



3.10 Estadística

Selecció de problemes preparada per Pedro Puig Casado, professor d'Estadística i Investigació Operativa del Departament de Matemàtiques de la Universitat Autònoma de Barcelona. tel.: 93 581 13 02, e-mail: ppuig@mat.uab.es.



Pollastres, mitjos pollastres i empreses igualitàries. Una de les crítiques de l'estadística que més sovint es veu als mitjans de comunicació està relacionada amb l'ús incorrecte de la mitjana aritmètica. Un d'aquests comentaris que podeu trobar és que si un individu es menja cada dia dos pollastres i un altre no se'n menja cap, de mitjana cadascun es menja un pollastre diari! Això el que ens diu és que la mitjana aritmètica només ens dona una informació parcial del conjunt de dades. Aquesta informació pot ampliar-se utilitzant altres mesures (estadístics descriptius), com per exemple la desviació tipus, els quartils, etc. Podeu veure alguna d'aquestes coses amb l'exercici següent,

- 3.10.1.** Tenim dues petites empreses cadascuna de les quals està formada pel director i nou treballadors. Fixeu vosaltres mateixos els sous de tots, de manera que el sou mitjà sigui el mateix per a cada empresa, però que en una d'elles el repartiment sigui menys igualitari que en l'altre.

Calculeu ara les desviacions tipus i els quartils i compareu-los.

Criptografia, què hi diu aquí? Edgard A. Poe en un conte titulat "L'enigma de l'escarbat d'or", descriu com el protagonista va desxifrar un text (escrit per un pirata) aparentment incomprensible fent servir un mètode estadístic. El pirata havia fet servir un sistema molt simple per a xifrar el seu missatge, consistent a substituir cada lletra de l'alfabet per un signe arbitrari. Per exemple, la A podria ser canviada per un *, la B per un {, etc. El protagonista de la nostra història tenia, doncs, un text ple de signes estranys amb un significat totalment desconegut. El que va fer per a desxifrar-lo és el mateix que us proposem en l'exercici següent,

- 3.10.2.** Digueu a un amic que us escrigui un missatge fent servir el mètode de xifratge descrit anteriorment, sense dir-vos quina és la clau però sí en quin idioma està escrit. A continuació feu una llista dels signes que surten en aquest text amb el nombre de vegades que apareix cadascun (això s'anomena distribució de freqüències). Tot seguit agafeu un text qualsevol d'una novel·la, diari, etc. escrits en el mateix idioma que el missatge xifrat, i feu també una distribució de freqüències de les lletres del text. Podem pensar que les proporcions d'aparició de cada lletra de l'alfabet són una característica a cada idioma. Per exemple, tant en català com en castellà la lletra més freqüent és la E, seguida de la A, etc. Fent servir aquesta informació i una mica d'enginy, tracteu de desxifrar el missatge. Com més llarg sigui el missatge, més fàcil serà desxifrar-lo! Consulteu també la secció 2.15.

El protagonista del conte de Poe ja sabia que el seu missatge xifrat estava escrit en anglès perquè coneixia la identitat del pirata. Si no sabem en quin idioma està escrit el missatge, podem deduir-lo a partir del mateix text en clau? En aquest proper exercici us proposem com fer-ho.

- 3.10.3.** Agafeu un text llarg qualsevol escrit en català, un altre en castellà, un altre en anglès, etc. i per a cadascun calculeu la mitjana i la desviació tipus del nombre de lletres que té cada paraula. Podem pensar que també aquestes dues quantitats són una característica de cada idioma. Si agafem el missatge en clau del nostre amic, que ara no ens ha dit en quin idioma ho ha escrit, encara que d'entrada no sapiguem el seu significat, sí que podem comptar el nombre de lletres que té cada paraula i calcular la mitjana i la desviació tipus. Amb aquesta informació podem deduir en quin idioma ha estat escrit.

Captura i recaptura, dels peixos als drogues. Tenim una població d'un cert tipus de peix que viu dins d'un estany. Volem conèixer el nombre de peixos que té aquesta població. Com podem "estimar" aquesta quantitat? Una idea molt simple és capturar uns quants peixos, marcar-los amb una tinta que no marxi amb l'aigua, deixar-los anar dins de l'estany i, passat un quant temps, tornar-ne a capturar.

Mirant el nombre de peixos d'aquesta segona captura que ja han estat marcats a la primera (individus recapturats), podem conèixer de manera aproximada el nombre total de peixos de l'estany. El següent exercici és un exemple de càlcul.

- 3.10.4.** A la primera captura hem agafat 24 peixos, els marquem i els tornem a l'estany. A la segona captura hem agafat 32 peixos, dels quals trobem cinc marcats anteriorment (aquests són els recapturats). Quin és el nombre aproximat de peixos que té l'estany? (Nota: us ha de donar 154.)

