

El mejor momento para los números

Santi García Cremades

(Universidad Miguel Hernández de Elche. España)

Resumen

Estamos en un mundo de datos, desde 2020, los medios de comunicación abren informativos, periódicos y noticias con números y más números. Veremos los errores cometidos por el camino, cómo entender el error en los propios datos y cómo explicar la información de la manera más didáctica y rigurosa.

Palabras clave

Estadística, ciencia de datos, coronavirus, pandemia, números, salud pública

Abstract

We are in a world of data, since 2020, the media open news, newspapers and news with numbers and more numbers. We will see the errors made along the way, how to understand the error in the data itself and how to explain the information in the most didactic and rigorous way.

Keywords

Statistics, data science, coronavirus, pandemic, numbers, public health

1. Día D pero no el momento M

Era nuestro momento, el de las Matemáticas, con M mayúscula y el de los matemáticos, con tilde. Era nuestro día, nuestro estreno, nuestra fiesta... Pero nooo, tenía que llegar un virus, un ser o no ser, una cosa que no se sabe si está viva o no, un agente microscópico para fastidiarnos a todos. Era sábado, que ya es un día especial (aunque todos los días son especiales por la paradoja de los números especiales, véase en otro libro sin pandemia mediante), pero no era un sábado cualquiera, era la primera vez en la historia que se celebraba el Día Mundial de las Matemáticas. Sí, el Estado de Alarma en España se declaró el 14 de marzo, que en formato anglosajón se escribe 3/14 y por tanto se conoce y se celebra el Día Del Número Pi por todo el mundo. Además, fue el día en que nos dejó Stephen Hawking y nació Albert Einstein, también nacieron Sierpinski, Pedro Duque y Akira Yoshizawa (uno de los padres del origami, que además falleció también un 14 de marzo, ¿casualidad?). Desde luego, es un día relacionado con la ciencia. Los físicos lo reivindican como suyo, pero en noviembre de 2019, en la 40ª Conferencia General de la UNESCO, se proclamó que cada 14 de marzo sería el Día Internacional de las Matemáticas. Este 2020 iba a ser el primer año en celebrarse, y estábamos preparados, pero se ve que el SARS-CoV-2 es un virus con vocación física y, por chingar, se nos chafó la fiesta. Bueno, ya en serio, lo que sí ha conseguido esta pandemia no es fastidiar la fiesta de las Matemáticas, sino crear una necesidad mundial de saber más de esta ciencia exacta, y añadiremos la Estadística, que es la ciencia de la aproximación. Se podría decir que 2020 es el año donde las matemáticas y la estadística han jugado el papel más relevante de toda la historia. Y todo empezó en el Día de Pi...

Las matemáticas nos están ayudando a dar certezas en un mar de incertidumbre pandémica. Todos necesitamos saber cómo afrontar esto, muchos necesitan tomar medidas adecuadas a cada situación y algunos no saben ni qué pensar entre un océano de cifras y variables. Medir es algo tan antiguo como el contar, y



como el escribir. Que esté con el homo sapiens civilizado desde casi siempre no quiere decir que sea un problema resuelto. De hecho, “cálculo” significa en latín “piedrecita”, concretamente la típica china que se te mete en el zapato y te hace parecer Chiquito de la Calzada en mitad de la calle. Sí, calcular el mundo que nos rodea es algo incómodo, por su imperfección, pero puede ser divertido, como Chiquito. Pero “Houston (y Wuhan), tenemos un problema”: las matemáticas son teóricas, viven en el mundo de lo abstracto, y cuando son aplicadas NUNCA cuadran las cuentas al 100%. Le pasó a la escuela de Pitágoras al intentar calcular el número Pi de forma exacta, le pasó al mismísimo Newton al intentar calcular el área bajo una curva, y nos pasa ahora al intentar entender y predecir los datos de esta maldita pandemia.

En cuanto hay datos recogidos, lo primero es mirar al becario, que lo informaticice todo y lo archive. Todo bien. Pero cuando hay datos, y más datos, un día y otro día, variables que cambian, problemas de medición, en seguida miran al matemático de cabecera y buscan una solución óptima a un problema abierto.

