



Índice

REVISTA DE ESTADÍSTICA Y SOCIEDAD

OCTUBRE 2021

83

la sociedad digital

Estadísticas en



Entrevista a Rafael Fernández Campos:

“La legislación puede regular, pero nunca sustituir la existencia de un conjunto de principios éticos societarios”

Entrevista a Carlos Ballano Fernández

“Las técnicas estadísticas van a ser absolutamente necesarias para la explotación de las nuevas fuentes de datos”

Índice

REVISTA DE ESTADÍSTICA Y SOCIEDAD

Edita: **Universidad Autónoma de Madrid**
Facultad de Ciencias Económicas y Empresariales. Módulo 6
Carretera de Colmenar Viejo, km 15 - 28049 Madrid
www.revistaindice.com
e-mail: revistaindice@revistaindice.com

Comité editorial

Director: **Diego Cano Soler**

Director adjunto: **Diego S. Garrocho Salcedo**

Consejo editorial

Antonio Berlanga de Jesús
Ángel Bizcarrondo Ibáñez
Antonio Camuñas Baena
Diego Cano Soler
Carmen María Casado Santana
Miguel Ángel de Castro Puente
José Carlos Díez Gangas
Mireia Farré Mallofré
Rafael Fernández Campos
Adolfo Gálvez Moraleda
Luisa Margarita García Ferruelo
Rafael Garesse Alarcón
Diego S. Garrocho Salcedo
Gregorio Gil de Rozas Balmaseda
Ignacio González Veiga
Juan José de Lucio Fernández
Donald Peña Martínez
Fidel Rodríguez Batalla
Juan Manuel Rodríguez Pío
Luis María Sáez de Jáuregui
Lis Paula San Miguel-Pradera
Lázaro Villada Ruiz

Colaboradores

Antonio Argüeso Jiménez
Emanuele Baldacci
Carlos Ballano Fernández
Fernando Bolívar García
Rafael Fernández Campos
Yolanda Gómez Menchón
Montserrat Guillén Estany
Elena Herrero-Beaumont
Fabio Ricciato
Ana Carmen Saura Vinuesa
Sirenia Vázquez Báez
Albrecht Wirthmann

Proyecto gráfico y cubierta: J. A. Alcalá

Composición y maquetación: JMR

Impresión y encuadernación: Estilo Estugraf Impresores, S.L.
Distribuye: INE

Fotos: Adobe Stock, Photodisc, Archivo INE

Publicación incluida en el programa editorial del Ministerio
de Asuntos Económicos y Transformación Digital.

En esta publicación se ha utilizado papel con certificación
FSC

Catálogo de publicaciones de la Administración
General del Estado

<http://publicacionesoficiales.bos.es/>

Acceso a la edición electrónica:

www.revistaindice.com (ISSN 1697-2325)

La revista Índice se edita mediante un convenio entre:



UAM

Universidad Autónoma
de Madrid

Colaboran:



Instituto de Estadística de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN Y CERVCA



Junta de
Castilla y León
Consejería de Hacienda
Dirección General de Estadística



Generalitat de Catalunya
Institut d'Estadística
de Catalunya



INSTITUTO GALEGO DE
ESTADÍSTICA

Depósito Legal: M-40226-2003 - ISSN: 1697-2325 - NPD: 095-20-011-6

***“En una época de engaño universal decir
la verdad es un acto revolucionario”***
George Orwell

Elena Herrero-Beaumont La Nota	2
Entrevista a Carlos Ballano Fernández	4
Emanuele Baldacci, Fabio Ricciato y Albrecht Wirthmann A Reflection on The Re(Use) of New Data Sources for Official Statistics	8
Yolanda Gómez Menchón Nuevo paradigma para la estadística oficial	12
Entrevista a Rafael Fernández Campos	15
Ana Carmen Saura Vinuesa Retos de la estadística en una sociedad altamente datificada	18
Antonio Argüeso Los estudios de movilidad del INE a partir de la telefonía móvil (2019-2021)	21
Fernando Bolívar García y Sirenia Vázquez Báez El uso del Big Data para la producción estadística. La experiencia de BBVA Research	24
Montserrat Guillén Estany Prevenir los accidentes antes de que sucedan con telemática	29
Reseña de publicaciones	32

La generación de datos procedentes de la actividad cotidiana de millones de personas interrelacionadas se constituye en fuente de primer orden para conocer no solo la fotografía estática de la sociedad sino también su previsible evolución. En este número trataremos las distintas oportunidades que ofrece el aprovechamiento de datos masivos para la elaboración de estadísticas oficiales y los desafíos que conlleva.

En La Nota, **Elena Herrero-Beaumont**, de Ethosfera, nos introduce en el tema con especial hincapié sobre el valor de la estadística para la sociedad. Entrevistamos a **Carlos Ballano**, subdirector del Departamento de Control de la Producción Estadística y Muestreo del INE, sobre los nuevos proyectos que se están llevando a cabo en el INE mediante el uso de datos masivos. A continuación, el artículo de **Emanuele Baldacci, Fabio Ricciato y Albrecht Wirthmann**, de Eurostat, trata sobre los cambios legislativos que se están iniciando en Europa para tener un marco legal que permita el acceso y uso a estos datos con las máximas garantías para los ciudadanos y las empresas propietarias de los datos. **Yolanda Gómez** y **Ana Carmen Saura**, del Gabinete de Presidencia del INE, nos detallan en sus artículos los temas tratados y las conclusiones de las “Jornadas sobre acceso a nuevas fuentes de datos” donde se ha debatido sobre la cooperación público-privada y entre diferentes organismos de la administración pública. Entre ambos artículos también entrevistamos a **Rafael Fernández**, presidente del Club CDOs de España & Latam, y le preguntamos sobre la ética del dato y las salvaguardas legales y operacionales que requiere un tratamiento garantista de los datos con la intimidad de las personas. En los últimos artículos del número presentaremos algunos ejemplos de tratamiento y explotación de grandes bases de datos. Así, **Antonio Argüeso**, del INE, detalla la elaboración de los estudios de movilidad a partir de la telefonía móvil, **Fernando Bolívar** y **Sirenia Vázquez**, de BBVA Research, las distintas aplicaciones llevadas a cabo en su institución a partir de los datos generados por sus clientes y **Montserrat Guillén**, de la Universidad de Barcelona, trata el uso de la telemática en la seguridad de la conducción de vehículos y en la tarificación de los seguros. ●

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 multidisciplinarios 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.

ENTREVISTA

Carlos Ballano Fernández

“LAS TÉCNICAS ESTADÍSTICAS VAN A SER ABSOLUTAMENTE NECESARIAS PARA LA EXPLOTACIÓN DE LAS NUEVAS FUENTES DE DATOS”



Actualmente, la actividad digital de individuos y organizaciones genera una gran cantidad de información que vemos cómo cada día es aprovechada por distintas empresas en la generación de nuevos productos y servicios. ¿En qué medida la estadística oficial puede aprovechar también estos datos para mejorar los productos que proporciona actualmente a la sociedad?

Hay que tener en cuenta que la estadística oficial tiene una importante experiencia en la utilización de información no procedente de encuestas para la generación de información. Desde hace años se utiliza la información procedente de ficheros administrativos con una finalidad estadística. Estos usos se extienden a distintos tipos de operaciones estadísticas:

- ▶ Para la formación de directorios, como es el caso del Padrón Municipal de Habitantes para el Marco de Población, o los ficheros que se utilizan para la formación del DIRCE.
- ▶ Otros para completar variables en encuestas, como es el caso de los datos de renta en la Encuesta de Condiciones de Vida, o de salario y discapacidad en el caso de la Encuesta de Población Activa.
- ▶ Sustituir la recogida directa. En el Censo Agrario 2020 la información de más de 650.000 explotaciones ha procedido de la explotación de la información de registros, fundamentalmente el FEGA del Ministerio de Agricultura, investigándose el resto mediante recogida directa; o en el caso de la Estadística Estructural de Empresas, en la que se utiliza la infor-

mación administrativa para disminuir el tamaño muestral.

- El Censo de Población y Viviendas 2021, que consta de una investigación exhaustiva a partir de ficheros administrativos y una operación en campo para estudiar las características esenciales de la población.
- Para el tratamiento estadístico de la no respuesta, como es la calibración a las cifras de población en las encuestas dirigidas a los hogares.

La utilización de esta información no es directa, sino que exige un trabajo de adecuación a los criterios estadísticos que, en el caso de operaciones importantes, puede significar un trabajo de varios años, como es el caso del uso en los Censos de Población y Viviendas o del Censo Agrario.

También exige en algunas ocasiones la utilización de técnicas estadísticas no estándares en las encuestas habituales por muestreo a través de marcos de lista; un ejemplo es el empleo de muestreo indirecto para la obtención de información a nivel de empresa estadística a partir de la información procedente de las unidades legales.

En la actualidad a la información anterior se añade la mayor disponibilidad de información procedente directamente del registro que queda de la actividad de los individuos, como es la procedente de telefonía móvil, tarjetas de crédito, reservas por internet de hoteles o apartamentos, compras por internet... El potencial estadístico de toda esta información se podría clasificar en dos categorías:

- Para proporcionar nuevos indicadores.
- Para, combinada con encuestas ya existentes, facilitar los desgloses temporales, espaciales o de otro tipo, proporcionando una mayor granularidad de la información generada por encuestas.

Como en el caso de los registros administrativos, es necesario un trabajo arduo para aproximar los conceptos estadísticos de interés en base a la información externa disponible. Es, por esto, que han aparecido en el ámbito de la estadística

oficial las ‘estadísticas experimentales’, que permiten un acercamiento de ‘aprender mientras se hace’, de forma que mientras se va ajustando los procedimientos se va ofreciendo un producto a la sociedad hasta que es un producto maduro que puede ser parte de la estadística oficial.

Como muestra de las posibilidades que ofrece esta información me referiré solo a las estadísticas experimentales del INE: la información sobre movilidad en base a la información de la telefonía móvil, estimación de la ocupación en establecimientos turísticos en base a la información de las plataformas digitales, estimación de los gastos de visitantes extranjeros en base a la información procedente de las tarjetas bancarias, o el estudio de viviendas turísticas en base a técnicas de *web scraping*.

La utilización de esta información no es directa, sino que exige un trabajo de adecuación a los criterios estadísticos que, en el caso de operaciones importantes, puede significar un trabajo de varios años

Además de estas estadísticas experimentales el INE está trabajando en otros proyectos como es la obtención de precios directamente de las grandes superficies de España, para su utilización en el IPC.

Obviamente, lo anteriormente mencionado no agota las posibilidades de utilización y las oficinas de estadística de nuestro entorno trabajan en proyectos en los que las nuevas fuentes de información ofrecen posibilidades de mejora en la estadística oficial. Un ejemplo relevante es el proyecto Big Data for European Statistics (BDES), lanzado por Eurostat, en el que el INE ha colaborado en un subproyecto, y en el que se han investigado en profundidad la potencialidad estadística de distintas fuentes de Big Data, como la telefonía móvil, las ofertas de trabajo *on line*, la obtención de informa-

ción de las empresas a través de sus páginas web, de los hogares a partir de los contadores eléctricos, de las transacciones financieras de distintos instrumentos y plataformas de pago, de la información de la tierra procedente de satélites y fuentes para el turismo.

Las oficinas de estadística de nuestro entorno trabajan en proyectos en los que las nuevas fuentes de información ofrecen posibilidades de mejora en la estadística oficial

Sin duda el aprovechamiento de las nuevas fuentes de datos supone un reto para las instituciones estadísticas. ¿Puede señalar los aspectos que en su opinión son los más importantes y novedosos en relación al uso de estas nuevas bases de datos?

En mi opinión son importantes los siguientes aspectos:

- ▀ Legal: al ser un fenómeno emergente hay que ver cómo se considera desde la legislación estadística, tanto europea como nacional. Es un asunto que se está tratando en ambos niveles y que tiene que ver con la gestión de la información que se considera un bien público desde el punto de vista de la estadística oficial.
- ▀ De infraestructura: es necesario pasar a infraestructuras informáticas que permitan mucha mayor capacidad de almacenamiento y de tratamiento de la información. Existe un proyecto de la Secretaría General de la Administración Digital en este sentido.
- ▀ De metodología estadística: sobre los procedimientos para transformar y utilizar los datos disponibles a información estadística relevante. Aquí hay que tener en cuenta la experiencia de las oficinas de estadística en el uso de información auxiliar, tanto para estimaciones por ca-

libración como en las técnicas estadísticas para estimación en pequeñas áreas.

- ▀ De medida de la calidad de la información: la información producida debe ser de utilidad para la toma de decisiones. En este aspecto hay que tener en cuenta que el control del proceso en las encuestas puede servir para corregir los sesgos de cobertura de las nuevas fuentes de datos. En algunos casos, las fuentes externas también pueden ayudar a detectar sesgos en las encuestas.
- ▀ De seguridad y preservación de la confidencialidad estadística y de protección de datos: los sistemas deben de satisfacer todas las condiciones que se establecen en el Esquema Nacional de Seguridad, la legislación nacional y europea sobre protección de datos y secreto estadístico.

