



Índice

REVISTA DE ESTADÍSTICA Y SOCIEDAD

JULIO 2022

86



Estadísticas COVID-19

Entrevista a Luis Enjuanes

Las epidemias y pandemias han emergido a lo largo de los años y podemos garantizar que seguirán emergiendo, lo que implica que tenemos que estar preparados

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 Póo

Luis María Sáez de Jáuregui

Lis Paula San Miguel-Pradera

Lázaro Villada Ruiz

Colaboradores

Sergio Alonso

Enric Álvarez-Lacalle

Miguel Ángel de Castro Puente

Martí Català

J. Alberto Conejero

Alfredo Cristóbal Cristóbal

Concha Delgado

Luis Enjuanes

Adolfo Gálvez Moraleda

Santi García Cremades

Margarita García Ferruelo

Iñaki Hernández

Amparo Larrauri

Inmaculada León

Daniel López

Eliseo Navarro

Nuria Oliver

Clara Prats

Pilar Requena

Rosa Ruggeri Cannata

Proyecto gráfico y cubierta: J. A. Alcalá

Composición y maqueta: 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.boe.es/>

Acceso a la edición electrónica:
www.revistaindice.com (ISSN 1697-2325)

La revista Índice se edita mediante un convenio entre:



Colaboran:



Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE TRANSFORMACIÓN ECONÓMICA, INDUSTRIA,
COMERCIO Y UNIVERSIDADES



Instituto de
Estadística de
Castilla-La Mancha



Junta de
Castilla y León

Corporación Económica Nacional
Dirección General de Presupuestos y Estadística



Generalitat de Catalunya
Institut d'Estadística
de Catalunya



INSTITUTO POLITÉCNICO DE
ESTADÍSTICA

Depósito Legal: M-46988-2003 - ISSN: 1696-8358 - N.º P: 926-20-011-6

*“El alma que hablar puede por los ojos,
también puede besar con la mirada”*
Gustavo Adolfo Bécquer

Santi García Cremades La Nota	2	<p>Ante la aparición del Coronavirus y la posterior pandemia mundial, la estadística pública y privada han respondido adaptando su forma de producir información y la propia información que generan, para ofrecer información actualizada y comprensible. Superados los tiempos más trágicos de la enfermedad, ha llegado el momento de analizar los elementos más significativos con los que la estadística ha respondido a estos retos.</p> <p>En la Nota, Santi García Cremades, matemático, profesor y divulgador, nos introduce sobre la complejidad de la medición del fenómeno de la pandemia. Entrevistamos en este número a Luis Enjuanes, Jefe del Laboratorio de Coronavirus en el Centro Nacional de Biotecnología del CSIC, sobre la situación actual de la pandemia, la vacuna que están desarrollando y lo que esta pandemia debería habernos enseñado. Los primeros artículos de esta edición, firmados por Miguel Ángel de Castro, Director General de Coordinación Estadística y de Estadísticas Laborales y de Precios, y Alfredo Cristóbal, Director General de Productos Estadísticos, del INE, nos explican los cambios organizativos y de producción que se han llevado a cabo en el INE en respuesta a la COVID-19 y cómo estos cambios han supuesto un impulso a la innovación estadística. A continuación, Rosa Ruggeri, desde Eurostat, nos detalla la información ofrecida por el European Statistical Recovery Dashboard. El artículo de Inmaculada León, Concha Delgado y Amparo Larrauri del Centro Nacional de Epidemiología del Instituto de Salud Carlos III, versa sobre la metodología del sistema de monitorización de la mortalidad diaria por todas las causas (MoMo) y su utilidad durante la pandemia. Margarita García, subdirectora general de Estadísticas Sociales Sectoriales del INE, nos explica los cambios derivados de la pandemia en la Estadística de Causas de Muerte. Desde BIOCÓM-SC y la Universidad de Oxford, Daniel López, Sergio Alonso, Martí Català, Enric Álvarez-Lacalle y Clara Prats nos detallan los modelos utilizados en su grupo de investigación para el seguimiento de la pandemia. También los investigadores que forman la iniciativa Valencia IA4COVID, J. Alberto Conejero y Nuria Oliver, entre otros, han realizado modelos predictivos que han servido de base para la toma de decisiones por parte de las AA. PP. A continuación, Eliseo Navarro y Pilar Requena de la Universidad de Alcalá, analizan cómo se han visto afectadas las tasas de mortalidad durante estos dos años. El artículo de Iñaki Hernández, Coordinador Sección de internet y EpData de Europa Press, describe la evolución de su base de datos para informar sobre la evolución de la pandemia. Por último, el artículo de Adolfo Gálvez, Subdirector General de Difusión Estadística del INE, detalla las estadísticas experimentales que se han desarrollado en el INE que se han visto impulsadas por las necesidades de información derivadas de la COVID-19. ●</p>
Entrevista a Luis Enjuanes	6	
Miguel Ángel de Castro Puente La respuesta del INE ante la crisis COVID-19. Cambios organizativos y de producción	9	
Alfredo Cristóbal Cristóbal La respuesta del INE ante la crisis COVID-19: impulso a la innovación estadística	13	
Rosa Ruggeri Cannata The European Statistical Recovery Dashboard	16	
Inmaculada León, Concha Delgado y Amparo Larrauri MoMo: una herramienta esencial durante la pandemia de COVID-19	20	
Margarita García Ferruelo La Estadística de Causas de Muerte en el contexto de la pandemia	23	
Daniel López, Sergio Alonso, Martí Català, Enric Álvarez-Lacalle y Clara Prats Matemáticas y COVID-19: monitorización, predicción, retos y perspectivas	28	
J. Alberto Conejero y Nuria Oliver La iniciativa Valencia IA4COVID	33	
Eliseo Navarro y Pilar Requena Tasas de mortalidad de la población española en 2020: efectos de la pandemia por covid-19 y de las medidas para contenerla	36	
Iñaki Hernández EpData: informando sobre el coronavirus	40	
Adolfo Gálvez Estadísticas experimentales en tiempos de COVID-19	43	
Reseña de publicaciones	48	

Historia de números en una pandemia

Santi García Cremades

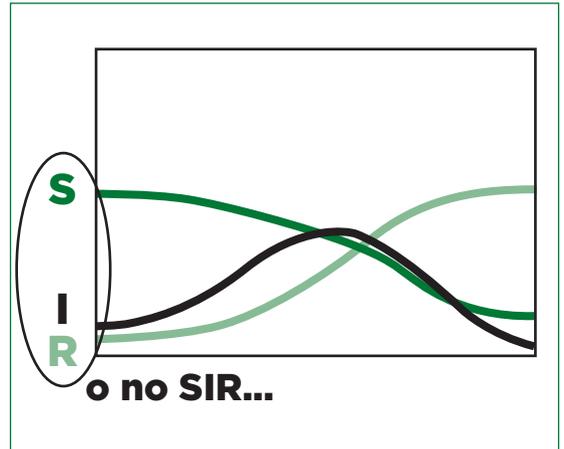
Matemático, profesor y divulgador científico

Modelos por doquier

Modelos hay infinitos, y la pregunta evidente es: ¿qué modelos usar en una epidemia? El modelo SIR fue el primero en levantar la mano, por razones claras, es el modelo que distribuye una población en tres categorías: Susceptibles, Infectados y Recuperados. Fácil, ¿no? Es un modelo basado en ecuaciones con derivadas, propuesto en 1926 por el escocés Anderson Gray McKendrick para estudiar la malaria. Hay muchas variantes, siendo el modelo SEIR (con la E de Expuestos) el más presente, pudiendo analizar a nivel teórico los efectos de las medidas restrictivas en el resto de variables. Este modelo tiene el problema evidente que si no tienes una buena estimación de los infectados (la población I), no obtenemos la información más relevante: cuánta gente puede estar inmunizada.

Los datos nos hablan, todos los días, a todas horas. Otra cosa es que no queramos oírlos, o seamos miopes a ellos y los veamos borrosos. Siempre me gustó ese concepto: las matemáticas son las gafas que nos faltan para ver con nitidez el mundo que nos rodea

Puesto que en tiempo real el modelo SIR parecía tener sus debilidades, salieron de golpe todos los modelos matemáticos que podía aportar cada familia. Podríamos destacar tres



grandes categorías: las cadenas de Markov, las series temporales y el *machine learning*.

Las **cadenas de Markov** tienen origen a principios del siglo XX, dentro de la teoría de la probabilidad. El nombre viene de Andrei Markov cuyo interés al crear las cadenas no tenía conexión con ningún fin práctico, salvo en otra de sus pasiones: la poesía. Es un método basado en probabilidades pasadas, para ponderar la probabilidad de un evento en el presente o futuro. Es un método muy aplicado para realizar predicciones como en meteorología, los buscadores de internet como Google o, claro está, en epidemias.

Las **series temporales** se explican por sí mismas: son series de datos en un intervalo de tiempo. Podemos marcar el comienzo del uso de estas técnicas en los años 20 del siglo XX, con George Udny Yule y con el *crack* del 29, donde saber cómo evoluciona una variable que se ve afectada por cientos de ellas tomó una gran importancia en economía (cuando empezaría la econometría). En estos modelos se pueden aplicar decenas de versiones: aditivos, multiplicativos, autorregresivos, con distintas frecuencias, etc. Este ha sido la base del modelo que utilicé para predecir la curva de fallecidos de abril a mediados de mayo, con mejor aproximación que el resto de métodos.

El **machine learning** es la herramienta más reciente y la menos matemática. Aunque Geoffrey Hinton introdujera el concepto por 1986, la aplicación y desarrollo han sido posteriores. Hay multitud de métodos, todos basados en el aprendizaje desde el entrenamiento de la máquina, es decir, de los datos pasados.

Con todos estos modelos, nos encontramos con distintas predicciones, cada una con un error distinto. Ahora, otra pregunta es ¿qué medir? ¿Qué variable? ¿Cuáles son los datos que más información nos dan? Los datos nos hablan, todos los días, a todas horas. Otra cosa es que no queramos oírlos, o seamos miopes a ellos y los veamos borrosos. Siempre me gustó ese concepto: las matemáticas son las gafas que nos faltan para ver con nitidez el mundo que nos rodea. No sé si tatuarme la frase o dejarla aquí, pero el concepto me parece acertado. Ahora, más que nunca, se habla de datos por todos lados. Tenemos los números cada día de infectados, fallecidos, recuperados, personas en la UCI, hospitalizados... y todo esto, por países, por comunidades, por provincias, incluso desagregados por edades. Tenemos muchos datos y, por tanto, mucho que estudiar. Vamos a ponernos esas gafas y entender qué

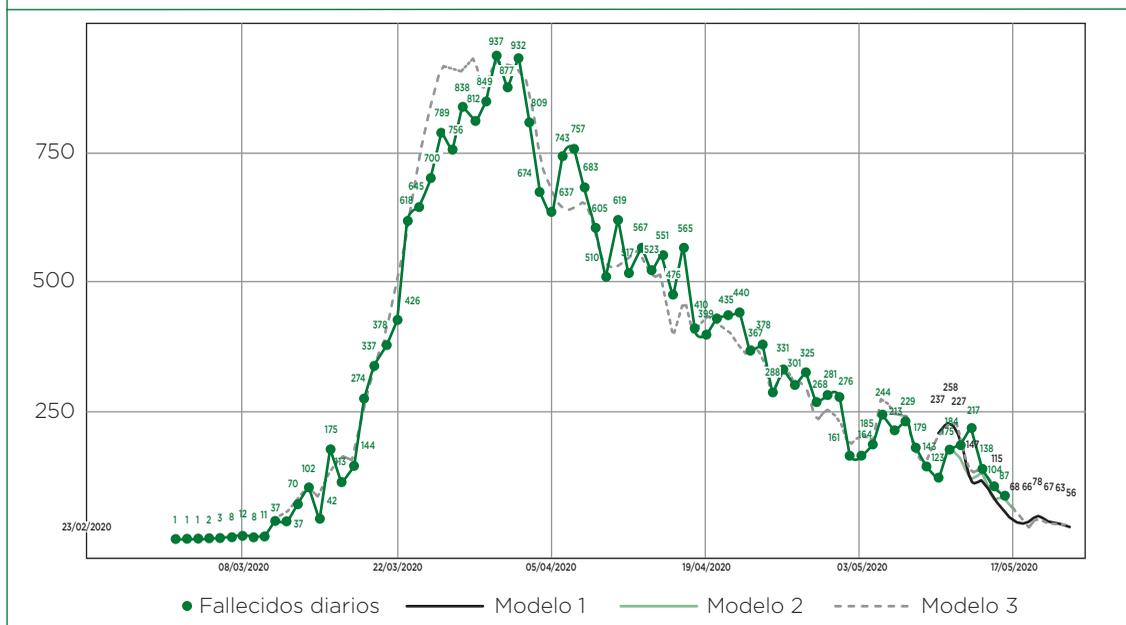
datos son claves en esta Guerra Mundial contra el SARS-CoV-2.

Exponencial y el país de las maravillas

Hay curvas por todos lados. A mí mismo me han llamado el “chico de la curva”, no sé si será por los kilos que estoy cogiendo (ya los soltaré por algún sitio). Es un tema complejo y, sobre todo, dramático, pero si algo nos han enseñado nuestros mayores es a aprender de la experiencia y a recuperarnos de los golpes de la vida. Entonces, vamos a aprender de los errores cometidos, de lo que se ha comunicado mal o no se ha entendido bien. Nos fijamos en las curvas de datos sanitarios. Empecemos primero por el inicio de la pandemia, el momento **exponencial**.

Al principio de un contagio, todas las cifras suben rápido, se duplican, se triplican e incluso se cuatuplican en un solo día, y eso no puede ser para siempre. A eso le llamamos el factor de contagio, en epidemiología se conoce como R_0 (el número de personas que se estiman que contagia cada persona infectada). Si ese factor de contagio fuese constante durante toda la pandemia, estaríamos perdidos, y estaríamos

FIGURA 1. SERIE TEMPORAL SOBRE LOS FALLECIDOS DE LA COVID-19 EN ESPAÑA



Fuente: Santi García Cremades.

hablando de una función exponencial. Se dice mucho la palabra “virus” y la palabra “viral”. Pero, ¿cómo es algo viral? ¿Crece tanto el virus como pensamos? Hablemos de exponencial. Sí, de la función exponencial, que tanta lata nos da en el Instituto. Si hay algo que es útil en Matemáticas es dar certeza de lo que uno puede intuir. Si hablamos de la magnitud de la función exponencial, no hay nada como poner un ejemplo visual para hacernos una idea. Es un ejemplo clásico, pero nos vamos a sorprender. Si tienes hipo, tranquilo que se te va a quitar.

Coge un folio. Coge el primer folio en blanco del libro, adelante, no estoy mirando. El grosor del folio es aproximadamente de 0,1 milímetros. Vamos a intentar doblarlo ahora. Una vez. Dos veces. Tres veces... Ya cuesta más. Se creía que ocho veces es lo máximo que podrías doblar el papel. Pues es mentira. La americana Britney Gallivan, cuando era tan solo una estudiante de secundaria, en 2002, demostró una fórmula que determinaba el tamaño del papel para poder doblarlo n veces, y hasta realizó la prueba doblando un papel doce veces. El récord actual es de trece dobleces (gente aburrida, qué sé yo). Vamos a ver qué anchura va cogiendo el folio según lo van doblando. Es sencillo, a cada doblado se dobla la anchura anterior, así que es una función exponencial en base 2. ¿Cuánto crees que llegaría el folio que tienes en la mano si pudieses doblarlo 22 veces (que por poder, no se puede, la física y sus límites)? Pues sería más alto que la Torre Eiffel, que mide 300 m de altura.

Aún tienes hipo, parece... A ver, ¿y si doblamos 42 veces el folio? Ahora la anchura es igual a 2 potencia 42 por la anchura del folio. Coge calculadora, que de cabeza no vas a poder. Te espero. Tendríamos ahora un total de 4.398.046.511.104 capas, más de 4 billones de capas (con “b”), equivalente a 439.805 kilómetros. Son muchísimos kilómetros, si dejásemos ese folio en el suelo, llegaríamos hasta a la Luna, que está a 384.400 km aproximadamente. Si quieres exponencialmente, podrás conseguirle a la persona amada la Luna, con tan solo 42 pasos. Y ya para rematar, si doblamos el folio 103 veces, el grosor sería mayor que el diámetro del Universo observable (93.000 millones de años luz) ...

¡Vaya poder, la exponencial! El coronavirus, sin embargo, parece crecer de manera mucho

más lenta que la exponencial, gracias a los sistemas estatales de Salud Pública y a la OMS, gracias a que tenemos defensas inmunitarias y a que tomamos prevenciones (que no es lo mismo que despertar el caos). Dejemos lo exponencial únicamente en el terreno matemático, por favor. Por poner un ejemplo, si cada contagiado hubiera contagiado a 2 personas cada día, en 33 días estaríamos más de 8 mil millones de personas contagiadas, más gente que hay en la Tierra. Entonces, vale, NO es exponencial, el factor de contagio es dinámico. Primer aprendizaje: crecíamos rápido, pero no exponencialmente.

Confinamiento “rules”

Hablemos ahora del confinamiento, aunque prometo no hablar del “Resistiré”. Según estadísticas de movilidad que sacó Google en marzo y abril, estuvimos entre los 4 países del mundo que más han respetado el confinamiento. Y eso tiene un efecto claro: el famoso **aplanamiento de la curva**. Está claro que necesitábamos aplanar la curva para no saturar el sistema sanitario y poder atender a los casos más graves. Hemos conseguido, entre todos, que unas cifras que crecían muy rápido (no exponencialmente, pero rápido) se ralentizaran. Esto se puede medir a través de una medida clásica en estadística, la curtosis. La curtosis mide cuánto de afilada o de aplanada es la curva que tenemos. Pues si nos confinamos el día 14 de marzo (el día Mundial de las Matemáticas, te recuerdo), el día 24 de marzo vimos un efecto directo en esta medida. La curtosis pasó de decirnos que estábamos en una curva afilada a un momento aplanado de la curva de infectados o fallecidos diarios. Por tanto, *QED*, queda demostrado, conseguimos aplanar la curva. Primero crecer rápido, y luego lento.

Ahora toca el momento clave, la luz al final del túnel. La curva no dejó de crecer hasta el 31 de marzo de 2020, lo que conocemos como el **punto máximo de la curva**. En modelos epidemiológicos, se entiende que sería el punto de inflexión de la pandemia, “de ahí p’abajo”. Bueno, eso es una medida a ojímetro, pero ¿podemos saber cuánto queda de estas cifras tan duras? Difícil, pues la curva no es simétrica, no

es una campana de Gauss. Además, el momento alto de las cifras de infectados y fallecidos se ha mantenido más tiempo del que nos gustaría, es una consecuencia de aplanar la curva, que la meseta dura más en el tiempo. Pero la bajada está ahí, la tendencia decreciente también está y las gafas de las matemáticas hacen ver una lucecita al final. Y ahora nos queda un momento clave: la victoria final.

Al final de esa primera ola, estamos ya desesperados, y hay algo más que el coronavirus en nuestras mentes: ¿cómo está afectando esto a nivel económico y psicológico? Aquí entra en juego un concepto que no tiene más de 100 años de formalismo matemático: ecuaciones simultáneas. Como decía Forrest Gump, “la vida es como una red: todo está conectado”. Vale, no es de Forrest, pero es una forma de entender el todo. En una red, cualquier movimiento afecta al total de los nodos de la red. Sin embargo, hay cosas que no se ven afectadas por cambios en la red, y otras que no. Por ejemplo, las horas de Sol afectan la actividad social de un lugar, la actividad social afecta a la economía, a la salud mental, y viceversa: la salud mental afecta a la economía, a la actividad social. Sin embargo, que tú trabajes más o que estés más feliz, poco lo importa al Sol. En este caso, la variable “horas de Sol” sería una variable exógena, una variable pasota, a su bola, y el resto son endógenas, como la red que decía Forrest Gump (en mi fantasía).

En epidemiología, casi todas las variables son endógenas, por tanto, cualquier decisión, tanto restrictiva como permisiva, tiene un efecto en TODAS las variables sanitarias, económicas y sociales. De ahí la complejidad del tema. Estamos en una red, donde si salvas vidas por COVID, causas pérdidas en otros factores. El confinamiento ha provocado una gran depresión económica y social, y para afrontar esos problemas, provocamos otras olas pandémicas.

