

Jesús BOUSO FREIJO
El paquete estadístico R. Cuadernos Metodológicos, 48
CIS, Madrid, 2013.

Pese a su importante contribución a la formación metodológica de los sociólogos hispanoparlantes, ninguno de los números de la colección *Cuadernos Metodológicos* del Centro de Investigaciones Sociológicas había llamado hasta ahora la atención de esta sección de reseñas. Prueba de su dilatada trayectoria son los casi cincuenta manuales editados desde el ya "prehistórico" primer cuaderno de Jacinto Rodríguez Osuna (1991) sobre métodos de muestreo. La variedad de temas y técnicas tratados y el sano pluralismo metodológico que promueve hacen de esta colección un recurso de inapreciable valor que ha ido consolidando un *corpus* de técnicas y métodos de investigación en castellano al que hemos tenido la fortuna de poder recurrir varias generaciones de sociólogos y sociólogas españoles y latinoamericanos.

Quizás esta ausencia se deba al carácter didáctico y aplicado de estos manuales, a su tono alejado de las extensas y profundas indagaciones que caracterizan otras publicaciones del CIS. Ciertamente, el trabajo de Bouso que voy a reseñar aquí no constituye un avance sustantivo para la sociología hispana ni aborda un tema de investigación inexplorado o de impacto. Pero tampoco pretende serlo: se trata, ni más ni menos, de un manual que persigue introducir y familiarizar a los sociólogos con R, el *software* estadístico

de código abierto y acceso gratuito con mayor desarrollo e implantación en los últimos años. Este programa ha despertado amplias expectativas por su potencia y versatilidad y el autor llega a plantear que puede convertirse en una "lengua franca" para el análisis estadístico (p.235). Sin embargo, dado que esta lengua franca es un lenguaje hecho por y para estadísticos, mis comentarios no se limitarán a consignar los méritos y debilidades del manual de Bouso, ya que también pretenden recoger una evaluación personal de las aptitudes de dicho programa para la labor cotidiana de los sociólogos y otros científicos sociales que desarrollan análisis de datos cuantitativos.

El libro de Bouso tiene, en este sentido, el mérito de ser el primer manual impreso en español que recoge una introducción a R específicamente orientada a las tareas de análisis que más frecuentemente realizamos los sociólogos: importar y exportar conjuntos de datos, etiquetar y depurar registros, estadística descriptiva, inferencia y estimación, muestreo y regresión. Al tratar estos aspectos el libro presta más atención a las variables no cuantitativas (nominales y ordinales) del que suele encontrarse en los, por otra parte, abundantes manuales introductorios y cursos gratuitos que podemos encontrar en Internet. La exposición de los contenidos está muy orientada a la

práctica, tratando los distintos procedimientos mediante ejemplos con datos descargables de la página web del CIS y dando por sentado que el lector tiene los conocimientos básicos de Estadística que atañen a cada procedimiento. Al final de cada capítulo se incluyen ejercicios para afianzar los conocimientos expuestos.

A mi juicio, un problema del libro es que pretende que el lector se sumerja rápidamente en el programa, asumiendo que aprenderá a nadar en dicho entorno a medida que avanza en los capítulos. La Introducción apenas abarca cinco páginas, tras las cuales se comienza instalando el programa y dando los primeros pasos (o brazadas, por seguir con el símil) con los vectores. Se echa en falta en la Introducción o el primer capítulo un apartado específicamente dedicado a la arquitectura general de R, que exponga con cierto detalle los interfaces, tipos de archivos y de objetos con que trabaja el programa, así como las peculiaridades de su funcionamiento respecto a los programas comerciales de uso habitual entre los sociólogos. En particular, falta una explicación más detallada de un rasgo clave de R, el hecho de que se trata de un lenguaje orientado a *objetos*. En ello reside el fundamento de la gran versatilidad de este programa: tanto los datos utilizados en los procesos de análisis como sus resultados se almacenan como objetos que pueden ser utilizados por sucesivos procesos. Esto permite que, una vez adquiridos los rudimentos de la gramática matricial del programa, podamos usar los resultados de, por ejemplo, una tabla de contingencia para desarrollar un gráfico de barras, un mapa temático o los factores de ponde-

ración de un conjunto de datos diferente. Este aspecto sólo es tratado de forma parcial en distintos momentos de los capítulos 1 y 4, en los que se presentan distintos tipos de objetos (vectores, matrices, conjuntos de datos, factores) y se asume que el lector irá adquiriendo con la práctica la forma de trabajar con ellos.

