

**UVIC**  
**LLIÇÓ**  
**INAUGURAL**  
**2005-2006**

**INCERTESA I CONEIXEMENT**

Dra. M. Luz Calle

## INCERTESA I CONEIXEMENT

Incertesa i coneixement són dos conceptes oposats, antagònics, però que molt sovint es veuen obligats a conviure. La seva gran rivalitat serà una constant que els acompanyarà en tot moment, cadascun amb l'objectiu d'apoderar-se de l'altre. L'estadística serà el pont entre aquest dos vells enemics.

### EL PAPER DE L'ESTADÍSTICA EN LA INVESTIGACIÓ

Moltes de les preguntes científiques que ens plantejem, referents tant a qüestions essencials de la vida com a qüestions cosmològiques, tenen a veure amb fenòmens aleatoris, anomenats també estocàstics, és a dir, fenòmens no reproduïbles, envoltats d'incertesa i de variabilitat. És per aquest motiu que l'estadística entra a jugar un paper molt important en la investigació científica.

L'estadística és la ciència que aborda l'estudi de la variabilitat, de la incertesa i de la presa de decisions davant d'aquesta incertesa. Té dos vessants diferenciats, d'una banda, és una part de la matemàtica aplicada, amb el seu propi camp d'estudi, i de l'altra és un instrument que utilitzen àmpliament altres ciències. L'estadística estudia com obtenir conclusions de la investigació empírica mitjançant l'ús dels models matemàtics. En aquest sentit és un pont entre els models matemàtics i els fenòmens reals.

Justament, la pròpia naturalesa de l'estadística de quantificar la incertesa ha propiciat una falsa i injusta imatge d'aquesta àrea de les matemàtiques i, malgrat la seva rellevància científica i pràctica, no sempre se li ha atribuït el reconeixement que es mereix. Aquesta desconfiança ve, d'una banda, de la confusió de pensar que la causa de la incertesa és el mateix mètode estadístic, quan el cert és que la incertesa és intrínseca al fenomen natural que s'està estudiant i, d'altra banda, del mal ús que de vegades es fa de l'estadística. Per a uns l'estadística és la ciència que oculta les diferències individuals a través de les mitjanes. Haureu sentit a dir allò que *L'estadística és la ciència que explica que si tu t'has menjat dos pollastres i jo cap, de mitjana ens n'hem menjat un cada un.* O allò que *Si un home té el cap al forn i els peus a la nevera, el cos està a una temperatura mitjana ideal.* Altres desconfien de l'estadística per la manipulació

que es pot fer de les dades. Potser haureu sentit allò de: *No hi ha conjunt de dades que, torturat adequadament, es resisteixi a confessar allò que es vol sentir*. Davant d'això, l'única cosa que puc dir és que no hi ha mètode estadístic que sigui capaç de vèncer la manca d'honestedat. L'únic antídote davant d'aquesta possible manipulació és tenir un coneixement bàsic dels mètodes estadístics. En aquest sentit és molt necessari que qualsevol estudiant d'una disciplina científica o que involucri l'experimentació empírica aprengui, des de l'inici dels seus estudis, el significat d'error experimental i de variabilitat. Serà l'única manera que en el futur en pugui fer un bon ús i pugui ser crític amb els resultats presentats per altres investigadors a les revistes especialitzades.

Justament, l'ús o abús que es fa de l'estadística a les revistes científiques és també un tema controvertit ja que és difícil trobar el punt d'equilibri entre una explicació excessivament formal del mètode utilitzat i una explicació massa superficial. He volgut transcriure-us el catàleg dels *set pecats capitals en la investigació mèdica* que, de forma molt irònica i perspicaç, va presentar J.L. Segú al V Encuentro Marcelino Pascua (1995), una trobada anual sobre salut pública:

**Luxúria:** La pateixen els estadístics que troben plaer a utilitzar la tècnica més complexa possible amb el propòsit secret d'impressionar i fer més difícil la vida als soferts editors i lectors. S'obliden que en els articles mèdics l'estadística ha d'estar al servei de la medicina.

