

*Los datos sin tapujos.
Cómo interpretar y difundir las estadísticas sociales*

Juan Ignacio Martínez Pastor

(Madrid, Los Libros de la Catarata, 2019)

La libertad de opinión de los ciudadanos es un rasgo que distingue a las sociedades democráticas. La pluralidad de pensamiento enriquece el debate y permite abordar problemas desde distintas perspectivas. La cuestión muchas veces no es tanto decidir qué medidas adoptar para afrontar un reto concreto, sino identificar con exactitud cómo es la realidad a la que nos enfrentamos. La preferencia por una serie de acciones u otras depende de las creencias que cada uno tenemos sobre lo que es más justo y sobre cómo debe funcionar una sociedad. Llegar a un acuerdo sobre esto último es en muchas ocasiones difícil, pero sin duda imposible si el diagnóstico de partida no es el mismo para todos. ¿A quién afecta más el paro? ¿En qué medida la delincuencia es más frecuente entre la población inmigrante? ¿Ha aumentado la desigualdad en la última década? ¿Quiénes han sido los más perjudicados por la crisis de 2008?

La finalidad de las ciencias sociales no es determinar qué políticas corresponden aplicar en cada situación. Su función es la de analizar con rigor científico la relación entre distintos fenómenos y, en el mejor de los casos, comprender las causas y las consecuencias que se derivan de ellos. En este sentido, el libro *Los datos sin tapujos. Cómo interpretar y difundir las estadísticas sociales* constituye una contribución de gran valor en unos tiempos en los que la proliferación de datos es inconmensurable y las posibilidades de información infinitas. El autor del libro, el profesor de Sociología de la UNED Juan Ignacio Martínez Pastor, comienza la obra reconociendo la inquietud que le provoca el modo en el que se difunden las estadísticas sociales. *Los datos sin tapujos* nace con la intención de proporcionar al lector las herramientas necesarias para sumergirse en el mundo de las cifras con el suficiente criterio para discernir la información veraz de la sesgada.

El autor de esta obra revisa algunos principios básicos de la estadística social. Si bien la naturaleza del libro podría resultar demasiado técnica y, por ello, reducir su potencial público al académico, una de las virtudes del texto es su capacidad de combinar explicaciones sencillas sobre conceptos estadísticos con ejemplos cercanos que ilustran cada idea. El carácter divulgativo del libro se plasma desde sus primeras páginas. La introducción arranca con un contundente titular de 2017 repetido con variaciones por distintos medios de comunicación: «El 96% de los nuevos desempleados que deja noviembre son mujeres». El impacto de esta noticia en las redes sociales, como cabría esperar, fue enorme, con declaraciones de líderes políticos y periodistas especializados denunciando la intolerable desigualdad laboral. En aquel momento, a Martínez Pastor la cifra le sorprendió lo suficiente como para consultar la fuente original y hacer sus propios cálculos, llegando a la conclusión de que el porcentaje real de nuevas paradas en aquel momento era del 51%.

La obra se compone de siete capítulos (el primero, introductorio) y unas breves conclusiones a modo de síntesis. Tras recordar que la información no implica necesariamente conocimiento y defender una actitud crítica ante los datos (a medio camino entre una actitud *naïf* y otra de desconfianza constante), el segundo capítulo se detiene en la importancia de las definiciones. Volviendo al principio, el diagnóstico de partida va a estar condicionado por cómo acotamos el fenómeno en cuestión. La clave está en que conceptos ampliamente utilizados, como los de desigualdad o pobreza, para analizarlos, necesitamos concretarlos en algo cuantificable. Conocer cómo se ha definido el objeto de estudio, por tanto, es decisivo a la hora de extraer conclusiones sobre el mismo. El autor ilustra la importancia de esta cuestión a partir de varios ejemplos. Uno de ellos es el que afectó a la definición del paro en 2002 como consecuencia de un nuevo reglamento de la Comisión Europea. A partir de ese momento, la condición de parado pasó a ser más restrictiva, requiriendo que el individuo buscara trabajo de forma activa y no simplemente con que estuviera inscrito en una oficina pública de empleo. El resultado de esta modificación supuso una reducción del paro en España de 634.000 personas. Un segundo ejemplo se centra en la confusión que habitualmente generan las cifras de paro juvenil en nuestro país. Por un lado, el autor advierte de que el rango de edad elegido para estudiarlo (16-24 años) está acotado arbitrariamente en su límite superior. Considerar joven a todos los menores de 30 años sería igualmente justificable (como ha hecho el Injuve), en cuyo caso el paro juvenil arrojaría cifras distintas. Por otro, abunda el desconocimiento sobre cómo se calcula la tasa de paro, en la que la población inactiva (como los estudiantes) no se tiene en cuenta. Si bien la tasa de paro entre los jóvenes españoles se aproxima al 50%, el porcentaje de jóvenes parados oscila en torno al 25%. Afirmar que la mitad de los jóvenes españoles se encuentra en el paro es un error mayúsculo repetido en medios de comunicación de máxima difusión.

