

USOS I ABUSOS DE L'ESTADÍSTICA

Escrit per:

Carles M. Cuadras

Carles M. Cuadras
Departament d'Estadística
Universitat de Barcelona

L'Estadística és una branca de les Matemàtiques, que brilla amb llum pròpia, i que és gairebé omnipresent en totes les Ciències Experimentals i Socials, degut a les seves múltiples aplicacions. En aquest article comentem alguns dels seus usos i abusos en Biologia, Psicologia i Medicina.

Esquemàticament, l'usuari de l'Estadística planteja un problema i segueix la següent seqüència:

Dades–Model–Resultats–Interpretació. (1)

Per diverses circumstàncies, l'ús de l'Estadística en la recerca, i per tant la seva presència en algunes revistes lligades a la investigació biomèdica, han augmentat considerablement. Per exemple, *Ecologia Aquatica* va passar d'un percentatge (ponderat) de 55% el 1977 a 93% el 1987. Aquest percentatge augmenta considerablement si considerem la seva continuació *Scientia Marina*, on pràcticament trobem estadística a cada article publicat. Pocs discuteixen la importància de l'Estadística, però, en fem un us correcte?

Veurem a continuació alguns casos històrics, i altres de recents, en els quals s'ha abusat o bé usat incorrectament l'estadística, en alguns dels passos de la seqüència (1), i en comentarem les conseqüències.

La comparació de cranis entre blancs i negres

Aquest és un típic exemple d'abús de l'Estadística per intentar provar quelcom que

l'investigador ja creu saber d'antuvi: que hi ha diferències de capacitat intel·lectual entre blancs i negres.

Els arguments a favor de les diferències vénen de molt lluny. Convenien, durant part del segle XIX, als partidaris de l'esclavitud i sovint estaven basats en la Bíblia, rememorant la maledicció de Cam, o de Canaan, el fill primogènit de Cam, fill de Noé. Segons la tradició bíblica, els descendents de Canaan estarien relacionats amb els negres. Per tant, els negres eren “diferents” per mandat bíblic i podien ser esclaus. Però la Ciència progressava i la teoria de l'evolució de Darwin va liquidar, pràcticament, les tesis creacionistes. Una altra cosa era l'ús que se'n feia: la “Ciència” va substituir la Religió en els arguments d'alguns a favor de l'esclavitud, i els partidaris del racisme tractaven de provar les seves idees a partir de criteris quantitius, emprant al seu favor “l'objectivitat” que proporcionen els nombres.

En una primera fase, que ocuparia la segona meitat del segle XIX, els investigadors justificaven la diferència qualitativa entre blancs i negres mitjançant la mesura biomètrica del crani. El màxim representant d'aquesta tendència va ser Paul Broca, que relacionava el volum del cervell amb la intel·ligència. En general, deia Broca, el cervell és més gran en els adults que en els vells, en els homes que en les dones, en les races superiors que en les races inferiors. La rèplica, ben *sui generis*, va venir per part de Pierre Gratiolet, que va arribar a afirmar que el cervell dels alemanys era més gran que el dels francesos, i, així doncs, això era una prova clara que la grandària del cervell no tenia res a veure amb la intel·ligència. Cal esmentar també a Bartomeu Robert, el famós Dr. Robert, que havia estat alcalde de Barcelona. Per tal de defensar la seva postura “regeneracionista” de Catalunya, el Dr. Robert va afirmar que el crani dels catalans era més gran. Com podia ser que, per a alguns, fossin inadmissibles les diferències

entre alemanys i francesos, o entre catalans i espanyols, i, en canvi, fossin naturals les que hi havia entre blancs i negres? Era una qüestió opinable? Un altre fet que refutava les teories de Broca era l'observació de la mida de cervells en autòpsies: homes eminents per la seva intel·ligència podien tenir un cervell relativament petit, mentre que criminals potiners el podien tenir més gran. La relació entre mida cerebral i intel·ligència era, per tant, insostenible i els esforços en demostrar-la, sovint basats en incorrectes interpretacions estadístiques, infructuosos.

