

Elosua Oliden, Paula y Juan Etxeberria Murgiondo (2012). *R Commander. Gestión y Análisis de Datos*. Madrid: La Muralla.

Bouso Freijo, Jesús (2013). *El paquete estadístico R*. Madrid: CIS.

Rodrigues-Silveira, Rodrigo (2013). *Representación espacial y mapas*. Madrid: CIS.

Lejos están los tiempos en que para realizar complejas operaciones estadísticas había que recurrir a una calculadora, cuando no a realizar manualmente los múltiples recuentos y cálculos que se requieren para obtener medias, desviaciones típicas o correlaciones, por no mencionar la ejecución de complejas regresiones logísticas, sofisticados modelos de ecuaciones estructurales o, incluso, simples tabulaciones de las respuestas en las encuestas. Tanto los profesionales de las encuestas, como los científicos de la estadística, como los estudiantes de múltiples disciplinas donde las técnicas de análisis de datos son esenciales encontraron en los paquetes estadísticos una inevitable herramienta para el trabajo. En concreto, en los años 70 se habían asentado tres grandes sistemas estadísticos con los que trabajaban los científicos sociales y de la salud: SAS, BMDP y SPSS. Los años 80 supusieron, como todos sabemos, una gran revolución en la informática con el nacimiento de los ordenadores personales. Con estas nuevas máquinas fuimos capaces de realizar complejos cálculos en nuestros despachos del lugar de trabajo, primero; en los de nuestras casas, a continuación, y en cualquier lugar, finalmente. Por ello, los originales sistemas estadísticos tuvieron que adaptarse a las nuevas máquinas, al tiempo que surgían otros nuevos conjuntos de programas estadísticos comerciales de alcance amplio como Stata, StatGraphics, Systat, SPAD, o de alcance más restringido como HLM, LIMDEP, UCINET, etc.

Por otro lado, los lenguajes con los que nos enfrentamos a las máquinas también sufrieron una gran transformación. Desde los clásicos Basic, Cobol y Fortran, la evolución llevó a herramientas más evolucionadas como Visual Basic, C++, Pascal, Java, ... Asimismo, surgieron potentes lenguajes de programación especializados en cálculos matriciales. Sintomáticamente, en el mismo año 1984 nacen Matlab y Gauss. Sin embargo, no será hasta el año 2000 en el que nace la versión 1.0 de R. Pero ¿qué es R? Como se define sus impulsores y divulgadores (<http://www.r-project.org>) R es un “lenguaje y un entorno para el cálculo estadístico y para los gráficos”. Se genera además como un proyecto GNU, basado en la filosofía del software libre, cuya idea básica es que el programa está abierto a cualquier modificación del usuario y puede distribuirse gratuitamente, sin restricción comercial alguna. Trece años más tarde, en la actualidad, ya ha visto una versión 3.0.2 en la se han incorporado más de 5.000 diferentes paquetes accesibles gratuitamente desde su propia página (<http://cran.r-project.org/web/packages/>). Ahora bien, como no es fácil el uso de esta herramienta, también se han creado otras herramientas de código abierto y gratuitas que nos permiten manejarlo con mayor facilidad. Entre ellas, destacan R commander y R-Studio. La primera de ellas permite trabajar con R a través de menús, mientras que la segunda nos permite un uso más amigable de R, al ser un entorno de desarrollo interactivo, que permite a los programadores crear sus programas con mucha más facilidad, que si tuviera que hacerlo empleando directamente el lenguaje de programación con sus propios recursos memorísticos.

Siendo este el panorama, no es extraño que ya estén apareciendo en el mercado bibliográfico español textos que nos ayuden a aprender esta herramienta tan potente y desarrollada que, al mismo tiempo, es gratuita. En consecuencia, vamos a comentar tres libros que han salido al mercado editorial¹ en los últimos dos años en dos de las colecciones más prestigiosas de títulos de carácter instrumental y técnico para la investigación científica: la de Cuadernos de Estadística, con 40 títulos en su haber y la de Cuadernos Metodológicos, que recientemente alcanzó su quincuagésimo número.