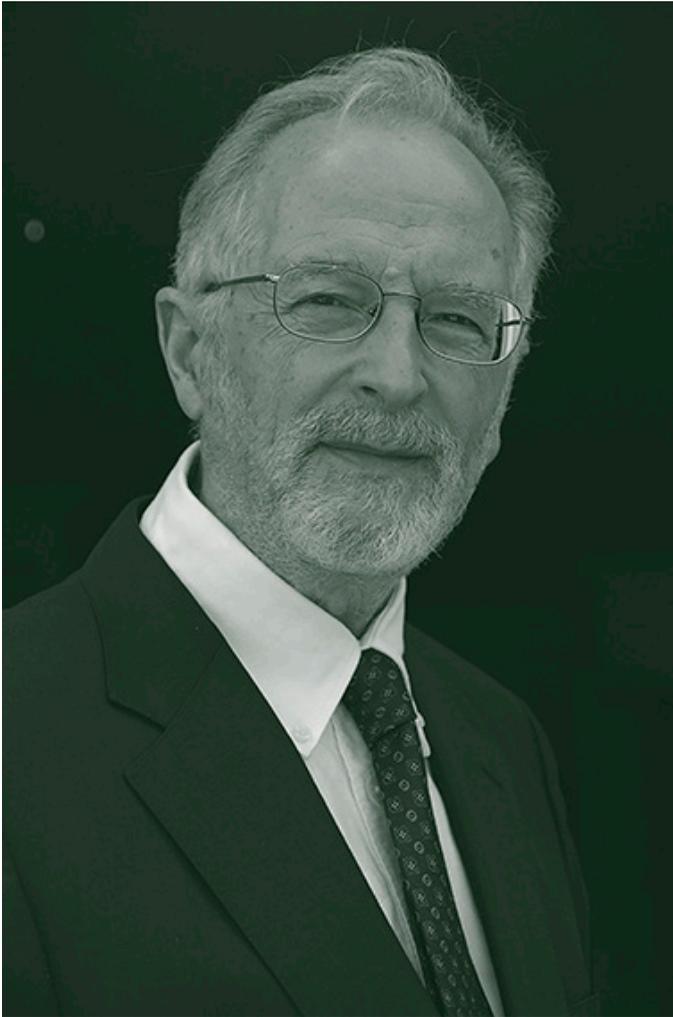
Olas van, olas vienen

Es natural bajar la guardia, en situación estable, cuando ya no se habla de datos dramáticos, buscamos también la vida pre-pandémica, la llamada “nueva normalidad”. Y este es



un problema para los matemáticos, pues el movimiento de poblaciones es un movimiento casi impredecible, lo que conocemos como movimiento caótico. En julio y agosto tuvimos el principio de una segunda ola. Bueno, antes habría que definir ola: cambio de tendencia positiva (de tendencia decreciente a creciente, o de tendencia constante a creciente) de la curva de contagios. Y esto ocurrió, y no se volvió a tomar ninguna restricción útil hasta octubre. Ahí hay un problema de asesoramiento matemático claro. No se supo remarcar un umbral de decisión en 3 meses, parecía como si no hubiéramos aprendido de la primera ola. Y, si bien es verdad que las olas pandémicas no tienen nada que ver, tampoco son tan diferentes estadísticamente hablando.

La clave está siempre en saber qué medir, cómo medir y, posteriormente, saber interpretar esas medidas. Algo estamos aprendiendo, seguro, sobre todo el gran valor que tiene la ciencia y, en particular, la ciencia más básica de todas: las **Matemáticas**. ●



ENTREVISTA

Luis Enjuanes

“LAS EPIDEMIAS Y PANDEMIAS HAN EMERGIDO A LO LARGO DE LOS AÑOS Y PODEMOS GARANTIZAR QUE SEGUIRÁN EMERGIENDO, LO QUE IMPLICA QUE TENEMOS QUE ESTAR PREPARADOS”

Dr. Enjuanes, los datos del coronavirus arrojan, todavía, cifras complejas. Los datos no parecen certificar que hayamos salido de la pandemia, pero, sin embargo, las administraciones y la población parecen no prestar demasiada atención a los datos. ¿Nos hemos cansado de estar permanentemente vigilantes? ¿En qué situación nos encontramos verdaderamente?

Probablemente la decisión de liberar a la población de medidas estrictas, como reducir el número de personas que acuden a una reunión o eliminar el uso de mascarillas en la mayor parte de los casos se deba a que la nueva variante del virus, la Omicron, que se ha extendido por todo el mundo de forma dominante, se comporta como atenuada, sobre todo en las personas que ya han sido vacunadas. Las autoridades

creen que ahora es más ventajoso dejar que esta variante se extienda y termine de inmunizar a toda la población. Claro que esto representa un cierto riesgo, particularmente para los no vacunados. Esta forma de proceder se ha impuesto en todos los países más adelantados.

De algún modo, parece obvio que la salida de la pandemia se celebra, sobre todo, en contextos occidentales o desarrollados. Sin embargo, una circunstancia de este tipo solo admitiría una aproximación global. ¿Hay una fecha previsible para poder confirmar que estamos fuera de peligro o, al menos, fuera del peligro que genera este virus?

Tal como se sugiere, para los países con una baja tasa de vacunación, no se pueden tomar

las mismas medidas. Se les tendría que ayudar a completar la vacunación de su población. Realmente fuera de peligro no estamos en ningún sitio, porque Omicron todavía puede causar patologías graves, pero la situación actual no tiene nada que ver respecto a cuando se inició la pandemia, un momento en que no había nadie vacunado.

La vacuna en la que trabajan usted y su equipo presenta muchas ventajas con respecto a las ya existentes. ¿Cuáles son esas características específicas y qué utilidad pueden tener a futuro?

Las ventajas de nuestra vacuna, según se ha evaluado en ratones y más recientemente en hámsteres, es que la inmunización intranasal con una sola dosis da una inmunización esterilizante, esto es, que los animales vacunados rechazan por completo la reinfección con dosis letales del virus, que son eliminadas en la puerta de entrada de nuestro organismo. Ello probablemente se debe a una combinación de factores, entre los que se encuentran: (i) la administración intranasal. Estos virus son respiratorios y nos invaden por la nariz, preferentemente. Por ello hay que inducir una inmunidad en mucosas respiratorias, algo que solo se induce administrando el antígeno localmente, en esas mismas mucosas, como hacemos en nuestro laboratorio; (ii) nuestra vacuna amplifica la cantidad de RNA que administramos, porque se trata de un RNA autoreplicativo que puede multiplicar el número de copias administradas por mil; y (iii) esta vacuna expresa varios antígenos del virus, no solo la proteína S, que es la única que llevan otras vacunas.

Del COVID-19 se han hecho lecturas políticas, sociales y hasta culturales, pero... ¿cuál es la principal enseñanza que, desde la ciencia, podemos extraer de estos años de pandemia?

La primera es que, como ha demostrado la historia científica, las epidemias y pandemias han emergido a lo largo de los años y pode-

mos garantizar que seguirán emergiendo, lo que implica que tenemos que estar preparados. ¿Cómo? Con redes de diagnóstico en los lugares de entrada más frecuentes, como China, África, Indonesia, etc. La segunda es aumentar el respeto por los medios naturales. La mayor parte de las epidemias se originan por brotes zoonóticos (paso del virus de otro mamífero inferior al hombre). La invasión de los espacios naturales facilita el contacto de las personas con los animales silvestres y el inicio de las epidemias.

La mayor parte de las epidemias se originan por brotes zoonóticos (paso del virus de otro mamífero inferior al hombre). La invasión de los espacios naturales facilita el contacto de las personas con los animales silvestres y el inicio de las epidemias

¿Por qué cree que en contextos desarrollados y con una aparente cultura científica suficiente han surgido tantos movimientos conspirativos y sospechosos con los resultados científicos?

Creo que convivimos personas con todo tipo de creencia religiosas y culturales que llevan en más casos de los deseados a posiciones personales por las que algunas personas han arriesgado sus vidas en guerras religiosas o civiles, o ante peligros por infecciones mortales para el hombre. El problema delicado en estas situaciones es cómo hacer que las personas adopten la postura que más les beneficia a ellos, sin coartar su libertad de elección, que frecuentemente representa un peligro para el resto de la sociedad. Estos temas, la circulación de bulos erráticos, y temas relacionados requieren un estudio urgente para establecer

las normas a cumplir y los medios sensatos para actuar por parte de la justicia y las autoridades.

Está demostrado que el potencial intelectual y creativo de los españoles es muy elevado, como se demuestra por la elevada aceptación de nuestros universitarios en los mejores laboratorios y universidades del mundo

Si pudiera formular una petición al regulador, ¿cuál sería?

Reforzar los medios oficiales de información en los canales de alta difusión, para distribuir una información fidedigna basada en principios científicos y éticos correctos. Incorporar a estas tareas a los mejores medios de información y profesionales, dado que son los que más fácilmente llegan a la sociedad.

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.

España en un país avanzado, pero con medios económicos limitados por el reducido potencial industrial de nuestro país. Está demostrado que el potencial intelectual y creativo de los españoles es muy elevado, como se demuestra por la elevada aceptación de nuestros universitarios en los mejores laboratorios y universidades del mundo. Pero el desarrollo industrial requiere buenos economistas y emprendedores, y centrarse en unos pocos objetivos para los que dispongamos de financiación y de alguna ventaja selectiva. Desgraciadamente, el café para todos no es bueno, porque si se defienden muchos proyectos no los podemos llevar adelante. Naturalmente, hay que tener en cuenta el apoyo a las clases más desfavorecidas, que en épocas de crisis siempre son las que corren un mayor riesgo. Dentro de 20 años, si nuestros políticos tienen sentido de Estado, España podría haber dado un salto hacia la modernidad muy importante. Ello requeriría de una elevada colaboración entre los partidos políticos, cosa que en este momento brilla por su ausencia en varios de ellos, unos de gran tamaño y otros más pequeños que, según creen, están llenos de sabiduría. ●

Diego S. Garrocho

LUIS ENJUANES

Luis Enjuanes ha trabajado en el campo de la virología durante más de 40 años, incluyendo más de 35 años en coronavirus. Su interés actual es el estudio de los mecanismos de replicación, transcripción, virulencia e interacción virus-huésped de los coronavirus. Enjuanes ha publicado más de 235 artículos en revistas internacionales y 58 capítulos de libros. Actualmente es Profesor de Investigación y Jefe del Laboratorio de Coronavirus en el Centro Nacional de Biotecnología del CSIC.

Enjuanes ha sido "Fogarty Visiting Fellow" en los Institutos Nacionales de Salud (NIH) en EE. UU., y Científico Visitante en el Centro de Investigaciones sobre el Cáncer (FCRC) de los Institutos Nacionales de EE. UU. (NIH).

Enjuanes es profesor de Virología de la Universidad Autónoma de Madrid y del Instituto Pasteur de París. Ha sido nombrado "Virólogo Senior Distinguido" por la Sociedad Española de Virología, Académico de la Real Academia de Ciencias Exactas, Físicas y Naturales, y Académico de la "Academia Norteamericana de Microbiología" y de la Academia Nacional de Ciencias de EE. UU. Enjuanes ha recibido la "Encomienda de la Orden Civil de Sanidad", la medalla al Mérito en la Investigación y en la Educación Universitaria concedido por el Ministerio de Ciencia e Innovación, y el Premio Nacional de Biotecnología 2022. Enjuanes es "Expert Consultant" de los NIH y de La Organización Mundial de la Salud. Ha sido Editor-en-Jefe de Virus Research.

La respuesta del INE ante la crisis COVID-19. Cambios organizativos y de producción

Miguel Ángel de Castro Puente

*Director General de Coordinación Estadística y de Estadísticas Laborales y de Precios.
Instituto Nacional de Estadística*

La grave situación generada con motivo de la pandemia mundial del año 2020 tuvo un efecto directo en la actividad desarrollada por el Instituto Nacional de Estadística (INE). Dos años después, este artículo repasa algunas de las características de este efecto, así como las medidas adoptadas para minimizar su impacto. Por último, se reflexiona sobre las lecciones aprendidas tras esta experiencia y el trabajo que ha de realizarse para que hechos como este influyan a futuro lo menos posible.

En marzo de 2020, ante las primeras disposiciones de restricción establecidas por el Gobierno, las actuaciones de los órganos de dirección del INE se enfocaron principalmente en adoptar medidas para garantizar las condiciones de seguridad y protección para los trabajadores del INE y, al mismo tiempo, arbitrar procedimientos organizativos que permitiesen cumplir los compromisos de producción estadística en los plazos acordados, tanto con los usuarios nacionales como con el Sistema Estadístico Europeo.

Cambios en los procesos de producción

Para valorar adecuadamente este reto de mantener la producción estadística con el mínimo impacto posible sobre los usuarios, debe tenerse en cuenta que las medidas adoptadas de restricción de movilidad y el cierre de las actividades no esenciales tuvieron consecuencias para todo el proceso de producción, desde la recogida a la difusión de los datos.

El INE utiliza desde hace años una combinación de métodos de recogida que se adapta en

cada caso a las características de la población a investigar y a la naturaleza de las operaciones. Aunque en los últimos años ha ido perdiendo peso la recogida directa de información por agentes entrevistadores, esta modalidad sigue teniendo un papel importante en algunas operaciones de especial relevancia como la Encuesta de Población Activa, donde la primera entrevista a cada hogar de la muestra se realiza de manera presencial. Igualmente una buena parte de la recogida de información necesaria para elaborar el Índice de Precios de Consumo se realiza directamente en los establecimientos.

En el primer estado de alarma y en menor medida en los siguientes, dependiendo del nivel de restricción en cada momento, se hizo un uso intensivo de otros métodos ya implantados en el INE, como la recogida telefónica (CATI) o por Internet (CAWI) que permitieron compensar la falta de recogida directa ante el cierre de establecimientos y las dificultades de localización o contacto con las empresas y los hogares. No obstante, a pesar de estas dificultades y con las soluciones establecidas, debe señalarse que en todo momento la colaboración de los informantes fue muy satisfactoria y una prueba de que la confianza depositada en la institución se ha mantenido intacta en esta pandemia. Por otra parte, la intensificación del uso de los registros administrativos y de las bases de datos privados de empresas sobre las que ya se trabajaba en el INE minimizaron los efectos de la disminución de información directa sobre la que basar las estimaciones.

Cualquier cambio en los sistemas de recogida como los descritos implica un análisis exhaustivo de los datos obtenidos para ver si existe algún efecto de los nuevos factores introducidos. El estudio de posibles sesgos, el refuerzo de las tareas de depuración, el análisis de la coherencia de las estimaciones, etc., supusieron una carga de trabajo añadida en la

fase de tratamiento de los datos que también fue necesario asumir. Estas tareas se realizaron por el personal del INE en condiciones de teletrabajo que en los primeros meses constituía, además, un reto organizativo importante, dado que en aquel momento la implantación de estos sistemas de trabajo en remoto no era una realidad sino que, precisamente, se estaba trabajando en un estudio piloto sobre la factibilidad de su implantación. Este proyecto piloto en el que se trabajaba -controlado, parcial- acabó siendo una experiencia generalizada a todos los trabajadores, con las menores excepciones posibles. Y funcionó.

Durante el tiempo de crisis y desde entonces, la demanda de los usuarios está configurando un nuevo panorama de necesidades estadísticas que se caracteriza por la búsqueda de informaciones más frecuentes, más detalladas y más puntuales

En la fase de difusión de los resultados también fue necesario adaptar los procedimientos habituales de trabajo con el fin de ofrecer toda la información relevante a los usuarios. Se trataba no solo de completar las series de datos sino facilitar los detalles para que los usuarios pudieran establecer comparaciones adecuadas. Por este motivo, durante un tiempo variable, en función de la naturaleza y la evolución de la operación estadística, el INE ha venido ofreciendo en sus notas de prensa información específica de las adaptaciones que se han realizado para apreciar el impacto de la crisis COVID-19 en los métodos de producción o en los resultados obtenidos. Estas publicaciones han ofrecido con transparencia la información disponible en cada momento y más pertinente en cada operación: tasas de respuesta obtenida en algunas encuestas, porcentajes de datos

imputados en el cálculo de algunos indicadores, explotaciones especiales relevantes para el estudio de la COVID-19... Además, se configuró en la web del INE desde las primeras semanas un apartado especial con información estadística disponible sobre aspectos demográficos, económicos y sanitarios sobre la COVID-19.

Todas estas actuaciones y sobreesfuerzos que, como se ve, afectan a todas las fases del proceso de producción estadística se han realizado en coordinación con el Sistema Estadístico Europeo. Más de 30 recomendaciones, guías o aclaraciones sobre el tratamiento y elaboración de indicadores se han publicado con referencia a esta crisis en todos los dominios estadísticos. Este trabajo conjunto ha permitido asegurar la comparabilidad de los resultados y mantener la calidad de las estadísticas en este contexto de máxima dificultad.

Cambios en las demandas de información y en la respuesta

Desde los primeros momentos de esta crisis, se pudo apreciar un incremento en la demanda de datos estadísticos por todo tipo de usuarios. Desde los ciudadanos y empresas que necesitaban datos con los que mitigar algunas de las incertidumbres que se presentaban hasta las instituciones encargadas de la toma de decisiones que solicitaban información sobre las que basar estas decisiones o investigadores que estudiaban el fenómeno desde distintos ámbitos.

Estas peticiones no siempre correspondían a datos disponibles en el sistema estadístico por lo que se identificaron pronto nuevas necesidades. En general, puede decirse que, durante el tiempo de crisis y desde entonces, la demanda de los usuarios está configurando un nuevo panorama de necesidades estadísticas que se caracteriza por la búsqueda de informaciones más frecuentes (estadísticas mensuales, semanales, incluso diarias), más detalladas (especialmente sensible es el detalle territorial de la información) y más puntuales (con menos tiempo desde el periodo de referencia hasta la publicación). Así, junto con las nuevas estadísticas o informaciones que trataban de dar respuesta a estas necesidades, otros indicadores

habituales ofrecían información específica para el contexto vivido, tal es el caso de la EPA o el IPC, que en coordinación con el Sistema Europeo, se centró en la medición de una cesta específica de los artículos esenciales más demandados en aquel momento.

Hay que decir que el proceso desde la identificación de una necesidad de los usuarios hasta la puesta en producción de una estadística oficial requiere habitualmente de un tiempo largo que se emplea para establecer un proceso de producción continua, asegurar la validez de los métodos empleados y garantizar la precisión y la comparabilidad de los datos.

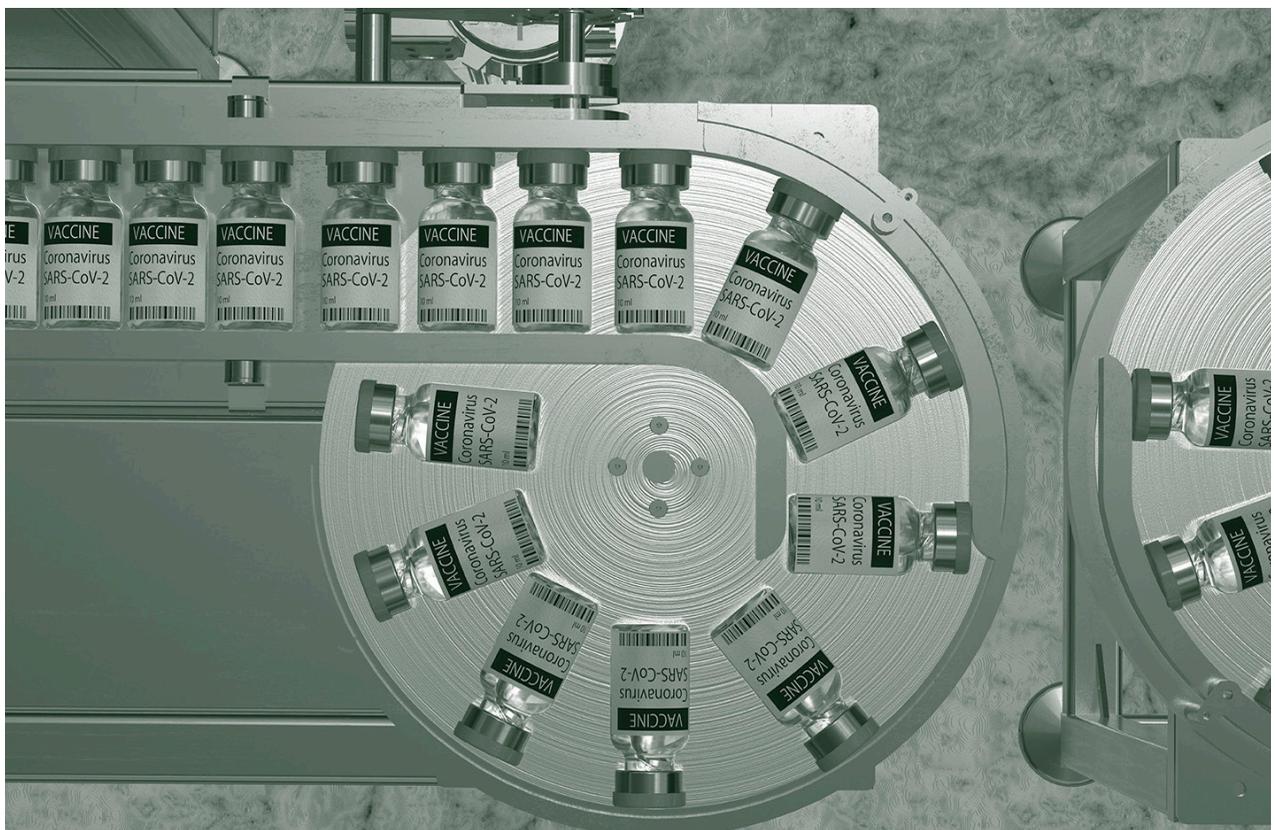
Estos desajustes entre la necesidad de la información y su disponibilidad ya habían sido identificados y señalados y el sistema estadístico ya venía trabajando en promover cambios que permitiesen respuestas más rápidas a nuevas situaciones. Algunas de las acciones que contemplaban este cambio se habían puesto en marcha en el marco de la estrategia denominada Vision 2020 del Sistema Estadístico Europeo y se han acelerado e impulsado en estos dos últimos años. La proliferación de estadísti-

cas experimentales que ponen a disposición de los usuarios resultados novedosos con mayor celeridad es hoy un hecho que permite ganar en agilidad y servicio mientras se consolidan los procesos de la estadística oficial con todas las garantías.

