

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.

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

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.

Participantes y desarrollo del curso

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. ●