Alfredo Cristóbal Cristóbal*

LECCIONES APRENDIDAS DE LA CRISIS PROVOCADA POR LA PANDEMIA. IMPACTO EN LA ESTADÍSTICA OFICIAL

El artículo describe la forma en la que la estadística oficial ha reaccionado ante un contexto muy difícil, marcado por la pandemia, para ofrecer información frecuente y más oportuna de la actividad socioeconómica, a través de estadísticas experimentales construidas tanto por una utilización masiva de registros administrativos como atendiendo a fuentes de datos privadas hasta ahora no utilizadas. Se analiza la sostenibilidad de este modelo y se discute el nuevo rol de las oficinas de estadística como gestores de datos.

Lessons learned from the crisis caused by the pandemic. Impact on official statistics

The article describes the way in which official statistics have reacted to a very difficult context, marked by the pandemic, to offer frequent and more timely information on socioeconomic activity, through experimental statistics constructed both by a massive use of administrative records as attending to private data sources until now not used. We analyze the sustainability of this model and we talk over the new role of statistical offices as data managers.

Palabras clave: *estadística experimental, registros administrativos, fuentes privadas, resiliencia, data steward.*
Keywords: *experimental statistics, administrative records, private sources, resilience, data steward.*
JEL: *A10, H12, H41, O35.*

1. Antecedentes

Hemos experimentado una crisis sin precedentes que, desafortunadamente, aún no ha terminado. Siendo honestos, si queremos extraer lecciones que hayamos aprendido durante este periodo y que ayuden a los sistemas estadísticos a encarar nuevas crisis en el futuro, hemos de ser críticos y, en primer lugar, reconocer las debilidades y vulnerabilidades que aquellos tenían antes de la pandemia.

La verdad es que los sistemas estadísticos no estaban preparados para esta pandemia. Así, gran parte de la respuesta dada por estos a las demandas generadas por los usuarios durante la crisis, especialmente en las primeras semanas, tuvo que ser bajo una perspectiva *ad hoc*.

Nos situamos al principio de la pandemia, en torno a mitad de marzo de 2020, momento en el que el COVID-19 comenzó a impactar con fuerza sobre la población, a partir de ese momento, podemos distinguir tres etapas. Una primera etapa, en la que las

* Director General de Productos Estadísticos.
Instituto Nacional de Estadística.
Contacto: alfredo.cristobal@ine.es
Versión de octubre de 2021.
https://doi.org/10.32796/ice.2021.923.7333

oficinas de estadística tomaron las primeras medidas internas para intentar garantizar la continuidad de la producción estadística. Una segunda etapa, en la que el cometido más importante era estimar los principales indicadores macroeconómicos (PIB, empleo, empresas…) en un contexto de incertidumbre, y así implementar medidas que permitieran mitigar los efectos de la crisis. Lo que era más importante para los estadísticos, en esta etapa, era ser conscientes de las nuevas demandas de los usuarios para ver de qué forma se podía dar una respuesta desde los sistemas estadísticos. Finalmente, una vez pasado el mayor impacto de la crisis y reducido el nivel de estrés, en una tercera etapa, los sistemas estadísticos deberían ser capaces de monitorizar la recuperación, utilizando para ello, tanto la información estadística producida habitualmente, como la desarrollada experimentalmente durante el periodo de crisis, analizando su sostenibilidad, de cara a dar respuesta rápida a las nuevas demandas de los usuarios.

En los apartados que se desarrollan a continuación, se describen las características que han definido cada una de las tres etapas mencionadas, ofreciendo al final una serie de conclusiones aprendidas durante este periodo que podrían guiar nuevas funciones para las oficinas de estadística nacionales.

2. Comienzo de la crisis. Primeras reacciones

En un primer momento, las oficinas de estadística implementaron medidas de emergencia, inicialmente de carácter temporal, con el objeto de que la producción estadística se viera afectada lo menos posible. Así, como primera consecuencia del confinamiento de la población, el personal del INE tuvo que realizar sus tareas en remoto, teletrabajando desde sus domicilios, sin poder acudir a las oficinas. Ello supuso un gran esfuerzo de logística, ya que una situación así no se había planteado en el pasado y la asignación de recursos no estaba enfocada en esa dirección. De igual manera, se modificaron los procedimientos de recogida de información (no se realizan entrevistas de forma física y se estimulan la recogida por otros canales como teléfono o internet) en determinadas operaciones estadísticas y se priorizan aquellas que se consideran clave para conocer el impacto de la crisis sobre la economía y la sociedad. Era el momento de evitar el colapso, evitar detener la producción.

De cara a no perder la comparabilidad espacial, era crucial que los sistemas estadísticos de los diferentes países, trabajaran de forma coordinada. Así, en el ámbito de la Unión Europea (UE), Eurostat tomó las riendas de la coordinación de los Estados miembros desde el inicio, estableciendo prioridades comunes en lo referente a la producción, concediendo cierta flexibilidad a aspectos que no estaban regulados y proporcionando orientación en muchos dominios estadísticos (empleo, indicadores de coyuntura económica, estadísticas macroeconómicas y de finanzas públicas, estadísticas sociales…). Ello contribuyó a la coordinación de los trabajos a desarrollar entre países.

Pero también fue crucial la coordinación institucional a nivel nacional. Así, todas las unidades del Sistema Estadístico Nacional trabajaron conjuntamente en una única dirección, coordinadas desde el INE (Instituto Nacional de Estadística).

La correcta comunicación a los usuarios de las acciones que se iban a desarrollar desde el INE era un elemento clave para que la sociedad fuera consciente del esfuerzo que se estaba realizando para garantizar la producción.

