

# Matemàtiques i estadística amb R

Estadística, àlgebra lineal i càlcul

Daniel Liviano Solís

Maria Pujol Jover

PID\_00203023

**Daniel Liviano Solís**

Professor dels Estudis d'Economia i Empresa de la Universitat Oberta de Catalunya. Doctor en Economia per la Universitat Rovira i Virgili.

**Maria Pujol Jover**

Professora dels Estudis d'Economia i Empresa de la Universitat Oberta de Catalunya. Doctora en Estudis Empresari- als per la Universitat de Barcelona.

Primera edició: setembre 2013

© Daniel Liviano Solís, Maria Pujol Jover

Tots els drets reservats

© d'aquesta edició, FUOC, 2013

Av. Tibidabo, 39-43, 08035 Barcelona

Disseny: Manel Andreu

Realització editorial: Oberta UOC Publishing, SL

*Cap part d'aquesta publicació, inclòs el disseny general i la coberta, pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera, ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, gravació, fotocòpia, o qualsevol altre, sense l'autorització escrita dels titulars del copyright.*

## Introducció

### La importància de l'anàlisi estadística i quantitativa

Tant les matemàtiques com l'estadística són fonamentals en molts plans d'estudi: física, enginyeries, ciències socials, turisme, etc. Per això mateix, és tant important conèixer la teoria que hi ha al darrere dels mètodes i tècniques incloses en aquestes disciplines com saber-la aplicar en la vida real. Aquest manual vol ser un complement al material teòric de les assignatures quantitatives que s'ofereixen en els diferents plans d'estudi de la Universitat Oberta de Catalunya (UOC).

És un fet que avui dia no es concep l'estudi d'una assignatura quantitativa (matemàtiques, estadística, física, etc.) sense un suport computacional que faciliti els càlculs. Això explica la importància que normalment adquireixen els programes informàtics en el moment en què els estudiants opten per matricular-se en aquesta mena d'assignatures. Per aquest motiu i perquè realment creiem que la millor immersió de l'estudiant en qualsevol programa informàtic és a partir la pròpia experiència, l'estructura d'aquest manual consisteix en la resolució de problemes que es plantegen amb l'ajuda d'un programa específic.

### El paquet estadístic R

En aquest sentit, recentment hi ha hagut un gran desenvolupament de programes informàtics destinats a fer anàlisis numèriques de dades empíriques. Els més interessants són els que es distribueixen sota una llicència *GNU General Public License*, és a dir, els que no impliquen cap cost econòmic i són una alternativa clara als programes informàtics privats més tradicionals que impliquen llicències de pagament per a utilitzar-los.

Un dels programes que cada dia guanya més popularitat en la comunitat científica i docent és R, gràcies al seu potencial i polivalència. R permet utilitzar un llenguatge de programació per a fer l'estudi que l'usuari vulgui: des d'obtenir gràfics i resultats de procediments estadístics, fins a la creació de noves funcions destinades a resoldre situacions concretes i la modelització mitjançant nous algoritmes dissenyats *ad hoc* per a un problema determinat.

R no es caracteritza únicament per tenir llicència lliure, i ser molt potent i polivalent, sinó que a més és un entorn molt fàcil d'instal·lar i és multiplataforma, és a dir, ofereix diverses interfícies que es poden adaptar a les preferències de cada usuari.

Si bé és cert que inicialment R es va concebre per a cobrir les necessitats de l'estadística com a disciplina, també permet el càlcul matricial i tècniques procedents d'altres disciplines. Per tant, actualment R ofereix una versatilitat que permet utilitzar-lo en qualsevol disciplina que requereixi manejar dades per a obtenir resultats.

Finalment, és important destacar que R evoluciona constantment, ja que la globalització i el treball col·laboratiu de milers d'usuaris, juntament amb el fet que es tracta d'un programa lliure, permeten agilitar-ne al màxim les actualitzacions i minimitzar el temps de llançament de noves versions. Això garanteix que l'usuari de R estigui constantment al dia.