Aquesta metodologia la podem aplicar a diversos problemes canviant la idea de "marcatge". Vegeu l'exemple següent.

Estimació del nombre d'individus que realment habiten en una ciutat o un país (Laplace, s. XVIII).

Si disposem del cens oficial, això pot constituir la primera captura. Una enquesta adient ens donarà la segona. Els enquestats que també apareguin al cens seran els recapturats.

Estimació del nombre de consumidors d'una certa droga en una ciutat o un país.

El conjunt d'individus que fan una determinada teràpia de desintoxicació pot constituir la primera captura. Els detinguts posteriorment durant un cert període de temps poden ser la segona. Els repetits a les dues "listes" seran els recapturats.

Cal no perdre amb les dades perdudes. Moltes vegades, revisant qüestionaris procedents d'enquestes, podem constatar que ens falten algunes dades de certs individus. Per exemple, d'un d'ells ens pot faltar l'edat, d'un altre el nombre de germans, etc. Si suposem que l'omissió es deguda a l'atzar i que no hi ha cap raó relacionada amb la natura de les mateixes dades, com per exemple seria el cas d'un que amaga el seu sou perquè és molt elevat, com podem recuperar d'alguna manera aquesta dada perduda o faltant? Una manera és substituint la dada faltant per la mitjana dels valors que tenim dels altres individus. Així, per exemple, si ens falta l'edat d'un d'ells, calcularem la mitjana de les edats dels altres i utilitzarem aquest valor en el seu lloc. A continuació veurem una propietat interessant d'aquesta manera de procedir en forma d'exercici.

- 3.10.5.** Tenim $n - 1$ dades i calculem la seva mitjana. Afegim aquest nou valor als $n - 1$ anteriors i així aconseguim un conjunt amb n dades. Quina serà la nova mitjana per a aquestes n dades?

Un dels exemples més antics de dades faltants apareix a l'Antic Testament, Nom. 1, on es fa l'anomenat cens d'Aaron. Es tracta d'un cens militar en què es compten el nombre de mascles, més grans de vint anys per a cadascuna de les dotze tribus d'Israel excepte per a la tribu de Levi (dada faltant). El motiu és que la tribu de Levi estava exclosa d'utilitzar armes perquè eren els encarregats dels serveis religiosos.

De l'avinguda Meridiana a Dunkerque (passant per París). El 1791, l'Assemblea Nacional francesa va decidir establir un sistema universal de mesures, anomenant "metre" a la unitat de longitud. Es va decidir definir-lo com $1/10\,000\,000$ de la distància entre el pol nord i l'equador mesurat en el meridià que passa per París. Per a portar a terme aquest projecte l'Assemblea va nomenar Delambre i Méchain com a responsables de mesurar la distància entre Dunquerque i Barcelona, amb el qual s'abastava uns 10^0 del meridià. Méchain va ser l'encarregat de realitzar les mesures a Catalunya, on va ser molt ben acollit pels intel·lectuals de l'època. A la plaça de les Glòries Catalanes hi ha un monument commemoratiu d'aquest afer i, de fet, l'avinguda Meridiana de Barcelona és el lloc per on passa el meridià Dunquerque-París-Barcelona.

Les dades procedents d'aquestes mesures van ser analitzades per Legendre fent servir un mètode que posteriorment ha tingut moltes aplicacions pràctiques; es tracta del mètode de mínims quadrats. L'exercici següent pot il·lustrar el procediment.

Les dades reportades per Delambre i Mechain al 1799, presentades per Laplace en el seu *Traité de Mécanique Céleste*, es poden observar a la següent taula,

Longituds dels arcs sobre el meridià de París.

Lloc	Latitud	Dist. des de Barcelona
Barcelona	45.958281	
Carcassonne	48.016790	52749.48
Évaux	51.309414	137174.03
París	54.274614	213319.77
Dunquerque	56.706944	275792.36

Les distàncies estan expressades en mòduls (1 mòdul=2 toeses, 1 toesa = 1.949 metres) i les latituds en graus centesimals.

- 3.10.6.** Anoteu en un paper els pesos i les alçades dels companys de classe i representeu-los en un gràfic cartesià. Observeu la tendència lineal del núvol de punts. Consulteu la secció 2.3. Calculeu la recta de millor ajustament al núvol de punts pel mètode dels mínims quadrats. Interpreteu-ne els resultats.

Una polèmica científica que s'ha fet famosa és la paternitat del mètode dels mínims quadrats. Si bé la primera publicació respecte a aquesta qüestió va ser deguda a Legendre en el context de la determinació del metre, Gauss va afirmar que ell havia fet servir el mètode uns quants anys abans...