I sondaggi che vengono condotti in Italia spesso non risultano attendibili poiché si basano su campioni estratti da liste di campionamento (ad esempio, gli elenchi telefonici) che escludono più della metà degli italiani. È quanto emerge da una ricerca condotta da Emanuela Sala e Roberto Lillini del Dipartimento di Sociologia e Ricerca Sociale dell’Università di Milano-Bicocca volta a indagare la qualità dei dati ottenuti attraverso indagini campionarie telefoniche (CATI). I risultati dello studio sono stati pubblicati sulla rivista International Journal of Public Opinion Research (“Undercoverage bias in telephone surveys in Europe. The Italian case”, doi:10.1093/ijpor/edv033).

Tra gli obiettivi della ricerca vi è innanzitutto la quantificazione del tasso di copertura telefonica, ovvero il calcolo della percentuale degli italiani esclusi dagli elenchi telefonici utilizzati ancora oggi per selezionare gli individui da intervistare durante i sondaggi; quindi la comparazione della qualità dei dati che si ottengono utilizzando diversi tipi di campione e infine la valutazione dell’efficacia di alcune procedure statistiche utilizzate per ridurre l’impatto dell’errore di copertura.

Premesso che, come detto, i campioni dei sondaggi vengono tuttora estratti dagli elenchi telefonici (le Pagine Bianche), lo studio ha messo in luce che al 2012 circa il 50% degli italiani è escluso da tali elenchi (di questi, il 32% non dispone di una linea fissa, mentre il 18% ha scelto di non comparire negli elenchi), e che i soggetti esclusi hanno caratteristiche socio-demografiche differenti rispetto a quelle dei soggetti inclusi negli elenchi.

L’identikit del campione fantasma

Gli esclusi rispondono perlopiù a queste categorie: maschi, residenti nel Sud Italia, soprattutto nelle isole maggiori, di età compresa tra i 25 e i 34 anni, con bassa scolarizzazione, separati o divorziati, e che abitano in affitto. Qualche cifra: rispetto agli abitanti del Centro Italia, coloro che risiedono al Sud hanno il 29% di probabilità in più di essere esclusi dagli i elenchi, e l’82% di probabilità in più per quanto riguarda coloro che abitano in Sicilia e Sardegna. Analogamente, paragonati a coloro che sono in possesso di un diploma di scuola superiore, i soggetti con solo la licenza media hanno il 32% di probabilità in più di non essere inclusi negli elenchi telefonici. Ancora, rispetto a coloro che sono sposati, gli individui separati o divorziati hanno il 60% di probabilità in più di essere esclusi dagli elenchi telefonici e quindi di non essere campionabili.

I ricercatori poi hanno voluto valutare l'impatto della copertura telefonica sulla qualità dei dati che si ottengono quando si conducono dei sondaggi, dividendo il dataset da loro analizzato (dati dell’Indagine Multiscopo dell’ISTAT) in due gruppi (soggetti inclusi nell’elenco telefonico e soggetti esclusi) e paragonando le risposte da loro fornite in merito a 13 domande riguardanti cinque aree tematiche: politica, salute, sicurezza, uso dei media e delle nuove tecnologie e tempo libero. Gli individui appartenenti ai due gruppi hanno dato spesso risposte differenti: ad es. coloro che sono esclusi dagli elenchi dichiarano di essere in condizioni di salute peggiori rispetto agli inclusi (vedi l’affermare di soffrire di una malattia cronica, 76,8% contro il 66,5%); gli esclusi fanno minore uso delle nuove tecnologie rispetto agli inclusi (vedi l'uso di internet, 55.5% contro il 61.8%); per quanto riguarda la politica, i soggetti esclusi si informano quotidianamente di meno (il 32,4% contro il 41,9%) e ascoltano meno i dibattiti politici (17.3% contro il 23%) di quanto non facciano gli inclusi.

Di conseguenza, vari sondaggi come quelli politico-elettorali potrebbero fornire un quadro poco veritiero delle opinioni degli italiani, i quali risulterebbero essere meno interessati alla politica o meno informati sui fatti politici di quanto i sondaggi possano far credere.

Infine, lo studio ha chiarito che anche le procedure statistiche spesso utilizzate per controbilanciare gli effetti dell’errore di copertura (i cosiddetti pesi statistici) sono poco efficaci. Ad esempio, pur utilizzando un sistema di pesi calcolato a partire da alcune caratteristiche socio-demografiche di base (età, sesso, area geografica di residenza e istruzione), le differenze nelle risposte fornite fra il gruppo degli inclusi e quello degli esclusi rimangono sostanzialmente invariate. Gli individui esclusi dall’elenco di campionamento mostrano sempre un 29% di probabilità in meno di seguire sui media i dibattiti politici rispetto ai soggetti presenti nell’elenco telefonico.

«Ci auguriamo», commenta Emanuela Sala, «che questo nostro lavoro possa contribuire ad aprire un serio dibattito anche in Italia sulla qualità dei dati campionari raccolti con il metodo dell’intervista telefonica e possa condurre allo sviluppo di nuove tecniche campionarie, come ad esempio il dual-frame sampling, tipo di campionamento che si basa su elenchi di telefonia fissa e mobile, o l’Address Based Sampling (ABS) basato su un elenco di indirizzi residenziali. Ricordiamo, a tal proposito, che proprio le indagini Multiscopo dell’ISTAT si basano su questo tipo di campionamento».

«Non dimentichiamoci», conclude Sala, «che i dati campionari raccolti con questi strumenti sono utilizzati anche dai policy makers; è perciò necessario lavorare insieme agli istituti di ricerca e ai sondaggisti al fine di migliorare la qualità dei dati raccolti».