Els estudis sobre cranis de Samuel Morton anaven en la mateixa direcció. Morton va "provar" que els cranis dels europeus eren els més grans, els anglesos els que més, i els més petits eren els dels aborígens africans i australians. Però si bé les dades eren correctes, els càlculs estadístics no. El paleontòleg Stephen Gould va provar que estaven manipulats o mal interpretats, ja que ometien excepcions o es basaven en cranis més petits per rebaixar la capacitat de les races que hom volia provar que eren inferiors. El final de l'època victoriana marcaria el declivi de la craniometria, i a principis del segle XX les legitimació de les diferències es buscava per altres camins. Els tests, els coeficients de correlació i l'anàlisi factorial, tècnica estadística molt més complicada i per tant més críptica, aportarien una nova perspectiva a la mesura de la intel·ligència i a l'estudi de les diferències entre grups.

El cas Burt

Analitzem ara el cas de Sir Cyril Burt, fervent partidari de considerar la intel·ligència hereditària. El seu principal argument es basava en el coeficient de correlació per al IQ (coeficient d'intel·ligència) entre bessons idèntics, edu-

cats juntament. La correlació era de 0,94. Burt va "aconseguir" una mostra de 53 bessons educats separatament i va obtenir una correlació de 0,87, una disminució bastant petita, que provava clarament que la intel·ligència era una característica hereditària, poc influïda per l'ambient. Però més tard, es va descobrir que la mostra era falsa, que no existien els 53 bessons educats separatament. I quan altres científics li van demanar les dades, Burt es va dedicar a inventar-se-les a fi i efecte que coincidissin amb les correlacions que havia publicat. També eren falses les seves "col·laboradores", dues dones que suposadament van recollir i elaborar les dades de l'estudi.

Aquest fals estudi va tenir greus conseqüències. Burt també era partidari de la teoria unifactorial, unidimensional, de la intel·ligència, l'anomenat factor G,

o dimensió general de la intel·ligència, que afirma que qualsevol test quantitatiu X verifica l'equació $X = aG + bU$, on G mesura el factor g (comú a tots els tests), U és el factor específic del test X i a, b són coeficients. La intel·ligència té, de fet, més dimensions, però Burt només n'acceptava i n'estudiava una, en el sentit que totes les habilitats mentals correlacionaven amb el factor G, que les dominava. Aquesta creença li permetia ordenar els individus de menys a més intel·ligents. Però, ves per on, resultava que les classes altes eren d'entrada més llestes que les classes rurals i treballadores. Per tant, els rics tindrien fills llestos i els pobres fills ximpls, perpetuant les diferències entre estrats socials. Aquestes conclusions no tindrien tanta importància si es limitessin a figurar en un estudi acadèmic. Però Burt, que va ser editor del *British Journal of Statistical Psychology* (1947-1963), tenia també el



càrrec de Psicòleg Oficial del London County Council. Les seves interpretacions del caràcter hereditari de la intel·ligència van justificar un ajut més gran per a les escoles de les classes altes que per a les de classes humils. L'argument era: com que els humils són menys intel·ligents, no cal invertir gaire en les seves escoles, seria malgastar els diners.

La rèplica "científica" a Burt vindria posteriorment de la mà del psicòleg americà Louis Thurstone, que estudiaria el model multifactorial i postulava més dimensions per a la intel·ligència. En Matemàtiques, és ben sabut que podem ordenar els punts de més petit a més gran al llarg d'una dimensió, però no ho podem fer pas en dues o més dimensions. Amb el model multifactorial ja no es podia fer cap ordenació. Meditem una mica: si ni tan sols podem ordenar els animals inferiors respecte a mesures biomètriques (Fig. 1), no és possible ni té cap sentit (científic) plantejar l'ordenació dels éssers humans segons la seva capacitat mental, si la intel·ligència té moltes dimensions.

Les interpretacions de Jensen

Arthur Jensen és un destacat psicòleg nord-americà, autor de "Bias in Mental

Testing" (1979), que va defensar –i encara defensa– que les diferències d'intel·ligència entre blancs i negres tenien una base hereditària. Naturalment, va ser qüestionat per estadístics com Jack Kaplan, per l'ús incorrecte que havia fet de l'Estadística. Jensen era essencialment partidari de la teoria del factor G, i defensava que l'herència, amb una influència de fins a un 80%, i no l'ambient, explicava la intel·ligència. Curiosament, va partir de les dades sobre bessons de Burt per justificar les seves teories. Com en el cas de Burt –que va tenir el suport del psicòleg Hans J. Eysenck, el qual afirmava que tot era un confabulació de l'esquerra que defensava la influència del medi–, Jensen també va ser defensat per altres estadístics. Hàbilment, Jensen responia a qüestions diferents d'aquelles per les quals havia estat criticat, i eludia donar explicacions sobre els seus errors. A aquest partidari convençut de les diferències d'IQ entre blancs i negres, Luca Cavalli-Sforza i Stephen Gould li van desmuntar els arguments, fent-li veure que l'ambient hi influïa, i molt. Jensen es va defensar de Gould titllant la seva crítica de marxista, en un estil semblant al adoptat per Eysenck quan va donar suport a Burt. A més, Sandra Scarr, psicòloga