En el tercer capítulo del libro se aborda la importancia, pero también las dificultades, que entraña una correcta comparación entre datos. Un número aislado puede ser poco informativo si no tenemos otras cifras que lo contextualicen. Por ejemplo, se sabe que en 2018 el coeficiente de Gini, que medía la distribución de ingresos, arrojaba una cifra del 0,33 para España. Cualquier persona que no esté familiarizada con este índice se puede preguntar si estamos ante un nivel alto o bajo de desigualdad. Siguiendo a Martínez Pastor, la mejor manera de interpretar ese dato es compararlo, siempre que se pueda, con otros retrospectivos o de países del entorno más próximo. Hacerlo nos dará una idea global del lugar en el que nos encontramos. En este sentido, las series históricas son de gran utilidad, pero hay que advertir que el periodo de tiempo seleccionado es crucial de cara a extraer conclusiones del momento actual. Un buen ejemplo de ello lo tenemos en la evolución de la ocupación en España. La destrucción de empleo desde la crisis de 2008 se llevó por delante cuatro millones de puestos de trabajo. Si bien la catástrofe es innegable, ampliar el foco temporal permite ver que la caída de la ocupación estuvo precedida por un periodo extraordinario de creación de empleo con ocho millones de puestos de trabajo en tan solo 13 años. Este capítulo finaliza con una reflexión sobre la visión que los científicos sociales tienen de los problemas sociales. Aunque la percepción tiende a ser pesimista, el autor del libro defiende que desde una perspectiva histórica las sociedades han progresado notablemente, algo que, por otra parte, no excluye una actitud crítica ante los problemas de nuestro tiempo.

El cuarto capítulo se centra en algunas técnicas básicas de medición. Comienza con la utilidad de un buen porcentaje y, en particular, con la forma de expresar un cambio de por-

centaje. Si la temporalidad en España pasa del 24% al 25% de un año a otro, podemos decir tanto que la tasa ha aumentado un punto porcentual como que ha crecido un 4%. Presentar el cambio en cualquiera de esas dos formas es correcto, pero para evitar equívocos el autor aconseja ofrecer siempre las cifras originales. En general, esta es una recomendación que se repite a lo largo del capítulo. La importancia de dar el dato base está en que, si la cifra de partida es muy grande, es muy difícil que se produzcan variaciones porcentuales abultadas, sin que ello signifique que el cambio no haya podido ser sustancial. El caso contrario ocurre cuando encontramos una gran variación porcentual, pero en realidad esconde un número muy bajo de casos y cualitativamente un cambio modesto. Para ilustrar lo anterior, Martínez Pastor recurre, entre otros ejemplos, a las estadísticas del fútbol. En la temporada 2011-12, el jugador del F. C. Barcelona, Thiago Alcántara, logró incrementar un 100% su marca goleadora con respecto a la temporada anterior. Cristiano Ronaldo, en aquel entonces delantero del Real Madrid, solo consiguió mejorarla un 15%. La conclusión que saquemos de estos números será errónea si no se aclara que Thiago pasó de marcar dos goles en la temporada 2010-11 a cuatro en la siguiente, mientras que el fenómeno portugués partía de 40 tantos y un año más tarde alcanzó la cifra de 46. El capítulo continúa revisando otras técnicas estadísticas básicas. Las tablas de contingencia nos ayudan a conocer la relación entre dos variables. Si lo que queremos saber es cómo la variable independiente se correlaciona con la dependiente, los porcentajes tienen que ser calculados sobre la primera de ellas y no sobre la segunda. Igualmente, la media y la mediana son medidas de tendencia central comúnmente utilizadas. Su capacidad informativa sobre una distribución es valiosa, pero hay que tener cuidado con los valores extremos, los cuales pueden distorsionar la imagen global del conjunto. Para tener en cuenta la dispersión de una muestra es necesario apoyarse en otras medidas como la desviación típica. Para finalizar, el autor introduce brevemente el uso de algunas medidas relativas como los *odds*.

