

La necesaria integridad estadística en la era de los grandes datos

Elena Herrero-Beaumont

La estadística es consustancial al buen gobierno. En la Constitución de los EE. UU. se prescribió la necesidad de hacer un censo cada década para que el legislador “pueda descansar sus argumentos en hechos”, en palabras de James Madison (*The Economist*, 2020). La estadística tiene una importancia sobresaliente en democracia, porque minimiza escenarios de incertidumbre, ofrece un marco para comprender el mundo y posibilita a los gobiernos tomar decisiones informadas. Como aseveran los autores del último artículo del presente número, la misión de la estadística es proporcionar a la sociedad “conocimiento sobre sí misma”.

Esta reflexión nos lleva a una idea central acerca del valor público del dato necesario para elaborar las estadísticas oficiales. La reutilización de los datos generados por los ciudadanos para la producción de mejores estadísticas es una manera de devolver a la ciudadanía ese valor público del dato. Esta idea central debería guiar a los gobiernos y a las compañías privadas en el desarrollo de las nuevas metodologías y algoritmos a partir de los grandes datos disponibles en las nuevas infraestructuras digitales.

El Covid-19 ha sido un gran acelerador de tendencias preexistentes, entre ellas, el desarrollo de estadísticas elaboradas por los gobiernos a partir de los grandes datos (o *Big Data*). Estas nuevas fuentes de datos, así como las nuevas metodologías y algoritmos necesarios para su procesamiento, plantean enormes desafíos y oportunidades que los expertos entrevistados y los autores de los sucesivos artículos que componen este nuevo número de Índice señalan.

Uno de los mayores desafíos que se presentan es la llamada vigilancia de alta definición. En la actual sociedad del dato, se agudizan los riesgos inherentes al tantas veces descrito

Gran Hermano Orwelliano en lo que respecta a las libertades individuales. El seguimiento virtual del ciudadano en todos y cada uno de sus comportamientos online a la hora de conversar con familiares y amigos, buscar información y comprar bienes y servicios plantea numerosas cuestiones de índole moral.

La información está siendo ya aprovechada por las compañías para identificar lo que sus dirigentes pueden llegar a definir como “comportamientos sospechosos” de sus trabajadores, o, en mayor medida, para aumentar la productividad (Sun, 2021). También hemos visto cómo Facebook vendió los datos asociados a 50 millones de usuarios de su plataforma a *Cambridge Analytica*, una firma de comunicación política contratada por el equipo de campaña de Donald Trump para desarrollar las estrategias electorales en las elecciones de 2016 (Granville, 2018).

Estos ejemplos materializan los riesgos que más preocupan a los expertos, que alertan de que las nuevas bases de datos utilizadas para alimentar los nuevos sistemas de inteligencia artificial (algoritmos y metodologías de procesamiento de datos) pueden derivar en una mayor desigualdad social y en una radicalización de los prejuicios raciales y de género (FT Editorial Board, 2020). Ya Darrell Huff escribió en 1954 un libro titulado “Cómo mentir con Estadísticas”, donde hablaba de cómo las estadísticas pueden convertirse en herramientas poderosas de publicistas y políticos (Westlake, 2020).

En un contexto de desconfianza del público hacia las instituciones se hace urgente trabajar en la integridad de las estadísticas oficiales, que han de verse más bien como “una fuerza para el bien” (Westlake, 2020). El economista Tim Harford ofrece algunos principios para caminar hacia esa integridad, de los cuales el más evidente es la transparencia del algoritmo.

El presente número de Índice comienza con una entrevista a Rafael Fernández Campos, el

Presidente del Club de Chief Data Officer, un cargo emergente en las organizaciones que se responsabiliza de la política y estrategia de datos. Para evitar los riesgos que hemos señalado más arriba, Fernández Campos ve muy necesario que los gobiernos establezcan mecanismos de control del uso masivo de datos, que las empresas se guíen por principios éticos y que la sociedad se haga responsable de su propia formación.

Carlos Ballano, Subdirector del Departamento de Control de la Producción Estadística y muestreo del INE, explica los aspectos legales, tecnológicos, metodológicos y de seguridad asociados a las técnicas estadísticas para la explotación de las nuevas fuentes de datos.

Antonio Argüeso, Subdirector General de Estadísticas Sociodemográficas del INE, analiza nuevas formas de afrontar las estadísticas en el ámbito sociodemográfico. Para ello utiliza el caso de los estudios de movilidad a partir de la telefonía móvil, donde se destaca los nuevos paradigmas de colaboración entre las oficinas estadísticas y los propietarios de datos.

Ana Carmen Saura, Subdirectora adjunta del Gabinete de Presidencia del INE, subraya como principal conclusión de un curso celebrado por el INE y la UIMP la importancia de disponer datos y estadísticas actualizados para la toma de decisiones. Fernando Bolívar y Sirenia Vázquez comparten la experiencia del BBVA Research en la utilización de datos para la producción estadística. Yolanda Gómez, del Gabinete de Presidencia del INE, subraya como principal objetivo de los servicios estadísticos lograr la colaboración con el sector privado. Y Montserrat Guillén Estany, profesora de la Universidad de Barcelona, detalla

cómo los datos pueden promover la seguridad en la conducción y prevenir los accidentes antes de que sucedan.

En un contexto de desconfianza del público hacia las instituciones se hace urgente trabajar en la integridad de las estadísticas oficiales, que han de verse más bien como “una fuerza para el bien”

Por último, Emanuele Baldacci, Director de Recursos, Fabio Ricciato, Oficina Estadística, y Albrecht Wirthmann, Jefe de la Unidad de Metodología, los tres del Eurostat, ahondan en los costes y beneficios derivados de la reutilización de las nuevas fuentes de datos para fines estadísticos. Para ello se hace necesario avanzar en tres aspectos: en un marco regulatorio como “condición sine qua non” para la futura producción estadística; en el co-desarrollo de las metodologías por parte de equipos multidisciplinares de expertos; la incorporación de la dimensión europea en los diferentes avances.

Todas las perspectivas contenidas en este número son una muestra de cómo los expertos avanzan en la generación de mejores estadísticas para lograr mejores políticas y una mejor sociedad informada. Esperamos así que en la era del Big Data la ciudadanía pueda, como diría Harford, “deleitarse con la utilidad que nos brinda la buena estadística” (Westlake, 2020). ●

Bibliografía

- Sun, Nikki (2021) “China’s Tech Workers Pushed to their Limits by Surveillance software”, The Financial Times.
- The Editorial Board (2020) “Blame not the Robot, but the Human Behind it”, The Financial Times.
- The Economist (2020) “America’s Census Looks Out of Date in the Age of Big Data”.
- Westlake, Stian (2020) “Tim Harford’s How to Make the World Add Up — a welcome guide to statistics”, The Financial Times.
- Granville, Kevin (2018) “Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens” The New York Times.
- Huff, Darrell (1954) How to Lie with Statistics W. W. Norton & Company.