



Lumley, T. (2010). *Complex Surveys. A Guide to Analysis Using R*. Hoboken, New Jersey: Wiley.

Este título es una de las últimas incorporaciones de la prestigiosa editorial Wiley a su serie sobre metodología de encuestas. En este caso se trata de un texto escrito por Thomas Lumley, quien actualmente ocupa una cátedra de Bioestadística en la Universidad de Auckland, en Nueva Zelanda, donde se trasladó recientemente tras desarrollar prácticamente toda su carrera profesional en la Universidad de Washington, en Seattle, a la que aún continúa vinculado como catedrático afiliado.

Lumley es conocido en el ámbito de la metodología de encuestas por haber diseñado el paquete *survey* para el entorno *R*. Dicho paquete proporciona herramientas para asistir al usuario en todas las etapas del desarrollo de una encuesta, desde el diseño al análisis final. En este texto, el autor cubre de nuevo todas las etapas de la encuesta, pero ahora desde un punto de vista metodológico aunque apoyándose fuertemente en el uso del software antes mencionado.

Se trata de un libro orientado a la práctica del análisis de encuestas, por lo que hace poco hincapié en los conceptos estadísticos subyacentes, que prácticamente supone conocidos por parte del lector. No obstante, el nivel es accesible a cualquier investigador del ámbito de las ciencias sociales con cierta experiencia en estadística aplicada. Queda patente la habilidad pedagógica del autor, resultado probablemente de su amplia experiencia docente y de su colaboración con profesionales de diferentes ámbitos, pero principalmente de la rama de ciencias de la salud.

Uno de los aspectos más interesantes de esta obra es el uso exclusivo del entorno *R*, quizás no tan habitual en este contexto como otros programas de tipo comercial. Precisamente, una de las ventajas de *R* es su libre disponibilidad y gratuidad de uso. Además, al tratarse de un entorno abierto, todas las novedades

metodológicas que se producen en el ámbito estadístico están, por lo general, disponibles inmediatamente en este software. Bien es cierto que la forma de trabajar es diferente a la habitual, pues la interacción con el sistema suele hacerse a través de un lenguaje algo más complicado que el usado por paquetes comerciales como SPSS. A cambio de esa complejidad, la flexibilidad y versatilidad son mucho mayores, lo que a la larga compensa con creces el esfuerzo inicial necesario para desenvolverse con soltura en ese entorno.

El libro comienza con un capítulo en el que se explican los conceptos básicos en los que se apoya el análisis de encuestas, y se da una breve introducción al lenguaje R. De manera acertada, el autor comienza explicando la captura de datos, ya sea en archivos de tipo texto o en formatos propios de paquetes comerciales habituales como SAS o SPSS. A lo largo del libro, van surgiendo de forma natural numerosas herramientas del software R. Mención especial hay que hacer a los gráficos, a los que se ha dedicado un capítulo completo, y que no en vano constituyen una de las principales fortalezas de este entorno.

Cada capítulo finaliza con un conjunto de ejercicios. Algunos de ellos se refieren al manejo de los conceptos estadísticos que aparecen en dicha unidad temática, pero principalmente se centran en la aplicación de las técnicas descritas, usando el paquete *survey* del entorno R. En este sentido, el libro es adecuado como referencia básica para un curso de postgrado sobre la práctica del análisis de encuestas.

De manera sistemática, se analizan todas las etapas del análisis y diseño de las encuestas, y se dan los ejemplos necesarios para ejecutarlas usando el paquete *survey*, abarcando temas desde el muestreo hasta la regresión, ya sea con variables numéricas o categóricas. Cabe destacar el tratamiento detallado de aspectos críticos como el manejo de datos faltantes, para lo que se usa el paquete *mitools*, del propio autor. Se dedica un capítulo completo a este tema fundamental.

Llama la atención la inclusión de un capítulo sobre inferencia causal, aspecto habitualmente ignorado en el análisis de las encuestas. Sin embargo, prácticamente se limita a estudiarlo desde el punto de vista de la regresión logística, citando solamente de pasada la herramienta fundamental en este ámbito, que son los llamados *modelos gráficos probabilísticos*. Sí que presta algo más de atención a este modelo en el capítulo sobre regresión con datos categóricos, aunque se echa de menos la descripción de alguno de los paquetes que ya hay disponible en R al respecto.

Es de agradecer la inclusión de cinco útiles apéndices. Los dos primeros de ellos proporcionan detalles técnicos desde un punto de vista estadístico. Los lectores con mayor inquietud en este sentido pueden profundizar aquí en los conceptos tratados de forma superficial durante el desarrollo de los capítulos. Los otros tres apéndices se dedican al entorno R. El primero de ellos consiste en una introducción básica al lenguaje de programación, con especial énfasis en la manipulación de datos y a la automatización de procesos sobre los mismos. De especial importancia es la descripción de las órdenes relativas a la repetición de tareas, ya que aunque en R es

posible usar bucles como en lenguajes de programación de propósito general, éstos pueden ser muy ineficientes en cuanto a tiempo de ejecución.

El siguiente apéndice se ocupa del tratamiento de grandes volúmenes de datos, que en el ámbito de las encuestas puede ser un hecho habitual. Precisamente, una de las críticas que algunos usuarios hacen al entorno R es que éste no es capaz de operar con archivos con una gran cantidad de registros, al no poder cargarlos en su totalidad en memoria. El autor explica de forma acertada cómo solventar este problema, mediante la funcionalidad de R que permite el acceso a bases de datos externas. Éste es un aspecto técnicamente complicado, pero necesario si se pretende realizar algún análisis sobre grandes archivos de datos de manera eficiente. El último apéndice, con el que se concluye el libro, está dedicado a sugerir posibles extensiones del paquete *survey* mediante la incorporación de diferentes modelos de regresión.

En resumen, se trata de un libro destacable por su aspecto práctico, donde todos los conceptos revisados se ponen en práctica usando el entorno R y, por tanto, se trata de una herramienta útil para el análisis de encuestas. No obstante, se requiere cierto conocimiento de estadística aplicada y de lenguaje R para sacar el máximo partido del mismo.

Antonio Salmerón Cerdán
Universidad de Almería