No obstante, esta versatilidad también tiene una vertiente negativa. Los usuarios habituados a los menús de los programas comerciales que integran todos los cálculos de un análisis determinado (p.e. componentes principales), encontrarán una dificultad añadida en R al tener que ejecutar varios procedimientos con sus correspondientes instrucciones (p.e. crear un objeto con el análisis, extraer un sumario de varianzas, crear otro objeto con las puntuaciones de los componentes, ordenar representaciones gráficas...) para obtener lo que en esos programas se obtiene con algunos menús. Ciertamente, existe la posibilidad de interactuar con el programa a través de un entorno visual de menús y ventanas muy similar al de los programas comerciales mediante el paquete R-Commander, al que el autor presta quizás demasiada atención mostrando reiteradamente en todos los epígrafes cómo ejecutar con este paquete los distintos procedimientos. Quizás una referencia más sucinta al modo de acceder a través de los menús de R-Commander a cada funcionalidad y a las opciones disponibles y no disponibles en ellos para cada comando habría dejado espacio para una explicación más sistemática de los argumentos y opciones disponibles en los distintos comandos. R-Commander puede ser una gran herramienta para aprender

código, ya que todos sus procedimientos son accesibles a través de la consola, pero reúne un conjunto muy limitado (aunque ampliable) de la creciente infinidad de funcionalidades y paquetes que va generando la comunidad de desarrolladores de R, de forma que acostumbrarse a usarlo puede suponer un obstáculo para familiarizarse con el código que, a poco que se profundice en el uso del programa, va a ser la única forma de trabajar con R. En este sentido, el autor apenas hace mención a los interfaces de usuario orientados a la programación de código (RStudio parece el más exitoso), que aportan toda una serie de funcionalidades para la escritura (texto predictivo, ayuda contextual, panel de objetos, gestión de *scripts*, etc.) que mejoran notablemente la experiencia con el programa.

Llegados a este punto, hay que decir que el término 'programa' se queda un poco corto para hablar de R. Se trata más bien de un entorno de programación sobre el que es posible ejecutar, al menos en potencia, cualquier tipo de análisis estadístico que haya sido previamente formalizado a través de paquetes (más de 5.000 en diciembre de 2013) o, llegado el caso, incluso desarrollar un nuevo paquete que cubra necesidades muy específicas o avanzadas. Bousa insiste en que la principal ventaja de R es que es un software gratuito, pero hay que hacer una precisión importante: instalar R no cuesta dinero, pero sacarle partido cuesta bastante trabajo. Por tanto, cabe preguntarse si realmente R aporta un valor añadido a los sociólogos que compense el esfuerzo que requiere el aprendizaje de sus rudimentos y el dominio de los paquetes básicos, más

aún cuando algunos programas comerciales han simplificado de forma importante la realización de análisis estadísticos. Considero que la respuesta dependerá mucho del uso que haga cada cual de la estadística en el ejercicio de la profesión. Para ilustrar esto, pondré tres casos ideales en los que cada cual podrá verse mejor o peor reflejado.

En primer lugar, cabe plantear un perfil profesional en el que el análisis estadístico juega un papel secundario (complementando a técnicas cualitativas) o se centra en el análisis descriptivo o exploratorio de datos de encuesta (estadísticos descriptivos, tablas de frecuencias, tablas de contingencia, representación gráfica) con un recurso esporádico a la estadística inferencial (contrastes de hipótesis) o multivariable (regresión logística, análisis factorial, conglomerados). No creo equivocarme al señalar que este puede ser el perfil mayoritario en la sociología española. En mi opinión R tiene poco que ofrecer a este tipo de usuarios. Ciertamente es un programa gratuito que permite, gracias a R-Commander, acceder a las funciones más usuales de transformación (importar/exportar, recodificar, calcular, filtrar) y análisis de datos con un mínimo de código. Sin embargo, ya he dicho que esta forma de operar tiene muchas limitaciones ya que, por ejemplo, cosas tan sencillas en otros programas como ponderar una muestra son imposibles sin utilizar el código. A esto se suma que las tablas de resultados de los análisis se obtienen como texto plano, de forma que llevarlas a un procesador de textos puede resultar engorroso. Para estos usuarios, R aún ofrecería ventajas en cuanto a la

boración de gráficos, un aspecto en el que la versatilidad y diversidad del programa permiten que supere a los programas comerciales. Pero, de nuevo, esta potencia solo está a nuestro alcance mediante código, algo que para este perfil de usuarios puede suponer un obstáculo insalvable. Bouso plantea en su capítulo final algunos contraargumentos a la dificultad que supone trabajar con código (p.238): que no es un lenguaje complejo una vez superadas las primeras etapas y que las funcionalidades de R-Commander pueden ampliarse mediante *plug-in* que permiten acceder a múltiples paquetes a través de su interfaz de usuario. Sin embargo, pocos considerarán que este esfuerzo valga la pena mientras los programas comerciales continúen siendo accesibles a través de licencias públicas (como las de las universidades) o copias piratas.