americana, va estudiar dades sobre negres adoptats i blancs en condicions similars i no va trobar diferències, com tampoc les va trobar l'anglesa Barbara Tizard comparant nens blancs i negres que provenien d'orfenats. Els resultats més sorprenents van ser el d'un estudi que comparava japonesos i nord-americans. Els primers guanyaven en 11 punts als segons, una diferència en IQ propera a la que s'havia trobat, suposadament, entre blancs i negres. Científics nord-americans no van acceptar aquesta inferioritat racial –com un segle abans els francesos no acceptaven les diferències amb els alemanys–, i van argumentar que la diferència era deguda a la qualitat superior de les escoles japoneses, que exigia molt més esforç als nens, sobretot per aprendre a escriure el seu complicat idioma. Altrament dit, la diferència era només ambiental.

Comentem ara alguns errors d'interpretació de Jensen. L'IQ tenia una mitjana i una desviació típica més altes en els blancs que en els negres. Segons Jensen, els negres eren més ximplers i més homogenis. Els blancs eren més llestos però amb més variabilitat, i aquesta distinció "només es podia explicar genèticament". En realitat, la diferencia

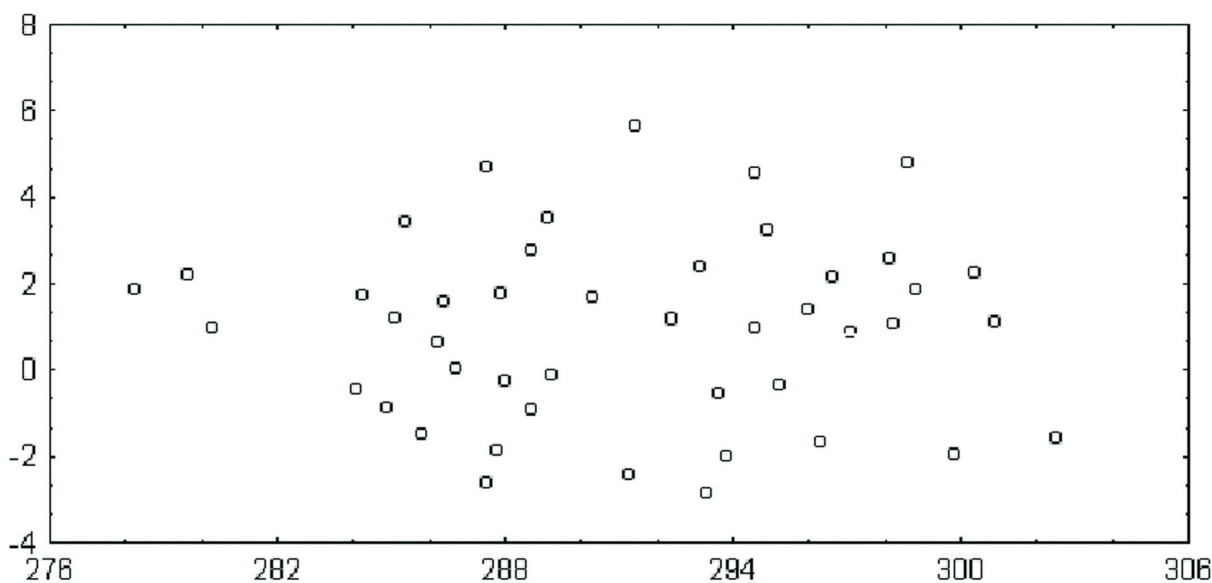


Fig.1. Representació de 49 pardals respecte dels dos primers eixos factorials (grandària i forma) sobre la base de 5 mesures biomètriques. Podem ordenar-los al llarg de l'eix horitzontal (grandària), però no és possible una ordenació tenint en compte les dues dimensions.

