



**FACULTAD DE ESTUDIOS ESTADÍSTICOS**

**GRADO EN ESTADÍSTICA APLICADA**

**Curso 2024/2025**

---

**Trabajo de Fin de Grado**

***TÍTULO: Análisis del efecto del COVID-19 en la mortalidad en España.***

***Alumno: Samuel Reyes del Nacimiento***

***Tutor: José Adolfo Gálvez Moraleda***

Junio (o Septiembre) de 2025



UNIVERSIDAD COMPLUTENSE  
MADRID

## Dedicatoria

*A ti, por creer en mí incluso cuando yo dudaba. Por enseñarme a celebrar  
cada pequeño paso; sin tu luz, nada de esto sería posible.*

*A vosotros, que convertisteis lo difícil en llevadero. Por llegar cuando más  
lo necesitaba, y por estar cada día.*

*A vosotras, por el tiempo que pasamos juntos y el apoyo que me disteis en  
mi peor momento.*

*A mis profesores, pues somos las cosas que aprendemos. En especial, a  
María Villeta, por ver en mí lo que aún no podía ver. Y a María Gamboa,  
por ayudarme a encontrar el camino que hoy elijo seguir.*

*Este trabajo es el final de una etapa, pero también el principio de algo  
nuevo. Os llevo conmigo y os deseo lo mejor. Gracias por todo.*

## Resumen

El presente trabajo analiza la evolución temporal de la mortalidad mensual en España con el objetivo de modelar su comportamiento y estudiar su cambio tras la pandemia de COVID-19. Se realiza una descomposición de las series temporales para identificar patrones estacionales y de tendencia, observándose una marcada estacionalidad anual sin una tendencia clara antes de la pandemia. Para detectar cambios estructurales en la serie, se aplica el test de Bai-Perron, identificando puntos de ruptura que coinciden aproximadamente con las fases de impacto del COVID-19. Posteriormente, se evalúa la estacionariedad de la serie pre-COVID mediante los tests ADF y KPSS, concluyendo una estacionariedad débil, y se ajustan dos modelos: Holt-Winters y SARIMA(0, 0, 1) × (1, 1, 0)<sub>12</sub>. El análisis de residuos, las métricas de error y los criterios de información (AIC, BIC) indican que el modelo SARIMA ofrece un mejor ajuste. A partir de este modelo se generan predicciones para el periodo post-COVID, permitiendo estimar la mortalidad esperada y comparar con la observada. Se identifican periodos de exceso de mortalidad, especialmente al inicio de la pandemia, seguidos de una convergencia progresiva hacia los valores esperados.

## Palabras clave

Mortalidad, España, COVID-19, series temporales, SARIMA, ETS.

## **Abstract**

This paper analyzes the temporal evolution of monthly mortality in Spain with the aim of modeling its behavior and studying its changes after the COVID-19 pandemic. A decomposition of the time series is performed to identify seasonal and trend patterns, revealing a marked annual seasonality without a clear trend prior to the pandemic. To detect structural changes in the series, the Bai-Perron test is applied, identifying breakpoints that approximately coincide with the phases of COVID-19 impact. Subsequently, the stationarity of the pre-COVID series is evaluated using the ADF and KPSS tests, concluding weak stationarity, and two models are fitted: Holt-Winters and SARIMA(0, 0, 1) × (1, 1, 0)<sub>12</sub>. Residual analysis, error metrics, and information criteria (AIC, BIC) indicate that the SARIMA model provides a better fit. Based on this model, forecasts are generated for the post-COVID period, allowing the estimation of expected mortality and comparison with observed values. Periods of excess mortality are identified, especially at the beginning of the pandemic, followed by a progressive convergence toward expected values.

## **Key words**

Mortality, Spain, COVID-19, time series, SARIMA, ETS.

## Índice general

Dedicatoria	I
Resumen	II
Palabras clave	II
Abstract	III
Key words	III
Índice de figuras	VI
Índice de cuadros	VIII
Capítulo 1. Introducción	1
1.1. El COVID-19 en España	1
1.2. Objetivos del estudio	2
1.3. Estructura del estudio	2
Capítulo 2. Marco teórico	4
2.1. Series temporales	4
2.2. Suavizado de series temporales	5
2.3. Modelado de series temporales	6
2.4. Contrastes estadísticos	8
2.5. Otras pruebas estadísticas: Detección de datos atípicos	9
Capítulo 3. Preparación de los datos	10
3.1. Descripción del conjunto de datos	10
3.2. Limpieza y verificación de los datos	11
3.3. Segmentación de los datos	11

Capítulo 4. Análisis exploratorio de los datos	12
4.1. Análisis exploratorio de los fallecimientos totales	12
4.2. Análisis exploratorio de las causas de fallecimiento	17
Capítulo 5. Análisis estadístico	23
5.1. Estructura de las series temporales	23
5.2. Implementación de modelos	27
5.3. Análisis de exceso de mortalidad	30
Capítulo 6. Discusión de los resultados	32
Capítulo 7. Conclusiones	34
Bibliografía	36
Anexos	37
ANEXO 1. Preparación de los datos	37
ANEXO 2. Análisis de enero de 2016	38
ANEXO 3. Gráficos de las series temporales	38
ANEXO 4. Modelos de series temporales	40
ANEXO 5. Librerías y funciones	42

## Índice de figuras

2.2.1. Suavizado de series temporales .....	5
4.1.1. Número de fallecimientos mensuales desde 2015 hasta 2023.....	12
4.1.2. Muertes mensuales por causa de muerte en el periodo pre-COVID.....	13
4.1.3. Distribución de los fallecimientos totales en enero.....	14
4.1.4. Suavizado del número de fallecimientos mensuales con medias móviles (ventana=12) .....	15
4.1.5. Suavizado del número de fallecimientos mensuales post-COVID.....	16
4.2.1. Porcentaje de fallecimientos (>5%) por causa en el periodo pre-COVID .....	17
4.2.2. Muertes mensuales por causa de muerte (excepto las 3 mayoritarias) en el periodo pre-COVID .....	18
4.2.3. Distribución de las causas de muerte en el periodo pre-COVID .....	19
4.2.4. Muertes mensuales por causa de muerte en el periodo post-COVID .....	19
4.2.5. Muertes mensuales por causa de muerte (excepto las 3 mayoritarias y $x_2$ ) en el periodo post-COVID.....	20
4.2.6. Distribución de las causas de muerte en el periodo post-COVID .....	20
4.2.7. Fallecimientos mensuales debido a la causa $x_2$ . .....	21
4.2.8. Número de fallecimientos mensuales sin $x_2$ .....	22
5.1.1. Descomposición clásica aditiva de los fallecimientos totales. ....	23
5.1.2. Descomposición clásica mediante LOESS de los fallecimientos totales. ....	24
5.1.3. Test de Bai-Perron.....	25
5.1.4. Representación de los puntos de ruptura del test de Bai-Perron .....	26
5.2.1. Análisis de los residuos .....	29

5.3.1. Predicciones del periodo post-COVID ..... 31

## Índice de cuadros

1. Resultados del test de Bai-Perron	25
2. Resultados del test de Dickey-Fuller aumentado	27
3. Resultados del test KPSS	28
4. Medidas de evaluación del modelo	29
5. Test de Ljung	30

## Capítulo 1

### Introducción

#### 1.1. El COVID-19 en España

La pandemia de COVID-19 supuso un punto de inflexión en múltiples ámbitos, desde la salud pública y ámbito social hasta la economía global. Uno de los efectos más evidentes fue el incremento en la mortalidad registrado en numerosos países. En España, en particular, uno de los países europeos más afectados durante las fases iniciales de la crisis sanitaria, la mortalidad global en abril de 2020 superó en más de un 75 % la registrada en el mismo mes del año anterior, según datos del Instituto Nacional de Estadística (INE). Las enfermedades cardiovasculares y oncológicas, en términos relativos, se mantuvieron como principales causas de muerte, pero las enfermedades infecciosas ascendieron al tercer lugar, con un incremento relativo del 1220'40 % respecto al año 2019. Otras causas, como las enfermedades respiratorias, mostraron descensos relativos (-11 %) .

Durante los años 2020 y 2021, numerosos estudios describieron cambios en las tasas de mortalidad atribuibles tanto a infección por SARS-CoV-2 (ya fuera por sus secuelas directas o por el desarrollo de patologías como el COVID persistente) como a los efectos indirectos derivados del colapso asistencial y las medidas de confinamiento. Además, investigaciones recientes han identificado asociaciones entre la infección por SARS-CoV-2 y un aumento del riesgo de desarrollar enfermedades inicialmente no vinculadas con la misma, que podrían situar esta infección como potencial factor de riesgo para el padecimiento de algunas enfermedades [1].

Más allá del impacto inmediato directo, una cuestión de interés es si la mortalidad en España ha retomado los patrones estacionales y estructurales previos a la pandemia o si, por el contrario, se ha establecido una “nueva normalidad”, con niveles, estacionalidades y/o tendencias distintas. Aunque algunos estudios internacionales han comenzado a abordar esta

cuestión, la mayoría se centran en países anglosajones y rara vez incluyen datos posteriores a 2021, lo que limita su aplicabilidad al contexto español [2, 3].

En consecuencia, persiste una laguna de conocimiento respecto a la evolución más reciente de la mortalidad por causas infecciosas y no infecciosas en España. Por ello, este trabajo busca contribuir al análisis de esta cuestión a través de un estudio basado en técnicas de análisis de series temporales. En particular, se plantea la siguiente pregunta: ¿la mortalidad en la España post-pandémica conserva el mismo comportamiento estacional y estructural que antes de 2020, o la pandemia ha supuesto un cambio permanente?

## **1.2. Objetivos del estudio**

El objetivo principal de este estudio es analizar la evolución de los patrones de mortalidad en España mediante el uso de series temporales, con el fin de identificar cambios entre el periodo pre-COVID y el periodo post-COVID. En concreto, se pretende observar si las variaciones anuales de los ciclos de mortalidad se han mantenido constantes o se han visto alteradas como consecuencia del impacto directo o indirecto de la pandemia.

Para ello, se emplean datos oficiales del número de fallecimientos mensuales en España, obtenidos del Instituto Nacional de Estadística (INE). Este estudio puede ser útil no solo para evaluar el impacto de la pandemia de COVID-19 en la salud pública, sino también para ayudar a planificar políticas sanitarias, adaptar modelos de predicción y entender cómo se modifican los sistemas poblacionales tras eventos disruptivos.

## **1.3. Estructura del estudio**

En el capítulo 2 se presenta el marco teórico, en el que se describen las técnicas estadísticas que se utilizarán en los capítulos posteriores.

En el capítulo 3 se presenta una descripción de los datos utilizados. Se detalla su origen, sus principales características y el proceso de preparación previo al análisis, incluyendo la limpieza y segmentación por periodos.

Los capítulos 4 y 5 están dedicados al análisis estadístico de los datos. Se comienza con un análisis exploratorio, se muestran gráficos y se crean los modelos de descomposición temporal que permiten identificar cambios estructurales en la mortalidad.

Finalmente, en los últimos capítulos se exponen los resultados obtenidos, realizando una reflexión crítica acerca de sus implicaciones, concluyendo así el trabajo.

## Capítulo 2

### Marco teórico

El desarrollo de este trabajo se realizó usando métodos aplicados a datos temporales, por lo que se procede a introducir las nociones básicas necesarias sobre las series temporales y sus métodos asociados [4, 5]:

#### 2.1. Series temporales

Una serie temporal es un conjunto de datos ordenados en el tiempo. A nivel matemático, dada una variable aleatoria  $Y$ , se denomina serie temporal a una sucesión ordenada en el tiempo de valores observados de la variable aleatoria  $Y$ ,  $\{Y_1, \dots, Y_t\}$ . Los componentes principales de una serie temporal son:

- Tendencia ( $T$ ): Desplazamiento a largo plazo de la serie temporal (por ejemplo, en la serie temporal de la temperatura mensual, un aumento año tras año indica una tendencia ascendente).
- Estacionalidad ( $S$ ): Comportamiento cíclico que se repite a intervalos regulares (por ejemplo, en la temperatura mensual, el aumento en verano y la disminución en invierno cada año representan estacionalidad anual).
- Error ( $\varepsilon$ ): Variabilidad no explicada o aleatoria.

