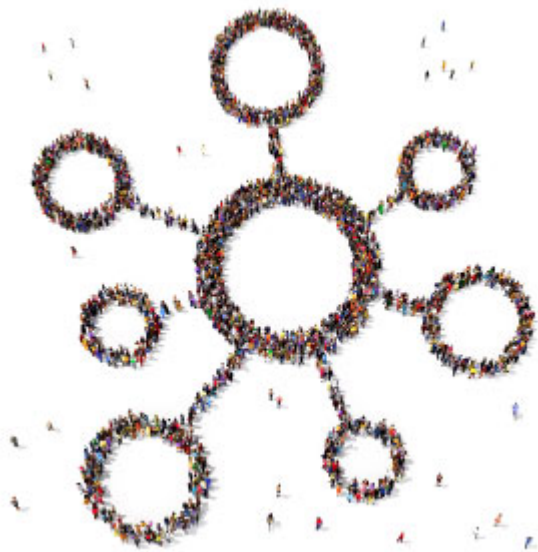


La statistica e i gruppi sociali

scritto da Giorgio Alleva | 25 Luglio 2017



La statistica e i gruppi sociali

Il Rapporto Annuale dell'Istat propone una lettura della situazione del Paese attraverso le lenti di un tema di rilievo nel dibattito corrente, valorizzando il nostro patrimonio informativo. Quest'anno abbiamo voluto analizzare l'articolazione sociale, mentre due anni fa ci eravamo concentrati sulla nuova geografia che emerge dalla revisione dei Sistemi locali, e l'anno scorso - 90° anniversario dell'Istituto - su una nuova tassonomia delle generazioni della nostra società.

L'analisi dell'articolazione sociale ha ricevuto grande attenzione, sia in ambito scientifico sia sui mezzi di informazione, a testimoniare l'importanza dell'argomento scelto dall'Istat. Non sono mancate le critiche e, anzi, soprattutto le critiche fanno assomigliare questo dibattito a quello che si sviluppò all'inizio del secolo scorso attorno alla fisica quantistica.

È un'analogia formale, ovviamente, dovuta al problema della misurazione. Nella fisica quantistica infatti non si potevano più usare le misure della fisica classica: di un elettrone se misuravi la velocità non potevi conoscere la posizione, e viceversa. Non valeva cioè più il concetto di traiettoria classica.

Un po' come succede adesso quando si cerca di rappresentare la società usando soltanto le variabili classiche dell'occupazione e del reddito. Il problema in fisica lo risolse Heisenberg introducendo un nuovo metodo di misurazione, ovvero la moltiplicazione delle indeterminazioni in maniera empirica, da qui il collegamento con "l'albero", un metodo che combinando più variabili associate al reddito in maniera empirica viene utilizzato nel Rapporto per determinare i gruppi sociali.

Questa analogia formale suggerisce che quella attuale è una sorta di società quantistica caratterizzata dalla frammentazione e dall'incertezza.

L'obiettivo di ricerca e l'unità d'analisi

L'analisi ha mirato a rappresentare la segmentazione della società in termini di reddito, e di alcune altre caratteristiche correlate alla generazione del reddito su base familiare.

L'unità d'analisi è la famiglia, perché le risorse disponibili vengono condivise al suo interno, ma nella procedura di suddivisione dell'universo delle famiglie italiane si è dovuto ridurre le configurazioni familiari possibili. Il risultato è che alcune variabili di segmentazione sono riferite all'individuo che ha il reddito più elevato nell'ambito della famiglia, ovvero la *persona di riferimento*.

Per la multidimensionalità degli elementi considerati e, insieme, la prospettiva centrata sulle risorse economiche a livello familiare nel Rapporto si parla di gruppi e non di classi sociali.

La posizione professionale – che è il criterio ovvio di suddivisione a priori della società in classi – resta un elemento importante dell'analisi, ma non l'unico, né la chiave di lettura fondamentale.

La scelta metodologica è stata realizzata attraverso un metodo statistico mirato a conseguire una classificazione delle famiglie residenti in gruppi. Ciascun gruppo è stato associato a una combinazione distintiva di caratteristiche comuni correlate col reddito familiare e non già con la previsione del reddito familiare in funzione di tali caratteristiche.

Disponendo di un'ipotesi relativamente forte (il reddito è influenzato da un insieme di altre variabili), si è scartato il campo delle tecniche di classificazione non supervisionata, quali la *cluster analysis*, eventualmente in combinazione con l'analisi fattoriale.

