

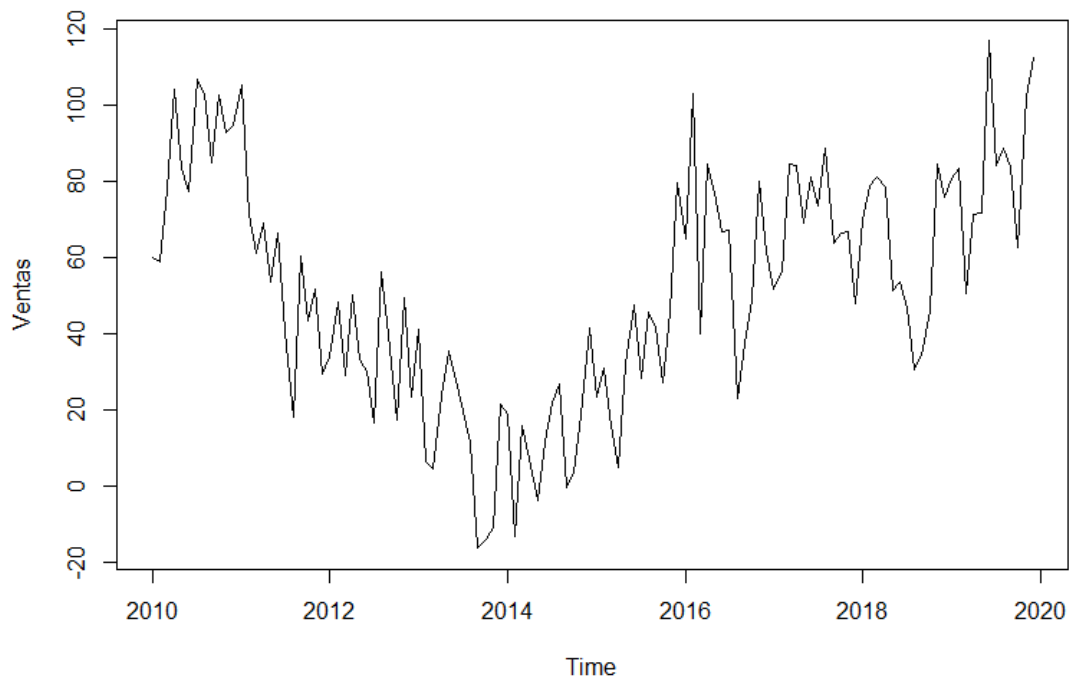
## SERIES TEMPORALES\_TIENDA DE ROPA

**Autora: Diana Cristina Rojas Romaní**

### **Análisis del código en R**

#### **1. Características estadísticas generales**

Se procede a nombrar la serie temporal con la terminación “.ts” y se grafica con el eje y: Ventas y eje x: Time. En la Gráfica 1 se observa una tendencia creciente de las ventas a través de los años y mayor dispersión de las ventas en los últimos años. Además, es probable que no sea una serie estacional debido a que los patrones no son repetitivos con el pico y el descenso de este; a pesar de ello, es necesario realizar más análisis que determinen ello.



Gráfica 1. Serie temporal de las ventas desde 2010 al 2019

#### **2. Descomposición de la serie temporal**

Se observa que en el primer componente del Gráfico 2, “observed”, un descenso en las ventas luego un ascenso gradual hasta el 2019; y tiene picos no uniformes entre ellos.

En el segundo componente del gráfico, “trend”, muestra primero una caída en el 2014 y luego una tendencia de aumento de ventas hasta el 2019.

Además, en cuando a “seasonal”, se observa picos repetitivos y similares entre sí por lo que se puede afirmar que es una serie de tiempo estacional.

En cuanto al último componente de la gráfica, “random”, se muestra el componente residual o ruido el cual representa la data cuyas fluctuaciones no son explicadas ni por la tendencia ni por la estacionalidad y están en el rango de 30 y -30.