en la variabilitat era només un “artefacte estadístic”, causat per la forma de passar a una escala que donava una mitjana 100 i una desviació típica de 15. Les dues escales no eren comparables. El resultat era diferent si es partia dels resultats dels blancs o dels resultats dels negres. D'altra banda, no es va tenir en compte la selecció feta en les dades sobre els negres, ja que, pel fet de viure generalment en unes condicions més precàries, els bebès negres no naixien amb les mateixes condicions sanitàries que els blancs, i per tant, hi havia més individus negres amb problemes des del naixement, cosa que provocava una clara diferència en “viabilitat”. Jensen ho va acceptar, però va observar que mentre hi havia una diferència en “viabilitat” mitjana equivalent a la meitat d'una desviació típica, la diferència en intel·ligència era d'una desviació típica. I això li donava força per afirmar que la diferència genètica que afavoria els blancs per sobre dels negres, era superior a la diferència ambiental. De nou, l'ús de l'estadística és incorrecte, ja que Jensen comparava les mitjanes, partint d'una injustificada “igualtat” de desviacions típiques de la variable “viabilitat” en blancs i en negres. No és correcte afirmar que la diferència de mitjanes és la meitat de σ si la desviació típica σ no és la mateixa a cada població (Fig. 2).

Jensen també va cometre errors d'interpretació en la predicció de la intel·ligència. Si un blanc tenia intel·ligència 120, els seus fills tenien intel·ligència 110 de mitjana. Si un negre tenia intel·ligència 120, els seus fills tenien intel·ligència 100, també de mitjana. I aquesta diferència, segons Jensen, no es podia explicar només per l'ambient. Sabem, per la teoria de la regressió a la mitjana de Galton, que els pares intel·ligents tenen fills intel·ligents però menys intel·ligents, en mitjana, que els pares. No obstant això, la predicció de la intel·ligència en els blancs seguia una equació diferent i, per tant, no era comparable a la predicció per als negres. Per exemple, si suposem una mitjana de 100 en blancs i de 80 en negres, i un coeficient de regressió de 0,5, tenim que $110 = 100 + 0,5(120-100)$ en els blancs i

$100 = 80 + 0,5(120-100)$ en els negres. La diferència entre 110 i 100 era d'esperar, estadísticament parlant, i no hi havia motius per atribuir-la a la genètica. A més, això no demostraria que la diferència no pogués ser també ambiental.

The Bell Curve

El llibre *The Bell Curve* (1994) del psicòleg Richard Herrnstein i el científic Charles Murray, un estudi sobre la comparació i distribució de la intel·ligència, va ser causa de gran polèmica. Dotzenes d'articles i moltes discussions per ràdio i televisió, donaren suport o posaren

existència del factor G. Però alguns, com Jack Kaplan, opinaven que si no té sentit mesurar habilitats físiques (córrer, saltar, nedar) i aleshores aplicar una anàlisi factorial, tampoc té sentit fer-ho amb els tests. A més Kaplan, que era professor d'estadística, va posar en dubte que els tests realment mesuressin la intel·ligència, qualificant l'estudi de l'IQ com una pseudo-ciència. Per si fos poc, va trobar un important error a *The Bell Curve*. Els autors del llibre afirmaven que si un grup, com ara els blancs, tenia una mitjana superior en IQ que un altre grup, per exemple els negres, i l'IQ estava positivament correlacionat amb una altra

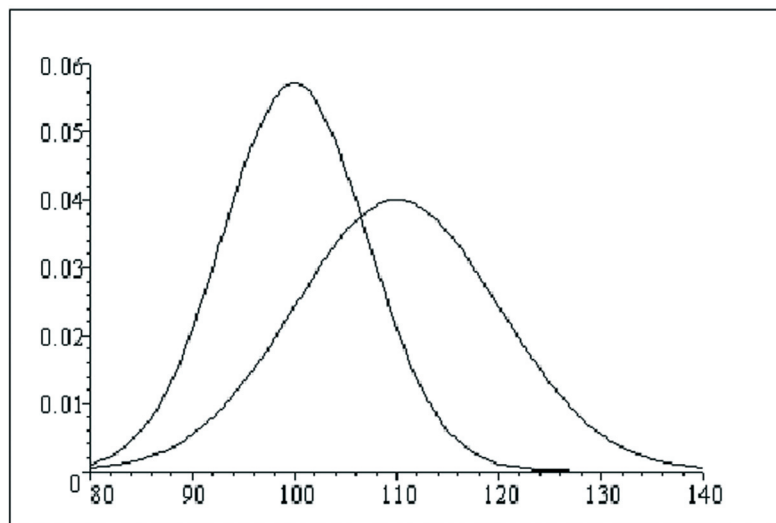


Fig.2. No és correcte mesurar les diferències entre mitjanes en termes d'una o mitja desviació típica si les desviacions típiques de les dues poblacions són diferents.

en dubte les seves teories i els resultats. Com el títol ens indica, és un estudi de la corba en forma de campana, fent clara referència a la distribució normal de la intel·ligència.