Junto a la producción directa de datos estadísticos, el INE ha colaborado en distintas iniciativas de otras instituciones como han sido el estudio EneCovid promovido por el Instituto de Salud Carlos III, los estudios de movilidad en los periodos de confinamiento promovidos por la Secretaria de Estado de Digitalización e Inteligencia Artificial o apoyo técnico puntual al Ministerio de Sanidad. El esfuerzo de los profesionales del INE en todo momento ha sido encomiable y su capacidad técnica ha quedado demostrada para proveer de datos útiles a la sociedad en circunstancias difíciles.

Lecciones aprendidas

Transcurridos ya más de dos años desde el inicio de esta situación puede mirarse con cierta



distancia y capacidad crítica estas actuaciones y extraer de esta experiencia algunas conclusiones, en muy diferentes ámbitos:

Operativo y logístico

- ▶ El personal de la institución, ante una situación excepcional ha demostrado ser consciente de su responsabilidad y ha realizado un esfuerzo por garantizar el servicio público que presta la Estadística Oficial.
- ▶ El sistema de trabajo en remoto funcionó y contribuyó a garantizar la continuidad de las tareas de la institución. Dicho sistema deberá limar progresivamente las pequeñas aristas que presenta y, con la debida organización teniendo en cuenta diferentes casos especiales, quedará implantado como un sistema de trabajo efectivo, eficiente y con resultados de calidad.
- ▶ Los informantes mantuvieron la colaboración con el INE, en unas condiciones muy difíciles para la mayoría de ellos. Precisamente, en este contexto se demostró que el contacto presencial no es estrictamente necesario, sino que hay que buscar aquel que más facilite la colaboración del informante y menos carga de respuesta le ocasione. Si en el caso de las empresas, el desarrollo de la captación de la información base mediante sistemas CATI (teléfono) o CAWI (internet) era muy grande, la pandemia propició la extensión de estos métodos de forma masiva a otras encuestas dirigidas a hogares, como la EPA. Fue un estudio piloto (obligatorio) que nos demostró que el uso de los teléfonos como medio de contacto era perfectamente válido, por lo que puede sustituir a los sistemas tradicionales presenciales, salvo en casos puntuales.

Técnicas estadísticas. Innovación

- ▶ La estadística oficial debe estar preparada para responder de forma rápida a nuevos paradigmas de trabajo y necesidades de los usuarios, ofreciendo datos con mayor frecuencia, mayor detalle

geográfico y funcional, y con suficiente calidad. Y dado que los recursos siempre serán limitados, la única vía para conseguirlo es la Innovación, innovación en métodos y procedimientos.

- ▶ Dentro de esta innovación estadística se encuentra, sin duda alguna, disponer de un marco que permita el acceso y uso de nuevas fuentes privadas de datos, registros administrativos..., los cuales se ha comprobado que facilitan la elaboración de las estadísticas oficiales en cualquier situación, incluso en las más adversas como fue la vivida en el año 2020.
- ▶ Disponer de infraestructuras y personal cualificado en tecnologías, además de personal conocedor de las últimas técnicas de estadística. Los costes de estas no han de verse como un mero gasto en las instituciones, sino como las necesarias inversiones para el progreso de la Estadística Oficial.

Coordinación institucional

- ▶ Reforzar las relaciones institucionales y el soporte legal para garantizar el acceso de los sistemas estadísticos a toda aquella información necesaria para desarrollar la actividad estadística: teléfonos de contacto, conjuntos de datos básicos relevantes para el estudio de los fenómenos que caracterizan nuestra sociedad.
- ▶ La coordinación realizada en el sistema estadístico europeo ha sido fundamental para adoptar medidas comunes que permiten asegurar la comparabilidad de la información.

En todos estos aspectos viene trabajando en los dos últimos años el INE con diferentes iniciativas de las que quiero destacar especialmente una, que es el estrechamiento en la colaboración con la Universidad y otros organismos de investigación, que se ha visto plasmado en la convocatoria de subvenciones a proyectos de investigación que aborden las líneas de trabajo que el INE define para lograr el desarrollo que la Estadística Oficial ha de tener en el futuro próximo. ●

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. ●

The European Statistical Recovery Dashboard

Rosa Ruggeri Cannata

National accounts methodology; Standards and indicators - unit C1, Eurostat, European Commission (Rosa.RUGGERI-CANNATA@ec.europa.eu)

The coronavirus disease (COVID-19) turned into a global pandemic by March 2020, striking Europe suddenly and with great force. It came with restrictions on mobility and public health measures, implemented to flatten the curve of infections and prevent healthcare systems from being overloaded, and these restrictions disrupted business activity across the European Union (EU). In this context, new data demands quickly emerged: there was an urgent need to track at high frequency what was happening from economic, social and environmental perspectives, in an uncertain situation. Official statistics were highly relevant for timely decision-making, with particular attention to monthly and quarterly data.

The answer of the European Statistical System

The European Statistical System demonstrated its *capacity to adapt quickly, prioritise work and respond to new demands during difficult times*, as required by Commissioner Gentiloni.

The answer of the European Statistical System has covered different developments: on top of ensuring continued production and dissemination of statistics even in the pandemic challenging conditions, Eurostat published a number of guidelines and methodological notes on statistical production and communication in the context of the COVID-19 crisis, and dedicated a section of its website to support statisticians. Moreover, Eurostat coordinated the development of the European Statistical Recovery Dashboard¹,

or recovery dashboard in short. The recovery dashboard put together a set of indicators that allows for tracking the recovery from the COVID-19 crisis, presenting them in an attractive and user-friendly way, together with a Eurostat commentary giving a multi-indicators view on the most recent developments.

The work started in July 2020, triggered by a request by the German presidency of the Council of the European Union, and the first issue of the dashboard was released in December 2020, with the support of Commissioner Gentiloni:

“Our priority in 2021 will be to successfully steer the European economy out of the unprecedented crisis caused by the COVID-19 pandemic. To take the right decisions, we need to have objective and timely data at our disposal.

That’s why today’s initiative from Eurostat and the other members of the European Statistical System is so important. The Recovery Dashboard will inform our actions as we work to recover and to shape a better life for tomorrow through NextGenerationEU”.

The main characteristics of the recovery dashboard were defined early on: it covers the EU member states and EFTA countries (when data are available); it is updated monthly; it consists of monthly and quarterly indicators, and it is improved over time.

Which indicators for measuring the recovery?

The design of the recovery dashboard started with the selection of indicators. Eurostat is committed to publish statistics of high quality, so quality aspects are important for an indicator to be selected. In some cases it was necessary to include ‘experimental statistics’.

Moreover, it was clear from the beginning that the dashboard should not be limited to economic indicators, but also cover social, environmental and health aspects.

When designing dashboards, the first challenge is the trade-off related to the number of indicators

¹ <https://ec.europa.eu/eurostat/cache/recovery-dashboard/>

selected: adding too many indicators creates the risk of overwhelming the user with too much information, whereas excessively restricting the choice could exclude relevant information.

On top of well-known macroeconomic indicators, such as *Gross domestic product and Inflation*, there was high interest in measuring the social impact of the pandemic, particularly the labour market, to the environmental dimension of the recovery, and to morbidity indicators.

With respect to mortality indicators, Eurostat decided not to concentrate on daily/weekly deaths to monitor the spread of the pandemic, but to offer a harmonised indicator to measure what was the “unusual” percentage of additional deaths. The Excess mortality indicator presented in the recovery dashboard does so by comparing the average mortality in a given month with the average of the same months in the pre-pandemic period 2016-2019.

The recovery dashboard also includes indicators concerning the most impacted sectors of the economy, such as the *number of flights* (commercial flights by reporting country, based on Eurocontrol data), and the *nights spent in tourist accommodation establishments*. Two indicators cover the environmental dimension: the monthly average concentration of Nitrogen dioxide (NO₂) in European capital cities, *Air quality*, and the quarterly *Greenhouse gas emissions by the economy*, covering emissions of carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O) and fluorinated gases. They were introduced into the dashboard after its launch, as part of its continuous improvement.

Focusing on the social dimension, there was a lot of attention on the status of employed and unemployed persons, in view of the relevance of the pandemic shock on the labour market. The statistical definition of being unemployed requires to be actively looking for a job. However, a significant proportion of those who had registered in unemployment agencies were no longer actively looking for a job due to the pandemic, as a consequence of, for example, being under confinement measures or no longer available for work if they had to take care of their children during a lockdown. Employed people were also impacted: they could work less hours due to restriction imposed by the enterprises,

often benefiting from short-time work schemes or other support policies put in place by governments. Moreover, the number of hours worked could have decreased without a direct impact on the unemployment rate. These issues generated interest in new indicators, such as the *labour market slack*, which measures unmet needs for employment by considering the sum of unemployed persons, underemployed part-time workers (those part-time workers who wish to work more), and persons either seeking work but not immediately available, or available to work but not seeking it. This indicator is therefore part of the recovery dashboard, together with three indicators on *labour market flows* showing the movements of individuals between employment, unemployment and economic inactivity.

Another interesting indicator is the economic sentiment indicator (ESI), a composite indicator known for its ability to track GDP growth. The ESI is a weighted average of the balances of replies to selected questions addressed to firms in five sectors (industry, services, consumers, retail and construction) in the context of the Business and Consumer Surveys managed by the Directorate-General for Economic and Financial Affairs of the European Commission. The ESI is of particular interest due to its timeliness: it is a monthly indicator available at the end of the reference month.

Looking at indicators for business, two more indicators are worth mentioning in relation to the impact of COVID-19: the first is the quarterly index of bankruptcy declarations, which measures the evolution of businesses that have started the procedure of being declared bankrupt.

The second interesting indicator is the quarterly index of registrations of new businesses, which measures the evolution of entered legal units in the business register at any time during the reference quarter. Those two indicators improved the monitoring of business demography; they are of high relevance not only for the monitoring of the business sector, but also for their potential impact on the labour market and the social consequences of the pandemic.

An analysis of all indicators is out of the scope of this article, and we invite the reader to have a look at the recovery dashboard for a complete list.

The Eurostat commentary

A monthly Eurostat commentary on the current situation is integrated into the recovery dashboard. The commentary describes the recent evolution of dashboard indicators, and it has been enriched by charts and graphics to highlight the main messages. Particular attention is given to comparing the current situation with the pre-pandemic one, and to show variability across countries. This is implemented by displaying in the graphs the highest and lowest values across member states, together with the linear evolution of the European Union indicators, and references to countries in the text.

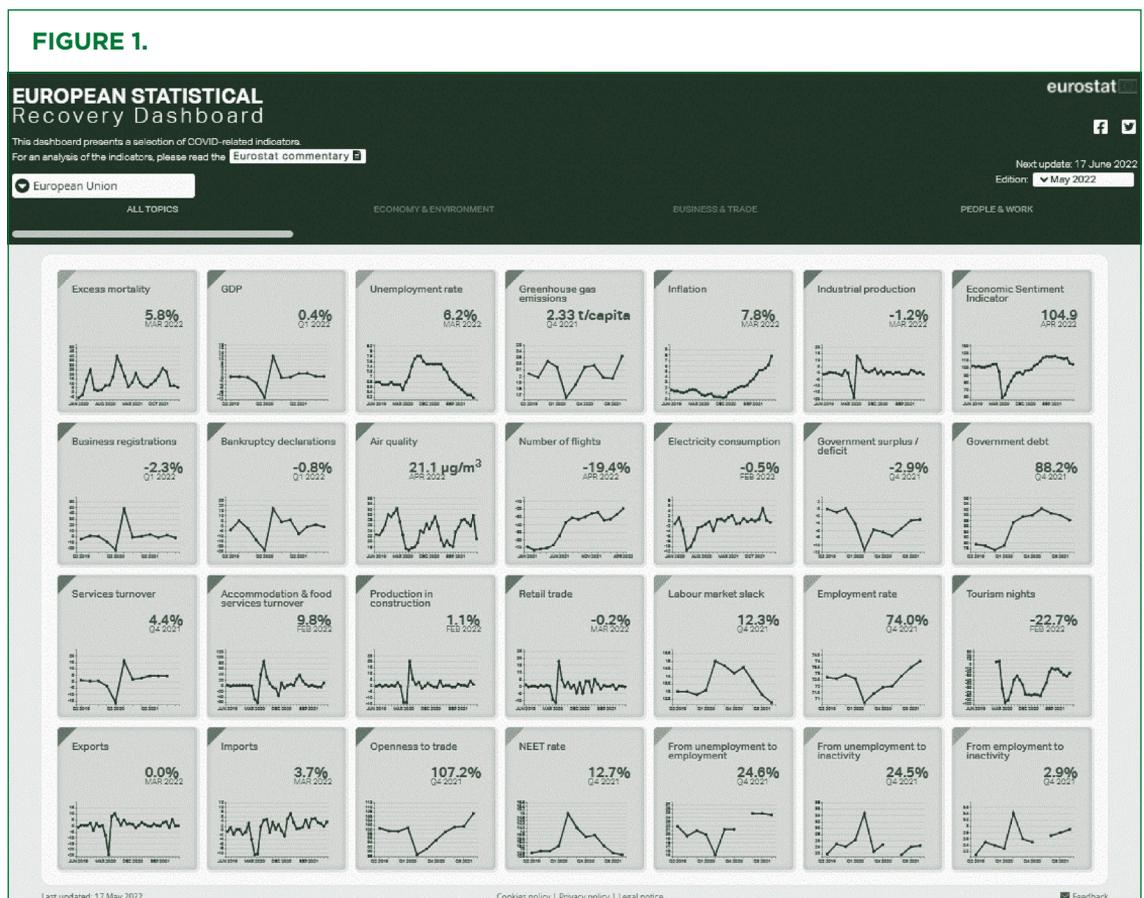
A graph displaying together some indicators and excessive mortality as a background, gives at once an overview on the position with respect to the pre-pandemic and with respect to the peaks in excess mortality. A table highlighting the indicators that are at pre-pandemic levels has

been also added. All those features complement the written analysis with a visual depiction of the recovery.

In order to draft a commentary, data must be frozen at a certain date for the sake of consistency between the data and the commentary. This is why the recovery dashboard is updated monthly, with no continuous update of indicators.

We recognise the value of having access to up to date indicator values, and we make this possible through links from the recovery dashboard's graphs to the source data in the Eurostat data browser. However, freezing the data allows to make a picture of the situation and comparing with the pre-pandemic period. Moreover, having a commentary permits to split data presentation into two levels: the main dashboard page is dedicated to interactive graphs, with the possibility of seeing longer/shorter time series, selecting countries and/or European aggregates to visually compare trends, exploring different data views by topic. In

FIGURE 1.



Source: <https://ec.europa.eu/eurostat/cache/recovery-dashboard/>

the commentary, we can present a multi-indicator analysis, together with data rebased at pre-pandemic levels, we can show graphs with lowest and highest values of indicators across countries at different periods, and we can show which indicators are back at pre-pandemic levels for the European Union, when such a comparison is possible. The user can then freely choose between a fast view of the indicators, or spending some time with analysis and more complex graphs.

A user-friendly tool for indicators' visualization

In developing the recovery dashboard, a key challenge was to design an attractive and appealing visualization tool. Eurostat invested in this direction and the result has been welcomed by users.

The recovery dashboard indicators are accessible either all together from the main page, or organised in three topics: Economy and environment, Business and trade, and People and work.

The user interface is based on linear graphs, giving an immediate image of indicators' trends; it allows the selection of one or more countries, and to keep the same country selection for all charts if so wished. In this way, users can easily compare developments across different countries and with the European aggregates. The user can also enlarge the time span shown in the default view by moving a time sliding banner, in order to display longer time series.

Moreover, the tool offers the possibility of consulting short methodological descriptions of the indicators, of accessing source datasets with downloading facilities for data, and accessing the Eurostat commentary.

Previous editions are retrievable via a menu, and the user is informed when the next update will be published. All graphs are downloadable as images, for example to be inserted in a report, and sharable via the most popular social platforms. Finally, yet importantly, Eurostat took care of guaranteeing a good responsiveness of the tool, and its portability. Eurostat also surveyed users by running a usability test, including in face-to-face interviews, receiving very positive feedback.

Continuous improvement

Eurostat is committed to a continuous improvement of the recovery dashboard. Since December 2020, a number of expansions in terms of coverage and functionalities have been introduced in order to better track the recovery and reflect user's needs, including those expressed in the user survey. The result has been a number of improvements both in indicators coverage and in the visualization tool. Some indicators can now be displayed in different units; this permits to look both at the latest developments, by showing for example the month on previous month growth rate, and to compare with the level before the pandemic, by looking at the index. A number of indicators have been added, and more are expected to become available, in order to keep the content in line with the latest statistical developments. Whilst the update process is completely transparent for the user, it is important to point out that developing and maintaining the recovery dashboard benefits from the contribution of many statisticians, across many statistical domains and including ICT experts, who actively engage and contribute to its improvement. ●

Useful references

- Eurostat page on COVID-19: <https://ec.europa.eu/eurostat/web/covid-19/overview>
- Statistics Explained article: Labour market slack - unmet need for employment - quarterly statistics: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Labour_market_slack_-_unmet_need_for_employment_-_quarterly_statistics
- Statistics Explained article: Hours of work and absences from work - quarterly statistics: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Hours_of_work_and_absences_from_work_-_quarterly_statistics#Impact_of_COVID-19_pandemic_and_potential_recovery

MoMo: una herramienta esencial durante la pandemia de COVID-19

**Inmaculada León, Concha Delgado,
Amparo Larrauri**

*Centro Nacional de Epidemiología. Ciber de Epidemiología y Salud Pública (CIBERESP).
Instituto de Salud Carlos III.*

El sistema de monitorización de la mortalidad diaria por todas las causas (MoMo) fue desarrollado en 2004 en el Centro Nacional de Epidemiología (CNE) del Instituto de Salud Carlos III (ISCIII), en el marco del «Plan de acciones preventivas contra los efectos de las temperaturas excesivas», coordinado por el Ministerio de Sanidad, para reducir el impacto sobre la salud de la población como consecuencia del exceso de temperatura, durante los meses estivales. El objetivo era identificar las desviaciones de mortalidad diaria observada con respecto a la esperada según las series históricas de mortalidad y comunicar al Ministerio de Sanidad los resultados para su investigación o para la puesta en marcha de las medidas de control.

MoMo se convirtió en el primer sistema que ofrecía información adicional a tiempo real, que complementaba los datos de defunciones por COVID-19 obtenidas de los sistemas de Vigilancia Epidemiológica, y dando una idea del impacto de la pandemia en la mortalidad de la población española

En 2009 se amplió su utilización durante todo el año a otras situaciones para estimar de forma indirecta el impacto de cualquier evento

de importancia en salud pública no relacionado con temperaturas excesivas. Desde entonces se utiliza como sistema complementario a otros sistemas de vigilancia como el sistema de la Vigilancia de la Gripe en España (SVGE) o el Sistema de Vigilancia de Infección Respiratoria Aguda (SIVIRA).

Su principal fuente de información son las defunciones diarias por todas las causas procedentes de los registros civiles informatizados, que se reciben en el CNE del Ministerio de Justicia. Actualmente se incluyen 3.999 registros civiles, con representación de todas las provincias españolas y que corresponden al 93% de la población española. Otra fuente de información es la mortalidad diaria por todas las causas, con cobertura nacional, procedente del Instituto Nacional de Estadística (INE).

A pesar de ser un sistema inespecífico que no permitía establecer causalidad directa en el análisis de mortalidad, MoMo ha constituido una herramienta esencial de vigilancia a tiempo real, para realizar estimaciones del impacto de diferentes eventos de importancia en Salud Pública en la mortalidad de la población. El ejemplo más cercano es su papel en la primera ola pandémica de COVID-19 en España, cuando MoMo se convirtió en el primer sistema que ofrecía información adicional a tiempo real, que complementaba los datos de defunciones por COVID-19 obtenidas de los sistemas de Vigilancia Epidemiológica, y dando una idea del impacto de la pandemia en la mortalidad de la población española. Entre marzo y mayo de 2020, MoMo estimó un exceso de mortalidad por todas las causas sin precedentes en España (45.010 excesos de defunciones por todas las causas). Dicho exceso fue considerablemente mayor que las defunciones confirmadas por COVID-19 notificadas a la Red Nacional de Vigilancia Epidemiológica (RENAVE) en el marco del sistema universal de vigilancia de COVID-19. Dicha diferencia se explica por la limitada capacidad diagnóstica de SARS-CoV-2 en Espa-

ña durante la primera ola pandémica, lo que dio lugar a una infranotificación de las defunciones debidas a COVID-19 en este periodo. Esta infranotificación ha sido común en otros países de nuestro entorno durante los primeros meses de la pandemia de COVID-19 y también es un aspecto frecuente en otras enfermedades como la gripe, cuya carga de mortalidad debida a las epidemias estacionales gripales, suele estimarse con modelos de excesos de mortalidad.