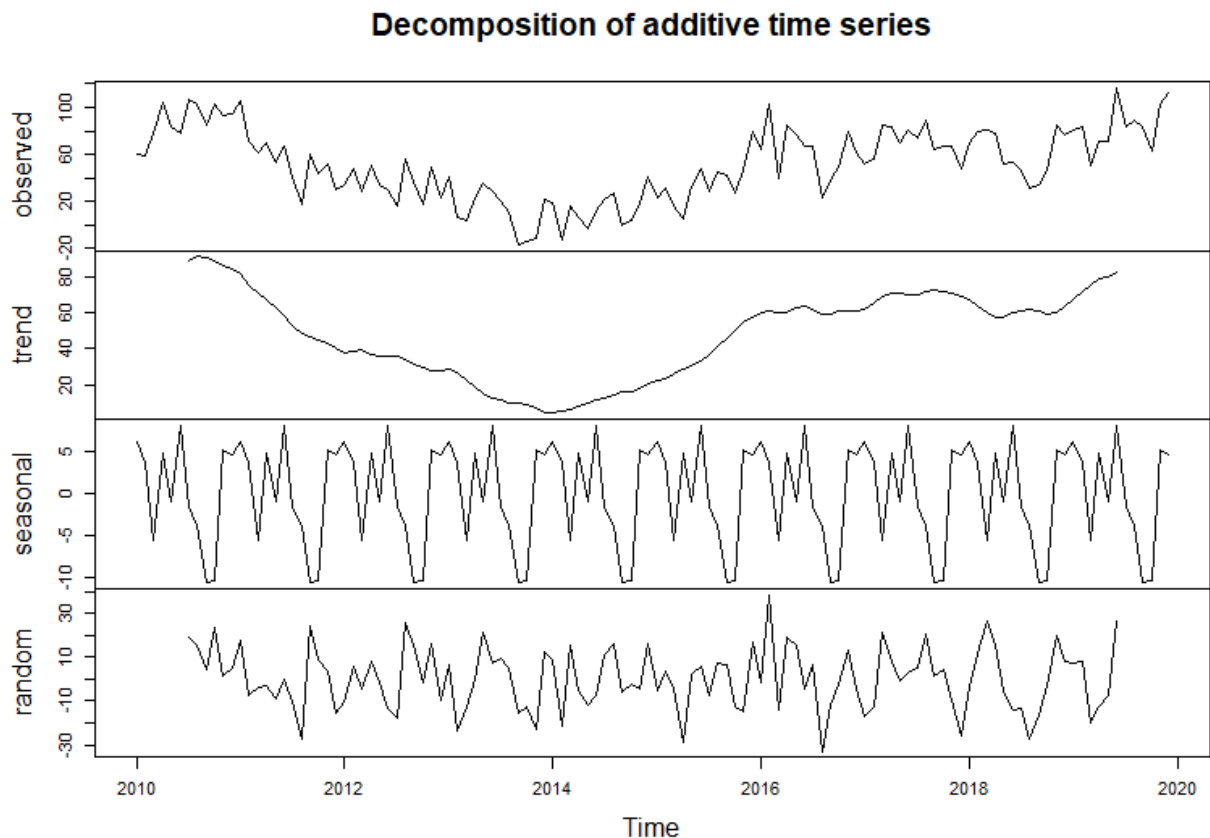


Gráfico 2. Descomposición de la serie temporal

### 3. Evaluación de estacionariedad - Test de Dickey-Fuller

Esta prueba indica la estacionariedad de una serie de tiempo, se observa que el valor de p-value es 0.7635 lo cual es mayor a 0.05 e indica que la serie no es estacionaria por lo que se tendrá que realizar la diferenciación.

#### Augmented Dickey-Fuller Test

```
data: ventas_mensual.ts  
Dickey-Fuller = -1.5499, Lag order = 4, p-value = 0.7635  
alternative hypothesis: stationary
```

Hasta ahora se conoce que la serie es estacional y no estacionaria, en los siguientes ítems se analizará ello para buscar el mejor modelo de serie temporal.

#### 4. Diferenciaciones

Luego de aplicar la diferenciación a la serie se grafica la serie temporal y se observa un comportamiento constante de la varianza y la media. La mayoría de los picos fluctúan entre -40 y 40 lo cual es constante. Sin embargo, en el año 2016 hay un pico que excedió dicho intervalo, pero ya que solo fue un pico aún así se puede considerar a este modelo como estacionario. Es decir, ya tenemos una serie estacionaria a la cual ya podemos aplicar los modelos ARIMA.

