La estacionalidad de la mortalidad en España. Sus causas y modelización.

VÍCTOR BARRIGA LUCAS
Actuario. Chief Actuary Spain and Portugal. RGA re Internacional Ibérica

El estudio de la mortalidad no puede dejar al margen un análisis más profundo de las causas subyacentes, así como los factores que influyen en la misma, como pueden ser factores socioeconómicos, demográficos o incluso geográficos.

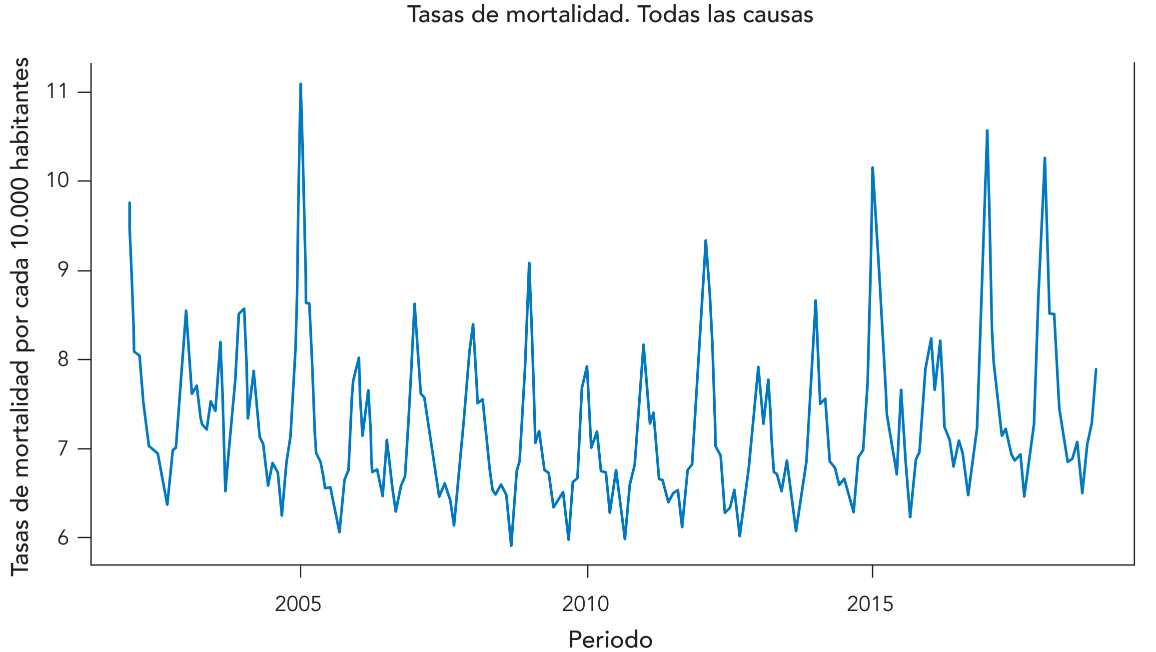
En este artículo nos centraremos en la estacionalidad de la mortalidad, cuáles son las causas de fallecimiento que principalmente contribuyen a ello, así como de su modelización.

Para comenzar, empezaremos definiendo qué entendemos por estacionalidad. La estacionalidad se refiere a un patrón o tendencia que se repite cada cierto tiempo, es decir, cada mes, cada trimestre o, como es el caso que queremos abordar, cada año.

De cara a realizar este análisis se han tomado en consideración únicamente los datos de defunción por causa y mes publicados por el Instituto Nacional de Estadística (www.ine.es), para el periodo comprendido entre los años 2002 y 2018.

En la Figura 1 aparecen las tasas de fallecimiento mensuales de España para el periodo 2002 – 2018, donde se puede observar un patrón estacional, en el

Figura 1: Gráfico de tasas de mortalidad por cada 10.000 habitantes



Fuente: *elaboración propia a partir de los datos del INE.*

que los meses invernales son aquellos donde la tasa de mortalidad es más alta, mientras que los meses de verano presentan tasas de mortalidad más bajas.

Esta mortalidad superior durante los meses correspondientes a la estación invernal puede afectar a las aseguradoras, y en especial a las aseguradoras de vida. Entender cuáles son las causas de la mortalidad estacional, así como las razones por las cuales en ciertos años la mortalidad estacional es mayor que la esperada, puede ser vital para la planificación de las compañías de (re)aseguros.

A pesar de que la mortalidad en los meses más fríos se espera que sea mayor, la magnitud de la estacionalidad varía año a año. Una manera de valorar el grado de estacionalidad puede ser a través de un ratio que compare los meses fríos con los meses cálidos. De este modo, para calcular el *Ratio Invierno - Verano* de un año concreto dividiremos el número de muertes de los meses diciembre, enero y febrero entre el número de muertes de los meses junio, julio y agosto, donde el número de muertes de diciembre corresponden al año anterior:

$$\text{Ratio Invierno} - \text{Verano} = \sum_{t=diciembre}^{febrero} \frac{\textit{Número de Muertes}_t}{\sum_{t=junio}^{agosto} \textit{Número de Muertes}_t}$$

En la Figura 2, podemos ver claramente que los meses de invierno experimentan niveles más altos de mortalidad con un Ratio Invierno - Verano superior a uno. También se puede observar que el grado de estacionalidad varía año a año, siendo en algunos años excepcionalmente alta.

