,42	100,00

Anche in questo caso, ci interessa la stabilità non in senso statistico ma interpretativo. Così, ad esempio, ci sembra di trovare sostegno a quell'idea che avevamo avuto, secondo cui l'appartenenza o meno al sistema di relazioni aveva un peso relativo in fase di assegnazione dei casi ai gruppi tradizionali. Questo peso viene, però, valorizzato nel momento in cui la seconda componente è calibrata e diventa, alla pari della prima, fattore determinante nell'incrocio delle possibili combinazioni. Insomma, immessa la variabile calibrata, è normale avere tipi tendenzialmente omogenei. Ma se andiamo a vedere come questi si incrociano con i tipi tradizionali, vediamo che, ad esempio, il gruppo 00 caratterizzato da doppia accettazione rientra, come era lecito aspettarsi, nel gruppo tradizionale dei giovani con sentimenti di accettazione; ma nella sponda dell'accettazione (parziale o non parziale) rientrano anche quei casi che nella *calibration* sono

caratterizzati da rifiuto dello straniero nel sistema delle relazioni, purché ci sia accettazione nello spazio materiale/simbolico. Situazione speculare per la non accettazione.

In poche parole, l'assegnazione alla *cluster analysis* è stata sostanzialmente guidata dalla posizione assunta dai casi su quello che viene definito "spazio materiale/simbolico"; il sistema di relazioni – che, di fatto, si riduce alla correlazione con il solo item riguardante la cerchia dell'amicizia – è quasi del tutto ininfluenza. Non nella calibrazione, però. Da qui, crediamo, scaturisce la caduta in termini di varianza riprodotta. Restano però valide le considerazioni positive sulla "*calibration + truth table*" come procedura, che abbiamo fatto precedentemente. Vediamo i dati laziali:

Tab. 5.54: *calibration * k-medie – Lazio, d19*

	A.	P. A.	P. N. A.	N. A.	Totale
Sì spazi materiali/simbolici	384	0	104	0	488
Sì sistema di relazioni (00)	78,69	0,00	21,31	0,00	100,00
Sì spazi materiali/simbolici	220	117	0	0	337
No sistema di relazioni (01)	65,28	34,72	0,00	0,00	100,00
No spazi materiali/simbolici	14	0	172	151	337
Sì sistema di relazioni (10)	4,15	0,00	51,04	44,81	100,00
No spazi materiali/simbolici	21	157	0	156	334
No sistema di relazioni (11)	6,29	47,01	0,00	46,71	100,00
Totale	639	274	276	307	1496
	42,71	18,32	18,45	20,52	100,00

Il dato è sostanzialmente confermato, anche se in termini leggermente minori. Alcuni casi segnati da rifiuto per lo spazio materiale/simbolico cadono nella sponda dell'accettazione, anche se parziale. Dato che i due gruppi parziali sono arrivati quasi a coincidere (tra l'altro, nei dati laziali il centro del secondo gruppo – parziale accettazione – finiva per cadere di poco nel semiasse positivo, quello del rifiuto), ci sembra un dettaglio trascurabile.

Possiamo considerare conclusa la parte relativa alla componente affettiva del pregiudizio etnico. Manca, ora, soltanto l'analisi della componente attiva; temendo di aver seriamente stancato il lettore, cercheremo di essere ancora più asciutti nell'esposizione.

5.5 LA COMPONENTE ATTIVA DEL PREGIUDIZIO ETNICO

La componente attiva del pregiudizio è stata indaga attraverso l'uso di una scala Likert a 12 items, anche questi a polarità invertita, costruiti anch'essi secondo il modello del differenziale semantico. Agli estremi del supposto *continuum* troviamo la coppia di espressioni verbali “approverei/disapproverei”; nessuna espressione verbale è associata alle caselle interne, che sono sei. Tale scala corrisponde alla domanda 20 del questionario, che riportiamo:

Tab. 5.55: scala Likert – componente attiva del pregiudizio etnico

20. Immagina di avere l'ultima parola sui seguenti provvedimenti da approvare in Italia. Quali approveresti e quali disapproveresti? (Per ciascuna frase indica il quadratino che meglio rappresenta la tua posizione personale rispetto all'una o all'altra risposta).								
a.	Nel tuo quartiere verranno costruiti dei luoghi di culto destinati a stranieri immigrati di confessioni religiose diverse da quella cattolica	Approverei	<input type="checkbox"/>	Disapproverei				
b.	La permanenza degli stranieri immigrati dovrà essere limitata esclusivamente ai periodi in cui lavorano	Approverei	<input type="checkbox"/>	Disapproverei				
c.	Sarà riconosciuto il diritto di voto agli stranieri immigrati che vivono in Italia da alcuni anni	Approverei	<input type="checkbox"/>	Disapproverei				
d.	Per tutti gli stranieri immigrati sarà più difficile ottenere il permesso di soggiorno	Approverei	<input type="checkbox"/>	Disapproverei				
e.	Tutti gli stranieri immigrati saranno rimandati nel loro paese d'origine	Approverei	<input type="checkbox"/>	Disapproverei				
f.	Le forze dell'ordine effettueranno controlli più severi nei quartieri ad alta concentrazione di stranieri immigrati	Approverei	<input type="checkbox"/>	Disapproverei				
g.	Per gli stranieri immigrati regolari sarà più facile ottenere la cittadinanza italiana ed i relativi diritti	Approverei	<input type="checkbox"/>	Disapproverei				
h.	Saranno severamente puniti i comportamenti non rispettosi nei confronti degli stranieri immigrati	Approverei	<input type="checkbox"/>	Disapproverei				
i.	Gli stranieri immigrati dovranno frequentare dei corsi per imparare ad adattarsi al nostro modo di vivere	Approverei	<input type="checkbox"/>	Disapproverei				
j.	Gli stranieri immigrati che trasgrediranno la legge saranno sottoposti a pene più severe rispetto agli italiani	Approverei	<input type="checkbox"/>	Disapproverei				
k.	Per gli stranieri immigrati sarà più facile far arrivare in Italia i membri della loro famiglia	Approverei	<input type="checkbox"/>	Disapproverei				
l.	Gli stranieri immigrati avranno l'opportunità di svolgere ogni tipo di professione	Approverei	<input type="checkbox"/>	Disapproverei				

La codifica è sempre la stessa, da 1 a 6, indipendentemente dalla polarità degli items. Escluse le poche risposte mancanti o errate, ricodifichiamo gli items negativi (*b, d, e, f, i, j*) e ci ritroviamo con un *range* che va da 1 (minimo

pregiudizio) a 6 (massimo). Applichiamo la tecnica dell'ACP ed andiamo subito a vedere le componenti estratte, prima per Roma e poi per il Lazio:

Tab. 5.56: Componenti estratte e varianza spiegata – Roma, d20

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,44	37,01	37,01	4,44	37,01	37,01
2	1,16	9,63	46,64	1,16	9,63	46,64
3	0,91	7,58	54,21			
4	0,80	6,65	60,87			
5	0,74	6,17	67,03			
6	0,68	5,66	72,69			
7	0,67	5,59	78,29			
8	0,61	5,11	83,40			
9	0,58	4,80	88,20			
10	0,52	4,37	92,57			
11	0,47	3,91	96,48			
12	0,42	3,52	100,00			

Tab. 5.57: Componenti estratte e varianza spiegata – Lazio, d20

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,19	34,92	34,92	4,19	34,92	34,92
2	1,27	10,56	45,48	1,27	10,56	45,48
3	0,97	8,06	53,54			
4	0,88	7,32	60,86			
5	0,76	6,32	67,18			
6	0,67	5,62	72,81			
7	0,64	5,34	78,15			
8	0,64	5,33	83,48			
9	0,57	4,79	88,27			
10	0,56	4,68	92,95			
11	0,47	3,94	96,88			
12	0,37	3,12	100,00			

Anche in questo caso, le variazioni sono trascurabili. Torniamo al criterio di estrazione per autovalore maggiore di uno, andremo quindi a considerare nell'analisi solo i primi due fattori. Nella ricerca sono stati presi in considerazione i primi tre fattori, ma ci sganciamo da questa scelta, anche perché, tra l'altro, non ci è dato sapere gli autovalori e la varianza riprodotta dei medesimi, in quanto non sono stati pubblicati. In più, come accadeva con la componente affettiva, nella ricerca le matrici non sono separate. Procediamo con l'interpretazione, andando a vedere la correlazione tra componenti ed items:

Tab. 5.58: Correlazione componenti-variabili – Roma, d20

Items	Correlazioni variabili-fattori	
	1	2
Costruzione di luoghi di culto per religioni diverse	0,65	
Permanenza degli stranieri solo per lavorare	0,59	-0,39
Diritto di voto agli stranieri residenti da alcuni anni	0,65	
Maggiore difficoltà nell'ottenere il permesso di soggiorno	0,63	
Espulsione di tutti gli stranieri	0,78	
Controllo più severi nei quartieri con più stranieri	0,62	-0,44
Cittadinanza più facile per gli stranieri regolari	0,48	
Severamente puniti i comportamenti contro gli stranieri	0,35	0,66
Corsi di adattamento allo stile di vita italiano per gli stranieri	0,43	-0,39
Pene più severe per gli stranieri rispetto agli italiani	0,65	
Maggiore facilità nel ricongiungimento familiare per gli stranieri	0,62	
Possibilità per gli stranieri di svolgere tutte le professioni	0,71	

Tab. 5.59: Correlazione componenti-variabili – Lazio, d20

Items	Correlazioni variabili-fattori	
	1	2
Costruzione di luoghi di culto per religioni diverse	0,63	
Permanenza degli stranieri solo per lavorare	0,56	-0,31
Diritto di voto agli stranieri residenti da alcuni anni	0,67	
Maggiore difficoltà nell'ottenere il permesso di soggiorno	0,59	
Espulsione di tutti gli stranieri	0,81	
Controllo più severi nei quartieri con più stranieri	0,58	-0,47
Cittadinanza più facile per gli stranieri regolari	0,38	
Severamente puniti i comportamenti contro gli stranieri	0,36	0,66
Corsi di adattamento allo stile di vita italiano per gli stranieri	0,34	-0,52
Pene più severe per gli stranieri rispetto agli italiani	0,67	
Maggiore facilità nel ricongiungimento familiare per gli stranieri	0,59	
Possibilità per gli stranieri di svolgere tutte le professioni	0,71	

Le due analisi combaciano, tutti gli items sono correlati positivamente alla prima componente, la scala è dunque fortemente unidimensionale. Tre items “punitivi” sono correlati negativamente alla seconda componente, mentre uno “protettivo” è correlato positivamente, in una combinazione che appare molto interessante.

A scopo di pura conferma, riportiamo l’orientamento dei semiassi sulla prima componente per ogni item, anche se risulta chiara la coerenza del fattore. Dato che non stiamo inserendo valori numerici, abbiamo deciso di mostrare una sola tabella, visto che tra Roma e gli altri capoluoghi di provincia non c’è, effettivamente, alcuna differenza.

Tab. 5.60: Interpretazione della prima componente – d20

	Asse neg	Asse pos
Costruzione di luoghi di culto per religioni diverse	Approvazione	Disapprovazione
Permanenza degli stranieri solo per lavorare	Disapprovazione	Approvazione
Diritto di voto agli stranieri residenti da alcuni anni	Approvazione	Disapprovazione
Maggiore difficoltà nell'ottenere il permesso di soggiorno	Disapprovazione	Approvazione
Espulsione di tutti gli stranieri	Disapprovazione	Approvazione
Controllo più severi nei quartieri con più stranieri	Disapprovazione	Approvazione
Cittadinanza più facile per gli stranieri regolari	Approvazione	Disapprovazione
Severamente puniti i comportamenti contro gli stranieri	Approvazione	Disapprovazione
Corsi di adattamento allo stile di vita italiano per gli stranieri	Disapprovazione	Approvazione
Pene più severe per gli stranieri rispetto agli italiani	Disapprovazione	Approvazione
Maggiore facilità nel ricongiungimento familiare per gli stranieri	Approvazione	Disapprovazione
Possibilità per gli stranieri di svolgere tutte le professioni	Approvazione	Disapprovazione

Come possiamo vedere, sul semiasse negativo sono approvati tutti i provvedimenti favorevoli agli stranieri e disapprovati quelli contrari. Questo semiasse sarà costituito dai giovani favorevoli all'inclusione; quello positivo, viceversa, dai favorevoli all'esclusione. Come nella ricerca (il capitolo è stato redatto da Antonio Regano), etichettiamo questa componente come **Inclusione/esclusione**. Vediamo la seconda componente:

Tab. 5.61: Interpretazione della seconda componente – d20

	Asse neg	Asse pos
Permanenza degli stranieri solo per lavorare	Approvazione	Disapprovazione
Controllo più severi nei quartieri con più stranieri	Approvazione	Disapprovazione
Severamente puniti i comportamenti contro gli stranieri	Approvazione	Disapprovazione
Corsi di adattamento allo stile di vita italiano per gli stranieri	Approvazione	Disapprovazione

Questa componente risulta immediatamente molto interessante. Sulla sponda negativa, si ha l'approvazione di provvedimenti che: permettono agli stranieri la permanenza per i soli periodi di lavoro, istituiscono controlli maggiori nei quartieri ad alta concentrazione straniera, impongono corsi di adattamento allo stile di vita italiano *ma* puniscono più severamente i comportamenti contro gli stranieri. Si trova in questa dimensione ampia traccia di quel fenomeno che è stato definito da Campelli *razzismo pragmatico*¹, una forma di razzismo nuova, sottile e

¹ Durante un intervento sul tema del razzismo in Italia, avvenuto nel giugno del 2003, presso la Protomoteca del Campidoglio, Roma.

molto moderna, che identifica lo straniero come mano d'opera a buon mercato, autorizzata ad entrare nei confini nazionali per lavorare e produrre ma isolata (si ripropone il tema del *ghetto*) e, ultimato il compito, eventualmente espulsa. Questa è una forma di integrazione che è stata definita *inclusione subordinata*: “quanti la perseguono sono favorevoli all’inclusione degli immigrati nel sistema produttivo e indifferenti o contrari alla loro integrazione culturale e politica” (Cotesta, 1999, p. 324). Abbiamo dunque denominato questa componente **Inclusione subordinata/egualitaria**. Procediamo con l’analisi dei gruppi.

5.5.1 CLUSTER ANALYSIS TRADIZIONALE E SEMISFOCATA

Optiamo sempre per una partizione a 4 gruppi. Vediamo i centri finali dei *cluster* scaturiti da tecnica a k-medie tradizionale:

Tab. 5.62: Centri finali dei gruppi – Roma, k-medie, d20

	1	2	3	4
Inclusione/esclusione	-0,91	-0,25	0,20	1,27
Inclusione subordinata/egualitaria	0,05	1,57	-1,00	0,39

Tab. 5.63: Centri finali dei gruppi – Lazio, k-medie, d20

	1	2	3	4
Inclusione/esclusione	-0,86	-0,67	0,59	1,28
Inclusione subordinata/egualitaria	1,25	-0,32	-0,92	0,84

Pur se con valori diversi, c’è una certa coerenza riguardo alla prima componente. Sulla seconda c’è un po’ di confusione in più. In generale ci sembra di poter dire che la tendenza ad un’inclusione subordinata prevalga leggermente nel contesto provinciale, mentre nel contesto metropolitano, in linea anche con quanto detto sulla “distanza cognitiva” riguardo alla componente cognitiva del pregiudizio, prevale una tendenza all’integrazione egualitaria.