Una de las características principales de una serie temporal es la estacionariedad: Una serie temporal se dice estacionaria si sus propiedades estadísticas (media, varianza y autocorrelación) se mantienen constantes en el tiempo. Esta propiedad es clave para aplicar ciertos modelos como los ARIMA, ya que garantizan que la estructura del proceso no varía. En la práctica, muchas series no son estacionarias, por lo que se suele aplicar una diferenciación (diferencias entre valores consecutivos) para estabilizar la media.

Un modelo clásico de series temporales se construye de la siguiente forma:

1. (Modelo aditivo):  $Y_i = T_i + S_i + \varepsilon_i; i = 1, \dots, t.$
2. (Modelo multiplicativo):  $Y_i = T_i \times S_i \times \varepsilon_i; i = 1, \dots, t.$

## 2.2. Suavizado de series temporales

Los datos reales, debido al componente aleatorio, muestran patrones muchas veces difíciles de interpretar. Por ello, se trata de “suavizar” esta aleatoriedad con lo que se denomina suavizado de series temporales. Con estas técnicas se busca que el comportamiento de los datos se ajuste a curvas más sencillas que se puedan interpretar con mayor facilidad.

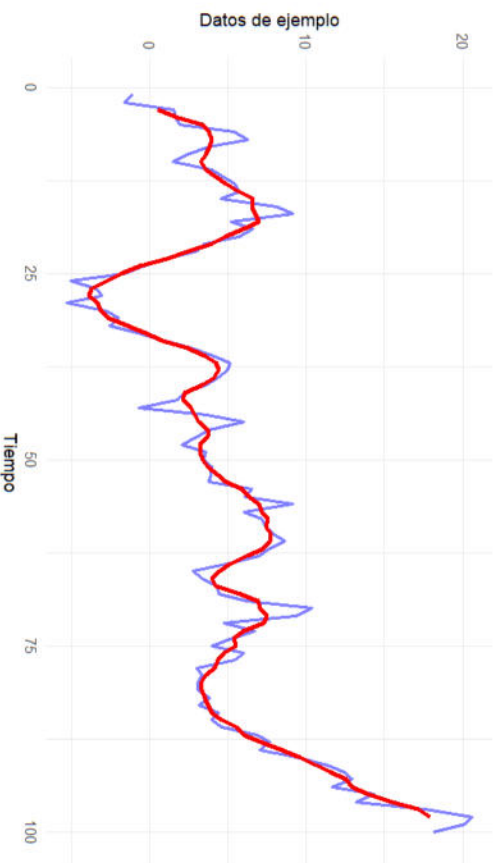


FIGURA 2.2.1. Suavizado de series temporales

### 2.2.1. Descomposición clásica de series temporales

Una primera forma de estudiar una serie temporal es, atendiendo a la descomposición clásica (Tendencia, Estacionalidad, Error), estimar cada uno de los componentes. De esta manera, se puede analizar cada uno de los componentes por separado, facilitando su interpretación.

### 2.2.2. *Seasonal-Trend decomposition using LOESS (STL)*

Similar a la descomposición clásica, pero estimando la tendencia con un método no paramétrico denominado LOESS (Locally Estimated Scatterplot Smoothing). Cada punto de la curva se estima de la siguiente forma:

1. Se selecciona una ventana de puntos cercanos a un punto  $x_0$ .
2. Se ajusta un polinomio (usualmente de grado 2 o 3) con esos puntos.
3. Se ponderan los puntos en función de su cercanía a  $x_0$  mediante pesos  $w_i(x_0)$ , que disminuyen conforme aumenta la distancia entre  $x_i$  y  $x_0$ .
4. La predicción de  $x_0$  será:  $\hat{y}(x_0) = \sum_{i=1}^n w_i(x_0) \cdot y_i$

Un parámetro clave en este método es el span, el cual indica la proporción de datos cercanos a  $x_0$  incluidos en la ventana. Valores pequeños de span muestran comportamientos más detallados de la serie temporal, mientras que valores más grandes muestran un comportamiento general. Es un método de descomposición especialmente útil para capturar tendencias no lineales, además de ser robusto en presencia de outliers si se ha escogido un span correcto.

### 2.2.3. **Medias móviles**

Una media móvil es una técnica de suavizado que consiste en calcular la media de un subconjunto de datos en ventanas temporales sucesivas. Es decir, cada punto de la curva ( $x_l$ ) se estima de la siguiente forma:

1. Se selecciona un tamaño de ventana  $k$  (usualmente impar).
2. Se realiza la media de un subconjunto de puntos de tamaño  $k$  centrado en el punto:

$$\hat{x}_l = \frac{1}{k} \sum_{j=l-m}^{l+m} x_j; \text{ donde } m = \frac{k-1}{2}.$$

## 2.3. **Modelado de series temporales**

El modelado de series temporales permite representar y predecir el comportamiento de una variable a lo largo del tiempo a partir de los valores conocidos. Para ello, se utilizan modelos que capturan componentes como la tendencia, la estacionalidad y la dependencia temporal. En este trabajo se emplean los siguientes modelos clásicos:

### 2.3.1. Modelos de series temporales

- Modelos ETS (Error, Trend, Seasonality): Este enfoque modela los componentes estructurales de la descomposición clásica de series temporales. Su estimación se basa en métodos de suavizado exponencial.
- Alisado triple de Holt-Winters: Variante de los modelos ETS que incluye una tendencia y una estacionalidad de forma explícita. Estos modelos son especialmente eficaces cuando se trabaja con series temporales con patrones estacionales regulares.
- Modelos ARIMA (AutoRegressive Integrated Moving Average): Estos modelos combinan tres componentes: una parte autorregresiva (AR) que utiliza los valores pasados de la serie, una parte integrada (I) que diferencia los datos para hacerlos estacionarios, y una parte de media móvil (MA) que incorpora errores pasados. Se denotan como  $ARIMA(p, d, q)$ , donde:
  - $p$  es el número de términos autorregresivos.
  - $d$  es el orden de diferenciación necesario para que la serie alcance la estacionariedad.
  - $q$  es el número de términos de la media móvil.
- Modelos SARIMA: Extienden los modelos ARIMA incorporando componentes estacionales. Se representan como  $SARIMA(p, d, q) \times (P, D, Q)_s$ , donde  $(P, D, Q)$  son los términos estacionales autorregresivo, de diferenciación y de media móvil, y  $s$  es la periodicidad (por ejemplo,  $s = 12$  para datos mensuales con periodo anual).

### 2.3.2. Validación de modelos de series temporales

Una vez ajustado un modelo, es fundamental evaluar su validez. Para ello se analizan los residuos del modelo (la diferencia entre los valores observados y los ajustados). Un buen modelo debe generar residuos sin estructura aparente, es decir, que se comporten como ruido blanco. Para comprobarlo se utilizan las siguientes herramientas:

- **Análisis gráfico de residuos:** Se representa la serie de residuos y sus autocorrelaciones (ACF y PACF). Cualquier patrón visible indicaría que el modelo no ha capturado toda la estructura temporal.
- **Test de Ljung-Box:** Contrasta la hipótesis nula de que los residuos no presentan autocorrelación. Un p-valor alto indica que el modelo ha absorbido la dependencia temporal de la serie.
- **Normalidad de residuos:** Aunque no siempre es un requisito, se analiza si los residuos se distribuyen aproximadamente de forma normal. Esto se puede comprobar mediante histogramas o tests como el de Shapiro-Wilk.
- **Medidas de precisión:** Se evalúa el ajuste mediante métricas como el error cuadrático medio (RMSE), el error absoluto medio (MAE) o el MAPE (error porcentual absoluto medio). Estas medidas son especialmente útiles en tareas de predicción.

#### **2.4. Contrastes estadísticos**

Los contrastes estadísticos permiten verificar supuestos sobre la serie temporal o sobre los modelos ajustados. En este trabajo se consideran los siguientes:

- **Test de Ljung-Box:** Contraste de independencia aplicado a los residuos del modelo. Evalúa si existen correlaciones significativas en los retardos, lo cual indicaría que el modelo no ha capturado toda la estructura temporal.
- **Test de Chow:** Permite detectar posibles cambios estructurales en la serie, es decir, puntos de ruptura donde las propiedades estadísticas cambian de forma significativa. Es útil para comprobar si un modelo ajustado a un periodo sigue siendo válido en otro. En este trabajo se implementará una generalización (test de Bai-Perron) que es un test que busca varios puntos de quiebre en una serie.
- **Test de Dickey-Fuller aumentado (ADF):** Test para evaluar la estacionariedad de una serie temporal, o desde otro punto de vista, permitir evaluar si es necesario diferenciar o no para que la serie temporal sea estacionaria.

## 2.5. Otras pruebas estadísticas: Detección de datos atípicos

En el contexto de este trabajo se van a tener en cuenta las siguientes técnicas:

- Test de Dixon: Método estadístico diseñado para detectar valores atípicos en muestras pequeñas ( $n \leq 30$ ), asumiendo normalidad en los datos. Se emplea comparando la diferencia entre un valor y sus vecinos más cercanos con el rango de todos los valores. Detecta los valores atípicos de uno en uno, por lo que es especialmente útil si se sospecha de un único valor atípico. El contraste de hipótesis asociado al test es el siguiente:

$$\begin{cases} H_0 : \text{No existen datos atípicos.} \\ H_1 : \text{Existe un dato atípico.} \end{cases}$$

- Z-scores: Medida estadística que indica cuántas desviaciones típicas se aleja un dato de la media del conjunto. Se calcula como la diferencia entre el valor y la media, dividida por la desviación típica. Valores superiores a 2 en valor absoluto se consideran potencialmente atípicos. Este método asume una distribución aproximadamente normal de los datos, por lo que su aplicación es más adecuada en muestras grandes o cuando se verifica dicha normalidad. En este trabajo, el uso de los z-scores es complementario para apoyar el análisis exploratorio de datos atípicos.
- Regla de Tukey: La regla de Tukey consiste en considerar un valor  $x$  atípico si se encuentra fuera del intervalo  $(Q_1 - 1,5 \times IQR, Q_3 + 1,5 \times IQR)$ , donde  $Q_1$  y  $Q_3$  son el primer y tercer cuartil de los datos e  $IQR$  el rango intercuartílico. Los gráficos de cajas y bigotes son la representación gráfica de esta regla, por lo que serán una herramienta más para la evaluación de datos atípicos. Además, no asume normalidad y es útil como criterio exploratorio robusto.