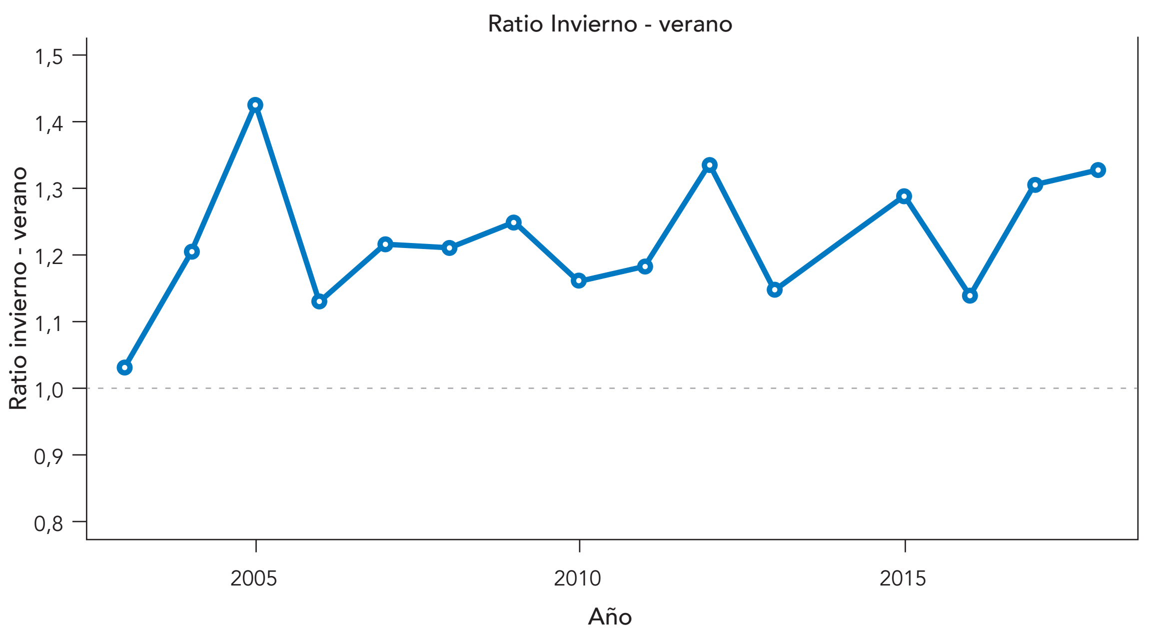
Llama la atención, por ejemplo, el año 2005, donde la gripe, convirtió este año en el año con más muertes en España hasta esa fecha. Las defunciones crecieron sobre todo en el primer trimestre del año un 21% más que el mismo periodo de 2004, mientras que la población únicamente creció un 2%.

Análisis de las causas de la estacionalidad en la mortalidad

Los factores que influyen en la mortalidad pueden tener un origen variado. Estas van desde factores intrínsecos, como pueden ser los hábitos y comportamiento de la persona, hasta factores externos, como la contaminación, el acceso a la medicina o la incidencia de la gripe.

En 2018, el 67% de las muertes en España fueron provocadas por las tres principales causas de mortalidad. En particular, las enfermedades del

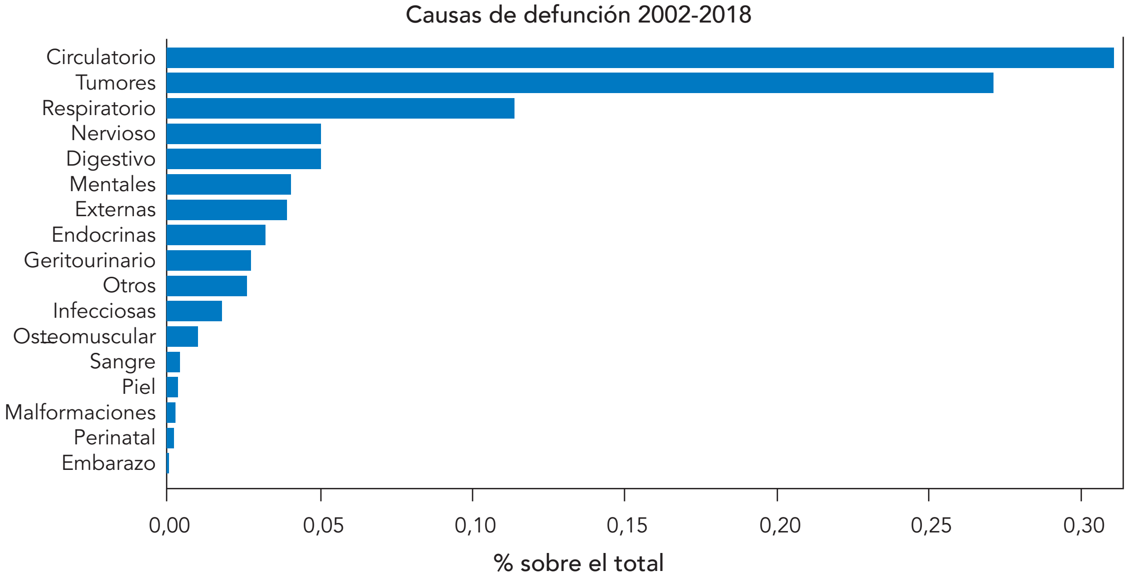
Figura 2: Gráfico del Ratio Invierno - Verano por año