Así, el 30 de marzo de 2020, el INE emitió un comunicado público en su web y en las redes sociales, con tres mensajes claros. Primero, el compromiso de continuar publicando estadísticas de la más alta calidad. Segundo, estableciendo la priorización de operaciones como el Índice de Precios al Consumo, las Cuentas Nacionales Trimestrales, los Indicadores de Coyuntura Económica y los Indicadores del Mercado de Trabajo. Tercero, agradeciendo la colaboración de los informantes y rogando a las empresas, hogares e individuos que continuaran proporcionando

información como lo habían hecho hasta la fecha. Se indicaba en dicho comunicado que, «… solo con su ayuda el INE podrá obtener la información necesaria para elaborar estadísticas que, posteriormente, guiarán decisiones de política económica y social, nacional y europea, facilitando así que se adopten las medidas más apropiadas en cada momento. Y todo ello en beneficio del conjunto de la sociedad». El comunicado terminaba con el deseo de mantener el calendario de difusión de estadísticas, si bien con el compromiso de anunciar debidamente cualquier revisión de dicho calendario como consecuencia del desarrollo de la pandemia. Esta fue la primera comunicación oficial del INE a la sociedad.

3. Respuesta de la estadística oficial a la crisis del COVID-19

Los gestores de política económica y, usuarios en general, demandaban información en tiempo real y sobre una base diaria o semanal, no superior; *la información que se proporciona hoy reemplaza la que se proporcionó ayer*, decían. Esto reveló una nueva debilidad de los sistemas estadísticos, pues no estaban preparados para cumplir las necesidades de los gestores políticos, que tuvieron que moverse en un contexto de incertidumbre causado por la ausencia de información. Las estadísticas mensuales o trimestrales elaboradas con los estándares de calidad más elevados no estuvieron disponibles durante las primeras semanas de la pandemia y los gestores de política económica no podían esperar para definir sus estrategias. Las primeras estadísticas oficiales referidas al mes de marzo (o al primer trimestre) no estuvieron disponibles hasta casi finales de abril, es decir, seis semanas después del comienzo de la pandemia. Los sistemas estadísticos, en general y, el INE, en particular, tuvieron que abordar este tema y comenzar a producir otro tipo de estadísticas, más allá de la estadística oficial, para dar respuesta a las necesidades de los usuarios y que las políticas diseñadas estuvieran basadas en datos.

Así, como medida inicial, el INE elaboró un tablero con los principales indicadores socioeconómicos de corto plazo oficiales (macroeconómicos, de precios, de empresas, de turismo, de hogares…) con el fin de que los usuarios pudieran seguir los efectos que estaba provocando la pandemia sobre el tejido económico y social del país.

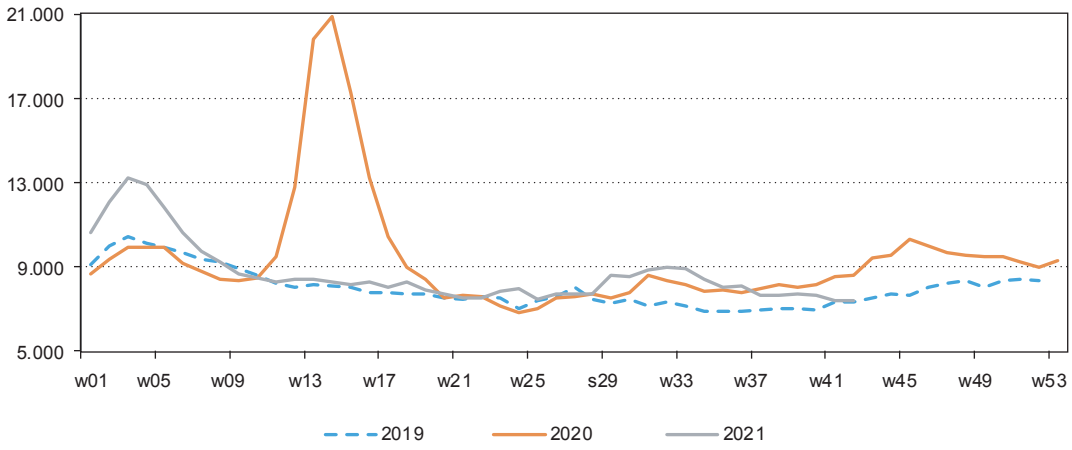
No obstante, la información solicitada por nuestros principales usuarios estaba centrada en la esfera económica y social. Los datos relativos al dominio de salud eran muy importantes, especialmente en las primeras semanas, de cara a conocer la evolución temporal y transmisión espacial del virus, así como su letalidad. Como ejemplos, uno de los temas de conocimiento más relevante en aquel momento era conocer hasta qué punto la movilidad diaria (*commuting*) de la población se había reducido en comparación al periodo anterior de la pandemia, de cara a controlar la expansión del brote entre territorios. Pero también, cuál era el efecto de los esquemas de los Expedientes de Regulación Temporal de Empleo (ERTE) interpuestos desde el inicio de la pandemia, sobre los indicadores macroeconómicos, como el Producto Interior Bruto (PIB) o sobre las finanzas públicas (ingresos, gastos, déficit de las Administraciones públicas). O también, cuántos negocios habían cerrado o estaban dormidos, en *stand-by,* esperando el retorno a la normalidad… En general, qué información de muy corto plazo se podía ofrecer para evaluar el comportamiento de los agentes económicos, empresas, hogares, Administraciones...

Desde el principio, en el INE teníamos claro que para cumplir con estas necesidades se debían investigar nuevas fuentes de información. De este modo, la respuesta fue innovar en fuentes que permitieran producir estadísticas, bajo el apelativo de experimentales, en dos direcciones claramente definidas.

Por un lado, haciendo un uso mucho más extensivo de los registros administrativos existentes. Pero también, tratando de acercarnos a nuevas fuentes de datos existentes en bases de datos privadas. En el

FIGURA 1

DEFUNCIONES SEMANALES



FUENTE: INE.

INE ya se estaban realizando antes de la pandemia intentos de aproximación a estas fuentes privadas para planificar algunas operaciones, pero las necesidades surgidas durante la pandemia impulsaron esta nueva línea de producción estadística.

Estadísticas experimentales basadas en el uso de registros administrativos

Desde el inicio de la pandemia, una información ampliamente demandada estaba relacionada con la mortalidad ocasionada por el virus. La estadística oficial del Movimiento Natural de la Población (MNP: nacimientos, defunciones, entre otras variables) se ofrecía a los usuarios con una periodicidad semestral. De esta forma, los primeros resultados provisionales de la mortalidad en el primer semestre de 2020 no hubieran estado disponibles hasta el mes de diciembre. Si bien esta estadística no detalla la causa del fallecimiento de las personas, al menos podría analizarse el exceso de mortalidad frente al observado los años anteriores, variable normalmente utilizada para investigar los efectos sobre la población en episodios de epidemias.