**Peresa:** La pateixen aquells metges als quals no els ve de gust utilitzar tècniques quantitatives per tal de presentar les seves dades.

**Supèrbia:** La pateixen aquells clínics amb coneixements mínims d'estadística i que es creuen autosuficients. Corren el risc d'utilitzar inadecuadament una tècnica estadística ja que els programes informàtics no distingeixen un ús adequat d'un d'inadequat.

**Enveja:** Plagi o còpia de mètodes utilitzats en altres articles. Practicada amb prudència pot ser beneficiosa i productiva.

**Ira:** Utilitza l'estadística amb enuig. El metge entén que necessita l'estadística, però es nega a aprendre-la.

**Gola:** Diversitat d'excessos. La cometen tant els bioestadístics com els metges quan es plantegen massa objectius, quan analitzen massa variables o quan extreuen massa informació.

**Avarícia:** Necessitat incontrolable d'augmentar el nombre de publicacions. La cometen ambdós col·lectius sota la pressió cada cop més acusada de

publicar i un desig irrefrenable de rendibilitzar tot el que fan per tal d'enreixar el currículum.

## BREU RESSENYA HISTÒRICA

Els orígens del que actualment entenem per estadística s'han de situar en els primers intents de formalitzar matemàticament dos conceptes inicialment sense connexió: el tractament de la incertesa i el tractament de dades.

Els conceptes d'atzar i incertesa són tan antics com la mateixa civilització. La humanitat sempre ha hagut de suportar la incertesa, per exemple, sobre el clima, el subministrament d'aliments i sobre molts altres factors de l'entorn, i s'ha esforçat a reduir-la per minvar-ne els efectes adversos.

També els jocs d'atzar tenen una llarga història. El llançament de l'astràgal es considera el joc d'atzar més antic, precursor del joc dels daus. L'astràgal és un os del taló de xais i altres vertebrats i s'ha trobat en grans quantitats en les excavacions arqueològiques més antigues. Sembla que s'utilitzaven tant per al joc com en cerimònies religioses. A poc a poc l'astràgal va ser reemplaçat pel dau i ja a l'antic Egipte els jocs de daus eren molt populars. L'atzar s'explicava a les civilitzacions antigues com un efecte de la voluntat divina. Els oracles i sacerdots de Grècia i de Roma utilitzaven el resultat obtingut en el llançament de quatre daus per predir el futur i la voluntat dels déus. Per exemple, l'aparició de la configuració Venus, és a dir, un 1, un 3, un 4 i un 6 en la tirada de quatre daus, representava una voluntat favorable dels déus. Ciceró (106-43 AC) en la seva obra *De Divinatione* reflexiona sobre l'atzar i sobre el que actualment entenem com a esdeveniments molt i poc probables:

*Es llancen quatre daus i resulta Venus, això és atzar; però també pensaràs que és atzar si en fer cent llançaments obtens Venus un centenar de vegades?*

Durant l'edat mitjana hi va haver una gran activitat científica i artística a l'Orient i el nom d'atzar podria tenir el seu origen en la paraula àrab *zahr*, que significa flor, per la flor que els àrabs pintaven en una de les cares dels daus.

Però els orígens de la teoria de la probabilitat, que serà el marc formal per quantificar i tractar matemàticament la incertesa, daten del 1654 a París, quan Chevalier de Mére, un gran jugador de l'època, planteja al matemàtic Blaise

Pascal (1623-1662) una sèrie de qüestions, d'entre les quals la més coneguda és el problema dels punts:

*Dues persones decideixen jugar una sèrie de jocs fins que una d'elles en guanyi sis. Cada jugador aposta la mateixa quantitat de diners i qui guanyi els sis jocs s'endurà tot el pot. Però suposem que, per alguna raó, els jugadors es veuen obligats a finalitzar la partida prematurament, quan el primer jugador portava guanyades 5 partides i el segon només 3. Com s'haurien de repartir la quantitat total apostada?*

Val a dir que la causa més probable per la qual els jugadors haurien d'acabar la partida abans d'hora seria l'arribada de la policia, ja que en aquella època el joc estava prohibit. La resposta correcta al problema dels punts és que el primer jugador rebí set vuitens del pot total, resultat que s'obté comptant en quantes de totes les possibles opcions de finalitzar el joc guanyaria el primer jugador.