## Preparación de los datos

### 3.1. Descripción del conjunto de datos

El conjunto de datos “Defunciones por causas (capítulos) por sexo”, obtenido del INE, contiene el total de muertes mensuales en España clasificadas según la causa de defunción, permitiendo diferenciar por la variable sexo. Estos datos corresponden al periodo desde enero de 2015 hasta diciembre de 2023 (Ambos incluidos). Las causas se encuentran codificadas de la siguiente forma:

$x_1$ = Todas las causas.	$x_{10}$ = Enfermedades del sistema digestivo.
$x_2$ = Enfermedades infecciosas y parasitarias.	$x_{11}$ = Enfermedades de la piel y del tejido subcutáneo.
$x_3$ = Tumores.	$x_{12}$ = Enfermedades del sistema osteomuscular y del tejido conjuntivo.
$x_4$ = Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan al mecanismo de la inmunidad.	$x_{13}$ = Enfermedades del sistema genitourinario.
$x_5$ = Enfermedades endocrinas, nutricionales y metabólicas.	$x_{14}$ = Embarazo, parto y puerperio.
$x_6$ = Trastornos mentales y del comportamiento.	$x_{15}$ = Afecciones originadas en el periodo perinatal.
$x_7$ = Enfermedades del sistema nervioso y de los órganos de los sentidos.	$x_{16}$ = Malformaciones congénitas, deformidades y anomalías cromosómicas.
$x_8$ = Enfermedades del sistema circulatorio.	$x_{17}$ = Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte.
$x_9$ = Enfermedades del sistema respiratorio.	$x_{18}$ = Causas externas de mortalidad.

### **3.2. Limpieza y verificación de los datos**

El conjunto de datos no contenía datos faltantes. Se comprobó que la suma de los fallecimientos por causa coincidía con el número total de fallecimientos por todas las causas, verificándose que no existían incoherencias en los datos (ver ANEXO 1. Preparación de los datos).

### **3.3. Segmentación de los datos**

Dado que el objetivo del análisis se centra en comparar la mortalidad pre-COVID y post-COVID, se han diferenciado los siguientes periodos temporales:

1. Periodo Pre-COVID: Desde enero de 2015 hasta enero de 2020.
2. Periodo COVID: Desde febrero de 2020 hasta diciembre de 2020.
3. Periodo Post-COVID: Desde enero de 2021 hasta diciembre de 2023.

Esta segmentación del tiempo se fundamenta en la siguiente información:

- El primer caso de COVID en España se registró el 31 de Enero de 2020.
- El estado de alarma en España finalizó durante el mes de Junio de 2020.

El periodo COVID se amplió más allá del final del estado de alarma para recoger la influencia inmediata de la pandemia, dado que se pretenden observar los efectos a medio y largo plazo. Por este mismo motivo, los datos del periodo COVID no serán tratados en este estudio.

## Análisis exploratorio de los datos

El primer paso en estudios de estas características es realizar una exploración de los datos previa al análisis estadístico, para comprender con mayor profundidad la estructura de los datos y conocer sus características principales. En este caso, dado que se está trabajando con series temporales, la visualización de los datos es fundamental para observar el comportamiento de la serie, y detectar características clave como su tendencia y estacionalidad.

### 4.1. Análisis exploratorio de los fallecimientos totales

Se comenzó analizando las series de tiempo de las muertes mensuales por todas las causas, para tener una comprensión global de cómo evoluciona la mortalidad en España:

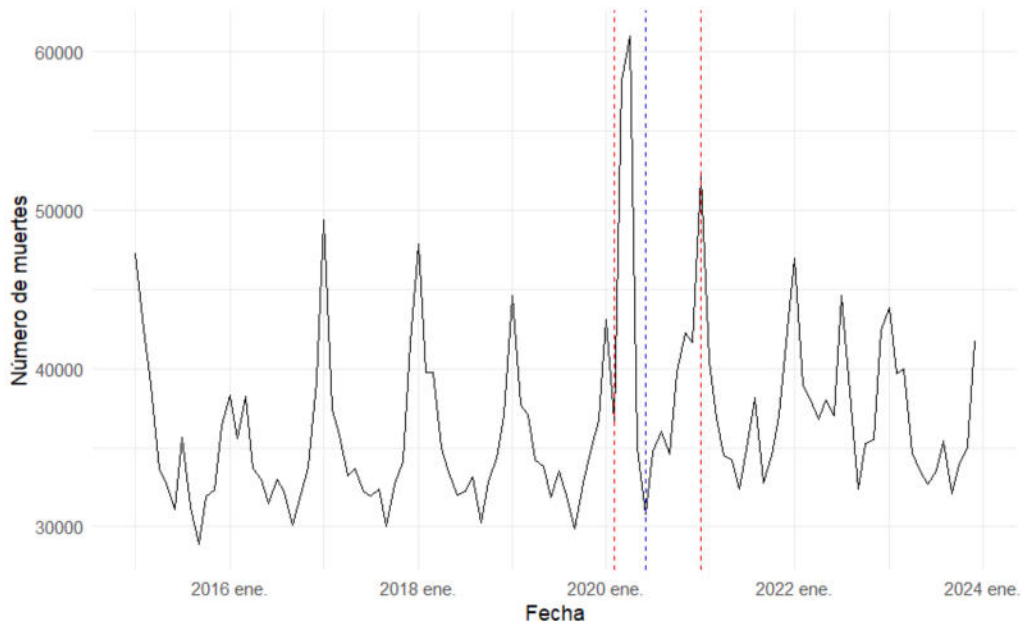


FIGURA 4.1.1. Número de fallecimientos mensuales desde 2015 hasta 2023

En la figura 4.1.1 se muestra el número de muertes mensuales en España desde 2015 hasta 2023. El interior de las líneas rojas corresponde al periodo denominado “Periodo COVID”, y la línea azul marca el final del estado de alarma en España.

Durante el periodo pre-COVID se observa un comportamiento estacional de carácter anual, con una anomalía destacable en enero de 2016. Además, no se aprecia una tendencia general clara. En contraste, el periodo post-COVID muestra un comportamiento más errático y sugiere una posible estacionalidad semestral.

Antes de continuar con la exploración conviene detenerse a estudiar qué sucede en 2016, con el fin de determinar si se trata de un dato atípico o si es el comportamiento normal de la serie. Lo primero será observar la evolución de los fallecimientos por causas, con el fin de comprobar si el defecto de mortalidad es generalizado o tiene su origen en alguna causa concreta.

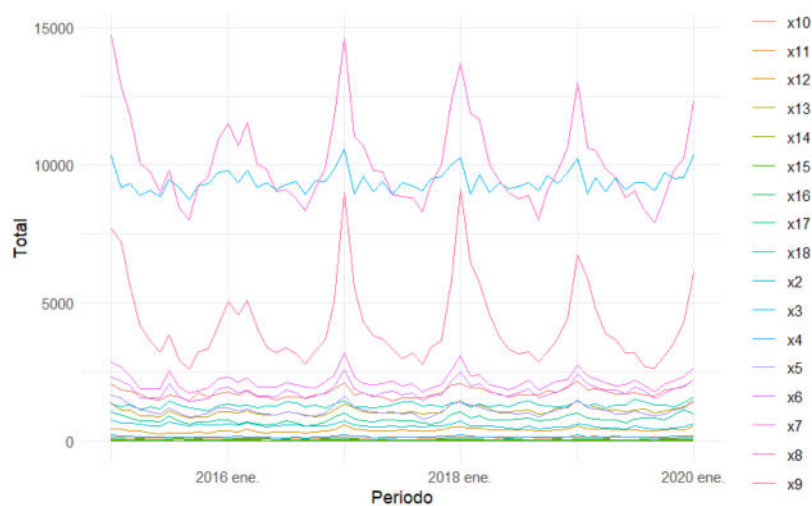


FIGURA 4.1.2. Muertes mensuales por causa de muerte en el periodo pre-COVID

La figura 4.1.2 parece indicar que el descenso en los fallecimientos es generalizado, siendo especialmente notorio en las causas que muestran mayor componente estacional. Para analizar si enero de 2016 respecto al resto de meses de enero es un outlier, se realizó un gráfico de cajas y bigotes para mostrar la distribución de los fallecimientos el mes de enero (por el

total de causas, dado que ninguna destaca sobre la otra), se calculó su z-score y se realizó el test de Dixon:

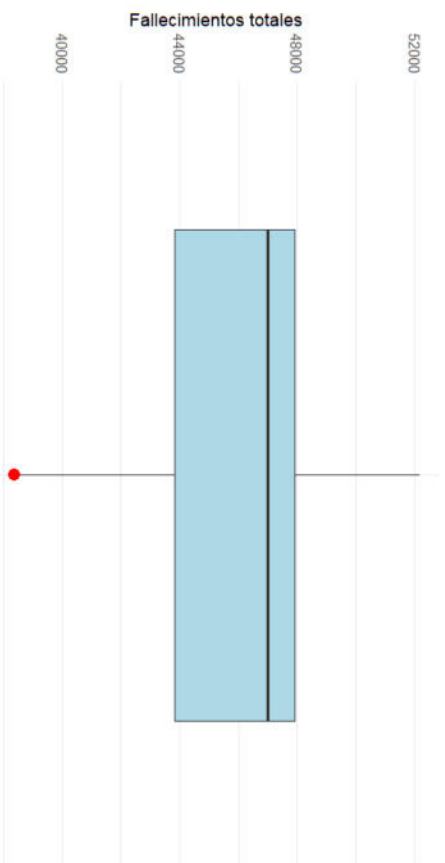
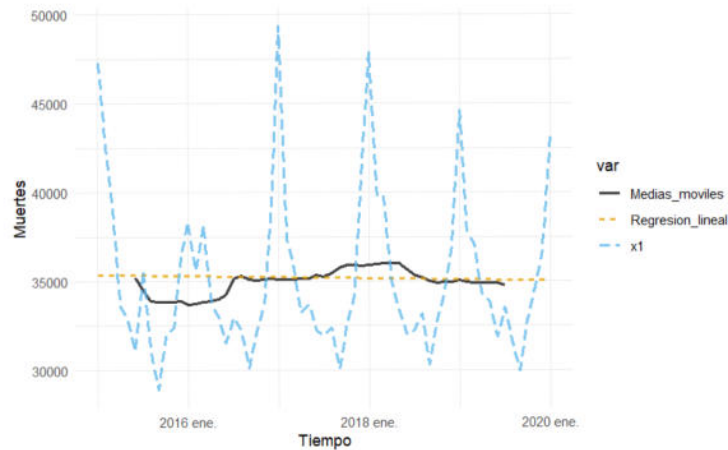


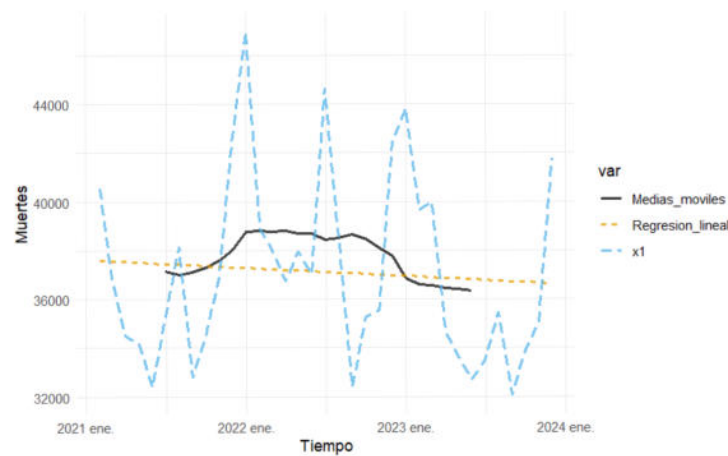
FIGURA 4.1.3. Distribución de los fallecimientos totales en enero.

En el gráfico de cajas y bigotes (Figura 4.1.3), donde la observación “2016” se representa con un punto rojo, se muestra claramente que este mes toma un valor muy alejado de lo esperado, aunque se encuentra dentro del rango inferior, lo que indica que es un valor que hay que analizar con detenimiento. Por otro, se obtuvo un z-score de -1’8963, y un p-valor en el test de Dixon de 0’2178, ambos indicando que no hay evidencia estadísticamente significativa para clasificar enero de 2016 como un dato atípico. Además, no existen registros de que 2016 fuera un año con alguna característica especial como para verse reducida de forma abrupta la mortalidad en el país. Teniendo todo esto en cuenta, no se considera enero de 2016 un dato atípico, y no será excluido del análisis.