Es importante subrayar que, al tratarse de un sistema inespecífico, las estimaciones MoMo no se pueden atribuir a un evento concreto, como muertes por COVID-19, sino también a la mortalidad indirecta de la pandemia, en la que factores como el aislamiento social, la dificultad de acceso a la atención médica, los cambios en la estructura social o las modificaciones en la actividad del sistema sanitario, entre otros, pueden haber contribuido a los excesos de mortalidad.

A partir de la tercera ola pandémica, las estimaciones MoMo de excesos de mortalidad por todas las causas, fueron mucho menores que en la primera ola pandémica, un efecto observado en otros países de nuestro entorno y que se atribuye a un posible efecto “cosecha” que suele observarse después de producirse una mortalidad significativa entre personas vulnerables, cuya salud ya estaba comprometida. Por otra parte, una mejora en la atención de salud y de los tratamientos contra COVID-19, así como el impacto de la vacunación COVID-19 en la población, que comenzó en España antes de la tercera ola pandémica, podrían haber contribuido a esta menor mortalidad, tanto de excesos MoMo como de defunciones confirmadas de COVID-19.

MoMo ha ido incorporando mejoras desde el inicio de su implementación, como automatizaciones parciales en los procesos, o correcciones de los modelos para asumir los cambios en las tendencias de mortalidad. Un cambio crítico se produjo en 2018, con una completa automatización del proceso de análisis y obtención de resultados, lo que permitió el desarrollo del panel MoMo en la Web del ISCIII (**Panel MoMo**¹) en el inicio de la pandemia de COVID-19. El panel MoMo ofreció diariamente

excesos de mortalidad por todas las causas, a tiempo real y en abierto, que ayudaron a estimar la gravedad de la emergencia pandémica en términos de mortalidad sobre la población española. MoMo puso en evidencia el impacto generalizado de la pandemia sobre las personas mayores de 79 años, especialmente en la primera ola pandémica (53% de los excesos de defunciones identificados en todos los grupos de edad), así como las diferencias geográficas de excesos de mortalidad por todas las causas en las distintas olas pandémicas. Desde el 10 de marzo de 2020, MoMo ha estimado un exceso de mortalidad de 108.020 defunciones por todas las causas en España (**Panel MoMo**²).

En 2022, MoMo ha actualizado su desarrollo metodológico y presentación para informar de manera más específica las estimaciones del exceso de mortalidad en España

En 2022, MoMo ha actualizado su desarrollo metodológico y presentación para informar de manera más específica las estimaciones del exceso de mortalidad en España. Tras esta actualización, los datos se presentan con dos enfoques diferenciados: uno Predictivo, basado en el Índice Kairós (**Panel Kairós**³), que ofrece alertas de excesos de mortalidad atribuibles al exceso o defecto de temperatura, y uno estimativo (**Panel MoMo**), que informa sobre excesos de mortalidad tanto por todas las causas, como atribuibles al exceso o defecto de temperatura.

El modelo MoMo inicial, utilizado para estimar las muertes esperadas, era un modelo de series temporales no paramétrico en el que se controlaba la tendencia mediante regresión lineal y la estacionalidad mediante medias móviles históricas de 5 días, sin la introducción de más variables independientes. Los cambios más relevan-

1 <https://cnecovid.isciii.es/momo.html>

2 https://momo.isciii.es/panel_momo/

3 <https://momo.isciii.es/kairos/>

tes en el MoMo actual han sido la utilización de modelos paramétricos y la inclusión de la temperatura, procedente de la Agencia Estatal de Meteorología, como variable independiente.

En la actualidad el panel MoMo se divide en:

- **Enfoque predictivo**, que proporciona alertas de excesos de mortalidad, pero no cuantifica estos excesos. En este modelo se incluye la información de los últimos diez años hasta el día en curso, mediante una ventana móvil, y no se elimina ningún *outlier* para poder ser lo más fiel posible a la realidad y realizar mejores predicciones; el año 2020 que se elimina por su patrón anómalo debido a la pandemia de COVID-19. Los resultados están disponibles en la web de **Panel Kairós**⁴, que ofrece alertas de mortalidad asociadas a un exceso o defecto de la temperatura mediante el Índice Kairós y sus tres niveles de riesgo de mortalidad atribuible a temperatura: Kairós 1, Kairós 2 y Kairós 3, según la probabilidad de que un exceso de mortalidad atribuible a temperatura de más del 10%, sea inferior al 40%, entre 40% y 60% y más de 40%, respectivamente. El índice Kairós se presenta para el día en curso y los cinco días siguientes, por ámbito poblacional, sexo y grupo de edad.
- **Enfoque estimativo**, que ofrece estimaciones diarias de exceso de mortalidad por todas las causas y la mortalidad atribuible al exceso o defecto de temperatura. En este modelo se incluyen los últimos diez años, a excepción del año en curso y el 2020 y se eliminan todos los *outliers* (observaciones por encima del intervalo de confianza al 99%). Los resultados de excesos de mortalidad por todas las causas y atribuibles tanto a exceso como defecto de temperatura, se presentan por ámbito poblacional (nacional, de comunidad autónoma y provincial), sexo y grupo de edad, y con diferente agregación temporal (diario, mensual o anual). Además, se ofrece información acerca del retraso

en la notificación de las defunciones y una pestaña de documentación en la que explica la metodología empleada. Toda la información puede consultarse en la web de **Panel MoMo**⁵.

Ambos modelos se basan en modelos mixtos por provincia GAM (*generalized additive model*), que utilizan regresión de Poisson para modelar la variable dependiente, defunciones diarias, ajustan la tendencia y estacionalidad mediante diferentes tipo de *splines*, e incluyen como otras variables independientes el impacto del exceso o defecto de temperatura a través de dos variables sintéticas *ato* y *f_ato*, definidas como número de grados en los que la temperatura observada está por encima o por debajo del umbral de disparo de la mortalidad por calor y frío, respectivamente, multiplicado por el número de días consecutivos previos, en los que se mantienen este efecto.

Además, se publican informes semanales de MoMo desde marzo de 2020 en los que se puede consultar los excesos de defunciones a nivel nacional, por comunidad autónoma y subgrupos poblacionales de edad y sexo. La información está disponible en la web de **Informes MoMo**⁶.

La actualización de MoMo permite ofrecer información más específica en el cálculo de posibles excesos de mortalidad, y ofrecer alertas de mortalidad por todas las causas y asociadas a excesos o defectos de temperaturas. De esta manera se fortalece la valoración global del impacto de diferentes situaciones de interés en salud pública, entre las que se incluyen fenómenos ambientales, la posible introducción de medidas de prevención y la mejora de actuaciones sociosanitarias, entre otras. En la actualidad, MoMo constituye una herramienta esencial de vigilancia a tiempo real, como sistema de alerta temprana para la acción en salud pública, y para realizar estimaciones del impacto de diferentes eventos de salud en la mortalidad de la población. ●

4 <https://momo.isciii.es/kairos/>

5 https://momo.isciii.es/panel_momo/

6 <https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/Enfermedades-Transmisibles/MoMo/Paginas/InformesSemanalesMoMo.aspx>

La Estadística de Causas de Muerte en el contexto de la pandemia

Cuando sobrevino la pandemia de la COVID-19 la información sobre mortalidad que se iba conociendo se refería a datos de mortalidad estimada (Sistema de Monitorización de la Mortalidad diaria -MoMo- y Estadística experimental sobre estimaciones semanales de defunciones) y de la mortalidad declarada a la Red Nacional de Vigilancia Epidemiológica (RENAVE).

Por una parte, la mortalidad declarada por COVID-19, aportada por el Ministerio de Sanidad a partir de los registros de las CC. AA., obedecía al criterio de fallecidos con prueba positiva sin distinguir la causa directa de estos fallecimientos y, por otra, la mortalidad estimada permitía establecer el exceso de mortalidad, sin distinguir tampoco las causas del fallecimiento.

Esta información, obtenida de manera rápida, fue esencial para llevar a cabo la vigilancia epidemiológica de la pandemia, pero con ella no se podía conocer de forma precisa el impacto real de la COVID-19 en la mortalidad. La fuente que proporciona la mejor estimación de la mortalidad atribuible a la pandemia es la Estadística de Defunciones según la Causa de Muerte.

Margarita García Ferruelo

*S. G. de Estadísticas Sociales Sectoriales.
Instituto Nacional de Estadística*

La Estadística de Defunciones según la Causa de Muerte ofrece información de forma muy precisa y de calidad pero no facilita resultados rápidos como se puso de manifiesto con la crisis de la COVID-19. El motivo está en el proceso estadístico, porque es muy complejo.

No obstante, conscientes de la importancia que tenía en aquel momento esta información para la sociedad se modificó el proceso para obtener un avance con los resultados de la primera ola (cinco primeros meses) que fueron difundidos en diciembre de 2020.

La variable fundamental que se investiga en esta estadística es la causa básica de defunción¹, que se selecciona siguiendo los criterios de la Clasificación Internacional de Enfermedades (CIE-10) a partir de las enfermedades que el médico describe en el Certificado Médico de Defunción (CMD).

El estudio de la mortalidad basado en la causa básica de muerte constituye por sí mismo una herramienta de gran valor, porque permite detectar cambios en las pautas de ciertas enfermedades. Las enfermedades cardiovasculares han ido reduciendo paulatinamente su peso relativo y, aunque se mantienen como primera causa de muerte, han ido cediendo importancia a otras causas como los tumores, las demencias o la enfermedad de Alzheimer. Con la aparición en el año 2020 de la COVID-19, esta enfermedad penalizó a otras enfermedades como las del sistema respiratorio, desplazándolas a favor de la COVID-19.

Cuando sobrevino la pandemia, la OMS incorporó la nueva enfermedad a la CIE-10 asignándole dos códigos diferentes para distinguir COVID-19 virus identificado (con prueba positiva) y COVID-19 virus no identificado (sospechoso), los cuales permitieron conocer la mortalidad causada directamente “por” COVID confirmada y “por” COVID sospechosa.

Por otra parte, desde el año 2016, además de seleccionar la causa básica también se proporciona información de cada una de las enfermedades informadas por el médico en el CMD y que han contribuido al fallecimiento (causas múltiples). Esta información supone un valor añadido esencial en el conocimiento de

1 La causa básica de la defunción se define como la enfermedad o lesión que inició la cadena de acontecimientos patológicos que condujeron directamente a la muerte o las circunstancias del accidente o violencia que produjeron la lesión fatal.

la mortalidad y en particular de la mortalidad asociada a la COVID-19. Si la Estadística se limitara al estudio de la causa básica, se perdería la oportunidad de identificar las asociaciones entre patologías y de conocer la verdadera dimensión de la mortalidad.

De esta manera se pudieron conocer también las defunciones en las que la COVID-19 había sido informada en el CMD pero sin constituir la causa directa de la muerte, es decir, la mortalidad “con” COVID confirmada y “con” COVID sospechosa.

El estudio de la mortalidad basado en la causa básica de muerte constituye por sí mismo una herramienta de gran valor, porque permite detectar cambios en las pautas de ciertas enfermedades.

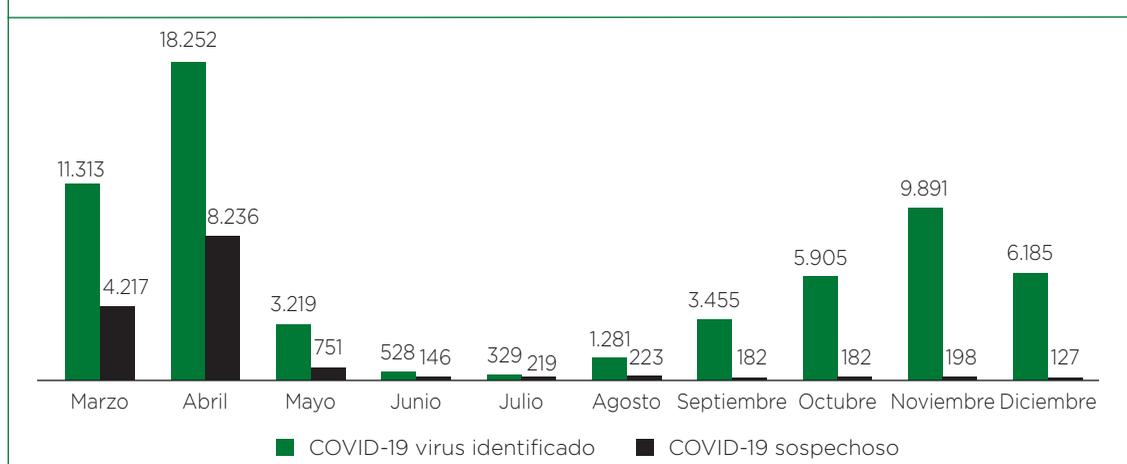
Igualmente, mediante las causas múltiples se obtuvo información sobre las complicaciones derivadas de la COVID-19 y las comorbilidades que presentaban las personas que fallecieron. La insuficiencia respiratoria y la neumonía fueron las complicaciones más frecuentes infor-

madas en los certificados de defunción de las personas que fallecieron “por” COVID-19 tanto en los casos de virus identificado como no identificado. Las principales comorbilidades en las muertes “por” COVID virus identificado fueron la enfermedad hipertensiva y la insuficiencia renal. En el caso de las muertes “por” COVID-19 virus no identificado (sospechoso) la principal comorbilidad fue la demencia.

Los resultados mostraron que en el año 2020 hubo 60.358 fallecimientos cuya causa de muerte fue COVID-19 virus identificado y otras 14.481 personas murieron con sospecha de COVID-19 por tener síntomas compatibles con la enfermedad (COVID-19 virus no identificado). Además, los médicos certificaron 8.275 defunciones en las que la muerte fue debida a otras causas, pero en las que la COVID-19 contribuyó al fallecimiento de dichas personas, como comorbilidad. En 3.770 casos los facultativos lo identificaron y en 4.505 casos no lo hicieron, pero sospecharon de su presencia por tener síntomas compatibles con la enfermedad.

Otra información igualmente relevante que proporcionó esta estadística sobre como influyó la pandemia en la mortalidad ha sido el lugar en el que ocurrió la defunción (hospital, residencia socio-sanitaria, domicilio, otro lugar), así como el efecto que tuvieron las medidas sociales adoptadas durante el estado de alarma en otras causas de muerte (accidentes de tráfico, suicidios, enfermedades cardiovasculares...).

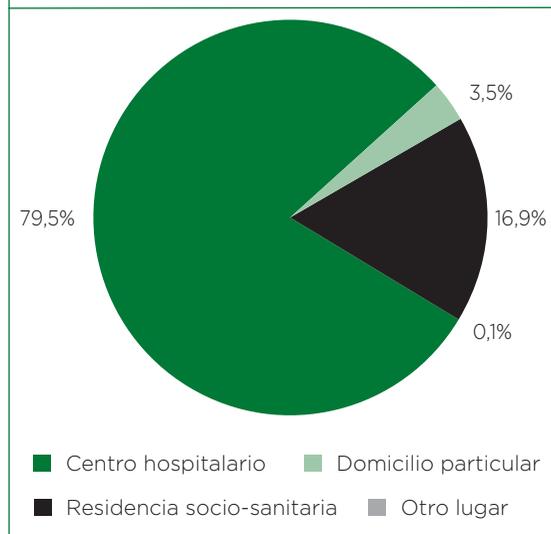
FIGURA 1. DEFUNCIONES DEBIDAS A COVID-19 POR MESES. Año 2020



Fuente: Estadísticas de Causas de Muerte. INE.

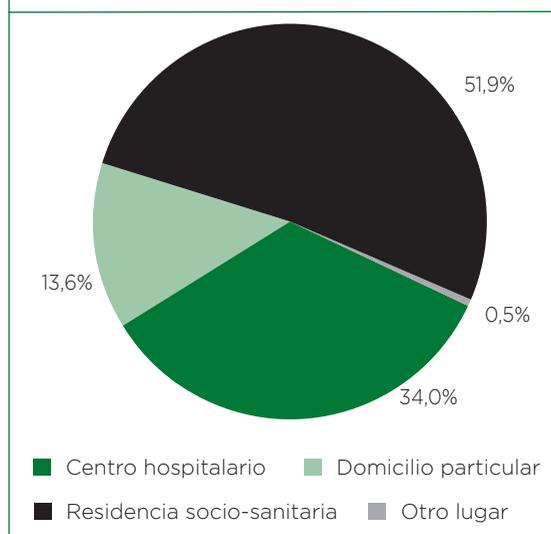
Durante 2020 las defunciones por COVID-19 virus identificado se produjeron, principalmente, en los centros hospitalarios (79,5%) mientras que el mayor número de fallecimientos por COVID-19 virus no identificado (sospechoso) se dio en las residencias socio-sanitarias (51,9%).

FIGURA 2. DISTRIBUCIÓN DE LAS DEFUNCIONES POR COVID-19 VIRUS IDENTIFICADO Y LUGAR DONDE OCURRIÓ LA DEFUNCIÓN



Fuente: Estadísticas de Causas de Muerte. INE.

FIGURA 3. DISTRIBUCIÓN DE LAS DEFUNCIONES POR COVID-19 VIRUS SOSPECHOSO Y LUGAR DONDE OCURRIÓ LA DEFUNCIÓN



Fuente: Estadísticas de Causas de Muerte. INE.

Y en cuanto a otras causas que reflejaron una tendencia distinta en 2020, hay que fijarse en los fallecimientos debidos a causas externas (muertes que se producen en circunstancias violentas o con sospecha de criminalidad). El suicidio se mantuvo como la primera causa de muerte externa, con 3.941 fallecimientos y se incrementó un 7,4% respecto a 2019.

El mayor incremento en el número de suicidios se dio en el mes de agosto, un 34%, y en los meses de abril y mayo, en pleno confinamiento, se reflejaron importantes descensos de defunciones por accidentes de tráfico (un 55,6% y un 51,0% menos que en 2019, respectivamente).

Al principio de este artículo se ha hecho referencia a la complejidad del proceso estadístico al que se someten los CMD. Este proceso se complicó especialmente durante la primera ola de la pandemia con el colapso en los Registros Civiles, el cierre de delegaciones provinciales del INE (donde llegan los CMD para su escaneo), un mayor volumen de certificados a procesar y la más importante, una enfermedad nueva.

La materia prima de la estadística es la información que el médico certificador de la muerte refleja en el Certificado Médico de Defunción (CMD). Estos certificados están preparados para el reconocimiento óptico, pero hay que tener en cuenta que los literales que figuran en los certificados son enfermedades lo que hace más difícil su reconocimiento y es necesario apoyarse en un diccionario de enfermedades que está compuesto por las enfermedades que los médicos han certificado a lo largo del tiempo (actualmente cuenta con casi 170.000 entradas diferentes).

Cuando apareció la nueva enfermedad COVID-19, el OCR no reconocía esos literales porque no estaban incorporados al diccionario lo que obligó a realizar una revisión manual de todos los CMD hasta que se pudieron incluir los términos utilizados en los primeros certificados para denominar la COVID (145 expresiones).

Una vez que la información literal de los certificados ha sido escaneada y revisada, el paso siguiente consiste en codificar las causas múltiples y seleccionar la causa básica de acuerdo con las reglas de la CIE-10.

La selección y codificación de la causa básica es la parte más complicada. El volumen 2 de la CIE-10 facilita las reglas de selección basadas en criterios médicos y buscando relaciones lógicas entre las enfermedades informadas en el CMD. Hasta 2014 esta selección se hacía manualmente introduciendo inevitablemente componentes de heterogeneidad en la interpretación de las reglas de la CIE-10. Para mejorar la comparabilidad se desarrolló en Europa el *software* IRIS, un sistema automático inteligente que no solo resuelve la codificación sino también la selección de la causa básica mediante algoritmos basados en las relaciones causales entre las enfermedades informadas en el CMD de acuerdo a las directrices de la CIE-10.

Para hacernos una idea de lo valiosa que es la herramienta IRIS frente a la selección manual, basta citar que el número de relaciones entre enfermedades que maneja este *software* supera los 29 millones.

Con la irrupción de la COVID-19 IRIS tuvo que modificar sus tablas de decisión para incluir el tratamiento de esta nueva enfermedad.