2. Pon un matemático en tu vida

Centrémonos en la COVID-19, que es el tema que nos toca por todos los poros. Las herramientas matemáticas para estudiar una epidemia son muy variadas y diferentes en su naturaleza. Son tiempos para que todos arrimemos el hombro, cada uno según su clavícula y escápula. En la ciencia más exacta hay familias, como si fuera esto Hogwarts y tú, apreciado lector que tienes la magia en las manos, podrías ser Harry Potter, elige bien:

- La familia de Álgebra, acostumbrados a trabajar entes abstractos e intangibles, vamos, los más locos de todos. Ellos aportan su sabiduría en ecuaciones y buscan la ecuación que más se ajuste a la evolución de las curvas de datos. Aportarán conceptos como linealidad, ajuste polinómico y exponencial, o cálculo matricial, con el que Gauss encontraría una forma de estimar relaciones entre variables y ecuaciones.

- La familia del Cálculo, aunque ellos quieren llamar los del “análisis matemático”, acostumbrados a calcular todo lo calculable, menos el cálculo de riñón. Pueden interpretar cualquier variable que cambie en el tiempo, y nos darán uno de los conceptos más relevantes: la variación. Aportarán las benditas derivadas (gracias Leibniz + Newton) con las que podemos entender los cambios de tendencia, los máximos, los mínimos y, lo más importante, las variaciones entre poblaciones (lo que dará lugar al modelo más famoso en epidemiología: el modelo SIR).

- La familia de la Geometría, que tienen a Rosy de Palma como musa. Acostumbrados a valorar las figuras por su estructura y su variabilidad, trabajan más en la parte microbiológica, pero nos darán conceptos como convexidad, que usaremos para conocer los puntos clave de los aumentos y descensos.

- La familia de la Topología, donde la distancia y el tamaño no importan (evitar chistes fáciles). Estudian cada figura según el número de agujeros que tienen, por ejemplo, una pelota y un televisor no tienen agujeros, pues para un topólogo lo mismo es jugar al fútbol que ver la tele, para ellos da igual, ven la misma estructura sin agujeros, y si ven un partido en la tele, pues ya se vuelven locos. En la pandemia, tenían poco que decir, así que se quedaron en casa pegando patadas a la tele, o viendo un balón de fútbol, algo así...

- Por último, la familia de la Estadística, tan denostados los pobres. Claro, antes la carrera de Matemáticas se llama “la carrera de exactas” (si sabes esto, es que tienes ya una edad), pero al estudiar también estadística, ya no podía llamarse así. La estadística vive del error, como un partido de la oposición. Son y serán la familia más relevante de estos momentos y, por tanto, como sombrero seleccionador que soy, te voy a mandar ahora a la familia de la Estadística.

NOTA: Por supuesto, de todo esto beben el resto de áreas y ciencias. La estadística es la más importante de todas, pues se basa en el control de la incertidumbre, que es lo más presente y lo más difícil de entender cada día. Vamos, la estadística es la que abre telediarios desde marzo de 2020, es la que más sufrimos y de la que más queremos saber.

¡Bienvenida/o a la familia de la Estadística! ¿Cómo te sientes? ¿Estás sentado/a? Prepárate para conocer el mayor error de nuestras vidas (le doy un poco de épica, pero sin música de Vangelis no es lo mismo, pónstela de fondo si eso). Quiero destacar que hablar de estadística no es dar datos sueltos y ya está. dar porcentajes o hablar de números del coronavirus sin un contexto es como decir palabras sin crear una oración con sentido. Si tenemos que plantear en qué momento estamos de la pandemia hay dos preguntas fundamentales: ¿qué variable nos da la información y cómo evoluciona esa variable? La primera pregunta era importantísima desde el principio y ha sido un gran problema. Nos interesa saber la variable “número de contagios”, parece evidente, pero ¿cómo medimos los contagios cuando el virus es desconocido?

