

Estadística

Josep Gibergans Bàguena
Àngel J. Gil Estallo
Carles Rovira Escofet

XP08/05057/02301



Universitat Oberta
de Catalunya

www.uoc.edu

Josep Gibergans Bàguena

Doctor en Ciències Físiques per la Universitat de Barcelona l'any 2001. És professor titular d'escola universitària del Departament de Matemàtica Aplicada III de la Universitat Politècnica de Catalunya. Des de l'any 1993 la seva activitat docent se centra en l'Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona en matèries de matemàtiques i estadística als estudis d'aquest centre. És consultor de la Universitat Oberta de Catalunya des de 1998.

Àngel J. Gil Estallo

Doctor en Ciències Matemàtiques per la Universitat de Barcelona des de l'any 1996. És professor titular d'escola universitària de la Universitat Pompeu Fabra des de 1991. La seva activitat docent se centra en temes de matemàtiques, estadística i informàtica als estudis d'Economia d'aquesta universitat. És consultor de la Universitat Oberta de Catalunya des de 1998.

Carles Rovira Escofet

Doctor en Matemàtiques per la Universitat de Barcelona. Professor titular al Departament d'Estadística de la Universitat de Barcelona. La seva recerca se situa en el camp dels processos estocàstics, principalment en l'estudi d'equacions diferencials estocàstiques.

Quarta edició: febrer 2009

© Josep Gibernans Bàguena, Àngel J. Gil Estallo, Carles Rovira Escofet

Tots els drets reservats

© d'aquesta edició, FUOC, 2009

Av. Tibidabo, 39-43, 08035 Barcelona

Disseny: Manel Andreu

Material realitzat per Eureka Media, SL

Dipòsit legal: B-2.291-2009

Cap part d'aquesta publicació, incloent-hi el disseny general i de la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric, com químic, mecànic, òptic, de gravació, de fotocòpia, o per altres mètodes, sense l'autorització prèvia per escrit dels titulars del copyright.

Introducció

És fàcil veure, només examinant el nostre entorn més immediat, la creixent importància de la multitud de dades estadístiques que ens envolten: estadístiques sobre preus, sobre habitatges, sobre ordinadors... També en la majoria dels llocs de treball l'anàlisi de dades està prenent una importància destacable, per exemple, en temes de control de qualitat, de *data mining*, etc. D'altra banda, en alguns temes més tècnics, com poden ser el processament de senyals i les comunicacions, són bàsics els anomenats models probabilístics.

A això s'uneixen les múltiples facilitats de càlcul estadístic que se'ns ofereixen avui en dia (paquets estadístics, pàgines web, etc.), que exigeixen, per tal de fer-ne un ús correcte, un bon coneixement de les tècniques estadístiques, de les seves possibilitats i de les seves limitacions.

L'estadística es preocupa d'estudiar situacions en les quals apareix l'atzar o, dit d'una altra manera, situacions en les quals prenem unes observacions en què hi ha elements fora del control de l'observador. La formulació d'una teoria matemàtica que ens ajudi a interpretar de manera coherent i científica aquest tipus de situacions és una eina fonamental en qualsevol tipus de recerca o estudi.

Tots hem sentit afirmacions com "Hi ha mentides i estadístiques" o "Amb l'estadística es pot demostrar qualsevol cosa". En aquest curs veurem que això no és cert. Potser es pot dir que amb l'estadística no som capaços de demostrar res, però comprovarem que és l'eina fonamental que ens proporciona tots els indicis de com és la realitat.

Les aplicacions de l'estadística les podeu trobar constantment: quan llegiu un diari, quan us arriben unes dades o uns gràfics a la feina, quan a la televisió parlen d'índexs d'audiència, etc.

Podem recordar el que va dir l'escriptor H.G. Wells:

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

Objectius

L'assignatura *Estadística* té com a objectiu general introduir les tècniques bàsiques estadístiques per a poder establir un pont entre els models matemàtics i els fenòmens reals, de manera que siguem capaços de determinar quina tècnica estadística es pot aplicar en cada situació, com s'ha d'aplicar i fins a quin punt podem estar segurs de les conclusions a què s'arriba.

Per a assolir aquest objectiu, s'ha distribuït l'assignatura en quatre parts. Tres parts tracten sobre la descripció i aplicació de diferents tècniques estadístiques: l'anàlisi de dades, la inferència estadística i els models de regressió. La quarta part, essencial per a entendre la metodologia pròpia de l'estadística, està dedicada a la teoria de la probabilitat, per a així poder donar sentit a les tècniques introduïdes i entendre'n les limitacions.

