

# Análisis estadístico de la evolución del COVID-19. Caso de Estudio: Madrid

Te puedes descargar las prácticas de cada día en nuestro repositorio público:

<https://github.com/bgonzalez380/Proyecto-Covid-2021>

## Datos 16/03/2021:

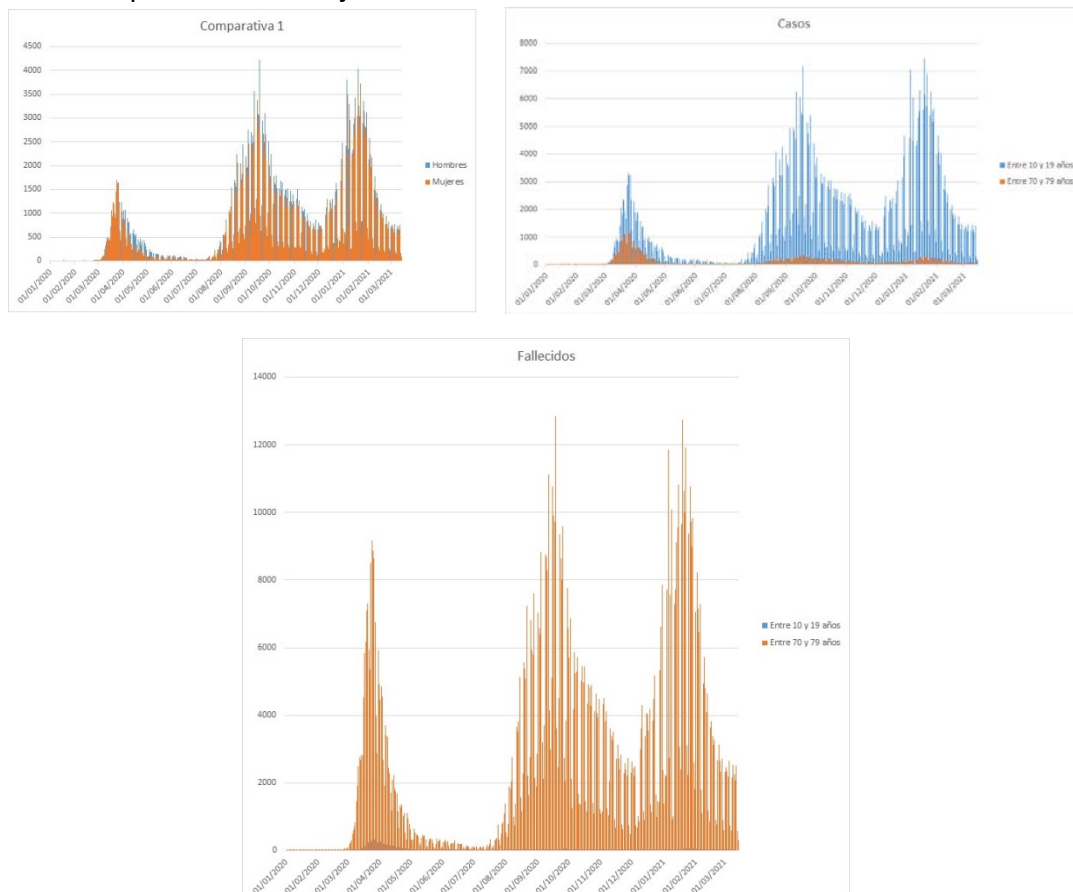
Los datos utilizados para este proyecto han sido sacados de la página web oficial del Instituto Nacional Carlos III (<https://cnecovid.isciii.es/covid19/>) Dichos datos estaban en formato “csv”. Para insertarlos en Excel, lo que hicimos fue abrir una nueva hoja de cálculo, y presionar la opción “datos” en el menú de arriba y, a continuación, “desde texto”.

Después, seleccionamos los datos descargados e hicimos un filtro (función filtro de Excel) para quedarnos solo con los de la comunidad de Madrid, representada con una M mayúscula en los datos.

Estos datos estaban separados por grupo de edad y sexo así que previamente tuvimos que sumar los valores de todas las categorías del mismo día para calcular el total de ese día.