El problema va intrigar Pascal, que el va compartir amb el brillant matemàtic Pierre Fermat (1601-1665). La correspondència entre Pascal i Fermat no solament va derivar en la resolució del problema dels punts, sinó que va establir els fonaments per a més resultats generals.

Durant el segle XVIII el càlcul de probabilitats es consolida gràcies a les contribucions d'importants matemàtics i físics. El factor impulsor més important d'aquestes contribucions són els problemes associats a la contrastació empírica de les lleis de la gravitació universal de la teoria de Newton (1687). Un dels primers problemes va ser el tractament dels errors de mesura. El matemàtic Daniel Bernoulli (1700-1782) va ser el primer a proporcionar una solució al problema d'estimar una quantitat desconeguda a partir d'un conjunt de mesures del seu valor que, degut a l'error experimental, presentava variabilitat. Però l'impuls més important d'aquesta època ve dels treballs de Pierre Simon de Laplace (1749-1827) i de Carl Friedrich Gauss (1777-1855) que contribueixen als fonaments de la construcció de models probabilístics i a la seva utilització per obtenir conclusions a partir d'unes dades. Laplace va introduir la primera definició formal de probabilitat i va desenvolupar la llei normal com a model per descriure la variabilitat dels errors de mesura. Gauss va resoldre de manera general el primer problema d'estimació en plantejar-se la predicció de la posició d'un planeta en un determinat instant en funció de les posicions de  $k$  cossos i del valor de certes constants desconegudes.

Però els avenços en la teoria d'errors es van restringir durant molt de temps al tractament de problemes de la física i l'astronomia i van tenir poca influència en altres àrees.

Durant la darrera dècada del segle XIX i el principi del segle XX, les ciències experimentals com l'agricultura, la genètica o la biologia en general van ser el motor i la font d'inspiració de molts mètodes estadístics que ara es consideren els fonaments de l'estadística moderna. Si l'obra de Newton va representar una revolució per a la física, la teoria de l'evolució de Darwin (1859) ho va ser per a la biologia. L'estadística va resultar ser la fórmula més apropiada per contrastar empíricament els mecanismes de la selecció natural. La producció de variabilitat es podria abordar amb el càlcul de probabilitats mentre que la selecció natural es podria contrastar amb l'estudi de la correlació entre determinades característiques genètiques i el grau de supervivència i descendència dels individus d'una espècie. El britànic Francis Galton (1822-1911) que, per cert, era cosí de Darwin, va ser el primer a destacar la necessitat d'utilitzar els mètodes estadístics per contrastar les teories de l'evolució. Galton va dedicar molts esforços a l'estudi de l'herència humana i va introduir el concepte de recta de regressió en comparar les alçades de pares i fills. La següent cita de Galton evidencia la seva admiració pel fet que l'aparent caos es pugui explicar mitjançant una llei de probabilitat, en aquest cas amb la llei normal.

*No conec res tan efectiu per impressionar la imaginació com la forma meravellosa de l'ordre còsmic, expressat per la llei de freqüència de l'error. Si aquesta llei hagués estat coneguda pels grecs, l'haurien personificada i deificada.*

Posteriorment, amb el propòsit de buscar un instrument formal per a l'exploració empírica del món, els també britànics Karl Pearson (1857-1936) i Ronald A. Fisher (1890-1962), considerats els pares de l'estadística actual, formalitzen les bases de la inferència estadística: elecció d'un model probabilístic, estudi de les seves propietats, estimació dels seus paràmetres i validació mitjançant proves de significació. En aquest moment queda clarament establert el lligam entre els models probabilístics i l'anàlisi de dades.