Durante mucho tiempo las técnicas de muestreo han sido nucleares para el diseño de las operaciones estadísticas. Estas técnicas contribuyen a disponer de información relevante, midiendo su calidad de manera científica y permiten asignar los recursos disponibles de forma óptima. En la época de las fuentes Big Data, ¿cree que disminuye la importancia del muestreo en favor de otras técnicas o seguirá siendo en el futuro un referente principal del trabajo de instituciones como el INE?

Como he comentado más arriba, mi opinión es que las técnicas estadísticas van a ser absolutamente necesarias para la explotación de las nuevas fuentes de datos. Y no me refiero solamente a que en algunos casos la magnitud de la información disponible haga necesario tomar muestras para su explotación. Aunque habrá indicadores que se deriven directamente de dichas fuentes, en muchos otros casos será la combinación de las nuevas fuentes y encuestas las que se deban utilizar para la producción de la información.

Por otra parte, considero adecuada la cita de Deming de 1950 que utiliza Lohr: 'El muestreo no es el reemplazo de una cobertura parcial por otra total, sino la ciencia y el arte de controlar y

medir la fiabilidad de la información estadística a través de la teoría de la probabilidad’.

Desde este punto de vista más amplio, creo que las nuevas fuentes de datos van a provocar avances en la teoría y práctica del muestreo.

Se entiende que el mayor aprovechamiento de los datos existentes redundará en una menor carga directa de solicitudes de información por parte de las instituciones estadísticas a los hogares y sobre todo a las empresas. Pero mientras tanto, ¿qué medidas adopta el INE para minimizar esta carga?

El INE es consciente de la carga estadística que genera. Por este motivo lleva trabajando desde hace décadas en técnicas para mitigar y distribuir esa carga.

En lo que atañe al diseño muestral aplicamos una técnica llamada de Números Aleatorios Permanentes, junto a una función de carga estadística, para obtener las muestras estructurales dirigidas a las empresas de la manera más equitativa posible, en el sentido de distribuir la carga estadística a las empresas, medida en el número de encuestas que recibe y en el tiempo estimado de realizar los cuestionarios, lo más uniformemente posible. Si una empresa es seleccionada de manera aleatoria para realizar una encuesta en un determinado año, dicha empresa tiene mucha menos probabilidad de que vuelva a ser seleccionada en otra encuesta dirigida a empresas en ese año, si por el tamaño no se investiga de forma exhaustiva.

También se lleva a cabo un control del número máximo de estadísticas en las que colabora una empresa en función de su tamaño.

Asimismo se utiliza la información administrativa disponible tal como he comentado más arriba. Un caso relevante es el uso del Impuesto de Sociedades para precargar los datos en la

cumplimentación a través de la web de la Estadística de Estructura Económica. Por motivos de calendario solo podemos utilizar la información en la muestra de esta encuesta que se lanza en septiembre, pero esperamos que en breve pueda extenderse el uso al total de la muestra de esta encuesta.

Acabamos nuestros encuentros pidiendo a los entrevistados un esfuerzo de imaginación. ¿Cómo ve la sociedad española dentro de 20 años? Denos un temor, una prioridad y un deseo para España.

Una forma de ver el futuro es mirar al pasado y proyectar la evolución percibida. Hace 20 años el INE contrató más de 40.000 personas para hacer el Censo de Población y Viviendas. En 2021 el presupuesto y el personal se han reducido sustancialmente debido a la disponibilidad de fuentes administrativas y de Big Data.

Si proyecto esta evolución 20 años en el futuro concibo que lo que habrá será unos sistemas de información integrados y georreferenciados que proporcionarán información a un nivel de granularidad temporal y espacial suficiente, una vez identificadas las potencialidades y metodologías de combinación de fuentes más adecuadas, con unos criterios de calidad también adecuados y preservando el secreto estadístico.

Un temor, mejor dos: la inmovilidad por un lado, y la adopción de novedades no suficientemente maduras para la producción, por otro.

Una prioridad: identificar las áreas de mejora que nos sitúen en una mejor posición para afrontar el futuro.

Un deseo: que sepamos transmitir a los nuevos estadísticos que se incorporarán en las próximas oposiciones la necesidad de mejorar el INE, de la misma forma que en su momento nos lo transmitieron a nosotros los que nos precedieron. ●

CARLOS BALLANO FERNÁNDEZ

Subdirector del Departamento de Control de la Producción Estadística y Muestreo del INE.

Funcionario del INE desde 1985. Ha trabajado fundamentalmente en asuntos relacionados con el muestreo y la recogida de datos.

Licenciado en Ciencias Matemáticas y en Ciencias Físicas.

Diploma de Estudios Avanzados en ‘Sistemas Estocásticos y su Control Óptimo’.

A Reflection on The Re(Use) of New Data Sources for Official Statistics

Emanuele Baldacci

Eurostat, Director of Resources

Fabio Ricciato

Eurostat, Statistical Officer

Albrecht Wirthmann

Eurostat, Head of Unit Methodology; Innovation in official statistics

The essential mission of official statistics is to provide information about the society, the economy and the environment to citizens, policy makers, researchers and economic actors, enabling them to take decisions and form opinions based on facts and evidence. In a nutshell, official statistics provide the society with “knowledge about itself”. Reliable, trustworthy, relevant and timely official statistics are among the enablers of a healthy democratic society.

Generally speaking, mandating the production of official statistics to specialised public bodies, i.e. statistical offices, impartial and independent from other public and private organizations, is a pre-requisite for ensuring trustworthiness of the statistical information. But what does the *production of official statistics* entail in practice?

In the pre-digital world, the only way to measure the society was “asking the people”. In census and surveys, what to ask (i.e., questionnaires) and *whom to ask* (i.e., samples) were determined directly by the statistical offices and primarily for statistical purposes, with the aim of producing impartial and representative information at minimum cost. With survey and census data the entire statistical production process, from data collection through data processing and dissemination of the final statistics, is performed entirely by the statistical office.

Survey and census data were later complemented by administrative data, collected by other public bodies (central and local administrations) primarily for administrative purposes, and then re-used by statistical offices for statistical purposes. With administrative

data, the data collection stage falls outside the statistical office but still remains within the public sector. Like survey and census data also administrative data are based on explicit declarations made by people (in the context of administrative procedures).

During the last two decades the digitalisation and *datafication* of our lives has changed many aspects of our behaviours. More and more of our actions, transactions and interactions are mediated by, directed to or at least sensed by digital systems and machines. Our online activities are natively embedded into the digital space, but also our physical activities leave digital traces via sensors and ‘smart’ devices. Each individual event in our lives – be it a purchase, a transaction, a movement, etc. – is now transformed into one or even more data points by some digital system(s). If ‘micro-data’ refers to the characteristics of an individual, the term ‘nano-data’ was proposed to refer to granular, behavioural data referring to individual events at sub-individual level. Private companies collect nano-data primarily for business purposes (e.g., for delivering services to their customers, or to gain more detailed knowledge about their needs and behaviours) and statistical offices are eager to reuse such data for producing more, better and timelier official statistics.

There is nowadays an increasing awareness that the data generated *by the citizens*, even if collected by private companies, have also a public value *for the citizens*. Such public value however is not being adequately ripped nowadays. Reusing such data for the production of better and timelier official statistics, that are then disseminated publicly and for free, is a way to “give back” part of the data value to the citizens themselves, to the extent that better statistics have the potential to lead to better policies and a better informed society.

The digitalisation of our lives has not only changed the way we produce data, but also the way we *consume* data and information. Citizens

and policymakers expect today to be informed in a more precise, comprehensive and timelier manner. They expect quantitative information (i.e., statistics) to describe an increasing number of phenomena. Official statistics must keep pace with such increased demand, and this is simply not possible by relying solely on traditional survey and administrative data. For statistical offices, plugging statistical production into the flow of machine-generated data is not only an *opportunity* driven by the availability of new data on the input side, but also a necessity in order to respond to the increased expectations and needs by the users of official statistics on the output side.

In most cases, new data sources will not replace entirely traditional ones but rather complement them. For example, some survey data may still be needed to calibrate and adjust the view offered by new data sources, e.g. to mitigate possible biases and incompleteness. In fact, the data from business companies typically refer to their customer base, that by definition a sub-population and cannot be considered representative of the whole target population, hence the need of making adjustments and calibration based on other auxiliary data. But a survey designed specifically to serve as auxiliary and complementary source for calibration purposes will be different (and likely less costly to implement) than a survey intended to serve as the sole primary data source. In other words, the opportunity to leverage new data sources involves a rethinking of the role of traditional data sources, not their complete dismissal.

Bringing new data sources into the process of statistical production requires solving a number of challenges and open issues. For some of them, the solutions will be different from those adopted in the context of traditional data sources.

The first issue concerns the legal and business aspects of data reuse. In most cases such new data are generated in the private sector. Whether such data represent a by-product or a central component of the business operations, they clearly have both a *private value for the business and a public value for the whole society*, and we need rules to govern their use in both domains. The European Commission is currently

working towards a new legal framework as part of the European Digital Strategy. The reuse of private data for public purposes, or business-to-government (B2G) data sharing, will be focus of the forthcoming Data Act proposal that is currently under preparation by the European Commission¹. For what concerns more specifically official statistics, Eurostat has recently taken the initiative to start a targeted revision process of Regulation 223/2009 on European Statistics with a view to adapt it to the new needs of official statistics. One prominent goal of such a revision is to enable the sustainable reuse of privately held data for the development, production and dissemination of official statistics in Europe. The ambition of Eurostat is to finalize the new Regulation by end of 2022.

The reuse of private data for public purposes, or business-to-government (B2G) data sharing, will be focus of the forthcoming Data Act proposal that is currently under preparation by the European Commission

From an operational point of view, the legal aspects of data reuse must be articulated with the technical aspects, i.e., translated into operational modalities defining how data held by one organization are accessed (i.e., reused) by another organization, and specifically by statistical offices. In many cases of practical interest, it will not be convenient, if at all possible, to move enormous amount of nano-data (or granular data) from the data holders to the statistical offices. In such cases, the input nano-data could be processed and aggregated, at least partially, at the premises of their respective data holders, with only the intermediate or

¹ See the ESS position paper <https://europa.eu/!KyUK9C> and the ESGAB opinion concerning the forthcoming Data Act proposal <https://europa.eu/!QQMb4>

final results passed to the statistical office. This approach reflects the idea that in statistical production *raw data are a means, not a goal*. Instead of moving raw data into the statistical office, it is often preferable to move methods and algorithms towards the point where data are generated. The statistical office will receive a lower volume of (aggregate) of data but with higher information content. This approach helps to minimise the risks.

Methodologies for new data sources need to be co-developed by multi-disciplinary teams where expert statisticians collaborate with specialists from the specific technological or business field relevant to the concerned data

The methods and algorithms by which the input data are processed, filtered and aggregated, including the part thereof executed automatically at the premises of the private data holders, should be fully open and standardised, so as to preserve methodological transparency and harmonisation across different data providers. Generally speaking, transforming machine-generated data into reliable statistics requires dealing with several sources of uncertainty (errors, bias, limits to temporal and spatial resolution, unknowns, etc.) that are specific to the system or technology from where they are generated, and do not always have an exact homologous in traditional human-generated data sources. For this reason, methodologies for new data sources need to be co-developed by multi-disciplinary teams where expert statisticians collaborate with specialists from the specific technological or business field relevant to the concerned data (e.g., telecom engineers for mobile network operator data). Compared to data that were designed primarily for statistics, the methodologies for processing reused data sources tend to be more complex. They need to cope with configuration-specific aspects of data generation that differ

somewhat across different sources (e.g., certain characteristics of the location data generated by the mobile network depends on the configuration of the network infrastructure and on the characteristics of the mobile customer base that vary across operators). They must be adapted more frequently in order to adapt to changes in the data generation mechanisms that follow the physiological evolution of technologies and markets (e.g., improvement of mobile network technology and changes of mobile usage due to different tariff schemes). All these aspects contribute to increase the cost and the complexity of methodological development for the reuse of new data sources in official statistics. In other words, going from survey and census data towards new machine-generated data sources, the dominant cost component of statistical production shifts from data collection to methodological development.

Also for this reason, the European dimension becomes extremely important for the development of new statistical methodologies. Pooling methodological development at the European level is more *efficient* in terms of resources, skills and capacities, as one methodology developed jointly by the ESS at the European level can be applied (possibly with customisations and limited adaptations) to all Member States, but it is also more *effective* in terms of better harmonisation and consistency of results. This is possible also because machine-generated data tend to display less national specificities that, say, administrative data, since the underlying technologies tend to be uniform across countries (e.g., smartphones and mobile networks standards are global, not country-specific). In many cases, they are generated by systems and platforms with an intrinsic multinational dimension (e.g., online platforms, credit card payment systems). In some business sectors, the level of data heterogeneity (both in format and semantic) across different data providers, also *within* the same country, prevails over difference *across* countries. That means, all countries would have to face basically the very same methodological and technical challenges, reinforcing the convenience of pooling methodological development at the European level. There may be still national

specificities to be taken into account though, e.g. on behavioural aspects (e.g., the number of mobile phones used by each individual may vary across different countries) but such differences can still be addressed in the context of a single methodological process conducted at the European level by leaving room to flexibility and customisations.