Fuente: *elaboración propia a partir de los datos del INE.*

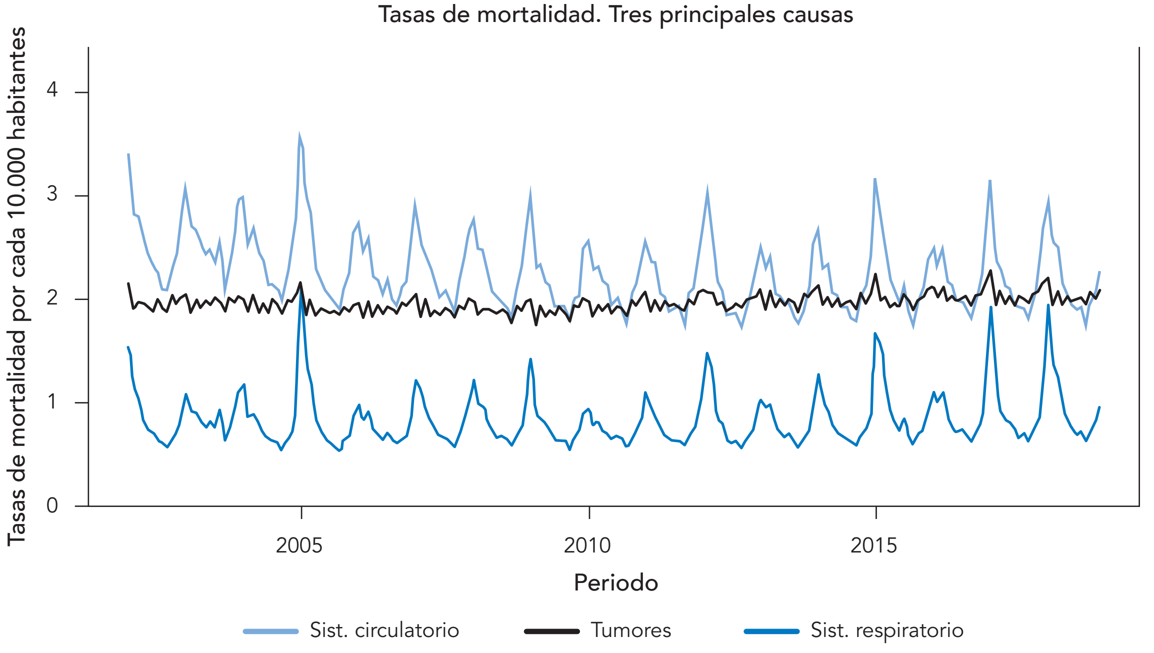
sistema circulatorio representaron el 28%, los tumores el 26%, y enfermedades del sistema respiratorio el 13%. Si miramos toda la serie (2002 - 2018) en la Figura 3, las principales causas de defunción no varían, aunque éstas representan el 70% del total.

Figura 3: Gráfico de las causas de defunción en España periodo 2002 - 2018



Fuente: *elaboración propia a partir de los datos del INE.*

Figura 4: Gráfico de tasas de mortalidad para las tres principales causas de mortalidad



Fuente: *elaboración propia a partir de los datos del INE.*

Centrándonos en los tres bloques principales, observamos que los fallecimientos por Tumores no presentan signos de estacionalidad, con unos ratios invierno-verano próximos a uno, mientras que aquellos provocados por Enfermedades de Sistema Circulatorio y Enfermedades del Sistema Respiratorio sí que lo presentan. Tanto en el caso de las defunciones por Enfermedades del Sistema Circulatorio, como por Enfermedades del Sistema Respiratorio, las tasas de mortalidad más altas las encontramos en los meses más fríos del año. En la Figura 4 se representan las tasas de mortalidad mensuales de las tres principales causas de defunción en España.

