

La respuesta del INE ante la crisis COVID-19: impulso a la innovación estadística

Alfredo Cristóbal Cristóbal

*Director General de Productos Estadísticos.
Instituto Nacional de Estadística*

La pandemia COVID-19 ha supuesto un cambio de paradigma en la estadística oficial. Cuando, en marzo de 2020, nos confinamos en nuestras casas, tanto la forma de elaborar estadísticas oficiales como las demandas de nuestros usuarios eran bastante diferentes a lo que lo son ahora.

Si bien ya se percibían ciertos cambios en los últimos años, antes de la pandemia las estadísticas oficiales se basaban casi exclusivamente en procedimientos que podríamos denominar tradicionales. Ello aseguraba un marco de calidad suficiente para que sus resultados pudieran utilizarse en el discurso económico y social. Nuestros usuarios estaban acostumbrados a una producción y una difusión estable de indicadores con un retraso en la disponibilidad de datos y unos niveles de desglose también suficientes para tomar decisiones en períodos estables.

Sin embargo, la pandemia hizo saltar todo por los aires. Por un lado, las demandas de nuestros usuarios cambiaron. Aspectos que estadísticamente no eran muy relevantes antes de la pandemia comenzaron a serlo durante ese período, de cara, por ejemplo, a tomar decisiones sobre el desarrollo del virus sobre la población. Pero, además, la información que se le pedía a la estadística tenía que tener un tiempo de producción mínimo y los datos sobre los que se sustentaba debían permitir su repetición frecuente y una distribución de resultados de elevado detalle. La producción estadística tradicional ni estaba orientada a generar ese tipo de información, que ahora se solicitaba, ni permitía alcanzar los niveles de puntualidad, frecuencia y granularidad que eran necesarios en ese momento para una rápida toma de decisiones basada en información.

La respuesta del INE tuvo que enfocarse para cumplir con estos aspectos que centraban las demandas de los usuarios, hecho que se tradujo en diferentes formas de innovación estadística, de las que destacaremos tres: la innovación en fuentes de información, en procesos estadísticos y en métodos de difusión y comunicación.

La producción estadística tradicional ni estaba orientada a generar ese tipo de información que ahora se solicitaba ni permitía alcanzar los niveles de puntualidad, frecuencia y granularidad que eran necesarios en ese momento para una rápida toma de decisiones basada en información

El resultado de la innovación ha sido la publicación de un conjunto de estadísticas denominadas experimentales, que han dado respuesta a las demandas de los usuarios desde el inicio de la pandemia y cuyo futuro estará ligado al interés que puedan seguir suscitando para los usuarios, pudiendo convertirse en estadística oficial si se mantiene su relevancia.

Innovación en fuentes de información

El procedimiento tradicional de obtención de datos para la elaboración de estadísticas ha estado basado en censos y en encuestas. Estas formas de recogida son costosas y los ciudadanos soportan una carga importante para proporcionarnos la información, si bien los datos que se obtienen pueden estructurarse de for-

ma más o menos sencilla en un procedimiento estadístico que permite obtener buenas estimaciones de las variables del fenómeno que se desea medir.

Sin embargo, el desarrollo de los registros administrativos, cada vez más numerosos, así como de las bases de datos privadas que contienen datos que los ciudadanos vamos dejando de forma consciente o inconsciente en nuestras gestiones del día a día, permite a los estadísticos el acceso a un tipo de información novedosa, sin tener que interaccionar con el informante y de un volumen y tamaño ingente. Aunque es cierto que la información que contienen no se ha recogido con efectos puramente estadísticos y puede estar desestructurada como para ser utilizada directamente en el método estadístico.

La innovación estadística basada en la explotación de registros administrativos durante la pandemia ha sido bastante prolija en el INE, dando lugar a diferentes estadísticas experimentales

La innovación estadística basada en la explotación de registros administrativos durante la pandemia ha sido bastante prolija en el INE, dando lugar a diferentes estadísticas experimentales basadas, por ejemplo, en la explotación de los Registros Civiles o en la de los registros en poder de la Agencia Tributaria o de la Seguridad Social y sus interacciones entre sí y con el Registro de Población. La disponibilidad de estas fuentes registrales para la realización de estadísticas está asegurada, bien por la legislación nacional, bien por la europea, por lo que la sostenibilidad de las estadísticas basadas en ellos está, asimismo, asegurada.

Pero también durante la pandemia se han realizado diferentes estadísticas experimentales basadas en la explotación de fuentes de datos privadas, como, por ejemplo, los estudios de movilidad diaria, estacional o de turismo basados

en el posicionamiento de dispositivos móviles. El acceso a estas fuentes de información es más complejo que en el caso de los registros, no pudiendo accederse a los microdatos y debiendo acordarse el nivel de acceso y metodología con el proveedor privado. Además, a día de hoy, la legislación que servirá de base para este acceso (a escala europea) está en construcción, por lo que es todavía incierto el grado de sostenibilidad de las estadísticas basadas en este tipo de fuentes.