El libro publicado por la editorial La Muralla, escrito por una profesora del área de Psicometría y un profesor del área de Métodos de Investigación y Diagnóstico en Educación, tiene como objetivo presentar a usuarios con escasos conocimientos de Estadística a elaborar e interpretar la estadística básica, haciendo uso de R commander. Para ello emplea algo menos de 200 páginas repletas de recortes de ventanas de menús o de resultados, así como una quincena de fórmulas de complejidad media del análisis estadístico.

El índice de *R Commander. Gestión y Análisis de Datos* se presenta dividido en dos partes: La primera, a su vez, consta de dos grandes apartados. El primero, compuesto en total por cinco capítulos, está dedicado a R y enseña a instalar este entorno estadístico, nos presenta algunos paquetes básicos del sistema y nos instruye acerca de cómo instalarlos, y termina centrándose en los primeros pasos para instalar y configurar el paquete R commander. El segundo apartado versa sobre la preparación de las bases de datos con R commander. Primero, se explica la introducción de datos desde los formatos más habituales, esto es, SPSS, ASCII, Excel,... Después, se aborda la modificación del fichero, centrándose solo en el filtrado y borrado de casos. A continuación, se dice cómo crear o modificar variables, incluyendo las transformaciones más importantes: tipificación, segmentación y conversión de variables numéricas en factores. Esta primera parte finaliza con dos extensiones útiles de R commander. Por un lado, Export, para transformar los resultados en ficheros LaTeX o HTML; por el otro, TeachingDemos, que contiene herramientas útiles para la enseñanza y el aprendizaje.

La segunda parte del número 40 de Cuadernos de Estadística muestra cómo obtener los más usuales procedimientos estadísticos que se enseñan en las aulas universitarias de ciencias sociales y biomédicas. También en esta segunda parte, se pueden distinguir dos partes diferenciadas. En la primera, se empieza con las tablas de frecuencias y los estadísticos básicos, desde la media a la correlación; posteriormente se consideran los test estadísticos y se finaliza con análisis multivariantes, incluyendo someramente análisis de escalas, factoriales, conglomerados, análisis de varianza y regresiones múltiples. En la segunda, se enseña a trabajar con gráficos, con distribuciones, así como distintos ejercicios para

¹ En inglés podemos reseñar que las principales editoriales como Springer, Sage y Wiley ya han sacado al mercado importantes títulos didácticos de R. Como ejemplo, citamos los textos de Crawley (2005), Field, Miles y Field (2012) y Shumacker, Randall y Tomek (2013), sin olvidarnos de otros ya clásicos como Ihaka, R., and Gentleman, R. (1996); Fox, J. (2002), y Venables, Smith y the R Development Core Team (2009).

demostrar en la práctica conceptos tan complejos como los intervalos de confianza, la potencia de las pruebas, la correlación o el límite central.

Se trata, pues, de un libro que pretende abarcar todos los conocimientos estadísticos de un estudiante de grado en carreras humanísticas o biomédicas. Obviamente, no se profundiza mucho en los temas, pero tiene el mérito de hacer una presentación muy didáctica de uno de las herramientas más temidas no sólo entre las alumnas y alumnos de letras, sino también entre sus profesores. Empezar desde el principio, antes de la instalación; emplear exclusivamente el paquete R commander, reduciendo todo el uso de R a la existencia de cuadros de diálogo de esta herramienta y centrándose en la versión en castellano de esta herramienta puede ser un comienzo excelente para poder sustituir los programas comerciales de estadística con precios exorbitados por esta prestigiosa herramienta de coste gratuito.

El segundo libro de esta reseña es el que ha aparecido en la colección de Cuadernos Metodológicos con el título de *El paquete estadístico R*. El mencionado volumen está centrado principalmente en el análisis de datos de encuestas y también tiene como cometido introducir a un inexperto usuario en las artes de la programación estadística. Ahora bien, su modo de proceder es distinto al de la obra anterior, pues en lugar de centrarse en R commander lo hace en el lenguaje y sintaxis propios de R, si bien en cada capítulo comunica al lector cómo podría hacer fácilmente con menús del mencionado paquete todo aquello que explicó previamente.