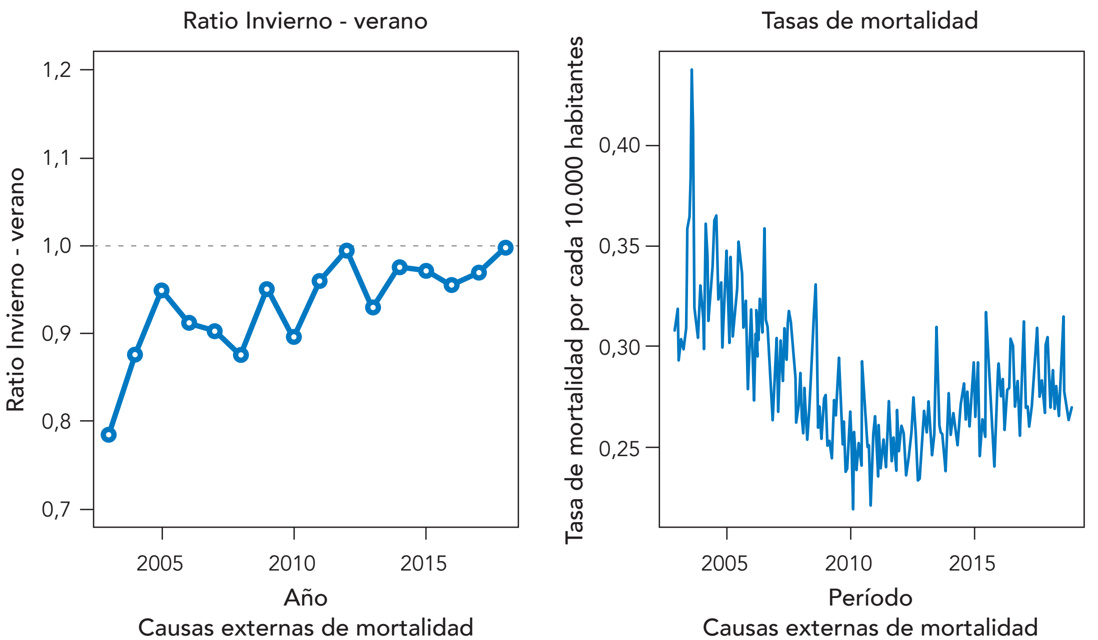
No siempre la estacionalidad presenta tasas de mortalidad más altas en los meses de invierno. Un ejemplo de ello son aquellas defunciones que se engloban dentro del epígrafe de Causas Externas de Mortalidad. Como se puede observar en la Figura 5, los Ratio Invierno - Verano para las Causas Externas de Mortalidad son menores de uno, lo cual indica un mayor número de fallecidos por estas causas en los meses más calurosos, y que coinciden con los periodos vacacionales más relevantes en España. Es destacable el verano de 2003, donde fallecieron 12.919 personas más que en el año 2002 durante los meses de junio, julio y agosto a causa de la ola de calor, y que supuso que la tasa de mortalidad en el verano de 2003 fuese un 13% superior a la del año anterior.

El análisis de las causas de la estacionalidad en la mortalidad no puede quedar completo sin tener en cuenta dos de los factores de riesgo más utilizados por las (re)aseguradoras de vida, como son la edad y el género. Mientras que por género no se observan grandes diferencias respecto al total, no ocurre lo mismo con la edad.

Para una comparación entre edades más efectiva, dividiremos los datos por bandas de edad, y posteriormente estandarizaremos las tasas mensuales de mortalidad a la media de cada tramo etario, y que se mostrarán en base 100. En la Figura 6 se muestran los datos comparativos tanto por géneros como por tramos etarios para el año 2018, y donde observamos cómo las edades más altas son aquellas que contribuyen más significativamente a la estacionalidad de la mortalidad en España, en especial los tramos etarios de 60 a 79 años y de más de 80 años, mientras que para las edades más jóvenes apenas se observa este fenómeno.

Por géneros, se ha seguido la misma metodología, pero en este caso no se aprecian diferencias de com-

Figura 5: Gráfico de los Ratio Invierno – Verano y tasas mensuales de mortalidad para las Causas Externas de mortalidad



Fuente: *elaboración propia a partir de los datos del INE.*

portamiento y ambos géneros tienen un comportamiento similar.

Series temporales y componentes de la mortalidad

Una serie temporal no es más que observaciones de una variable aleatoria a lo largo del tiempo, es decir, un conjunto de datos ordenados en el tiempo y que pueden presentar correlación entre ellos.

En nuestro caso, vamos a disponer de un conjunto ordenado de datos de mortalidad mensuales por causa de fallecimiento que previamente hemos derivado con los datos del Instituto Nacional de Estadística. Además, utilizaremos R como software estadístico para ilustrar cómo podemos analizar las series temporales de tasas de mortalidad mensuales.

Una vez preparados los datos, que hemos grabado en un fichero delimitado por comas (Datos.csv), lo primero que haremos será leerlos utilizando la función read.csv(). El siguiente código de R nos permitirá cargar nuestros datos y guardarlos en el objeto datos. El fichero que está siendo cargado contiene las tasas mensuales de mortalidad para todas las causas de defunción entre los años 2002 a 2018:

```
> datos <- read.csv("Datos.csv")
```