Por último, también se ha mencionado al principio del artículo que esta Estadística ofre-

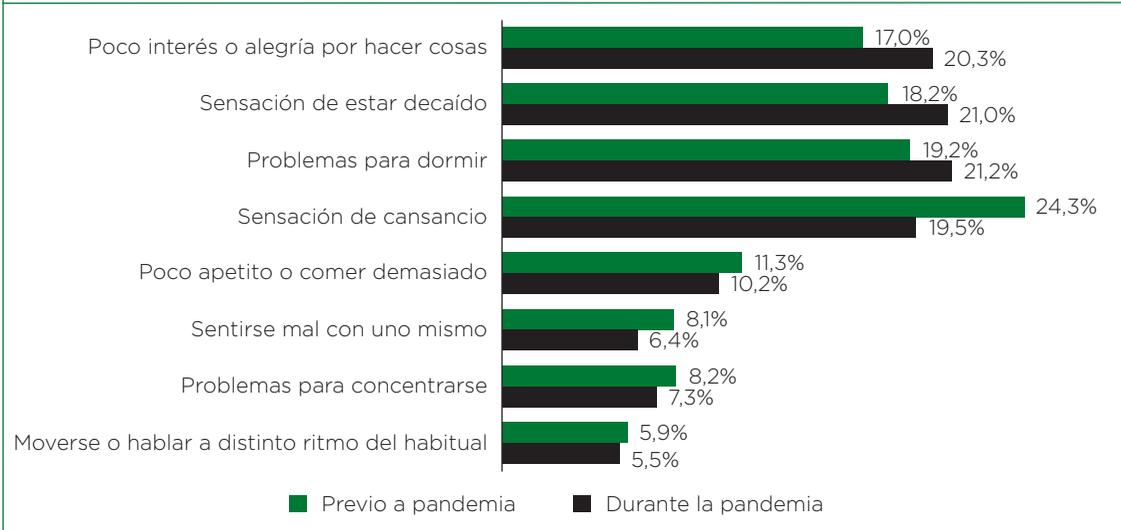
ce información de forma precisa y de calidad pero no facilita resultados rápidos. La razón fundamental se encuentra en el certificado médico oficial de defunción, un documento en papel que por una parte tiene un recorrido administrativo (Registro Civil) antes de poder ser explotado estadísticamente y por otra su explotación estadística requiere del escaneo masivo y de revisión del OCR.

La solución se encuentra en implantar el CMD electrónico (CMDe). En efecto esto permitirá recopilar la información de defunciones en tiempo real y reducir significativamente los plazos de la Estadística de Causas de Muerte. Además, esta misma información alimentará con la misma rapidez el sistema de Registro Civil y el sistema sanitario. Por otra parte, desde el punto de vista administrativo el CMDe dotará de mayor garantía al acto médico.

El INE ha desarrollado la aplicación CMDe en colaboración con el Consejo General de Colegios Oficiales de Médicos y trabaja en su implantación en coordinación con el proyecto de digitalización del Registro Civil previsto en la Ley 20/2011 de Registro Civil.



FIGURA 4. INDICADORES DE SALUD MENTAL. Población de 15 y más años



Fuente: Encuesta Europea de Salud.

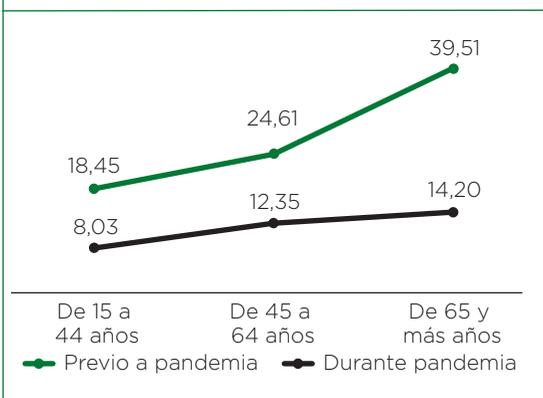
Hasta ahora solo se ha hablado del efecto de la pandemia sobre la mortalidad, pero también se ha sentido en otras dimensiones de la salud como la salud mental o el uso de los servicios sanitarios.

La última Encuesta Europea de Salud, que se realizó entre julio 2019 y julio 2020, ha permitido obtener indicadores en dos períodos diferenciados: previo a la pandemia (julio 2019-febrero 2020) y durante la pandemia (marzo 2020-julio 2020).

En los primeros meses de la pandemia se incrementaron los problemas de salud mental. Aumentó el número de personas que declararon tener poco interés o alegría por hacer cosas y los problemas para dormir.

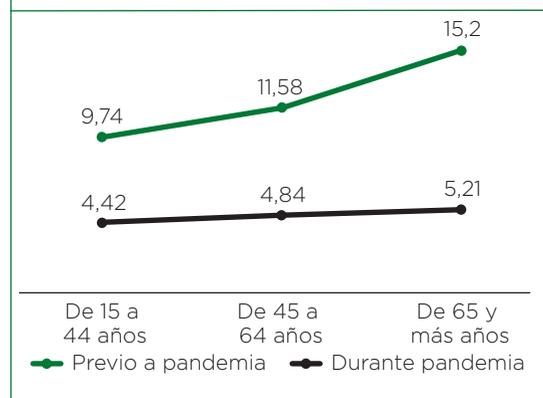
El consumo de medicamentos, tanto recetados como no recetados, disminuyó en la primera ola de la pandemia y la asistencia sanitaria también se resintió en ese periodo, tanto las consultas al médico de familia como al médico especialista. ●

FIGURA 5. CONSULTAS AL MÉDICO DE FAMILIA. Porcentaje de población de 15 y más años



Fuente: Encuesta Europea de Salud.

FIGURA 6. CONSULTAS AL MÉDICO ESPECIALISTA. Porcentaje de población de 15 y más años



Fuente: Encuesta Europea de Salud.

Matemáticas y COVID-19: monitorización, predicción, retos y perspectivas

Daniel López¹, Sergio Alonso¹, Martí Català², Enric Álvarez-Lacalle¹, Clara Prats¹

¹ *Computational Biology and Complex Systems (BIOCOM-SC), Departamento de Física, Universitat Politècnica de Catalunya*

² *Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford*

La pandemia ha sido, y aún sigue siendo, una catástrofe que ha afectado a todas las poblaciones del mundo. No había expertos en pandemias, hemos aprendido mucho, pero, probablemente, si tenemos una nueva pandemia deberemos seguir aprendiendo.

BIOCOM-SC es un grupo de investigación de la *Universitat Politècnica de Catalunya* que, antes de la pandemia, tenía una línea de investigación en epidemiología matemática y simulación de tuberculosis, malaria, enfermedad de Chagas y otras enfermedades infecciosas. Cuando en diciembre de 2019 se comienzan a hacer públicos diariamente los datos de la epidemia en China, por simple curiosidad científica constatamos que algunos modelos empíricos nos describían correctamente el comportamiento observado y permitían hacer predicciones a corto plazo.

De hecho, en aquellos primeros meses vimos cómo la evolución temporal de la epidemia dependía especialmente de las medidas no farmacológicas, cuyos efectos no se podían prever porque no se disponía de datos anteriores. En consecuencia, en aquel momento los modelos matemáticos clásicos como los modelos SIR o SEIR no tenían aún una buena capacidad predictiva, ya que dependían de múltiples parámetros y procesos que no se conocían con detalle. Por otro lado, tampoco se podían utilizar herramientas de inteligencia artificial ni de estadística multivariable porque no existían datos históricos

y todo cambiaba muy rápidamente. En cambio, los modelos empíricos, como el modelo de Gompertz, nos permitían realizar predicciones a corto plazo con una fiabilidad aceptable.

Sin apenas darnos cuenta, en marzo del 2020 nos encontramos haciendo informes diarios para la *Comisión Europea* (DG-CONNECT) y colaborando con la *Agència de Qualitat i Avaluació Sanitàries de Catalunya de la Generalitat de Catalunya*. Nuestro trabajo ha girado desde entonces alrededor de tres ejes: (1) el desarrollo de índices y herramientas para la monitorización, (2) las predicciones a corto plazo (1-3 semanas) y (3) el análisis de varios aspectos relacionados con la pandemia mediante el uso de modelos computacionales.

La pandemia nos hace patente que, como investigadores, vivir en el siglo XXI es un privilegio. Desde los inicios de la crisis hemos dispuesto de actualizaciones diarias de muchos datos abiertos de incidencia, de muertes, de vacunación... Al trabajarlos hemos descubierto sus problemas, carencias y limitaciones, pero los datos siguen siendo nuestra herramienta básica para objetivar y cuantificar la situación pandémica en cada momento, comprender mejor su dinámica y trabajar para contribuir en su gestión. De hecho, las herramientas de modelización han evolucionado en paralelo a la calidad de los datos, que han acompañado también el incremento de conocimiento sobre SARS-CoV-2. En apenas dos años, de una situación inicial de emergencia absoluta con poco conocimiento, pocos datos y una falta de herramientas globales de monitorización y predicción, hemos evolucionado hasta una situación de aparente control funcional gracias a la campaña de vacunación masiva, donde conocemos mucho mejor este virus y su dinámica y donde, ahora sí, ya podemos trabajar con modelos mecanicistas basados en este conocimiento.

Monitorización de la pandemia

Para monitorizar la situación epidemiológica es necesario disponer de indicadores simples que sean informativos. En los primeros meses de pandemia, en ausencia de vacunación, la evolución de los casos nos anticipaba unívocamente la afectación hospitalaria y la mortalidad asociada a SARS-CoV-2. Era clave poder conocer en cada momento no solo el nivel de incidencia (indicador de afectación), sino también la tendencia al alza o a la baja de dicha incidencia (indicador de velocidad).

Uno de los indicadores de afectación más utilizados ha sido la *Incidencia Acumulada en 14 días* (IA_{14}), que son los casos nuevos diagnosticados las dos semanas previas por 100.000 habitantes. Este indicador está poco afectado por la variabilidad diaria y, además, al ser relativo permite comparar situaciones de países o regiones distintas. La ventana de 14 días es la que se estimó inicialmente como el intervalo máximo en el que una persona puede ser infecciosa, aunque generalmente este período es menor. En estas estimaciones iniciales se estableció

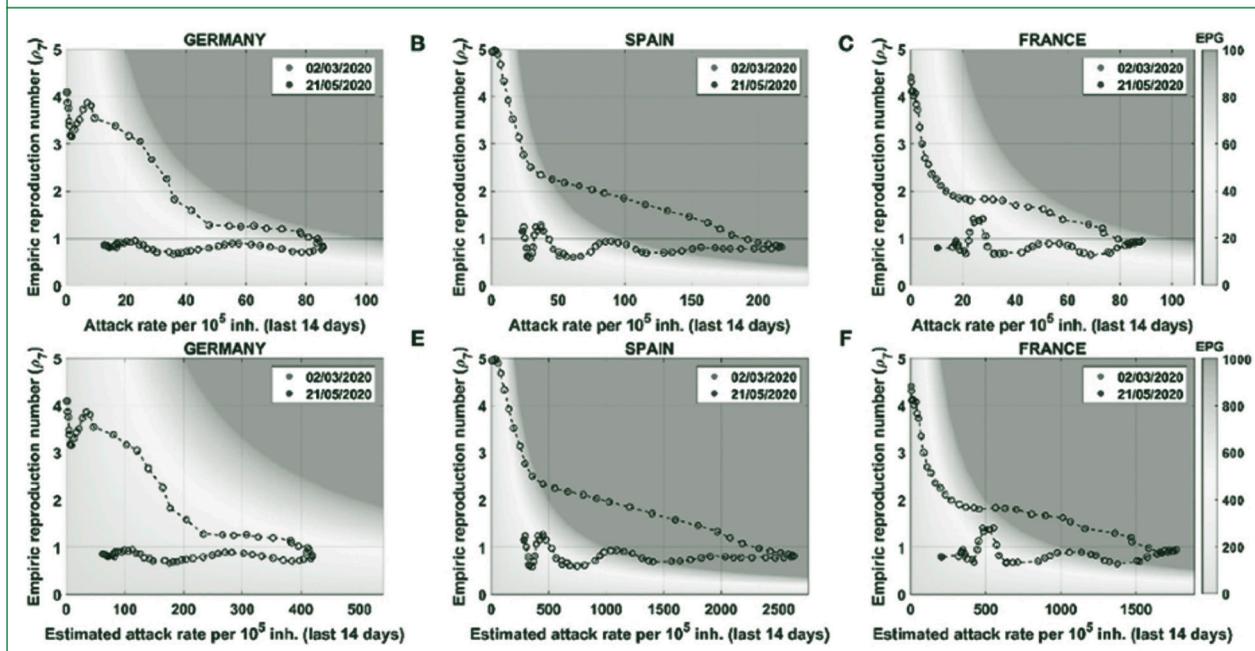
el intervalo entre la infección y el momento de máxima infecciosidad de una persona alrededor de 5 días. Comparando los casos nuevos diarios de un día con los de cinco días anteriores podíamos evaluar, en promedio, cuántas personas habría infectado cada nuevo caso diagnosticado [1]. Este valor puede ser indicativo del *Número Reproductivo Efectivo* (R_t), es decir, del promedio de nuevos casos generados por un caso índice en un cierto momento, un indicador de la velocidad a la que se está propagando una epidemia.

Considerar simultáneamente la velocidad en que se está propagando la infección conjuntamente con la intensidad de la afectación es una buena medida de la gravedad de la situación. Una forma simple de hacerlo es mediante el producto de R_t por IA_{14} :

$$EPG = R_t \cdot IA_{14}$$

EPG son las iniciales en inglés de *Effective Potential Growth*, que se podría traducir como crecimiento efectivo potencial. Este valor, conocido en Cataluña como “índice de riesgo de rebrote” por su efectividad en la detección de cambios de tendencia [2],

FIGURA 1. DIAGRAMAS DE RIESGO DE DIFERENTES PAÍSES CORRESPONDIENTES A LAS PRIMERAS SEMANAS DE LA EPIDEMIA EN EUROPA.



Fuente: [1].



amplifica o relativiza una cierta incidencia en función de si estamos en fase expansiva ($R_t > 1$) o regresiva ($R_t < 1$).

Con R_t , IA_{14} y EPG disponíamos de tres parámetros para evaluar la situación pandémica en determinado país o región, pero necesitábamos además alguna herramienta para poder comprender o mostrar la evolución del nivel de riesgo a lo largo del tiempo. Diseñamos un diagrama que resolvía esta necesidad, el *Diagrama de riesgo*, donde se representa R_t en el eje de ordenadas y IA_{14} en el eje de abscisas (figura 1). La distancia al origen de coordenadas es una buena medida de la situación de “riesgo” en que se encuentra la población [1]. Además, este diagrama nos permite ver la situación actual y compararla con el pasado. Utilizando una escala de colores, facilitábamos una interpretación intuitiva del diagrama.

En situaciones epidémicas, realizar predicciones de calidad es importante para ayudar a los responsables de salud pública y a los gestores del sistema de salud a tomar decisiones con criterios lo más objetivos posibles

Los índices y el diagrama descritos fueron diseñados en la situación de emergencia que describíamos al principio, buscando ese equilibrio entre la simplicidad y la utilidad. Existen otros índices con propiedades similares, capaces también de dar información sobre el nivel de afectación y la velocidad de propagación. Con la vacunación masiva y con la proliferación de las variantes Ómicron han cambiado los niveles de riesgo asociados a la circulación del virus, ya que la afectación en cuanto a casos graves se ha reducido drásticamente. De hecho, en la situación actual los indicadores más relevantes para la monitorización son aquellos relacionados con las poblaciones vulnerables y los casos graves.

Predicción

En situaciones epidémicas, realizar predicciones de calidad es importante para ayudar a los responsables de salud pública y a los gestores del sistema de salud a tomar decisiones con criterios lo más objetivos posibles. Cabe remarcar que las predicciones aportan una información relevante pero no la única necesaria para tomar decisiones, ya que se deben tener en consideración otros aspectos como las consecuencias sociales y económicas de las medidas, además de otros indicadores.

Como comentábamos al principio de este artículo, inicialmente las predicciones no se podían realizar con modelos mecanicistas ni con

métodos estadísticos avanzados. Los métodos empíricos, extremadamente simples pero independientes del conocimiento del sistema, sí que permitían realizar predicciones a 1-3 semanas. Después de evaluar diferentes opciones, decidimos utilizar la función de crecimiento de Gompertz [3]. Es una función muy parecida a la ecuación logística. Nos permitía predecir el comportamiento de la incidencia acumulada y el número de casos nuevos a lo largo del tiempo [4]. El método matemático de predicción lo fuimos mejorando progresivamente, por ejemplo, incrementando la ponderación de los últimos días en ciertos períodos o teniendo en consideración el patrón semanal del registro de casos de cada país o región. Comprobamos que con la función de Gompertz podíamos realizar predicciones a corto plazo de calidad, no solo para la incidencia, sino también para otros valores realmente importantes como el número de hospitalizados, el número de casos en unidades de cuidados intensivos o el número de fallecidos. Estas predicciones podían realizarse incluso para zonas geográficas correspondientes a un

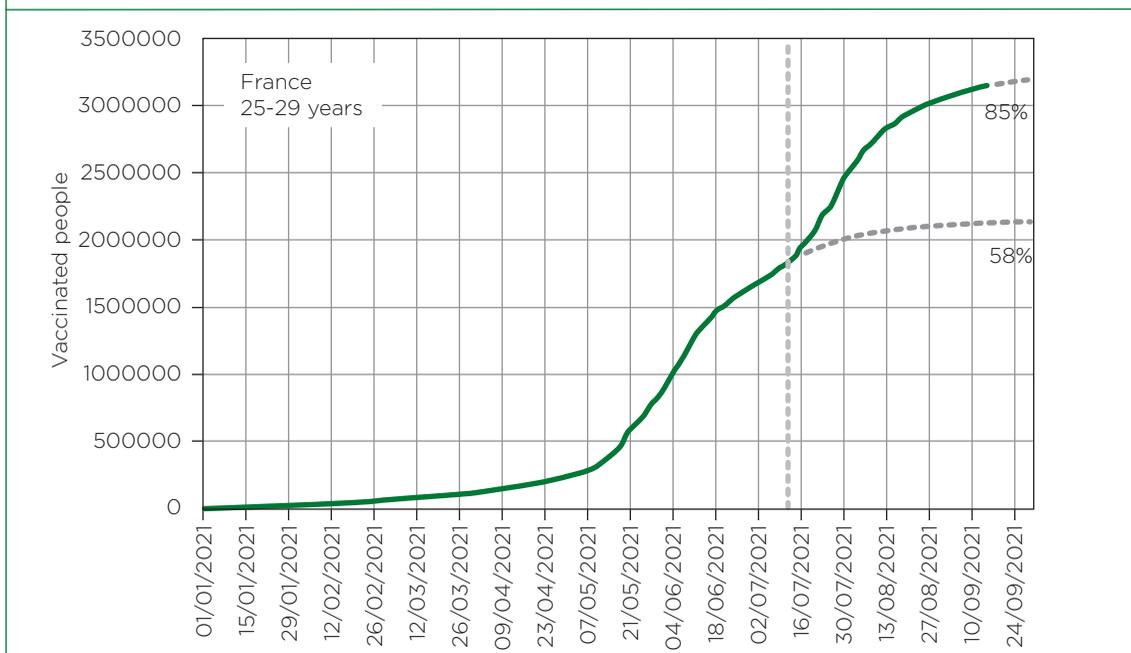
hospital de referencia, permitiendo así una optimización de la organización del hospital.

Retos

Los retos durante la pandemia han sido muchos y diferentes en cada una de las etapas que hemos pasado. En colaboración con expertos de los distintos ámbitos de estudio, hemos utilizado modelos estadísticos para estimar la afectación real de la primera ola [5], modelos estocásticos para comprender ciertos aspectos de la dinámica de transmisión en las escuelas [6], modelos basados en agentes para estudiar las distintas estrategias de vacunación [7], o modelos deterministas para estudiar y predecir los procesos de sustitución de las variantes [8], entre otros.

Una de las preguntas que intentamos responder fue sobre el efecto de la implantación del certificado COVID-19 en la campaña de vacunación francesa en verano de 2021. La evolución del porcentaje de población vacunada depende de diversos factores: de la disponibilidad de

FIGURA 2. EVOLUCIÓN DE LA VACUNACIÓN EN FRANCIA EN LA FRANJA DE EDAD ENTRE 25 Y 29 AÑOS. LA LÍNEA VERTICAL CORRESPONDE AL DÍA DE ANUNCIO POR PARTE DEL GOBIERNO FRANCÉS DE LAS NUEVAS CONSECUENCIAS DE NO ESTAR VACUNADO



Fuente: [3].

vacunas, de la infraestructura organizativa para realizar las vacunaciones y del compromiso de la población. Comprobamos que el modelo de crecimiento restringido [3], variando linealmente a lo largo del tiempo el valor máximo, nos describía correctamente las campañas de vacunación realizadas, como observamos en el ejemplo de la figura 2. Este simple modelo permitía prever cuál sería el valor máximo alcanzado, y mostraba en algunos casos la necesidad de promocionar de forma importante la vacunación. El ejemplo francés es excelente, a partir de cierto momento las autoridades explican que la vacunación será imprescindible para muchas actividades y, automáticamente, cambia el comportamiento.