From the discussion above it should be clear that reusing new data sources for official statistics has a cost. There are costs in developing new methodologies, there are costs in deploying IT infrastructures that implement these methodologies and there are organisational costs in establishing and implementing agreements between the statistical officers and data holders. Regardless of how they are distributed across the involved actors, these costs must be assessed against the benefits that more comprehensive, detailed and timely statistics will bring to the society. Such cost-vs-benefit analysis must be conducted for each kind (or class) of new data sources and periodically reassessed. As a matter of fact, some kinds of data sources are not yet sufficiently mature to be used reliably as a source of official statistics, or anyway involve costs that would be repaid by a proportionate benefit, but they may become appealing in the future. In other words, statistical system must be able to select pragmatically

the subset of data sources for which *the gain is worth the pain*. Data sources generated by early stage technologies that are not yet sufficiently widespread would likely bring little benefit. Data sources generated by systems characterised by large degrees of heterogeneity and technological dispersion (many proprietary systems instead of few common standards) would likely involve too high costs. The most appealing data sources for official statistics are generated in mature sectors, with high penetration and low technological heterogeneity and the potential for reuseability in different statistical domains.

In conclusion, this short article has attempted to provide a coherent “big picture” of the issues that the European Statistical System is facing in order to move forward on the reuse of new data sources for official statistics. We must move in parallel on multiple fronts notwithstanding their interdependencies. Achieving an appropriate legal framework for data reuse is the most compelling *conditio sine qua non* future statistical production can benefit from certain kinds of new data sources. We as statisticians need to (co-) develop new methodologies for data that are very different from the ones we are familiar with, and in so doing we must seek the collaboration by experts from other domains. On each of these fronts, joining efforts at the European level will be key to success. ●



Nuevo paradigma para la estadística oficial

Yolanda Gómez Menchón

Gabinete de Presidencia del INE

En las últimas décadas hemos visto cómo los avances tecnológicos han ido creciendo e invadiendo todas las facetas de nuestras vidas. Nos hemos digitalizado, ya no entendemos la vida sin navegar por internet, sin el correo electrónico, sin las aplicaciones móviles o las redes sociales, por ejemplo. Y esta digitalización de nuestra sociedad está impulsando el desarrollo de unos servicios públicos más cercanos al ciudadano y más eficientes.

La estadística no es una excepción, toda la información que generan estas nuevas formas de comunicación social, así como el llamado internet de las cosas, es potencialmente relevante para las estadísticas oficiales. Hoy día es un hecho ya indiscutible que la utilización de nuevas fuentes de datos y, en especial de los Big Data, en la estadística oficial supone un cambio de paradigma y conlleva la necesidad de adaptar los enfoques de la producción estadística, las capacidades y competencias de los estadísticos, los métodos de producción y las tecnologías utilizadas buscando que nuestras estadísticas sigan siendo significativas y relevantes, es decir, manteniendo el alto grado de confianza del que disfrutamos por parte de la sociedad.

En este contexto, el Gabinete de la Presidencia del Instituto Nacional de Estadística (INE) ha organizado las **Jornadas sobre acceso a nuevas fuentes de datos**¹ que a lo largo de tres sesiones virtuales han puesto sobre la mesa los aspectos más relevantes relacionados con el uso de nuevas fuentes de datos para la estadística oficial. En las sesiones participaron, junto al Presidente del INE, Juan Manuel Rodríguez Póo, el Director de Recursos de Eurostat, altos

cargos de la administración española, incluida la Directora de la Agencia Española de Protección de Datos, junto con expertos en gestión de datos de empresas, prensa y del mundo académico. En la primera conferencia celebrada en mayo se debatió sobre **“El nuevo paradigma para la estadística oficial”**, centrándose la sesión de junio, **“Estadística y sociedad”**, en el impacto de la estadística oficial en la vida de los ciudadanos y en la sociedad en general, mientras que la sesión de octubre estuvo focalizada en el intercambio de nuevas fuentes de datos entre empresas y administración bajo el título **“Nuevas fuentes de datos y estadísticas oficiales: B2G4S”**. Las jornadas han sido muy bien acogidas con una asistencia media de 250 personas.

Durante las sesiones se puso de manifiesto la solvencia del INE y los servicios estadísticos en la elaboración de las estadísticas oficiales ya que están sujetos a sólidos principios, criterios y procedimientos establecidos tanto en la legislación nacional como europea. De la misma manera, para poder hacer estadísticas fiables y pertinentes es necesario que el INE pueda acceder a todo tipo de datos, ya sean de identificación personal directa, indirecta o anonimizados, de cualquier persona física o jurídica, pública o privada, dentro de un marco de garantía de preservación de la privacidad y la seguridad de los datos.

Las oficinas de estadística abrieron ya hace unos años un debate a nivel mundial sobre el futuro de la estadística en un mundo globalizado y altamente tecnológico. El INE de España ha estado presente en muchos de estos debates y ha participado en las iniciativas que se han lanzado, sobre todo a nivel europeo, para avanzar hacia la modernización de los sistemas estadísticos. A través de las mismas se ha podido constatar que los beneficios y oportunidades del uso de nuevas fuentes de datos son numerosos, por un lado, se obtienen estadísticas más oportunas, pertinentes y relevantes

¹ <https://ine.es/ss/Satellite?L=0&c=Page&cid=1259953169112&p=1259953169112&pagename=INE%2FINELayout>

(sin que ello implique pérdida de transparencia, independencia o imparcialidad) lo que conlleva, a la larga, una reducción de la carga sobre los informantes y una disminución en los costes de producción. Por otro lado, la estadística oficial puede actuar como catalizador de la innovación en otras áreas del sector público. Además, el uso de estos datos por las autoridades estadísticas reduce el riesgo de alta concentración de datos personales en pocas grandes empresas, lo que supone mejor información y más transparente. Y, por supuesto, las nuevas estadísticas implican una mayor oportunidad y calidad en la toma de decisiones hacia el ciudadano.

El año 2020 y el actual año en curso han supuesto un impulso importante en lo referente a la modernización de la producción estadística y a las iniciativas en el uso de nuevas fuentes de datos. La trágica situación de pandemia mundial llevó aparejada una necesidad de nuevos datos para la toma de decisiones sobre la evolución de la pandemia y el INE pudo tener mayor acceso a datos privados y a registros administrativos con este fin. Estas experiencias han constituido en cierta manera un laboratorio de datos acelerado que está siendo muy útil para el desarrollo de nuevas estadísticas oficiales. Aunque el INE ya estaba trabajando en esta línea, un proyecto que se puso en marcha hace unos años y se está desarrollando en la actualidad es, por ejemplo, el Censo 2021² elaborado por primera vez solo con registros administrativos. También había ya empezado a elaborar las llamadas estadísticas experimentales³ basadas en nuevas fuentes de datos, como el Atlas de distribución de renta de los hogares o los estudios de movilidad a partir de la telefonía móvil, entre otros, junto con nuevos proyectos en fase de desarrollo en los que se están utilizando nuevas fuentes de datos y probando una nueva metodología. Estos proyectos innovadores han sido y son posibles gracias a la colaboración entre administraciones y, por supuesto, a la inestimable colaboración del sector privado. Esta

colaboración para el acceso a nuevas fuentes de datos debe ir en aumento en beneficio de todos y debe ejercitarse, tal y como recordó el Presidente del INE, respetando los principios de especialidad, igualdad de condiciones, proporcionalidad, confidencialidad estadística y coste-beneficio.

Las oficinas de estadística abrieron ya hace unos años un debate a nivel mundial sobre el futuro de la estadística en un mundo globalizado y altamente tecnológico

En las sesiones también se comentó el impulso que el INE ha dado al debate institucional en nuestro país sobre el futuro de la estadística oficial en relación con las nuevas fuentes de datos y los nuevos roles que pueden asumir los servicios estadísticos en una sociedad datificada. El Consejo Superior de Estadística encargó a un grupo de expertos el estudio del papel de los servicios administrativos como Data Stewards, esto es, como administradores de datos y el Comité Interterritorial de Estadística creó el Grupo de Trabajo para el estudio de la creación de un Nodo de Interoperabilidad Estadística Nacional mientras que en el seno del Foro Regional de Estadística se está estudiando la Estadística en el ecosistema de datos. Recientemente el INE junto con la Agencia Tributaria, el Banco de España y la Seguridad Social han puesto en marcha un proyecto conjunto para diseñar un sistema colaborativo de datos para fines científicos de interés público.

Sin embargo, aunque se ha andado mucho camino en estos últimos años aún quedan cuestiones por resolver sobre el uso del Big Data en la estadística. En primer lugar, ahora somos conscientes de la importancia de transmitir a la sociedad qué es la estadística, cómo funcionamos y porqué necesitamos, al igual que cualquier ciudadano, adaptarnos a los nue-

² <https://www.ine.es/censos2021/>

³ <https://www.ine.es/experimental/experimental.htm>

vos tiempos. Comunicación y transparencia es por lo tanto uno de los retos fundamentales y para afrontar este reto de una manera uniforme el Sistema Estadístico Europeo ha adoptado un “Plan de Comunicación” que cada país irá adaptando a su propia cultura e idiosincrasia. De hecho, las Jornadas sobre Acceso a Datos organizadas por el INE han sido parte de este Plan de Comunicación más exhaustivo.

En segundo lugar, es necesario adaptar la legislación existente, las nuevas reglas deben tener como objetivo maximizar el beneficio para toda la sociedad equilibrando los derechos e intereses legítimos de todas las partes involucradas. Varias acciones se han puesto en marcha en paralelo para hacer frente a este reto. Por un lado, a nivel europeo se están tomando iniciativas para que las necesidades estadísticas estén convenientemente recogidas en la futura Ley del Dato Europea y en la Ley estadística europea. Por otro lado, a nivel nacional, el Plan Estadístico Nacional en vigor ya contempla como una de las líneas estratégicas para este periodo la reforma de nuestra ley estadística, la Ley de la Función Estadística Pública, en particular para recoger estos aspectos.

En tercer lugar, es necesario, para modernizar la producción estadística, adaptar los conocimientos y capacidades de nuestro personal a los nuevos retos y mejorar las infraestructuras tecnológicas. Las “Medidas de Racionalización de la producción estadística” aprobadas por el INE, junto con la adaptación del temario de las pruebas de acceso al Cuerpo de Estadísticos del

Estado, el diseño de nuevos cursos de formación o de nuevas herramientas, son algunas de las medidas dirigidas a alcanzar este objetivo.

Por último, es importante que en todo momento garanticemos el mismo nivel de calidad del que gozan las estadísticas que ya hemos calificado de “tradicionales”. Para ello el INE sigue participando en los estudios piloto a nivel europeo en los que se testan nuevas metodologías, a la vez que sigue avanzando en esta línea con las estadísticas experimentales buscando la colaboración de los medios académicos y del sector privado para avanzar conjuntamente y alcanzar el objetivo común, que no es otro que la mejora de nuestra sociedad.

Las jornadas se cerraron con un mensaje clave, el objetivo de los servicios estadísticos es conseguir colaboraciones con el sector privado de larga duración y que sean beneficiosas tanto para el sector privado como para el interés público. En definitiva, la estadística debe poder utilizar las fuentes de datos públicas y privadas para transformarlas en información relevante y fiable devolviendo así el valor de esos datos a la sociedad. Las estadísticas oficiales son un bien público y el uso de estos datos para las estadísticas oficiales significa “devolver” a los ciudadanos y a toda la sociedad parte de su valor, en forma de información fiable. El INE debe seguir avanzando en esta línea y, como destacó el Director de la Oficina del Dato, *“el INE deberá tener un papel protagonista en la estrategia del dato, tanto a nivel nacional como europeo y será el actor clave para su implantación en España”*. ●



ENTREVISTA

Rafael Fernández Campos

“LA LEGISLACIÓN PUEDE REGULAR, PERO NUNCA SUSTITUIR LA EXISTENCIA DE UN CONJUNTO DE PRINCIPIOS ÉTICOS SOCIETARIOS”

Eres Presidente del Club Chief Data Officer. ¿En qué consiste este club?, ¿cuáles son su misión y objetivos? Agrupáis una figura profesional de relativa reciente implantación en las empresas españolas. ¿Cuál es nuestra situación en el cuidado y gestión del dato en relación a nuestros pares europeos?

El Club nace de la necesidad de los responsables de datos de las principales compañías de nuestra geografía de disponer de un foro donde resolver problemas y crecer juntos. Este objetivo ha ido evolucionando y en la actualidad, además de celebrar eventos, se genera contenido de alto valor para la comunidad del dato, diseñamos formación en colaboración con escuelas de negocio y se han desarrollado dos herramientas de ayuda a la estrategia de datos: el dataMat, un cuestionario de autoevaluación del grado de madurez de las compañías en términos de datos, y el dataToolkit, un *gaming* que ayuda a los responsables de datos a implantar una estrategia de datos sostenible en sus compañías. Actualmente el Club cuenta entre sus miembros con responsables de 120 compañías de España y Latinoamérica.