Paral·lelament al desenvolupament de la inferència estadística, l'escola matemàtica russa fa grans contribucions a la teoria de la probabilitat. Destaquen especialment Andrei A. Markov (1856-1922), responsable dels anomenats processos o cadenes de Markov i Andrei N. Kolmogorov (1903-1987) que formalitza axiomàticament el concepte de probabilitat.

Durant la primera meitat del segle XX s'estén l'aplicació dels mètodes estadístics a àrees ben diverses. En enginyeria Walter A. Shewart (1891-1968) introdueix els primers mètodes de control de qualitat i Norbert Wiener (1894-1964) i Claude E. Shannon (1916-2001), pare de la teoria de la informació, són pioners en l'aplicació dels mètodes estadístics en control de processos i codificació de senyals. Els problemes de predicció i extracció de senyals donen lloc al desenvolupament de la teoria dels processos estocàstics (Wiener i Kolmogorov). En aquesta època neix l'econometria com a conseqüència de l'aplicació dels mètodes estadístics a l'economia.

A partir dels anys 50 es considera que comença l'època moderna de l'estadística. La formalització dels mètodes estadístics va representar un avenç molt important en la investigació científica i des d'aleshores s'ha convertit en una eina fonamental i indispensable per a la investigació, tant en ciències experimentals com observacionals. Durant la segona meitat del segle XX s'han proposat models com el de regressió logística, que ha resultat ser pràcticament indispensable en la majoria d'estudis mèdics. S'ha desenvolupat la teoria de sèries temporals, d'una gran rellevància en econometria i estadística industrial. S'ha avançat molt en estadística espacial mitjançant la formulació de models estadístics per a l'estudi de factors que varien en el temps i en l'espai i que tenen una gran importància en estudis mediambientals.

## RECERCA EN ESTADÍSTICA

Per il·lustrar en què consisteix la recerca estadística em centraré en una de les seves àrees d'estudi: l'anàlisi de la supervivència.

L'estadística no solament s'enfronta a la incertesa deguda a la variabilitat entre individus, també s'enfronta a tipus d'incertesa més complexos relacionats amb la dificultat d'observar la mesura que es vol estudiar. Això succeeix molt especialment quan l'interès és l'estudi del temps fins que es produeix un determinat esdeveniment. Molt sovint aquests esdeveniments són adversos, com la mort, el diagnòstic d'un càncer, la caducitat d'un aliment o la fallada d'un determinat mecanisme. Però també poden ser positius, com recuperar-se d'una malaltia, trobar feina o tenir un fill. O difícils de classificar, com per exemple, una separació matrimonial. En qualsevol cas es parla de dades de supervivència independentment del tipus d'esdeveniment.

La característica que distingeix les dades de supervivència és que la variable d'interès és un temps i per tant no es mesura com la majoria de variables. En general les variables es mesuren instantàniament, mentre que en supervivència, per obtenir la mesura, cal esperar que passi el temps i es produeixi l'esdeveniment d'interès. El fet que el temps es mesuri seqüencialment té com a conseqüència el que es denomina censura, que es produeix quan es disposa d'informació incompleta sobre els temps de supervivència d'alguns individus.

El tipus més habitual de censura és l'anomenada censura per la dreta, que es produeix quan al final del període de seguiment alguns individus encara no han experimentat l'esdeveniment d'interès. Per a aquests, es diu que el seu temps de supervivència està censurat per la dreta, ja que no sabem el seu valor exacte, però sabem que és superior al temps de seguiment. Per exemple, si l'esdeveniment d'interès és la mort després del diagnòstic d'una malaltia greu i l'estudi té una durada de dos anys, pot passar que després dels dos anys de seguiment alguns pacients encara siguin vius. No coneixem el temps de supervivència d'aquests pacients, però sabem que és superior als dos anys. Aquesta manca d'informació precisa comporta problemes tècnics considerables que fan que el tractament de dades de supervivència s'hagi de dur a terme amb una metodologia estadística específica que anomenem anàlisi de la supervivència.