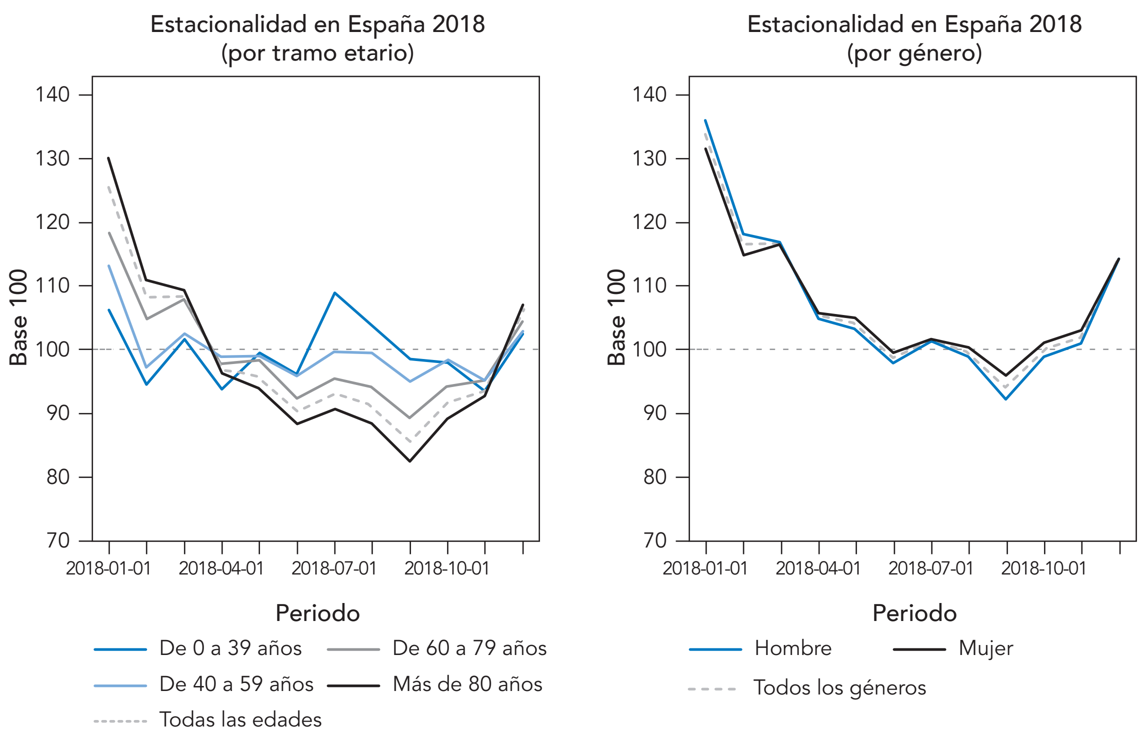
Para que R entienda que se trata de una serie temporal utilizaremos la función ts(), creando el objeto ts_datos, donde indicaremos además la frecuencia y la fecha de comienzo de los datos.

```
> ts_datos <- ts(datos, frequency = 12, start = c(2002,1))
```

De este modo, tendremos cargados los datos de mortalidad mensuales como una serie temporal en nuestro software estadístico.

Cuando analizamos datos, una de las primeras tareas que solemos realizar es la visualización de los mismos, lo cual nos ayudará a entenderlos de una manera rápida. Utilizaremos las siguientes líneas de código para genera un gráfico de líneas, muy similar a los mostrados al inicio del artículo y en el cual observaremos sus potenciales patrones, entre ellos el de su estacio-

Figura 6: Gráfico de estacionalidad en España 2019 por tramo etario y por género



Fuente: elaboración propia a partir de los datos del INE.

nalidad. Hay varias librerías que se pueden utilizar en R para la realización de gráficos, aunque la función que utilizaremos para este propósito será la función *plot()*.

> plot(ts_datos)

Suelen distinguirse cuatro elementos o componentes en una serie temporal:

- Tendencia
- Componente Cíclica
- Componente estacional
- Variaciones residuales

La *Tendencia*, es un componente de larga duración, y que nos señala la dirección en la que se mueve el fenómeno observado, en nuestro caso la mortalidad, y que representa la evolución general del mismo. Si la única componente que se presentara en una serie temporal fuese la Tendencia, el método natural para estimarla sería la regresión de mínimos cuadrados, mediante el cual ajustaríamos la recta, polinomio, exponencial o tendencia logística, a la nube de puntos proporcionada por la serie temporal.

La *Componente Cíclica*, son movimientos con amplitud y periodos, por lo general, superiores a un año y no constantes. Es decir, son movimientos que se repiten de forma periódica, pero que no son tan regulares como los de la Componente Estacional.

La *Tendencia* y la *Componente Cíclica* son a menudo combinadas, refiriéndose a ellas como Componentes Extra-estacionales o directamente como Tendencia, como veremos más adelante cuando descompongamos nuestra serie temporal.

La *Componente Estacionales* son variaciones o fluctuaciones que se producen con pautas de comportamiento periódico, generalmente conocido y constante, e inferior al año.

Por último, las *Variaciones Residuales*, también conocidas como ruido, son pequeñas perturbaciones sin regularidad alguna y de una amplitud relativamente débil. Éstas representan el resto de la serie temporal no capturadas por las otras tres componentes.

La función *stl()* nos va a permitir descomponer una serie temporal en sus tres diferentes componentes, ya que como comentamos anteriormente, tanto la componente Tendencia como la Componente Cíclica se muestran combinadas bajo la etiqueta de Trend. Una descomposición de la serie temporal puede ayudarnos a comprender la serie temporal. La función *stl()* hace uso de una regresión polinómica para la descomposición y extraer las componentes Estacional y Tendencia de la serie temporal. El siguiente código de R realiza la descomposición de la serie temporal de las tasas mensuales de mortalidad y grafica sus componentes.