Vediamo ora la prossima tavola sintetica, che riporta i punteggi medi e le deviazioni standard dei vari gruppi sugli items, seguendo, come al solito, l’andamento che vede prima la tipologia tradizionale ed a seguire le tipologie semi-sfocate, ordinate per aumento di sfocatura. Di nuovo, riportiamo in

immediata successione, i valori sulla matrice di Roma e su quella delle altre province:

Tab. 5.64: Media e deviazione standard delle tipologie sulla batteria – Roma, d20

Roma – d20	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	M	σ	M	σ	M	σ	M	σ
Gruppi								
Giovani favorevoli a provvedimenti di inclusione	2,14	1,28	1,87	1,13	1,90	1,13	1,89	1,12
Giovani parzialmente favorevoli a provvedimenti di inclusione	2,92	1,71	3,12	1,55	3,10	1,50	3,18	1,50
Giovani parzialmente favorevoli a provvedimenti di esclusione	3,38	1,50	3,16	1,46	3,16	1,46	3,12	1,45
Giovani favorevoli a provvedimenti di esclusione	4,53	1,58	4,54	1,52	4,50	1,52	4,48	1,52
Totale	3,16	1,81	3,12	1,75	3,11	1,72	3,10	1,70

Tab. 5.65: Media e deviazione standard delle tipologie sulla batteria – Lazio, d20

Lazio – d20	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	M	σ	M	σ	M	σ	M	σ
Gruppi								
Giovani favorevoli a provvedimenti di inclusione	2,25	1,51	2,24	1,44	2,27	1,41	2,23	1,36
Giovani parzialmente favorevoli a provvedimenti di inclusione	2,45	1,33	2,46	1,32	2,46	1,32	2,46	1,28
Giovani parzialmente favorevoli a provvedimenti di esclusione	3,76	1,51	3,74	1,49	3,69	1,48	3,66	1,46
Giovani favorevoli a provvedimenti di esclusione	4,45	1,58	4,40	1,57	4,23	1,56	4,10	1,56
Totale	3,14	1,79	3,14	1,75	3,10	1,70	3,08	1,67

Come ci aveva suggerito la lettura dei centri finali dei gruppi, vediamo la conferma di una non trascurabile differenza tra il contesto metropolitano e quello provinciale. Nel Lazio il gruppo favorevole a provvedimenti di inclusione è caratterizzato da valori medi che tendono a rimanere stabili e non particolarmente bassi, mentre il valore medio sullo stesso gruppo, a Roma, cala vistosamente con l'aumento della sfocatura, così come cala la sua deviazione standard. Un'altra differenza: i valori dei gruppi interni nel Lazio tendono a differenziarsi, a Roma tendono ad assomigliarsi. Potrebbe essere corretta la nostra interpretazione, per cui la seconda componente si fa sentire più in provincia, in direzione di una tendenza all'integrazione subordinata, quindi all'aumento di pregiudizio generalizzato.

Vediamo ora le prossime due tavole sintetiche che riportano distribuzioni di frequenza, stabilità e varianza media riprodotta relative alle quattro differenti tipologie:

Tab. 5.66: Distribuzioni di frequenza, stabilità ed età – Roma, d20

Roma – d20	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	N	% val	N	% val	N	% val	N	% val
Giovani favorevoli a provvedimenti di inclusione	487	33,77	343	25,17	320	25,04	288	24,43
Giovani parzialmente favorevoli a provvedimenti di inclusione	190	13,18	244	17,90	228	17,84	225	19,08
Giovani parzialmente favorevoli a provvedimenti di esclusione	449	31,14	487	35,73	463	36,23	436	36,98
Giovani favorevoli a provvedimenti di esclusione	316	21,91	289	21,20	267	20,89	230	19,51
Totale valido	1442	100,00	1363	100,00	1278	100,00	1179	100,00
Mancante di sistema	0	-	79	5,48	164	11,37	263	18,24
Iterazioni	16		19		22		30	
Stabilità	-		82,32		82,47		79,90	
Età quadro	31,45		32,69		31,56		30,32	

Tab. 5.67: Distribuzioni di frequenza, stabilità ed età – Lazio, d20

Lazio – d20	K-medie		Fuzzy I		Fuzzy II		Fuzzy III	
	N	% val	N	% val	N	% val	N	% val
Giovani favorevoli a provvedimenti di inclusione	262	17,24	271	18,89	242	18,33	201	17,28
Giovani parzialmente favorevoli a provvedimenti di inclusione	564	37,11	501	34,91	474	35,91	410	35,25
Giovani parzialmente favorevoli a provvedimenti di esclusione	415	27,30	387	26,97	348	26,36	314	27,00
Giovani favorevoli a provvedimenti di esclusione	279	18,36	276	19,23	256	19,39	238	20,46
Totale valido	1520	100,00	1435	100,00	1320	100,00	1163	100,00
Mancante di sistema	0	-	85	5,59	200	13,16	357	23,49
Iterazioni	11		10		10		14	
Stabilità	-		95,82		94,47		93,29	
Età quadro	31,48		30,62		28,03		27,14	

Intanto sono piuttosto impressionanti le differenze fra le due distribuzioni. Il gruppo più numeroso a Roma è quello dei giovani favorevoli all'approvazione di provvedimenti di inclusione. Questo gruppo va però fortemente ridimensionandosi con l'aumento della sfocatura, tanto che nell'ultima tipologia, a fronte di 263 casi non assegnati, il primo gruppo diminuisce di 199 unità. I gruppi più numerosi in provincia sono quelli interni, dove con molte sfumature convivono inclusione ed esclusione. Sembra comunque confermarsi la divisione tra metropoli e provincia sul tratto della distanza cognitiva. I cittadini metropolitani sono, in qualche misura, stranieri essi stessi. La stabilità è molto alta nel Lazio, meno a Roma. La semi-sfocatura, in questo caso, non funziona in

termini di varianza riprodotta, soprattutto nel Lazio: a Roma ci sono due piccoli aumenti ed una piccola diminuzione; nel Lazio solo diminuzioni, ragguardevoli nelle ultime due tipologie.

5.5.2 CALIBRATION

Proseguiamo ora con l'ultima calibrazione su questa componente. Il processo inizia dalla scelta dei *target set* e delle soglie. I criteri sono ancora una volta: appartenenza al gruppo portatore di pregiudizio e ausilio dei percentili. Vediamo di seguito i gruppi di riferimento e le statistiche descrittive delle componenti che andiamo a calibrare:

1. Inclusione/esclusione → Giovani favorevoli a provvedimenti di esclusione;
2. Inclusione subordinata/egualitaria → Giovani favorevoli ad una integrazione subordinata.

Tab. 5.68: Statistiche descrittive e percentili delle due componenti

ROMA		FAC1_20	FAC2_20
N	Valid	1442	1442
	Missing	0	0
Minimum		-1,94	-2,43
Maximum		2,54	4,62
Percentiles	10	-1,31	-1,23
	50	-0,05	-0,06
	90	1,44	1,29

Tab. 5.69: Statistiche descrittive e percentili delle due componenti

LAZIO		FAC1_20	FAC2_20
N	Valid	1520	1520
	Missing	0	0
Minimum		-1,96	-2,36
Maximum		2,64	4,27
Percentiles	10	-1,23	-1,21
	50	-0,07	-0,06
	90	1,39	1,33

Queste sono dunque le impostazioni per la calibrazione delle componenti sulla matrice di Roma:

$$c201 = \text{calibrate}(f201, 1.44, 0, -1.31)$$

$$c202 = \text{calibrate}(f202, -1.23, 0, 1.29)$$

Queste sono, invece, quelle relative ai dati delle altre province:

$$c201 = \text{calibrate}(f201, 1.39, 0, -1.23)$$

$$c202 = \text{calibrate}(f202, -1.21, 0, 1.33)$$

Le tipologie che derivano dalla combinazione di appartenenza e non appartenenza, secondo il modello delle *truth tables*, sono le seguenti:

Tab. 5.70: Distribuzioni di frequenza, eta – Roma, calibration, d20

	N	% Val
Giovani favorevoli a provvedimenti di inclusione ma con integrazione subordinata (01)	381	26,77
Giovani favorevoli a provvedimenti di esclusione ma con tendenze di integrazione subordinata (11)	365	25,65
Giovani favorevoli a provvedimenti di inclusione con integrazione egualitaria (00)	360	25,30
Giovani favorevoli a provvedimenti di esclusione senza tendenze di integrazione subordinata (10)	317	22,28
Totale valido	1423	100,00
Missing	19	1,32
Totale	1442	-
Eta quadro	32,57	

Tab. 5.71: Distribuzioni di frequenza, eta – Lazio, calibration, d20

	N	% Val
Giovani favorevoli a provvedimenti di inclusione ma con integrazione subordinata (01)	419	28,14
Giovani favorevoli a provvedimenti di esclusione ma con tendenze di integrazione subordinata (11)	373	25,05
Giovani favorevoli a provvedimenti di inclusione con integrazione egualitaria (00)	370	24,85
Giovani favorevoli a provvedimenti di esclusione senza tendenze di integrazione subordinata (10)	327	21,96
Totale valido	1489	100,00
Missing	31	2,04
Totale	1520	-
Eta quadro	31,60	

Ci ripetiamo per l'ultima volta. Le tipologie scaturite dall'uso congiunto di *calibration* e *truth table* mostrano ancora tutto il loro potenziale investigativo, grazie alla struttura a forma di *property space* che le caratterizza. Una struttura che permette di indagare puntualmente le combinazioni di concetti e, grazie a questa tecnica, di associare a tali tipi intensionali l'assegnazione automatica di parti di campioni o popolazioni, grazie ad una tecnica rodada e, in seguito allo sviluppo di un ottimo *software* come fsQCA, facilmente accessibile. Anche in quest'ultimo caso, abbiamo una tipologia molto ben distribuita, una caduta irrisoria delle risposte e dei tipi già interpretati e maneggevoli. In più, l'età quadro, anche se di poco, aumenta.