Se procede a realizar un suavizado de las series temporales con medias móviles, tratando de eliminar el efecto estacional y analizar con mayor claridad la tendencia. Se tomará una ventana de 12 meses, dado que es la estacionalidad que muestran los datos:



(A) Periodo pre-COVID



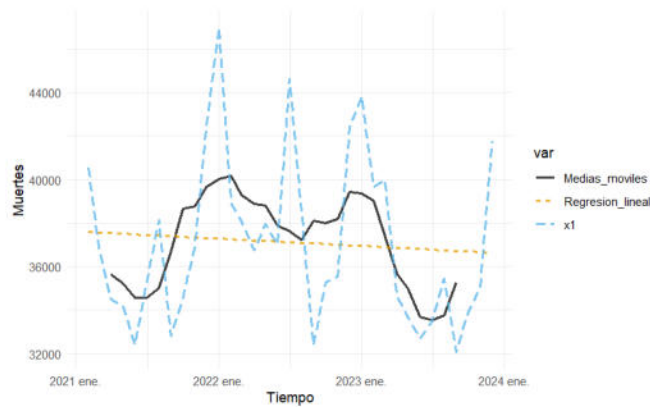
(B) Periodo post-COVID

FIGURA 4.1.4. Suavizado del número de fallecimientos mensuales con medias móviles (ventana=12)

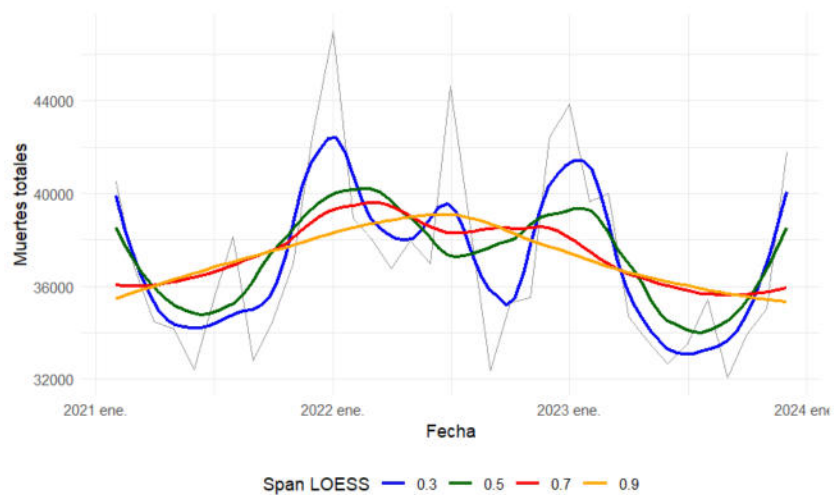
En el periodo pre-COVID se puede apreciar una aparente falta de tendencia lineal en los datos. Las perturbaciones observadas pueden ser las fluctuaciones del azar, además de la bajada inicial, que se corresponde al periodo de comportamiento inusual observado el enero de 2016.

En segundo lugar, el periodo post-COVID muestra lo que sugiere una posible tendencia lineal descendente, aunque no concluyente. Sin embargo, la forma de meseta que toma la curva suavizada podría estar indicando algún tipo de tendencia no lineal. Se procede a

comprobar esto con otros métodos de suavizado (LOESS), y se repite el mismo gráfico pero ahora con una ventana de 6 meses, para analizar el comportamiento estacional de este tipo:



(A) Suavizado con medias móviles (ventana=6)



(B) Suavizado con método LOESS

FIGURA 4.1.5. Suavizado del número de fallecimientos mensuales post-COVID

En el primer gráfico (Figura 4.1.5 (A)) puede observarse como la curva suavizada muestra únicamente estacionalidad anual, por lo que si que había cierta estacionalidad cada 6 meses. Sin embargo, que entre los años 2022 y 2023 no se produzca una bajada más pronunciada, dado que la estacionalidad cada 6 meses debería estar corregida, indica que en este periodo hubo un aumento inusual en los fallecimientos.

El suavizado con el método LOESS (Figura 4.1.5 (B)) muestra, a medida que se aumenta el span, una tendencia ascendente hasta mediados de 2022 y una tendencia descendente tras esto, apoyando la idea de que en este periodo hubo un exceso de mortalidad. Todo esto plantea la idea de que no hay una estacionalidad cada 6 meses, sino que en julio de 2022 hubo un pico en los fallecimientos de forma puntual.

En cualquier caso, es evidente que el comportamiento en ambos periodos es diferente, lo que sugiere explorar posibles motivos profundizando en las causas de muerte.

#### 4.2. Análisis exploratorio de las causas de fallecimiento

Se comenzó analizando el periodo pre-COVID. Volviendo a observar la figura 4.1.2, lo primero que llama la atención es que solo tres de las 17 posibles causas de muerte suponen la mayoría de las muertes mensuales. Las dos primeras son  $x_8$  y  $x_3$ , seguidas por  $x_9$  (las cuales son “Enfermedades del sistema circulatorio”, “Tumores” y “Enfermedades del sistema respiratorio”). La media de fallecimientos mensuales durante el periodo pre-COVID fue 35200’31 personas al mes, de las cuales un 67’42 % fallecieron por las 3 causas mencionadas. Se procedió a eliminarlas del gráfico para observar el resto de causas con mayor facilidad:

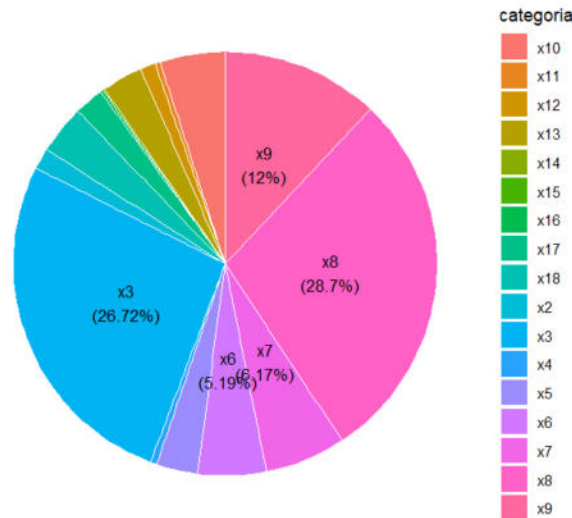


FIGURA 4.2.1. Porcentaje de fallecimientos (>5%) por causa en el periodo pre-COVID

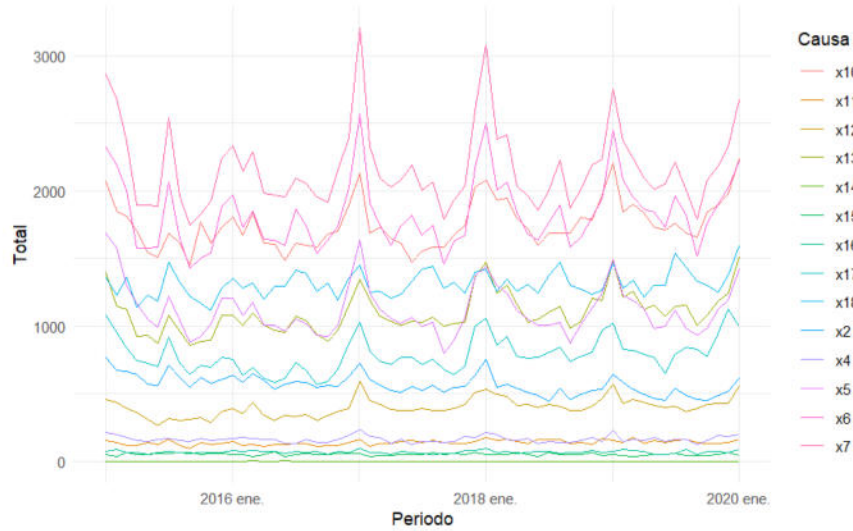


FIGURA 4.2.2. Muertes mensuales por causa de muerte (excepto las 3 mayoritarias) en el periodo pre-COVID

La visualización conjunta de ambos gráficos diferencia dos tipos de causas, las que tienen estacionalidad anual ( $x_6, x_7, x_8, x_9, \dots$ ) y las que se mantienen a lo largo del tiempo sin grandes variaciones ( $x_2, x_3, x_{15}, \dots$ ). Además, se cumple que las causas que producen más fallecimientos en general son las que presentan estacionalidad anual, haciendo evidente que su suma también tenga estacionalidad anual. Por otro lado, no parece que alguna de las causas de fallecimiento muestre picos pronunciados en mitad de periodo, sugiriendo la idea de que el despunte que se apreciaba en los fallecimientos post-COVID pueda ser consecuencia de lo ocurrido en el periodo COVID. Por último, se muestra un resumen de los estadísticos descriptivos de cada causa de muerte:

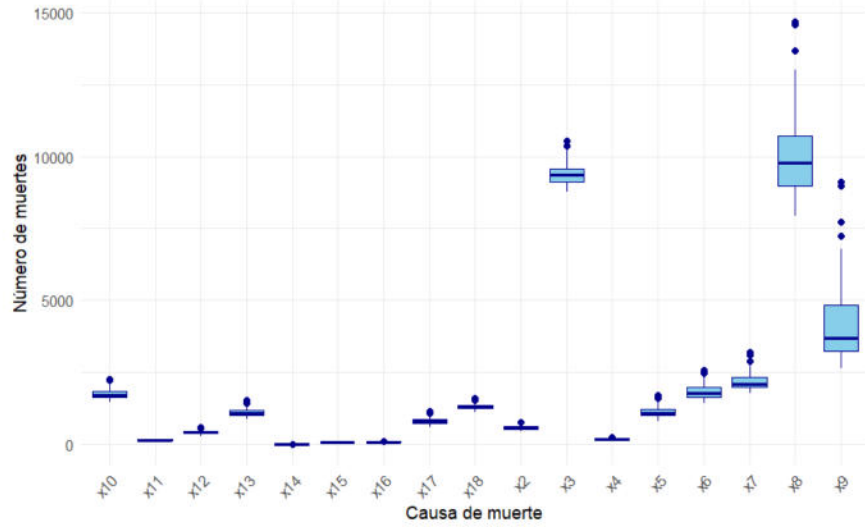


FIGURA 4.2.3. Distribución de las causas de muerte en el periodo pre-COVID

Con todo esto, se procede a recopilar la misma información del periodo post-COVID para comparar los resultados:

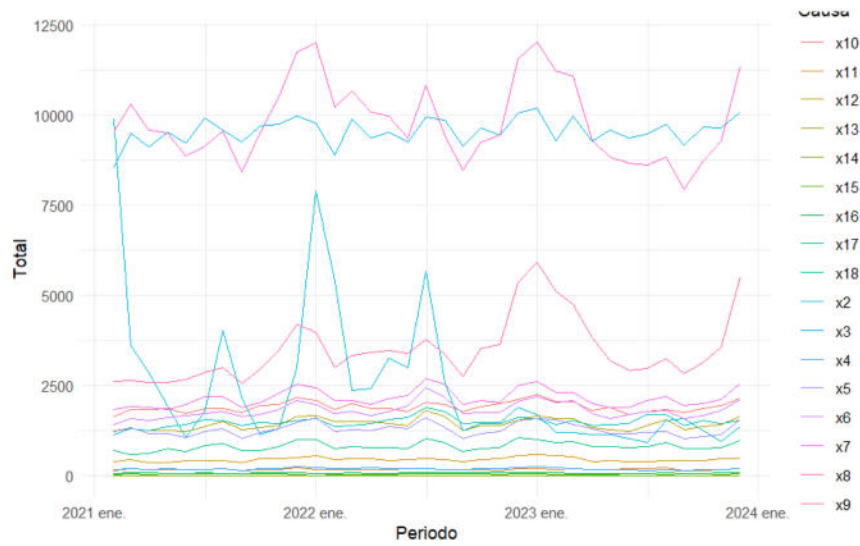


FIGURA 4.2.4. Muertes mensuales por causa de muerte en el periodo post-COVID

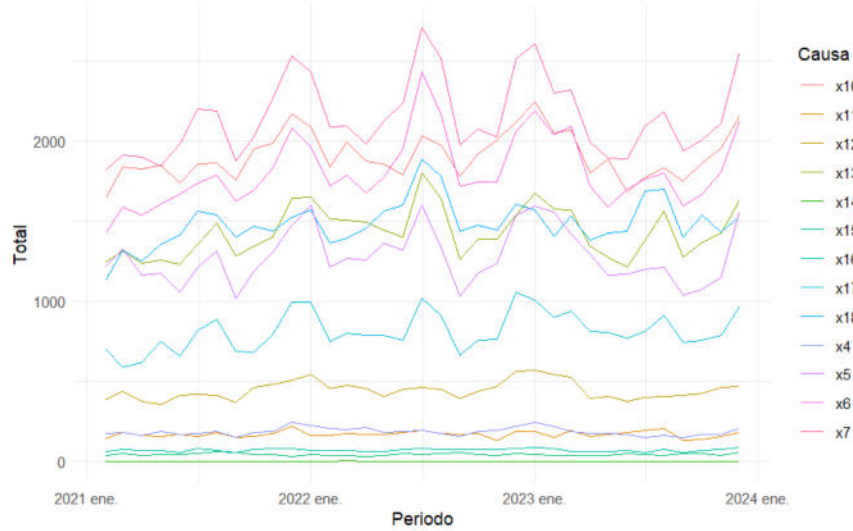


FIGURA 4.2.5. Muertes mensuales por causa de muerte (excepto las 3 mayoritarias y  $x_2$ ) en el periodo post-COVID

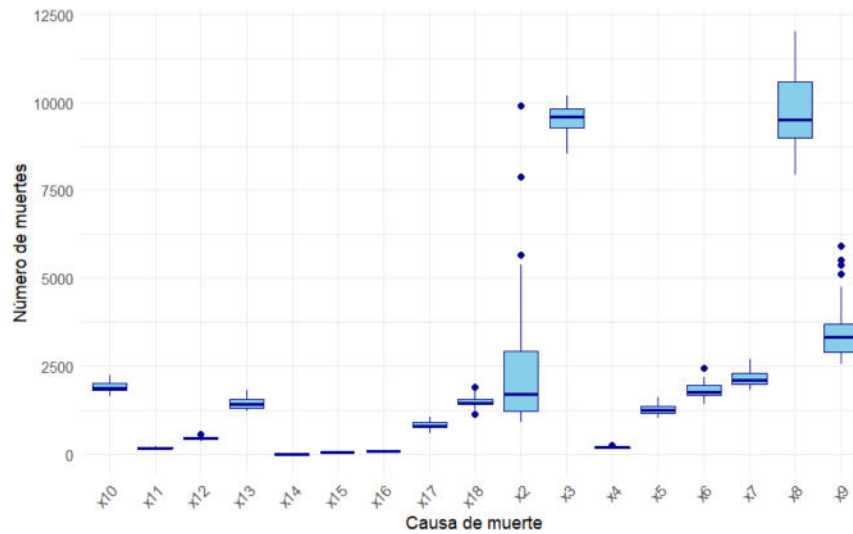


FIGURA 4.2.6. Distribución de las causas de muerte en el periodo post-COVID

En cuanto a los primeros gráficos (Figuras 4.2.3 y 4.2.4), las diferentes causas de muerte siguen un comportamiento similar al observado en el periodo pre-COVID, pero con desplazamientos en el patrón estacional más lentos (en vez de haber picos de muertes en enero, existen picos más amplios abarcando diciembre-febrero). Sin embargo, llama la atención la

causa  $x_2$ , que muestra un comportamiento completamente diferente al resto tanto en el gráfico de su serie temporal como en la representación de su distribución en los diagramas de cajas y bigotes (Figura 4.2.5).

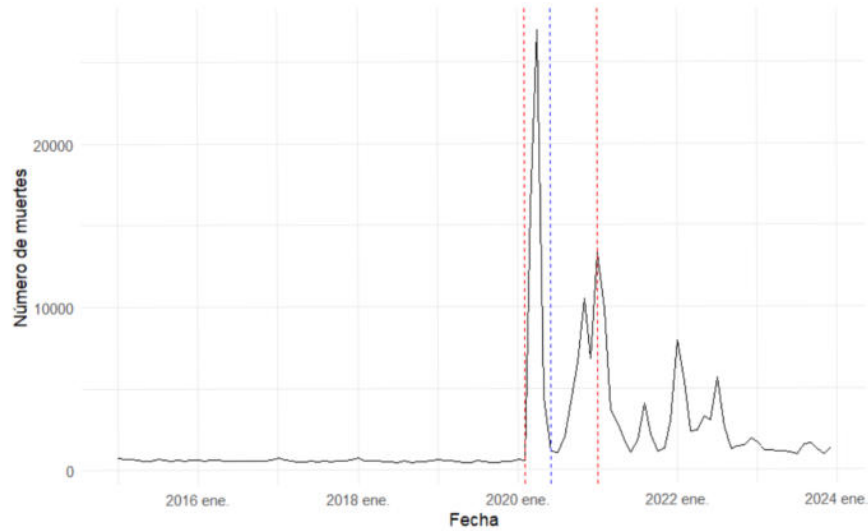


FIGURA 4.2.7. Fallecimientos mensuales debido a la causa  $x_2$ .

Recordando que  $x_2$  representa los fallecimientos por enfermedades infecciosas y parasitarias, dado que la infección por SARS-CoV-2 pertenece a este grupo, es coherente la evolución temporal que se muestra. El comportamiento claramente descendente desde el periodo COVID plantea la hipótesis de que el comportamiento inusual de  $x_2$  se debe a que aún se está diluyendo el efecto que causó la pandemia de COVID-19 en la mortalidad española.

Para explorar la idea anterior, se procede a recalculer los fallecimientos por todas las causas, pero esta vez excluyendo a  $x_2$ :

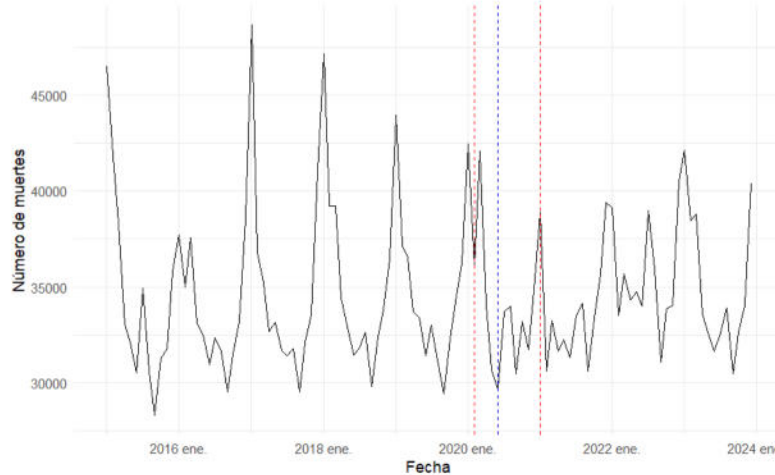


FIGURA 4.2.8. Número de fallecimientos mensuales sin  $x_2$ .

Se puede ver como ahora la serie temporal parece seguir un comportamiento homogéneo a lo largo del tiempo. Se aprecia un ligero descenso de la mortalidad en el periodo COVID, lo cual es coherente con la hipótesis de desplazamiento de mortalidad, según la cual parte de los fallecimientos por SARS-CoV-2 podrían haber anticipado muertes que, en ausencia de la pandemia, habrían ocurrido por otras causas [3].

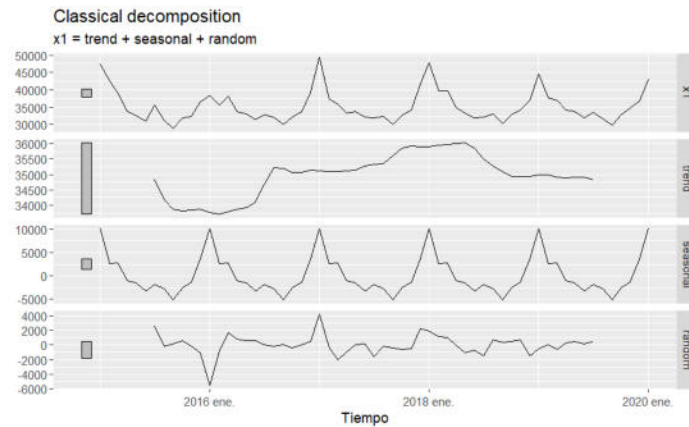
Teniendo en cuenta toda la información obtenida del análisis exploratorio, surge la siguiente hipótesis principal a la que se buscará dar respuesta en los apartados siguientes:

Los patrones de mortalidad post-COVID están volviendo a la normalidad pre-COVID, pero aún existe un efecto en la mortalidad heredado de la pandemia.

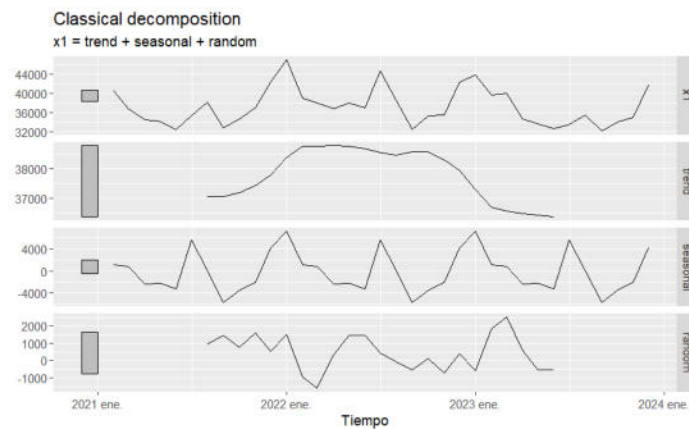
## Análisis estadístico

### 5.1. Estructura de las series temporales

Lo primero que se planteó fue un estudio formal de la tendencia de las series, lo cual se realizó mediante una descomposición clásica aditiva y una descomposición clásica estimando la tendencia mediante regresión LOESS (Descomposición STL).