En este sentido, se estudió qué pasaría si en lugar de esperar a recibir la información de los boletines de defunción se utilizaba la información grabada en el registro INFOREG (Informatización de los Registros Civiles) del Ministerio de Justicia, que tiene una cobertura aproximada del 93 % de los Registros Civiles de España y, cuyos resultados se pueden obtener en torno a diez días después de producido un evento (defunción). El resultado fue que con ese mínimo retraso de diez días, el INE podía ofrecer información semanal (provisional) del número de defunciones ocurridas y, por lo tanto, rehaciendo la serie histórica, estimar el exceso de mortalidad provocado por la pandemia, con base semanal. Dado que el cubrimiento del registro no era similar en todas las provincias, se tuvo que diseñar un modelo para poder ofrecer datos, no solamente nacionales sino por comunidades autónomas y provincias (Figura 1).

El análisis pormenorizado de los certificados oficiales de defunción, permitió ofrecer datos provisionales de fallecidos por COVID-19 en la primera ola de la pandemia (entre enero y mayo de 2020) antes de terminar el año 2020. Así, la Estadística de Defunciones por Causa de Muerte, avanzó estos resultados un año con respecto a la fecha prevista de publicación y dos años con respecto a la fecha límite establecida en el Reglamento europeo que regula esta operación. Atendiendo a las indicaciones de la Organización Mundial de la Salud, no solamente se estimaron datos de fallecidos en los que los facultativos habían detectado la presencia del virus sino también aquellos fallecidos en los que no se había detectado por prueba médica pero que presentaron síntomas compatibles con el COVID-19.

Otra de las variables demográficas que se vio afectada significativamente fue el número de nacimientos. Utilizando la misma metodología que para las defunciones, se ha podido elaborar una estadística experimental para ofrecer mensualmente, con un desfase mínimo, el número de nacimientos en España y sus comunidades autónomas y provincias. Como hecho notable, merece la pena observar que, entre los meses de diciembre de 2020 y febrero de 2021, nueve meses después de la fase inicial de la pandemia, el número de nacimientos se redujo considerablemente, llegando a superar caídas del 20 %.

Una de las principales preocupaciones de los estadísticos, dedicados a las operaciones estadísticas sobre empresas durante las primeras semanas de la pandemia, era la de distinguir entre negocios que se habían cerrado definitivamente, de aquellos que lo habían hecho a consecuencia de la legislación temporal que entró en vigor durante el confinamiento, pero que volverían a operar de una manera más o menos normal, una vez que se terminara la temporalidad de dichos actos legales. Este hecho afectaba enormemente a las muestras de los indicadores de coyuntura económica, que habían sido diseñadas previamente a la pandemia, en el último trimestre de 2019. Era muy difícil contactar con los establecimientos que estaban cerrados temporalmente por ley y se temía, como así se comprobó después, un notable incremento de la falta de respuesta, que era necesario abordar de alguna forma, de cara a separar aquellos establecimientos cuya falta de respuesta denotaba un «cero» temporal en las ventas, de aquellos que habían cerrado definitivamente y que deberían ser tratados de otra manera en el ámbito de la gestión muestral.

Este hecho tuvo, si cabe, más efecto en los indicadores del mes de marzo, ya que los establecimientos habían estado abiertos de una manera más o menos normal durante la primera quincena del mes. Los cuestionarios estadísticos llegaron a dichos establecimientos durante el mes de abril, momento en el que estaban cerrados al público, dificultando el contacto y realmente conocer la evolución del negocio durante el mes de marzo.

Este hecho impulsó dos estadísticas experimentales nuevas basadas en la explotación de registros administrativos. Por un lado, proporcionar información con una mayor frecuencia de la demografía empresarial, de forma que se pueda conocer con mayor rapidez la evolución de los cierres de negocios, atendiendo a su no presencia en los registros administrativos y, por otro, ofrecer información, aunque sea parcial, basada en registros de los indicadores de coyuntura económica con una mayor frecuencia y mínimo desfase.

La explotación exhaustiva de los registros de la Agencia Tributaria y de la Seguridad Social, tanto para personas físicas como jurídicas ha permitido una aproximación a la demografía de empresas desde una perspectiva coyuntural (trimestral). Así, es posible ofrecer información trimestral sobre el número de empresas creadas, empresas cerradas o incluso analizar la evolución temporal de cohortes de empresas.

Como puede verse en la Tabla 1, de los 1,2 millones de empresas que había registradas a principios de 2020, un 12 % habían desaparecido un trimestre más tarde, en las primeras semanas de la pandemia. De la misma manera, del *stock* inicial, solamente un 80 % permanecían vivas un año después. Estos datos

TABLA1

SUPERVIVENCIAS OBSERVADAS EN EL TIEMPO

Empresas

Cohorte		En 1 trimestre		En 2 trimestres		En 3 trimestres		En 4 trimestres	
	Número	Número	%	Número	%	Número	%	Número	%
1 ene 2020	1.190.870	1.050.404	88,20	1.014.398	85,18	983.011	82,55	952.502	79,98
1 abr 2020	1.102.738	1.059.378	96,07	1.022.614	92,73	986.485	89,46	950.014	86,15
1 jul 2020	1.141.253	1.085.053	95,08	1.033.148	90,53	991.293	86,86	966.184	84,66
1 oct 2020	1.149.469	1.080.193	93,97	1.031.022	89,70	1.000.340	87,03	–	–
1 ene 2021	1.136.561	1.074.320	94,52	1.035.575	91,11	–	–	–	–
1 abr 2021	1.137.018	1.087.341	95,63	–	–	–	–	–	–