> descomponer<-stl(ts_datos, s.window= "periodic")

> plot(descomponer)

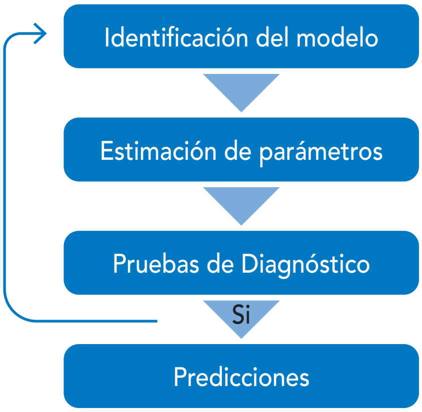
La Figura 7 muestra la descomposición de los componentes de la serie temporal. El primero de los gráficos denominado data, muestra los datos originales utilizados, seguido por la *Componente Estacional*, seasonal. En este gráfico, se observan dos picos adicionales al de enero, que corresponden con los meses de marzo y agosto, los cuales suelen coincidir con los meses más relevantes de vacaciones en España. El siguiente gráfico se refiere a la componente *Tendencia*, bajo la etiqueta trend, donde se pueden observar algunas de las características que hemos comentado con anterioridad, como por ejemplo, el incremento de fallecidos en el verano de 2003, el incremento de muertes en el invierno de 2005, además de la ligera tendencia creciente desde el año 2013 en adelante. Para finalizar, en el último de los gráficos nos encontramos las *Variaciones residuales*, denominadas en el gráfico como remainder.

Modelos ARIMA aplicados a Serie Temporal de tasas mensuales de mortalidad en España

Existe una amplia literatura sobre la aplicación en series temporales de modelos autoregresivos integrados de medias móviles, también conocidos por sus siglas en inglés ARIMA(p,d,q), Autoregressive Integrated Moving Average model, y cuyos parámetros son p, d y q. R tiene numerosos procedimientos para estimar, validar y predecir con estos modelos.

A menudo, las series temporales poseen un componente estacional que se repite en todas las observaciones. Para hacer frente a la estacionalidad, los procesos ARIMA(p,d,q) han sido generalizados, estableciendo los modelos SARIMA(p,d,q)(P,D,Q)[m], Seasonal Autoregressive Integrated Moving Average model, donde P, D, y Q son los parámetros de la Componente Estacional y m es el periodo de la estacionalidad.

El análisis de una serie temporal y su ajuste a un proceso ARIMA, o SARIMA, suele realizarse en varias etapas.



Primero debemos determinar los menores valores de p, d, y q del proceso ARIMA, así como los de P, D y Q en caso de un proceso SARIMA, que son necesarios para explicar la serie temporal, esta primera etapa se denomina Identificación del modelo.

Después debemos estimar los parámetros del modelo y contrastar si son o no significativos. La estimación se puede realizar utilizando métodos de máxima verosimilitud.

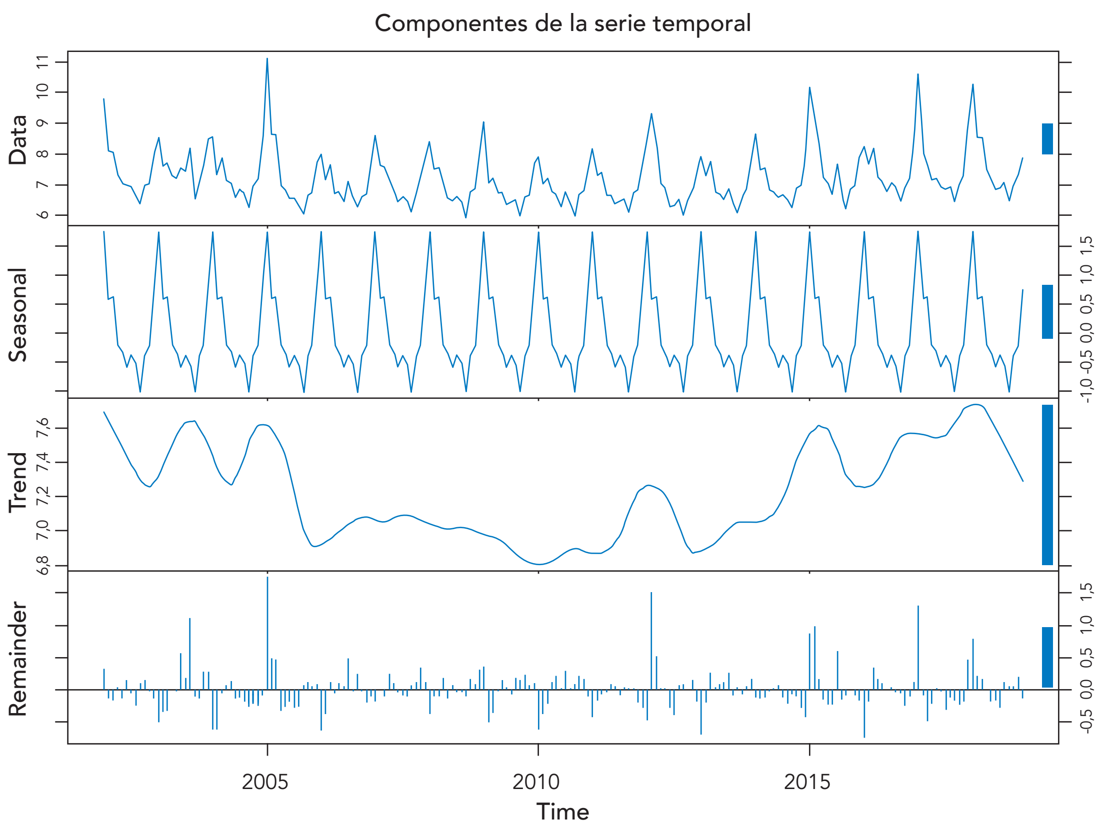
Posteriormente deberemos chequear si el modelo ajusta adecuadamente a los datos mediante pruebas de diagnóstico.

