

PEC 8 - Aproximación de funciones y regresión (II)

Fecha de entrega: 22/05/2022



Descripción del problema

Las epidemias y pandemias son objetos de estudio muy importantes ya que afectan directamente al desarrollo de las actividades habituales de una sociedad. Su tratamiento se basa en analizar las características particulares de una determinada enfermedad o infección, estudiar su propagación y proponer medidas para su control. Es evidente que los factores que influyen en la evolución de una pandemia están fuertemente relacionados entre ellos, contando a su vez con un gran componente de incertidumbre. Esto implica que su tratamiento matemático sea altamente complejo, necesitando de una amplia variedad de herramientas numéricas y de un estudio multidisciplinar.

Las agencias de salud pública, tanto nacionales como internacionales, suelen recabar datos de muy diversa índole relacionados con la evolución de las epidemias (pandemias) que ocurren en todo el planeta. Por lo tanto, una gran cantidad de información heterogénea está disponible públicamente para su tratamiento. Basada en ella se pueden desarrollar modelos estadísticos que expliquen el comportamiento de epidemias pasadas, ayuden a desarrollar políticas de control para epidemias activas y permitan predecir posibles escenarios futuros.

Disponéis de datos relacionados con la pandemia de la COVID-19, una enfermedad producida por el virus SARS-CoV-2. En particular, para esta actividad se utilizan los datos correspondientes a España, extraídos de la Organización Mundial de la Salud. Se os proporciona el fichero *WHO-COVID-19-global-data-SPAIN.csv*, con datos de evolución de la COVID-19 desde el 03/01/2020 al 03/09/2021. Para importar los datos se os proporciona un *script* de R, que facilita la correcta lectura del fichero anterior, dada una fecha de inicio, una fecha de fin y una etiqueta. Los días se numerarán de manera ascendente y consecutiva, siendo el día 1 el correspondiente a la fecha de inicio.

1. Modelización de la mortalidad

El crecimiento observado en las curvas de nuevos contagios y muertes de la COVID-19 es claramente no lineal, creciendo mucho más rápido a medida que pasan los días. Se pide:

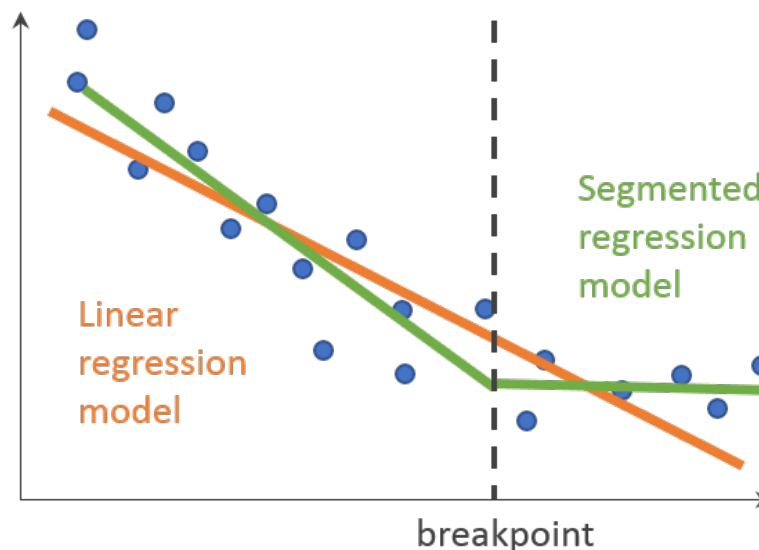
1. Utilizar alguna de las técnicas de linealización para ajustar los datos de nuevas muertes (etiqueta *New_deaths*) observados entre las fechas 15/12/2020 y 01/02/2021, correspondientes a la tercera ola de pandemia en España. Justificar la elección.
2. Realizar el mismo ajuste mediante regresión lineal ($n = 1$).
3. Representar gráficamente sobre los datos originales los ajustes obtenidos mediante los modelos

anteriores. Comentar razonadamente los resultados obtenidos.