Autónomos

Cohorte		En 1 trimestre		En 2 trimestres		En 3 trimestres		En 4 trimestres	
	Número	Número	%	Número	%	Número	%	Número	%
1 ene 2020	2.997.941	2.807.861	93,66	2.748.666	91,69	2.673.302	89,17	2.597.708	86,65
1 abr 2020	2.938.754	2.872.111	97,73	2.784.268	94,74	2.700.506	91,89	2.641.443	89,88
1 jul 2020	2.962.399	2.860.164	96,55	2.767.434	93,42	2.704.508	91,29	2.637.789	89,04
1 oct 2020	2.963.105	2.856.926	96,42	2.787.463	94,07	2.714.054	91,59	–	–
1 ene 2021	2.964.241	2.884.077	97,30	2.800.408	94,47	–	–	–	–
1 abr 2021	2.993.411	2.898.267	96,82	–	–	–	–	–	–

FUENTE: INE.

muestran el duro efecto del inicio de la pandemia sobre las empresas. Como puede verse también en la Tabla 1, a partir del mes de abril, de la cohorte inicial de cada trimestre solamente se pierden entre un 4 % y un 6 % en los primeros tres meses. Al cabo de un año, el número de empresas que se pierden se sitúa en el entorno del 15 %.

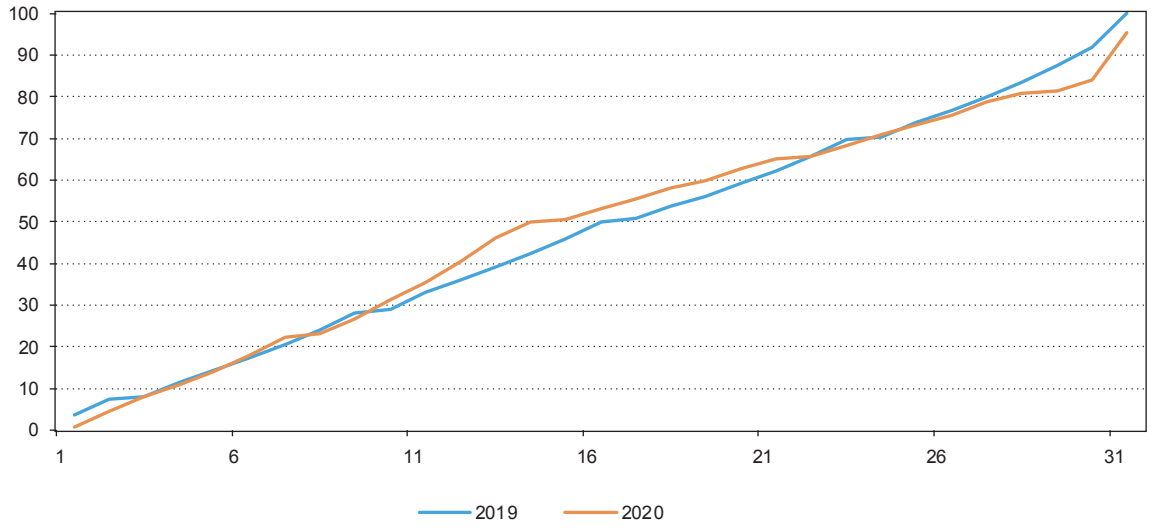
Estos efectos han sido más notables en las empresas que en los autónomos. Si partimos del *stock* de tres millones de autónomos a primeros de 2020, en el primer trimestre se perdieron un 6,3 % y, al cabo de un año habían desaparecido algo más del 13 %. Para los trimestres posteriores puede verse que en los primeros tres meses se pierden entre un 2,5 % y un 3,5 % y que, al cabo de un año, las pérdidas son inferiores al 10 %.

De la misma manera, la explotación de un nuevo registro de la Agencia Tributaria, el denominado Sistema Inmediato de Información (SII) ha permitido estimar diariamente y con un retraso mínimo, la cifra de negocio (ventas) de las grandes empresas de comercio al por menor.

En efecto, a través de este nuevo sistema, los contribuyentes obligados al SII (y aquellos que voluntariamente decidan utilizarlo) deben suministrar el detalle

FIGURA 2

VENTAS ACUMULADAS AL DIA *t* DE LOS GRANDES MINORISTAS
(En %)



FUENTE: INE basado en datos AEAT (Agencia Estatal de Administración Tributaria).

de los registros diarios de facturación a través de la Sede electrónica de la Agencia Tributaria, en un plazo máximo de cuatro días hábiles.

En el caso del comercio minorista, la cifra de negocio de estas empresas representa más de la mitad del total, siendo notable su presencia en el caso del comercio al por menor en establecimientos no especializados (más del 80 %), en el caso de comercio no presencial (comercio electrónico fundamentalmente, más del 63 %) o en el comercio de equipos de tecnologías de la información y las comunicaciones, en el que representa casi el 60 % del total.

Por tanto, el indicador diario del comercio de estas empresas es, de por sí, un indicador relevante para el seguimiento de la actividad del comercio minorista, pudiendo utilizarse como *proxy* para una estimación rápida del Índice de Comercio al por Menor (ICM).

Dadas las diferentes pautas de facturación que tienen las empresas, se observa que el indicador diario ofrece una cierta estabilidad de las ventas del día *t* en torno a diez días después. Esto determina que el indicador de un mes cualquiera se tendría el día 10 del mes siguiente, entre 15 y 20 días antes que esté disponible el ICM para ese mismo mes.

Pero la gran potencia de este indicador es su frecuencia diaria y puede comprobarse analizando las ventas minoristas del mes de marzo de 2020, en el inicio de la pandemia. Como puede verse en la Figura 2, que representa las ventas acumuladas al día *t* de las grandes empresas de comercio minorista en los meses de marzo de 2019 y 2020, hasta el día 10 de marzo de 2020, las ventas diarias acumuladas presentaban una evolución más o menos similar a la del año precedente. Sin embargo, entre los días 11 y 15 del mes, las ventas se disparan. Fundamentalmente se debieron a ventas de productos alimenticios y de primera necesidad en establecimientos que quedaron prácticamente desabastecidos debido al fuerte volumen de compras

realizadas por los hogares por motivo de precaución, en los días previos al confinamiento.

Asimismo, puede verse que, una vez comenzado el confinamiento y como consecuencia del cierre de muchos establecimientos minoristas, las ventas se van reduciendo de forma paulatina hasta acabar el mes con un volumen de facturación menor que el del año precedente.