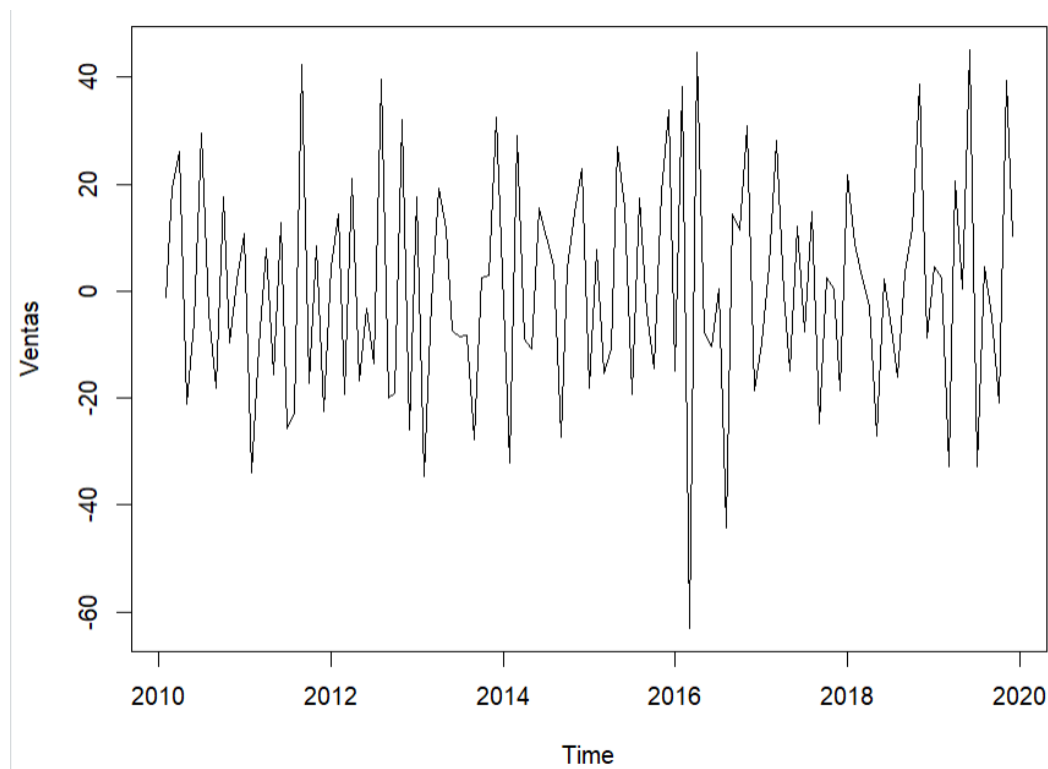


Gráfico 3. Serie temporal diferenciada

También se puede corroborar si la serie es estacionaria o no realizando de nuevo el Test de Dickey-Fuller.

#### Augmented Dickey-Fuller Test

```
data: ventasdiff
Dickey-Fuller = -6.5693, Lag order = 4, p-value =
0.01
alternative hypothesis: stationary
```

Luego de realizar el test, se observa un p-value (0.01) menor a 0.05 lo cual indica que la serie sí es estacionaria.

## 5. Análisis de autocorrelación

En la gráfica 4 se observa la correlación de la serie con los lags en la cual se tiene una disminución lenta de la autocorrelación a medida que aumenta los lags. Esto indicaría que la serie tiene al menos un componente estacional, y por esa disminución un poco lenta se podría señalar una posible no estacionariedad.