La generalización de la sociedad del dato comporta riesgos y oportunidades en su uso. Conocer más implica poder servir mejor, pero también dominar con mayor capacidad. ¿Andamos sobre el filo de la navaja? ¿Hacia dónde podemos caer?

Cualquier avance tecnológico o científico es susceptible de ser utilizado para el bien o para el mal, por supuesto suponiendo que fuéramos capaces de encontrar una definición consen-



suada de lo que es el bien y el mal. El uso masivo de los datos no es diferente de otros avances. No es ni bueno ni malo en sí mismo, sino en función del propósito que persigamos con su uso. Hay compañías éticamente modélicas en el uso masivo de datos y, sin embargo, existen otras cuyos objetivos empresariales pisotean cualquier principio moral básico.

En este sentido yo estoy de acuerdo con aquello que decía Hobbes de que la moral es la flor más perfecta pero que para crecer necesita hacerlo a la sombra de una espada levantada.

El uso masivo de los datos no es diferente de otros avances. No es ni bueno ni malo en sí mismo, sino en función del propósito que persigamos con su uso

Es por ello que la mejor forma de que las compañías respeten una serie de principios morales es regular el uso ético de los datos, unido a la constatación de que los comportamientos inmorales no compensan, es decir, que las sociedades expulsan del mercado aquellas compañías que conculcan la ética aceptada. No obstante, para ello la sociedad debería estar suficientemente formada, lo que nos llevaría a otros dilemas que sobrepasan la pregunta que me formuláis.

La salvaguarda de los valores democráticos en lo público y la dignidad en lo privado consiste en retomar el orden ético en la vida pública tanto de las administraciones como de las empresas. Ética y dato deben aprender a mezclar bien. ¿Cuáles son los próximos compromisos que deben afrontar las empresas y las administraciones en este campo? ¿Los ves posibles?

Como comentaba anteriormente, los retos son relevantes y deben integrar actuaciones por parte de los tres actores principales:

Por un lado, los gobiernos deben ser capaces de encontrar mecanismos de control del uso ético de los datos, no solo a nivel de cada país, sino en ámbitos supranacionales, ya que las compañías que más datos gestionan son globales. Adi-

cionalmente, han de proveer a la ciudadanía de unos conocimientos básicos en términos del uso responsable de nuestros datos.

En segundo lugar, las empresas se enfrentan al mayor de los dilemas: elegir entre generar valor para el accionista y respetar unos principios morales por el bien de la sociedad. La legislación puede regular, pero nunca sustituir la existencia de un conjunto de principios éticos societarios. La acción social, como una suerte de “activismo ético”, puede ayudar a contraponer un poder frente a otro y limitar a las compañías más predatoras.

En último lugar, la sociedad debe asumir una responsabilidad doble. Por un lado, la de su propia formación y, por otro, la certeza de que la mayoría de los datos personales de los que disponen las grandes multinacionales se los hemos proporcionado nosotros mismos. La ley nunca puede subsanar la falta de responsabilidad individual.

En tus libros planteas el dato como activo personal que pueden y deben rescatar los ciudadanos. Esto es fácil de declarar, pero algo más complejo de gestionar para el particular. ¿Qué ayuda es necesario ofrecer para facilitar la recuperación del dato por parte del particular?

La ciudadanía, empezando por los niños, debe ser consciente del inmenso poder que otorgamos a las compañías cuando compartimos datos personales. Las grandes tecnológicas realmente no pelean por el dinero, pelean por el poder. Y ese poder se lo damos cada vez que exponemos nuestras vidas y, lo que es peor, no reaccionamos ante conculcaciones flagrantes de la privacidad.

Sin embargo, para ello la formación es fundamental. Realmente es la vieja sentencia (de Jesucristo) recogida por San Juan “la verdad os hará libres”, actualizada y demostrada de forma incontestable, ya que, en este caso, la ignorancia nos convierte en esclavos de la manipulación.

En todo caso, son una minoría (aunque muy extendida) las compañías que actúan de forma irresponsable. La mayoría de ellas comprende que, sin ética, no hay futuro posible.

Las revoluciones industriales y tecnológicas han tenido también influencia sobre las estructuras sociales y la organización institucional.

Con el desarrollo tecnológico y económico del dato, ¿esperas cambios sociales?

Los cambios sociales suelen ser lentos y, en todo caso, producirse a muy diferentes ritmos en función de multitud de factores como la geografía, los sectores productivos, la clase social... Claramente estamos asistiendo a un cambio en las formas de interacción social y en las fuentes de generación de valor, si bien no veo muy factible que todo ello vaya a tener un alto impacto en la distribución de la riqueza generada o del poder.

No obstante, en mi opinión, los países que sean capaces de fortalecer la creación de oportunidades para todos, el incentivo al emprendimiento y la creatividad unido a la protección de los más desfavorecidos, conseguirán salir vencedores de la disrupción digital. Y para ello, nada mejor que la cooperación entre países.

Acabamos nuestros encuentros pidiendo a los entrevistados un esfuerzo de imaginación. ¿Cómo ve la sociedad española dentro de 20 años? Denos un temor, una prioridad y un deseo para España.

Soy un optimista irredento, así que mi visión para los próximos 20 años es positiva. Creo que si algo

ha demostrado el ser humano en su historia es que es capaz de crear el futuro que se proponga.

Los países que sean capaces de fortalecer la creación de oportunidades para todos, el incentivo al emprendimiento y la creatividad unido a la protección de los más desfavorecidos, conseguirán salir vencedores de la disrupción digital

Respondiendo a tu pregunta, mi principal temor es que no aprovechemos los desarrollos tecnológicos derivados del uso masivo de datos para disminuir las desigualdades existentes. Es una oportunidad única la que se nos presenta.

Por este motivo, la prioridad que veo impostergable es la formación de la ciudadanía para vivir en un mundo digitalizado. La transformación de la fuerza laboral pasa por un esfuerzo inmenso de reorientación de tareas basado en una formación continua, para lo cual la colaboración de las empresas es crítica.

Y en cuanto al deseo, pues que seamos capaces de castigar la incapacidad de nuestros dirigentes para alcanzar acuerdos de país. No es culpa de los políticos, es culpa de los que les elegimos. ●

RESEÑA BIOGRÁFICA



Rafael Fernández Campos, doctor *cum laude* en Economía y Hacienda Pública por la Univ. Autónoma de Madrid y licenciado en Administración de Empresas por la misma universidad. Máster en Dirección de Entidades Financieras por la Univ. Francisco de Vitoria.

Chief Data Officer de Bankia entre 2015 y 2021, actualmente es el Director de Estrategia Informacional y Reporting Regulatorio de CaixaBank.

Desde el año 2019 ocupa el cargo de Presidente del Club de CDOs de España & Latam.

Es co-autor del libro *El Valor del Dato* (2020) (www.valordeldato.es), donde los autores proponen una metodología para valorar el impacto de la gestión de datos en el valor de una compañía.

Tiene una experiencia de más de 25 años en Banca, como Director Financiero y Director de Control de Gestión, es decir, siempre vinculado a la información y al dato.

Desde un punto de vista académico, es Profesor en la Escuela de Finanzas Aplicadas, Coordinador Académico de la Certificación en Gobierno y Estrategia del Dato del Instituto de Empresa, además de profesor en otras escuelas de negocio, conferenciante, articulista y escritor.

Asimismo, es miembro del Consejo Asesor de la Revista Índice y de la Revista de Actuarios de España, así como miembro del Consejo Académico de Ethosfera.

Retos de la estadística en una sociedad altamente datificada

Ana Carmen Saura Vinuesa

Subdirectora adjunta del Gabinete de la Presidencia del INE

El Gabinete de la Presidencia del Instituto Nacional de Estadística, en colaboración con la Universidad Internacional Menéndez Pelayo (UIMP), ha organizado el curso “Retos de la Estadística en una Sociedad Altamente Datificada” del 7 al 9 de julio de 2021, en formato híbrido en la sede de la UIMP del Palacio de la Magdalena de Santander.

El acceso a las nuevas fuentes de datos con fines estadísticos permitiría a las oficinas de estadística responder a los nuevos retos que se plantean y ofrecer a la sociedad la información que necesita, con la suficiente actualidad y desagregación

El objetivo del curso era exponer a los asistentes los retos y oportunidades asociados al uso del Big Data en el ámbito de la estadística, conocer la metodología y herramientas estadísticas necesarias para la integración del Big Data en el proceso de producción de las estadísticas oficiales, presentar casos concretos del uso de Big Data para la elaboración de estadísticas y debatir sobre el papel de la estadística oficial en este contexto.

La inauguración corrió a cargo de la Vicepresidenta Primera y Ministra de Asuntos Económicos y Transición Digital, que destacó la importancia de disponer de estadísticas y datos actualizados para la toma de decisiones.

Presentación del Presidente

Seguidamente, el Presidente del INE dirigió unas palabras de bienvenida a los asistentes, tanto a los que participaban presencialmente como a los alumnos *on-line*, y realizó una presentación acerca de “La Estadística oficial en España: oportunidades y nuevos retos”. Durante su exposición, señaló que vivimos en una sociedad datificada donde cada acción que realizamos genera datos, la mayoría de los cuales se recogen por entidades privadas. Estos datos, que se conocen como Big Data, ofrecen muchas oportunidades para las estadísticas oficiales como son el elevado nivel de detalle y la actualidad de la información.

Además, Rodríguez Póo destacó que en las difíciles circunstancias que estamos viviendo con motivo de la pandemia, la información estadística es más necesaria que nunca para tomar las decisiones adecuadas en cada momento. Se constata asimismo, la necesidad de disponer de datos de alta calidad, con gran puntualidad y elevado nivel de detalle, lo que supone un reto para las oficinas de estadística. El acceso a las nuevas fuentes de datos con fines estadísticos permitiría a las oficinas de estadística responder a los nuevos retos que se plantean y ofrecer a la sociedad la información que necesita, con la suficiente actualidad y desagregación.

Comentó las iniciativas europeas que se están llevando a cabo para potenciar el uso de fuentes privadas en la producción de estadísticas, entre las que se puede destacar la Hoja de Ruta de las Estadísticas Inteligentes Fiables o la posición común del Sistema Estadístico Europeo sobre la futura Ley Europea del Dato. A nivel nacional, mencionó algunas estadísticas experimentales del INE elaboradas a partir de nuevas fuentes de datos.

También señaló que en este nuevo ecosistema de datos se debería buscar el máximo beneficio para la sociedad, pero respetando al mismo tiempo los derechos e intereses de las

partes interesadas. Es decir, se debería buscar un equilibrio entre el derecho de los ciudadanos a su privacidad y el derecho a estar informados.

El Presidente del INE indicó que es fundamental garantizar un acceso sostenible a las nuevas fuentes de datos para las oficinas de estadística, pues estas gozan de independencia en el desarrollo, producción y difusión de datos, con lo que los datos privados se utilizarían para producir estadísticas de alta calidad.

En definitiva, garantizar el acceso sostenible a nuevas fuentes de datos para los productores de estadísticas oficiales empoderaría a los ciudadanos en su conjunto y redundaría en beneficio de toda la sociedad.

Por otro lado, mencionó que el uso de Big Data también conlleva importantes retos para los estadísticos que se pueden resumir en: garantizar un acceso sostenible a datos en poder de manos privadas, disponer de la capacidad tecnológica para procesar una ingente cantidad de datos y formular métodos estadísticos que infieran resultados significativos.

Retos y oportunidades en diversas áreas estadísticas

Durante el curso se expusieron los retos y oportunidades asociados al uso del Big Data para la producción de estadísticas en distintos ámbitos como la economía, la población, el turismo o los Indicadores de la Agenda 2030. Para ello, se contó con expertos de reconocido prestigio.

El Premio Nacional de Estadística y catedrático emérito de la Universidad Carlos III, Daniel Peña, presentó nuevos enfoques para el Análisis de Series temporales masivas.

La Directora del Departamento de Cuentas Nacionales del INE, M^a Antonia Martínez Luengo, expuso la situación actual y los retos de la estadística económica, como la medición de la globalización, desafío para el que Eurostat recomienda la creación de una Unidad de Grandes Empresas en los INE, encargada de proporcionar, a todos los dominios estadísticos, datos coherentes procedentes de los grupos empresariales multinacionales para la elaboración de estadísticas.



La sesión sobre “Los censos de población de 2021 y el futuro de las estadísticas socio-demográficas a partir de nuevas fuentes” fue impartida por Antonio Argüeso, Subdirector General de Estadísticas Sociodemográficas del INE. Este presentó las características generales del Censo 2021 que se basa exclusivamente en registros administrativos, lo cual permite publicar nuevas variables que no se habían recogido en censos anteriores por no sobrecargar el cuestionario. De manera complementaria, se va a llevar a cabo una Encuesta de Características

Esenciales de la Población y Viviendas dirigida al 1% de la población, cuyo objetivo es obtener información sobre las características sociodemográficas de la población, de las formas de convivencia y de las viviendas, que no se recoge en registros.