El indicador descrito está relacionado con las ventas de grandes empresas en el comercio al por menor, pero esta metodología puede extenderse también a otras actividades de la industria y los servicios, por lo que desde el INE se promoverán estadísticas experimentales en estas otras materias para complementar el indicador existente.

Estadísticas experimentales basadas en fuentes privadas

Con motivo de innovar en el proceso de elaboración de los Censos de Población y Viviendas 2021 y, más concretamente, con los aspectos relacionados con la movilidad de la población, el INE venía desarrollando un proyecto para implementar indicadores de movilidad basados en la información proporcionada por los Operadores de Redes Móviles de telefonía en referencia al posicionamiento de los dispositivos móviles de la población. El estudio planteaba tres posibles indicadores; un indicador, de análisis de la movilidad diaria de la población (*commuting*); un segundo indicador, relacionado con el primero, sobre cuál es la población de las ciudades en los periodos de día y de noche y, finalmente, un tercer indicador, relativo a la movilidad estacional de la población (en periodos vacacionales). Los dos primeros se realizaron en el mes de noviembre de 2019 y el tercero contempló fechas fijadas de los meses de julio y agosto de 2019 (verano) y de diciembre (navidades).

Para definir movimientos entre territorios, se dividió el territorio nacional en algo más de 3.200 áreas, representando cada una, una población media de unas 15.000 personas. Las grandes ciudades comprendían varias áreas, relacionadas con sus distritos y los territorios menos poblados se unieron para evitar problemas de protección de datos. Los operadores de telefonía proporcionaban al INE directamente la información, implementando la metodología que el propio INE elaboró al respecto.

Así, en el primero de los estudios, relativo al *commuting* o movilidad diaria, los operadores, a grandes rasgos, proporcionaban información sobre dónde (en qué área) se situaban los dispositivos móviles durante la noche, sin moverse, identificándose esta como el área de residencia y dónde se encontraban los mismos móviles durante el día (en qué área o en qué áreas), identificándose esta o estas como las áreas de actividad de la población.

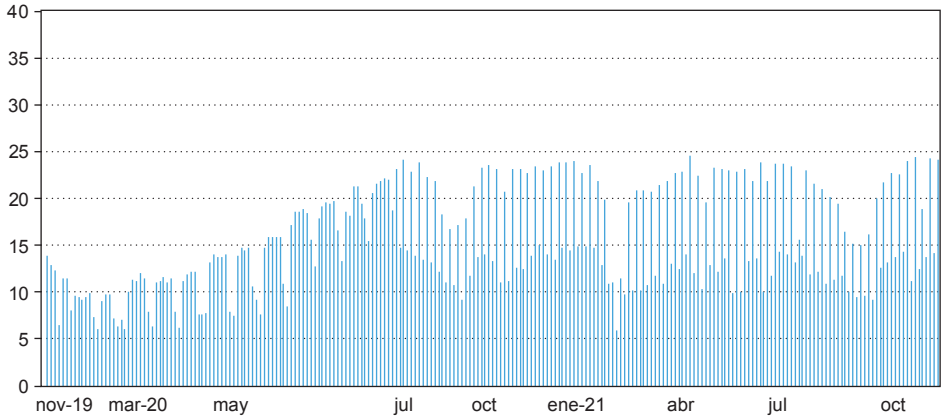
En noviembre de 2019, como resultado general promedio, el 29,2 % de la población abandonaba su área de residencia durante el día para realizar actividades (trabajar, estudiar…). El porcentaje más elevado se observaba en Palencia o Córdoba, por encima del 75 % y el más reducido en Yecla o Teruel, con movimientos de apenas un 6 % de la población. Pero la información proporcionada por el estudio iba mucho más allá de estas cifras, pues se conocía el origen y destino de la población, es decir, se podía conocer dónde realizaban actividad durante el día los residentes, por ejemplo, en Segovia o bien de dónde procedían (de qué área) todas las personas que estaban realizando actividad diaria, por ejemplo, en Madrid. Todo ello a un nivel de desglose de más de 3.000 áreas, es decir con un grado de granularidad muy detallado.

Este indicador fue clave en el inicio de la pandemia, pues el movimiento cotidiano de la población se redujo drásticamente como consecuencia del confinamiento y de la reducción al mínimo de la actividad, cerrando toda actividad no esencial.

El INE reaccionó inmediatamente repitiendo este estudio con una frecuencia diaria desde finales de marzo. Así, no solamente se podía observar la reducción de la movilidad, como *proxy* de la actividad, sino

FIGURA 3

COMUNIDAD DE MADRID. PERSONAS QUE SALEN DE SU ÁREA DE RESIDENCIA DIARIAMENTE (En %)



FUENTE: INE basado en datos de operadores de telefonía móvil.

entre qué áreas se estaba produciendo. Esta información fue muy valorada por los gestores ministeriales debido a su utilidad, entre otros aspectos, para analizar la transmisibilidad del virus entre territorios.

El indicador se mantuvo con frecuencia diaria hasta finales de junio, ya comenzada la desescalada y, a partir de ese momento, se repitió dos veces por semana, uno en día laborable (miércoles) y otro en día de fin de semana (domingo), dado que se observaba que no había diferencias sensibles entre los cinco días de la semana entre sí, pero sí entre aquellos y los dos del fin de semana.

Los resultados para la Comunidad de Madrid aparecen en la Figura 3:

Así, la movilidad diaria en la Comunidad de Madrid se redujo desde el 36,7 % de la población, en noviembre de 2019, a valores por debajo del 15 % en el inicio del confinamiento, e incluso por debajo del 10 % en el periodo de cierre de toda actividad no esencial, momento en el que alcanza el mínimo de la serie. Posteriormente, los valores se han ido recuperando paulatinamente hasta situarse en valores próximos al 25 %, casi 12 puntos menos que los registros previos a la pandemia. La recuperación de la actividad aún no es completa. El incremento del teletrabajo y la sustitución de reuniones de trabajo presenciales por videoconferencias están detrás de este resultado.