Y por último, en el caso de que las pruebas de diagnóstico sean satisfactorias, entonces podremos hacer predicciones. En caso contrario, tendremos que volver a iniciar el proceso.

Para facilitarnos la tarea vamos a utilizar la librería de R *forecast*, la cual incluye la función *auto.arima()* que realiza un ajuste automático de un modelo SARIMA a una serie temporal dada. De este modo, y en pocas líneas de código podremos crear nuestro modelo. El código de R para hacer esto se muestra a continuación, donde previamente habremos cargado la librería *forecast*.

```
> library(forecast)
```

Figura 7: Gráficos de Descomposición de Componentes Serie Temporal



Fuente: *elaboración propia.*

```
> ajuste1 <- auto.arima(ts_datos, stepwise = FALSE, approximation = FALSE)
```

Es habitual que un modelo no ajuste muy bien a una serie muy larga, y hace necesario hacer ajustes por periodos de tiempo, siendo especialmente interesante obtener un buen ajuste para el periodo final si lo que queremos es hacer predicciones. Es por ello, que a la hora de crear el modelo hemos seleccionado los últimos periodos de la serie, desde 2015 al 2018.

```
> seriefinal <- window(ts_datos, start = c(2015, 12), end = c(2018, 12))
```

```
> ajuste2 <- auto.arima(seriefinal, stepwise = FALSE, approximation = FALSE)
```

Las pruebas de diagnóstico del modelo estimado puede llevarse a cabo mediante la función *tsdiag()*. El código de R para hacer las pruebas de diagnóstico se muestra a continuación, y el cual generará los gráficos de la Figura 8.

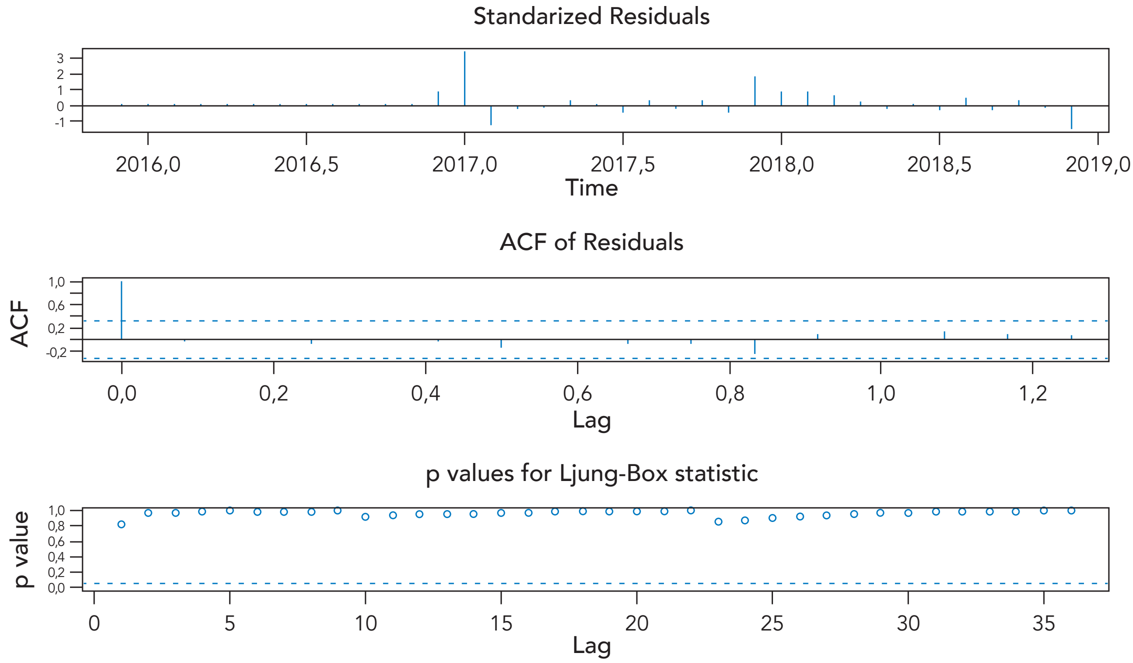
```
> tsdiag(ajuste2)
```

Para chequear la validez de los modelos, la función *tsdiag()* nos grafica los residuos del modelo, un correlograma de los residuos, y calcula el estadístico Ljung-Box para testear la hipótesis nula de que los residuos no están autocorrelacionados. La Figura 8 muestra los residuos estandarizados del modelo, la función de autocorrelación de los residuos y los p-valores del estadístico Ljung-Box a diferentes retardos o lags. En nuestro caso, el estadístico Ljung-Box es no significativo a diferentes retardos, así como la función de autocorrelación (ACF) de los residuos no muestra una correlación significativa, lo que nos lleva a considerar como adecuado el modelo ajustado. Cabe señalar que la función *tsdiag()* exagera los p-valores para la prueba de Ljung-Box, pero como nuestro modelo es relativamente simple, esta exageración es menor.