Si è dunque prescelta la tecnica nota come 'albero di decisione'. Tra i vantaggi, quello di evitare ipotesi a priori sulla distribuzione della popolazione (tecnica non-parametrica) e di produrre un output relativamente facile da interpretare e da presentare visivamente. La fonte primaria utilizzata è la Rilevazione Eu-Silc sul reddito e le condizioni di vita delle famiglie, che l'Italia conduce annualmente dal 2004 su un campione di 29 mila famiglie per un totale di quasi 70 mila componenti, intervistati su base individuale. È un quindi un campione di ampiezza tale da consentire analisi robuste a un livello fine di disaggregazione.

Il riferimento è in parte obbligato (Eu-Silc è la fonte primaria per i redditi), ma anche suggerito dalle disponibilità di variabili atte a caratterizzare sia il principale percettore di reddito sia la tipologia familiare.

Nell'analisi, la variabile dipendente scelta è dunque il reddito medio familiare equivalente. Quali variabili indipendenti si sono invece considerate: situazione professionale, cittadinanza, titolo di studio, sesso ed età della persona di riferimento della famiglia, oltre al numero di componenti della famiglia e il tipo di comune di residenza.

Va segnalato che a fini classificatori non sono state utilizzate tutte le variabili disponibili nell'indagine, perché si è voluto che la classificazione ottenuta fosse poi applicabile anche ad altre fonti statistiche (tra le altre, le indagini sulle forze di lavoro o le indagini multi-scopo sugli aspetti di vita quotidiana), un fatto necessario per realizzare analisi tematiche che fossero riferite agli stessi gruppi sociali. L'algoritmo utilizzato, xCHAID (Exhaustive Chi-Square Automatic Interaction Detector), adotta la metafora dell'albero: i gruppi sono le foglie, i punti in cui lo spazio si biforca sono nodi, connessi dai rami. Nel nostro caso, l'obiettivo è quello di dividere lo spazio multidimensionale delle famiglie italiane in insiemi distinti e non sovrapposti.

Ad ogni passo, l'algoritmo sceglie la suddivisione che massimizza l'omogeneità della variabile dipendente all'interno dei nuovi gruppi creati, rispetto a tutti gli altri possibili. Si ripete l'algoritmo ricorsivamente (ogni gruppo ottenuto a un certo passo diventa il nodo da dividere al passo successivo).

La tecnica adottata non predetermina il numero di gruppi possibili, ma questo numero cresce esponenzialmente: ci siamo fermati a 9, intervenendo anche a 'potare' l'albero come la letteratura

suggerisce di fare, per trovare un equilibrio tra l'esigenza di sintesi – un numero limitato di gruppi ne facilita la lettura – e quella di avere una moltitudine di gruppi molto omogenei al loro interno.

La statistica è utile per operare la sintesi di milioni di comportamenti quotidiani e decisioni individuali; il ritratto collettivo non è meno vero del primo piano del singolo.

I risultati ottenuti

Si capisce quindi come, una volta selezionate le variabili esplicative, a definire i gruppi sia l'algoritmo e non una scelta a priori del ricercatore.

Queste sono risultate: la partecipazione al lavoro della 'persona di riferimento' nella famiglia ('principale percettore di reddito' secondo Eu-Silc), la professione svolta e il tipo di contratto, la cittadinanza, la dimensione familiare, il titolo di studio conseguito. Sono tutte variabili con un riscontro nella letteratura e tradizionalmente considerate nelle analisi dell'Istat.

La loro combinazione e il ruolo che svolgono nel processo di definizione dei gruppi permette però di individuare elementi che aggiungono informazione: il processo di classificazione fa emergere con chiarezza due configurazioni.

Nei gruppi a reddito più basso, oltre alla presenza di uno straniero, i ruoli di maggior rilievo lo giocano la dimensione familiare (oggi bastano 4 persone per rendere svantaggiata una famiglia) e l'occupazione della persona di riferimento. Nei gruppi a reddito più alto è sempre il titolo di studio a operare la differenza.

La variabilità residua che sussiste nei gruppi permette di apprezzare come persone appartenenti a famiglie con caratteristiche simili possano generare redditi eterogenei e, viceversa, come un reddito simile possa essere conseguito da persone e famiglie con caratteristiche e percorsi di vita diversi.

Prospettive ulteriori

Il percorso intrapreso nel Rapporto annuale 2017 rappresenta per l'Istituto il punto di partenza di un filone di ricerca. Già oggi possiamo riportare alcuni risultati ulteriori rispetto a quelli riassunti nel Rapporto stesso, che ci auguriamo possano contribuire ad alimentare il dibattito. Ne riassumo qui le linee principali.

- Un quadro descrittivo riassuntivo che permette di confrontare i nove gruppi rispetto alle singole variabili utilizzate nell'analisi, anche al di là di quelle classificatorie. Le caratteristiche dei gruppi rispetto alle variabili illustrative contribuiscono a delinearne il profilo e offrono ulteriori elementi di validazione dell'analisi.