Perspectivas

La vacunación y las altas incidencias de los últimos meses han cambiado drásticamente la situación. La pandemia no ha acabado, pero el número de casos graves o de muertes se ha reducido de forma importante. Poder prever los efectos a lo largo del tiempo de la disminución de la inmunidad poblacional es un nuevo reto. Los problemas a afrontar desde la estadística y desde la modelización matemática vuelven a ser nuevos. Debemos adaptarnos a nuevos sistemas de vigilancia epidemiológica y desarrollar nuevos indicadores, nuevos análisis. ●

Referencias

- [1] Català, M., Marchena, M., Conesa, D., Palacios, P., Urdiales, T., Alonso, S., Álvarez-Lacalle, E., López, D., Cardona, P. J., Prats, C. Monitoring and Analysis of COVID-19 Pandemic: The Need for an Empirical Approach. *Front Public Health*. 2021 Jul 8;9:633123. doi: 10.3389/fpubh.2021.633123
- [2] Català, M., Coma, E., Alonso, S., Álvarez-Lacalle, E., Cordomi, S., López, D., Fina, F., Medina-Peralta, M., Prats, C., Prieto-Alhambra, D. Risk Diagrams Based on Primary Care Electronic Medical Records and Linked Real-Time PCR Data to Monitor Local COVID-19 Outbreaks During the Summer 2020: A Prospective Study Including 7,671,862 People in Catalonia. *Front Public Health*. 2021 Jul 5;9:693956. doi: 10.3389/fpubh.2021.693956
- [3] Gamito, S. Growth models and their use in ecological modelling: an application to a fish population. *Ecological Modelling* 1998, 113, 83-94. doi: 10.1016/S0304-3800(98)00136-7
- [4] Català, M., Alonso, S., Álvarez-Lacalle, E., López, D., Cardona, P.J., Prats, C. Empirical model for short-time prediction of COVID-19 spreading. *PLoS Comput Biol*. 2020 Dec 9;16(12):e1008431. doi: 10.1371/journal.pcbi.1008431.
- [5] Català, M., Pino, D., Marchena, M., Palacios, P., Urdiales, T., Cardona, P.J., Alonso, S., López-Codina, D., Prats, C., Álvarez-Lacalle, E. Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers. *PLoS One*. 2021 Jan 7;16(1):e0243701. doi: 10.1371/journal.pone.0243701
- [6] Alonso, S., Català, M., López, D., Álvarez-Lacalle, E., Jordan, I., García-García, J.J., Fumadó, V., Muñoz-Almagro, C., Gratacós, E., Balanza, N., Varo, R., Millat, P., Baro, B., Ajanovic, S., Arias, S., Claverol, J., de Sevilla, M.F., Bonet-Carne, E., García-Miquel, A., Coma, E., Medina-Peralta, M., Fina, F., Prats, C., Bassat, Q. Individual prevention and containment measures in schools in Catalonia, Spain, and community transmission of SARS-CoV-2 after school re-opening. *PLoS One*. 2022 Feb 16;17(2):e0263741. doi: 10.1371/journal.pone.0263741
- [7] Català, M., Li, X., Prats, C., Prieto-Alhambra, D. The impact of prioritisation and dosing intervals on the effects of COVID-19 vaccination in Europe: an agent-based cohort model. *Sci Rep*. 2021 Sep 22;11(1):18812. doi: 10.1038/s41598-021-98216-0
- [8] Català, M., Martró, E., Noguera-Julian, M., Andrés, C., Cardona, P.J., Prats, C., Alonso, S., Álvarez, E., et al. Prediction and monitoring of the Delta variant substitution process in Catalonia (Spain). In: Analysis and prediction of COVID-19 for EU-EFTA-UK and other countries, 2021, Report 254, pp. 9-14. Disponible en: <https://upcommons.upc.edu/handle/2117/355115>

La iniciativa Valencia IA4COVID

J. Alberto Conejero

Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València (aconejero@upv.es)

Nuria Oliver

Ellis Alicante (nuria@ellisalicante.org)

A principios de 2020, pudimos observar cómo el virus del SARS-CoV-2 se fue extendiendo por todo el mundo de manera imparable. En paralelo a los esfuerzos realizados en la asistencia sanitaria para atender a un número de pacientes que crecía de manera exponencial, se comenzaron a utilizar modelos matemáticos y de inteligencia artificial para predecir la evolución de la pandemia. Para contener la propagación del coronavirus, la mayoría de países y regiones del mundo aplicaron distintas medidas no farmacológicas (*Non Pharmaceutical Interventions, NPIs*) como el cierre de fronteras, la suspensión de las clases o los confinamientos de la población.

En un esfuerzo pionero a nivel mundial, la Generalitat Valenciana creó en marzo de 2020 un grupo de expertos en Ciencias de Datos en la lucha contra COVID-19. El principal objetivo de dicho grupo era apoyar la toma de decisiones de políticas públicas durante la pandemia a través del análisis de datos. Los algoritmos y métodos empleados nos ayudaron a entender el estado de la pandemia en cada momento y a predecir su impacto a corto y medio plazo. Además, nos permitieron predecir la posible evolución de la curva pandémica bajo distintos escenarios de aplicación de medidas de confinamiento y de restricciones de movilidad a diferentes niveles, así como entender y modelar el impacto de la pandemia en la vida de las personas a través de una gran encuesta ciudadana (COVID19impactsurvey.org).

Durante dos años (marzo 2020 a abril 2022), el trabajo de este grupo se ha vertebrado en torno a 4 ejes:

1. Modelado de la movilidad humana a gran escala gracias al análisis de datos agregados

y anonimizados de movilidad compartidos por el INE¹, gracias a un acuerdo de colaboración entre el INE y las tres operadoras móviles más grandes en España (Telefónica, Vodafone y Orange).

2. Desarrollo de modelos epidemiológicos computacionales para predecir la evolución de la curva pandémica.

3. Desarrollo de modelos predictivos de ocupación hospitalaria en planta y en UCI.

4. Elaboración y difusión de una iniciativa de ciencia ciudadana a través de la macroencuesta *covid-19 impact survey*². Con más de 700.000 respuestas, es una de las mayores del mundo. Nos ha permitido entender³ el impacto de la pandemia en la vida de las personas⁴.

En paralelo a los esfuerzos realizados en la asistencia sanitaria para atender a un número de pacientes que crecía de manera exponencial, se comenzaron a utilizar modelos matemáticos y de inteligencia artificial para predecir la evolución de la pandemia

De todos estos trabajos, describimos en este artículo los modelos epidemiológicos basados en inteligencia artificial desarrollados en el contexto del *Pandemic Response Challenge*⁵, organizado por la Fundación XPRIZE⁶ y patrocinado por la empresa tecnológica Cognizant. Dicho certamen tuvo lugar desde noviembre de 2020 hasta marzo de 2021 y estaba formu-

1 https://www.ine.es/covid/covid_movilidad.htm

2 <https://covid19impactsurvey.org/>

3 <https://www.nature.com/articles/s41598-022-05041-0>

4 <https://www.nature.com/articles/s41598-021-97645-1>

5 <https://www.xprize.org/challenge/pandemicresponse>

6 <https://www.xprize.org/>

lado en un contexto de innovación abierta. Su objeto consistía en desarrollar modelos que predijeran la evolución de la pandemia bajo ciertos escenarios de aplicación de medidas no farmacológicas y proponiendo la aplicación de los que tenían menor coste para un mismo grado de contención. Dicha competición constaba de dos fases:

- ▶ En la primera fase había que desarrollar un sistema de inteligencia artificial que pudiera predecir el número de casos de COVID-19 en 236 países y regiones del mundo, a 180 días en el futuro y teniendo en cuenta las NPIs adoptadas en cada uno de esos países y regiones. Dichos modelos eran entrenados a partir de la combinación de las series temporales de casos y de NPIs aplicadas en cada lugar del mundo desde el inicio de la pandemia hasta la fecha en la que se entrenaban los modelos. Ambos conjuntos de datos se encontraban disponibles en el Oxford COVID-19 Government Response Tracker (OxCGRT)⁷.
- ▶ A partir de las predicciones de los modelos de la primera fase, había que desarrollar un prescriptor que propusiera diferentes escenarios de contención de la pandemia, combinando en cada caso diferentes NPIs y diferentes niveles de aplicación de las mismas. Dentro de estos escenarios había que ofrecer escenarios de contención de diferentes niveles, tanto blandos como intermedios y duros. Del mismo modo, se deseaba que los planes fueran óptimos en el sentido de que, para un mismo nivel de dureza, fueran los que tuvieran un menor coste económico y social.

El equipo Valencia IA4COVID⁸ del que fuimos miembros junto a otros 12 investigadores e investigadoras de universidades y centros de investigación de la Comunitat Valenciana, fue declarado ganador mundial de la competición, siendo la primera vez que un equipo

español ganaba un certamen de la Fundación XPRIZE.

Nuestros modelos predictivos de la primera fase estaban basados en una arquitectura de redes neuronales profundas que combinaba varias capas de convolucionales con redes neuronales recurrentes de tipo LSTM bidireccionales. Nuestros modelos funcionaron muy bien en esta fase estando en el top 3 en la suma de mejores modelos en cada país o región. No obstante, lo más destacado es que las predicciones de nuestros modelos fueron compartidas con la Generalitat Valenciana a partir de la tercera ola la pandemia que se inició en diciembre de 2020. Dichas predicciones fueron de gran utilidad en la toma de decisiones de políticas públicas para contener la evolución pandemia, en unas circunstancias críticas con un constante crecimiento del número de casos y con una presión hospitalaria cada vez mayor sin que la vacunación hubiera prácticamente comenzado.

En la segunda fase del reto desarrollamos un sistema recomendador de medidas no farmacológicas o NPIs que mostrara hasta 10 posibles políticas públicas distintas, mostrando su efecto en las curvas de número de casos confirmados, así como su coste. Dichos escenarios debían tener el mejor balance entre su coste y el número de casos de COVID-19 que resultaría si se aplicaran. El hecho de que no hubiera un único plan prescrito permitía que las autoridades gubernamentales pudieran decidir de entre ellos los más adecuados teniendo presente otras circunstancias como la situación económica y el agotamiento de la población. Las intervenciones no farmacológicas consideradas incluían medidas para reducir la movilidad, el fomento del teletrabajo, el cierre parcial o total de centros educativos, así como limitar la actividad económica, las concentraciones multitudinarias o limitar las reuniones familiares.

Al margen del Pandemic Response Challenge, nuestro modelo Valencia IA4COVID⁹ de predicción del número de casos confirmados y de prescripción de planes de intervención de NPIs fue presentado en el mayor con-

⁷ <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>

⁸ <https://ellisalicante.org/xprize>

⁹ https://ellisalicante.org/xprize_valencia

greso europeo sobre aprendizaje automático: la Conferencia Europea sobre Aprendizaje Automático y Principios y Práctica del Descubrimiento del Conocimiento en Bases de Datos (ECML PKDD 2021)¹⁰. Allí recibimos el premio al mejor artículo en ciencias de datos aplicadas¹¹. Gracias al apoyo de la Generalitat Valenciana, actualmente nos encontramos estudiando el impacto de la vacunación y de las diferentes variantes del virus a lo largo de la pandemia, con el fin de adaptar los modelos predictivos a estas circunstancias. Asimismo, nos encontramos estudiando con mayor exactitud los costes asociados a cada nivel de aplicación de las NPIs, de manera que los modelos puedan ofrecer medidas mucho más precisas.

Para nosotros ha sido muy gratificante trabajar durante estos casi 2 años en el desarrollo de estos modelos, colaborando estrechamente con las administraciones públicas de nuestro entorno, así como involucrando a la población en una iniciativa de ciencia ciudadana que permitió disponer de información de primera mano sobre la evolución de la pandemia, sobre todo en los momentos más duros y cuando la disponibilidad de datos era limitada.

Los miembros de Valencia IA4COVID son Nuria Oliver (ELLIS Alicante) y J. Alberto Conejero (Universitat Politècnica de València) (colíderes del equipo); Francisco Escolano y Miguel Ángel Lozano (Universidad de Alicante); Miguel Rebollo, Miguel Ángel García-March, Oscar Garibo, Eloy Piñol, Víctor de Elena y Manuel Portolés (Universitat Politècnica de València); Kristina Polotskaya, Aurola Mula, Elisa Espin (Universidad Miguel Hernández) y Adolfo López (FISABIO).

Este trabajo ha sido financiado por FONDO SUPERA COVID-19 Santander-CRUE (CD4COVID19 2020-2021), Fundación BBVA for SARS-CoV-2 research (IA4COVID19 2020-2022) y por la Generalitat Valenciana, proyecto VALENCIA IA4COVID (GVA-COVID19/2021/100). ELLIS Alicante cuenta con financiación a través de un convenio de colaboración firmado en 2021 (Convenio Singular 2021) entre la Fundación Unidad ELLIS Alicante y la Generalitat Valenciana (Conselleria d'Innovació, Universitats, Ciència i Societat Digital, Direcció General para el Avance de la Sociedad Digital). ●

¹⁰ <https://2021.ecmlpkdd.org/>

¹¹ https://link.springer.com/chapter/10.1007/978-3-030-86514-6_24

Para saber más...

- De Nadai, M., Roomp, K., Lepri, B. et al. The impact of control and mitigation strategies during the second wave of coronavirus infections in Spain and Italy. *Sci Rep* 12, 1073 (2022). <https://doi.org/10.1038/s41598-022-05041-0>
- Lozano, M.A., Orts, Ò.G., Piñol, E., Rebollo, M., Polotskaya, K., García-March, M.A., Conejero, J.A., Escolano, F., Oliver, N. Open Data Science to Fight COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge (2021) *Lecture Notes in Artificial Intelligence*, 12978 LNAI, pp. 384 - 399. https://doi.org/10.1007/978-3-030-86514-6_24
- Martínez-García, M., Rabasa, A., Barber, X. et al. Key factors affecting people's unwillingness to be confined during the COVID-19 pandemic in Spain: a large-scale population study. *Sci Rep* 11, 18626 (2021). <https://doi.org/10.1038/s41598-021-97645-1>
- Pandemic Response Challenge organized by the XPRIZE Foundation and sponsored by Cognizant <https://www.xprize.org/challenge/pandemicresponse>
- Oxford Covid-19 Government Response Tracker (OxCGRT) <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>
- VALENCIA IA4COVID https://ellisalicante.org/xprize_valencia

Tasas de mortalidad de la población española en 2020: efectos de la pandemia por COVID-19 y de las medidas para contenerla

Eliseo Navarro

Universidad de Alcalá. (eliseo.navarro@uah.es)

Pilar Requena

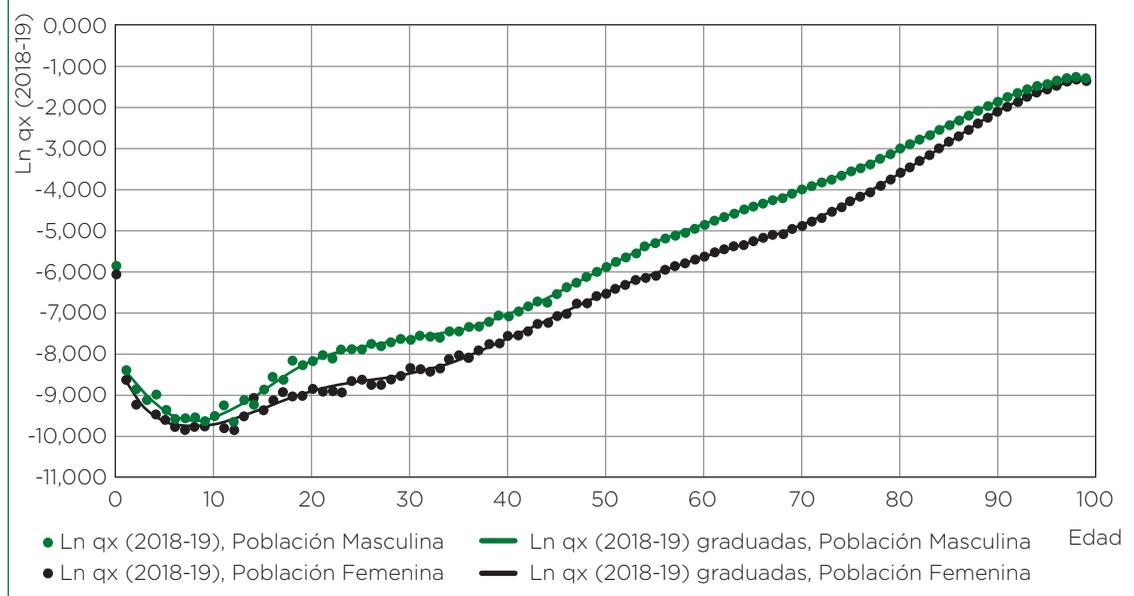
Universidad de Alcalá. (pilar.requena@uah.es)

En la literatura científica, se han realizado trabajos que señalan la edad y el sexo como los principales factores de riesgo de mortalidad por COVID-19, de forma que los varones de edades avanzadas habrían sufrido mayor mortalidad que los más jóvenes, y a su vez, a todas las edades, mayor mortalidad que la sufrida por la po-

blación femenina. La estructura de edades de la población se ha considerado un factor clave en la mortalidad por COVID-19. Estos trabajos tienen en común planteamientos metodológicos basados en la utilización del número de fallecimientos durante el período de estudio considerado.

Sin embargo, la tasa de mortalidad es la medida más apropiada para realizar un análisis retrospectivo del efecto de la pandemia por COVID-19 sobre la mortalidad del año 2020. Aplicando metodología actuarial es posible estimar las tasas de mortalidad para cada edad, correspondientes a dos períodos

FIGURA 1. TASAS BRUTAS DE MORTALIDAD (PUNTOS) Y GRADUADAS (LÍNEA) DE LA POBLACIÓN ESPAÑOLA MASCULINA Y FEMENINA DURANTE EL PERÍODO 2018-19



Fuente: Elaboración propia con datos obtenidos del INE y Eurostat.

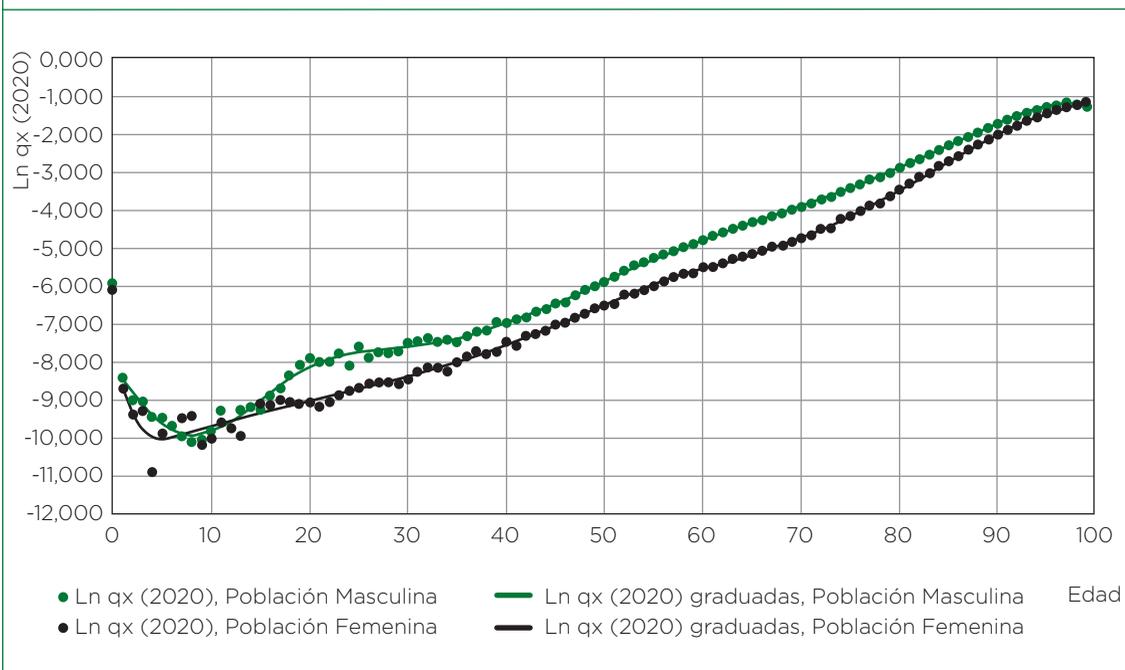
consecutivos (períodos 2018/19 y 2020), para las poblaciones masculina y femenina. A continuación, se procede a su comparación para la obtención de la variación relativa de las tasas de mortalidad de un período en relación con el período anterior. Los resultados obtenidos evidencian un efecto heterogéneo de la pandemia por COVID-19 y de las medidas utilizadas para su contención, sobre los distintos tramos de edades de la curva de mortalidad española, tanto para la población masculina como femenina. Destaca la importante disminución de las tasas de mortalidad infantil durante el año 2020 en relación con el período anterior, y el incremento relativo de la mortalidad de las mujeres entre los 25 y 45 años. A partir de los 70 años aproximadamente, ha habido un incremento relativo de la mortalidad muy similar en la población masculina y femenina, cuya magnitud tiende a disminuir a partir de los 80-85 años. Estos dos tramos de edades de la curva han sido los que han sufrido el mayor incremento relativo en su mortalidad durante el período 2020.

Estimación de las tasas de mortalidad para los períodos 2018/19 y 2020¹.