Una vez tuvimos los datos filtrados y preparados, hicimos dos gráficos en dos distintas hojas (comparativa 1 y 2). En el primero comparamos el número de casos entre los hombres y las mujeres, y en el segundo el número de fallecidos y casos entre los contagiados con edades comprendidas entre los 10 y 19 años y los 70 y 79 años.

Como podemos observar, (colores) el número de casos es mayor entre los jóvenes, sin embargo, los fallecidos ancianos superan los de los jóvenes.



## Datos 23/03/2021:

El 23 de marzo actualizamos los datos, descargándonos los que subieron ese día. Estos contenían el número de contagiados, hospitalizados, ucis y defunciones hasta el 21 de marzo.

Nosotros hicimos el mismo filtro que hicimos con los datos anteriores, cogiendo solo los de Madrid, y los pegamos en una hoja a parte (hoja 3). A continuación, cogimos nuestra selección, y la copiamos y pegamos en una hoja nueva (hoja 3). Después de esto calculamos la frecuencia acumulada (se expresa con el siguiente símbolo:  $F_i$ ) del número de casos, ucis, defunciones y hospitalizados. Para calcularla, sumamos el número actual del valor que estemos tratando con el anterior, y así con todos los valores, para obtener el número total de ucis, casos, defunciones y hospitalizados que había en todos los días.

El último día en el que teníamos los datos disponibles, el 21 de marzo de 2021, la frecuencia acumulada para cada variable era la siguiente:

Casos: 611.587

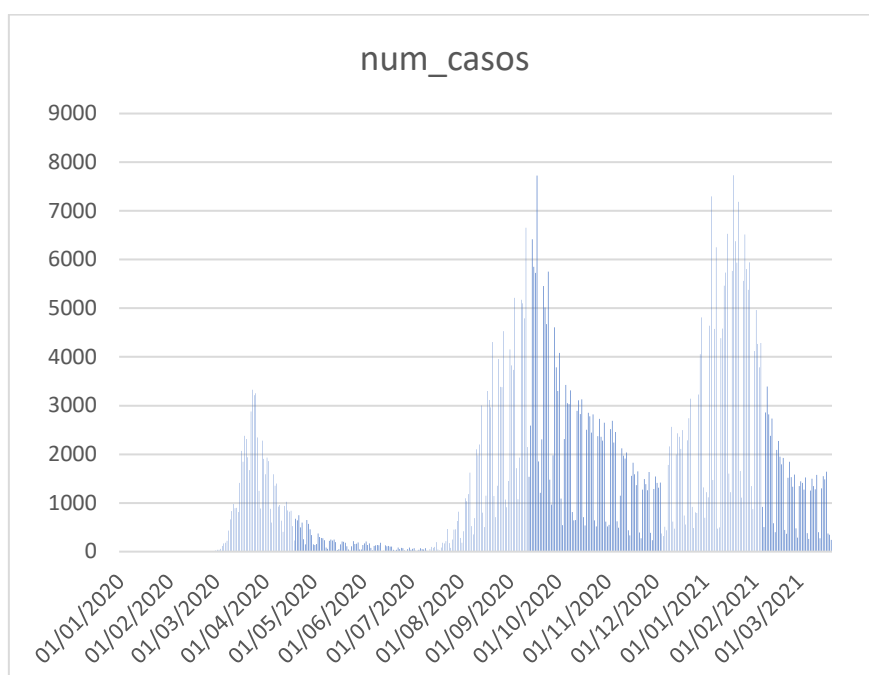
Ucis: 6.641

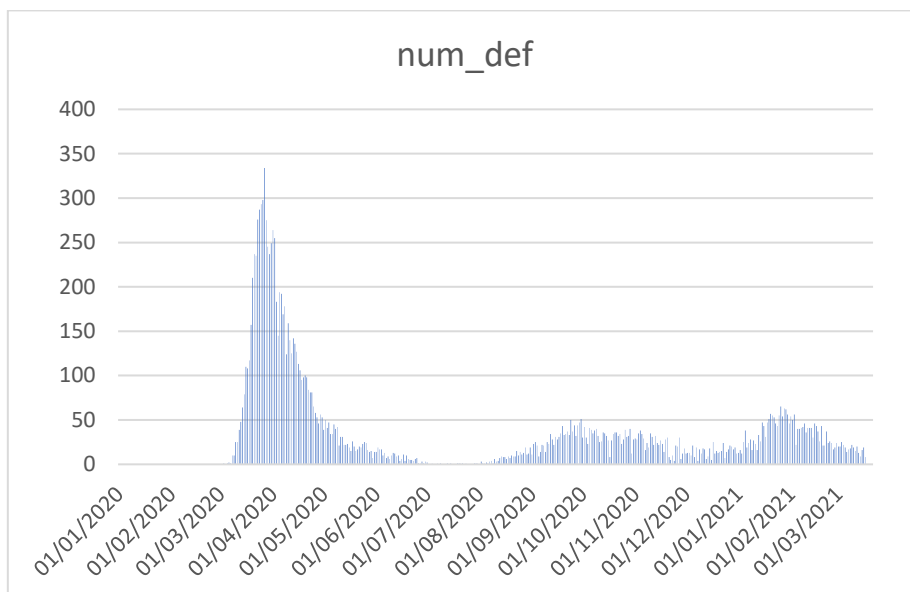