En febrero arrancó la primera ola en Asia y Europa, poco después en el resto del mundo. Pero en marzo no teníamos la capacidad de detección suficiente para saber el verdadero número de personas infectadas. Por tanto, en la primera ola, la curva más usada en la primera ola no fue la de contagios, sino la de hospitalizados y fallecidos. Pero ni siquiera en la forma de contar fallecidos por COVID-19 había consenso entre países. La detección no es completa, nunca lo es en el mundo real, y el número real de fallecidos causados por esta enfermedad sigue estando bajo estudio. Pero eso no quita que no podamos estudiar la evolución de las cifras. Ahí fue cuando los modelos matemáticos entraron en acción. Los modelos matemáticos son algo más perfectos que los hijos de los futbolistas y las influencers, así que ojito...

3. ¿Qué y cómo medir a tiempo real?

Podemos medir de todo, vamos a dar una ristra de medidas que están con nosotros desde esta pandemia:

- Número de contagios. Es complejo, al principio no hubo buena detección del virus en los primeros meses, a partir de mayo algo mejor, pero siempre con errores. Se pueden identificar a través de distintas pruebas, donde dominan el test PCR y el test de antígenos (ver capítulo XXX). Pero también está el tema del rastreo, a más rastreo, más porcentaje de asintomáticos encontramos, y la forma de rastrear es una mezcla de estadística aleatorizada y búsqueda de núcleos de contagiosidad. Un lío, pero hemos mejorado mucho con los meses.
- Número de hospitalizados. Esta variable es de las más homogéneas a la hora de medir la enfermedad, por dos motivos: desde casi el principio se establecieron unos protocolos para identificar sintomatología leve de COVID y la forma en que nos afecta la enfermedad sigue siendo prácticamente la misma. Así que será una curva de datos muy interesante a tener en cuenta para estudiar los estados de la epidemia.
- Número de UCI. Las camas UCI han sido el gran problema sanitario. Tratar un caso grave en el momento oportuno es a menudo salvar o no la vida de un paciente, y aquí ha habido momentos en los que los respiradores no aparecía, o las camas no existían, o el personal sanitario era insuficiente. Un drama que hemos vivido en la primera ola, y casi repetimos en la segunda ola europea. Por tanto, analizar esta curva en el tiempo se vuelve algo impreciso por los picos de saturación.
- Número de fallecidos. Es el dato con el que más se especula y un tema que SIEMPRE quedará abierto. Tenemos que tener claro que el número exacto jamás se sabrá. Nos estamos aproximando a través de estudios, pero la causalidad de una defunción nunca es algo trivial, y determinar cuánta gente ha fallecido por la COVID o por falta de atención, por ejemplo, es casi imposible. El INE está trabajando en ello, y realiza informes semestrales desde la entrada de la



Democracia en España, y el Mo-Mo también ha descrito de forma diaria la evolución del “exceso de mortalidad”, que es muy diferente del concepto “fallecidos reales por COVID”.