5.6 UNA BREVE ANALISI COMBINATA

Siamo sostanzialmente giunti alla fine del nostro lavoro. Appena prima di chiudere, vogliamo fare un'altra brevissima applicazione, totalmente diversa da quelle viste finora. Abbiamo, fin qui, calibrato le varie componenti estratte tramite ACP. Vogliamo ora trattare gli items in modo diverso, non attraverso riduzione fattoriale, ma attraverso indice additivo.

Ricordando che abbiamo a che fare con delle scale Likert, ci mettiamo subito in guardia dal trarre conclusioni indebite citando le parole di Marradi:

non si possono sommare numeri ordinali (1 + 2 fa 3 ma primo più secondo non fa terzo); se si sommano, si stanno usando come numeri cardinali le etichette numeriche attribuite a semplici espressioni di (dis)accordo (Marradi, 2007, p. 147).

Ciò detto, procediamo spinti da semplice curiosità. Il primo passo è la costruzione degli indici additivi. Una volta ripuliti gli items, esattamente come avevamo fatto prima di trattarli con tecniche fattoriali, dobbiamo semplicemente sommarli. Avremo così tre indici che hanno i seguenti *range*:

Componente cognitiva: 25...150;

Componente affettiva: 8...48;

Componente attiva: 12...72.

Come *target set* generico per tutte e tre le variabili calibrate che formeremo, poniamo quello dei “*Giovani conflittuali*”.

Scegliamo, dunque, le soglie di calibrazione sempre con l’uso del 10° e 90° percentile; in più questa volta utilizzeremo il 50° percentile come punto di *crossover*. Vediamo le statistiche relative ai due contesti oggetto d’analisi e le rispettive regole di calibrazione:

Tab. 5.72: Statistiche descrittive e percentili degli indici additivi

Roma		Indice additivo d16	Indice additivo d19	Indice additivo d20
N	Valid	1254	1351	1344
	Missing	188	91	98
Minimum		28	8	12
Maximum		150	48	72
Percentiles	10	53	10	21
	50	88	21	37
	90	123	38	57

cia16=calibrate(ia16,123,88,53)

cia19=calibrate(ia19,38,21,10)

cia20=calibrate(ia20,57,37,21)

Tab. 5.73: Statistiche descrittive e percentili degli indici additivi

Lazio		Indice additivo d16	Indice additivo d19	Indice additivo d20
N	Valid	1357	1449	1451
	Missing	163	71	69
Minimum		30	8	12
Maximum		150	48	72
Percentiles	10	54	9	22
	50	86	20	37
	90	120	37	55

cia16=calibrate(ia16,120,86,54)

cia19=calibrate(ia19,37,20,9)

cia20=calibrate(ia20,55,37,22)

Non abbiamo rinominato i tipi, abbiamo preferito lasciarli espressi sotto forma di codice numerico, ricordando che ogni codice corrisponde ad una combinazione. Il tipo 010, ad esempio, sarà un tollerante sulla componente cognitiva, conflittuale sulla componente affettiva e tollerante sulla componente attiva. Vediamo la distribuzione su Roma e sul Lazio, associata all'eta quadro (che abbiamo calcolato considerando tutti gli items delle tre scale– 45 items – e facendo la media):

Tab. 5.74: Distribuzioni di frequenza, eta – Roma, calibration, i.a.

		Frequency	Percent	Valid Percent
Valid	(111)	392	27,18	37,69
	(000)	372	25,80	35,77
	(011)	59	4,09	5,67
	(010)	55	3,81	5,29
	(100)	51	3,54	4,90
	(101)	45	3,12	4,33
	(110)	35	2,43	3,37
	(001)	31	2,15	2,98
	Total	1040	72,12	100,00
Missing	System	402	27,88	
Total		1442	100,00	
Eta quadro	33,36			

Tab. 5.75: Distribuzioni di frequenza, eta – Lazio, calibration, i.a.

		Frequency	Percent	Valid Percent
Valid	(000)	425	27,96	37,09
	(111)	402	26,45	35,08
	(011)	63	4,14	5,50
	(100)	59	3,88	5,15
	(010)	52	3,42	4,54
	(101)	49	3,22	4,28
	(001)	48	3,16	4,19
	(110)	48	3,16	4,19
	Total	1146	75,39	100,00
Missing	System	374	24,61	
Total		1520	100,00	
Eta quadro	31,92			

Il valore dell'eta quadro è indicato a solo scopo indicativo, non ne deduciamo nulla, visto che gli indici additivi non sono variabili cardinali. Le distribuzioni sono quanto mai diverse da quelle viste finora. Il motivo è presto

detto: abbiamo, fin qui, calibrato componenti principali, che sono ortogonali per definizione. L'uso dello spazio geometrico/concettuale che abbiamo osservato è dovuto, appunto, a questo. Possiamo vedere, ora, come cambia la rappresentazione grafica in 3d rispetto alle ultime tipologie:

Fig. 5.9: Rappresentazione in 3d – calibration, i.a., angolo (111)

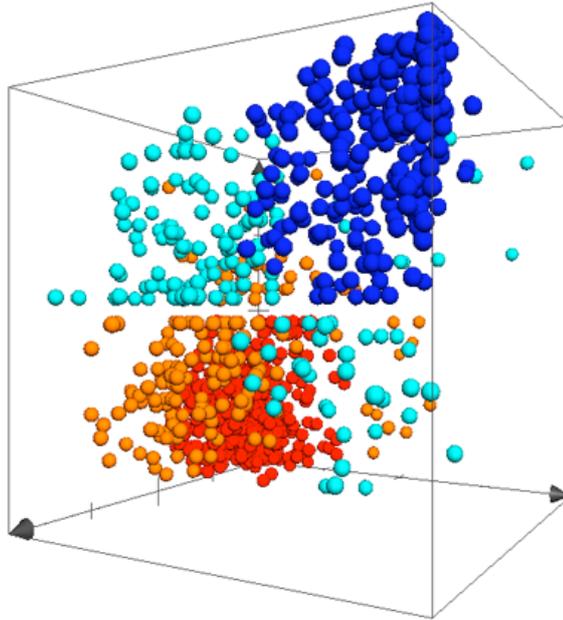
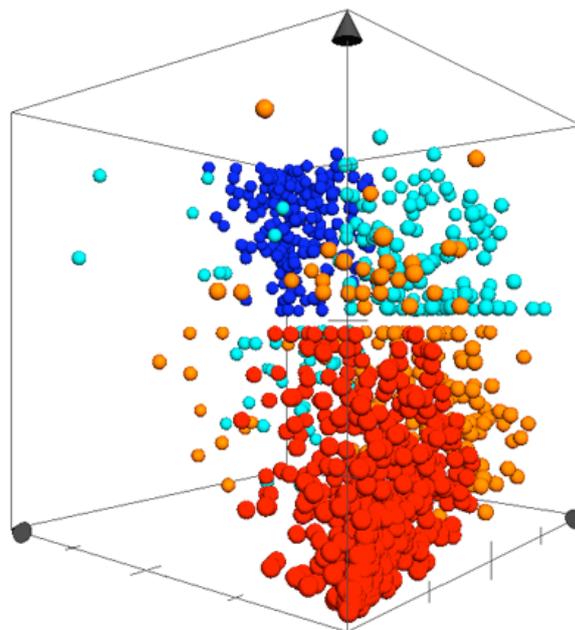


Fig. 5.10: Rappresentazione in 3d – calibration, i.a., angolo (000)



Il gruppo 000, quello dei tolleranti su tutte le scale, è colorato in rosso; i conflittuali, 111, in blu; in arancio i gruppi dove si alternano due 0 e un 1; in celeste, due 1 e uno 0. I casi si dispongono in modo molto interessante, sostanzialmente giacciono sulla diagonale che va dal massimo al minimo conflitto. È una rappresentazione singolare, ma – per i motivi già detti – non significativa. Diverso sarebbe se si riuscisse ad utilizzare variabili cardinali.