Con los datos estadísticos sobre Población publicados por el INE, y los datos de fallecimientos por edad y sexo publicados por Eurostat, se han calculado las tasas brutas de mortalidad para los períodos 2018/19 y 2020. Con el fin de eliminar la aleatoriedad inherente al número de fallecimientos, se ha procedido a la graduación de las tasas brutas de mortalidad para los dos períodos. La graduación de la curva de mortalidad permite encontrar una función dependiente de la edad, lo suficientemente suavizada para describir el comportamiento de las tasas de mortalidad, y cuyo ajuste permita obtener

¹ En el documento “Descriptive Analysis of the Impact of COVID-19 and Anti-Pandemic Measures on Mortality Rates in the Spanish Population in 2020 Using Graduation Techniques” (March 22, 2022) realizado por los autores puede consultarse el detalle de los cálculos realizados y la bibliografía utilizada. Disponible en <https://ssrn.com/abstract=4063837>.

FIGURA 2. TASAS BRUTAS DE MORTALIDAD (PUNTOS) Y GRADUADAS (LÍNEA) DE LA POBLACIÓN ESPAÑOLA MASCULINA Y FEMENINA DURANTE EL PERÍODO 2020



Fuente: Elaboración propia con datos obtenidos del INE y Eurostat.

estimaciones precisas de la tasa de mortalidad a la edad x . Con este doble propósito, el procedimiento de graduación utilizado ha sido no paramétrico, específicamente mediante funciones spline cúbicas con nudos equidistantes (8 nudos para las edades de 1 a 99 años). La presentación de los resultados obtenidos para el logaritmo de las tasas brutas de mortalidad y las tasas graduadas se presenta en las figuras 1 y 2. La figura 1 contiene los resultados para el período 2018-19, para la población española masculina y femenina. Y la figura 2 contiene los mismos resultados para el período 2020.

La calidad de la graduación se ha probado mediante el test de signos y el test de rachas para determinar si el número de desviaciones positivas y negativas están distribuidas de forma aleatoria a lo largo de la curva de mortalidad. También se ha analizado el número de tasas de mortalidad graduadas que caen fuera de los intervalos de confianza, calculados estos al 95% y al 99%.

Los resultados de los test de signos y de rachas, tanto en la graduación realizada para el período 2018/19 como para el período 2020 son satisfactorios, por lo que es posible afirmar que no hay tramos de la curva de mortalidad en los que las tasas de mortalidad estén sistemáticamente sobreestimadas o subestimadas. Sin embargo, el número de tasas graduadas fuera de los intervalos de confianza excede el número esperado. Se ha observado que las desviaciones se concentran en algunas edades en las que ocurren cambios abruptos en las tasas de mortalidad brutas, mostrando un comportamiento errático. Para el período 2018-19 este comportamiento se ha observado en la población masculina, en el intervalo de edades entre los 50 y 60 años. Para el período 2020, en la población masculina, entre las edades 20 y 26 años, y entre 72 y 81 años. Y para la población femenina, las desviaciones están en el intervalo de 73 a 84 años, dando resultados muy satisfactorios en el resto de las tasas de la curva de mortalidad. El hecho de que coincidan las desviaciones en ambas poblaciones entre los 72 y 84 años, puesto que solo dependen de los datos y no de la graduación, da lugar a pensar que pueda existir algún problema con los datos.

Comparación de tasas de mortalidad del período 2020 en relación con las tasas de mortalidad del período 2018-19

La tendencia de disminución de las tasas de mortalidad española observada durante las últimas décadas (a excepción del SIDA en los años 90) ha sido alterada en 2020 por la aparición de la pandemia. Para realizar un análisis detallado del impacto de la pandemia sobre la curva de mortalidad en el año 2020, se han comparado directamente las tasas de mortalidad del período 2018-19 y 2020. Si la ratio entre las tasas de mortalidad del período 2020 y del período 2018-19, ya sean tasas brutas o tasas graduadas, es superior a 1, significa que la mortalidad del período 2020 se ha incrementado en relación con la mortalidad del período 2018-19. La figura 3 muestra los resultados de la ratio con tasas graduadas, para la población masculina y femenina.

Los principales resultados y conclusiones de la comparación son:

- ▶ La intensa disminución de la mortalidad infantil durante el año 2020 en relación con el período 2018-19, un shock con forma de joroba invertida, de diferente intensidad en la población masculina y femenina. Entre los varones la máxima disminución relativa ha sido aproximadamente el 25%, y ha tenido lugar alrededor de los 9 años, pero la disminución se ha producido entre 1 y 19 años. Entre las mujeres el comportamiento ha sido diferente. Ha existido una disminución muy intensa entre 1 y 9 años aproximadamente, alcanzando la disminución relativa máxima, aproximadamente 40%, alrededor de los 4 años. Entre los 9 y los 14 años se produce un ligero incremento de la mortalidad, pero no hay una variación relativa significativa. Sin embargo, desde los 14 hasta aproximadamente los 26 años se produce otra disminución relativa de mortalidad con forma de joroba invertida.
- ▶ En la población masculina entre los 25 y 45 años ha habido un incremento relativo de la mortalidad entre el 7,2% y el 9,2%. Hay una segunda sección entre los 45 y 60 años, con una joroba inver-

tida, con una disminución relativa mínima alrededor de los 55 años. Desde los 60 años comienza un incremento relativo que se estabiliza alrededor de los 75 años aproximadamente. A partir de entonces, en las edades más avanzadas, el incremento relativo disminuye.

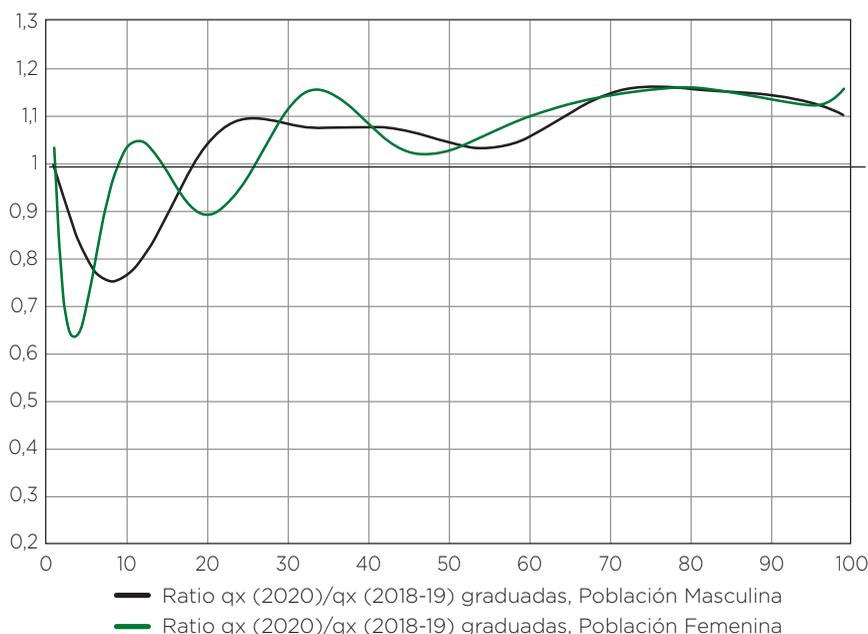
- En la población femenina, el comportamiento ha sido diferente. Entre los 26 y 45-50 años ha habido un repunte significativo en las tasas de mortalidad, con un pico aproximado del 15% alrededor de los 34 años. La magnitud del impacto de la mortalidad en este tramo de edades ha sido más severa en la población femenina que en la masculina. Si bien este fenómeno necesita ser investigado, es posible que sea debido a las ocupaciones de las mujeres. En la estadística publicada por el INE sobre ocupados por sexo y rama de actividad, puede comprobarse que, en el año 2020, el 15,4% (en número 1.350.400) de las mujeres estaban ocupadas en actividades sanitarias y de servicios sociales frente al 4%

(en número 418.900) de los hombres. A su vez, el 50% de estas mujeres tenían edades comprendidas entre los 30 y 49 años.

- A partir de los 50 años, en la población femenina vuelve a incrementarse gradualmente la mortalidad relativa, y a partir de los 68-70 años ese incremento relativo se estabiliza, y coincide con el de la población masculina. E igual que en la población masculina, a las edades más avanzadas el incremento relativo tiende a disminuir.

La pandemia por COVID-19 y las medidas utilizadas para su contención han afectado de forma heterogénea a distintos grupos de edades de la curva de mortalidad, y a la población masculina y femenina. En las edades infantiles ha habido una mejora de mortalidad relativa. En las edades avanzadas ha habido un incremento relativo, con un impacto prácticamente similar en hombres y mujeres. Para las demás edades destaca el importante incremento relativo de mortalidad de la población femenina en el intervalo de 26 a 45-50 años. ●

FIGURA 3. RATIO ENTRE LAS TASAS DE MORTALIDAD GRADUADAS DEL PERÍODO 2020 Y DEL PERÍODO 2018-19, POBLACIÓN MASCULINA (LÍNEA NEGRA) Y POBLACIÓN FEMENINA (LÍNEA VERDE)



Fuente: Elaboración propia con datos obtenidos del INE y Eurostat.

EpData: informando sobre el coronavirus

Iñaki Hernández

*Coordinador Sección de internet y EpData.
Europa Press*

La última vez que lo comprobamos, llevábamos más de doce versiones del script de R que usamos en EpData para ‘escrapear’ el pdf del Ministerio de Sanidad con los datos del balance del coronavirus en España. Cada cambio de criterio de Sanidad, cada dato que decidíamos incorporar a nuestro seguimiento, nos obligaba a retocar el código y añadir una nueva versión hasta superar la docena. Y eso fue antes de que el Gobierno decidiera dar un giro, una vez más, a su estrategia de seguimiento y anunciara que dejaba de medir la incidencia para la población general. Ahora ya, hemos perdido la cuenta.

Es una anécdota, pero de algún modo ejemplifica cómo esta pandemia nos ha obligado a todos a adaptarnos e ir aprendiendo sobre la marcha. Cuando decidimos cubrir desde EpData el nuevo brote de coronavirus, se trataba de una noticia limitada a China. De hecho, la página web que lanzamos para seguir la enfermedad se llamaba ‘Coronavirus en China, en gráficos’.

Durante semanas, fuimos testigos de cómo los datos de China iban empeorando. Con dificultades por la poca transparencia del país, lo fuimos reflejando en gráficos. Luego llegó Italia. Y de repente nos encontramos encerramos en casa, con un proyecto que se llamaba coronavirus en China, pero que, de una manera inimaginable, al menos inimaginable hace menos de un mes, nos afectaba directamente.

Fueron días complicados, en los que la información llegaba “a cuentagotas”. La primera versión del balance de Sanidad, de febrero de 2020 –ahora ya vamos por unos 600 balances– daba información sobre la incidencia, nuevos casos y fallecidos, pero no ofrecía información desagregada ni una serie en la que ver la evolución ni datos en formato reutilizable.

Rápidamente vimos la necesidad de ofrecer a los medios y usuarios información de lo que estaba pasando en un formato abierto. Creamos series de datos que actualizamos diariamente con ese balance. Según la situación empeoraba, y los balances de Sanidad ganaban en complejidad, nosotros también íbamos sofisticando nuestra oferta de datos en abierto.

La gente buscaba desesperadamente información sobre la enfermedad. El tráfico de la web se disparó con búsquedas de personas anónimas, periodistas e investigadores que usaban las series que íbamos construyendo día a día para sus reportes. Medios locales y nacionales comenzaron a usar nuestros gráficos para insertarlos en las noticias sobre la pandemia.

La falta de información, al principio, y los cambios de criterio obligados por la situación después, nos forzaron a nosotros también a ir tomando decisiones difíciles. Cuando se empezó a debatir sobre la necesidad de mirar a la cifra de fallecidos totales, ante la limitación que suponía el número de fallecidos confirmados de COVID-19 por el colapso del sistema, recurrimos a los registros civiles para construir una nueva serie. Cuando se demostró la buena correlación que existía entre los desplazamientos y los repuntes de casos, incorporamos esa información. Aprendíamos y crecíamos sobre la marcha.

No lo tuvimos que construir desde cero. EpData nació en 2017 como la apuesta de Europa Press por el periodismo de datos. Financiado por la iniciativa Digital News Initiative (DNI, por sus siglas en inglés) de Google, la idea de EpData es trasladar al mundo del periodismo de datos la filosofía de una agencia de noticias.

Mucho antes de que se hablara de periodismo de datos en España, los periodistas de la agencia de noticias ya trabajaban con estadísticas. Informaban del paro, de la evolución de casos de violencia de género o de la incidencia de la gripe, por poner tres ejemplos dispares, y contextualizaban el dato puntual, noticioso, con sus propias series estadísticas, recopiladas

durante años y guardadas en ocasiones en libretas o cuadernos en cajones.

Lo que hicimos al crear EpData fue volcar todo ese conocimiento que ya existía en la agencia en una gran base de datos compartida y crear sobre ella algunas herramientas que creíamos que podían ser útiles para visibilizar y contextualizar la información estadística, como los gráficos interactivos. Porque no es lo mismo informar de que el paro sube en X personas, que poner esa subida en contexto y mostrar un gráfico con la tendencia histórica.

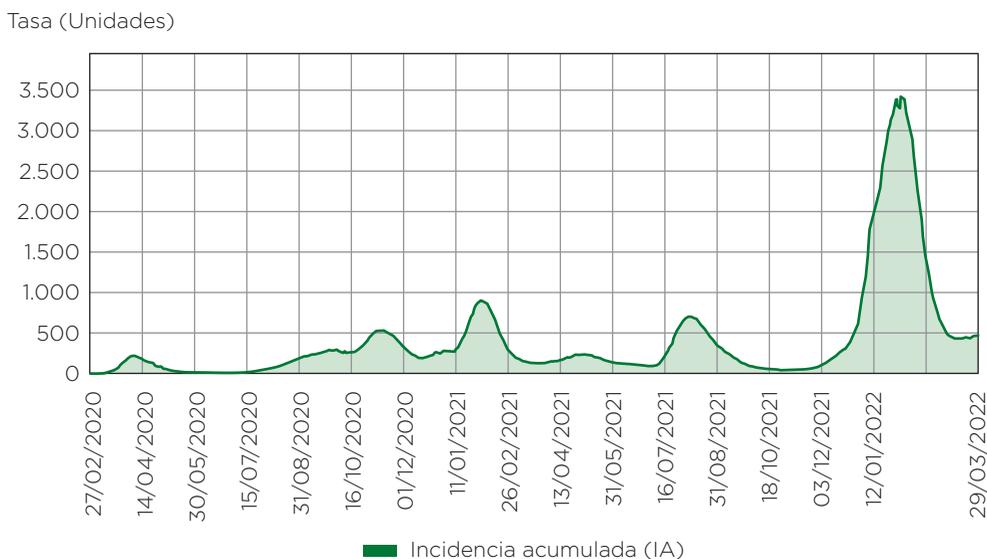
Igual que no es lo mismo decir que la incidencia sube un miércoles en 30 puntos, que ofrecer una serie de datos de libre uso para que cada uno pueda hacer sus propios análisis. En nuestro trabajo con el COVID-19 es algo que aprendimos rápido. Éramos conscientes de que los datos se veían afectados por cosas ajenas a la evolución de la pandemia. Por ejemplo, notamos que existía una estacionalidad en los datos que dependía no de las subidas y bajadas de las sucesivas olas que golpeaban al país, sino simplemente del número de pruebas que se hacían. Aprendimos que cualquier subida en la incidencia de casos iba seguida de una subida en las hospitalizaciones y en las muertes (al

menos hasta la llegada de las vacunas), tratamos de explicar, muchas veces, la diferencia entre las muertes notificadas y los fallecidos por fecha de defunción...

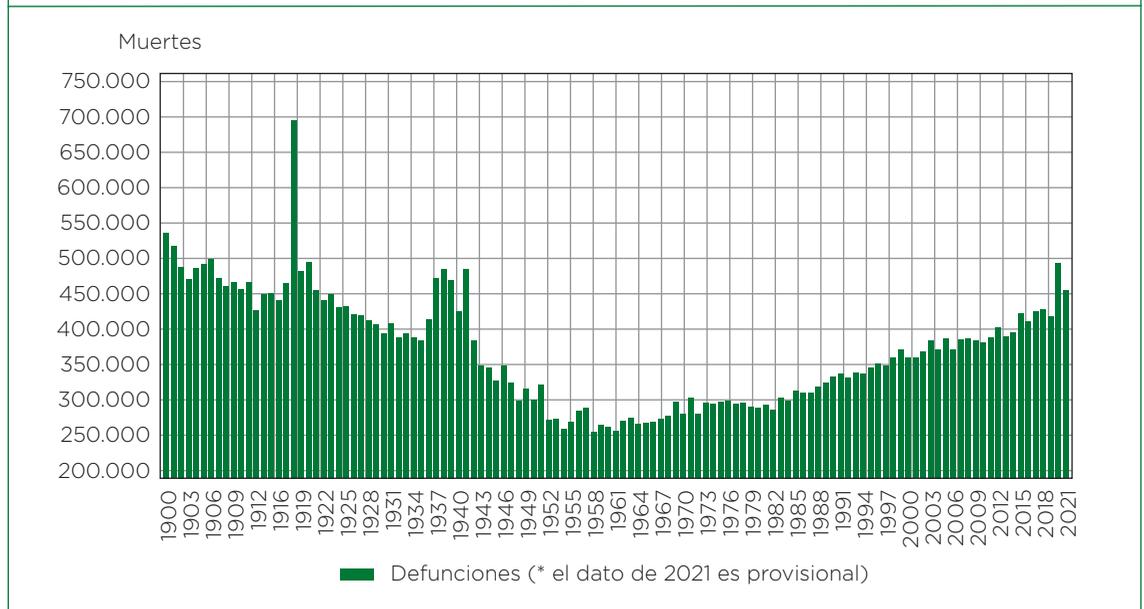
Ofrecer el dato en crudo a la gente siempre es peliagudo. “Suben los casos en 100.000”. ¿Es eso mucho? ¿Es poco? Depende del contexto, de donde se venga, de en qué punto se esté de la pandemia, del número de test... Posiblemente una de las principales virtudes del proyecto fue precisamente eso, que ofrecía contexto al lector para que él mismo decidiera. De nuevo, una premisa del periodismo de agencia: ante todo informar, sin opinar.

Casi de manera orgánica, desde las primeras series que construimos de casos y fallecidos, fuimos levantando un entramado crecientemente complejo para ofrecer más y más contexto: incidencia, número de test, casos en otros países, número de desplazamientos, muertes por cualquier causa, datos desagregados por comunidades autónomas, información local, el impacto económico, número de sanciones por saltarse el confinamiento, impacto en las residencias de ancianos... Las temáticas se multiplicaban según la pandemia iba golpeando y la administración pública refinaba su oferta de

FIGURA 1. TASA DE CASOS DE CORONAVIRUS POR CADA 100.000 HABITANTES EN LOS ÚLTIMOS 14 DÍAS EN ESPAÑA



Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad, www.epdata.es

FIGURA 2. MUERTES OCURRIDAS EN ESPAÑA DESDE 1900Fuente: INE, www.epdata.es

información. Allí dónde no llegaban los datos públicos, hacíamos solicitudes de información a Transparencia.

Teníamos modelos en los que inspirarnos. Ejemplos como el excepcional trabajo de agregación y liberación de datos realizado por 'Our World in Data' o por la Universidad Johns Hopkins, del que nos hemos beneficiado medios e

investigadores de todo el mundo. La pandemia ha servido para arrojar luz sobre las carencias en los sistemas de recopilación de datos de algunos países y también para mostrarnos el camino hacia sistemas de información pública mucho más detallados y en formato reutilizable, que se benefician de las nuevas tecnologías y del ambiente colaborativo propio de Internet para enriquecer y mejorar la capacidad de la opinión pública de fiscalizar a los gobiernos y empresas privadas.

Nada tiene que ver el primer balance que Sanidad ofreció allá por febrero de 2020 sobre el coronavirus con el número 600 por el que va ya. La información ha mejorado mucho. El Instituto Nacional de Estadística (INE) ha hecho un esfuerzo tremendo, que desde el mundo del periodismo se agradece, por ofrecer nuevas estadísticas experimentales, lanzadas en tiempos inimaginables antes de que surgiera la epidemia, con la intención de mejorar la transparencia sobre la enfermedad. Pero todavía queda camino de mejora. En EpData compartimos esa apuesta por la transparencia, por seguir aprendiendo. Y ahora que parece que la pandemia ha tomado otro curso, seguiremos adaptándonos, aprendiendo, enfocados en el objetivo de ofrecer la mejor información, en pro de la transparencia. ●

Para saber más...

Algunos datos sobre EpData:

- Fue creado en 2017 como la apuesta por el periodismo de datos de EpData.
- Contiene más de 60 millones de datos de cientos de fuentes públicas y privadas en temáticas tan diversas como empleo, población, violencia de género, salud o resultados empresariales entre otras muchas.
- La información estadística en EpData está actualizada y validada por un grupo de periodistas de la agencia de noticias Europa Press.
- Todas las estadísticas agrupadas en EpData son accesibles a través de un registro gratuito.

Estadísticas experimentales en tiempos de COVID-19

Adolfo Gálvez

S. G. de Difusión Estadística.

Instituto Nacional de Estadística

En 2019 el INE comenzó la publicación de estadísticas experimentales, proyectos de operaciones estadísticas que aún en desarrollo, cuentan con aspectos innovadores en aspectos como el uso de nuevas fuentes de datos, nuevos métodos estadísticos o la forma de difundir sus resultados.