- Positividad. La positividad es el porcentaje de positivos frente al total de pruebas diagnósticas de una enfermedad. Fácil, ¿verdad? Pues la OMS consensuó que la tasa de positividad preocupante en la pandemia era del 5%, y ahí es cuando los políticos usan la estadística a su favor. Ya decía Churchill: “sólo me fío de las estadísticas que he manipulado”. En verdad, esto de la manipulación estadística no es más que un mal uso de la estadística. En la positividad se ha jugado con reducir el número de test, o aumentar un tipo de test (más del tipo clínico) en personas asintomáticas, no voy a señalar a nadie, pero por estos motivos, la positividad no es la variable más homogénea a valorar entre países, comunidades, etc.
- Tasa de letalidad / mortalidad. Un tema interesante: ¿cuántas personas contagiadas fallecen por esta maldita enfermedad? Esa es la tasa de letalidad, el número de fallecidos por una enfermedad dividido entre el número total de contagiados por esa enfermedad. Actualmente en la COVID-19, oscila entre el 1% y el 5% del total de contagiados, evidentemente, hay que distinguir entre tipos de población (otro melón para un estadístico). La tasa de mortalidad es otra cosa: el número de fallecidos por una enfermedad en el total de la población, y de eso se encarga cada semestre el Instituto Nacional de Estadística, en este caso, estamos por debajo del 0,15% de la población española (que ya es un disparate).
- Medias móviles. Si miramos los datos cada día nos volvemos locos, hay subidas y bajadas durante una semana, por lo general. Para entender realmente si algo está subiendo o bajando, usamos algo que suaviza ese vaivén: promediar cifras durante un tiempo. Uno puede hacer un examen bueno, pero la nota del curso es una agrupación de notas durante todo un año. Con la pandemia, por la forma de recoger y trabajar los datos, ha estado muy presente la media móvil semanal, y con esta variable los modelos que hablamos ya tienen un poder de precisión muy alto.
- Incidencia Acumulada. Este concepto no lleva mucho con nosotros y es de lo más natural. No podemos comparar Madrid con Murcia en datos absolutos, la población tiene que tenerse en cuenta. Pues con una simple división ya lo tendríamos, lo que viene a ser una densidad. La Incidencia Acumulada es la suma de varios días del número de contagiados (7 o 14 días), dividido entre la población estudiada (municipio, provincia, comunidad, país, etc.). Como no hay tantos contagios para que esa cifra sea estudiada con porcentajes (y menos mal), se multiplica por 100.000. Así, tenemos un dato a comparar entre poblaciones distintas. La IA (que yo pensaba que era Inteligencia Artificial, pero no) es la variable más presente desde julio.

Al menos ya sabemos qué medir, y de todas las variables (y muchas otras que no comentamos), el Instituto de Salud Global de Harvard recomienda la Incidencia Acumulada desde el 1 de julio de 2020, para tener datos suavizados (donde disminuya el error de medición en un día concreto) y en función de una población (para algo está la densidad). La primera ola fue muy empinada, pero algo hemos avanzado cuando el resto de olas son más estables, más fáciles de estudiar y, por tanto, más fáciles para tomar decisiones en función de datos fiables.

4. El negacionismo is coming

Este año 2020 no lo esperábamos ninguno. Para un matemático era un simple año autobiográfico. Sí, como suena. Fíjate bien, coge las cifras 2-0-2-0:

- El primer 2 describe el número de ceros de 2020.
- Después viene el 0, que es el número de unos de 2020.
- El otro 2 es el número de doses que tiene 2020.
- Y finalmente, el 0 es el número de treses.

Puede ser un juego divertido ir buscando números autobiográficos, aunque difícil, pues no hay muchos números así: 1210, 2020, 21200, 3211000, 42101000, 521001000, 6210001000, y así ya vemos la construcción de los siguientes números con esta propiedad.

Yo no sé si 2020 nos define, pero ya se podría ir a freír espárragos, aunque después nos huela la orina (que, por cierto, es por el ácido asparagúxico). Lo que sí que parece es que está definiendo a algunos, sobre todo, para mal. Estamos en la era en la que el humano está más conectado con el mundo de la historia. Por supuesto, existen paradojas: “cuanto más conectados en global, más desconectados en local”, vamos, que como diría mi abuela “estamos ensimismados con el movris” y a veces parecemos sacados de *The Walking Dead*, solo que, en lugar de humanos, comemos bits. Estamos todos conectados y tenemos, por primera vez en generaciones, un problema común a todos. Ya estaba el Cambio Climático alertándonos que la política mundial nada tenía que ver con los consejos de la ciencia. El virus ha destapado este telón del todo.

Lo que no esperábamos, al menos un servidor, era ver en tiempos de virus y emergencia mundial, negacionistas. Unos niegan los datos de la OMS, otros de los datos de los profesionales de la Sanidad de este país, otros los efectos de las mascarillas, las PCR, y otros, aún más creativos, del propio virus. Novak Djokovic lo tiene hasta en el propio nombre. Un Orden mundial, un poder supremo, cada uno tiene sus teorías... Un orden sí que haría falta. Al menos empezando por lo individual. Me niego a dedicar más espacio de esta columna a los negacionistas, a riesgo que esto no me convierta en uno de ellos. Decía Aristóteles que “la sabiduría es un adorno en la prosperidad y un refugio en la adversidad”. Estamos en tiempos adversos, desde luego. Es cierto que la ciencia no tiene todas las respuestas, de hecho, se ha visto como hemos tenido más preguntas que respuestas desde principios de año (autobiográfico, recuerdo), y es ahora cuando vemos más eureka que errores. Los errores son necesarios en cualquier aprendizaje, y no hay aprendizaje sin descubrimiento, ergo no fallar implicaría no descubrir nada.