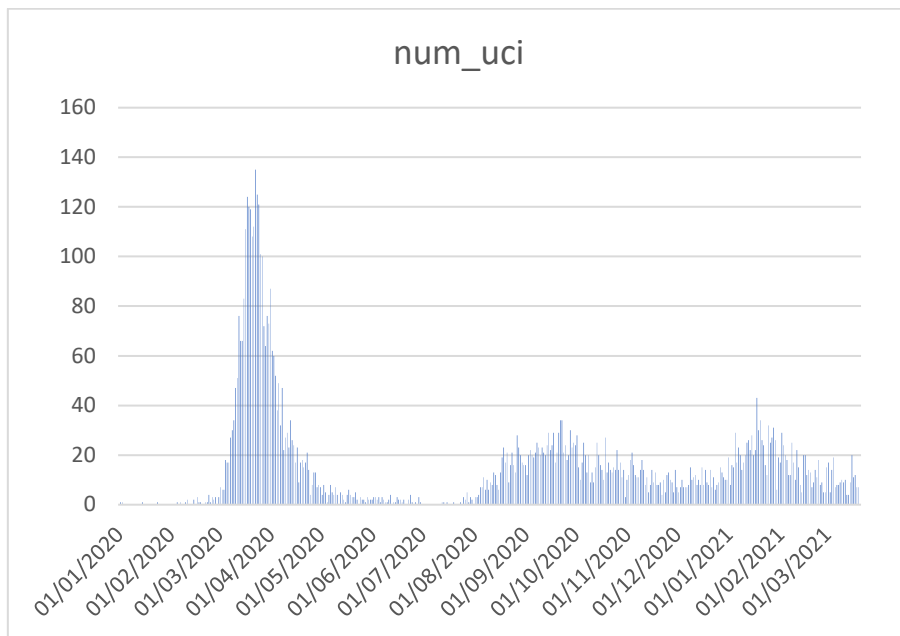
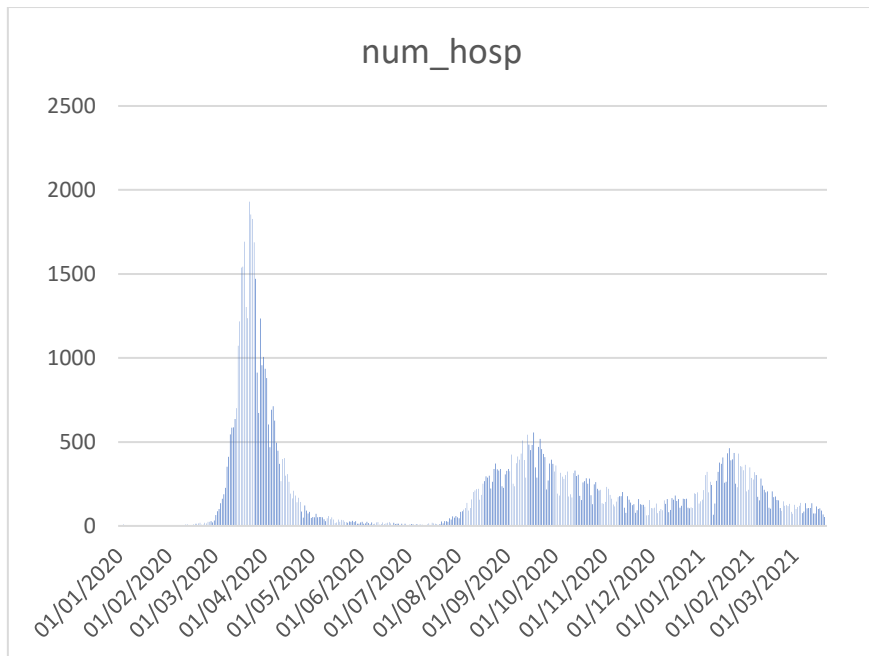
Hospitalizados: 90877