No hay que olvidar que el uso experimental de este tipo de información ha permitido ir mucho más lejos en el conocimiento de fenómenos estadísticos. Por ejemplo, en el caso del turismo, las estadísticas tradicionales nos permiten conocer cuáles son las preferencias de los turistas extranjeros a la hora de elegir su destino principal de vacaciones en España. Pero hay regiones que no son, usualmente, destino principal de los turistas y que reciben, en forma de excursiones, un buen número de aquellos. El impacto del turismo en dichas regiones aparece claramente infravalorado en las estadísticas tradicionales. Sin embargo, disponer, por ejemplo, de datos de gasto de los turistas a través de sus tarjetas bancarias en el sitio en el que realmente se ha producido el gasto mejora la visibilidad del fenómeno en esas regiones.

Innovación en procesos

La disponibilidad de grandes bases de datos, tanto en el ámbito público como en el privado, ha venido asimismo acompañada del desarrollo de técnicas para su gestión, como las técnicas de *machine learning* o de inteligencia artificial. Ello ha permitido desarrollar algoritmos que ayuden en la modelización estadística para la elaboración de indicadores.

Así, por ejemplo, predecir el resultado de un indicador mensual con las observaciones recogidas en el trabajo de campo durante los primeros diez, quince o veinte días con una precisión máxima permite disponer de resultados avanzados del indicador que pueden utilizarse con seguridad en contextos de estadísticas de síntesis, como las cuentas nacionales.

Asimismo, la utilización de técnicas de aprendizaje artificial para completar variables no ob-

servadas de un registro basadas en la información suministrada por una muestra permitirá, por ejemplo, disponer de un registro completo de empresas y todas sus variables que regulatoriamente son requeridas en el ámbito de las encuestas estructurales de empresas.

En este campo va a ser muy relevante el desarrollo de la ciencia de datos. La posibilidad que podamos tener desde el INE de captar científicos de datos en nuestras oposiciones de acceso a los cuerpos estadísticos permitirá avanzar en la innovación en procesos.

Innovación en difusión y comunicación

La comunicación de datos estadísticos tiene que ser lo más próxima posible al ciudadano o, dicho de otro modo, los ciudadanos tienen que verse reflejados en las estadísticas de una manera lo más cercana posible. Las estimaciones nacionales o incluso al nivel de comunidad autónoma o provincia quedan muy lejos del ciudadano y hay que ampliar el foco para acercarnos lo más posible a él. Datos a escala municipal e incluso a escala infra municipal son mucho más apreciados.

Este hecho exige una innovación también en nuevas formas de difusión que acerquen la estadística lo más posible al ciudadano.

Por ejemplo, desde el punto de vista geográfico, no es lo mismo una renta por hogar media en una provincia que en un municipio de dicha provincia o incluso en un distrito de este, dado que los datos medios pueden estar contaminados por distribuciones diferentes, como en este ejemplo, en la renta media por hogar en la capital de la provincia, que puede ser muy diferente a la de sus municipios. Cruzando los datos del registro de población con el registro de rentas personales y de hogares del IRPF se puede obtener esta diferenciación a escala geográfica muy granular, por ejemplo, al nivel de sección censal.

Una representación gráfica en forma de mapas a escala lo más pequeña posible, con una elevada granularidad, permitirá diferenciar las características de la población en entornos geográficos reducidos, con respecto al fenómeno económico o social que se desee medir.

Conclusiones

Los tres tipos de innovación que el INE viene realizando desde la pandemia están muy ligados entre sí y tienen su origen en la disponibilidad de grandes bases de datos, públicas y privadas.

La disponibilidad de esta información permite innovar experimentalmente en nuevas estadísticas, en procesos novedosos que tienen su origen en la gestión de grandes volúmenes de datos aunque no estén estructurados (ciencia de datos) y ambos tienen su reflejo en una difusión y comunicación de datos mucho más cercana al ciudadano.

Los tres tipos de innovación que el INE viene realizando desde la pandemia están muy ligados entre sí y tienen su origen en la disponibilidad de grandes bases de datos, públicas y privadas

Si bien el acceso a estas bases de datos no está, a día de hoy, garantizado en todo momento para el INE, especialmente en el caso de las bases de datos privadas, la toma en consideración de este hecho dentro de la legislación nacional y europea permitirá una mejor accesibilidad a esta información. Asimismo, el INE nunca va a entrar en competencia comercial con los tenedores privados de esas bases de datos. Por ello, la práctica de acuerdos con estos proveedores hará que poco a poco se entienda el papel del INE y de la estadística oficial en lo que se refiere a la utilización de esa información en un contexto de servicio público. Al fin y al cabo, este hecho es innovador para el INE pero también lo es para los tenedores de esa información.

Asegurada la sostenibilidad de esta información y, por lo tanto, de las estadísticas que, a día de hoy, tienen el carácter de experimentales, el camino para continuar innovando en productos, procesos y formas de difusión estará más abierto y la producción estadística del INE podrá basarse cada vez más en este tipo de información. No hay vuelta atrás posible. ●