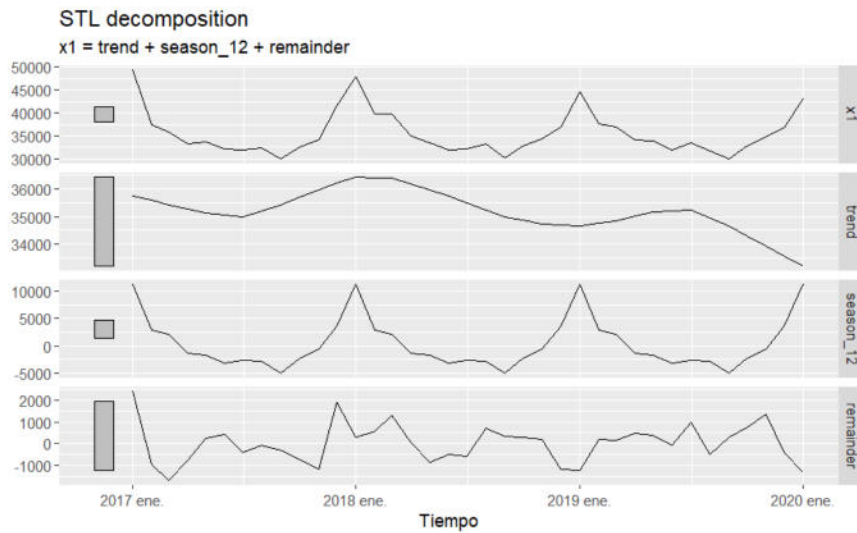


(A) Periodo pre-COVID

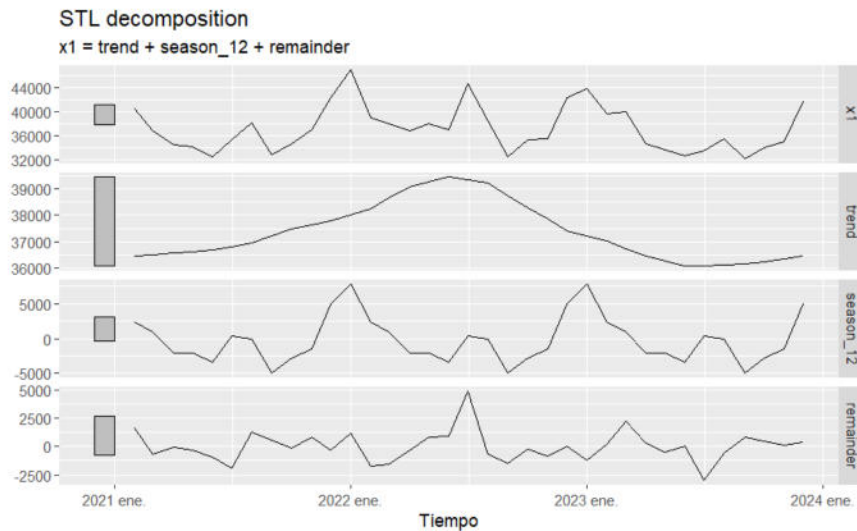


(B) Periodo post-COVID

FIGURA 5.1.1. Descomposición clásica aditiva de los fallecimientos totales.



(A) Periodo pre-COVID



(B) Periodo post-COVID

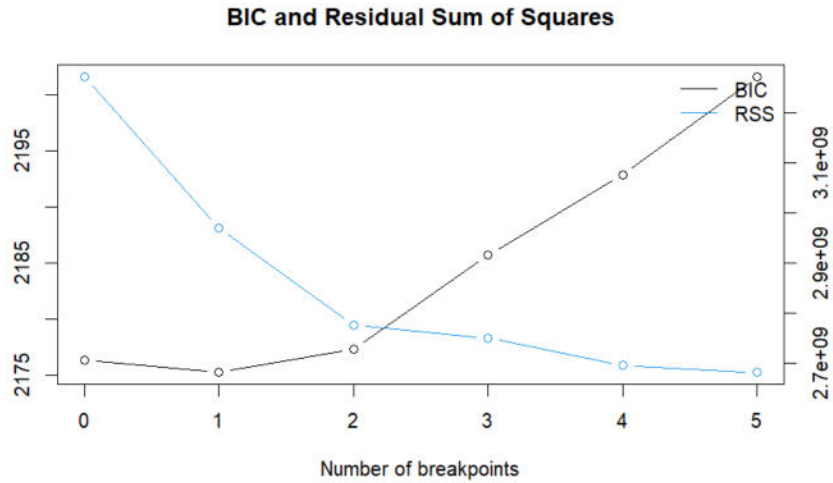
FIGURA 5.1.2. Descomposición clásica mediante LOESS de los fallecimientos totales.

El período pre-COVID presenta un componente estacional anual claro. No se detecta una tendencia aparente significativa en los datos, y las oscilaciones observadas en la tendencia podrían deberse a factores externos como temperatura o causas sociales.

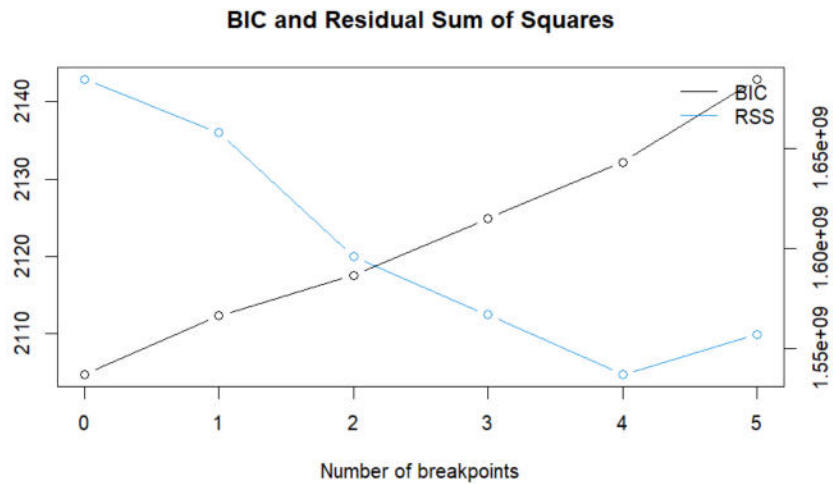
Tras esto, se plantea el test de Bai-Perron (Test de Chow generalizado) en la serie total y la serie sin  $x_2$ , con el objetivo de detectar puntos de ruptura estructurales:

CUADRO 1. Resultados del test de Bai-Perron

Serie	Número de puntos de ruptura	Puntos de ruptura
Serie total	Se detectaron 2 puntos de ruptura	dic 2019, abril 2021
Serie sin $x_2$	Se detectaron 2 puntos de ruptura	mayo 2020, nov 2021



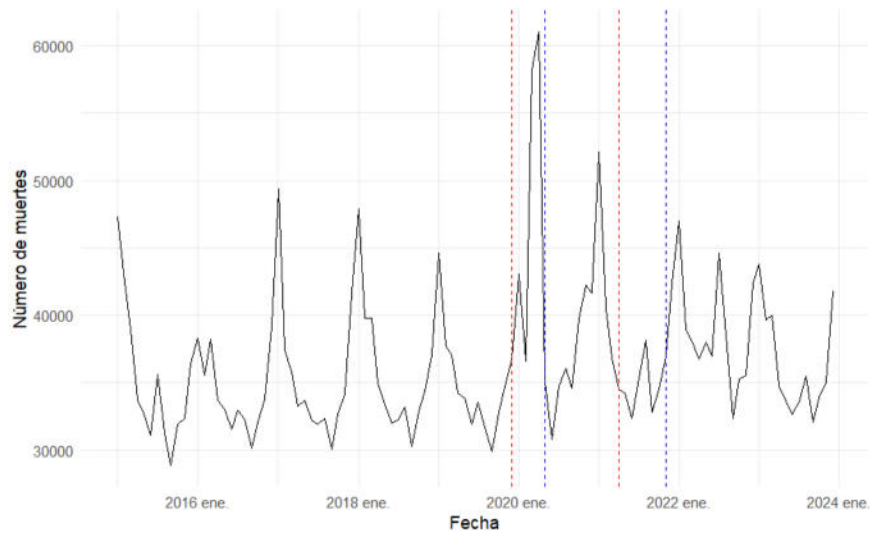
(A) Serie del total



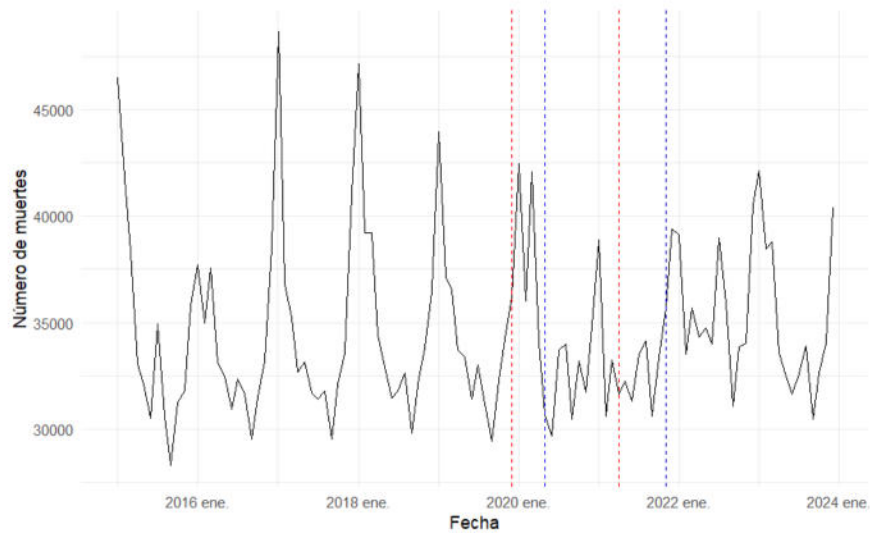
(B) Serie sin  $x_2$

FIGURA 5.1.3. Test de Bai-Perron

En la figura 5.3.1 se muestran los resultados del test de Bai-Perron de forma visual. En ambos casos se detectan dos puntos de ruptura, aunque no coinciden. Como se ilustra en el cuadro 1, estos puntos de ruptura son diciembre de 2019 y abril de 2021 para la serie del



(A) Serie del total



(B) Serie sin  $x_2$

FIGURA 5.1.4. Representación de los puntos de ruptura del test de Bai-Perron total; y mayo de 2020 y noviembre de 2021 para la serie sin  $x_2$ . Para una mejor interpretación de estos resultados, en la figura 5.1.4 se muestran de nuevo ambas series, marcando en rojo los puntos de ruptura de la serie total y en azul los puntos de ruptura de la serie sin  $x_2$ :

- En la serie total (Figura 5.1.4 (A)) se observa como los puntos de quiebre coinciden de forma aproximada con el periodo COVID. Aunque abril de 2021 se aleje de la fecha final del periodo COVID, durante el análisis exploratorio se observó que quizá

el periodo COVID había que extenderlo unos meses más. Estos resultados proponen un candidato a final de periodo COVID.

- En cuanto a la serie sin  $x_2$  (Figura 5.1.4 (B)), se observa un periodo COVID retrasado. Dado que no se observaron diferencias en las causas de fallecimiento, acudiendo a los datos mostrados en la introducción, este periodo podría representar el descenso en la mortalidad debido al aumento del COVID.

## 5.2. Implementación de modelos

Para modelar la serie pre-COVID y poder estimar la mortalidad esperada en el periodo posterior, se analizaron la estacionariedad y ajuste de modelos de series temporales. Se aplicó el test de Dickey-Fuller aumentado (ADF), con hipótesis:

$$\begin{cases} H_0 : \text{Serie no estacionaria} \\ H_1 : \text{Serie estacionaria} \end{cases}$$

Valor crítico $\tau$			Estadístico del contraste
1 %	5 %	10 %	
-3'46	-2'88	-2'57	-2'35

CUADRO 2. Resultados del test de Dickey-Fuller aumentado

Como el estadístico es mayor que todos los valores críticos, no hay evidencia estadística significativa para rechazar la hipótesis de no estacionariedad. Sin embargo, dada la baja potencia del test en muestras pequeñas (61 observaciones), se complementó con el test KPSS, cuya hipótesis es:

$$\begin{cases} H_0 : \text{Serie estacionaria} \\ H_1 : \text{Serie no estacionaria} \end{cases}$$

Nivel KPSS	p-valor
0'0318	> 0'1

CUADRO 3. Resultados del test KPSS

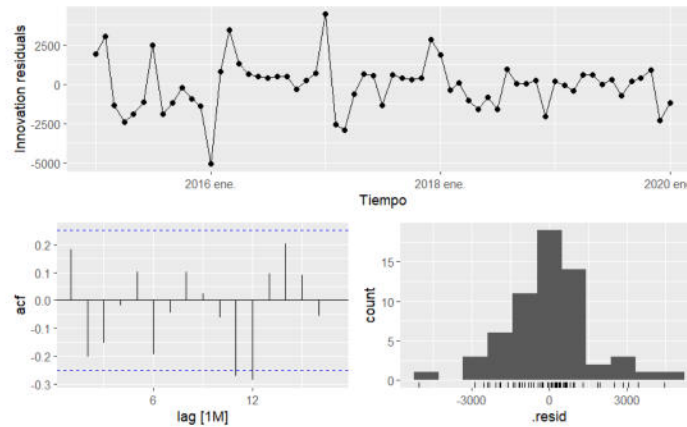
En este caso tampoco hay evidencia suficiente para rechazar la hipótesis nula, pero por como está construido este contraste esto implicaría que se puede considerar estacionaria. Considerando ambos tests, el análisis visual previo y la ausencia de eventos disruptivos durante el periodo pre-COVID, se opta por considerar la serie estacionaria, teniendo en cuenta la incertidumbre en la decisión.