Defunciones: 14465

También hicimos una tabla con el pico de cada ola con los casos, hospitalizados, ucis y defunciones, ayudándonos de un gráfico para encontrar el valor máximo más fácilmente. Este es el resultado:

	Primera ola	Segunda Ola	Tercera Ola
casos	24/03/2020	18/09/2020	19/01/2021
hospitalizados	23/03/2020	18/09/2020	20/01/2021
ucis	23/03/2020	22/09/2020	18/01/2021
defunciones	27/03/2020	30/09/2020	26/01/2021
casos	3328	7724	7729
hospitalizados	1930	556	391
ucis	135	34	43
defunciones	334	51	65

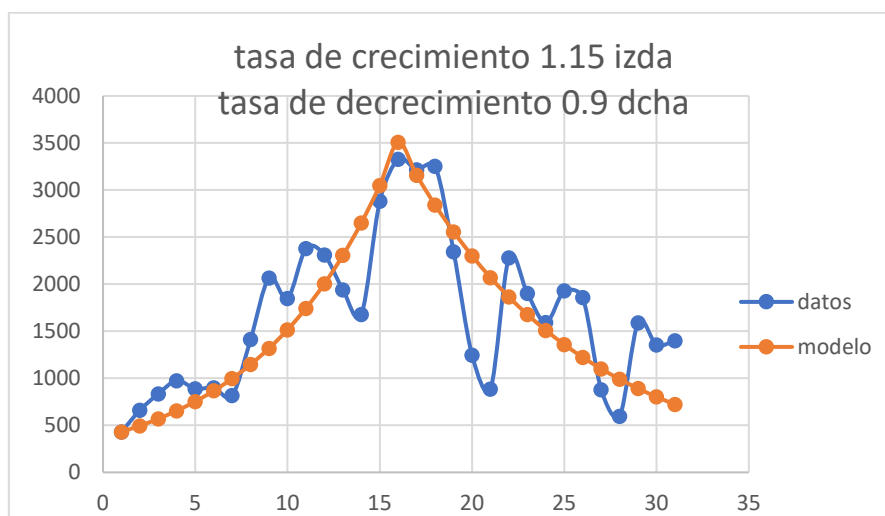




30/03/2021

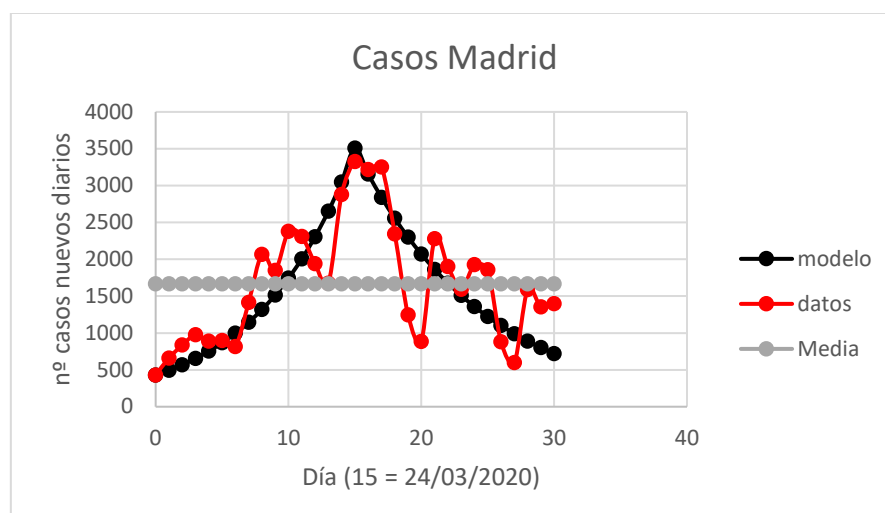
El día 30 de marzo, en un nuevo libro de Excel, pusimos los números del 0 al 30 en una columna, siendo estos los días de un mes. A continuación, adjuntamos los datos de N.º de casos de Madrid sólo de los días 09/03/2020 al 08/04/2020, ya que el 24/03/2020 fue el pico de la primera ola y esos son los días previos y posteriores.

Después, elaboramos el número reproductivo de estos días, poniendo en la primera mitad un número mayor que 1 (crecimiento de los días 0 al 15) y menor que 1 en la otra (decrecimiento en los días 16 al 30). Para ello, los datos oficiales los insertamos en un gráfico de dispersión. En una nueva columna, escribimos en la primera celda el número de casos del día 0 de los datos oficiales (431), y en la siguiente pusimos una fórmula que consiste en multiplicar el dígito de la celda anterior por el número reproductivo correspondiente a la nueva. Esta fórmula es aplicada a todas las celdas de la columna hasta el día 30. Los datos de esta columna los añadimos en una nueva serie en el gráfico de dispersión creado anteriormente, para poder contrastarlos. El objetivo principal fue hacer coincidir el pico de los datos de nuestro modelo con el de los datos oficiales, por lo que editamos el número reproductivo hasta que fue muy similar.

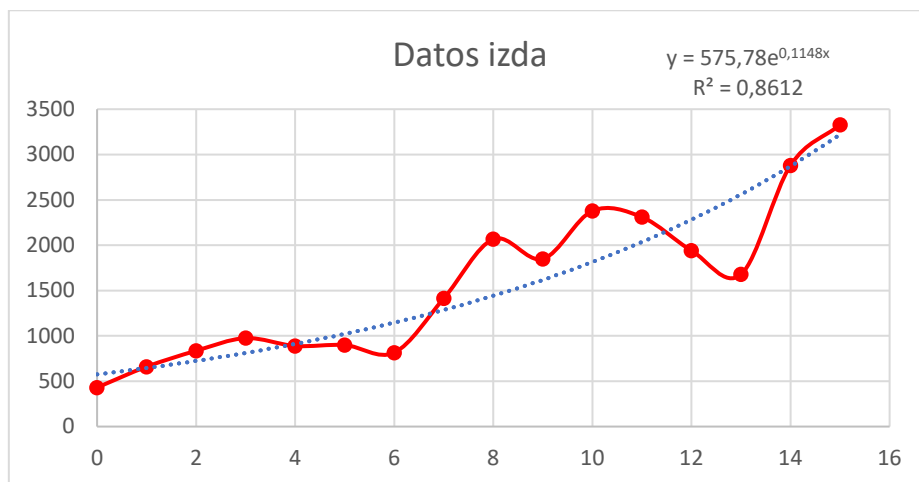


06/04/2021

Calculamos el error cuadrático medio entre el modelo y los datos oficiales, restándole al día 0 del modelo el correspondiente de los datos oficiales, y elevándolo al cuadrado, para después aplicar la fórmula al resto de los días. Calculamos también el error cuadrático medio para el peor modelo (asociado a la constante que pasa por la media de los datos oficiales) que resultó ser del orden de tres veces superior al del modelo ajustado.