Por su parte, el responsable de análisis de Big Data en BBVA Research, Álvaro Ortiz Vidal Abarca, realizó una exposición sobre “Big Data y Economía en Tiempo Real y Alta Definición”. Comentó las razones estructurales para el desarrollo de Big Data en la actualidad como son la mejora de algoritmos, los desarrollos tecnológicos y la explosión de datos. Describió cómo calculaban algunas magnitudes de contabilidad nacional (consumo privado, inversión y salarios) a partir de la información de las transferencias bancarias, si bien las Cuentas Nacionales presentan las ventajas de equilibrio entre oferta y demanda, estabilidad, representación y desestacionalidad.

contabilidad nacional que desean medir. La firmante de este artículo presentó el uso de Big Data para la producción de los Indicadores de la Agenda 2030, prestando especial atención al cálculo de indicadores a partir de observaciones de la Tierra a nivel internacional.

Adicionalmente, se contó con la participación de académicos de reconocido prestigio en el ámbito estadístico. Wenceslao González, Catedrático de Estadística de la Universidad de Santiago de Compostela, realizó una exposición sobre “New Statistical Models in the Big Data Context”. Por su parte, Rosa Lillo, Catedrática de Estadística e Investigación Operativa de la Universidad Carlos III, expuso los principales aspectos de su estudio “Variable selection in high dimensional data. Applications in genetic and finance.” En tercer lugar, el profesor de Estadística e Investigación Operativa de la Universidade da Coruña, Ricardo Cao, compartió los resultados de sus investigaciones en el campo de “Nonparametric inference for big-but-biased data”.

En el acto de clausura el Presidente del INE señaló que para una institución como el INE, cuya misión es ofrecer estadísticas de alta calidad a la sociedad, este tipo de cursos son de gran utilidad para avanzar en la modernización de la producción de estadísticas y en la mejora de la eficiencia del servicio público que prestamos a toda la sociedad.

Este tipo de cursos son de gran utilidad para avanzar en la modernización de la producción de estadísticas y en la mejora de la eficiencia del servicio público que el INE presta a toda la sociedad

En la mesa redonda sobre “Nuevas estadísticas para la medición de fenómenos emergentes”, Belén González, Subdirectora General de Estadísticas de Turismo y Ciencia y Tecnología, describió las mejoras implantadas en las estadísticas de turismo como el Gasto realizado por los visitantes extranjeros en sus visitas a España obtenido de los pagos realizados a través de terminales de punto de Venta, así como retiradas de efectivo en cajeros. El representante del BBVA Research señaló que dedican bastantes recursos a limpiar, representar y validar las grandes cantidades de información de la que disponen en el Banco, así como a verificar que sus resultados son similares a la magnitud de

En el curso participaron 45 alumnos, de los cuales 30 lo hicieron presencialmente y 15 on-line. La organización del curso facilitó un chat para poder plantear cuestiones a los ponentes. Entre los participantes se encontraban expertos del INE y de los servicios de estadística de las comunidades autónomas y alumnos de universidades.

El desarrollo del curso ha sido muy satisfactorio en las actuales condiciones marcadas por la pandemia y los alumnos han podido familiarizarse con los nuevos retos que presenta el Big Data para la producción de estadísticas al mismo tiempo que estrechar lazos con los compañeros de otras unidades. ●

Participantes y desarrollo del curso

Los estudios de movilidad del INE a partir de la telefonía móvil (2019-2021)

Antonio Argüeso

Subdirector General de

Estadísticas Sociodemográficas (INE)

En abril de 2020, en pleno confinamiento de la población por la pandemia de Covid-19, el INE publicaba por primera vez datos sobre movilidad de la población a partir de la telefonía móvil. Esta publicación era consecuencia de un estudio que se había iniciado en 2019 y que, obviamente no se había diseñado para tal fin. En realidad, se trataba de un elemento más en la preparación de los censos de población y viviendas 2021.

Los censos, que se realizan cada 10 años, constituyen una operación estadística de mucha tradición (se remontan al siglo XVIII). El censo de 2011, como ya ocurriera en ediciones anteriores, introducía preguntas que trataban de ofrecer información tanto de los flujos diarios entre zonas (número de personas que se desplazan en un día laborable entre el lugar donde residen y el lugar donde estudian o trabajan) como de determinar la población que, residiendo en un municipio, tiene vinculación con otro porque pasa en él al menos 15 días al año (población vinculada).

El censo de 2021 se enfrentaba a un problema importante a la hora de actualizar esos datos de movilidad, al basarse completamente en el tratamiento de registros administrativos, sin enviar cuestionarios a los hogares. La información sobre movilidad cotidiana, es decir, las matrices de flujos origen-destino (residencia-trabajo) y las de población vinculada, no se pueden obtener, en principio, de registros administrativos. Con el fin de evitar la pérdida de esta información, muy demandada por los usuarios, en torno a 2015, se empezó a explorar la posibilidad de utilizar la telefonía móvil como fuente alternativa.

Los datos de movilidad son, por naturaleza, muy detallados en el territorio. No importa conocer movilidad por CC. AA. sino cuantificar y caracterizar los desplazamientos entre municipios e incluso dentro de ellos y la información de la telefonía móvil parecía ideal para estos fines.

El INE mantuvo diversas reuniones con los operadores de telefonía móvil (OTM) para analizar la forma de disponer de esta información y la metodología a emplear. Se llegó finalmente a un punto de encuentro con ellos que cristaliza en el estudio de movilidad de 2019, ahora llamado EM-1.

El estudio original EM-1 se puso en marcha a finales de 2019, con un objetivo muy específico: se trataba de construir matrices de movilidad cotidiana (residencia-trabajo) que sirviera de fuente para delimitar la forma de las áreas de influencia de las ciudades, medida esta a partir de observar cuantas personas de municipios cercanos se desplazan cotidianamente a la ciudad. El trabajo se basaba en datos agregados (recuentos totales de flujos origen-destino) de los tres principales OTM del país.

Para medir la movilidad cotidiana se tomaba como referencia una semana concreta de noviembre de 2019, que se suponía “normal” a estos efectos.

El ámbito poblacional está constituido por los teléfonos móviles de población residente en España de los tres principales operadores, es decir, de teléfonos móviles del plan nacional de numeración telefónica. Se excluyen teléfonos de numeración extranjera, que operan en España en *roaming*, normalmente en manos de turistas.

La unidad mínima de observación geográfica es el “**área de movilidad**”. Esta unidad geográfica se creó específicamente para este proyecto. Se buscaba tener divisiones administrativas con un mínimo de 5.000 habitantes aunque algunas pocas áreas quedan muy ligeramente por debajo de ese umbral. Con esto se

pretendía garantizar que el número de usuarios de cualquier OTM en cada área fuera suficiente para el análisis de movilidad entre áreas. Las áreas se construyeron así:

- ▶ Dentro de cada provincia, los municipios menores de 5.000 habitantes se agrupan geográficamente para alcanzar al menos ese umbral de población.
- ▶ Cada uno de los municipios de entre 5.000 y 50.000 habitantes constituye un área.
- ▶ Los municipios de más de 50.000 habitantes se pueden desagregar en distritos o barrios, siempre con un umbral superior a esos 5.000 habitantes (unas 1.150 áreas para toda España). En el caso más extremo, Madrid, se divide el municipio en los 128 barrios que la componen dado que todos superan el umbral de 5.000 habitantes contemplado.

El número exacto de áreas formado finalmente fue de 3.214, con un tamaño medio de unas 15.000 personas, o de unos 12.000 teléfonos sumando los tres operadores.

El proyecto original EM-1 incluía concretamente tres de matrices de movilidad:

La matriz de movilidad cotidiana es de origen-destino, formada por N (3.214) áreas $\times$ N áreas. Se toman cuatro días consecutivos (lunes a jueves) de una semana de noviembre de 2019. Para cada día observado, en cada elemento “fila A-columna B” de esta matriz figuraría el número de terminales que se desplazan del área de residencia (A) al área de destino (B). Se calcula una sola matriz como promedio de los cuatro días.

Para la construcción de esta matriz, en primer lugar se debe determinar el área de residencia de cada uno de los **teléfonos móviles de cada OTM**: aquella en donde el teléfono móvil se encuentra durante más tiempo entre las 22:00 horas del día anterior y las 06:00 horas durante el día observado.

Asignada para cada teléfono móvil el área de residencia, se intenta determinar el área de destino cotidiano: donde se encuentra durante más tiempo el terminal en el horario de 10:00 a 16:00 del día observado.

La segunda matriz, de tamaño $N \times 12$, simplemente contaba cuantos terminales se encontraban en cada una de las 3.214 áreas en doce momentos distintos de cada día observado. Era una forma complementaria a la matriz 1 para conocer la movilidad cotidiana. Se toman las mismas fechas y de nuevo se calcula el promedio.

La tercera matriz también tiene tamaño $N \times N$, como la primera, pero se dedica a flujos “estacionales” y se obtiene una para cada uno de cuatro días de 2019 elegidos que permiten medir cómo se localiza la población fuera de su lugar habitual de residencia en fechas señaladas (20 de julio, 15 de agosto, 24 de noviembre por suponer un fin de semana normal de un mes valle y el día de Navidad, 25 de diciembre).

Para cada día elegido, en la casilla fila A-columna B figura el número de personas que, residiendo en el área A, pernoctan en el área B durante ese día, pero para ello se debe determinar esa área de residencia con información histórica: aquella en donde el teléfono móvil se encuentra durante más tiempo en horario nocturno (entre las 00:01 y las 06:00 horas) de los últimos dos o tres meses (el número de días es distinto según el operador pero como mínimo se toman 60 días). Es decir, para determinar el área de residencia es necesaria información de los meses anteriores.

El área donde “pernocta” cada teléfono móvil es donde se le encuentra más frecuentemente durante el periodo desde las 22:00 horas del día anterior a las 06:00 en esa fecha. Este método proporciona un área de pernoctación a todos los teléfonos presentes en territorio español en esas fechas.

El INE recibía de los operadores las matrices elaboradas ya agregadas. No se recibe en ningún momento información individual de ningún tipo sino solo de matrices de resultados finales elaborados por cada uno de los OTM que ofrecen recuentos a nivel de área de movilidad. Además, siempre que el número de observaciones sea inferior a 10 o 15 en un área, para un operador dado, el resultado vendrá censurado en origen. Por tanto, solo se podrían ofrecer datos de movilidad origen-destino para áreas en las cuales al menos uno de los tres OTM cuente más de ese umbral.

Estaba previsto publicar los resultados de este proyecto a finales de abril de 2020, pero a mediados de marzo todo se torció. La llegada

del brote de Covid-19 produjo un confinamiento total de la población y se hizo de pronto necesario disponer de información precisa y actual sobre la movilidad de la población para conocer el impacto de las medidas de confinamiento bajo el estado de alarma. Esto hizo que se planteara de forma urgente reconducir el estudio de movilidad original, que estaba a punto de concluir, y se pusiera en marcha el llamado **estudio EM-2**, que medía exclusivamente movilidad cotidiana.

Los operadores disponían de un algoritmo para ofrecer datos diarios de movilidad cotidiana porque estaban trabajando en el EM-1. Se les pidió usar ese mismo algoritmo, no para unos días concretos de noviembre de 2019 sino durante el confinamiento. Por las circunstancias excepcionales en que se desarrolló este estudio los operadores de telefonía entregaron información casi diaria durante el estado de alarma (16 de marzo a 20 de junio).

Una vez terminado el primer estado de alarma, ya en junio de 2020, se publicó el estudio de movilidad original EM-1. Y dado que la pandemia se ha mantenido con mayor o menor intensidad, se han seguido publicando estas matrices de movilidad con frecuencia semanal durante 2020 (**estudio EM-3**) y 2021 (**estudio EM-4**).

En conjunto, sumando todos los estudios disponemos de una serie homogénea de matrices de movilidad cotidiana, al menos dos días por semana, que abarcan el periodo 16 de marzo de 2020 - 30 de diciembre de 2021 y tienen como referencia de la “normalidad” una semana de noviembre de 2019.

Es importante destacar que la fuente empleada es la localización de los teléfonos móviles a partir de las antenas de telefonía por lo que no se puede determinar con total precisión la ubicación de un teléfono. El margen de error es desde decenas de metros en zonas urbanas a centenares de metros, incluso kilómetros, en zonas rurales. Así, los movimientos que se observan entre áreas adyacentes deben tomarse con cautela porque pueden estar provocados por el error que se comete al fijar la posición de cada terminal, lo que puede dar lugar a que se cuente un terminal que está inmóvil en dos áreas adyacentes diferentes en distintos momentos.

Con todo ello, se puede observar como el porcentaje de personas que en horario de trabajo se

ausentaba de su área de residencia, alcanzó un valor en noviembre de 2019 que no se ha vuelto a alcanzar en casi ninguna área de movilidad durante todo 2020 y 2021. La movilidad se vio lógicamente muy afectada durante los meses del confinamiento (marzo-junio de 2020) pero aún a finales de 2021 sigue en niveles inferiores. Probablemente las nuevas pautas de teletrabajo ayuden a explicar este comportamiento.