Se van a considerar los siguientes modelos:

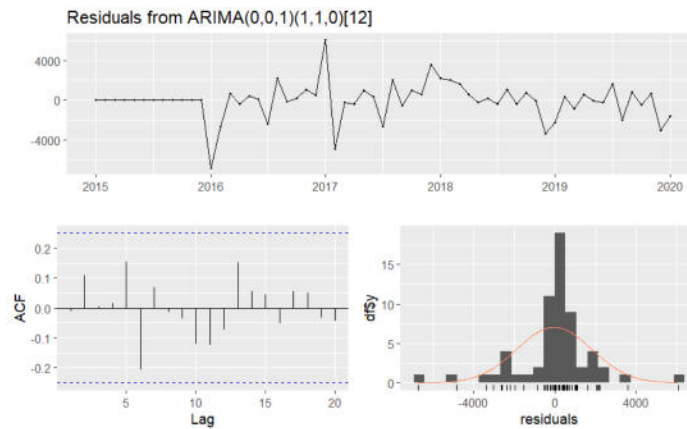
1. Modelo de Holt-Winters (alisado triple).
2. SARIMA(0, 0, 1)  $\times$  (1, 1, 0)<sub>12</sub>.

En la Figura 5.2.1 se muestran los análisis de residuos para ambos modelos, incluyendo distribución temporal, distribución estadística y función de autocorrelación (ACF).

En el modelo Holt-Winters (Figura 5.2.1 (A)) se observa un patrón oscilante en la ACF residual. Aunque las autocorrelaciones no superan claramente los límites de significancia, los lags 11 y 12 superan ligeramente estos límites, indicando posible autocorrelación estacional. Esto se refleja en el test de Ljung-Box (Cuadro 5), con un p-valor de 0'06, sugiriendo dependencia residual. Además, los residuos no evidencian claramente normalidad, lo que podría afectar la inferencia.



(A) Modelo de Holt-Winters



(B) Modelo SARIMA(0, 0, 1) × (1, 1, 0)<sub>12</sub>

FIGURA 5.2.1. Análisis de los residuos

En el modelo SARIMA (Figura 5.2.1 (B)) los parámetros fueron determinados de forma automática mediante el software R con la función “auto.arima()” (Ver ANEXO 4. Modelos de series temporales). Este presenta menores autocorrelaciones residuales, ninguna cercana a los límites de significancia. La distribución residual es más próxima a la normalidad y el test de Ljung-Box arroja un p-valor de 0’5719, sin evidencia de autocorrelación. Esto indica que el modelo SARIMA se ajusta mejor a los datos.

Modelo	MAE	RMSE	MAPE	MASE	AIC	BIC
Holt-Winters	1159’631	1586’333	3’1763	0’7118	1183’803	1219’688
SARIMA(0, 0, 1) × (1, 1, 0) <sub>12</sub>	1122’56	1824’97	3’0328	0’6891	895’67	901’35

CUADRO 4. Medidas de evaluación del modelo

Al comparar las métricas de error (Cuadro 4), el modelo SARIMA presenta un MAE (Error Medio Absoluto) y un MAPE (Error Porcentual Absoluto Medio) ligeramente mejores que el modelo Holt-Winters, aunque el RMSE (Raíz del Error Cuadrático Medio) es mayor en SARIMA. En términos de criterios de información, el modelo SARIMA obtiene valores de AIC y BIC considerablemente inferiores, lo que indica un mejor equilibrio entre ajuste y complejidad del modelo.

Respecto al test de Ljung-Box aplicado a los residuos, los estadísticos y p-valores obtenidos fueron:

Modelo	Estadístico	p-valor
Holt-Winters	35'4434	0'06
<b>SARIMA(0,0,1)×(1,1,0)<sub>12</sub></b>	8'5844	0'5719

CUADRO 5. Test de Ljung

En conclusión, considerando tanto la evaluación gráfica de residuos, el test de autocorrelación y las métricas de error, el modelo SARIMA(0, 0, 1) × (1, 1, 0)<sub>12</sub> se presenta como una mejor opción para modelar la serie temporal objetivo.

### 5.3. Análisis de exceso de mortalidad

Con el modelo seleccionado ajustado a datos pre-COVID, se predice la mortalidad esperada en el periodo post-COVID para comparar con la observada:

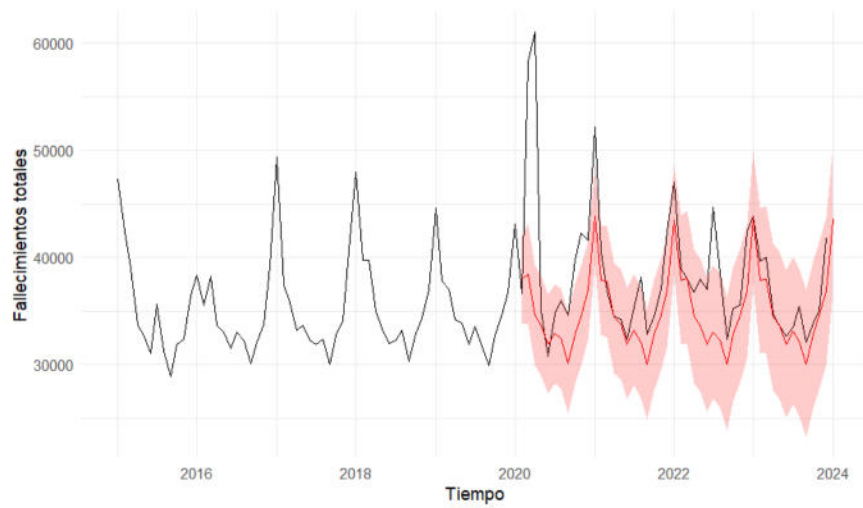


FIGURA 5.3.1. Predicciones del periodo post-COVID

En la Figura 5.3.1 se muestra la serie temporal original, la predicción del modelo (línea roja) y su intervalo de confianza (sombra roja, nivel 95 %). Se observa que con el tiempo la predicción se ajusta a la serie observada, apoyando la hipótesis de estabilización tras el COVID.

## Discusión de los resultados

Los resultados obtenidos en este trabajo permiten evaluar el impacto del COVID-19 en el comportamiento de la mortalidad en España desde una perspectiva temporal. El modelo seleccionado,  $SARIMA(0, 0, 1) \times (1, 1, 0)_{12}$ , mostró un comportamiento adecuado para capturar la estructura de la serie pre-COVID, tanto en términos de ajuste como de independencia de residuos. Es destacable, además, que el propio modelo, al no requerir diferenciación en la parte no estacional ( $d=0$ ), sugiere una estructura estacionaria en la serie, lo que refuerza las conclusiones obtenidas mediante los tests formales (ADF y KPSS), aunque estos fuesen ambiguos debido al tamaño limitado de la muestra (61 observaciones).

Una limitación importante en el desarrollo del trabajo es el limitado número de años post-COVID. Diversas investigaciones han mostrado un aumento en la incidencia de nuevas enfermedades de carácter crónico (Por ejemplo, el COVID crónico) y de las ya existentes [6, 7]. Por lo tanto, la aparente vuelta a los patrones normales de mortalidad quizá sea transitoria, dado que no ha habido tiempo suficiente para notar un aumento significativo en el número de fallecimientos por estas nuevas causas. Esto hace necesario tanto la realización del mismo trabajo con una mayor muestra de datos (dentro de 5/10 años) como la realización del mismo trabajo pero con datos de incidencias, dado que ya deberían ser observables los nuevos patrones (en caso de haberlos).

Desde un punto de vista metodológico, el estudio también está limitado por la naturaleza agregada de los datos. La ausencia de desagregación por edad o causa específica de muerte impide identificar patrones diferenciados entre poblaciones vulnerables (como las personas mayores) y otras menos afectadas por la pandemia. Contar con datos más detallados permitiría modelar de forma más precisa el impacto directo e indirecto del COVID-19 sobre la mortalidad.

En resumen, aunque el modelo y el enfoque propuestos ofrecen una visión útil sobre el exceso de mortalidad tras la pandemia, las limitaciones en los datos disponibles y el momento temporal del análisis sugieren que cualquier conclusión debe ser considerada con cautela. La tendencia a la recuperación del patrón de mortalidad previo podría consolidarse o, alternativamente, esconder transformaciones estructurales que sólo serán visibles a largo plazo.

## Conclusiones

El objetivo principal de este trabajo fue realizar un análisis comparativo de los patrones de mortalidad en España antes y después del inicio de la pandemia de COVID-19. A partir del análisis exploratorio de las series temporales correspondientes al periodo pre y post-COVID, se formularon varias hipótesis preliminares:

1. El periodo COVID podría haberse extendido más allá del periodo considerado.
2. Las diferencias más notables en la mortalidad se atribuyen principalmente a la categoría “Enfermedades infecciosas y parasitarias”
3. Persisten efectos directos de la pandemia sobre los niveles de mortalidad.
4. Se observa una tendencia progresiva hacia la recuperación de los patrones de mortalidad previos.

El análisis formal posterior, compuesto por la descomposición de las series temporales y la posterior implementación de un modelo  $SARIMA(0, 0, 1) \times (1, 1, 0)_{12}$ , ofreció evidencia empírica en apoyo de estas hipótesis. En particular, se confirmó que los puntos de cambio estructural detectados se alinean con un periodo de impacto más prolongado del COVID-19. Asimismo, los resultados sugieren un efecto de “desplazamiento de mortalidad”, consistente con la disminución observada en otras causas de fallecimiento durante el periodo más crítico de la pandemia.

Finalmente, aunque los resultados apuntan hacia una reanudación del comportamiento normalizado de la mortalidad en los últimos años, este estudio pone de manifiesto la necesidad de continuar analizando en profundidad las consecuencias a largo plazo de la pandemia. La falta de datos desagregados, por ejemplo por edad o por causa específica de muerte por COVID-19, limita la capacidad para realizar interpretaciones más precisas y restringe el

alcance de estudios como el presente. No obstante, este trabajo contribuye a la comprensión inicial de la evolución post-pandémica de la mortalidad en España, y abre la puerta a investigaciones futuras más detalladas.

## Bibliografía

- [1] Gaudet LA, Pillay J, Saba S, Zakaria D, Cheta N, Gardiner H, et al. Associations between SARS-CoV-2 Infection and Incidence of New Chronic Condition Diagnoses: A Systematic Review. *Emerging Microbes & Infections*. 2023 Dec;12(1):2204166.
- [2] De Ridder D, Uppal A, Rouzinov S, Lamour J, Zaballa ME, Baysson H, et al. SARS-CoV-2 Infection and the Risk of New Chronic Conditions: Insights from a Longitudinal Population-Based Study. *International Journal of Environmental Research and Public Health*. 2025;22(2):166.
- [3] Lee WE, Woo Park S, Weinberger DM, Olson D, Simonsen L, Grenfell BT, et al. Direct and Indirect Mortality Impacts of the COVID-19 Pandemic in the United States, March 1, 2020 to January 1, 2022. *eLife*. 2023 Feb;12:e77562.
- [4] Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice (3rd Ed)*. Monash University, Australia;.
- [5] Golemund HWaG. *Welcome | R for Data Science*;.
- [6] Hallek M, Adorjan K, Behrends U, Ertl G, Suttorp N, Lehmann C. Post-COVID Syndrome. *Deutsches Arzteblatt International*. 2023 Jan;120(4):48-55.
- [7] Xie Y, Xu E, Bowe B, Al-Aly Z. Long-Term Cardiovascular Outcomes of COVID-19. *Nature Medicine*. 2022 Mar;28(3):583-90.