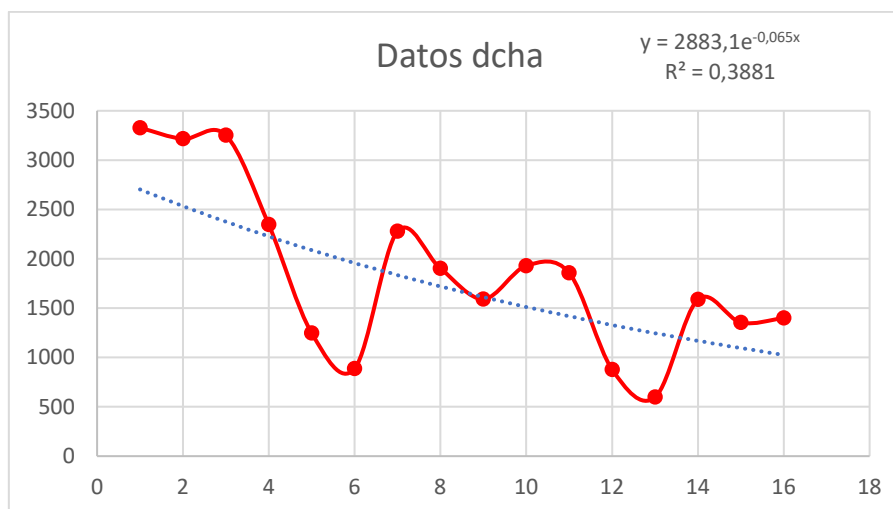


Luego, añadimos un nuevo gráfico de dispersión con los datos de los primeros 15 días de los datos oficiales, y le adjuntamos una línea exponencial desde el menú agregar línea de tendencia.



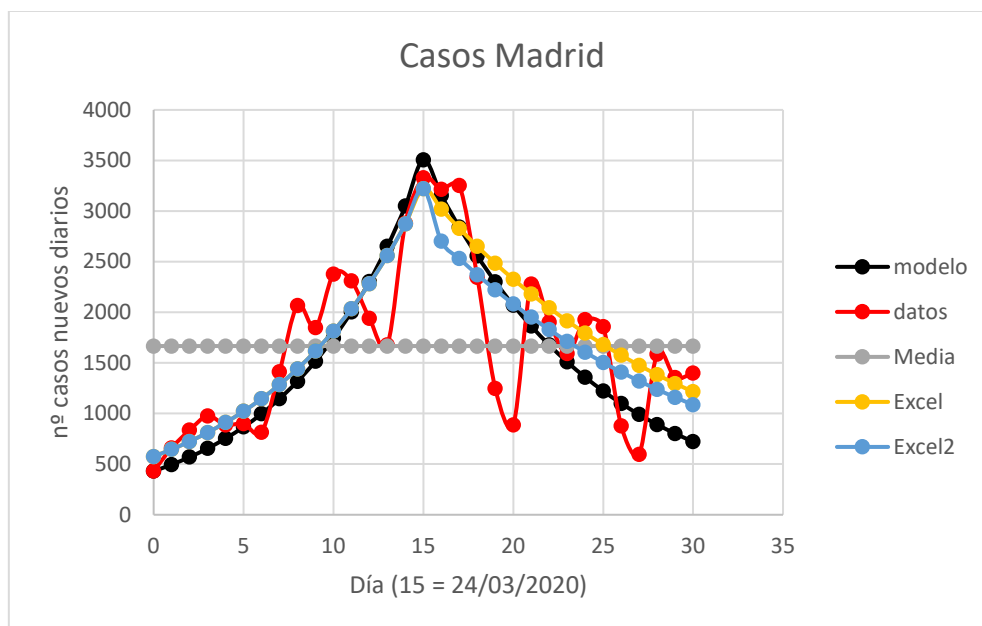
**13/04/2021**

Hicimos lo mismo después del pico.



Aprendimos entonces lo que era el número de Euler "e" y su interpretación como crecimiento y decrecimiento exponencial.