Además, en la Figura 3 pueden contemplarse nítidamente las diferentes olas de la pandemia; la primera, al inicio del confinamiento; la segunda, en el verano de 2020; la tercera, en torno al inicio de 2021, momento en el que se superpuso también el efecto del temporal *Filomena* y las posteriores cuarta y quinta ola, en la primavera y el verano de 2021.

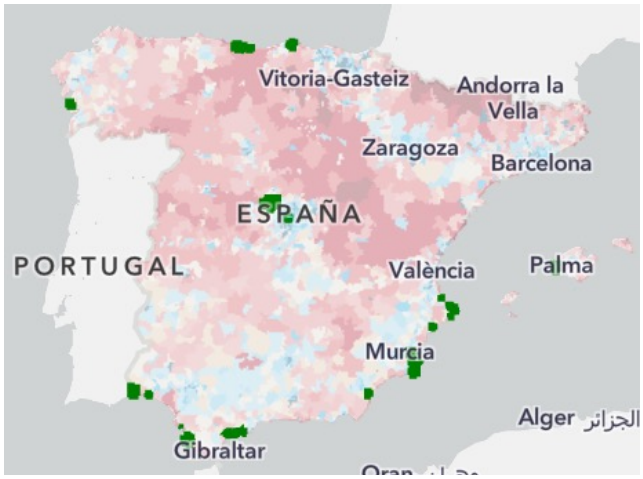
Como se ha comentado, esta estadística experimental tiene un alto grado de granularidad. Así, se puede conocer, cada día, dónde se encuentran los residentes en una ciudad o un distrito de esta, hecho que permite explotar esta información en el dominio turístico. Por ejemplo, la Figura 4 muestra dónde se encuentran los residentes

FIGURA 4

LOCALIZACIÓN DE LOS RESIDENTES DE VALLEHERMOSO (MADRID)

15 de agosto de 2019

15 de agosto de 2020



FUENTE: INE.

en un distrito de Madrid (Vallehermoso) los días 15 de agosto de 2019 y de 2020.

Naturalmente, esto no es más que una foto fija en la que se representa dónde está la población en un día determinado, pero no ofrece información sobre cómo han llegado a su destino esas personas o si ese destino es el único que han realizado en su viaje o no. Por lo tanto, esta metodología puede ampliarse para ofrecer datos sobre los viajes de los residentes en España y todas las etapas realizadas, tanto en España como en el extranjero, pues los operadores de telefonía pueden también ofrecer esa información. De la misma manera, se podría aplicar la metodología a los no residentes que llegan a España, para conocer los destinos de todos sus viajes y las etapas realizadas. El INE está preparando otra estadística experimental, en el ámbito del turismo, en la que ofrecer esta información.

Durante la pandemia, el INE ha explotado también bases de datos de pagos con tarjetas de crédito y de débito. Estos datos, que han sido muy útiles en el ámbito de las cuentas nacionales como indicador de seguimiento de la actividad y del consumo, también se han erigido en un indicador de coyuntura aplicable en distintos dominios, como es el caso del comercio al por menor y del turismo.

En efecto, aunque representa una perspectiva de análisis diferente, si se compara la evolución de las ventas diarias con el gasto con tarjetas de crédito en todos los establecimientos de comercio minorista, la correlación entre las dos series es muy elevada.

Naturalmente, las ventas minoristas representan una perspectiva de oferta y únicamente relacionada con cierto tipo de establecimientos mientras que las tarjetas representan una perspectiva de demanda, si bien únicamente relativa a las compras realizadas con este medio de pago.

Pero también esta información de medios de pago ha servido para realizar otra estadística experimental en el ámbito del turismo. La Encuesta de Gasto Turístico de no residentes (EGATUR) que realiza mensualmente

el INE se basa en una operación muestral (junto con la Estadística de Movimientos Turísticos en Fronteras, FRONTUR) que ofrece información del gasto realizado por los no residentes en sus viajes y excursiones a España atendiendo a su destino principal. Naturalmente, la dimensión de la muestra y el hecho de considerar únicamente el destino principal del viaje permite desagregar el gasto según la nacionalidad del turista con un desglose no muy elevado de países (seis) y únicamente para las seis comunidades autónomas más turísticas. Utilizar esta información muestral conjuntamente con los datos de gasto realizado por no residentes con tarjetas de crédito y débito, emitidas fuera de España, ha permitido ampliar de forma considerable el desglose de la información ofrecida por EGATUR. Además, el hecho de que se atienda únicamente al destino principal del viaje hace que todos los gastos se asignen a dicho destino, quedando fuera del ámbito de la estadística el gasto realizado en otros puntos que no corresponden al destino principal del viaje y penalizando informativamente a aquellas comunidades que no son destinos principales de los viajes de los no residentes.

De esta manera, con periodicidad trimestral, el INE ofrece información de gasto para todas las comunidades autónomas españolas con un desglose de 34 nacionalidades y zonas geográficas.

Así, por ejemplo, se ha podido estimar el gasto turístico de no residentes, en la Comunidad Autónoma de La Rioja, usualmente no destino principal de viajes de no residentes, pero territorio en el que aquellos realizan gasto en excursiones o en el transcurso de sus etapas. Así, en 2020, una quinta parte de dicho gasto lo realizan ciudadanos franceses, seguidos de los estadounidenses o los del Reino Unido, nacionalidades más representativas. Si bien se ha estimado el gasto también de ciudadanos de los Emiratos Árabes Unidos, de Turquía o de Corea (en torno al 0,2 % del total en cada caso), incluso el gasto también de ciudadanos de Filipinas, India, Arabia Saudí o Israel, aunque dichos datos quedan cubiertos por el secreto estadístico debido al escaso número de observaciones.

Un fenómeno turístico que ha evidenciado un gran desarrollo en los últimos años y del que no existía información estadística directa es el relativo a las viviendas de uso turístico. Esta actividad forma parte de la denominada economía colaborativa, en la que los propietarios individuales de viviendas alquilan la totalidad de la vivienda o parte de ella a otros usuarios, anunciándose a través de plataformas privadas de alojamiento turístico.