El contenido de este segundo texto es muy similar al del primero, aunque más reducido en operaciones estadísticas. Sin embargo, profundiza en el lenguaje de R, en lugar de centrarse exclusivamente en los menús de este. Por ello, el número de capturas de pantalla es muchísimo menor que en el libro anterior; pero a cambio, ofrece múltiples ilustraciones de la consola de R, así como de las órdenes propias de R. A pesar de que incluye muchas de ellas, el libro no es un manual típico de R, que va de orden en orden explicando sus múltiples opciones, sino que las va haciendo aparecer a medida que se van necesitando para cumplir las funciones que el analista de encuesta necesita para cumplir su cometido de tratar las respuestas de los entrevistados.

La primera parte de este Cuaderno Metodológico trata de la preparación de la información. El primer capítulo enseña cómo instalar R y como llevar a cabo una primera sesión con este entorno de programación. A tal efecto, nos introduce en los vectores, las matrices, los operadores elementales, las hojas de datos (*data frame*) y los tipos de objetos que en ella se pueden incluir (lógicos, factores, numéricos, y textuales). El capítulo acaba con la presentación de R commander y con las instrucciones para su instalación. Es esta una constante en este libro, pues casi todos sus capítulos terminan con el mismo contenido de su comienzo, empleando este paquete que, como hemos visto en la anterior obra, facilita en gran medida el empleo de R. Los otros dos capítulos de esta primera parte versan sobre la construcción del conjunto de datos de trabajo a partir de fichero ASCII, Excel o SPSS, y sobre cómo modificar esa base de datos, creando variables, modificándolas, depurando datos y unificando bases de origen distinto.

La segunda parte se centra en el análisis y se divide en cinco capítulos. El primero se dedica a la estadística básica. Sin embargo, comienza con dos operaciones

elementales de R: la conversión de variables numéricas a factores y el empleo de bucles para realizar operaciones repetitivas de modo simple. A continuación, pues, relata cómo obtener medidas de tendencia central, posición, dispersión y forma. Seguidamente, se detiene en las distribuciones de frecuencia y continúa con las tablas de contingencia. Lo que no se entiende muy bien es por qué ha optado por realizarlas con variables numéricas, en lugar de con factores, ya que tal como las explica, se pierden las etiquetas. El capítulo 5, segundo de la segunda parte, está dedicado a las pruebas de hipótesis abarcando tanto contrastes sobre medias dependientes e independientes, como proporciones y varianzas. Asimismo, contempla pruebas paramétricas y no paramétricas, como Wilcoxon y Kruskal-Wallis. El siguiente capítulo, el sexto, se concentra sobre los gráficos. De todos son sabidas las grandes posibilidades gráficas de R. Sin embargo, como es propio de un libro introductorio, Bouso, su autor, se centra en los histogramas, los gráficos de barras y de sectores, los de caja, para terminar con la representación gráfica de serie temporales, puesto que las nubes de puntos o diagramas de dispersión son tratados con criterio en el capítulo de la regresión.

Posiblemente, el capítulo de esta obra que menos satisfacción genere sea el de las técnicas de muestreo. Básicamente, porque el analista de encuesta se enfrenta más a menudo con la tarea de tener que ponderar su muestra, que con la de realizar una muestra u otorgar los pesos. Sin embargo, el autor dedica 10 páginas a estas últimas y tan solo media a la ponderación. Tan poca es la importancia que le otorga que solo habla de medias ponderadas, cuando lo importante en las encuestas es más bien el análisis de los porcentajes. Termina la obra con un capítulo que incluye tanto la regresión múltiple como la logística, lo que confirma que es un libro más bien pensado para enseñar R, sea con código, sea con los menús de R commander, a estudiantes o profesionales que están dando sus primeros pasos en estadística, más bien que a aquellos que ya conocen un programa estadístico y quieren pasarse a otro.