En segundo lugar, es posible pensar en un perfil mucho más centrado en la sociología cuantitativa y condicionado por la creciente sofisticación matemática de las herramientas de análisis en este campo. Estoy pensando en el tipo de usuario que emplea técnicas de modelización avanzadas, utiliza grandes conjuntos de datos o realiza análisis específicos que los paquetes básicos de los programas de análisis estadístico convencionales no incluyen. Para estos especialistas, R es una solución que probablemente les permita integrar en un único entorno todas esas herramientas que actualmente utilizan en aplicaciones *shareware* o comerciales separadas, facilitando enormemente la obtención de gráficos de gran calidad y permitiéndoles un mayor control sobre los distintos procesos de cálculo. Además,

el bagaje matemático-formal que les se presupone indudablemente facilitará el aprendizaje inicial de los rudimentos del código. Creo que todo aquel que se sienta mínimamente identificado con este perfil y aún no use R debe empezar a plantearse que, tarde o temprano, va a verse prácticamente obligado a usarlo. Sin embargo, quizás sea este usuario el que quede menos satisfecho con el manual de Bouso y que, por ejemplo, agradecería un “mapa” más detallado de las funcionalidades de los paquetes básicos y de las que aportan otros paquetes relevantes, algo así como una guía que muestre en qué paquetes es posible encontrar las principales técnicas de análisis. En este sentido, por ejemplo, Bouso no menciona que R también dispone de paquetes muy interesantes para realizar análisis de contenido, análisis de redes sociales o análisis espaciales.

Por último, quiero plantear un tercer perfil: el del estudiante de sociología que, al iniciarse en el análisis estadístico, no cuenta con un bagaje previo con el que comparar la experiencia de usar R pero, se supone, tiene interés en aprender a manejar un paquete de análisis estadístico que le capacite para acceder al mercado laboral de nuestra profesión. Si se le preguntará, este hipotético alumno respondería que prefiere un programa que sea fácil a uno que sea difícil, pero esa misma lógica le llevaría a preferir no estudiar estadística inferencial ni modelos de regresión. Es por ello que el criterio sobre los contenidos formativos se deja a los profesores y es a ellos a quien se dirige este párrafo. Independientemente de con cuál de los dos perfiles anteriores se identifiquen, espero que lo dicho hasta ahora

haya servido para que estén preguntándose si sus alumnos deberían aprender R. Considero que la respuesta debe ser afirmativa, aunque otra cosa es que sea factible. Con respecto a lo primero caben varios argumentos: el ahorro de recursos que supondría para las universidades en estos tiempos de recortes y escasez sería el más evidente, pero creo que es más convincente plantear que la educación superior debe capacitar al alumnado para seleccionar y renovar sus herramientas analíticas y no para ser clientes de una empresa que decida por ellos qué herramientas son accesibles o no. Con respecto a lo segundo, como he dicho, la gratuidad de R es solo aparente: introducir R en el currículo de los estudiantes de sociología requiere un esfuerzo colectivo de formación y reciclaje profesional para el que la Universidad española no parece contar con recursos ni motivación suficiente, por lo que su implementación dependerá de la iniciativa de los equipos docentes de cada facultad, aunque aquí hay que resaltar la gran cantidad de recursos online disponibles sin coste económico. Para este fin el manual de Bouso puede ser de gran ayuda ya que, deliberadamente o no, trata

la gran mayoría de los contenidos prácticos de las asignaturas de estadística de los grados de sociología, aportando ejemplos y ejercicios que pueden servir de inspiración para la práctica docente.

En definitiva, nos encontramos ante un manual que, aunque a algunos parecerá excesivamente simple y orientado a la práctica, puede servir para dar nuestro primeros pasos con este programa o, de cara a la docencia, como material de apoyo para introducir R en los contenidos de prácticas de estadística u otras asignaturas en las que el análisis estadístico tenga cierto protagonismo. Sin embargo, quienes comiencen a trabajar con R pronto se darán cuenta de que el libro apenas cubre el principio del recorrido, pero que la mayor ventaja de este programa —y la mejor forma de profundizar su aprendizaje— es su comunidad de usuarios y la inmensa cantidad de recursos en línea fácilmente accesibles gracias a la filosofía de acceso abierto que inspira todo el proyecto R.

PABLO RODRÍGUEZ-GONZÁLEZ
Universidad de La Laguna
prodriigg@ull.edu.es