

I **gruppi sociali** nel Rapporto Istat 2017



Relazione del Presidente
Giorgio Alleva
Istat | Istituto nazionale di statistica

SEMINARIO PROGRAMMA

ROMA

14 LUGLIO 2017

ORE 10.00

ISTAT | AULA MAGNA

VIA CESARE BALBO 14

L'esperienza del Rapporto Annuale 2017 e la giornata di studio di oggi

Introduzione di Giorgio Alleva

Buongiorno a tutti. Vi ringrazio per avere accettato l'invito a partecipare a quest'evento, dal quale contiamo di ricevere nuovi stimoli per proseguire e migliorare il lavoro fatto per il Rapporto Annuale di quest'anno.

A riguardo, questa mattina abbiamo il privilegio di poter contare su un vero *parterre des roix*:

- **Antonio Schizzerotto**, sociologo autorevole e punto di riferimento indiscusso per lo studio delle classi sociali in Italia;
- **Paolo Giudici**, specialista nell'analisi statistica multidimensionale e nella estrazione del valore da grandi volumi di dati;
- **Rosella Rettaroli**, demografa, che ha offerto numerosi contributi all'analisi della popolazione e dei comportamenti sociali nel nostro paese;
- **Andrea Brandolini**, economista e grande esperto sui temi della disuguaglianza e della povertà dal punto di vista teorico e della misurazione;
- **Giuseppe De Rita**, che ha avuto il merito di sapere interpretare da molti decenni i problemi emergenti e le correnti profonde che muovono la società italiana..

Ringrazio infine Dario Di Vico, che coordinerà i lavori e che, con un articolo sul *Corriere* all'indomani della presentazione del Rapporto, è stato tra i primi a cogliere il senso delle novità che abbiamo introdotto quest'anno.

Ogni anno, 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: due anni fa il *filo conduttore* è stato la nuova geografia che emerge dalla revisione dei Sistemi locali, e l'anno scorso, 90° anniversario dell'Istituto, una nuova tassonomia delle generazioni che compongono la società italiana. Quest'anno, come già sapete, ci siamo prefissi di rappresentare e analizzare l'articolazione sociale del nostro Paese.

L'attenzione ricevuta in ambito scientifico e sui mezzi di informazione, comprese le critiche, testimoniano l'importanza dell'argomento che abbiamo scelto di affrontare.

Come Presidente dell'Istat, quest'anno ho avuto l'occasione di presentare e discutere il Rapporto Annuale e la sua chiave di lettura in sedi molto diverse, dalla Società Italiana di Economia, Demografia e Statistica, allo Studio Ambrosetti, alle Università LUMSA e di Palermo, fino al Comune di Firenze.

E ritengo il dibattito generatosi attorno ai nostri gruppi sociali particolarmente stimolante.

Se mi è permesso un paragone, le critiche soprattutto mi hanno fatto venire in mente il dibattito che all'inizio del secolo scorso si è sviluppato 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, e perdonate se questo mi ha portato ad un collegamento con il nostro albero, un metodo che combina più variabili associate al reddito in maniera empirica per determinare i gruppi sociali.

Questa analogia formale mi ha fatto pensare di essere in una sorta di società quantistica che – evidentemente – è governata dalla frammentazione e dall'incertezza. Nel dibattito fra i fisici il più ostile a questo cambiamento fu un gigante come Einstein.

Questo mio intervento introduttivo vuole fornire alcune coordinate per il dibattito. Per questo, illustrerò brevemente l'impianto concettuale e le scelte metodologiche che abbiamo seguito nel Rapporto.

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

Anzitutto, è bene esplicitare che 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. D'altra parte, nella procedura di suddivisione dell'universo delle famiglie italiane è stato necessario ridurre le configurazioni familiari possibili. Per questo, alcune variabili di segmentazione sono riferite alla sola *persona di riferimento*, cioè l'individuo che ha il reddito più elevato nell'ambito della famiglia.

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.

2. La scelta metodologica

Passando alla scelta metodologica, la ripartizione delle famiglie in gruppi sulla base dei livelli di reddito e di variabili correlate al reddito è stata realizzata attraverso un metodo statistico.

Con l'intento di mettere in relazione i redditi percepiti con caratteristiche correlate suggerite dalla letteratura e dall'osservazione empirica, a orientare nella scelta del disegno sperimentale hanno concorso due considerazioni.

Da un lato, l'interesse principale era verso una classificazione delle famiglie residenti in gruppi, ciascuno associato a una combinazione distintiva di caratteristiche comuni correlate col reddito familiare, piuttosto che la previsione del reddito familiare in funzione di tali caratteristiche. Dall'altro, 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.

3. Indagini e variabili utilizzate

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.

4. L'algoritmo: xCHAID (Exhaustive Chi-Square Automatic Interaction Detector)

L'algoritmo utilizzato 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.

Poiché è computazionalmente proibitivo calcolare tutte le partizioni possibili, a 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).

È importante sottolineare che la procedura non garantisce un'omogeneità elevata in assoluto rispetto alla variabile dipendente, ma l'omogeneità massima possibile a ogni passo.

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.

Come in qualsiasi analisi fondata su un modello, nella fase di validazione è infatti compito del ricercatore accettare i risultati valutandone la robustezza.

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.

5. I risultati ottenuti

È importante sottolineare che, una volta selezionato l'insieme delle variabili esplicative, quelle che suddividono via via i gruppi e l'ordine in cui intervengono è definito dall'algoritmo, e non una scelta a priori del ricercatore.

Queste sono risultate: la partecipazione al lavoro della 'persona di riferimento' della 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.