Els partidaris del llibre opinaven que, de la mateixa manera que podem ordenar individus per la seva talla, o per la seva presència física, també els podem ordenar segons la seva intel·ligència. Altres replicaven que la mesura de la intel·ligència és bàsicament una mesura de l'aprenentatge escolar i que, a més, depèn de molts factors: socials, econòmics, hàbits de lectura i condicions de salut. De fet, l'ordenació en una escala lineal és deguda a la controvertida

variable Y, que mesurava una habilitat no intel·lectual, però amb la mateixa mitjana a ambdós grups, aleshores hi hauria una superioritat individual. Concretament, si un blanc i un negre tenien el mateix IQ, caldria esperar que el blanc obtingués un valor superior per a la variable Y. La superioritat en IQ arrossegaria la superioritat en Y. L'afirmació era errònia: en ésser l'IQ del negre relativament superior respecte a la mitjana del seu grup, el valor predit d'Y donat un IQ superarà el valor predit per al blanc. És fàcil provar-ho dibuixant uns eixos IQ-Y i dues rectes paral·leles (Fig. 3). Per a un mateix IQ, la recta de l'esquerra (negres, N) dona una Y més alta que la recta més a la dreta (blancs, B).

Ignorar una variable

Un error molt comú en la pràctica de l'Estadística apareix quan hom ignora alguna variable de relleu. Durant la guerra de Cuba entre Espanya i els Estats Units (1898), la propaganda americana animava els joves a allistar-se, al·legant que morien 9 de cada mil homes entre els soldats expedicionaris i en canvi 16 de cada mil homes a Nova York. Allistar-se "augmentava" la probabilitat de viure. En un estudi fet els anys 1972-74 a Newcastle upon Tyne (Gran Bretanya), s'investigava l'efecte del tabac en 1.314 dones, amb problemes de tiroides i cor. Es va registrar la supervivència de les dones estudiades al cap de 20 anys.

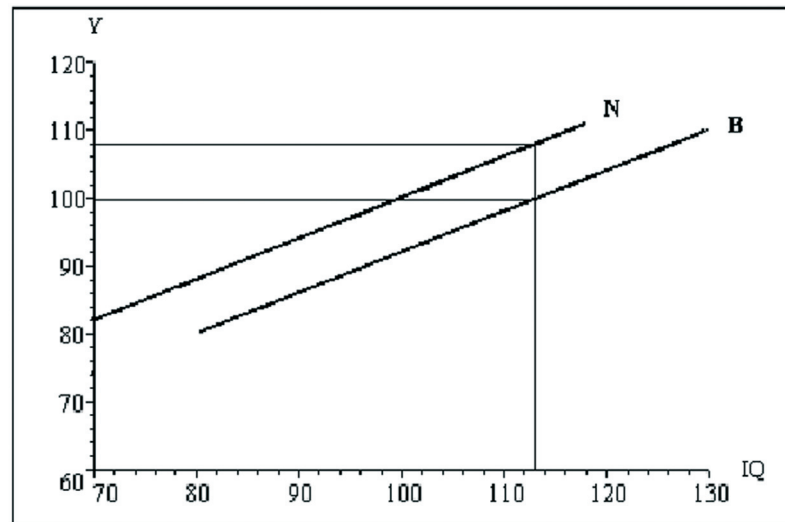


Fig.3. Un IQ superior en mitjana no implica una predicció superior en una altra variable Y. Encara que l'IQ mig sigui superior (recta de regressió B), la predicció d'Y per a un mateix IQ és més gran en el grup N (recta de regressió N).



Fumar va resultar "beneficiós", ja que només el 24% de les fumadores van morir, contra el 31% de les no-fumadores. Com en el cas dels soldats, l'error d'interpretació era degut a ignorar l'edat, principal causa de la mortalitat. La majoria de dones grans no fumava. Per exemple, de les dones de 65 anys o més, van morir el 85% però només en fumava el 20%. En canvi, de les de menys de 35 anys, van morir el 3% però fumaven el 46%.