Pero, además, se dispone de la matriz 3 (de movilidad estacional) para cuatro días concretos de 2019, 2020 y 2021, que nos permite también conocer como se ha comportado el turismo nacional, algo muy demandado también durante la pandemia. Las fechas de 2020 son 18 de julio, 15 de agosto, 22 de noviembre y 25 de diciembre y las de 2021 fueron 17 de julio, 15 de agosto, 21 de noviembre y 25 de diciembre.

De nuevo se observa como, en general, los flujos mayores se dan en las fechas de 2019 y los más bajos de observan en 2020. En 2021 se ha recuperado solo parcialmente la movilidad perdida.

Más allá del análisis general de la movilidad en situación de pandemia, estos estudios ofrecen una radiografía muy precisa de los movimientos origen-destino en España, a unos niveles mucho más detallados que los que ofrecían los censos de 2011. Esto permite dibujar las áreas funcionales en torno a las ciudades y ofrece información para innumerables aproximaciones de análisis tanto de movilidad cotidiana como estacional. Es pronto para extraer conclusiones en ese sentido porque la información a analizar es mucha. Todos los resultados de estos estudios están a disposición de los usuarios en la sección de Estadísticas Experimentales¹ del INE. Además de las matrices de datos que se han descrito, se ofrece un visor de mapas que facilita el análisis de la información.

Lo que sí se puede ya afirmar es que para el INE este estudio ha constituido una nueva forma de afrontar las estadísticas en el ámbito sociodemográfico, apostando por el uso de nuevas fuentes de datos masivos (Big Data) y todo ello bajo un nuevo paradigma de colaboración con los propietarios de estas fuentes de datos que se ha mostrado satisfactorio. ●

1 <https://www.ine.es/experimental/experimental.htm>

El uso del Big Data para la producción estadística. La experiencia de BBVA Research

Fernando Bolívar García

Sirenia Vázquez Báez

BBVA Research

Introducción

La era digital y la innovación tecnológica están revolucionando el paradigma industrial, incluyendo el sector financiero. El flujo de información procedente de la actividad cotidiana de los individuos genera un manantial de datos que ofrece la oportunidad de ampliar nuestra comprensión sobre el comportamiento de los individuos, la economía y la sociedad. Cuando leemos las noticias en internet, buscamos algún término o palabra en algún buscador, participamos en redes sociales, hacemos una compra o transacción financiera de manera electrónica, añadimos información a este universo de datos. Una vez anonimizados y procesados, estos datos masivos plantean nuevos retos y ofrecen soluciones de valor e innovadoras a preguntas relevantes que no podían ser respondidas por datos tradicionales. Esto es lo que denominamos en BBVA Research como “Economía en Tiempo Real y Alta Definición”.

Volumen, Velocidad y Variedad

Las características principales de **volumen, velocidad y variedad**, son claves en la definición y en el entendimiento del Big Data. El volumen hace referencia a la cantidad de información recopilada no solo por las organizaciones, sino toda la generada por redes sociales, búsquedas online, datos geolocalizados, interacciones con los clientes, datos recogidos durante los procesos de negocio, etc. En definitiva, un tamaño inmenso de datos estructurado y no estructurado

cuya gestión y análisis es todo un reto y brinda la oportunidad de enriquecer las bases de datos tradicionales con datos de alta dimensión.

La **velocidad**, se entiende como la rapidez con que los datos son generados, almacenados y procesados. Cada vez más, el tiempo es una variable fundamental para la comprensión y anticipación de eventos en el entorno económico y financiero. La alta frecuencia se convierte en una importante herramienta de diagnóstico y alerta temprana para detectar rápidamente cambios de comportamiento, especialmente en el cambiante panorama actual.

Por último, y no menos importante, el concepto de **variedad** recoge la amplia gama de información disponible en nuestro entorno. Los datos clásicos en formato fila-columna han dejado de ser los protagonistas, actualmente los datos desestructurados como el texto, imágenes, videos o audios están ganando terreno. Esta diversidad supone un incremento cuantitativo y cualitativo de la información disponible tanto para el análisis social, económico o financiero.

Estos tres pilares básicos del Big Data requieren además de ciertos atributos que trasladen su potencial en utilidad para el analista. Entre ellos, podemos añadir conceptos como los de **veracidad, valor y visualización**. A medida que los datos crecen, es necesario controlar la integridad y fiabilidad de la información recogida ya que de esto dependerá principalmente el acierto de las decisiones. Asimismo, la abundancia de datos puede difuminar el propósito de cualquier estrategia Big Data, por lo que es primordial mantener el foco de convertir la información en conocimiento como factor diferencial de manera rentable y a la vez eficiente. Finalmente, uno de los puntos clave y definitivos es el modo en el que la información obtenida es presentada y transmitida ya que de

esto dependerá mayoritariamente el alcance y el valor agregado generado.

Ventajas y desafíos

El crecimiento explosivo de datos recogidos y el avance tecnológico imparable brindan un sinfín de oportunidades que conllevan una serie de ventajas competitivas, pero no están ajenos a inconvenientes.

A nivel agregado, dicha información puede transformarse en indicadores que midan sentimiento, comportamiento o actividad económica. Dichos indicadores permiten conocer en tiempo real lo que está ocurriendo, lo que proporciona un margen extraordinario a las autoridades, empresas e individuos para la toma de decisiones. Además, nos indican con elevado detalle los grupos de individuos que necesitan una respuesta individualizada (Políticas Inteligentes). La crisis sanitaria del Covid-19 es un buen ejemplo de ello, pues la información generada a gran escala fue de vital importancia para, entre otras cosas, observar en tiempo real el impacto de las medidas de contención de la pandemia sobre hogares, empresas y la economía en su conjunto. De hecho, ha sido durante esta crisis del Covid que se ha desencadenado una mayor demanda de información en tiempo real y alta granularidad, ya sea para contar con indicadores alternativos o generar modelos de alta frecuencia.

Además de la inmediatez, a un coste menor que otros procedimientos se puede obtener una imagen con alta granularidad, en alta definición, en términos geográficos, sectoriales o temporales. Esto abre la posibilidad de plantearse preguntas que hasta ahora era imposible responder empíricamente. Un ejemplo de ello es la información generada a través de las transacciones financieras, la cual ha abierto nuevos caminos para entender mejor el comportamiento de hogares, empresas e individuos. Como mencionan Baker y Kueng (2021), “la granularidad de los datos y el hecho de que a menudo están emparejados con otras cuentas financieras de las que es titular el individuo, permiten formular preguntas sobre la totalidad del balance o los flujos financieros

del hogar. Los investigadores pueden controlar mejor por posibles variables de confusión y rechazar explicaciones alternativas para una pregunta determinada”[1].

Por otro lado, la infraestructura tecnológica actual está democratizando el dato permitiendo la colaboración entre distintas disciplinas, la combinación de fuentes y de formatos consiguiendo un conocimiento más profundo del mercado. Estas ventajas de aplicación empiezan a experimentarse no solo entre organizaciones privadas sino por organismos públicos como institutos nacionales de estadística, bancos centrales y centros académicos de investigación pues complementan y enriquecen la información que recaban con métodos más “tradicionales”.

El crecimiento explosivo de datos recogidos y el avance tecnológico imparable brindan un sinfín de oportunidades que conllevan una serie de ventajas competitivas, pero no están ajenos a inconvenientes

Sin embargo, como toda tecnología, el Big Data conlleva una serie de desafíos que deben ser tenidos en cuenta. Uno de los principales retos es la calidad o la escasez de los datos, así como el muestreo y la representatividad de los mismos. Otros factores relevantes a tener en cuenta, además del problema de la dimensionalidad, son la multicolinealidad y el sesgo intrínseco, ya que los grandes volúmenes de datos no eluden directamente los problemas de inferencia habituales de estadística. Por último, trabajar con datos diarios hace que el problema de desestacionalización de los datos sea más complejo. Existen también desafíos “no técnicos” como garantizar en todo momento la privacidad y la confidencialidad de los datos que

1 Baker, S. R., & Kueng, L. (2021). *Household Financial Transaction Data* (No. w29027). National Bureau of Economic Research.

juegan un papel fundamental en la seguridad jurídica de los individuos. Finalmente, existen desafíos de seguridad, dada la exposición de los datos a potenciales ciberataques.

Con todas estas y otras ventajas y desventajas, ¿cómo podemos los economistas beneficiarnos de la revolución del Big Data?, ¿abre el Big Data nuevos temas de investigación para los economistas financieros? o ¿nos permiten responder preguntas tradicionales de forma novedosa y más reveladora?

Aplicaciones BBVA Research

En BBVA Research llevamos varios años haciendo uso de la información y herramientas disponibles para ilustrar la utilidad del Big Data dentro del campo social, económico y financiero. El potencial es grande, pero las dificultades no son menores, pues para hacerlos útiles, los datos deben pasar por distintas etapas, con retos en cada una de ellas y siempre preservando la anonimidad de los individuos que realizan estas operaciones.

Nada ha cambiado aquí, el principal reto al que se enfrenta el investigador es el de plantearse la pregunta adecuada

El primer reto al que se enfrenta BBVA Research al trabajar con Big Data es asegurar una buena materia prima. En esencia, esto supone mantener una infraestructura adecuada para ingestar, gestionar, procesar y anonimizar la información. Este proceso es complejo y costoso aunque una vez garantizado los resultados superan con creces el coste. Una vez que la información está disponible, el siguiente reto es identificar exactamente el dato que se está buscando, procesarlo y extraerlo, ya sea de fuentes

internas o externas. Nada ha cambiado aquí, el principal reto al que se enfrenta el investigador es el de plantearse la pregunta adecuada.

Una vez que la pregunta y los datos están disponibles, comenzamos la fase de procesamiento, tratamiento y validación. La duración de esta etapa dependerá de la calidad de la información (existencia de nulos, dispersión o baja calidad del dato). Hay que tener en cuenta que en muchos casos no existe un horizonte temporal lo suficientemente largo, ya sea para aplicar técnicas estadísticas convencionales a una serie de tiempo generada con Big Data o para incorporar estos datos a los modelos de pronósticos que suelen utilizar horizontes de tiempo amplios. Así, para lidiar con ello es necesario aplicar no solo los métodos estadísticos tradicionales sino también nuevas metodologías, como por ejemplo análisis de lenguaje natural, regularización, reducción de la dimensionalidad, modelos de aprendizaje automático, entre otros. La última etapa del proceso corresponde a la narrativa y visualización que son fundamentales para explicar lo que hacemos y por qué lo hacemos.

El resultado del proceso mencionado son algunos de los indicadores en tiempo real y alta definición como los de consumo e inversión en tiempo real. En el primer caso, utilizando las transacciones financieras estudiamos el comportamiento agregado de los consumidores gracias a la información agregada y anonimizada de los pagos con tarjeta BBVA y tarjetas ajenas en terminales de punto de venta españoles. Concretamente, dicha información nos ha permitido analizar el impacto del gasto en España durante la actual pandemia de la Covid-19, así como las políticas aplicadas para controlarlo a nivel diario. Los metadatos de las transacciones también nos permiten estudiar la variación de los comportamientos en función de la geografía, los sectores y el modo de venta (por ejemplo, *online/offline*). Llegamos a la conclusión de que estos datos son capaces de captar las tendencias relevantes de gasto y, lo que es más importante, lo hacen en tiempo real.

En el segundo caso, podemos utilizar las transacciones entre empresas para desarrollar un índice para medir la inversión en tiempo real y alta definición. Este indicador es innovador pues utili-

za datos agregados de transferencias financieras empresa-empresa e individuo-empresa. En particular, se utilizan aquellas transferencias en las cuales las empresas vendedoras son productoras de bienes y servicios de inversión. Dado que la información está desagregada por sector de actividad económica y está geolocalizada, podemos conocer con detalle el tipo de inversión (Construcción, Maquinaria y Equipo o Bienes Intangibles) y la región donde esta se produce.

Estos indicadores tienen una alta correlación (de 0.60 o superior) con los componentes de la inversión de cuentas nacionales y, dada la alta frecuencia de los datos utilizados, es posible conocer la evolución de la inversión con anticipación. Gracias a ello, se pudo observar el impacto que tuvo sobre la inversión la interrupción de las cadenas globales de valor generada por la pandemia. Del mismo modo, conforme se ha normalizado la situación, se observa una recuperación.