Oltre alla situazione professionale, sempre importante, e la presenza di uno straniero in famiglia, spesso penalizzante, nei gruppi a reddito più basso i ruoli di maggior rilievo lo giocano la dimensione familiare (non sono svantaggiate soltanto le famiglie numerose, ma anche quelle di 4 persone, che fino a qualche anno fa erano la norma) e il fatto che la persona di riferimento sia occupata (che rappresenta invece un vantaggio). Nei gruppi a reddito più alto, è sempre il titolo di studio a operare la differenza.

Il titolo di studio, inoltre, rimanda a una forma di trasmissione ereditaria dell'appartenenza sociale. Anche il titolo di godimento dell'abitazione si presta a essere considerato un'approssimazione della ricchezza delle famiglie (che l'Istat non rileva direttamente) e della sua trasmissione intergenerazionale.

La variabilità residua che sussiste nei gruppi permette di apprezzare come persone appartenenti a famiglie con caratteristiche simili (condizione professionale e titolo di studio della persona di riferimento, titolo di godimento dell'abitazione, e così via) possano generare redditi eterogenei e, viceversa, come un reddito simile possa essere conseguito da persone e famiglie con caratteristiche e percorsi di vita diversi.

6. Prospettive ulteriori

Infine ci tengo a sottolineare che 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.

Senza sottrarre tempo alla discussione di oggi, 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.

Ad esempio, si osserva che 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. Ancora, come riflesso della partizione dovuta alla condizione professionale, che nei tre gruppi di "anziani" nella nostra classificazione, la quota di ultra-settantenni varia da meno del 40% per le pensioni d'argento fino 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; o infine, l'intreccio tra livelli di istruzione ed esito professionale-reddituale, che è solo mediato dalla classificazione, per cui il profilo delle famiglie a basso reddito con stranieri si avvicina a quello dei giovani blue collar. Potranno essere considerate anche altre caratteristiche familiari.

- Per valutare la **robustezza territoriale** dell'analisi proposta e le differenze degli elementi prevalenti nelle diverse aree del Paese, la stessa metodologia (xCHAID su Eu-Silc 2015) è stata applicata distintamente per Centro-Nord e Mezzogiorno (ricordo, per inciso, che nel Rapporto si è volutamente esclusa dall'analisi la variabile ripartizionale).

Questo esercizio mostra che nel Centro-Nord imprenditori e dirigenti si separano immediatamente dal resto delle famiglie, che successivamente si partizionano in base a cittadinanza e titolo di studio. Nel Mezzogiorno, invece, impiegati e pensionati da lavoro rientrano tra le famiglie a reddito più elevato e la cittadinanza è molto meno rilevante (in ragione della minore presenza relativa di famiglie straniere tra i residenti)

- Altri aspetti interessanti riguardano le caratteristiche della povertà: a riguardo, segnalo che ieri sono stati diffusi i dati relativi al 2016, che comprendono anche i dati sulla povertà assoluta per i nove gruppi sociali identificati nel Rapporto, riportati anch'essi nel Dossier, che con riferimento al triennio 2014-2016 evidenziano un aumento dell'incidenza della povertà per le famiglie a basso reddito con stranieri e una sostanziale stabilità per gli altri gruppi sociali.

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 più salienti di questo esercizio comparativo sono la forte crescita relativa del gruppo delle famiglie con stranieri (la cui quota in termini individuali è aumentata dal 4,9 al 7,8%) e, in misura minore, delle *anziane sole e giovani disoccupati* (dal 7,4 all'8,7%) tra quelli a basso reddito, nonché, tra i ceti più agiati, dei gruppi delle famiglie della *classe dirigente* (dal 6,9 al 7,6%) e degli *impiegati* (dal 19,6 al 20,3%), mentre si riduce l'importanza di tutti gli altri gruppi e, in particolare, delle *famiglie tradizionali della provincia* (dal 7,6 al 6%). Questi aspetti sono illustrativi di alcuni cambiamenti strutturali intervenuti nel tempo, dall'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).

L'albero di decisione nel 2008 presenta molte similitudini rispetto al 2015, mostrando però alcune differenze che ben si spiegano con l'aumento della popolazione straniera, l'invecchiamento della popolazione e gli effetti della crisi economica sulle famiglie di occupati.

Nel 2008, come nel 2015, la prima partizione dell'universo delle famiglie avviene in base alla situazione professionale della persona di riferimento. Si osserva, tuttavia, che nel 2015 le famiglie con persona di

riferimento impiegata, lavoratore in proprio o con pensione da lavoro vanno nel “ramo” a reddito più elevato, in cui nel 2008 confluiscono invece solo dirigenti, quadri e imprenditori. Inoltre, nel 2008 la seconda partizione avviene sempre in base al titolo di studio (mentre nel 2015 per le famiglie a basso reddito entra in gioco anche la cittadinanza). Nel 2008, infine, non si identifica nessun gruppo definito in base alla cittadinanza e, soprattutto, non vi sono gruppi di famiglie con persona di riferimento ritirata dal lavoro, mentre nel 2015 vi sono ben tre gruppi di “anziani”.

L'analisi preliminare condotta sul 1988, pure se con le dovute cautele per le differenze tra fonti e variabili utilizzate (consumi anziché reddito; disoccupazione autodichiarata vs. ILO; mancanza di informazione sulla cittadinanza) evidenzia che la variabile di maggior rilievo nello spiegare le differenze nelle spese familiari (cioè quella che determina la prima partizione) era rappresentata dal titolo di studio della persona di riferimento, mentre la condizione professionale interveniva solamente al quarto passaggio, dopo il numero di componenti della famiglia e l'età della persona di riferimento.

Molto altro c'è da fare e siamo naturalmente aperti a ogni proposta e suggerimento che possa emergere a partire dal seminario di oggi per i quali ringrazio anticipatamente.