## Anexos

### ANEXO 1. Preparación de los datos

El fichero de datos se descargó en formato .xlsx desde la web del INE. Se descargó posicionando la variable “Periodo” en las filas y la variable “Causa de muerte” en las columnas. El documento se renombró como “muertes\_total.xlsx”, y se editó la casilla A7 introduciendo la palabra “Tiempo” para posteriormente leerla en R como nombre de variable. Tras la descarga de los datos, se recodificó el nombre de las variables como se comentó en el apartado 3.1.

Para poder trabajar con series temporales, se transformó la variable “Tiempo” a formato fecha. Una vez realizados los ajustes necesarios, se procedió a realizar las comprobaciones mencionadas:

---

#### Algoritmo 1 Preparación de los datos

---

```
```{r datos}
library(readxl)

# Se añadió la palabra 'Tiempo'
datos <- read_excel("C:\\R\\muertes_total.xlsx",
                    range = "A7:S115",
                    col_names = TRUE)
...

```{r Codificacion}
n <- length(names(datos)) - 1
nombres <- list()
for (i in 1:n){
  nombres[[paste0("x", i)]] <- names(datos)[i + 1]
  names(datos)[i + 1] <- paste0("x", i)
}
enf <- function(var){
  ##Función que devuelve el nombre de la enfermedad
  return(nombres[[var]])
}
rm(n, i)
...

```{r Tiempo}
datos$Tiempo <- yearmonth(datos$Tiempo, format = '%YM%m')
...

```{r Comprobaciones}
any(is.na(datos))

total_suma_0 <- datos %>% select(starts_with('x')) %>% sum()
total_todas_causas <- sum(datos$x1)
total_suma_0 == 2 * total_todas_causas

rm(total_suma_0, total_todas_causas)
...`
```

---

Por último, se realizó la segmentación temporal de los datos y se pasaron a formato “serie temporal” de R:

---

**Algoritmo 2** Transformación a series temporales de los datos

---

```

{r Segmentacion}
cortes <- yearmonth(c("2014 Dec", "2020 Feb", "2021 Feb", "2024 Jan"))
datos$Periodo <- cut(as.numeric(datos$Tiempo),
                    breaks = as.numeric(cortes),
                    labels = c("Pre-COVID", "COVID", "Post-COVID"),
                    right = FALSE)

pre_covid <- datos %>%
  filter(Periodo == "Pre-COVID") %>%
  select(-Periodo)

post_covid <- datos %>%
  filter(Periodo == "Post-COVID") %>%
  select(-Periodo)
rm(cortes)

{r Formato_Serie_Temporal}
datos_ts_pre <- pre_covid %>%
  as_tsibble(index = Tiempo)
datos_ts_post <- post_covid %>%
  as_tsibble(index = Tiempo)

```

**ANEXO 2. Análisis de enero de 2016**

Para realizar el análisis de dato atípico, lo primero fue separar los datos de los eneros del resto para manejarlos de forma sencilla. Tras esto, se realizó el gráfico de cajas y bigotes con la función “ggplot()”. El z-score se calculó de forma manual, y el test de Dixon cuenta con una función específica, “dixon.test()”:

---

**Algoritmo 3** Análisis de enero de 2016

---

```

{r 2016}
eneros <- datos %>%
  filter(format(Tiempo, "%m") == "01") %>%
  mutate(año = format(Tiempo, "%Y"))

ggplot(eneros, aes(x = "", y = x1)) +
  geom_boxplot(fill = "lightblue") +
  geom_point(data = eneros %>% filter(año == "2016"), aes(y = x1), color = "red", size = 3) +
  labs(title = "",
       y = "Fallecimientos totales en enero", x = "") +
  theme_minimal()

media_eneros <- mean(eneros$x1)
sd_eneros <- sd(eneros$x1)
enero_2016 <- eneros %>% filter(año == "2016") %>% pull(x1)

z_score <- (enero_2016 - media_eneros) / sd_eneros
cat('z-score: ', z_score)
cat("\n=====\n")

dixon.test(eneros$x1)

```

**ANEXO 3. Gráficos de las series temporales**

En este apartado se presentan los códigos de algunos de los gráficos del trabajo, siendo posible realizar el resto cambiando los datos y realizando ajustes mínimos:

---

### Algoritmo 4 Gráfico de la serie total

---

```
```{r serie_total}
ggplot(datos,aes(x = Tiempo, y = x1)) +
  geom_line() +
  geom_vline(xintercept = as.Date(yearmonth("2020 Feb")),
            color = 'red', linetype = 'dashed') +
  geom_vline(xintercept = as.Date(yearmonth("2020 Jun")),
            color = 'blue', linetype = 'dashed') +
  geom_vline(xintercept = as.Date(yearmonth("2021 Jan")),
            color = 'red', linetype = 'dashed') +
  labs(x = 'Fecha', y = 'Número de muertes') +
  theme_minimal()
```
```

---

### Algoritmo 5 Suavizado mediante LOESS de la serie total post-COVID

---

```
```{r}
ggplot(post_covid, aes(x = Tiempo, y = x1)) +
  geom_line(alpha = 0.4) +
  geom_smooth(aes(color = "0.3"), method = "loess", span = 0.3, se = FALSE) +
  geom_smooth(aes(color = "0.5"), method = "loess", span = 0.5, se = FALSE) +
  geom_smooth(aes(color = "0.7"), method = "loess", span = 0.7, se = FALSE) +
  geom_smooth(aes(color = "0.9"), method = "loess", span = 0.9, se = FALSE) +
  scale_color_manual(
    name = "Span LOESS",
    values = c(
      "0.3" = "blue",
      "0.5" = "darkgreen",
      "0.7" = "red",
      "0.9" = "orange"
    )
  ) +
  labs(
    x = "Fecha",
    y = "Muertes totales",
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")
```
```

---

### Algoritmo 6 Serie por causas de fallecimiento del periodo post-COVID

---

```
```{r}
post_covid %>%
  select(-x1) %>%
  pivot_longer(cols=starts_with('x'),
              names_to="Causa",
              values_to="Total") %>%
  ggplot(aes(x=Tiempo,y=Total,color=Causa,group=Causa)) +
  geom_line() +
  labs(title="Evolución de las muertes en función de su causa",
       x="Periodo",y="Total") +
  theme_minimal()
```
```

---

### Algoritmo 7 Distribución de las causas de muerte en el periodo pre-COVID

---

```
```{r}
pre_covid %>%
  select(-x1) %>%
  pivot_longer(cols = -Tiempo,
              names_to = "Causa",
              values_to = "Muertes") %>%
  ggplot(aes(x = Causa, y = Muertes)) +
  geom_boxplot(fill = "skyblue", color = "darkblue") +
  labs(x = "Causa de muerte",
       y = "Número de muertes") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
```

---

**Algoritmo 8** Descomposición clásica de la serie total post-COVID

---

```
```{r}
decomp_clasica_pre <- datos_ts_pre |>
  model(cts = classical_decomposition(x1 ~ season(12), type = "additive"))

decomp_clasica_pre |>
  components() |>
  autoplot()
```
```

---

**Algoritmo 9** Media móvil de la serie total pre-COVID

---

```
```{r}
total1 <- pre_covid %>%
  select(Tiempo, x1)
k <- 12
total1$x1_linear <- predict(total1 |> lm(formula = x1 ~ Tiempo), total1)
total1 <- total1[order(total1$Tiempo), ]
total1$x_smooth_12ma <- stats::filter(total1$x1, filter = rep(1/k, k))

ggplot(total1 |>
  pivot_longer(cols = -Tiempo, names_to = "var", values_to = "values")) +
  geom_line(aes(x = Tiempo, y = values, color = var, linetype = var),
    linewidth = 0.9, alpha = 0.75) +
  ggthemes::scale_color_colorblind() +
  theme_minimal()
```
```

## ANEXO 4. Modelos de series temporales

En este apartado se presentan los códigos principales para generar los resultados del análisis estadístico:

---

**Algoritmo 10** Prueba de Bai-Perron para la serie total pre-COVID

---

```
```{r}
data <- datos |>
  as.data.frame()

modelo <- breakpoints(data$x1 ~ 1)
plot(modelo)
summary(modelo)

cat("=====\n")

datos$Tiempo[c(33, 49)]
```
```

---

**Algoritmo 11** Pruebas de estacionariedad para la serie total pre-COVID

---

```
```{r}
df <- ur.df(datos_ts_pre$x1, type = "none", lags = 12)
summary(df)

kpss <- kpss.test(datos_ts_pre$x1)
kpss
```
```

---

### Algoritmo 12 Modelo de Holt-Winters de la serie total pre-COVID

---

```
```{r Holt-Winters}
fit_pre <-
  datos_ts_pre |>
  model[alisado_triple = ETS(x1 ~ error("A") + trend("A") + season("A"))]

fit_pre |>
  gg_tsresiduals()

fit_pre |>
  augment() |>
  features(.resid, ljung_box, lag = 24)

accuracy(fit_pre)
glance(fit_pre)
```
```

---

### Algoritmo 13 Modelo SARIMA (automático) de la serie total pre-COVID

---

```
```{r SARIMA}
serie <- ts(datos_ts_pre$x1, start = c(2015, 1), frequency = 12)

modelo <- auto.arima(serie, seasonal = TRUE)
summary(modelo)
checkresiduals(modelo)
```
```

---

### Algoritmo 14 Gráfico con las predicciones del modelo SARIMA

---

```
```{r}
serie <- ts(datos_ts_pre$x1, start = c(2015, 1), frequency = 12)
modelo <- auto.arima(serie, seasonal = TRUE)
futuro <- forecast(modelo, h = 4*12)
futuro_df <- data.frame(
  Tiempo = time(futuro$mean),
  mean = as.numeric(futuro$mean),
  lower_80 = as.numeric(futuro$lower[,1]),
  upper_80 = as.numeric(futuro$upper[,1]),
  lower_95 = as.numeric(futuro$lower[,2]),
  upper_95 = as.numeric(futuro$upper[,2])
)
datos_df <- datos %>% select(Tiempo, x1)

ggplot() +
  geom_line(data = datos_df, aes(x = as.Date(Tiempo), y = x1)) +
  geom_line(data = futuro_df, aes(x = as.Date(Tiempo), y = mean), color = "red") +
  geom_ribbon(data = futuro_df, aes(x = as.Date(Tiempo), ymin = lower_95, ymax = upper_95), alpha = 0.2, fill = "red") +
  theme_minimal() +
  labs(title = "", y = "Fallecimientos totales", x = "Tiempo")
```
```

## ANEXO 5. Librerías y funciones

En este apartado se muestra un resumen de las librerías y funciones usadas durante el trabajo:

---

### Algoritmo 15 Librerías utilizadas

---

```
```{r, warning = FALSE, message = FALSE}
library(seasonal)
library(psych)
library(zoo)
library(tidyverse)
library(tsibble)
library(ggplot2)
library(scales)
library(viridis)
library(gridExtra)
library(feasts)
library(tsibbledata)
library(fable)
library(forecast)
library(tsfeatures)
library(tseries)
library(urca)
library(dplyr)
library(ggpubr)
library(EnvStats)
library(outliers)
library(strucchange)
```
```

---