Alguns d'aquests estudis es poden abordar amb mètodes de supervivència ben consolidats, com el model de Cox, proposat l'any 1972 per Sir David Cox, sens dubte una de les figures més importants actualment en estadística. El model de Cox o model de riscos proporcionals és ara mateix un dels models més utilitzats en la investigació clínica. L'aplicació d'aquestes metodologies més consolidades, que es troben implementades en la majoria de paquets estadístics, molt sovint les realitzen directament els metges o científics responsables de l'estudi.

Altres estudis més complexos, com el que els exposo a continuació, requereixen la utilització de metodologia més específica o, fins i tot, del desenvolupament de nova metodologia. En aquests casos es fa necessària la intervenció de matemàtics o estadístics.

Es tracta d'un assaig clínic encara en curs, anomenat projecte TIBET, que porta a terme la Fundació Lluita contra la Sida a l'Hospital Germans Trias i Pujol de Badalona. Actualment els tractaments combinats d'antiretrovirals tenen una gran efectivitat i permeten que molts pacients infectats amb el virus de la Sida convisquin amb la malaltia amb una qualitat de vida prou bona. Però l'ús con-

tinuat d'aquests tractaments també té efectes adversos. Pot provocar toxicitats importants o resistència a algun dels medicaments. Per aquest motiu l'objectiu d'aquest estudi és explorar els beneficis de la interrupció controlada del tractament. És a dir, determinar si la interrupció controlada és segura, si redueix el risc de desenvolupar toxicitat o resistència als medicaments i si, en definitiva, millora la qualitat de vida dels pacients. Durant la interrupció es controla regularment l'evolució de l'estat immunològic dels pacients que es mesura mitjançant dos marcadors, el nombre de cèl·lules CD4 a la sang i el valor de la càrrega viral. Quan algun d'aquests dos marcadors traspasa un cert llindar establert de seguretat el pacient ha de tornar a iniciar el tractament. L'assaig porta ja gairebé dos anys de seguiment i les dades de què disposem per a cada pacient són: els temps consecutius que el pacient ha pogut estar sense tractament, els temps de tractament que ha necessitat entre dues interrupcions consecutives, l'evolució del nombre de cèl·lules CD4 i de la càrrega viral. Des del punt de vista estadístic l'estudi presenta moltes dificultats metodològiques: requereix l'estudi conjunt de diversos temps consecutius per a cada individu, l'estudi de la dependència entre els diferents temps, la relació entre els temps i els marcadors immunològics, etc.

La complexitat de situacions com aquesta és l'estímul de bona part de la recerca en estadística. En primer lloc cal trobar la millor formulació matemàtica per a un determinat problema científic o mèdic; la millor formulació en el sentit que ens permeti extreure la màxima informació i de la forma més precisa possible. Per aconseguir això cal un coneixement prou profund de l'entorn del problema que només s'aconsegueix amb una bona comunicació amb els responsables de l'estudi. Un cop plantejat el model cal estudiar-ne les propietats mitjançant la teoria de la probabilitat.

Però la recerca en estadística no acaba en la formulació teòrica del model. Volem que tingui una utilitat i donar resposta al problema concret plantejat. Això s'aconsegueix amb l'estimació dels paràmetres del model i la seva validació mitjançant proves de significació. Però, com que es tracta de noves propostes metodològiques per a resoldre problemes complexos, en la majoria dels casos no podem obtenir les estimacions del model utilitzant programari existent i cal programar els algorismes específics d'aquella nova metodologia. La comunitat estadística cada vegada és més conscient de la necessitat que aquests esforços particulars en desenvolupament de programes es puguin divulgar i compartir, i amb aquest objectiu s'ha creat el projecte R (<http://www.r-project.org/>), un entorn a internet on es comparteixen les noves contribucions estadístiques.

