

Matemáticas y estadística con R

Estadística, álgebra lineal y cálculo

Daniel Liviano Solís

Maria Pujol Jover

PID_00203024

**Daniel Liviano Solís**

Profesor de los Estudios de Economía y Empresa de la Universitat Oberta de Catalunya. Doctor en Economía por la Universitat Rovira i Virgili.

**Maria Pujol Jover**

Profesora de los Estudios de Economía y Empresa de la Universitat Oberta de Catalunya. Doctora en Estudios Empresariales por la Universitat de Barcelona.

Primera edición: septiembre 2013

© Daniel Liviano Solís, Maria Pujol Jover

Todos los derechos reservados

© de esta edición, FUOC, 2013

Av. Tibidabo, 39-43, 08035 Barcelona

Diseño: Manel Andreu

Realización editorial: Oberta UOC Publishing, SL

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Introducción

La importancia del análisis estadístico y cuantitativo

Tanto las matemáticas como la estadística son fundamentales en muchos planes de estudio: física, ingenierías, ciencias sociales, turismo, etc. Por eso mismo, es tan importante conocer la teoría subyacente a los métodos y técnicas incluidas en estas disciplinas como saberla aplicar en la vida real. Este manual pretende ser un complemento al material teórico de las asignaturas cuantitativas que se ofrecen en los distintos planes de estudio de la Universitat Oberta de Catalunya (UOC).

Es un hecho que hoy en día no se concibe el estudio de una asignatura cuantitativa (matemáticas, estadística, física,...) sin un soporte computacional que facilite los cálculos a realizar. De ahí la importancia que normalmente adquieren los programas informáticos en el momento en que los estudiantes optan por matricularse de este tipo de asignaturas. Por este motivo y porque realmente creemos que la mejor inmersión del estudiante en cualquier programa informático es a partir de su propia experiencia, la estructura que se sigue en este manual consiste en la resolución de problemas planteados con la ayuda de un programa específico.

El paquete estadístico R

En este sentido, recientemente ha habido un gran desarrollo de programas informáticos destinados a realizar análisis numéricos de datos empíricos. Los más interesantes son que se distribuyen bajo una licencia *GNU General Public License*, es decir, que no implican coste económico alguno y que son una clara alternativa a los programas informáticos privados más tradicionales que implican licencias de pago para su uso.

Uno de los programas que está ganando cada día mayor popularidad en la comunidad científica y docente es R, debido a su potencial y polivalencia. R permite el uso de un lenguaje de programación para realizar el estudio que el usuario desee: desde obtener gráficos y resultados de procedimientos estadísticos, hasta la creación de nuevas funciones destinadas a resolver situaciones concretas y la modelización mediante nuevos algoritmos diseñados *ad hoc* para un problema determinado.

R no se caracteriza únicamente por tener licencia libre, ser muy potente y polivalente. Además, se trata de un entorno de muy fácil instalación y es multiplataforma, es decir, que ofrece diversas interfaces que se pueden adaptar a las preferencias de cada usuario.

Si bien es cierto que R fue inicialmente concebido para cubrir las necesidades de la estadística como disciplina, también permite el cálculo matricial y técnicas procedentes de otras disciplinas. Por tanto, en la actualidad R ofrece una versatilidad que permite su uso en cualquier disciplina que requiera el manejo de datos para obtener resultados.

Finalmente, es importante destacar que R está en constante evolución, ya que la globalización y el trabajo colaborativo de miles de usuarios, junto con el hecho de tratarse de un programa libre, permiten agilizar al máximo sus actualizaciones y minimizar el tiempo de lanzamiento de nuevas versiones. Esto garantiza que el usuario de R esté constantemente al día.

La interfaz R-Commander

Aunque R representa un paquete estadístico muy completo y versátil, el usuario que desee aprender su lenguaje de programación deberá enfrentarse a una curva de aprendizaje que puede llegar a ser complicada. Por eso mismo, se han desarrollado diversas aplicaciones basadas en R que simplifican enormemente su uso, ya que ofrecen un entorno gráfico basado en ventanas y menús desplegables similar a programas como SPSS o Minitab.

De entre las diferentes alternativas, en los siguientes módulos se ha optado por la interfaz gráfica R-Commander. El motivo de esta elección es que R-Commander es la interfaz más popular y utilizada en la mayoría de universidades. Dicha interfaz está especialmente indicada para aquellos estudiantes de grados no técnicos que cursen asignaturas de contenido cuantitativo, ya que les permite implementar los conocimientos teóricos adquiridos sin tener que aprender el lenguaje de programación de R. La gran ventaja de R-Commander respecto a otras interfaces es que en todo momento aparece en pantalla el código subyacente. Por este hecho, R-Commander también se recomienda a aquellos estudiantes que deseen aprender el lenguaje de R y agilizar el proceso de aprendizaje.

Estructura de este manual

Este manual se compone de cinco módulos cuyo contenido específico se detalla en la siguiente sección.

Una descripción general de cada uno de ellos se muestra a continuación:

- **Módulo 1.** Contiene la introducción al programa R, empezando por la instalación y repasando los principales aspectos de su lenguaje de programación. El módulo finaliza con una sección dedicada a la instalación y componentes de R-Commander.
- **Módulo 2.** Cubre los aspectos de R relacionados al álgebra lineal y al cálculo. Esto incluye operaciones algebraicas básicas, transformaciones matriciales, diferencia-

ción e integración, optimización, ecuaciones diferenciales y análisis de variable compleja. A lo largo del módulo, a través de varios ejemplos, se ofrecen instrucciones sobre la construcción de gráficos con código R.

- **Módulo 3.** Representa una introducción al análisis estadístico descriptivo y gráfico mediante ejemplos ilustrativos. La primera sección se centra en la interfaz R-Commander, mientras que el resto del módulo se realiza únicamente mediante instrucciones con código de R.
- **Módulo 4.** Se centra en el estudio de las distribuciones de probabilidad (discretas y continuas), así como su aplicación en la inferencia estadística (intervalos de confianza y contrastes de hipótesis). Todo el módulo se realiza usando R-Commander.
- **Módulo 5.** Está dedicado al análisis multivariante (ANOVA, componentes principales y análisis clúster) y a la introducción de los modelos de regresión simple y múltiple. Igual que el módulo anterior, los ejemplos se trabajan utilizando R-Commander.

Contenidos

Módulo didáctico 1

El entorno estadístico R

Daniel Liviano Solís

Maria Pujol Jover

1. Primeros pasos con R
2. Sintaxis y programación
3. La extensión R-Commander

Módulo didáctico 2

Álgebra lineal y cálculo con R

Daniel Liviano Solís

Maria Pujol Jover

1. Álgebra vectorial y matricial
2. Funciones
3. Cálculo diferencial e integral
4. Optimización
5. Análisis de variable compleja

Módulo didáctico 3

Análisis de datos y estadística descriptiva con R y R-Commander

Daniel Liviano Solís

Maria Pujol Jover

1. Estadística descriptiva con R-Commander
2. Análisis del mercado de trabajo en España
3. Análisis demográfico en Cataluña

Módulo didáctico 4

Distribuciones de probabilidad e inferencia estadística con R-Commander

Daniel Liviano Solís

Maria Pujol Jover

1. Distribuciones de probabilidad
2. Inferencia estadística

Módulo didáctico 5

Modelos de regresión y análisis multivariante con R-Commander

Daniel Liviano Solís

Maria Pujol Jover

1. Modelos de regresión
2. Análisis de la varianza (ANOVA) y tablas de contingencia
3. Análisis de componentes principales y análisis clúster

Bibliografía

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