El libro avanza con un quinto capítulo dedicado a la correlación y la causalidad. En primer lugar, comienza aclarando que si bien para que exista causalidad entre dos fenómenos es imprescindible que haya correlación, esta última no conduce automáticamente a una relación de causa-efecto. Aquí nos adentramos en el delicado terreno de las relaciones espurias entre variables. Un hecho social puede parecer ser la causa de otro, pero en verdad estar provocado por un tercer fenómeno que no se ha tenido en cuenta. En las ciencias sociales las relaciones causa-efecto son complicadas de detectar en parte debido a la dificultad de diseñar experimentos semejantes a los realizados en las ciencias naturales. Una estrategia para lograrlo consiste en el uso de técnicas estadísticas, como el análisis de regresión, que permiten aislar el efecto neto de distintas variables independientes sobre otra dependiente. De este modo es posible conocer en qué medida un fenómeno influye sobre otro manteniendo todo lo demás constante. Martínez Pastor reconoce que estas técnicas no están exentas de problemas a la hora de identificar verdaderas relaciones de causalidad. En primer lugar, porque ninguna base de datos contiene todos los factores que potencialmente pueden influir sobre un hecho en particular. En segundo lugar, porque encontrar una correlación no nos dice nada sobre los mecanismos que realmente están funcionando para que una variable (por ejemplo, tener padres con estudios universitarios) influya positivamente en otra (por ejemplo, tener ingresos altos). Por último, porque la correlación entre dos fenómenos puede esconder problemas de endogeneidad, es decir, que el efecto de una variable sobre otra se produzca también a la inversa. En estos casos es difícil desentrañar qué funciona como causa y qué como efecto. El capítulo presenta distintas soluciones metodológicas para gestionar los anteriores problemas. Entre ellas, el autor

destaca el valor de los datos longitudinales, es decir, de bases de datos con información de un mismo individuo a lo largo del tiempo.

En el sexto capítulo del libro se abordan algunos debates académicos ligados a la estadística social. Muchas de las encuestas que se utilizan en las ciencias sociales se realizan a una muestra representativa de la población de estudio. Cuando los resultados que se obtienen a partir de esa muestra son estadísticamente significativos asumimos que se pueden inferir al conjunto de la población. A este respecto, Martínez Pastor recomienda cierta cautela, pues una relación no significativa estadísticamente podría serlo sustantivamente. Asimismo, confirmar la hipótesis nula —es decir, encontrar que una variable no influye sobre otra— ha de considerarse un hallazgo valioso, aunque la investigación social tienda a buscar justamente relaciones entre fenómenos. El autor advierte de otras cuestiones metodológicas a tener en cuenta, como la necesidad de que las variables varíen si queremos conocer el efecto de cada una de ellas; los casos de multicolinealidad, que tienen lugar cuando dos variables independientes cambian al mismo tiempo —algo que impide diferenciar el efecto neto de cada una de ellas sobre la variable dependiente—; o el peligro de la falacia ecológica, que se produce cuando asumimos erróneamente comportamientos individuales a partir de la correlación de datos agregados. El capítulo finaliza revisando las dificultades de predicción del futuro mediante proyecciones a partir de la evidencia empírica existente, así como la importancia de diferenciar datos de stocks y de flujos si no queremos caer en errores groseros como el del 96% descrito en el capítulo introductorio.

El libro cierra con un séptimo capítulo dedicado a la fiabilidad de las encuestas. ¿De verdad es posible saber lo que opinan millones de personas entrevistando únicamente a unas pocas? En este capítulo Martínez Pastor repasa las principales premisas estadísticas que deben cumplirse para que la pregunta anterior se responda de manera afirmativa. Para ello se centra en dos aspectos clave, en que la muestra debe ser representativa del conjunto de la población de estudio y en que los márgenes de error nos permiten conocer de manera aproximada cómo es la población. Como en el resto de la obra, ambas ideas son explicadas con claridad y apoyadas por ejemplos que contribuyen a una mejor comprensión del asunto. *Los datos sin tapujos* es un libro para un público amplio, desde profesionales de la comunicación que están acostumbrados a difundir información estadística hasta académicos en distintas etapas de su carrera. También para todas aquellas personas que tienen interés en los cambios que acontecen en nuestra sociedad. La imposibilidad de digerir toda la información que tenemos a nuestro alcance nos obliga a ser selectivos y, como el propio Martínez Pastor dice, a separar con ojo crítico la paja del trigo.

por Jacobo MUÑOZ-COMET

Universidad Nacional de Educación a Distancia (UNED)

jmcomet@poli.uned.es