Due esempi per capire: i 2/3 delle famiglie a basso reddito con stranieri sono in affitto, contro una quota del 19% per l'insieme delle famiglie, che scende al di sotto del 10% nel caso dei due gruppi più benestanti. Oppure, facendo una ripartizione professionale si vede che la quota di ultra-settantenni nei tre gruppi "anziani" della nostra classificazione è molto variabile: da meno 40% per le "pensioni d'argento" a oltre il 60% per gli "operai in pensione", mentre in altri quattro gruppi le persone di riferimento in queste coorti di popolazione sono praticamente assenti.

- Per valutare la robustezza territoriale dell'analisi proposta e le differenze degli elementi prevalenti nelle diverse aree del Paese, la stessa metodologia è stata applicata distintamente per Centro-Nord e

Mezzogiorno (nel Rapporto si è volutamente esclusa dall'analisi la variabile ripartizionale).

E' emerso che nel Centro-Nord imprenditori e dirigenti si separano immediatamente dal resto delle famiglie che poi si partizionano in base a cittadinanza e titolo di studio, mentre nel Mezzogiorno la cittadinanza è molto meno rilevante e tra le famiglie a reddito più elevato ci sono impiegati e pensionati da lavoro.

- Altri aspetti interessanti riguardano le caratteristiche della povertà: recentemente sono stati diffusi i dati relativi al 2016, che comprendono anche i dati sulla povertà assoluta per i nove gruppi sociali identificati nel Rapporto che evidenziano un aumento dell'incidenza della povertà per le famiglie a basso reddito con stranieri e una sostanziale stabilità per gli altri gruppi sociali¹.

Un tema centrale è poi la dinamica temporale dei gruppi. Per questo sono state proposte diverse elaborazioni:

- Per cogliere l'evoluzione dimensionale dei gruppi individuati per il 2015, questi sono stati applicati ai dati d'indagine del 2008 (nel Rapporto questo esercizio è riferito ai soli quinti di reddito). I risultati evidenziano alcuni cambiamenti dovuti all'aumento della popolazione straniera, alla caduta del lavoro autonomo e dell'occupazione operaia.
- Per esplorare gli elementi più significativi nella trasformazione della società, la metodologia xCHAID è stata applicata ai dati dell'indagine Eu-Silc 2008, cioè prima del manifestarsi della crisi, e all'indagine sui consumi del 1988 (utilizzando la spesa familiare come proxy del reddito).

Molto altro c'è da fare e l'Istat rimane aperta a ogni proposta e suggerimento che possa emergere dal dibattito.

I gruppi sociali in Italia: un dibattito di neodemos

Il cambio delle società sviluppate, negli ultimi decenni, è stato vorticoso e ha mutato profondamente diversi punti di riferimento tradizionali per sociologi ed economisti. E' sparita la classe contadina, si è diluita la distinzione tra attività manuali e intellettuali, si sono modificati i tempi del lavoro. Si sono disarticolati gruppi e ceti sociali, riorganizzandosi in combinazioni diverse. Esistono ancora gruppi (classi) sociali dai confini netti e stabili? A queste domanda ha provato a rispondere il Rapporto Istat 2017² che, utilizzando una tecnica statistica inferenziale, è partito da una serie di variabili (reddito, professione, istruzione e molte altre) per ricostruire gruppi di famiglie relativamente omogenei. Un approccio *ex post* diverso da quello *ex ante* comunemente adottato nelle analisi sociali. Metodo e risultati hanno riscosso sia approvazioni che critiche nel mondo degli studiosi³, alle quali ha risposto, nella pagine di Neodemos, il Presidente dell'Istat Giorgio Alleva[4]. Il dibattito così aperto ha trovato una sua continuazione lo scorso 14 luglio, con un Seminario[5] promosso dallo stesso Istat, con contributi di molti studiosi interessati alla questione. Pubblichiamo oggi un articolo di Giorgio Alleva che sintetizza la relazione tenuta in apertura del Seminario.

Note

¹ **Marzio Barbagli, Chiara Saraceno, Antonio Schizzerotto** - L'Istat ora vuole eliminare le classi sociali, Lavoce.info, 23 Maggio 2017 e Carlo Barone, Struttura sociale e disuguaglianze: due consigli

per l'Istat. Neodemos 30 Maggio 2017.

² **Giorgio Alleva** - Classi e gruppi sociali nel Rapporto Istat 2017, Nodemos, 26 Maggio 2017,

³ <https://www.istat.it/it/archivio/201757>

⁴ Cfr. <http://www.istat.it/it/archivio/202338>

⁵ <https://www.istat.it/it/archivio/199318>