No quisiera terminar de hablar de estos dos libros sin recordar que R, aunque sea calificado como un paquete estadístico, es sobre todo un lenguaje y un entorno de programación, compuesto por muchos paquetes. A este respecto en las dos obras reseñadas hasta el momento, parecen que lo tratan como tal, como un sustituto de SPSS, SAS, Stata o cualquier otro paquete. Pero en realidad R es otra cosa y quizá el lector de ambos libros se queda con una idea equivocada de lo que pueda ser R, pues en ninguno de ellos se explica bien lo que es una función o un objeto, aspectos elementales del producto, si se quiere hacer un productivo uso de él.

Finalmente, se incluye en esta reseña el último número de la colección de Cuadernos Metodológicos, el 50, que versa sobre representación espacial a través de mapas y ha tenido como acierto, emplear como plataforma para el aprendizaje el lenguaje R. Por ello, se añade a esta revisión bibliográfica.

Representación Espacial y Mapas contiene cuatro apartados. El primero se constituye con los conceptos generales para la representación visual. En el segundo se habla de las fuentes y de la preparación de los datos. El tercero enseña cómo se confeccionan los mapas y cómo se representan en estos los datos. En el cuarto, se presentan distintos tipos de mapas temáticos, como la coropletas, el de densidad, los de símbolos, los isopléticos, los anamórficos y los de flujos. Finalmente, el cuaderno

expone las bases conceptuales y metodológicas útiles para análisis e interpretación de los fenómenos sociales con dimensión espacial.

Ahora bien, lo que destaca en el cuaderno es que todas las preparaciones de las bases de datos, los mapas y análisis que realiza están expuesto en código R. Y, para el lector que desconoce este entorno, se ofrece en el apéndice una guía de introducción a R y a otro de los auxiliares que permiten un entorno más amigable que el que proporciona este entorno de programación. Se llama R-Studio y es altamente recomendado para aquellas personas familiarizadas con Stata, pues le proporciona una interfaz muy similar para trabajar con los paquetes y funciones de R.

De las tres obras que aquí se han comentado, es esta la más especializada y, por ello, la que está menos al alcance de estudiantes de grado. Sólo se recomienda su empleo a quienes quieran hacer mapas o análisis espaciales y no tenga recursos para adquirir un software comercial de alto precio. Quien desee aprender a realizar estas tareas, lo que sí requerirá es tiempo para familiarizarse con unos nuevos conceptos; pero sobre todo para emplear correctamente el código necesario para realizar las operaciones deseadas. Quizá no hubiera sido mala idea, al hilo de esta publicación, haber creado nuevas funciones que hubieran facilitado la realización de mapas al usuario menos proclive a escribir instrucciones de sintaxis completa, repleta de opciones o argumentos.

En definitiva, parece que R se está asentando en la academia. Cada vez son más los artículos de revistas de prestigio que han sido confeccionados empleando esta herramienta. En el área de las ciencias sociales aún estamos algo rezagados. Pero la aparición de estos tres libros en el mercado editorial español no orientado a economistas, matemáticos o científicos de la salud, presagia que no tardaremos en transmitir su conocimiento y uso a las nuevas generaciones de científicos sociales.

Referencias

- Crawley, M. J. (2005). *An Introduction to R*. Chichester: Wiley
- Field, A.; Miles, J., and Field, Z. (2012). *Statistics using R*. London: Sage.
- Fox, J. (2002). *An R and S-Plus Companion to Applied Regression*. London: Sage.
- Ihaha, R., and Gentleman R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- Muechen, R. A., and Hilbe, J.M. (2010). *R for Stata Users*. New York: Springer.
- Muechen, R. A. (2011). *R for SAS and SPSS Users*. New York: Springer
- Schumacker, R., and Tomek, S. (2013). *Understanding Statistics Using R*. New York: Springer.
- Venables, W.N.; Smith, D.M., and the R Development Core Team (2009). *An Introduction to R*. Bristol: Network Theory.

Modesto Escobar Mercado
Universidad de Salamanca