

Estadística: val la pena saber-ne?



Pere Grima

Universitat Politècnica de Catalunya

Sovint quan es parla d'estadística les “vibracions” no són bones. Amb les matemàtiques pot passar una cosa semblant (molta gent pensa que costa d'entendre, que són molt abstractes, fins i tot algú pot pensar que costa trobar-hi aplicacions...) però almenys tothom està d'acord en que les matemàtiques són serioses. Quan es parla de futbol i es diu que un equip és “matemàticament campió” es vol dir que segur que ho serà, passi el que passi, no hi ha cap dubte. Però una altra cosa és l'estadística. Que l'estadística digui que aquest equip guanya sempre a casa no vol dir que aquesta vegada també guanyi, només són “estadístiques”.

I si l'estadística és similar a les matemàtiques quant a dificultat (els llibres d'estadística també van plens de números i de fórmules) però sembla que els seus resultats no són segurs i no te'n pots refiar... val la pena aprendre estadística?

Què és l'estadística?

Estarem d'acord en el fet que, per prendre decisions, ja sigui en l'àmbit de la sanitat pública (què cal fer per a lluitar contra una determinada malaltia?), l'administració pública (es pot dir que hi ha hagut irregularitats en unes eleccions?) o gestió del medi ambient (cal protegir un cert tipus d'animal?) sempre és millor basar-se en un coneixement objectiu de la realitat que no pas en prejudicis, interessos o suposicions, i coneixement objectiu vol dir dades que reflecteixin la realitat tal com és.

L'estadística és la disciplina que s'encarrega d'estudiar com recollir les dades (quantes? de quina manera per assegurar la seva representativitat?) i com analitzar-les per poder respondre a les preguntes que ens plantejem. Moltes persones identifiquen l'estadística només amb aquells aspectes que més apareixen en els mitjans de comunicació, però, més enllà de percentatges, sondejos electorals i indicadors econòmics, l'estadística juga un paper fonamental en molts camps del coneixement, com ara el control de qualitat, el descobriment de nous medicaments, la biologia, la sociologia, els estudis relacionats amb el medi ambient, l'economia, els estudis de mercat i molts d'altres entre els que també s'inclou la física i l'explicació de l'estructura de la matèria. Veiem alguns exemples.

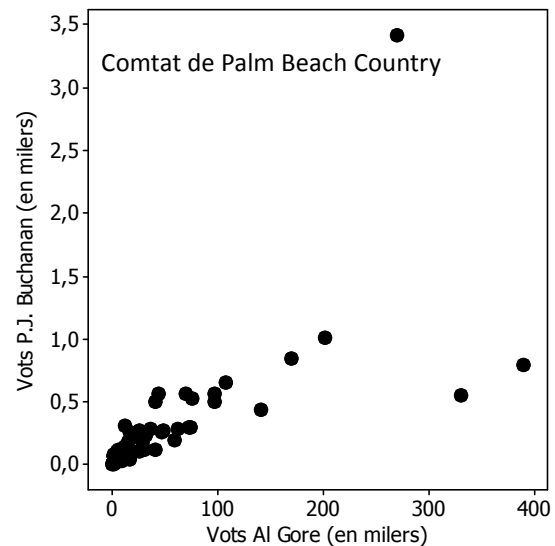
Estadística i salut, un cas històric: Còlera a Londres

A finals d'agost de 1854 es va desencadenar una terrible epidèmia de còlera al centre de Londres. Va produir més de 500 morts en només 10 dies i el pànic es va apoderar de la població. John Snow¹ era un metge escèptic amb les teories que dominaven en aquesta època i segons les quals el còlera era provocat pel *miasma*, un gas que se suposava que es despenia dels cossos en descomposició.

La realitat que ell observava no quadrava amb la teoria i va sospitar que segurament l'explicació havia de ser una altra. Com a bon



Fragment del mapa de la zona on es va produir l'epidèmia, realitzat pel Dr. Snow. S'indica la posició de les fonts (PUMP). Les ratlletes representen el nombre de morts a cada casa.



Vots de Patrick J. Buchanan enfront dels d'Al Gore en cadascun dels 67 comtats de l'Estat de Florida. El comtat de Palm Beach és el de la part superior.

científic, va decidir estudiar la realitat de manera objectiva i, entre altres coses, va aconseguir una llista dels morts per còlera. Quan va situar el lloc de residència de cada un d'ells sobre un mapa de la zona va observar que estaven agrupats al voltant d'una font que va considerar sospitosa. També hi havia excepcions, però les va analitzar una per una: va comprovar que algunes de les persones que van morir fora de l'àrea d'influència d'aquesta font tenien el costum d'agafar l'aigua d'allí, perquè la consideraven millor. També van emmalaltir uns nens que vivien lluny, però la font estava en el camí de casa a l'escola.

Amb un complet i detallat estudi el Dr. Snow va convèncer les autoritats que la causa de l'epidèmia estava en aquesta font situada a Broad Street. El 7 de setembre es va treure la maneta que obria l'aixeta i el nombre d'afectats va caure en picat. Mai més no hi van haver epidèmies de còlera a Londres i ràpidament aquesta amenaça va desaparèixer d'Europa. I va ser amb esperit científic, buscant dades i analitzant-les de manera intel·ligent -fent servir l'estadística- com es va aconseguir.

