Series Temporales

Parte práctica

Introducción

Una serie temporal es una secuencia de datos que han tenido lugar a lo largo del tiempo. La información que se describe en estas series está ordenada cronológicamente y designa un valor a cada punto del tiempo. Es de interés general el análisis de estas series para poder estimar qué va a suceder en el futuro, y de este modo realizar una predicción relativamente acertada.

En este estudio se va a proceder a analizar dichas series mediante diversas herramientas, para posteriormente evaluar resultados y valorar su fiabilidad. Todas ellas tendrán lugar en el entorno RStudio, usando la librería “tseries” (<https://cran.r-project.org/web/packages/tseries/tseries.pdf>), lo cual facililtará sustancialmente el análisis de las Series temporales que se presentan. El proceso a llevar se describe en el siguiente apartado, pero primero se realizará una breve descripción de las herramientas a utilizar para definir el contexto del trabajo.

Metodología Box-Jenkins

Esta metodología se basa en realizar ciertas transformaciones sobre la serie temporal, para así poder aplicar otros métodos. Para ello, divide la serie en tres características: Tendencia T(t), Estacionalidad S(t) y una componente irregular E(t).



Las transformaciones a realizar son eliminar la tendencia y la estacionalidad de la serie. Luego hacer la componente E(t) estacionaria y aplicar métodos paramétricos. Para ello existen diversas técnicas de eliminación.

Las técnicas de tratamiento de la tendencia más comunes son de tres tipos. La primera se refiere a una Estimación funcional, en la cual se aproxima la tendencia de la serie con una función representativa. Es la que se va a utilizar en este estudio, concretamente la estimación mediante un modelo lineal. Luego existe la técnica de Filtrado, en la que se calcula la media por áreas temporales, obteniendo así la evolución de la media a lo largo del tiempo. Por último, está el método de Diferenciación, en el que se diferencia la señal un número de veces hasta que se descubra su tendencia.

Por otra parte, el análisis de la estacionalidad se lleva a cabo mediante la visualización de la gráfica ACF. Ésta muestra una gráfica sinusoidal cuando existe dicha característica en la serie. Para calcular el parámetro de Estacionalidad es necesario estimar (según los datos) el periodo que poseen. Para ello, basta con observar la gráfica o conocer alguna característica de los datos (como por ejemplo, que se trata de unos datos recogidos mensualmente).

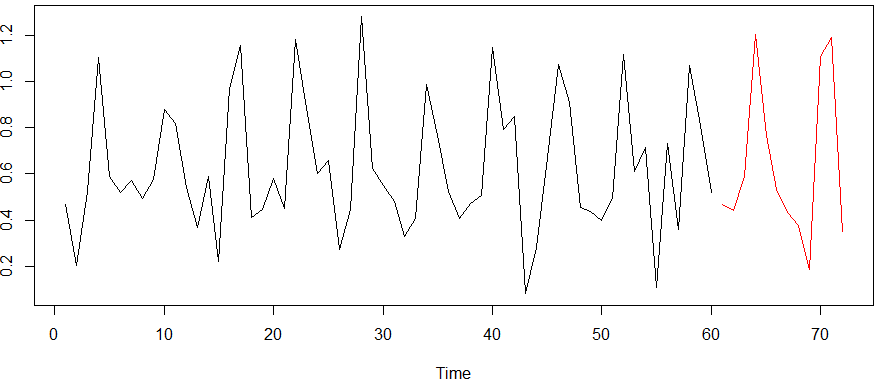
El cálculo de la estacionalidad es simple. Se basa en calcular las medias de los valores dependiendo de en qué ciclo de la gráfica se encuentran. Por ejemplo, en caso de que el periodo sea 12, para el ciclo 1, la media se calcula con los valores {0, 12, 24, 36, …}.

Por último, se utilizará el modelo ARIMA para generar las predicciones sobre la serie temporal. Este modelo usa las variaciones y regresiones de los datos con el objetivo de encontrar patrones, para luego usarlo en predicciones futuras. Se requiere de una serie estacionaria para que el modelo funcione, las cual se obtendrá utilizando las herramientas anteriormente descritas.