5.7 CONCLUSIONE

Abbiamo ripetuto le nostre considerazioni più di una volta, per cui saremo estremamente rapidi. Partendo dal presupposto che aderiamo ad una visione *fuzzy* delle cose, abbiamo provato ad applicare questa visione alla struttura concettuale della classificazione, che costituisce la base della conoscenza, scientifica e non.

L'avvicinamento è stato graduale, l'approdo più immediato era rappresentato dall'applicazione di qualcuno dei molti algoritmi *fuzzy* esistenti per la *cluster analysis*. Avevamo provato questo tentativo già in nostri lavori precedenti. Dopo un iniziale entusiasmo, ci eravamo raffreddati. Per due motivi: il primo era una generale insoddisfazione verso le tecniche di *cluster analysis tout court*. Il secondo era il fatto che, in questo modo, l'aspetto *fuzzy* interveniva solo in un secondo momento, dopo che era già stata fatta un'analisi dei gruppi bivariata, con il solo ruolo secondario di ripulire i gruppi dalle assegnazioni troppo forzate. Ruolo meritorio, certo, ma speravamo in qualcosa in più.

La scoperta della *calibration* (unita all'uso di *truth table*) ci ha offerto questa possibilità. La *fuzziness* interviene in un momento decisivo, andando a modificare le variabili stesse che useremo per classificare, e non la classificazione ultimata. Apprezziamo le misure calibrate per la loro intelligibilità ed utilità. Con questa procedura, riusciamo a costruire tipologie *bypassando* le tecniche di *cluster analysis*, che ci lasciano leggermente insoddisfatti. In più, costruiamo tipologie *intensionali* e controlliamo il significato di esse attraverso uno strumento tanto semplice quanto fondamentale ed utile come lo *spazio di attributi*.

Per tutti questi motivi, senza dimenticare il rendimento in termini di varianza riprodotta che spesso è stato molto buono ed alcune volte ottimo, ci riteniamo pienamente soddisfatti dell'utilizzo della *calibration* – anche al di fuori dal contesto metodologico dove essa si è sviluppata – come mezzo per classificare. Esprimiamo, infine, la speranza che questo tentativo non rimanga isolato e che, premesse le critiche a cui qualsiasi lavoro di un “apprendista ricercatore” va incontro, qualcuno trovi questo percorso interessante e degno di un'altra possibilità.

RIFERIMENTI BIBLIOGRAFICI

- AA.VV., 1987, *Nuovo dizionario enciclopedico Sansoni*, Firenze, Sansoni
- AA.VV., 1993, *Classificazione*, The New Encyclopaedia Britannica, Micropaedia
- ABEL T., 1948, *The operation called Verstehen*, American Journal of Sociology, LIV
- ADANSON M., 1763, *Familles des plantes*, Paris.
- AGNOLI M.S., 1997, *Concetti e pratica nella ricerca sociale*, 2^a ed., Milano, Franco Angeli
- AGNOLI M.S., 2004, *Il disegno della ricerca sociale*, Roma, Carocci
- AGNOLI M.S., (a cura di), 2004, *Lo straniero in immagine*, Milano, Franco Angeli
- AGODI M.C., 1999, *Lazarsfeld e la "natura" della classificazione nelle scienze sociali*, Sociologia e ricerca sociale, 58-59
- AGODI M.C., 2001, *Pensare/classificare: appunti sugli obiettivi della classificazione nelle scienze sociali*, Sociologia e ricerca sociale, 64
- ALLPORT G. W., 1954, *The Nature of Prejudice*, Cambridge, Addison Wesley; tr. it., 1973, *La natura del pregiudizio*, Firenze, La Nuova Italia
- AMISANO E., RINALDI G. e PAMPANIN E., 2002, *Sono equidistanti le categorie di una scala Likert?*, in Marradi A. e Gasperoni G., 2002, *Costruire il dato 3. Le scale Likert*, Milano, Angeli, pp. 95-122
- ANDERBERG M., 1973, *Cluster analysis for application*, New York, Academic Press
- AURELI E. e MIDDEI F., 1997, *Dispense sulla clusterizzazione sfocata*, Roma, Centro Stampa Ateneo

- BAILEY K.D., 1973, "*Monothetic and polythetic typologies and their relation to conceptualization, measurement and scaling*", *American Sociological Review*, 38, 1, pp. 18-33
- BAILEY K.D., 1984, "*A three-level measurement*", *Quality and Quantity*, 18, pp. 225-45
- BAILEY K.D., 1994, *Typologies and taxonomies: an introduction to classification techniques*, Thousand Oaks, Sage Publication
- BAUMAN Z., 1999, *La società dell'incertezza*, Bologna, Il Mulino
- BEER S., 1949, *Classificazione*, Enciclopedia Cattolica, Firenze, Sansoni
- BERGER J. E., ZELDITCH M., 1968, *Review of T. Parsons' Sociological Theory and Modern Society*, in "*American Sociological Review*", XXXIII, 3, pp. 446-450.
- BERGER P. L., LUCKMANN T., 1969, *La realtà come costruzione sociale*, Bologna, Il mulino
- BEZDEK J. C., 1981, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York
- BIORCIO R., 1993, *L'analisi dei gruppi*, Milano, Angeli
- BLALOCK H. M., 1960, *Social Statistics*, New York; tr. it., 1970, *Statistica per la ricerca sociale*, Bologna, il Mulino
- BLASHFIELD R. K., 1980, *The growth of cluster analysis: Tryon, Ward, and Johnson*, *Multivariate Behavioral Research*
- BLASHFIELD R. K., ALDENDERFER M. S., 1978, *The literature on cluster analysis*, *Multivariate Behavioral Research*
- BOUDON R. e BOURRICAUD F., 1982, *Dictionnaire critique de la sociologie*, Paris, Press Universitaires de France; tr. it., 1991, *Dizionario critico di sociologia*, Milano, Armando
- BRADY H. E., COLLIER D. (a cura di), 2004, *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham, Md.: Rowman & Littlefield
- BRUNELLI C., 2000, *La classificazione fuzzy nella ricerca sociale: pregi e limiti di un nuovo strumento per l'analisi dei dati*, *Sociologia e ricerca sociale*, 61

- BRUNELLI C., 2001, *La logica fuzzy nell'analisi dei gruppi: criteri e possibilità*, Sociologia e ricerca sociale, 64
- BRUSCHI A., 1990, *Conoscenza e metodo*, Milano, Mondadori
- BRUSCHI A., 1993, *La metodologia povera. Logica causale e ricerca sociale*, Roma, Nuova Italia Scientifica
- CALOGERO G., 1968, *I fondamenti della logica aristotelica*, Firenze, La nuova Italia
- CAMPELLI E., 1991, *Il metodo e il suo contrario*, Milano, Angeli
- CAMPELLI E., (a cura di), 1999, *T.S. Khun: come mutano le idee sulla scienza*, Milano, Franco Angeli
- CAMPELLI E., 2001, *Tohu va-vohu. Note non tecniche sul problema della classificazione*, Sociologia e ricerca sociale, 64
- CAMPELLI E., 2002, *Prefazione*, in Cipollini R. (a cura di), 2004, *Stranieri. Percezione dello straniero e pregiudizio etnico*, Milano, FrancoAngeli
- CAMPELLI E., 2004, *Prefazione*, in Agnoli M.S., (a cura di), 2004, *Lo straniero in immagine*, Milano, Franco Angeli
- CAPECCHI V., 1964, *I modelli di classificazione e l'analisi della struttura latente*, in "Quaderni di Sociologia", XIII, 3, pp. 289-340
- CAPECCHI V., 1966, *Typologies in Relation to Mathematical Models*, in "Ikon", n. 58 (luglio-settembre), pp. 1-62
- CAPECCHI V. e MOELLER F., 1968, *Some Applications of Entropy to the Problems of Classification*, in "Quality & Quantity" II, 1-2, pp. 63-84
- CARITAS DI ROMA, 2003, *Immigrazione. Dossier statistico 2003*, Roma, Anterem
- CECCHINI A., 1967, *Classificazione*, Grande Dizionario Enciclopedico Utet, fondato da Pietro Fedele, Terza edizione interamente riveduta e accresciuta, Torino, Utet
- CIPOLLINI R., 2000, *Lo straniero come forma sociale*, in Macchietti S. S. (a cura di), *Itinerari pedagogici e culturali. Scritti in onore di R. Vallini*, Siena, Cantagalli, pp. 445-470
- CIPOLLINI R. (a cura di), 2002, *Stranieri. Percezione dello straniero e pregiudizio etnico*, Milano, FrancoAngeli