Con los ajustes automáticos del Excel empalmamos los modelos exponenciales de la izda y de la dcha de dos formas distintas: empalme con las fórmulas de las líneas exponenciales generadas automáticamente con el excel y empalme con las tasas de crecimiento y decrecimiento calculadas a partir de dichas exponenciales para activar el procedimiento manual. La idea es que a partir de estas dos líneas exponenciales, produjimos un nuevo modelo. Para ello, hicimos el producto de los datos de la misma: 575,78 por el número de Euler “e”, elevado a 0,1148 que multiplica al N.º de día del mes. Luego, aplicamos esta fórmula al resto de días hasta el pico. Sacamos, en una columna nueva, el número reproductivo de estos datos mediante una división, la segunda celda del modelo entre la primera. Lo mismo hicimos con la línea exponencial de la derecha que tiene exponente negativo. Observamos que las tasas de crecimiento y decrecimiento del proceso automático se parecían a las que nosotros habíamos encontrado manualmente.



**20/04/2021**

Aunque podíamos comparar todos los modelos con el error cuadrático medio, aprendimos a expresar con números cuán bueno es el modelo restando a 1 el cociente entre el error cuadrático medio de un modelo ajustado dividido por el peor error cuadrático medio asociado a la constante (coeficiente de determinación  $R^2$ ). Si multiplicamos la cifra resultante por 100, nos da el porcentaje de fiabilidad de nuestro modelo. El mejor modelo es el azul con un 0.67 aunque el modelo negro que habíamos ajustado manualmente no se queda lejos con un 0.6. Al modelo asociado a la Media le corresponde el valor mínimo de 0. Al modelo que pasa por todos los datos le corresponde el valor máximo de 1.

**27/04/2021**

Lo primero que hicimos fue coger el número de casos (sólo los de Madrid) y lo pegamos en Excel. Sanidad dice que un infectado empieza a contagiar 5 días después de su infección (hay margen de error, por ello hicimos los cálculos a partir de tres días consecuentes); por tanto, contamos desde el día que hay un contagiado cinco días, así que en el 5º día se muestra el N.º de contagiados a partir del contagiado del día 1 (aplicamos esto a todos los valores).

directa. En nuestro grupo, atendiendo a que una persona infectada tarda aproximadamente 5 días a ser **infecciosa**, definimos el *número reproductivo empírico*,  $p$ , como el cociente de los casos nuevos dividido por los casos nuevos de hace 5 días<sup>2</sup>. Utilizamos

## Indicadores de la situación, la tendencia y el riesgo epidemiológico

En los datos epidemiológicos se utilizan parámetros como el número reproductivo básico ( $R_0$ , valor esperado al inicio de una epidemia) o el número reproductivo específico ( $R_t$ , variable a lo largo de la epidemia). Estos números determinan el número promedio de nuevos contagios por cada caso. Si es mayor que 1, el número de casos nuevos diarios va en aumento; si es menor que 1, el número de casos nuevos diarios disminuye. Dada las dificultades de utilizar los modelos clásicos y de determinar valores característicos de la epidemia, discutidas en la Introducción, se utiliza un indicador empírico que puede mostrar una información similar. Definimos Rho ( $\rho_t$ ) como el número reproductivo empírico:

$$\rho_t = \frac{N_{new}(t-1) + N_{new}(t) + N_{new}(t+1)}{N_{new}(t-6) + N_{new}(t-5) + N_{new}(t-4)} \quad (2)$$

donde  $N_{new}$  son los casos nuevos en un cierto día. Para mi-

A partir de estos 6 datos que escogimos como muestra, calculamos el N.º reproductivo empírico del virus. Después, añadimos un gráfico de dispersión para ver la evolución de la tasa de contagios. A continuación, insertamos una línea exponencial para observar hacia la tendencia del N.º de infectados. Calculamos la ecuación en el gráfico, numeramos los días desde el primer valor del N.º reproductivo empírico hasta el pico del N.º de contagiados real; utilizamos esa numeración como la variable dependiente de la ecuación del gráfico, y con ello obtenemos el N.º reproductivo empírico de nuestro modelo.

Luego, buscamos el pico de nuestro modelo; este ocurriría cuando la unidad del N.º reproductivo pasa por el valor 1. En este caso, el pico sería entre el 21 de marzo y el 22 de marzo (día 21 y 22 del eje horizontal). El margen de error entre el pico real (24 de marzo) y el pico del modelo es muy pequeño.

