

Cuinar les dades per no enganyar la ciutadania

«És especialment irresponsable publicar dades sense corregir com si fossin dades corregides»

Generalment, quan es publica una enquesta electoral, els partidaris de les forces que queden més mal parades tendeixen a no creure's les dades. És una tendència ben descrita per la psicologia: tendim a resistir-nos a acceptar la informació que no ens agrada, i en canvi som molt poc crítics i ens creiem a la primera quan ens explica el que volem sentir.

Aquesta setmana ha passat amb el baròmetre de novembre del CIS, dependent del govern espanyol, que han generat força comentaris i polèmica. En aquest cas, però, està justificada perquè són unes dades que es desvien molt de la majoria d'enquestes que s'han anat publicant. Entre d'altres coses, per exemple, col·locaven a Junts per Catalunya com a quarta força, fins i tot per sota dels Comuns a les properes eleccions al Parlament. És un resultat força estrany, que no es correspon amb el que raonablement podem esperar ni amb el que van dient les enquestes que s'han anat publicant.

Quan això passa de seguida apareixen les hipòtesis, diguem-ne així, de la conspiració. El comentari més habitual és que les dades han estat ?cuinades? perquè tinguessin un resultat determinat i així, influir en les eleccions. Això és evident que hi ha vegades que passa. Però com diu la famosa ?navalla de Hanlon (https://ca.wikipedia.org/wiki/Navalla_de_Hanlon) ?, no hem d'atribuir mai a la malícia allò que pot ser explicat adequadament per l'estupidesa. En aquest cas, més que estupidesa és manca de professionalitat del CIS, en primera instància, i dels mitjans que van difondre les dades sense una mínima anàlisi crítica en segona instància.

Perquè, de fet, el problema amb les dades d'aquesta enquesta del CIS és, precisament, que no han estat cuinades. S'han oferit a la ciutadania les dades d'intenció de vot (afegint-hi la simpatia per assignar els que no responen) tal qual vénen de l'enquesta. Això és molt problemàtic, i fa que les dades no es puguin interpretar en cap cas com una fotografia raonable de la intenció de vot de l'electorat català a dia d'avui. Per això en contra del comentari habitual d'aquests dies, el problema no és un excés de ?cuina? sinó una manca de ?cuina? en aquestes dades.

La raó per la qual és problemàtic és relativament fàcil d'entendre. La mostra de persones enquestades no és necessàriament representativa del conjunt de la població. Fonamentalment, perquè hi ha molta gent que rebutja respondre enquestes, o que és difícil de localitzar a casa. Com que la gent que respon té unes característiques sociopolítiques diferents a les de la gent que no respon, el resultat és que la mostra té biaixos que cal corregir. La correcció d'aquests problemes (mitjançant diversos procediments estadístics) és el que popularment s'anomena ?cuina?, i és necessària. Difondre les dades sense corregir aquests errors sistemàtics que apareixen en el procés de mostreig és irresponsable, i porta a equívocs que, tant qui fa enquestes com qui les difon haurien d'evitar.

A més, això s'afegeix a un canvi clau que ha fet el CIS a les seves enquestes aquest any 2020: en lloc de fer-les en persona, enviant enquestadors als domicilis, les fa per telèfon. La causa que addueixen és la pandèmia, tot i que altres institucions com el CEO català, o l'ICPS han realitzat aquests dies enquestes domiciliàries. Les mostres telefòniques són cada cop més problemàtiques, perquè el percentatge de gent que no respon és molt gran, i el perfil de la gent que respon és cada cop menys representatiu.

En el context actual fer enquestes és cada cop més difícil, precisament per la caiguda tan

forta en les taxes de resposta. És un fenomen internacional. Gràcies als avenços en els procediments de correcció dels biaixos, de fet, les enquestes electorals són en general tan predictives (o tan poc predictives) avui com fa uns anys.

Per això és especialment irresponsable publicar dades sense corregir com si fossin dades corregides. És evident que en els procediments de correcció poden sorgir nous biaixos, intencionals o no. Però la solució és la transparència i la publicitat dels procediments. En cap cas deixar de corregir unes dades que incorporen molts errors sistemàtics. Tant el CIS com el CEO publiquen les micro-dades de l'enquesta, cosa que permet d'alguna manera avaluar els procediments emprats. Però el proper pas hauria de ser la publicació de tot el codi que s'hagi aplicat per arribar a l'estimació final, de manera que la ciutadania pugui avaluar i conèixer els procediments emprats. És important perquè sabem que les enquestes poden tenir influència en el comportament de la gent, i per això les institucions públiques que en fan han de ser totalment transparents. Com a mínim, amb els estàndards que s'exigeix en les publicacions científiques.