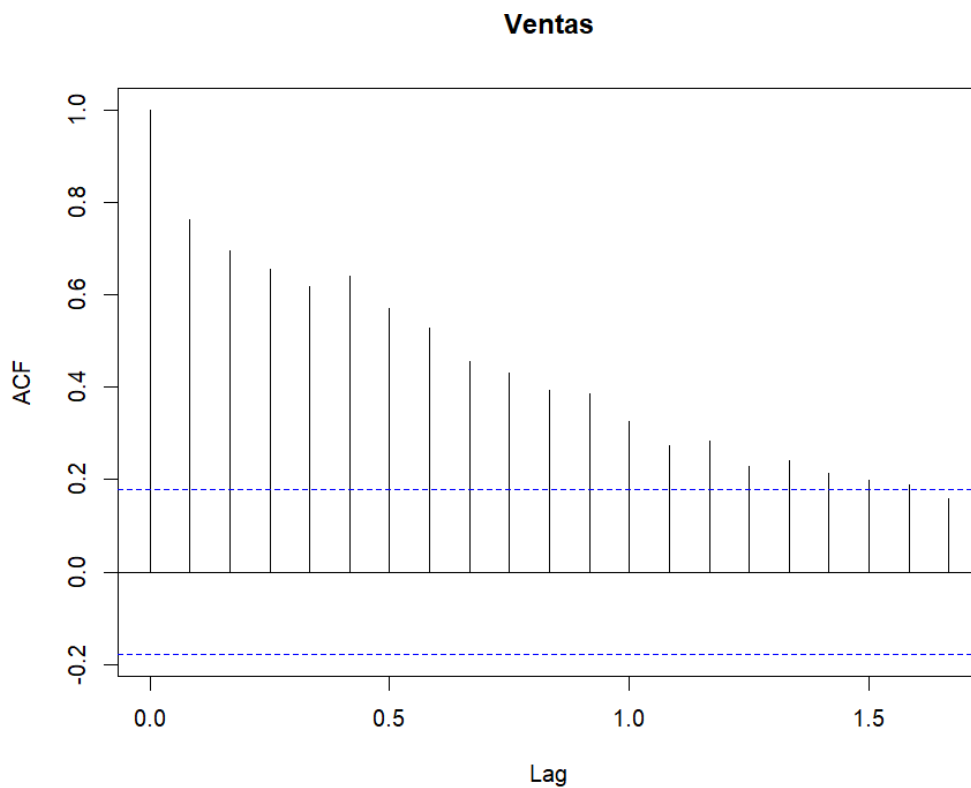


Gráfico 4. Función de autocorrelación ACF (modelo autorregresivo AR)

En la gráfica 5 se observa la autocorrelación parcial en la cual se observa 2 lags y casi todas las verticales están dentro de las líneas puntuadas de color azul. Es decir, hay un cierto corte brusco en los rezagos ya que se observa que el primer vertical es mucho más alto que las demás por lo que se podría evaluar un modelo con  $AR=1$ . Aún así, también se puede evaluar con otro orden ya que hay un segundo vertical entre 0.2 y 0.4 lo cual podría sugerir un  $AR=2$ .

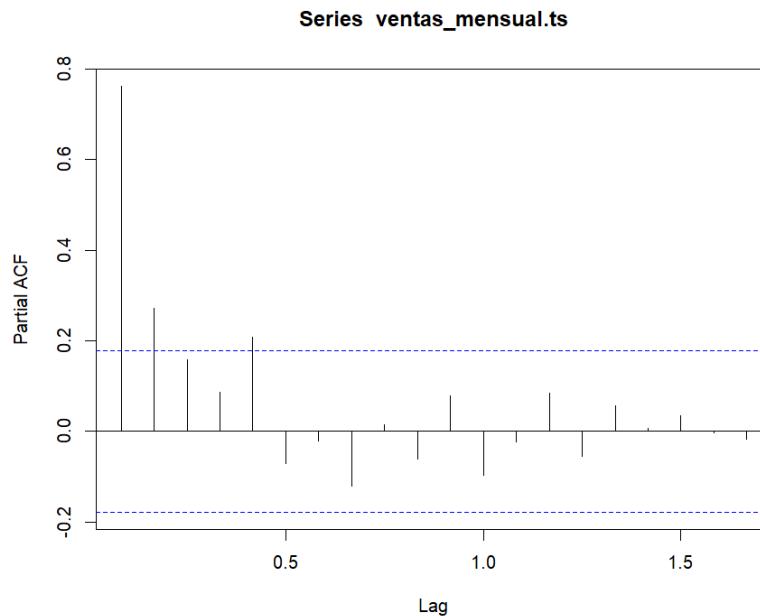


Gráfico 5. Función de autocorrelación parcial

## 6. Determinación del orden autorregresivo y medias móviles

Luego de realizar una diferenciación, se graficó el ACF Y ACF parcial para determinar el orden de ambos. De la gráfica 6 se tiene dos verticales significativas y la tercera ya se encuentra dentro del intervalo por lo que se indica el valor  $q = 2$ .

Por su parte, de la gráfica 7 se tiene solo una vertical significativa y la segunda ya está dentro del intervalo por lo que se considera un valor  $p = 1$ .

