Dieci buoni motivi per non credere (troppo) ai sondaggi – 6. Errore di copertura e 7. Errore di mancata risposta

written by Paolo Natale | 11 Luglio 2022

I due tipi di distorsione che vedremo ora fanno entrambi parte del cosiddetto errore di rilevazione; al contrario dell'errore di campionamento, sempre presente ma calcolabile, questo tipo di errore potrebbe teoricamente non esserci (peraltro solo in teoria…) ma non è comunque mai calcolabile: non sappiamo dunque di quanto sbagliamo nella generalizzazione delle nostre stime.

L'errore di copertura si verifica in diverse fasi dell'indagine e per diversi motivi. Può essere dovuto alla mancanza di un elenco completo dell'universo dei possibili intervistati, da cui estrarre il campione, alla impossibilità di reperimento di una quota significativa di individui campionati o al loro rifiuto preventivo di rispondere alle domande. Nel primo caso nulla si può fare, mentre negli altri due è possibile ovviare al problema estraendo preventivamente ulteriori campioni, chiamati "di riserva", che prenderanno il posto dei soggetti irreperibili, sostituendoli quindi con nuovi individui reperibili e/o disponibili a collaborare. Ma è ovvio che così facendo introduciamo una (possibile) fonte di distorsione a volte piuttosto grave: come possiamo sapere se l'individuo sostituito risponderebbe in maniera simile al suo sostituto? Solitamente, quando si pubblicano i risultati di un sondaggio, in caratteri molto piccoli, a volte quasi invisibili, viene inserito anche il dato relativo al numero delle sostituzioni, che molto spesso sono almeno tre-quattro volte più numerose delle interviste effettivamente effettuate, come nel caso (tipico) qui riportato.

Indagine condotta con tecnica mista CATI-CAMI-CAWI su un campione di 1200 soggetti maggiorenni residenti in Italia (4566 non rispondenti) il 7 giugno 2022.

Il che significa molto semplicemente che l'Istituto di ricerca è riuscito a intervistare 1200 soggetti facendo quasi 6000 tentativi! Quei quasi cinquemila individui irreperibili o non disponibili avrebbero dato risposte simili? Può darsi di sì o può darsi di no; non lo sappiamo né potremo mai saperlo.

—

L'errore di mancata risposta si verifica invece quando una parte di coloro che vengono intervistati non vogliono rispondere ad alcune delle domande del questionario, oppure si dichiarano indecisi sulla risposta da fornire oppure ancora non sanno semplicemente cosa rispondere. Prendiamo ad esempio il caso più ricorrente, un sondaggio cioè sulle intenzioni di voto, con una numerosità campionaria di 800 casi: negli ultimi mesi, il numero di intervistati che si dichiara indeciso, che non vuole rispondere o che opta per l'astensione risulta spesso pari se non superiore al 40% del campione. Una parte di loro, diciamo intorno al 25%, non si recherà effettivamente alle urne, ma il restante 15% ci andrà e avrà un certo comportamento di voto, che potrebbe ribaltare completamente i risultati che ci vengono proposti. Come possiamo sapere se chi oggi non indica la propria preferenza elettorale si distribuirà, il giorno delle votazioni, esattamente come chi l'ha invece indicata? Ma non solo. Il problema ulteriore è che le stime che noi leggiamo si riferiscono non più a 800, ma soltanto a 480 individui (con un errore di campionamento – peraltro – molto superiore a quello indicato), che ci dicono siano rappresentativi della popolazione elettorale italiana. Saranno davvero rappresentativi di tutti gli elettori oppure soltanto di quelli più sicuri del proprio voto e/o propensi a dichiararlo?

Paolo Natale

**estratto del volume “Sondaggi”, in uscita nel prossimo autunno presso Laterza*