Todo tiene un error: la medición de datos no es exacta, las medidas de prevención más severas tienen sus márgenes de error, las PCR ofrecen falsos positivos y negativos (como cualquier otra prueba) ... Lo importante de esto es saber cómo medimos el error. En esto los matemáticos llevamos 200 años trabajando, bueno, realmente llevamos conviviendo con la visión del error desde que somos *Sapiens*, pero no lo formalizamos hasta principios del siglo XIX, con el Príncipe de las Matemáticas: el gran Carl Friedrich Gauss. A él le debemos el estudio del error gracias a su pasión por la astronomía. Gauss estudiaba el movimiento de planetas y estrellas, y fue capaz de predecir con gran precisión la trayectoria del planeta enano Ceres usando logaritmos. Cuando comprendió que el error se acumulaba, normalmente cerca de la medición esperada, a veces fallamos por exceso y otras por defecto, pero prácticamente todo se parecía cuando se estudiaba el error. Así empezó a intuir su famosa Campana de Gauss, y así tenemos ahora un teorema que nos dice que, en promedio, todo es Normal, es decir, todo se puede distribuir según la campana de Gauss, si tenemos el número de datos suficiente.

Ahora vienen las vacunaciones, y con ellas los antivacunas, pero eso lo dejo para el capítulo XXX. Seguramente, cuando se le denominó Normal a esta curva no pensaron en los negacionistas. Y cuando pensaron en números autobiográficos, no creo que pensaran en lo que iba a pasar en 2020. Desde luego, a mí no me define, más bien me engorda. Quiero terminar con un llamamiento a la prudencia individual y social, y ya que no nos podemos citar todos ahora, acabaré con una cita de Edmund Burke:

Nadie ha cometido un error mayor que el que no hizo nada porque sólo podía hacer un poco.

Vale, no hemos celebrado el 1er Día Internacional de las Matemáticas, pero nos hemos puesto de moda, lo que parecía un oxímoron histórico, una contradicción en sí misma, dos palabras que no podrían estar jamás en la misma oración. “Matemáticas” y “moda”, qué cosas. Bueno, están los “modelos matemáticos”, y la “moda estadística” ... No sé, igual nos tenemos que poner las gafas



El mejor momento para los números

S. García Cremades

matemáticas más tiempo y seguir teniendo presente que vivimos en un océano de errores donde el único oxígeno está en seguir aprendiendo. No sé si me ha quedado un poco Paolo Coleho o Jacques Cousteau, pero me he quedado a gusto, eso sí. Prudencia, paciencia y mucha mucha ciencia...

Bibliografía

García-Cremades, S., Morales-García, J., Hernández-Sanjaime, R. et al. Improving prediction of COVID-19 evolution by fusing epidemiological and mobility data. Sci Rep 11, 15173 (2021). <https://doi.org/10.1038/s41598-021-94696-2>

Webgrafía

Informes COVID-19 (ISCIII)

<https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx>

Country & Technical Guidance - Coronavirus disease (COVID-19)

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance-publications>

How comparable is COVID-19 mortality across countries?

<https://analysis.covid19healthsystem.org/index.php/2020/06/04/how-comparable-is-covid-19-mortality-across-countries/>

Santiago García Cremades. Profesor asociado del área de Estadística e Investigación Operativa, dentro del Departamento de Estadística, Matemática e Informática de la Universidad Miguel Hernández de Elche (Alicante). Divulgador de las matemáticas con cuyo trabajo ha contribuido a la comprensión de los datos estadísticos por el que ha recibido diversos premios.