Parte teórica

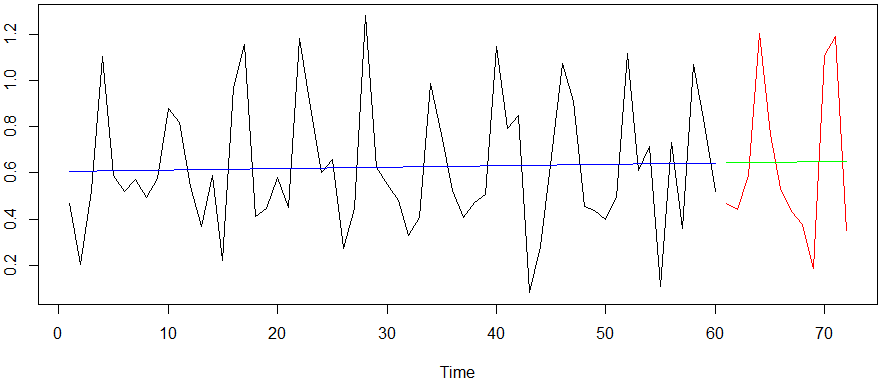
El objetivo de este apartado es analizar la serie que se presenta y generar un modelo ARIMA de predicción. En la serie temporal se almacena el número de ventas de un producto durante 6 años, guardados mes a mes. Por tanto, se puede deducir que existe un total de 72 registros de datos, los cuales representan los 12 meses del año de los 6 años nombrados.

Se dividirá la serie en dos partes, una de entrenamiento y otra de test. El primero poseerá 5 años de información (60 datos) y el segundo 1 año (12 datos). Además, de estos números se peude sacar una conclusión a priori: es altamente probable que la estacionalidad sea de un periodo de 12 (anual).



Los pasos a seguir son los siguientes: Comprobación y eliminación de la tendencia, comprobación y eliminación de la estacionalidad, transformación a serie estacionaria y, finalmente, búsqueda del mejor modelo ARIMA.

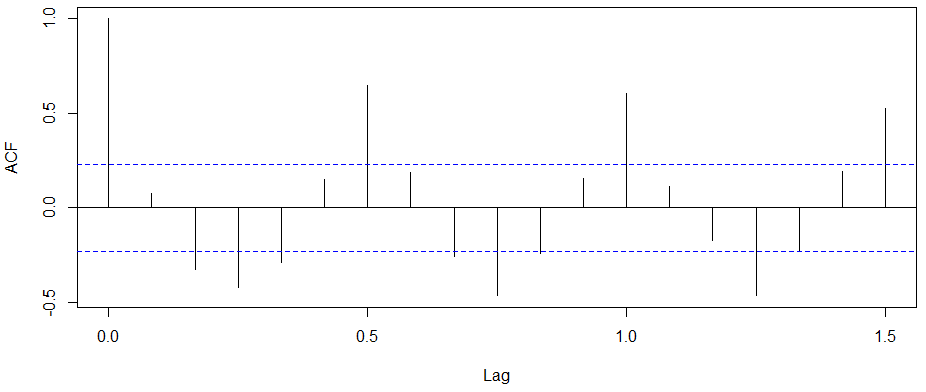
Primero, se procederá con la comprobación y la posible eliminación de la tendencia. Para ello se pondrá en el punto de mira la gráfica la serie. De este modo se determinará visualmente si existe alguna tendencia notable, lo cual es notable que no. Luego se aplicará un contraste de hipótesis en el cual se transformará la serie a un modelo lineal, observando así la pendiente de dicho modelo.



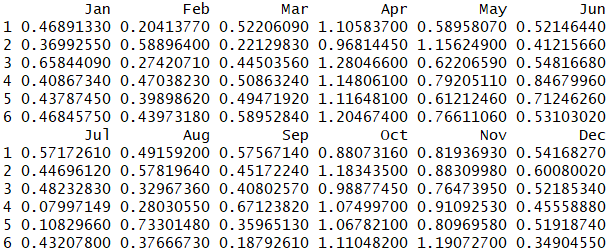
Se puede observar que no existe una tendencia clara. Para comprobar la fiabilidad de esta afirmación se puede realizar un test de Jarque-Bera sobre el error residual. De este modo se sabrá si el modelo lineal se ajusta correctamente a la serie. Los resultados arrojados para el entrenamiento y el test son los siguientes:

Luego, se procederá a eliminar la estacionalidad de la serie en caso de que existiese. Para comprobar su existencia se utilizará el método de ACF, el cual muestra una gráfica que nos indicará las variaciones de los datos. Si ésta muestra periodo sinusoidal significará que existe estacionalidad.