4. Determinar el mejor modelo en base al cálculo del coeficiente de correlación r . Comentar razonadamente los resultados obtenidos.
5. Uno de los usos más habituales de los modelos estadísticos es el de predecir posibles escenarios futuros. Empleando los modelos ajustados en los apartados anteriores, predecir el número de muertes que se alcanzarían 15 días después de la última fecha utilizada en caso de dejar la evolución de la pandemia sin control, es decir, en caso de no aplicar ninguna medida que mitigue su propagación. Comentar razonadamente los resultados obtenidos.

2. Detectar los cambios de tendencia

La regresión segmentada (o por segmentos) es una técnica estadística que consiste en separar los datos disponibles atendiendo a la observación de relaciones lineales en distintos tramos de datos. Esta técnica es muy útil para detectar los puntos en los que se produce un salto brusco en la magnitud observada o un cambio de tendencia en la evolución de los datos. En la siguiente figura podemos ver un ejemplo de regresión segmentada.



Se pide:

1. Emplear la regresión segmentada sobre los datos (todos los disponibles, es decir, desde el 03/01/2020 hasta el 03/09/2021) de la curva de contagios acumulados (etiqueta *Cumulative_cases*). Seleccionar un número de segmentos y los rangos de datos que se asignan a cada uno simplemente mediante observación.
2. Representar gráficamente la regresión segmentada obtenida sobre los datos. Comentar razonadamente los resultados obtenidos.
3. Realizar el ajuste de la misma curva empleando regresión lineal básica ($n = 1$) y regresión lineal polinómica de grado 2 ($n = 2$). Representar estos dos ajustes sobre los datos y comparar los resultados obtenidos con la regresión segmentada.
4. A partir de este estudio, determinar los instantes en los que se produce un cambio de tendencia en la evolución de los contagios, es decir, los puntos de corte de las rectas obtenidas para dos segmentos consecutivos. Comentar razonadamente los resultados obtenidos.
5. Repetir los apartados anteriores empleando los datos disponibles de la curva de muertes acumuladas (etiqueta *Cumulative_deaths*).

Criterios de corrección y puntuación de los apartados

Esta PEC tendrá un valor de **10 puntos** repartidos como sigue.

- Ejercicio 1. Este ejercicio tendrá un valor de **5 puntos**:
 - Tarea 1. La justificación de la elección de la técnica de linialización se valorará con **0.5 puntos** y el ajuste mediante la misma se valorará con **0.5 puntos**. **Total 1 punto.**
 - Tarea 2. El cálculo correcto de los coeficientes de la regresión lineal se valorará con **0.5 puntos**. **Total 1 punto.**
 - Tarea 3. La representación gráfica se valorará con **0.5 puntos** (0.25 puntos para cada modelo) y el comentario razonado se valorará con **0.5 puntos**. **Total 1 punto.**
 - Tarea 4. El cálculo correcto de los coeficientes de correlación se valorará con **0.5 puntos** (0.25 puntos para cada modelo) y el comentario razonado se valorará con **0.5 puntos**. **Total 1 punto.**
 - Tarea 5. La predicción del escenario futuro mediante ambos modelos vale **0.5 puntos** (0.25 puntos para cada modelo) y el comentario razonado vale **0.5 puntos**. **Total 1 punto.**

- Ejercicio 2. Este ejercicio tendrá un valor de **5 puntos**:
 - Tarea 1. La implementación correcta de la regresión segmentada se valorará con **1 punto. Total 1 punto.**
 - Tarea 2. La representación de la regresión segmentada se valorará con **0.5 puntos** y el comentario razonado se valorará con **0.5 puntos. Total 1 punto.**
 - Tarea 3. El ajuste mediante regresión lineal se valorará con **0.25 puntos** y mediante regresión lineal polinómica se valorará con **0.25 puntos**. La representación vale **0.25 puntos** y la comparación razonada de resultados vale **0.25 puntos. Total 1 punto.**
 - Tarea 4. El cálculo correcto de los puntos de corte vale **0.75 puntos** y el comentario razonado vale **0.25 puntos. Total 1 punto.**
 - Tarea 5. La repetición de los apartados anteriores con la curva de muertes acumuladas se valorará con **1 punto. Total 1 punto.**

Referencias

- [1] Howard, J. P. (2017). Computational methods for numerical analysis with R. Nueva York: Chapman & Hall/CRC.
- [2] Leita Rodríguez, A.; Salvador Mancho, B.; Sancho Vinuesa, T. (2022). Aproximación de funciones y regresión. PID_00285421.