### La interfície R-Commander

Tot i que R representa un paquet estadístic molt complet i versàtil, l'usuari que en vulgui aprendre el llenguatge de programació haurà de fer front a una corba d'aprenentatge que pot arribar a ser complicada. Per això mateix, s'han desenvolupat diverses aplicacions basades en R que en simplifiquen l'ús enormement, ja que ofereixen un entorn gràfic basat en finestres i menús desplegable similar al de programes com l'SPSS o el Minitab.

Entre les diferents alternatives, en els mòduls següents s'ha optat per la interfície gràfica R-Commander. El motiu d'aquesta elecció és que R-Commander és la interfície més popular i utilitzada en la majoria d'universitats. Aquesta interfície està especialment indicada per als estudiants de graus no tècnics que cursen assignatures de contingut quantitatiu, ja que els permet implementar els coneixements teòrics que han adquirit sense haver d'aprendre el llenguatge de programació d'R. El gran avantatge d'R-Commander respecte a altres interfícies és que el codi subjacent apareix contínuament en pantalla. Per aquest motiu, R-Commander també es recomana als estudiants que volen aprendre el llenguatge d'R i agilitar el procés d'aprenentatge.

### Estructura d'aquest manual

Aquest manual consta de cinc mòduls amb un contingut específic que es detalla en la secció següent.

A continuació es mostra una descripció general de cada un:

- **Mòdul 1.** Conté la introducció al programa R, començant per la instal·lació i repassant els aspectes principals del seu llenguatge de programació. El mòdul finalitza amb una secció dedicada a la instal·lació i als components d'R-Commander.
- **Mòdul 2.** Cobreix els aspectes d'R relacionats amb l'àlgebra lineal i el càlcul. Això inclou operacions algebraiques bàsiques, transformacions matricials, diferenciació i integració, optimització, equacions diferencials i anàlisi de variable

complexa. Al llarg del mòdul, amb diversos exemples, s'ofereixen instruccions sobre la construcció de gràfics amb codi R.

- **Mòdul 3.** Representa una introducció a l'anàlisi estadística descriptiva i gràfica mitjançant exemples il·lustratius. La primera secció se centra en la interfície R-Commander, mentre que la resta del mòdul consta únicament d'instruccions amb codi d'R
- **Mòdul 4.** Se centra en l'estudi de les distribucions de probabilitat (discretes i contínues), i en la seva aplicació en la inferència estadística (interval de confiança i contrastos d'hipòtesis). Tot el mòdul s'elabora utilitzant R-Commander.
- **Mòdul 5.** Està dedicat a l'anàlisi multivariant (ANOVA, components principals i anàlisi clúster) i a la introducció dels models de regressió simple i múltiple. Igual que en el mòdul anterior, els exemples es treballen utilitzant R-Commander.

## Continguts

Mòdul didàctic 1

### **L'entorn estadístic R**

Daniel Liviano Solís

Maria Pujol Jover

1. Primers passos amb R
2. Sintaxi i programació
3. L'extensió R-Commander

Mòdul didàctic 2

### **Àlgebra lineal i càlcul amb R**

Daniel Liviano Solís

Maria Pujol Jover

1. Àlgebra vectorial i matricial
2. Funcions
3. Càlcul diferencial i integral
4. Optimització
5. Anàlisi de variable complexa

Mòdul didàctic 3

### **Anàlisi de dades i estadística descriptiva amb R i R-Commander**

Daniel Liviano Solís

Maria Pujol Jover

1. Estadística descriptiva amb R-Commander
2. Anàlisi del mercat de treball a Espanya
3. Anàlisi demogràfic a Catalunya

Mòdul didàctic 4

### **Distribucions de probabilitat i inferència estadística amb R-Commander**

Daniel Liviano Solís

Maria Pujol Jover

1. Distribucions de probabilitat
2. Inferència estadística

Mòdul didàctic 5

**Models de regressió i anàlisi multivariant amb R-Commander**

Daniel Liviano Solís

Maria Pujol Jover

1. Models de regressió
2. Anàlisi de la variància (ANOVA) i taules de contingència
3. Anàlisi de components principals i anàlisi clúster

## Bibliografia

**Gibernans Bàguena, J.; Gil Estallo, À. J.; Rovira Escofet, C.** (2009). *Estadística*.  
Barcelona: Material didàctic UOC.