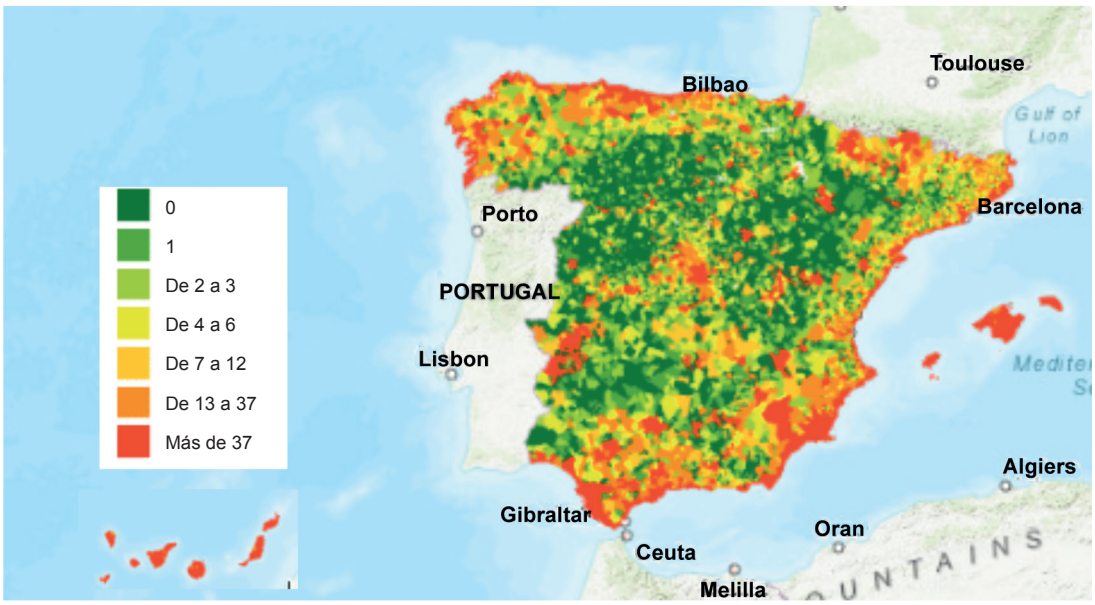
El INE ha puesto en marcha una estadística experimental en la que, accediendo a las bases de datos de las principales plataformas privadas que se anuncian en España, a través de técnicas de *web scraping*, obteniendo información del número de alojamientos de vivienda turística, así como de su capacidad. La información obtenida es tremendamente granular, pues llega al nivel, no solo de municipio, sino de sección censal. La Figura 5 muestra, por tonos de colores, las viviendas turísticas a nivel nacional, pudiéndose contemplar los territorios en los que más abunda este tipo de alojamiento, cerca de las costas, en los dos archipiélagos y las grandes ciudades en el interior.

El proceso de obtención de datos de la web es muy minucioso en la descripción de cada una de las viviendas, con el objeto de definirla e identificarla unívocamente del resto. Este proceso es muy importante, dado que un propietario puede anunciar la misma vivienda en dos o más plataformas. Así, se ha diseñado un proceso de eliminación de duplicados para que la estimación final del número de viviendas fuera robusto.

Como resultado, en agosto de 2020 se estimaron más de 320.000 viviendas turísticas, albergando más de 1,6 millones de plazas. Alicante, Málaga e Illes Balears fueron las provincias con un mayor número de viviendas ofertadas. En los municipios de La Oliva (Fuerteventura), Pollença (Illes Balears) y Begur (Girona) más del 20 % de las viviendas del municipio eran viviendas turísticas. En el Distrito Centro de Madrid, casi el 9 % de sus viviendas son turísticas (28 % en las inmediaciones de la Puerta del Sol). En el Distrito de la Ciutat Vella de Barcelona, más del 7 % (41 % en las inmediaciones de la Plaza Catalunya).

FIGURA 5

VIVIENDAS TURÍSTICAS POR MUNICIPIO
(Febrero de 2021)



FUENTE: INE.

Está previsto utilizar también esta técnica de *web scraping* para la obtención automática de precios, desde la web, para la elaboración del IPC.

También en el ámbito del IPC, en 2020 y fruto de acuerdos con determinadas grandes cadenas de supermercados e hipermercados, se comenzaron a utilizar los denominados *scanner data* en el proceso de elaboración del índice, es decir, utilizando la información procedente de las bases de datos de las empresas, que contienen el registro de las ventas realizadas de cada uno de sus productos en la línea de caja de todos sus establecimientos (unidades vendidas e ingresos). La información que se incorporará, de este modo, al IPC, es la relativa a productos de gran consumo (alimentos envasados, bebidas, productos de limpieza y mantenimiento del hogar, productos de parafarmacia, alimentos y productos para mascotas y artículos de cuidado personal). De esta manera, la incorporación de estos datos procedentes de bases privadas permitirá la eliminación de la recogida de precios en aquellos establecimientos de las cadenas incorporadas, agilizando la elaboración del índice y mejorando su calidad y oportunidad. Esta mejora metodológica se incorporará al IPC en el marco de la nueva base 2021 del índice.

4. Sostenibilidad de las estadísticas experimentales

Pasado un tiempo desde el inicio de la crisis y después de haber abordado las necesidades más urgentes de los usuarios, es tiempo de monitorizar la recuperación económica. Esta tarea no solo se debe realizar

mirando únicamente a los indicadores de la estadística tradicional, sino que debe ser apoyada e impulsada por las estadísticas experimentales desarrolladas y ejecutadas durante este periodo de crisis.

Monitorizar la recuperación significa elaborar tableros de indicadores, representativos de los diferentes dominios, que deben estar compuestos por estadísticas oficiales tanto como sea posible, para garantizar la comparabilidad, pero también por indicadores que, bajo el apelativo de estadísticas experimentales han servido a los usuarios durante los pasados meses, estadísticas basadas tanto en información pública como privada.