Descriurem a continuació els objectius més concrets a assolir en cadascuna de les parts mencionades:

1. Estadística descriptiva, introducció a l'anàlisi de dades i mostreig (mòduls 1 i 2). En aquesta part bàsicament s'introdueixen els conceptes que més endavant permetran fer un treball estadístic: definició de la variable d'interès, recollida i organització de dades i càlcul de paràmetres que resumeixin el comportament d'aquesta variable. També s'insisteix en la representació gràfica de les dades com a primer pas en el seu estudi.
2. Probabilitat i variables aleatòries (mòduls 3, 4 i 5). En aquesta part es descriu el model matemàtic dels fenòmens aleatoris que "suporta" la teoria estadística i que permet de construir els models estadístics d'una manera rigorosa: la teoria de probabilitats. S'hi estudia amb profunditat el concepte de variable aleatòria i s'hi presenten les més usals, entre les quals destaquen per la seva importància la distribució normal i les seves propietats. S'hi presenta la distribució t i també es tracta l'aproximació de la binomial a la normal com a pas previ per a presentar el teorema del límit central. També s'hi introdueix el concepte de control de qualitat i les seves aplicacions.
3. Inferència i modelització estadística (mòduls 6, 7, 8 i 9). En aquesta part es presenta la teoria que permet treure conclusions sobre una població a partir d'una mostra (un subconjunt dels individus de la població) i precisar amb quins marges de confiança són vàlides aquestes afirmacions. Com a aplicació de les distribucions mostrals i del teorema del límit central es defineixen els intervals de confiança per a la mitjana i la proporció d'una població, i s'insisteix en la interpretació del resultat que s'obté en calcular un d'aquests intervals. Finalment, s'hi introdueixen les distribucions necessà-

ries per a dur a terme la inferència estadística (concretament, la F de Snedecor i la χ^2 quadrat), i s'hi estudien bàsicament els contrastos d'hipòtesis sobre mitjanes i variàncies per a una o dues poblacions; aquests contrastos permeten, dins les limitacions de l'anàlisi estadística, decidir entre dues hipòtesis mútuament excloents referides als possibles valors de la mitjana, la proporció o la variància dels valors d'un conjunt de dades.

4. Regressió lineal múltiple i anàlisi de la variància (mòduls 10, 11 i 12). En aquesta part es desenvolupa la metodologia que permet explicar una variable com a combinació lineal d'unes altres, cosa que dóna lloc als anomenats models de regressió (ja sigui simple, on s'utilitza una única variable explicativa, o múltiple, on tenim més variables explicatives). També s'hi ha de fer inferència sobre algun dels paràmetres del model i entendre el concepte de bondat de l'ajustament. Finalment, s'hi introdueix la comparació de mitjanes entre diverses poblacions mitjançant l'anomenat ANOVA.

Com veieu, són uns objectius ambiciosos, que ens donaran una rica panoràmica de la pràctica i la metodologia de l'estadística.

Continguts

Mòdul 1

Estadística descriptiva

Àngel J. Gil Estallo

1. Tipus de dades i la seva representació gràfica
2. Mesures de centre i propietats
3. Mesures de dispersió

Mòdul 2

Mostreig

Àngel J. Gil Estallo

1. Mostreig

Mòdul 3

Probabilitat

Àngel J. Gil Estallo

1. Introducció a la probabilitat
2. Combinatòria i tècniques de recompte
3. Probabilitat
4. El teorema de Bayes

Mòdul 4

Variables aleatòries

Carles Rovira Escofet

1. Introducció a les variables aleatòries. Variables aleatòries discretes
2. Esperança i variància
3. Algunes distribucions discretes
4. Variables aleatòries contínues
5. Algunes lleis contínues. La llei normal
6. Processos estocàstics

Mòdul 5

Teorema del límit central

Carles Rovira Escofet

1. La distribució de la mitjana mostral
2. El teorema del límit central

Mòdul 6

Intervals de confiança

Àngel J. Gil Estallo

1. Introducció als intervals de confiança. El cas de la mitjana aritmètica
2. Intervals de confiança per a la proporció

Mòdul 7

Contrast d'hipòtesis

Carles Rovira Escofet

1. Introducció al contrast d'hipòtesis
2. Contrastos sobre la mitjana i sobre la proporció

Mòdul 8

Contrast de dues mostres

Josep Gibergans Bàguena

1. Contrastos sobre la diferència de mitjanes
2. Contrastos sobre la diferència de proporcions

Mòdul 9

Contrast de variàncies

Josep Gibergans Bàguena

1. Contrast de la variància
2. Comparació de variàncies

Mòdul 10

Regressió lineal simple

Josep Gibergans Bàguena

1. El model de regressió simple
2. La qualitat de l'ajust
3. Inferència en la regressió

Mòdul 11

Regressió lineal múltiple

Josep Gibergans Bàguena

1. El model de regressió múltiple
2. La qualitat de l'ajust
3. Inferència en la regressió lineal múltiple

Mòdul 12

L'anàlisi de la variància (ANOVA)

Josep Gibergans Bàguena

1. L'anàlisi de la variància (ANOVA)