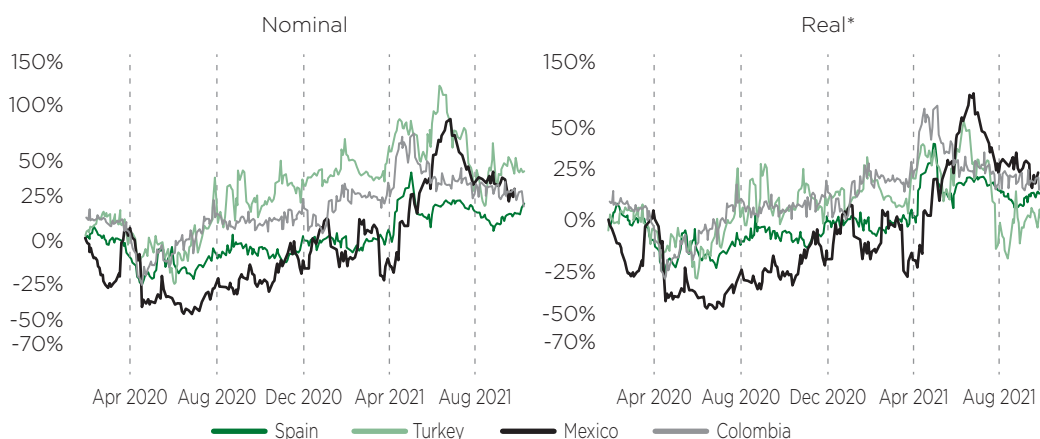
Una aplicación especial de estos indicadores es el caso de la inversión en bienes intangibles, para los cuales existe escasa información pública para medir su composición y comportamiento. Este componente puede tener, para el caso de España, un gran potencial para monitorear el efecto

de políticas de sostenibilidad y de asignación de recursos para la inversión, como es el caso de los Fondos de Recuperación para Europa.

Pero también hemos desarrollado estudios temáticos o focalizados en un tipo de consumo. Como aplicación derivada del gasto de consumo, podemos estimar el gasto turístico nacional en España mediante el gasto realizado con tarjetas BBVA fuera de la provincia habitual de residencia. Esto nos permite monitorizar los flujos de gasto turístico dentro de España a nivel agregado, por comunidades autónomas y por provincias, de forma diaria. De nuevo, la alta frecuencia y granularidad de los datos de tarjeta permiten avanzar y complementar las cifras oficiales de turismo nacional, que como se está demostrando, está siendo clave en el proceso de recuperación. Del mismo modo, conseguimos detectar los efectos permanentes y temporales de la crisis en las preferencias de los turistas españoles durante las limitaciones de movilidad, así como la llegada de turistas extranjeros.

Además de los datos propios del banco, utilizamos un conjunto amplio de información no estructurada que nos permite entender mejor el entorno mediante la aplicación de técnicas

FIGURA 1. INDICADORES DE INVERSIÓN BIG DATA: INVERSIÓN TOTAL.
(% interanual acumulado 28 días)



Indicadores de inversión estimados a partir de las transacciones monetarias entre empresas y hogares a las empresas nacionales que producen los activos fijos definidos por el código NACE. "Series reales deflactadas por los precios de producción, excepto Perú, cuyo deflactor es el índice de precios mayoristas y España, que utiliza el deflactor implícito del componente de la inversión del PIB por el lado de la demanda. Forthcoming (Carvalho et al. 2021).

Fuente: BBVA Research.

Como aplicación derivada del gasto de consumo, podemos estimar el gasto turístico nacional en España mediante el gasto realizado con tarjetas BBVA fuera de la provincia habitual de residencia

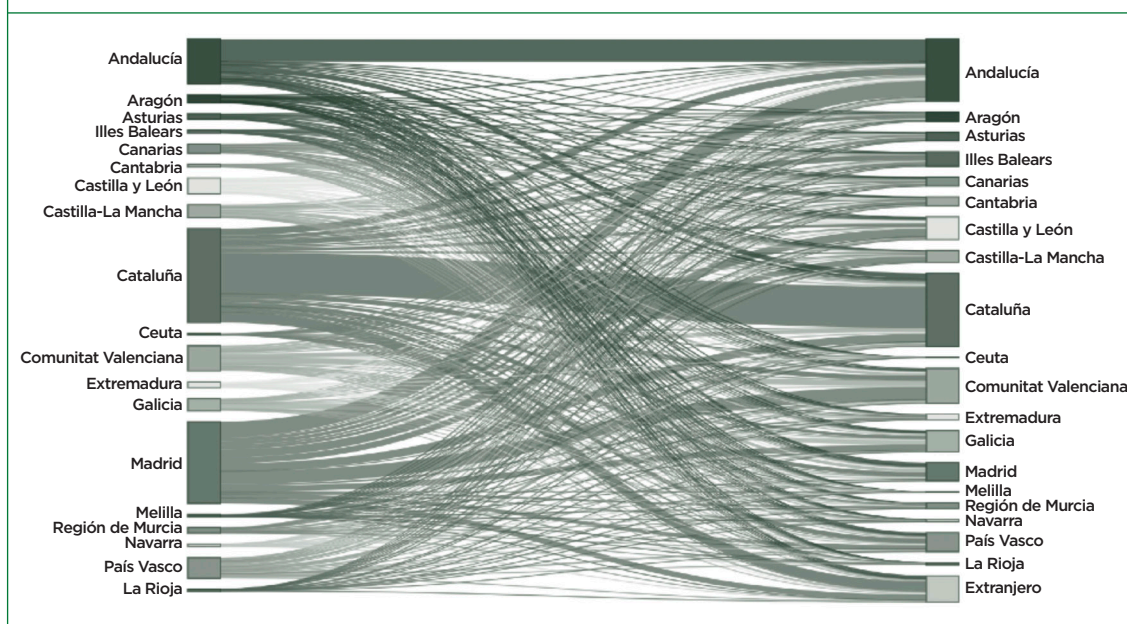
de analítica avanzada. Para entender la evolución y percepción de algunos eventos sociales o geopolíticos empleamos información procedente de redes sociales (Twitter), medios de comunicación a escala mundial (GDELT) o el volumen de búsquedas realizadas en internet (Google Trends). Estos datos generados por la interacción social nos han permitido estudiar temas económicos de gran calado como el estancamiento del comercio mundial y el retroceso de la globalización, así como entender el marco de la sostenibilidad o el universo Fintech. Del mismo modo, hacemos uso de la informa-

ción publicada en la web, por ejemplo, el análisis de la narrativa de los bancos centrales en sus comunicados de prensa y actas de política monetaria nos ayuda a entender las estrategias de comunicación, y a anticipar sus decisiones.

Reflexión

Los indicadores mencionados son solo un ejemplo del uso de datos masivos de transacciones financieras. BBVA Research ya ha trabajado en otras aplicaciones, como la medición de impacto de la política monetaria en tiempo real, el desarrollo de indicadores para medir la evolución del sector exterior y de los distintos sectores de actividad económica, o la construcción de tablas insumo producto en alta frecuencia. Asimismo, desde BBVA Research entendemos estas fuentes como un activo clave para tener respuestas rápidas a eventos de interés dada su alta frecuencia, así como para comprender la percepción social de los mismos, puesto que recogen la opinión generada por la sociedad, actores particulares, empresas e instituciones. ●

FIGURA 2. FLUJOS DE GASTO PRESENCIAL CON TARJETA REALIZADO FUERA DE LA PROVINCIA HABITUAL DE RESIDENCIA DE COMUNIDAD ORIGEN A DESTINO (Porcentaje de gasto sobre el total, Julio-Agosto 2021)



Fuente: BBVA Research a partir de BBVA.

Prevenir los accidentes antes de que sucedan con telemática

Montserrat Guillén Estany

Universitat de Barcelona

Nos encontramos en una nueva era de la seguridad en la conducción. Las tecnologías que nos acercan al vehículo autónomo unidas a la telemática, entendida como una capacidad sin precedentes de registrar información sobre objetos en movimiento, están transformando cómo nos sentamos al volante. Concebidos en las últimas décadas, los asistentes de conducción se desarrollaron de la mano de la ingeniería, proporcionando innovaciones que abarcan desde el facilitar el aparcamiento hasta el control automático de velocidad. Sin embargo, la telemetría se basa en los datos y ha sido directamente promovida por las aseguradoras, pero se está convirtiendo en una oportunidad para la prevención de accidentes.

Mediante sensores a bordo inerciales o integrados en los teléfonos móviles, la telemática permite acceder a los datos sobre un vehículo cuando se desplaza, midiendo velocidad, aceleración, orientación, ángulo de giro o distancia al vehículo precedente. A dichos factores se le puede añadir información del motor, como las revoluciones por minuto, u otra como el uso activo del móvil mientras se conduce. Una pregunta que solemos hacernos es si conocer la posición exacta del vehículo puede tener cierto interés. Si lo tiene, por ejemplo en la gestión de flotas o la optimización de redes de transporte, pero para el grueso de conductores es un dato sensible y privado por lo que muchos se declaran reacios a que se graben sus posiciones exactas. En esta materia, los contrastes entre países son evidentes, mientras algunas compañías estadounidenses reconocen abiertamente usar, e incluso comercializar, la ubicación para trazar hábitos de consumo, en Europa la resistencia a la intrusión es mucho mayor. En España, los datos telemáticos de conducción no graban coordenadas, pero sí se suelen filtrar según el tipo de vía por el que se circula, básicamente para identificar excesos o promedios de velocidad, sin guardar localización.

La riqueza en volumen y velocidad de recogida de información en telemetría supone un reto para los científicos de datos y una oportunidad para mejorar nuestra seguridad. Permite disponer de mediciones a tiempo real en intervalos inferiores a un segundo, pero exige almacenar millones de cadenas de observación que, como es fácil de entender, requieren una analítica de compresión que maximice la información en el mínimo espacio.

La riqueza en volumen y velocidad de recogida de información en telemetría supone un reto para los científicos de datos y una oportunidad para mejorar nuestra seguridad

¿Qué datos telemáticos se almacenan en la práctica?

Los analistas de riesgos en seguros suelen extraer indicadores de comportamiento de los conductores a un nivel de agregación temporal relativamente amplio. Tradicionalmente, se utilizan estadísticos de resumen anual, ya que el precio del seguro suele calcularse por anualidades. La variable más utilizada en los seguros de automóviles telemáticos es el total de kilómetros recorridos en un año. Por consiguiente, aunque la telemetría pueda llegar a analizar qué ocurre cada momento, en realidad, en la práctica se utilizan promedios o indicadores acumulados, para clasificar a los conductores. En muchos países entre los que se encuentra España ya es posible contratar un seguro de pago por kilómetro (que en inglés

se conoce como UBI, *Usage-based insurance*), una modalidad muy interesante para aquellos conductores que efectúan un reducido número de desplazamientos y que por lo tanto, prefieren pagar una tarifa básica, generalmente muy económica, más un coste por cada kilómetro recorrido, en lugar de una tarifa anual. Se estima que este tipo de conductores pueden reducir su factura hasta un 40% con una modalidad de seguro UBI.

*Una pequeña parte del flujo de datos,
y especialmente el segmento
inmediatamente previo a un accidente,
puede constituir un indicador de
advertencia para prever que un
accidente esté a punto de ocurrir*

¿Cómo se relacionan los datos con los accidentes?

Con la mayor disponibilidad de información telemática, los desafíos para el análisis de riesgos de accidentes son distintos a los tradicionales. El análisis de riesgos consiste en descubrir qué parte de los datos puede informar y anticipar la ocurrencia de un siniestro, y cómo encontrarla. En la telemetría de conducción, los datos son muy poco informativos o repetitivos, por lo que no se hallan fácilmente indicios de comportamiento anómalo. Sin embargo, una pequeña parte del flujo de datos, y especialmente el segmento inmediatamente previo a un accidente, puede constituir un indicador de advertencia para prever que un accidente esté a punto de ocurrir. Esos pocos segundos pueden ser esenciales para salvar vidas y mejorar la protección.

Por desgracia, el estudio de accidentes de tráfico y los resultados científicos obtenidos respecto al análisis de la secuencia inmediatamente anterior a un accidente no ayudan demasiado a nivel general, excepto para el caso en el que se detecta somnolencia en el movimiento de los ojos. Aunque con un éxito mo-

derado, a nivel de experimental se ha podido verificar que determinados conductores tienen más capacidad de reacción que otros, y su propia destreza da lugar a una accidentabilidad inferior al resto.

Para los conductores con telemetría que circulan en España, algunas decenas de miles, se han efectuado diversos análisis, y lo que sí se ha concluido es que después de un accidente, y muy especialmente si este es grave, se produce un cambio de comportamiento muy significativo en cuanto a la velocidad media: se conduce más despacio. Sin embargo, este resultado, que era completamente esperable porque un conductor que ha sufrido un accidente suele mostrarse más prudente que otro que no hace sufrido ningún percance, se produce a posteriori y, por lo tanto, no sirve como predictor.

¿Qué conocimiento se extrae de los datos?

Las aseguradoras están convencidas de que la recopilación de datos telemáticos proporciona información valiosa y útil sobre sus clientes, aunque todavía se desconoce el retorno económico global asociado al coste de instalar la inteligencia a bordo y gestionar el Big Data que esta es capaz de generar. Estudios recientes [Gao et al. (2019), Guillén et al. (2019), Würthrich (2017)] confirman que los datos telemáticos sustituyen perfectamente a los factores tradicionales de tarificación, es decir, los indicadores que tradicionalmente se registraban al suscribir un contrato de seguro para establecer perfiles de riesgo y precios (antigüedad de carné, tipo de vehículo, etc.). Además, algunos estudios [Barry y Charpentier (2020), Geyer et al. (2020)] han descubierto que en solo unas pocas semanas de monitorización con dispositivos telemáticos, es relativamente sencillo clasificar a los conductores según patrones.