Aunque estos proyectos no tienen aún la consideración de estadísticas oficiales, sus resultados se ponen a disposición de los usuarios con el objetivo de que puedan valorar si se ajustan a sus necesidades y detectar posibles mejoras que puedan incorporarse en ellos.

Las necesidades de información detectadas con motivo de la crisis COVID-19 han impulsado la publicación de nuevas estadísticas experimentales desde 2020. Estos estudios han permitido el seguimiento de algunos fenómenos sanitarios, demográficos y económicos que han sido objeto de atención en este periodo.

Se presentan a continuación de forma esquemática las estadísticas experimentales publicadas hasta la fecha de cierre de este número. Algunas de estos proyectos se han incorporado ya al programa de estadísticas oficiales siendo incluidas en los instrumentos de planificación habituales.



Estimaciones de defunciones semanales

Objetivo: Ofrecer estimaciones rápidas sobre el número de defunciones ocurridas cada semana a partir de las inscripciones registradas en los Registros Civiles informatizados, así como su comparación con los datos históricos de defunciones desde el año 2000. Esto permite interpretar los datos con una perspectiva necesaria, dada la variabilidad que presentan las defunciones a lo largo del tiempo.

Metodología: Estimación realizada a partir de la información administrativa para corregir la falta de cobertura y retrasos en la grabación de los fallecimientos. Las estimaciones se refieren al lugar de inscripción de la defunción.

Contenido de la publicación: Defunciones semanales por sexo y grupos quinquenales de edad desde el año 2000.

Detalle geográfico: Nacional, CC. AA., provincia, isla.

Nota: Actualmente es una estadística oficial que forma parte del Programa Anual 2022.

Estimación mensual de nacimientos

Objetivo: Ofrecer estimaciones rápidas sobre el número de nacimientos ocurridos cada mes, así como su comparación con los datos históricos de nacimientos.

Esto permite interpretar los datos con una perspectiva histórica necesaria, dada la variabilidad y paulatina disminución que presentan los nacimientos en los últimos años.

Metodología: La estimación se realiza a partir de la información administrativa para

corregir la falta de cobertura y retrasos en la grabación de los nacimientos. Las estimaciones se refieren al lugar de inscripción del nacimiento.

Contenido de la publicación: Nacimientos mensuales por grupo de edad de la madre desde el año 2016.

Detalle geográfico: Nacional, CC. AA., provincia.

Nota: Actualmente es una estadística oficial que forma parte del Programa Anual 2022.

Estudios de movilidad a partir de la telefonía móvil

Objetivo: Medir la movilidad de la población entre áreas a partir de la posición de teléfonos móviles.

Con motivo del brote de COVID-19 en marzo de 2020, este estudio de movilidad se ha realizado en distintos periodos ofreciendo información relevante en relación con las medidas de restricción de movilidad adoptadas en esta crisis.

Metodología: Se divide el territorio nacional en 3.214 “áreas de movilidad” específicas para el proyecto, cada una de ellas formada por un mínimo de 5.000 habitantes y una media de casi 15.000 habitantes. Esta es la unidad mínima de observación sobre la que las operadoras de telefonía móvil remiten al INE la información

agregada y anonimizada para su integración y publicación.

Contenido de la publicación: El proyecto original EM-1 con referencia en 2019 incluye tres tipos de matrices de movilidad:

- ▶ Matriz 1: movilidad cotidiana.
- ▶ Matriz 2: población de día y de noche.
- ▶ Matriz 3: población estacional.

En otras realizaciones de este estudio se ofrecen únicamente las matrices 1 y 3.

Se ofrece la consulta de datos de cada una de las matrices y los mapas que permiten el análisis de los flujos de movimientos de población de unas áreas a otras.

Detalle geográfico: CC. AA., provincias, áreas de movilidad.

Atlas de distribución de renta de los hogares

Objetivo: Construcción de indicadores estadísticos de nivel y distribución de renta de los hogares a nivel municipal e inframunicipal, a partir del enlace de información del INE con datos tributarios, fundamentalmente de la AEAT, pero también conteniendo información de las Haciendas Forales.

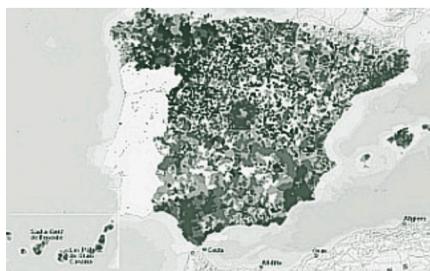
Metodología: Combinación de datos entre fuentes tributarias y el fichero estadístico de población. En esta

combinación se realizan los cálculos de renta media para distintas unidades estadísticas (persona, hogar, unidad de consumo) e indicadores de desigualdad.

Contenido de la publicación: Se ofrece la consulta de mapas y descarga de datos sobre los indicadores que comprende esta operación.

Detalle geográfico: Municipio, distrito, sección censal

Nota: Actualmente es una estadística oficial que forma parte del Programa Anual 2022.



Indicador multidimensional de calidad de vida

Objetivo: Obtener un indicador compuesto de medición de calidad de vida que se presenta con carácter experimental, sometido al escrutinio de los usuarios con el fin de mejorar su metodología y culminar, deseablemente, en un indicador multidimensional de calidad de vida estable dentro de la producción oficial de INE.

Metodología: El IMCV se construye a partir de los Indicadores de calidad del INE, que ofrecen una visión panorámica de la calidad de vida en España, mediante la elección de un conjunto amplio pero limitado de indicadores que cubren las distintas dimensiones consideradas: condiciones materiales de vida, trabajo, salud, educación, ocio y relaciones sociales,

seguridad física y personal, gobernanza y derechos básicos, entorno y medio ambiente y, por último, experiencia general de la vida. Los indicadores se combinan en cada dimensión según el método el método de Mazziotto-Pareto ajustado (AMPI). Los indicadores AMPI de cada dimensión se agregan en un único indicador de calidad de vida por media aritmética, es decir, dando a todas las dimensiones el mismo peso.

Contenido de la publicación: Se ofrecen datos y herramientas de visualización que permiten analizar la evolución del IMCV a nivel nacional y por CC. AA. Además, se ofrece la posibilidad de que el usuario pueda construir su propio IMCV ajustando las ponderaciones de las distintas dimensiones de manera personalizada.

Detalle geográfico: Nacional, CC. AA.

Índice de Precios de Vivienda en Alquiler

Objetivo: Disponer de un indicador de la evolución anual de los precios de la vivienda alquilada como residencia habitual de los hogares.

Metodología: La fuente de información utilizada para su cálculo es la base de datos proporcionada por la AEAT, que procede del modelo 100 de las declaraciones del IRPF, completada con datos de la Dirección General del Catastro sobre características de las viviendas como la superficie o el tipo de edificio. Cada año, se recoge la información declarada a efectos tributarios de los bienes inmuebles sobre los que se han declarado ingresos por arrendamiento de vivienda para su uso habitual.

El sistema de cálculo del IPVA está basado en la combinación de dos elementos básicos: los precios de las viviendas en alquiler, y las ponderaciones, o importancia relativa de cada tipología de vivienda según el gasto que los hogares realizan en el alquiler de viviendas de esa tipología respecto al gasto total en alquiler de vivienda. La combinación de estos dos elementos para obtener el IPVA se realiza mediante la fórmula del índice de Laspeyres en cadena.

Contenido de la publicación: Índice y variación anual de precios según distintas

características de las viviendas alquiladas (tipo de edificación, superficie) desde el año 2011.

Detalle geográfico: Nacional, CC. AA., provincias, municipios mayores de 10.000 habitantes, distritos municipales en capitales de provincia.

Actualmente, el ámbito geográfico lo constituye el territorio fiscal común, es decir, todo el territorio nacional excepto Comunidad Foral de Navarra y País Vasco.



Coyuntura demográfica de las empresas

Objetivo: Dar respuesta a la necesidad de información de mayor frecuencia, especialmente durante la crisis generada por la pandemia COVID-19, sobre los fenómenos vinculados al dinamismo empresarial, así como su caracterización mediante variables de clasificación, especialmente las vinculadas a los expedientes de regulación temporal de empleo (ERTE), actividad económica, tamaño o características socio-demográficas para los trabajadores autónomos.

Metodología: Se aplica de forma simplificada la metodología desarrollada en la Demografía Armonizada de Empresas para la obtención de indicadores estadísticos.

La información de base utilizada para la obtención de estos indicadores es el Registro de Cuentas de Cotización y el Registro de Trabajadores Autónomos (RETA) de la Seguridad Social. Estas fuentes son complementadas con ficheros sobre trabajadores acogidos a ERTE del Servicio Público de Empleo Estatal. También se

ha utilizado información estadística derivada del Padrón Municipal y del DIRCE.

Contenido de la publicación: Se distinguen dos poblaciones: unidades legales empleadoras y trabajadores autónomos. Se obtienen cuadros comparativos trimestrales de los fenómenos demográficos estudiados (altas, bajas, reactivaciones, permanencias). Además, se ofrece estudio de cohortes para la observación de estos fenómenos a lo largo del tiempo. Los resultados están clasificados por actividad económica y en el caso de trabajadores autónomos por edad y sexo. Además, existen variables relacionadas con la situación de las unidades respecto a los ERTE.

Detalle geográfico: Nacional, CC. AA., provincia, municipio

Estimación de la ocupación en alojamientos turísticos a partir de datos de plataformas digitales

Objetivo: Complementar la información de las encuestas tradicionales de ocupación en el ámbito del grupo Alojamientos turísticos y otros alojamientos de corta estancia que es el que tiene una mayor presencia en las plataformas digitales y en la que el fenómeno de la economía colaborativa (origen de estas plataformas) tiene un mayor impacto.

Metodología: La información se obtiene a nivel europeo mediante un acuerdo de Eurostat con las principales plataformas del sector de alojamientos turísticos Airbnb, Booking, Expedia Group y Tripadvisor. El INE recibe los datos nacionales mediante un acuerdo bilateral con Eurostat. Los datos corresponden a un amplio abanico de tipologías alojativas que va desde los apartamentos turísticos, albergues turísticos, casas rurales, así como las viviendas de uso turístico. Se excluyen hoteles y campings.

Contenido de la publicación: Se ofrecen datos mensuales y anuales de alojamientos turísticos ocupados, noches ocupadas, número de pernoctaciones y estancia media, clasificados por residencia del viajero. Se dispone de información de enero de 2018.

Detalle geográfico: Nacional, CC. AA., provincia, punto turístico.

Medición del comercio diario al por menor de grandes empresas

Objetivo: Servir de complemento al Índice de Comercio al por Menor, ofreciendo información relativa a un subconjunto de empresas, de forma más agregada, pero con una periodicidad diaria y una frecuencia quincenal.

Metodología: La información de base son las declaraciones del Sistema de Información Inmediata de las empresas cuya actividad principal es el comercio al por menor, que la Agencia Estatal de Administración Tributaria envía al INE. Sobre esta información se calculan los indicadores relevantes en cada periodo.

Contenido de la publicación: Se ofrecen tasas de variación mensuales y anuales, así como porcentajes y medias móviles que permiten el seguimiento diario del comercio al por menor en estas empresas.

Detalle geográfico: Nacional.

Distribución del gasto realizado por los visitantes extranjeros en sus visitas a España

Objetivo: Proporcionar información trimestral sobre el gasto en destino real realizado por los visitantes extranjeros en sus visitas a España, con el fin de obtener una mayor desagregación que la proporcionada por la Encuesta de gasto turístico (EGATUR), tanto de los países de residencia como de las comunidades autónomas de destino.

Metodología: La información de base es la Encuesta de gasto turístico a la que se ha in-



Medición del número de viviendas turísticas en España y su capacidad

Objetivo: Estimar el número de alojamientos de vivienda turística que hay en España, así como su capacidad.

Metodología: La información se extrae mediante web scraping de las tres plataformas más utilizadas de alojamiento turístico en España. Sobre los alojamientos extraídos, primero se seleccionan las viviendas turísticas de acuerdo con la normativa en esta materia de cada comunidad autónoma, y después, se eliminan aquellas viviendas presentes en más de una plataforma mediante un algoritmo que elimina duplicados.

Contenido de la publicación: Tablas y mapas de número de viviendas turísticas, así como plazas disponibles y porcentaje de viviendas turísticas sobre el total de viviendas.

Detalle geográfico: Nacional, CC. AA., provincia, municipio, distrito, sección censal.

corporado información auxiliar de transacciones bancarias realizadas a través de tarjetas por los visitantes no residentes en España en sus viajes o excursiones en nuestro país. Estas transacciones incluyen las realizadas a través de tarjeta de manera presencial (pagos realizados a través de Terminal Punto de Venta o TPV), así como retiradas de efectivo en cajeros. Esta información permite asignar gastos al territorio en el que se realiza dicho gasto ampliando la información sobre destinos recogida en la encuesta de gasto turístico.

Contenido de la publicación: Se ofrecen datos trimestrales y anuales sobre distribución de gasto en destino por CC. AA. y por países de residencia de los turistas, así como gasto medio diario y por visitante.

Detalle geográfico: Nacional, CC. AA.

Medición del turismo a partir de teléfonos móviles

Objetivo: Estimar los movimientos de los turistas residentes y no residentes con un nivel de granularidad geográfica superior al de las operaciones FRONTUR y ETR, y con una mayor oportunidad

Metodología: Los ficheros enviados al INE mensualmente por las operadoras de telefonía móvil contienen información agregada y anonimizada sobre viajes y pernoctaciones en el extranjero realizados por turistas residentes (turismo emisor), así como los realizados por turistas y excursionistas extranjeros (turismo receptor), una vez se han definido los conceptos turísticos en términos de los eventos registrados por las compañías. Las estimaciones se realizan a partir de la información depurada de estos ficheros y la integración de los datos de las diferentes operadoras.

Contenido de la publicación: Se ofrecen datos y herramientas de visualización que permiten analizar la evolución del turismo receptor y emisor.

Detalle geográfico: Nacional, CC. AA., provincias y municipios. Además, se ofrece información de procedencia y destino por países y continentes. ●

INDICADORES URBANOS. EDICIÓN 2022

Descarga gratuita a través de la web del INE

A finales de los 90, se inicia el proyecto europeo Urban Audit, con el fin de recopilar información estadística que permita comparar la calidad de vida de las principales ciudades europeas. Desde entonces se ha venido desarrollando en diversas fases o ciclos de recogida de datos de aproximadamente 3 años cada uno.

La recopilación de la información corresponde fundamentalmente a los institutos nacionales de estadística de los estados miembros que se implican en el proyecto. En la actualidad se recoge información sobre más de 900 ciudades de la UE para las cuales contiene 171 variables y 62 indicadores. Para los ámbitos supramunicipal o submunicipal se recogen solo una parte de esos indicadores.

La publicación Indicadores Urbanos actual es una selección y adaptación del contenido del proyecto Urban Audit, actualmente denominado "data collection for sub-national statistics (mainly cities)". De la lista de indicadores recopilados para el proyecto europeo se ha elegido un conjunto limitado de 39 indicadores que cubren distintos dominios, con información desde 2010. Se ofrecen datos para las ciudades seleccionadas en el proyecto europeo, las conurbaciones o áreas donde el entorno urbano de la ciudad se expande más allá de sus límites administrativos, las zonas supramunicipales definidas por su influencia laboral (Áreas Urbanas Funcionales) y para zonas submunicipales, equivalentes a la partición administrativa en distritos. Más información en la página web Eurostat Spatial-Units.

ESPAÑA EN CIFRAS

Descarga gratuita a través de la web del INE

Papel. 64 págs. 2,66 € (IVA incluido)

España en Cifras es una publicación que se caracteriza por su carácter divulgativo y su estilo directo y sencillo para explicar y exponer los fenómenos sociales, económicos y demográficos que suceden en nuestro país. Sus breves explicaciones acompañadas de tablas, representaciones gráficas o mapas que complementan y simplifican la información, aportan una visión general sobre nuestro país y su situación en el entorno europeo.

Los datos recogidos provienen tanto del Instituto Nacional de Estadística (INE) como de otras fuentes oficiales, nacionales e internacionales. Queremos agradecer la valiosa colaboración prestada por todas las instituciones que han facilitado la información, así como a los informantes por su participación y colaboración con el INE, tan importante como necesaria puesto que es la base de nuestros resultados.

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 ABRIL A JUNIO DE 2022**INEbase. Marzo 2022**

Contabilidad Nacional Anual de España. 4º trimestre 2021.

Cuentas trimestrales no financieras de los sectores institucionales.

Contabilidad Nacional Trimestral de España. 4º trimestre 2021.

Encuesta Coyuntural sobre Stock y Existencias. 4º trimestre 2021.

Encuesta de Comercio Internacional de Servicios. 4º trimestre 2021.

Encuesta de Turismo de Residentes. 4º trimestre 2021.

Encuesta trimestral de coste laboral. 4º trimestre 2021.

EPA. Variables submuestra. 2021.

Estadística de ejecuciones hipotecarias. 4º trimestre 2021.

Estadística sobre transporte ferroviario. 4º trimestre 2021.

Índice de coste laboral armonizado (ICLA). 4º trimestre 2021.

Índice de Precios de Vivienda (IPV). 4º trimestre 2021.

Índice de Precios del Sector Servicios. 4º trimestre 2021.

Padrón de Españoles Residentes en el Extranjero (PERE). 01/01/2021

INEbase. Febrero 2022

Descarga gratuita a través de la web del INE

Contenido:

EPA. Resultados anuales. 2021.

Estadística sobre el uso de la Biotecnología. 2020.

Indicadores Urbanos. Edición 2022

Descarga gratuita a través de la web del INE.

España en cifras

Descarga gratuita a través de la web del INE.

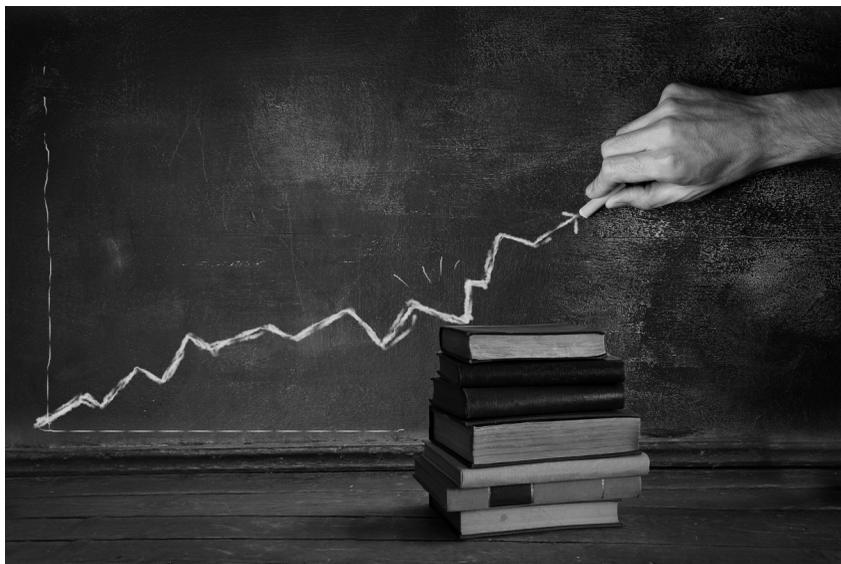
Papel. 64 págs. 2,66 € (IVA incluido).

La Península Ibérica en cifras

Descarga gratuita a través de la web del INE.

En el próximo número:

Estadísticas culturales



El pensamiento económico distingue entre bien privado, cuyo consumo agota un agente; bien común, a disposición de todos de manera decreciente según es consumido, y bien público, que permite ser consumido por varios agentes sin que disminuya la cantidad disponible. A partir de estas definiciones es tentador calificar a la Cultura como bien público de una comunidad, dado que está a disposición de todos sus miembros.

Hay que confesar que el efecto de la cultura sobre las personas que forman la sociedad supera los límites estrechos que hemos definido: en efecto, la cultura no solo sigue disponible para un miembro de la comunidad después de su disfrute por otro; se engrandece y manifiesta su mejor versión cuanto mayor es la vivencia individual de la misma. La cultura forma parte del bien común que nos explica santo Tomás de Aquino, que se erige simultáneamente y engrandece el fin de cada una de las personas que existen en la comunidad, sea esta una familia, una región o un país.

En el próximo número intentaremos definir los contornos de un fenómeno que trasciende lo económico, teniendo efectos sobre el bienestar medible; que supera lo local, definiéndose en un lugar concreto; y que se eleva de lo individual, aun brotando en tantas ocasiones de hombres y mujeres enfrentados en soledad a los retos que plantea la existencia de belleza, bien y verdad. El ministerio de Cultura y sus servicios estadísticos tendrán un papel protagónico que se verá completado por la mirada internacional y los gestores locales de facetas concretas del hecho cultural.

Oferta de empleo público 2020, 2021 y 2022

Cuerpo Superior
de Estadísticos
del Estado

223 PLAZAS

303 PLAZAS

Cuerpo de Diplomados
en Estadística
del Estado

Cuerpo General
Administrativo de la
Administración del Estado,
especialidad estadística

222 PLAZAS

INE

Instituto Nacional de Estadística

www.ine.es/ine/oposiciones