- CIPOLLINI R., 2004, *La componente cognitiva del pregiudizio etnico: cognizioni, credenze e conoscenze sugli stranieri*, pp. 74-121, in Agnoli M.S., (a cura di), 2004, *Lo straniero in immagine*, Milano, Franco Angeli
- CLIFF N., 1988, *The eigenvalues-greater-than-one rule and the reliability of components*, Psychological Bulletin
- COHEN M. R. e NAGEL E., 1934, *An Introduction to Logic and Scientific Method*, New York
- COLOMBO E., 1999, *Rappresentazioni dell'Altro. Lo straniero nella riflessione sociale occidentale*, Milano, Guerini
- CORBETTA P., 1999, *Metodologia e tecniche della ricerca sociale*, Bologna, Il Mulino
- COTESTA V., 1999, *Sociologia dei conflitti etnici*, Roma-Bari, Laterza
- CRONQVIST L., 2005, *An Introduction to Multi-Value QCA*. Didactical Paper, Mimeo, Marburgo
- DE FELICE e DURO, 1993, *Vocabolario italiano*, Torino, S.E.I./Palombo
- DE GIACOMO P., RESNIK S., PIERRI G., 1980, *Psicologia medica e psichiatria clinica e dinamica*, Padova, Piccin
- DeMEUR G., RIHOUX B., 2002: *L'analyse quali-quantitative comparée*, Louvain-La-Neuve, Bruyland Academi
- DEWEY J., 1938, *Logic, the theory of inquiry*, New York, Holt and Co.; tr. it., 1974, *Logica, teoria dell'indagine*, Torino, Einaudi
- DILTHEY W., 1949, *Introduzione alle scienze dello spirito*, Torino, Einaudi
- DI FRANCO G., 1997, *Tecniche e modelli di analisi multivariata dei dati. Introduzione all'applicazione per la ricerca sociale*, Roma, Seam
- DURKHEIM E., 1963, *Le forme elementari della vita religiosa*, Milano, Comunità
- DURKHEIM E., 1969, *Il suicidio; L'educazione morale*, Torino, UTET
- DURKHEIM E. e MAUSS M., 1901, *De quelques formes primitives de classification*, in "L'année sociologique", VI, pp. 1-71 (tr. ingl.: *Primitive Classification*, London 1963).
- EINSTEIN A. e INFELD L., 1938, *The Evolution of Physics*, New York, Simon and Schuster; tr. it, *L'evoluzione della fisica*, Torino, Boringhieri, 1965

- EDWARDS A. W. F., CAVALLI-SFORZA L. L., 1965, *A method for cluster analysis*, Biometrics
- ERMINIO D., 2001, *Contributi della logica fuzzy alla classificazione*, Sociologia e ricerca sociale, 64
- FAGGIANO M. P., 2004, *La componente affettiva del pregiudizio etnico: sentimenti ed emozioni verso gli stranieri immigrati*, pp. 182-206, in Agnoli M.S., (a cura di), 2004, *Lo straniero in immagine*, Milano, Franco Angeli
- FERNANDEZ J. W., 1986, *Persuasion and Performance: The Play of Tropes in Culture*, Bloomington, Indiana University Press
- FERRAROTTI F., 2000, *La tentazione dell'oblio. Razzismo, antisemitismo e neonazismo*, Roma-Bari, Laterza
- FEYERABEND P.K., 1984, *Contro il metodo. Abbozzo di una teoria anarchica della conoscenza*, Milano, Feltrinelli
- FINDLAY J. N., 1994, *Platone: le dottrine scritte e non scritte*, Milano, Vita e pensiero
- FLECK L., 1983, *Genesi e sviluppo di un fatto scientifico : per una teoria dello stile e del collettivo di pensiero*, Bologna, Il mulino
- FOUCAULT M., 1966, *Les mots et les choses*, Paris, Editions Gallimard; tr. it., 1996, *Le parole e le cose. Un'archeologia delle scienze umane*, Milano, Rizzoli
- FOX J., 1982, *Selective Aspects of Measuring Resemblance for Taxonomy*, in HUDSON, H. C. (ed.), *Classifying Social Data*, San Francisco, pp. 127-151
- FRAIRE M., 1994, *Metodi di analisi multidimensionale dei dati*, Roma, Cisu
- GALTUNG J., 1967, *Theory and Methods of Social Research*, London, Allen & Unwin
- GELLNER E. A., 1964, *Concepts*, in Gould J. e Kold W. L. (a cura di), *A Dictionary of the Social Sciences*, Glencoe, Free Press, pp. 120-121
- GIDDENS A., 1976, *New rules of Sociological Method: a Positive Critique of Interpretative Sociologies*, London, Hutchinson; tr. it., 1979, *Nuove regole del metodo sociologico*, Bologna, Il Mulino

- GIL F., 1981, *Sistematica e classificazione*, Enciclopedia Einaudi
- GLASER B. G., 1978, *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*, Mill Valley
- GOERTZ G., 2003, *The Substantive Importance of Necessary Condition Hypotheses*, in: Goertz G., Starr H., (a cura di), *Necessary Conditions*, Lanham, ecc.: Rowman & Littlefield Publishers, pp. 65-94
- GOERTZ G., 2005, *Necessary Condition Hypotheses as Deterministic or Probabilistic: Does It Matter?*, in *Qualitative Methods. Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, 3 (1), pp. 22-27
- GOERTZ G., MAHONEY J., 2005, *Two-Level Theories and Fuzzy Set Analysis*, in *Sociological Methods and Research*, 33 (4), pp. 497-538
- GOLDTHORPE J. H., 1997: *Current Issues in Comparative Macrosociology: A Debate on Methodological Issues*, in *Comparative Social Research*, 16, pp. 1-26
- GREENBERG J. M., 1957, *The Nature and Uses of Linguistic Typologies*, in "International Journal of American Linguistics", XXIII, 2, pp. 68-72
- HARTIGAN J. A., 1975, *Clustering algorithms*, New York Wiley
- HEMPEL C.G., 1952, *Fundamentals of concept formation in empirical science*, The University of Chicago Press; tr. it., 1976, *La formazione dei concetti e delle teorie nella ricerca empirica*, Milano, Feltrinelli
- HEMPEL C.G., 1961, *Fundamentals of Taxonomy*, in Zubin J. (a cura di), *Field Studies in Mental Disorder*, New York, Grune & Stratton
- HEMPEL C.G., 1965, *Fundamentals of Taxonomy*, in HEMPEL, C. G., *Aspects of Scientific Explanation*, Glencoe, pp. 137-154; tr. it. 1965, *Aspetti della spiegazione scientifica*, Milano
- HEMPEL C. G., OPPENHEIM P., 1936, *Der Typusbegriff im Lichte der neuen Logik*, Leyden
- HENNIG W., 1950, *Grundzüge einer Theorie der phylogenetischen Systematik*, Berlin, Deutscher Zentralverlag
- HENNIG W., 1979, *Phylogenetic Systematics*, Urbana, University of Illinois Press