Parece claro que las estadísticas experimentales deberían evolucionar en tres diferentes líneas de cara a ser utilizadas en el ámbito socioeconómico. Por un lado, se debe analizar cuáles deben adquirir cierta estabilidad, pudiendo convertirse en estadísticas oficiales o no. Por otro lado, analizar hasta qué punto, algunas estadísticas experimentales pueden servir como indicadores auxiliares para estimaciones rápidas de indicadores coyunturales (*nowcasting*). Finalmente, una tercera línea en la que analizar la posibilidad de que algunas estadísticas oficiales existentes puedan modificar sus diseños como operaciones, combinando elementos tradicionales como el muestreo con elementos experimentales.

En cualquiera de los casos, hoy en día parece más segura la sostenibilidad de estadísticas experimentales basadas en la explotación de información administrativa. No obstante, para las operaciones basadas en información privada, merecerá la pena esperar a la entrada en vigor de la Ley de Datos Europea (*European Data Act*) que podría ayudar a conseguir también la estabilidad en estos casos.

5. Conclusiones. Lecciones aprendidas

Hoy en día y echando la vista atrás, podemos decir que el INE, así como otras oficinas de estadística, ha mostrado una gran resiliencia, una gran capacidad para sobrevivir, aprender y reinventar ideas en un contexto muy difícil. Institucionalmente, ahora es momento de aprender de todo lo bueno que se ha hecho. Por un lado, eliminar las vulnerabilidades que teníamos al principio de la pandemia, siendo capaces de reaccionar muy rápidamente a las demandas de los usuarios y producir otros datos diferentes de los tradicionales, más rápidamente, con mayor frecuencia e innovando en fuentes estadísticas, con el objetivo de facilitar la toma de decisiones basada en datos. Pero, además, mantener e incluso reforzar las fortalezas que tenemos como oficina de estadística y que nos hemos ganado durante muchísimos años. Todo ello con el fin de decidir cuál debe ser el rol que debe desempeñar el INE en el futuro, más allá de la producción estadística.

Hemos de ser conscientes de que, durante este tiempo, el INE ha podido acceder a todos los registros administrativos a los que ha solicitado acceso para ofrecer información a los usuarios.

Además, hemos podido convenir con algunas empresas del sector privado (operadores de telefonía para disponer de datos de posicionamientos móviles; grandes cadenas comerciales, para disponer de *scanner data,* con las que mejorar las estadísticas de consumo y precios; y operadores de tarjetas de crédito y débito) en muchos casos atendiendo a su responsabilidad social corporativa, para tener acceso a información de sus grandes bases de datos con el objetivo de elaborar nuevas estadísticas.

Hemos podido acceder a toda esta información porque la oficina de estadística es una institución independiente e imparcial, que genera confianza, no solo en los gestores o propietarios de las bases de datos sino en la sociedad en general, pues conocemos perfectamente el valor de los datos y sabemos cómo preservar su confidencialidad. También los investigadores, como potenciales usuarios directos de los datos, nos ven como el tercero seguro que puede proporcionarles información confidencial de identificación indirecta para sus investigaciones, que no pretenden otra cosa que generar valor añadido para la sociedad.

Así, más allá de nuestro papel como productores de estadísticas, nos hemos convertido, de facto, en un gestor de datos de la Administración pública y ese debe ser el reto que guíe el comportamiento del INE en el futuro. Deberemos revisar nuestro marco legal para realizar esta labor, pero merecerá la pena hacerlo, pues son muchos los usuarios que confían en nosotros para gestionar los datos como bien público (*policy-makers*, académicos, investigadores, científicos, periodistas, etc.).

De no haber existido esta pandemia, es posible que las oficinas de estadística se hubieran movido en el medio plazo hacia este nuevo rol, quién sabe, pero lo que es cierto es que la pandemia nos ha dado una gran oportunidad para evolucionar rápidamente en este sentido y permitir convertirnos en el gestor de datos de la Administración, el *data steward*.

No hay vuelta atrás.

Referencias bibliográficas

ESS, European Statistical System. (2021). *European Statistical System (ESS) position paper on the future Data Act proposal. Access to privately held data is urgently needed for producing new, faster, more detailed official statistics*. https://ec.europa.eu/eurostat/web/european-statistical-system/governance-bodies/essc

European Commission. (2021). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Drawing the early lessons from the COVID-19 pandemic.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0380

INE, Instituto Nacional de Estadística. (2020a). *Comunicado relativo a la actividad del INE ante la emergencia sanitaria con motivo del COVID-19.* https://www.ine.es/ine/comunicado_ine_covid19.pdf

INE, Instituto Nacional de Estadística. (2020b). *Estudio EM-1 de movilidad a partir de la telefonía móvil.* Proyecto técnico. https://www.ine.es/experimental/movilidad/exp_em1_proyecto.pdf

INE, Instituto Nacional de Estadística. (2020c). *Medición del comercio diario al por menor de grandes empresas (CDMGE)*. Proyecto Técnico. https://ine.es/experimental/cdmge/

INE, Instituto Nacional de Estadística. (2020d). *Medición del número de viviendas turísticas en España y su capacidad*. Proyecto Técnico. https://ine.es/experimental/viv_turistica/experimental_viv_turistica.htm

INE, Instituto Nacional de Estadística. (2020e). *Distribución del gasto realizado por los visitantes extranjeros en sus visitas a España según comunidad autónoma de destino real y país de residencia*. Proyecto Técnico. https://www.ine.es/experimental/gasto_tarjetas/exp_gasto_tarjetas_proyecto.pdf

INE, Instituto Nacional de Estadística. (2021a). *Estimación del número de defunciones semanales durante el brote de Covid-19 (EDeS)*. Proyecto Técnico. https://ine.es/experimental/defunciones/experimental_defunciones.htm

INE, Instituto Nacional de Estadística. (2021b). *Coyuntura Demográfica de Empresas (CODEM)*. Metodología y Líneas Operativas. https://ine.es/experimental/codem/exp_codem_proyecto.pdf

INE, Instituto Nacional de Estadística. (2021c). *Estudios de movilidad a partir de la telefonía 2020-2021*. Proyecto Técnico. https://www.ine.es/experimental/movilidad/experimental_em4.htm

INE, Instituto Nacional de Estadística. (2021d). *Estimación mensual de nacimientos (EMN).* Proyecto Técnico. https://www.ine.es/experimental/nacimientos/experimental_nacimientos.htm