Por último, y seguramente sea incluso de mayor interés, una vez construido el modelo, se pueden realizar predicciones mediante la función *forecast()* de la librería de R con el mismo nombre. El código de R para hacer esto se muestra a continuación, y utilizaremos la función *plot()* para construir la Figura 9.

```
> estimacion <- forecast(ajuste2)
```

Figura 8: Gráficos de resultados de las Pruebas de diagnóstico



Fuente: *elaboración propia.*

> plot(estimacion)

En la Figura 9 se muestran los parámetros del modelo SARIMA(0,0,1)(1,1,0)[12], y en color azul las predicciones para los dos años posteriores al último periodo disponible, así como los intervalos predicción del 80% y 95% que por defecto nos calcula la función *forecast()*.

Debemos tener en cuenta que este tipo de modelos no capturan eventos extraordinarios, como una pandemia, por lo que se espera que nuestra predicción para 2020 se desvíe de las cifras reales.

Conclusiones

La mortalidad en España presenta estacionalidad propia del hemisferio norte, donde los meses de invierno hay una mayor mortalidad que en los meses de verano.

Esta estacionalidad se debe en gran parte a que dos de las tres principales causas de mortalidad presentan esta característica. En particular, enfermedades del sistema circulatorio y del sistema respiratorio.

Existen otras causas con estacionalidad con signo contrario, como son las causas externas de mortalidad, donde la tasa de mortalidad es más elevada en los meses de verano.

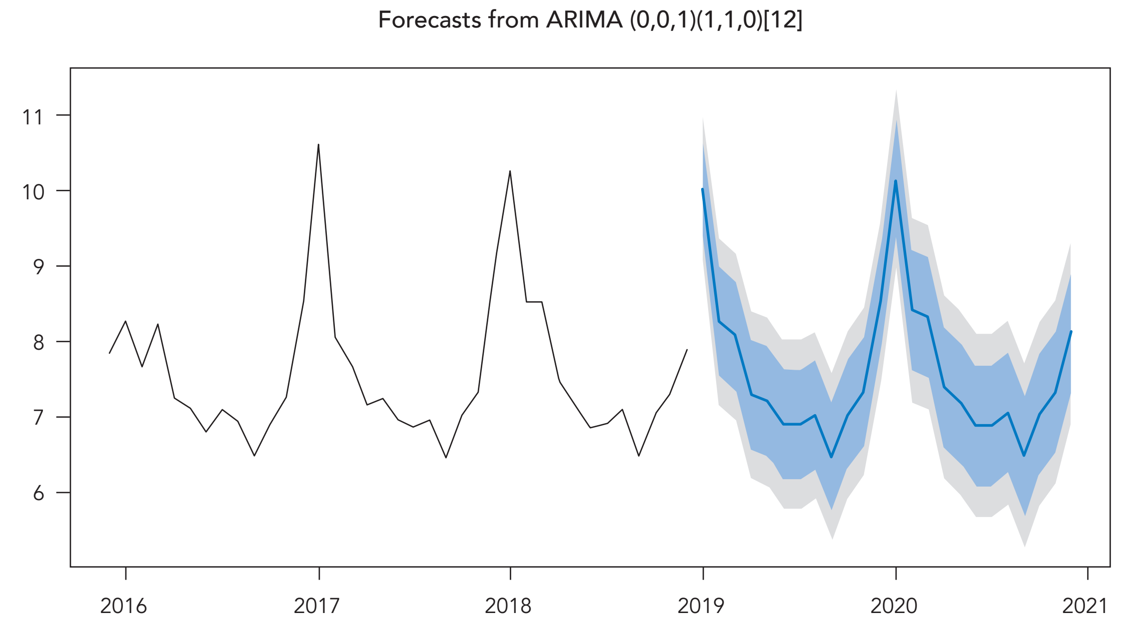
Si bien por género no se aprecian grandes diferencias, sí se observan por edad, donde las edades más altas tienen una estacionalidad de la mortalidad más acusada.

La magnitud de la estacionalidad no es constante en el tiempo. Una manera de cuantificarla es a través del ratio invierno-verano.

Con el uso de herramientas estadísticas modernas y potentes, podremos analizar una serie temporal, en nuestro caso las tasas de mortalidad mensuales de España, observar y comprender las componentes de esta, y realizar una modelización con modelos autoregresivos estacionales de medias móviles (SARIMA) con la finalidad de realizar predicciones, las cuales pueden ser de interés a las compañías de (rea)seguros.

Finalmente, debemos tener en cuenta que este tipo de modelos no captura eventos extraordinarios, como la situación actual de la pandemia de provocada por el Covid-19, y nuestras predicciones para 2020 se desvíen de las cifras reales. ●

Figura 9: Gráfico de predicción



Fuente: *elaboración propia.*