- HUANG Z., 1998, *Extensions to the K-means Algorithm for Clustering Large Datasets with Categorical Values*. Data Mining and Knowledge Discovery, 2, p. 283-304.
- IACOVACCI G., 1995, *Sull'utilizzo del metodo delle c-medie sfocato per la classificazione dei comuni italiani a seconda del grado di urbanità e ruralità*, in Statistica Applicata, vol. 7, n. 1, pp. 33-48.
- JAIN A. K., DUBES R. C., 1988, *Algorithms for clustering data*, Prentice Hall Advanced Reference Series
- JAIN A. K., MURTY M. N., FLYNN P. J., 1999, *Data clustering: a review*, ACM computing surveys
- JEVONS W. S., 1874, *The principles of science : a treatise on logic and scientific method*, London, Macmillan
- JORDAN N., 1965, *The Asimmetry of "Liking" and "Disliking": A Phenomenon Meriting Further Reflection and Research*, in Public Opinion Quarterly, XXIX, 2 (estate), pp. 315-322
- KAISER H. F., 1960, *The application of electronic computers to factor analysis*, Educational and psychological measurement
- KING P. J., MAMDANI E. H., 1977, *The application of fuzzy control systems to industrial processes*, Automatica
- KUHN T.S., 1962, *The structure of scientific revolutions*, Chicago, University of Chicago press; tr. it., 1978, *La struttura delle rivoluzioni scientifiche. Come mutano le idee della scienza*, Torino, Einaudi
- KOSKO B., 1991, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall
- KOSKO B., 1993, *Fuzzy thinking: the new science of fuzzy logic*, Hyperion; tr. it., *Il fuzzy-pensiero. Teoria e applicazioni della logica fuzzy*, Milano, Baldini&Castoldi
- LAITIN D., 1992, *Language Repertiores and State Construction in Africa*, New York, Cambridge University Press
- LAKOFF G., 1987, *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*, Chicago, University of Chicago Press

- LATOUR B., 1987, *Science in action: how to follow scientists and engineers through society*, Milton Keynes, Open university press; tr. it., 1998, *La scienza in azione: introduzione alla sociologia della scienza*, Torino, Edizioni di Comunita
- LAZARSELD P.F., 1937, *Some Remarks on the Typological Procedures in Social Research*, in "Zeitschrift für Sozialforschung", VI, pp. 119-139
- LAZARSELD P.F., 1967, *Metodologia e ricerca sociale*, Bologna, Il mulino
- LAZARSELD P.F. e BARTON A., 1951, *Qualitative measurement in the social science: classification, typologies and indices*, in Lerner D., Lasswell H.D. (eds), *The policy sciences*, Stanford, Stanford University Press; tr. it., *Classificazioni, tipologie e indici*, in LAZARSELD P.F., 1967
- LIKERT R., 1932, *The Method of Constructing an Attitude Scale*, in *A Technique for the Measurement of Attitudes*, in *Archives of Psychology*, monogr. N. 140, pp. 44-53
- LIVOLSI M., (a cura di), 1992, *Il pubblico dei media*, Firenze, La Nuova Italia
- LOCKE J., 1690, *An Essay Concerning Human Understanding*, London, Basset; tr. it., 1951, *Saggio sull'intelletto umano*, Bari, Laterza
- LOFLAND L., 1973, *A World of Strangers. Order and Action in Urban Public Space*, New York, Basic Books
- MacQUEEN J. B., 1967, *Some Methods for classification and Analysis of Multivariate Observations*, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297
- MACKIE J. L., 1974, *The cement of the universe: a study of causation*, Oxford, Clarendon Press
- MAHONEY J., 2003, *Strategies of Causal Assessment in Comparative Historical Analysis*, in Mahoney J., Rueschemeyer D. (a cura di), *Comparative Historical Analysis in the Social Sciences*, Cambridge, Cambridge University Press, pp. 337-372
- MAHONEY J., 2000, *Strategies in Causal Inference in Small-N Analysis*, in: *Sociological Methods and Research*, 28, 4, pp. 387-424

- MALINOWSKI B., 1944, *A Scientific Theory of Culture and Other Essays*, Chapel Hill
- MAMDANI E. H., 1977, *Applications of fuzzy set theory to control systems: a survey*, Fuzzy Automata and Decision Processes
- MARRADI A., 1984, *Concetti e metodi per la ricerca sociale*, Firenze, La giuntina
- MARRADI A., 1984a, *Teoria: una tipologia dei significati*, in *Metascienza. Riflessioni sulla conoscenza scientifica*, Sociologia e Ricerca Sociale, n. 13, pp. 157-81
- MARRADI A., 1991, *Misurazione e scale: qualche riflessione e una proposta*, in Cardano M., Miceli R. (a c. di), *Il linguaggio delle variabili*, Torino, Rosenberg & Sellier
- MARRADI A., 1993, *Classificazioni, tipologie, tassonomie*, Roma, Enciclopedia delle scienze sociali, Enciclopedia italiana fondata da Giovanni Treccani, II
- MARRADI A., 1994, *Referenti, pensiero e linguaggio: una questione rilevante per gli indicatori*, Sociologia e ricerca sociale, XV, 43, pp. 137-207
- MARRADI A., 2000, *La tipologia da Aristotele alle scienze umane moderne*, in AA.VV., *Una facoltà nel Mediterraneo*, Milano, Giuffré, 2000, pp. 183-201
- MARRADI A., 2002, *Le scale Likert e la reazione all'oggetto*, in Marradi A. e Gasperoni G. (a cura di), 2002, *Costruire il dato 3. Le scale Likert*, Milano, Angeli, pp. 15-52
- MARRADI A., 2007, *Metodologia delle scienze sociali*, Bologna, Il Mulino
- MARRADI A. e GASPERONI G. (a cura di), 2002, *Costruire il dato 3. Le scale Likert*, Milano, Angeli
- MAUCERI S., 2003, *Per la qualità del dato nella ricerca sociale. Strategie di progettazione e conduzione dell'intervista con questionario*, Milano, Angeli
- MAY R. W., 1982, *Discriminant Analysis in Cluster Analysis*, in HUDSON, H. C. (ed.), *Classifying Social Data*, San Francisco, pp. 39-55.
- MENDELEEV D. I., 1994, *Il sistema periodico degli elementi*, Roma, Teknos

- MICELI R., 1995, *Appartenenze multiple*, Torino, Eidos
- MILL J. S., 1843, *A System of Logic, Ratiocinative and Inductive*, London, Longmans
- MORICONI A., 2000, *La classificazione nelle scienze sociali*, Roma, Edizioni Seam
- MORLINO L., 2005, *Introduzione alla ricerca comparata*, Bologna, Il Mulino
- MOSCOVICI S., 1984, *Il fenomeno delle rappresentazioni sociali*, in Farr M., Moscovici S., 1984, *Social Representations*, Cambridge, Cambridge University Press; tr. it., 1989, *Le rappresentazioni sociali*, Bologna, Il Mulino
- OSHERSON D. e SMITH E., 1981, *On the Adequacy of Prototype Theory as a Theory of Concepts*, in "Cognition", IX, 1, pp. 35-56
- NIETZSCHE F., 1972, *Sull'utilità e il danno della storia per la vita, Considerazioni inattuali, II*, in *Opere di Friedrich Nietzsche*, a cura di G. Colli e M. Montanari, Milano, Adelphi
- OSGOOD C. E., 1952, *The Nature and Measurement of Meaning*, in *Psychological Bulletin*, XLIX, 3 (maggio), pp. 197-237
- PAPPIS C. P., MAMDANI E. H., 1977, *A Fuzzy Logic Controller for a Traffic Junction*, *IEEE Transactions on Systems, Man and Cybernetics*
- PARSONS T., 1951, *The Social System*, Glencoe, Free Press
- PAVSIC R., PITRONE M. C., 2003, *Come conoscere opinioni e atteggiamenti*, Acireale, Bonanno
- PAWSON R., 1989, *A Measure for Measures: A Manifesto for Empirical Sociology*, New York, Routledge
- PIAGET J. e INHELDER B., 1959, *La genèse des structures logiques élémentaires chez l'enfant: classifications et sériations*, Neuchâtel
- PITRONE M. C., 1984, *Il sondaggio*, Milano, Angeli
- POPPER K., 1959, *The logic of scientific discovery*, London, Hutchinson; tr. it., 1970, *Logica della scoperta scientifica*, Torino, Einaudi
- POPPER K., 1969, *Scienza e filosofia*, Torino, Einaudi
- QUINE W. V., 1981, *Theories and things*, Harvard University Press
- RADFORD A. E. e altri, 1974, *Vascular Plants Systematics*, New York

