

# Modelo de regresión logístico de 5 parámetros

## Propuesta metodológica

Para modelar los datos relativos al virus COVID-19 en España y en sus Comunidades Autónomas (unidades de las que se dispone de información), se propone el empleo del modelo de regresión logístico de 5-parámetros, también conocido como modelo de Richards (Richards, 1959). Este sencillo modelo, originalmente propuesto por el grupo italiano StatGroup-19 (<https://statgroup-19.blogspot.com/>) permite modelar el crecimiento de una variable  $Y$  en relación a otra variable  $X$ . En nuestro caso la variable  $Y$  sería el número acumulado de alguna de las variables proporcionadas por el Ministerio de Sanidad, Consumo y Bienestar Social, recogidos en el siguiente repositorio de GitHub, <https://github.com/datadista/datasets/tree/master/COVID%2019>. En particular, la variable  $Y$  que consideramos puede ser:

- Número de fallecimientos.
- Número de nuevos casos.
- Número de altas.
- Número de ingresos en UCI.
- Número de enfermos hospitalizados.

Con respecto a la variable explicativa  $X$  de la que dependen las respuestas anteriores, esta mediría el momento (tiempo) en el que se registran las distintas mediciones. Es decir: el día en el que se han publicado los datos de la variable respuesta. En los últimos informes esta variable explicativa se corresponde con aproximadamente las 11:00 del día en el que se publican los datos, que son consolidados a las 21:00 del día anterior.

## Modelo de Richards

A partir de la variable respuesta  $Y$  (número acumulado) y la variable explicativa  $X$  (fecha), el modelo de regresión logístico de 5-parámetros, se puede formular del siguiente modo:

$$Y = B + \frac{T - B}{(1 + 10^{b(x_0 - X)})^s}$$

Cabe destacar que, a mayores del buen ajuste que se consigue con este modelo, los parámetros tienen una interpretación clara.

- $B$  es la asíntota inferior.
- $T$  es la asíntota superior.
- $b$  es el factor de pendiente o pendiente de Hill.
- $x_0$  es el valor de  $X$  en el punto de inflexión de la curva.
- $s$  es el coeficiente de asimetría.

Traduciendo esto términos al lenguaje común, mientras este modelo sea válido,  $T$  nos diría el **número máximo de casos** que nos esperamos encontrar en el futuro y  $x_0$  sería el valor para el cual se alcanza el conocido como “**pico**” de la curva, donde observaríamos una desaceleración en el número de casos.

## Simplificaciones del modelo y predicción

Con respecto al modelado, este modelo de regresión logístico podría ser simplificado, considerando el modelo de regresión logístico de 4-parámetros, esto es, su versión simétrica (que fijaría  $s = 1$ ). La selección del modelo más adecuado se realizará mediante medidas de bondad de ajuste ponderadas.

Una vez se decide el modelo a emplear y se estiman los parámetros mencionados anteriormente, el modelo de regresión nos permite predecir cuál va a ser el comportamiento de la variable de interés  $Y$  en los próximos días, substituyendo el valor de  $X$  por el día en el que estemos interesados. Además, las predicciones se puede acompañar de intervalos de confianza (para la media condicionada). Mediante técnicas de remuestreo se explorará la obtención de medidas de variabilidad para los estimadores de los parámetros y las predicciones.