## LES PERSPECTIVES DE FUTUR

Si els inicis del segle XX van ser una època daurada per a l'estadística gràcies als grans avenços que es van produir en les ciències experimentals, en aquest moment, a principis del segle XXI, ens tornem a trobar en un moment clau per a aquesta disciplina. En aquests darrers anys s'han produït una sèrie d'avenços científics i tecnològics relacionats majoritàriament amb la capacitat de recollir i emmagatzemar grans quantitats de dades, que fan que la rellevància científica i pràctica de l'estadística experimenti novament un creixement espectacular. A mesura que la tecnologia i la instrumentació augmenten, també ho fa la quantitat de dades recollides i la velocitat a què es poden recollir. Podem trobar moltes altres àrees on els investigadors s'enfronten cada vegada més a la necessitat d'una anàlisi massiva de dades que supera la capacitat d'anàlisi de les tècniques estadístiques habituals.

Aquesta situació ha provocat que en els darrers anys emergís espectacularment una nova filosofia per al tractament de dades, l'anomenada «minería de dades», en anglès *Data Mining*. Es tracta d'un enfocament menys formal consistent en una sèrie de tècniques computacionals per a l'extracció automàtica de possibles relacions o patrons existents en les grans bases de dades. De fet la minería de dades es troba en la frontera de l'estadística i de les tecnologies de la informació i s'hi utilitzen tècniques pròpies de la intel·ligència artificial, la gestió de bases de dades o el reconeixement de patrons.

La diferència essencial entre els mètodes estadístics i la minería de dades radica en com es tracta el procés natural que ha generat les dades. L'estadística segueix l'estructura del mètode científic en el sentit de plantejar una hipòtesi o model per a aquest procés. Aleshores, seguint un procediment formalment contrastat, s'accepta o es rebutja la hipòtesi en base a les observacions. En els mètodes de minería de dades, en canvi, el procés que ha generat les dades es considera com una caixa negra, desconegut i complex. S'utilitzen tècniques algorísmiques que, sense necessitat de modelar aquest procés, busquen relacions i estructures entre les dades.

Malgrat que moltes de les tècniques utilitzades en la minería de dades han estat proposades des de departaments d'estadística, el cert és que fins ara la comunitat estadística en general ha mirat amb molt de recel aquest enfocament alternatiu. El motiu és que generalment el funcionament de les tècniques de minería de dades no té al darrere una justificació teòrica i matemàtica ferma.

Això ha fet que la majoria d'estudis que implementaven tècniques de mineria de dades es portessin a terme exclusivament des de departaments d'informàtica o enginyeria. Sortosament, cada vegada més la comunitat estadística està prenent consciència de la necessitat de participar en tot aquest procés i no quedar-ne al marge. És indiscutible que l'èxit en aquest tipus d'estudi requerirà la col·laboració de totes les parts. S'hauran de crear equips multidisciplinaris que integrin les diferents perspectives, tant estadístiques com algorísmiques.

### EL TRACTAMENT DE DADES BIOLÒGIQUES POSTGENÒMIQUES

Aquesta situació la trobem molt especialment en el context de la biologia i la medicina. Els progressos recents en biologia molecular i en genètica obren completament noves àrees d'investigació en què l'estadística pot jugar un paper molt rellevant. El Grup de Recerca en Modelització de Sistemes Biològics de la UVic, al qual pertanyo, treballa en dos projectes en aquest àmbit:

El primer projecte s'emmarca en el camp de la genòmica evolutiva i té com a objectiu l'estudi de les bases genètiques del llenguatge humà. La revista *Nature* va publicar el mes de setembre passat un primer esborrany del genoma del ximpanzé i unes primeres conclusions de la comparació amb el genoma humà. Tot i que allò que ha transcendit més a l'opinió pública és la gran similitud entre els dos genomes –només un 1,23% de bases diferents–, no s'ha fet prou èmfasi en fet que aquest petit percentatge en termes absoluts representa una diferència d'aproximadament 35 milions de bases. L'estudi profund de quines d'aquestes diferències van representar realment un canvi evolutiu entre les dues espècies i les implicacions funcionals d'aquests canvis són els grans reptes que ens planteja aquest estudi i que encara cal respondre. Per abordar aquestes qüestions serà necessària la utilització de mètodes estadístics eficients de comparació de seqüències, per exemple els mètodes basats en processos ocults de Markov.

El segon projecte té a veure amb l'estudi de la relació entre malaltia i genètica. Se situa en el marc de la farmacogenòmica i l'epidemiologia genètica, que tenen com a objectiu l'aplicació dels coneixements moleculars al disseny de nous fàrmacs i de tractaments mèdics individualitzats, tenint en compte el perfil genètic de cada pacient. El component genètic de malalties complexes, com el càncer, malalties cardiovasculars, diabetis, etc., ve determinat per petits canvis combinats en diferents gens i per la seva interacció amb l'ambient. L'objectiu

és identificar aquests petits canvis genètics que estan associats a una major susceptibilitat a desenvolupar la malaltia. El nostre grup participa en un d'aquests estudis. Es tracta d'un dels estudis més grans sobre càncer de bufeta, en el qual participen l'Institut Municipal d'Investigació Mèdica (IMIM) de Barcelona, el Centre de Regulació Genòmica (CRG) de Barcelona, el Centro Nacional de Investigaciones Oncológicas de Madrid i la Universitat de Vic. L'objectiu de l'estudi és la identificació de variacions genètiques associades al desenvolupament i la progressió del càncer de bufeta, de forma independent o en interacció amb factors ambientals. Aquest objectiu representa un repte estadístic important com a conseqüència del gran nombre de factors genètics que cal analitzar simultàniament.

En resum, en els propers anys les grans quantitats de dades produïdes per la biologia molecular i la variabilitat de la resposta humana a les intervencions mèdiques seran el motor de bona part de la investigació futura en metodologia estadística.

Per acabar voldria dirigir les darreres paraules d'aquesta lliçó inaugural als estudiants. Aquesta ha estat una de les primeres lliçons d'aquest curs, però en queden moltes més. Aproveiteu-les i gaudiu-ne, perquè, tal com va dir Louis Pasteur (1822-1895) fent referència al factor sort en els descobriments científics:

*L'atzar només afavoreix les ments preparades*

Moltes gràcies i molt bon curs.

## REFERÈNCIES

- Abraira, V. i altres. *Mesa redonda: La estadística en la investigación Médica*. Qüestió. Vol 25(1): 121-156, 2001.
- Altman, D.G.; Bland J.M. *Time to event (survival) data* British Medical Journal, vol 317: 468-469, 1998.
- Hastie, T.; Tibshirani, R.; Friedman, J. *The elements of Statistical Learning. Data mining, Inference, and Prediction*. Springer-Verlag, NY, 2001.
- Peña, D. *Estadística. Modelos y Métodos 1. Fundamentos*. Alianza Universidad Textos, 1986.
- Thornton-Wells, T.A.; Moore, J.H.; Haines J.L. *Genetics, statistics and human disease: analytical retooling for complexity*. Trends in Genetics, vol 20, n. 12, 640-647, 2004
- Zeger, S.L.; Diggle, P.J.; Liang, K. *A Cox Model for Biostatistics of the Future*. <http://www.bepress.com/jhubiostat/paper32>, 2004





**Universitat de Vic**

Carrer de la Sagrada Família, 7

08500 Vic. Barcelona

Tel. 93 886 12 22

Fax 93 889 10 63

[www.uvic.es](http://www.uvic.es)