Por los anteriores motivos, aunque con mucha discreción, se espera que las aseguradoras vayan generando paulatinamente indicadores de estilo de conducción con el objetivo de trasladarlos a tarifas de los seguros de automóviles más personalizadas. La competencia entre aseguradoras significa que cada una de ellas utilizará sus propias herramientas y, por lo ge-

neral, la capacidad predictiva de dichos estos factores de precios basados en telemetría no se revelará a los competidores. También es posible que se requiriera algún tipo de regulación para esta “competencia algorítmica” en la tarificación [Cevolini y Esposito (2020)].

¿Cómo puede aplicarse la telemetría a la reducción de la siniestralidad?

La especificidad de la analítica de riesgos reside en el hecho de que no solo permite mejorar la evaluación de siniestros en sí, sino que la abundancia de datos perfila a los clientes más allá del “cliente medio”. Dicha metodología contribuye a redefinir procedimientos de fijación de precios clásicos utilizados por las compañías de seguros, especialmente cuando los datos que tienen a su disposición son datos de comportamiento.

Estudios recientes [Guillen et al. (2021)] proponen que el seguro se componga de dos bloques: i) una prima base dependiendo de características generales (factores tradicionales como la potencia del vehículo o la zona de conducción), y ii) una prima variable en función de factores específicos como la distancia recorrida y el estilo de conducción. Guillen et al. (2020)

son los primeros en apuntar al análisis de los “cuasi accidentes” (en inglés, near-misses), es decir, eventos como frenadas o acelerones que pueden indicar la presencia de un peligro inminente, incluso si este no fue un siniestro, sino que pudo evitarse. Algunos resultados más recientes efectuados este mismo año apuntan a que los excesos de velocidad detectados por la telemetría podrían llegar a penalizarse mediante una sobreprima o sanción por cada kilómetro recorrido con una velocidad superior a la permitida. Alternativamente, se podrían incrementar precios si el patrón de velocidad medio no se adecúa al trayecto y densidad de tráfico.

Para la seguridad vial, la telemetría ya incorporada en los automóviles es una oportunidad de trasladar el radar a cada vehículo y podría llegar a materializarse en una alianza en la que compañías de seguros y autoridades trabajaran en un mismo objetivo. Para las compañías, hay que buscar los incentivos y la rentabilidad de un sector que está cambiando, para prestar un servicio de prevención y seguridad a los conductores informándoles sobre cómo mejoran su estilo al volante. Para la seguridad vial, sería un eslabón más en el objetivo “cero accidentes”.

Un consejo final: dejarse monitorizar. Creer en la tecnología y en su capacidad para evitar accidentes. ●

BIBLIOGRAFÍA

- Barry, L. y A. Charpentier, 2020, “Personalization as a promise: can big data change the practice of insurance?” *Big Data & Society* 7:1.
- Cevolini, A. y E. Esposito, 2020, “From pool to profile: social consequences of algorithmic prediction in insurance” *Big Data & Society* 7:2, 1-11.
- Gao, G., S. Meng y M. V. Wüthrich, 2019, “Claims frequency modeling using telematics car driving data” *Scandinavian Actuarial Journal* 2019:2, 143-162.
- Geyer, A., D. Kreamslehner y A. Mürmann, 2020, “Asymmetric information in automobile insurance: evidence from driving behaviour” *Journal of Risk and Insurance* 87:4, 969-995.
- Guillen, M. y A. Cevolini (2021) “Using risk analytics to prevent accidents before they occur – the future of insurance” *Journal of Financial Transformation*, en prensa.
- Guillen, M., J. P. Nielsen, M. Ayuso y A. M. Pérez-Marín, 2019, “The use of telematics devices to improve automobile insurance rates” *Risk Analysis* 39:3, 662-672.
- Guillen, M., J. P. Nielsen y A. M. Pérez-Marín, 2021, “Near-miss telematics in motor insurance” *Journal of Risk and Insurance*, 88, 569-589.
- Guillen, M., J. P. Nielsen, A. M. Pérez-Marín y V. Elpidorou, 2020, “Can automobile insurance telematics predict the risk of near-miss events?” *North American Actuarial Journal*, 24:1, 141-152.
- Wüthrich, M. V., 2017, “Covariate selection from telematics car driving data” *European Actuarial Journal* 7:1, 89-108.

ENCUESTA DE CONDICIONES DE VIDA (ECV). AÑO 2020

Descarga gratuita a través de la web del INE

La Encuesta de Condiciones de Vida (ECV) es una operación estadística anual dirigida a hogares que se realiza en todos los países de la Unión Europea. Se trata de una estadística armonizada que está respaldada por el Reglamento (CE) N° 1177/2003 del Parlamento Europeo y del Consejo de 16 de junio de 2003 relativo a las estadísticas comunitarias sobre la renta y las condiciones de vida.

La ECV se realizó por primera vez en 2004. Su objetivo principal es proporcionar información sobre la renta, el nivel y composición de la pobreza y la exclusión social en España y permitir la realización de comparaciones con otros países de la Unión Europea. Para ello se recogen los ingresos del año natural anterior a la entrevista. Además, se recogen otras muchas preguntas sobre condiciones de vida que se refieren al momento de la entrevista. Por ello las variables sobre ingresos de la encuesta de 2020 se refieren al año 2019 mientras que las demás preguntas se refieren a 2020.

La ECV 2020 ha sido realizada por el Instituto Nacional de Estadística (INE) en colaboración con el Instituto de Estadística de Cataluña (IDESCAT) en el ámbito de su comunidad autónoma.

ENCUESTA DE PRESUPUESTOS FAMILIARES. 2020

Descarga gratuita a través de la web del INE

La Encuesta de Presupuestos Familiares (EPF) suministra información anual sobre la naturaleza y destino de los gastos de consumo, así como sobre diversas características relativas a las condiciones de vida de los hogares.

Los gastos de consumo se refieren tanto al flujo monetario que destina el hogar al pago de determinados bienes y servicios de consumo final, como al valor de los bienes percibidos en concepto de autoconsumo, autosuministro, salario en especie, comidas gratuitas o bonificadas y alquiler imputado a la vivienda en la que reside el hogar (cuando es propietario de la misma o la tiene cedida por otros hogares o instituciones). Los gastos se registran en el momento de adquisición, independientemente de que el pago sea al contado o a plazos.

DIRECCIONES Y TELÉFONOS DE INTERÉS

INE-Avenida de Manoteras, 50-52 - 28050 Madrid. www.ine.es

Atención a usuarios

Tfno.: 91.583.91.00

Fax: 91.583.91.58

Consultas: www.ine.es/infoine

Lunes a jueves de 9 a 14 y de 16 a 18 horas

Viernes de 9 a 14:30 horas

Índice-Librería del INE

Tfno.: 91.583.94.38

Fax: 91.583.45.65

E-mail: indice@ine.es

Lunes a viernes de 9 a 14:30 horas

Biblioteca

E-mail: biblioteca@ine.es

PUBLICACIONES EDITADAS POR EL INE DE JULIO A SEPTIEMBRE DE 2021

INEbase. Julio 2021

Descarga gratuita a través de la web del INE.

Contenido:

- Contabilidad Nacional Trimestral de España. 2º trimestre 2021.
- Encuesta Anual de Coste Laboral. 2020.
- Encuesta de Condiciones de Vida. Módulo 2020. Sobreendeudamiento y ahorro.
- Encuesta de Condiciones de Vida. 2020.
- Encuesta del Gasto de la Industria en Protección Ambiental. 2019.
- EPA. Flujos de la Población Activa. 2º trimestre 2021.
- EPA. Resultados Trimestrales. 2º trimestre 2021.
- Estadística de Vida Laboral de las Personas con Discapacidad. 2019.
- Estadística sobre Transporte Ferroviario. 2020.
- Indicadores de Alta Tecnología. 2019.
- Indicadores de Confianza Empresarial. 3º trimestre 2021.
- Indicadores del Sector TIC. 2019.

INEbase. Junio 2021

Descarga gratuita a través de la web del INE.

Contenido:

- Cifras de Población. 01/01/2021. Datos provisionales.
- Contabilidad Nacional Anual de España. 1º trimestre 2021. Cuentas Trimestrales No Financieras de los Sectores Institucionales.
- Contabilidad Nacional Trimestral de España. 1º trimestre 2021.
- Encuesta Anual de Estructura Salarial. 2019.
- Encuesta Coyuntural sobre Stock y Existencias. 1º trimestre 2021.
- Encuesta de Comercio Internacional de Servicios. 1º trimestre 2021.
- Encuesta de Presupuestos Familiares. 2020.
- Encuesta de Turismo de Residentes. 1º trimestre 2021.
- Encuesta Trimestral de Coste Laboral. 1º trimestre 2021.
- EPA. Módulo 2020. Accidentes Laborales y Problemas de Salud Relacionados con el Trabajo.
- Estadística de Adquisiciones de Nacionalidad Española. 2020. Datos provisionales.
- Estadística de Ejecuciones Hipotecarias. 1º trimestre. 2021.
- Estadística de Migraciones 2020. Datos provisionales.
- Estadística de Productos en el Sector Comercio. 2019. Datos definitivos.
- Estadística de Productos en el Sector Servicios. 2019. Datos definitivos.
- Estadística de Profesionales Sanitarios Colegiados. 2020.
- Estadística de Variaciones Residenciales. 2020.
- Estadística Estructural de Empresas: sector comercio. 2019. Datos definitivos.
- Estadística Estructural de Empresas: sector industria. 2019. Datos definitivos.
- Estadística Estructural de Empresas: sector servicios. 2019. Datos definitivos.
- Estadística sobre Litigios en Arrendamientos Urbanos. 2020.
- Indicadores Demográficos Básicos. 2020. Datos provisionales.
- Índice de Coste Laboral Armonizado. ICLA. 1º trimestre 2021.
- Índice de Precios de Vivienda (IPV). 1º trimestre 2021.
- Índice de Precios del Sector Servicios. 1º trimestre 2021.
- Índice de Precios del Trabajo. 2019.
- Movimiento Natural de la Población 2020. Datos provisionales.

España en cifras

Descarga gratuita a través de la web del INE.

En el próximo número:

Inclusión y dependencia: estadísticas de discapacidad



Las sociedades modernas y democráticas se caracterizan por reconocer y asegurar la **dignidad de la persona en cuanto tal**, en su condición misma de persona, con independencia de las circunstancias, ideas, origen, edad y cualquier otro tipo de características específicas y diferenciadoras que pueden adornarla. Las sociedades avanzadas sentimos el mandato previo y director del **consenso social de asegurar la mejor calidad de vida de todos** como exigencia esencial, facilitando las herramientas que puedan ayudar a superar los efectos adversos de las circunstancias accidentales.

Los organismos internacionales tienen presente la necesidad de mejorar la calidad de los datos y su capacidad de comparación internacional para facilitar el diseño y evaluación de políticas para personas con discapacidades, tanto en los ámbitos globales como locales. También la Unión Europea y el gobierno nacional plantean estrategias para afrontar los efectos de la discapacidad, estrategias que se ven complementadas por las administraciones autonómicas y locales.

Junto con las fuentes públicas procedentes de organismos, estado, regiones y municipios, que permiten todas ellas conocer la discapacidad, así como los esfuerzos en su inclusión, buscaremos y recogeremos las que proceden del sector privado, con un énfasis específico en las del **tercer sector**, que se manifiesta particularmente activo en este campo.



Vecindarios

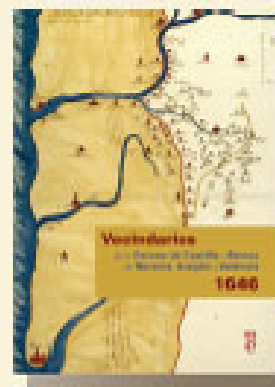
de la Corona de Castilla y reinos de Navarra, Aragón y Valencia

1646

Realizados a instancias de Felipe IV a lo largo de los años 1646-1647, obedecen a los acuerdos tomados en las Cortes de los distintos reinos, con el fin de proveer y distribuir recursos entre los vecinos objeto de tributación, ante las acuciantes necesidades financieras y militares por las que atravesó la Corona a mediados del turbulento siglo XVII.

El lector interesado encontrará en esta publicación una recopilación de datos cuantitativos (en algunos casos, de los que se da cuenta en la introducción, poco o nada explorados) de lugares y vecindad de la Corona de Castilla y por primera vez, para este periodo histórico, de otros territorios peninsulares de la Monarquía Hispánica, reinos de Navarra, Aragón y Valencia.

Las fuentes documentales sobre las que se ha trabajado provienen del Archivo General de Simancas, Archivo General de Navarra, Real Academia de la Historia y página web Asociación Raíces Reino de Valencia.



- 658 páginas.
- Con abundantes ilustraciones cartográficas de gran valor histórico.
- Precio: 20,43 euros (IVA incluido).