Estadística i Política: Eleccions presidencials als EUA

Un dels aspectes més comentats en els mitjans de comunicació sobre les últimes eleccions americanes (novembre de 2012) és

com els models estadístics i, especialment, els d'un matemàtic anomenat Nate Silver, han estat capaços de predir els resultats amb molta precisió². Però ara voldria comentar els resultats electorals de l'any 2000 amb el demòcrata Al Gore contra el republicà George Bush, que va guanyar Bush, però amb un resultat molt ajustat i també molt discutit. A l'Estat de Florida, amb uns 6 milions d'electors, Bush va guanyar per 537 vots de diferència, i qui guanyava en aquest Estat tenia la majoria necessària per a ser president. Hi va haver impugnacions i els jutges van haver de decidir. Sense entrar en consideracions jurídiques, el gràfic mostra els vots que va obtenir Al Gore contra els d'un altre candidat, Patrick J. Buchanan, en cadascun dels 67 comtats de l'Estat de Florida³.

El primer que salta a la vista és que el valor de Palm Beach no segueix el patró general. Els punts es presenten agrupats marcant una tendència segons la qual a Palm Beach es podien atribuir a Buchanan uns 1500 vots, però va obtenir-ne 3411. Alguna cosa singular havia de tenir Palm Beach, però no hi havia cap raó perquè Buchanan tragués en aquest comtat un percentatge de vots molt superior als altres. Ell mateix i el seu equip van declarar que obtenir 1000 vots era una previsió optimista. Aviat va estar clar que la singularitat era el disseny de la papereta utilitzada per a votar en aquest comtat. Calia perforar un cercle segons el candidat escollit

però l'assignació de cercles a cada candidat es prestava a confusió i moltes persones, segurament més de 2000, van votar a Buchanan quan en realitat volien votar a Al Gore. Una simple anàlisi de les dades -fent servir l'estadística- mostra que alguna cosa devia passar. No és clar que els jutges que van decidir ho tinguessin en compte.

Estadística i Ecologia: Quants peixos hi ha en un llac

Comptar quants peixos hi ha en un llac no sembla una tasca fàcil, especialment si és gran i d'aigües tèrboles, però els biòlegs saben com fer-ho. Utilitzant tècniques estadístiques, és clar. Un mètode molt utilitzat és l'anomenat de *pesca i repesca* (o, en general, de *captura-recaptura*, perquè no només serveix per a peixos). El procediment és el següent:

- Pescar una mostra de peixos, marcar-los i tornar-los a l'aigua.
- Deixar passar un temps (poden ser uns dies) fins que sigui raonable considerar que els peixos marcats s'han dispersat per tot el llac, i tornar a pescar una altra mostra (la *repesca*) d'un nombre que no necessàriament ha de ser igual al de la pesca.
- Fer els càlculs: Si al llac hi ha N peixos i en marquem M , la proporció de marcats és M/N . A la repesca capturem C peixos que podem considerar una mostra representativa de tots els peixos del llac i entre ells trobem R marcats. És raonable considerar que la proporció de peixos marcats a la segona mostra és similar a la proporció de peixos marcats al llac, és a dir: $M/N \approx R/C$, de manera que el nombre de peixos que hi ha al llac estarà al voltant de: $N \approx (M \cdot C)/R$.

Però... què vol dir "estar al voltant de"? Basant-se en unes hipòtesis raonables i amb arguments matemàtics, la teoria estadística s'encarrega de respondre aquesta pregunta i també de trobar estimadors més precisos, tot i que també més complicats de deduir. De tota manera, per fer-nos una idea de la precisió d'aquesta estimació podem fer un petit programa d'ordinador que simuli el procés sobre un nombre de peixos fixat per nosaltres. Amb aquest procediment, es pot repetir la pesca i repesca tantes vegades com es vulgui i veient el nombre de peixos

que s'hauria estimat en cada simulació es pot determinar la magnitud dels errors que s'hi cometen i amb quina freqüència s'hi presenta cadascun d'ells.

Més aplicacions

Hi ha llibres dedicats a explicar casos d'aplicació de l'estadística en àmbits diferents. Uns de molt coneguts són els intitolats *Statistics: A Guide to the Unknown*. El primer dels dos volums de què consta ha estat traduït al castellà i publicat per Alianza Editorial (crec que està exhaurit). Més fàcil de trobar és el que va editar la Facultat de Matemàtiques i Estadística de la Universitat Politècnica de Catalunya (UPC), amb exemples d'aplicació basats en projectes de fi de carrera i explicats pels mateixos autors. N'hi ha una versió en català (*Fent servir l'Estadística*) i en castellà (*Estadística en Acción*) i ambdues es poden abaixar de www.upccommons.upc.edu ⁴.

I encara més

Fa un temps vaig preguntar a un estudiant de la Diplomatura d'Estadística de la UPC (ara Grau) per què havia decidit estudiar estadística. Em va dir que volia saber com es cuina la informació que ens donen, saber valorar si unes dades són fiables i tenir esperit crític per a valorar la informació que ens arriba. Doncs, sí, senyor. També per a això serveix l'estadística.



NOTES

¹ Mes informació sobre John Snow i l'epidèmia de còlera de 1854 es pot trobar a: <http://www.ph.ucla.edu/epi/snow.html>. D'aquesta pàgina web he tret el mapa.

² Veure, per exemple, El País del 13 de novembre de 2012: http://elpais.com/elpais/2012/11/12/opinion/1352747017_116039.html

³ La idea està treta de D.S. Moore: "Learning from Data", que presenta aquest mateix gràfic (a "Statistics: A Guide to the Unknown", de Roxy Peck et al., 4a edició). Les dades per reproduir el gràfic les he tret de: <http://uselectionatlas.org/RESULTS/>

⁴ "Estadística en acción": <http://upccommons.upc.edu/e-prints/handle/2117/7915> "Fent servir l'estadística": <http://upccommons.upc.edu/e-prints/handle/2117/7854>