Un altre exemple: suposem que no hi ha correlació entre mesures del crani i pigmentació, però prenem i mesquem dades d'una població alpina i una altra nòrdica. Aleshores apareixerà una falsa correlació deguda a la separació de les dues poblacions. Suposem ara que una malaltia hereditària està produïda per dos gens A i B indistingibles, i que les edats d'aparició en el pare i en el fill són independents dintre d'una mateixa família (mateix gen). Si considerem tota la població, com que hi ha dos gens, detectarem una falsa correlació, que ens farà creure que hi ha dependència. Sovint, l'usuari dels mètodes estadístics arriba a conclusions errònies pel fet d'ignorar que tenim dues o més poblacions mesclades, o que hi ha una variable no tinguda en compte i que influeix en els resultats.

Petita estadística dels errors estadístics

Presentem finalment algunes dades estadístiques sobre equivocacions en l'ús de l'Estadística, tema que ha estat analitzat des de fa temps.

S. M. Gore i els seus col·laboradors (1977) van estudiar els errors en 62 articles que contenien

estadístiques de 13 números consecutius del *British Medical Journal*. Els errors eren sobre la inadequada descripció de les dades, sobre confusions en la independència estadística, aleatorització incorrecta, errors en la t de Student o en la khi-quadrat. Van trobar que 30 no contenien errors però 32 sí. Aquests 32 estaven distribuïts en 19 articles amb 1 error, 11 articles amb 2 errors i 2 articles amb 3 errors.

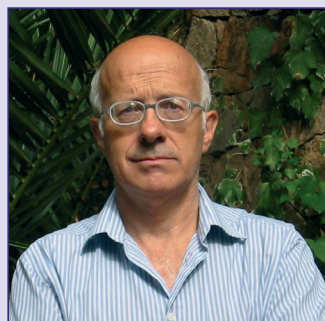
M. Martín i els seus col·laboradors van fer un estudi semblant (1982), revisant 595 articles publicats a *Medicina Clínica*, entre 1971 i 1979. Van trobar que, si bé la utilització de l'Estadística havia augmentat molt, passant del 20% (1971) al 50% (1979) dels articles, la proporció de treballs estadísticament inadequats també havia augmentat, essent el 43% (1971) i el 51% (1979).

H. Wulff i el seus col·laboradors (1987) va interrogar metges sobre la comprensió de conceptes bàsics ($p < 0,05$, desviació típica, error estàndard, coeficient de correlació), i van concloure que el coneixement que en tenien era molt limitat, malgrat que la majoria dels metges accepten que entendre l'Estadística és necessari.

Finalment, E. García-Berthou i C. Alcaraz (2004) van analitzar possibles incongruències entre proves estadístiques i valors p a tots els articles dels volums 409-412 de *Nature* (2001) i a 12 articles escollits a l'atzar dels volums 322-323 de *British Medical Journal*. Cal tenir en compte que, per a molts investigadors, la interpretació dels resultats biomèdics (si hi ha significació o no) es basa en el grau de significació o valor p. En general, si $p < 0,05$ hom accepta que hi ha significació. Els autors van detectar incongruències de càlcul (transcripció i d'arrodoniment) en el 11% dels articles, i algun error en 1 de cada 3 articles. D'altra banda, si no hi ha significació, convé saber que el valor p és un valor uniformement distribuït entre 0 i 1, i tant pot donar 0,572 com 0,034. Per tant, les conclusions en funció del grau de significació p s'haurien de prendre amb cautela.

Referències

- L. L. Cavalli-Sforza (1999) *¿Quiénes somos? Historia de la diversidad humana*. Ed. Crítica, Barcelona.
- C. M. Cuadras (2003) *Report. Una narració científica*. EUB, Barcelona.
- S. J. Gould. (1997) *La falsa medida del hombre*. Ed. Crítica, Barcelona.
- C. R. Rao (1994) *Estadística y verdad*. PPU, Barcelona. 2ª Ed., Pub.UB, 2004.



Carles M. Cuadras (Figueres, 1945) és llicenciat (1968) i doctor (1973) en Ciències Matemàtiques. Analista del Seminari Matemàtic del CSIC (1974), professor agregat (1979) i catedràtic (1983) d'estadística de la Universitat de Barcelona, ha publicat i editat 9 llibres i més de 100 treballs de recerca. Ha visitat les universitats de Londres, Paris VI, Pittsburgh, Wrocław, Tartu, McGill, Penn State, Simon Fraser, Stanford i La Frontera. Ha organitzat un congrés nacional i dos congressos internacionals. Ha treballat en diferents temes: probabilitats, estadística, anàlisi multivariant i aplicacions a les ciències naturals i socials.