- RAGIN C., 1987, *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*, University of California Press
- RAGIN C., 1991, *Issues and Alternatives in Comparative Social Research*. E.J. Brill. Edited collection; also published as special issue of *International Journal of Comparative Sociology* 32 (1 and 2)
- RAGIN C., 1992, *What Is a Case? Exploring the Foundations of Social Inquiry*. Edited collection, with Howard S. Becker. Cambridge University Press
- RAGIN C., 1994, *Constructing Social Research: The Unity and Diversity of Method*. Pine Forge Press
- RAGIN C., 1994, *Formal Methods of Qualitative Analysis*. Edited collection, with Larry Griffin; special issue of *Sociological Methods and Research* 23 (1)
- RAGIN C., 2000, *Fuzzy-Set Social Science*, University of Chicago Press
- RAGIN C., 2004: *Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research*, in Brady H. E., Collier D. (a cura di), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham, Md.: Rowman & Littlefield, pp. 123-138
- RAGIN C., 2005, *Fuzzy Sets*. Edited collection, with Paul Pennings; special Issue of *Sociological Methods and Research* 33 (May).
- RAGIN C., 2007, *Fuzzy Sets: Calibration Versus Measurement*, COMPASSSS Working Papers, Mimeo, Tucson
- RAGIN C., 2008, *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. (edited volume, with Benoît Rihoux).
- RAGIN C., 2008, *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press
- RAGIN C., 2009, *Handbook of Case-Based Research*. (edited volume, with David Byrne). Sage (in production).
- RAGIN C., BECKER H. S., (a cura di), 2001, *What is a Case? Exploring the Foundations of Social Inquiry*, Cambridge, Cambridge University Press
- RAGIN C., DRASS K., DAVEY S., 2006. *Fuzzy-Set/Qualitative Comparative Analysis 2.0*. Tucson, Arizona: Department of Sociology, University of Arizona.

- REALE G., 2000, *Introduzione a Aristotele*, 11^a ed., Roma, Laterza
- REGANO A., 2004, *La componente attiva del pregiudizio etnico: quali provvedimenti per gli stranieri immigrati?*, pp. 207-232, in Agnoli M.S., (a cura di), 2004, *Lo straniero in immagine*, Milano, Franco Angeli
- REICHENBACH H., 1938, *Experience and Prediction*, Chicago, University of Chicago Press
- RICOLFI L., 1989, *Sul mito delle k Italie. Una critica delle tipologie degli italiani*, Polis, III, 3, pp. 445-78
- RICOLFI L., 1992, *Helga*, Milano, Angeli
- ROETHLISBERGER F.J. e DICKSON W.J., 1939, *Management and the worker*, Cambridge, Massachusetts, Harvard University Press, pp. 248-268
- ROKKAN S., 1975, *Dimensions of State Formation and Nation Building: A Possible Paradigm for Research on Variations Within Europe*, in Tilly C. (a cura di), *The Formation of Nation States in Western Europe*, Princeton, Princeton University Press.
- SANDRI G., 1969, *Contributo all'analisi dei procedimenti classificatori*, Revue internationale de sociologie, 5, 3, pp. 473-80
- SANDRI G., 1969, *On the Logic of Classification*, in "Quality & Quantity" III, 1-2, pp. 80-124
- SARTORI G., 1970, *Concept Misformation in Comparative Politics*, in "American Political Science Review", LXIV, 4, pp. 1033-53
- SARTORI G., 1971, *La politica comparata: premesse e problemi*, in Rivista Italiana di Scienza Politica, I, 1, (aprile), pp. 7-66
- SARTORI G., 1979, *La politica. Logica e metodo in scienze sociali*, Milano, SugarCo
- SARTORI G., 1984, *Foreword*, in Sartori G. (a cura di), *Social Science Concepts. A Systematic Analysis*, London, Stage
- SAUSSURE F. de, 1916, *Cours de linguistique générale*, Paris
- SCHEFFLER I., 1967, *Science and Subjectivity*, Indianapolis; tr. it., 1983, *Scienza e soggettività*, Roma
- SCHUTZ A., 1979, *Saggi sociologici*, Torino, UTET

- SEAWRIGHT J., COLLIER D., 2004, *Glossary of Selected Terms*, in Brady H. E., Collier D. (a cura di), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham, Md.: Rowman & Littlefield, pp. 273-313
- SELYE H., 1964, *From Dream to Discovery*, New York, McGraw-Hill
- SMELSER N. J., 1976, *Comparative Methods in the Social Science*, Englewood Cliffs, Prentice-Hall
- SMITHSON M., 1987, *Fuzzy set analysis for behavioral and social sciences*, New York, Springer-Verlag
- SMITHSON M., VERKUILEN J., 2006, *Fuzzy set theory applications in the social sciences*, Thousand Oaks, Sage
- SKOCPOL T., 1984, *Emerging Agendas and Recurrent Strategies in Historical Sociology*, in Skocpol T. (a cura di), *Visions and Methods in Historical Sociology*, Cambridge/New York, Cambridge University Press, pp. 356-391
- SNEATH P. H. A., 1957, *Some Thoughts on Bacterial Classification*, in “Journal of General Microbiology”, XVII
- SOKAL R. R., 1961, *Distance as a Measure of Taxonomic Similarity*, in “Systematic Zoology”, X, 1: 70-79
- SOKAL R. R. e SNEATH P. H. A., 1963, *The principles of numerical taxonomy*, San Francisco, W.H. Freeman
- SPENCER H., 1892, *The Principles of Sociology*, London
- STATERA G., 1994, *Logica dell'indagine scientifico-sociale*, Milano, Angeli
- STOUFFER S.A. et al., 1949, *The American Soldier*, vol. I e II, Princeton N.J., Princeton University Press
- TABBONI S., 1986, *Vicinanza e lontananza. Modelli e figure dello straniero come categoria sociologica*, Milano, Angeli
- TIRYAKIAN E. A., 1968, *Typologies*, in International Encyclopedia of the Social Sciences, vol. XVI, London & New York, pp. 177-185
- THOLEN D. J., 1989, *Asteroid taxonomic classifications* in Asteroids II, pp. 1139-1150, University of Arizona Press
- TORALDO DI FRANCIA G., 1986, *Le cose e loro nomi*, Roma-Bari, Laterza
- TRYON R.C., 1939, *Cluster Analysis*, New York, Mac Graw Hill

- TRYON R. C. e BAILEY D. E., 1970, *Cluster Analysis*, Mac Graw Hill
- VINK M., Van VLIET O., 2007, *Not Quite Crisp, Not Yet Fuzzy? Assessing the Potentials and Pitfalls of Multi-Value QCA*, Mimeo, Maastricht/Leiden
- WAGEMANN C., 2007, *QCA e 'Fuzzy Set Analysis'*. *Che cosa è e che cosa non è*, in *Rivista italiana di scienza politica*, XXXVII, 3, pp. 385-410
- WAGEMANN C., SCHNEIDER C., 2007: *Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy Sets*, Mimeo, Firenze/Budapest
- WALDNER D., 2005, *It Ain't Necessarily So – Or Is It?*, in *Qualitative Methods. Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, 3 (1), pp. 27-29
- WEBER M., 1904, *Die Objektivität sozialwissenschaftlicher und sozialpolitischer Erkenntnis*, in *Archiv für Sozialwissenschaft und Sozialpolitik*, XIX, pp. 22-87; tr. it., *L'"oggettività" conoscitiva della scienza sociale e della politica sociale*, in WEBER M., 1958, *Il metodo delle scienze storico-sociali*, a cura di P. Rossi, Torino, pp. 53-141
- WILSON D.E., REEDER D.M., 2005, *Mammal Species of the World. A Taxonomic and Geographic Reference*. 3a edizione, Johns Hopkins University Press
- WINCH P., 1972, *Il concetto di scienza sociale e le sue relazioni con la filosofia*, Milano, Il Saggiatore
- WINCH R.F., 1947, "Heuristic and empirical typologies: a job for factor", *American Sociological Review*, 12, pp. 68-75
- WOOD M. M., 1934, *The Stranger. A Study in Social Relationships*, New York, Columbia University Press
- ZADEH L.A., 1965, *Fuzzy Sets*, *Information and Control*, 8, 338-353
- ZIFF P., 1960, *Semantic Analysis*, Ithaca, Cornell University Press