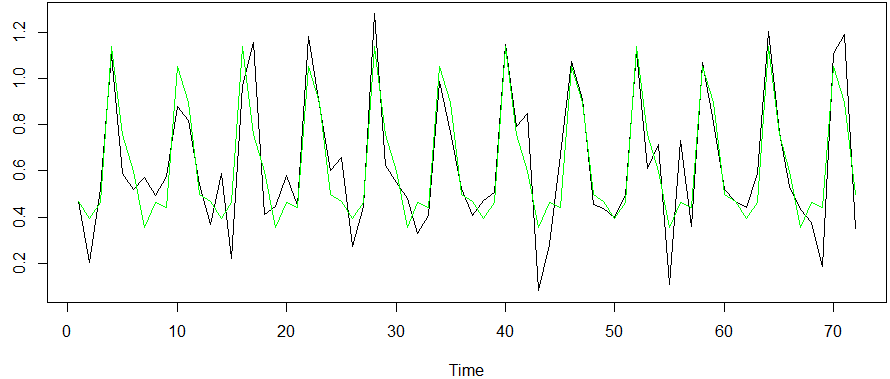


Se puede observar que existe dicho periodo sinusoidal. Ahora es necesario definir un tiempo de periodo. Como se comentaba previamente, existe una alta sospecha de que la estacionalidad posee un periodo de 12. Primero se observarán los datos con este periodo definido para comprobar su veracidad.

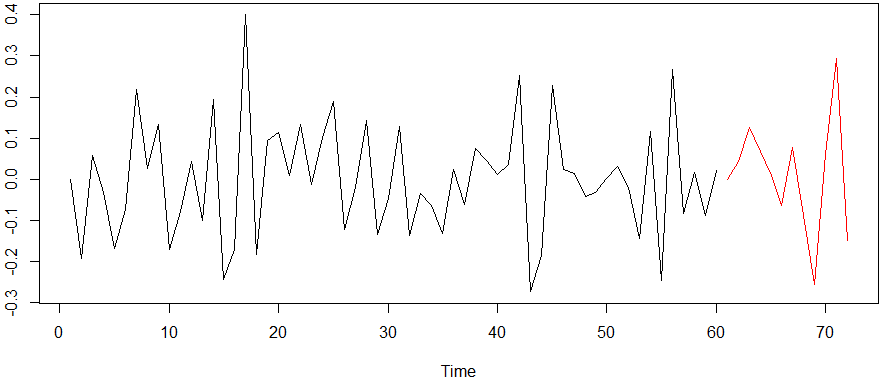


Tal y como se preveía, cada mes posee una varianza en sus datos baja, siendo todos estos similares año por año. Por tanto, queda definido como 12 el periodo de estacionalidad de la serie.

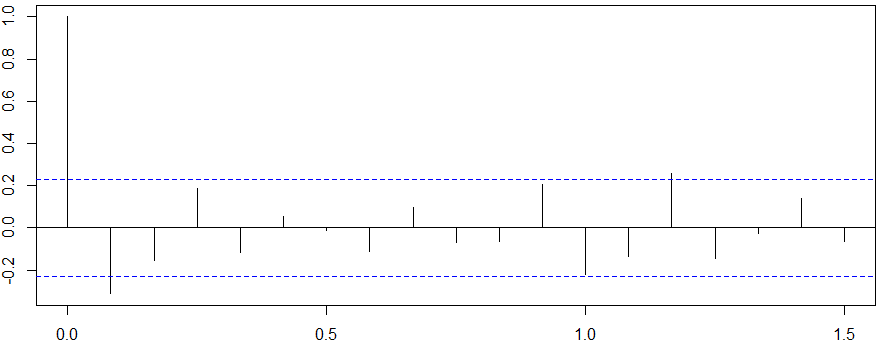
Al existir dicha estacionalidad, es necesario eliminarla. Para ello, se calculará la media de cada mes y luego se le restará dicha media a cada dato. Esta resta se aplicará tanto a los datos de entrenamiento como los de test. Comprobamos si la estacionalidad obtenida se ajusta a la serie temporal original:



Tras su posterior eliminación, la serie queda tal que así:

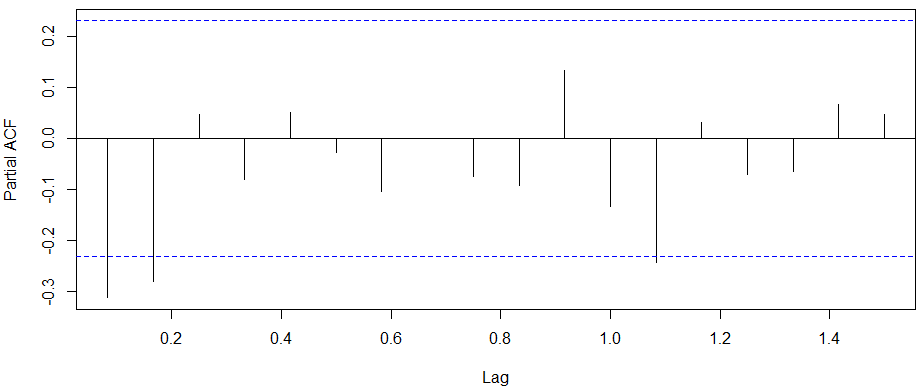


Por último, se comprobará si la serie es estacionaria, lo cual es requisito indispensable para que el modelo ARIMA funcione correctamente. Esto se confirma observando de nuevo el comportamiento de la serie nueva sobre la gráfica ACF. Para que sea estacionaria debe decrecer rápidamente y conservar ese estado.



Es estacionaria, por lo que ya se puede proceder a la selección del modelo ARIMA. Este modelo posee 2 parámetros a ajustar, el primero sólo se debe moldear en caso de que se haya aplicado diferenciación para hacer la serie estacionaria (en el paso anterior). Ya que esto no ha ocurrido, este parámetro se dejará a cero. Simplemente se procederá a modificar el segundo parámetro, que tiene que ver con el grado del método ARIMA.

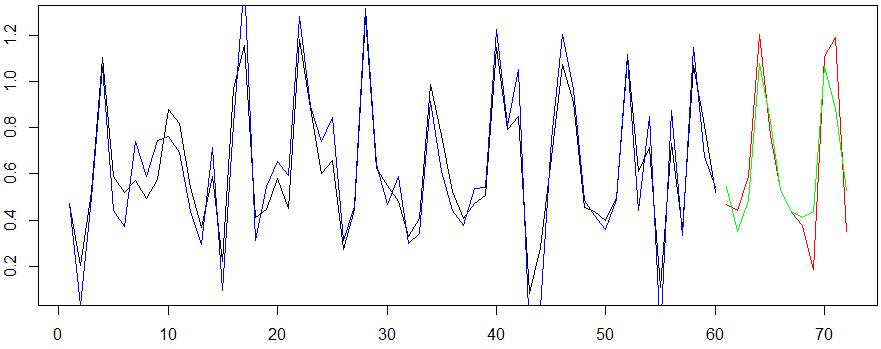
Para decidir sobre este valor, se observarán las gráficas ACF y PACF, en concreto, la segunda:



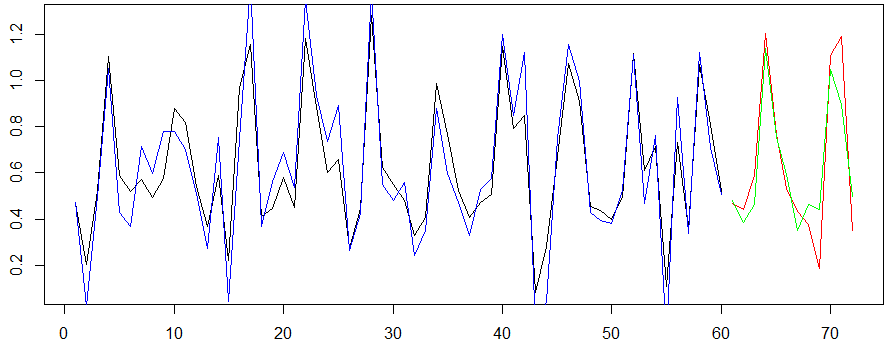
Se puede observar que existen dos primeros picos iniciales, que luego posiblemente se repite ese patrón 12 valores más adelante. Sin embargo, también es posible que el patrón sea de 2, ya que se cambia la dirección de los picos cada 2 valores en muchos de los casos.

Por tanto, los modelos escogidos son de grado 2 y de grado 12.

* ARIMA (12,0,0):
  + Error RSS en ajuste (entrenamiento) con ARIMA: 0.8123648
  + Error RSS en predicción (test) con ARIMA: 0.2357965
  + Error medio de predicción: 1.28226



* ARIMA (2,0,0):
  + Error RSS en ajuste (entrenamiento) con ARIMA: 0.9163813
  + Error RSS en predicción (test) con ARIMA: 0.2184429
  + Error medio de predicción: 1.258441



Se puede observar que las gráficas obtenidas y los errores son muy similares, por tanto, el ajuste de valores no hace que varíe mucho la tasa de acierto del modelo. Se puede concluir que en este caso no es necesario un alto preprocesamiento de los datos. Tampoco es necesario un modelo extremadamente complejo para poder predecir correctamente.