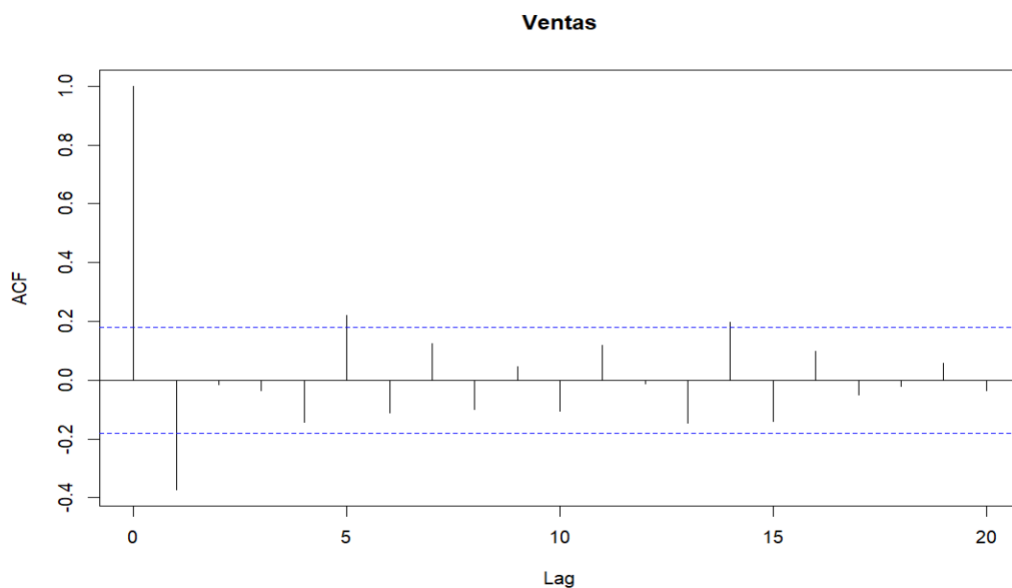


Gráfico 6. Función de autocorrelación ACF luego de la diferenciación

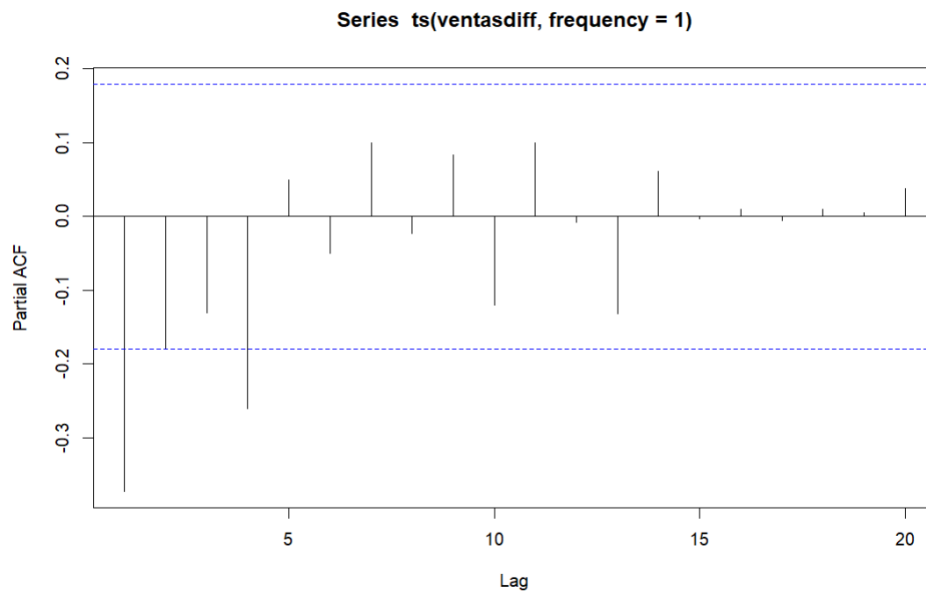


Gráfico 7. Función de autocorrelación parcial luego de la diferenciación

Con estos valores de  $p$ ,  $d$  y  $q$  tenemos los componentes no estacionales, sin embargo, una mejor forma de hallar un mejor modelo SARIMA con los 6 parámetros y el menor AKAIKE es usando una función `auto.arima` ya que esta contempla en su interior un análisis más completo con los 6 parámetros dándonos un mejor modelo.

## 7. Aplicación del modelo

Antes de aplicar el modelo SARIMA adecuado, observamos que la serie temporal tiene varianzas muy diferentes al igual que sus medias (Gráfico 1). Por ello se hace uso del logaritmo a la serie temporal el cual busca estabilizar la varianza y con ello mejorar el modelo. Además, esto ayuda bastante al modelo ya que los modelos SARIMA ya suponen que la serie tenga una varianza constante durante el tiempo medido.

Entonces, se usa la función `auto.arima` para hallar el mejor modelo que incluye los parámetros más adecuados a fin de obtener una predicción eficiente. En la siguiente imagen se muestra los resultados del Summary del modelo aplicado en R.

```

> summary(modelo_sarima)
Series: log(ventas_mensual.ts)
ARIMA(2,1,1)(0,0,1)[12]

Coefficients:
      ar1      ar2      ma1      sma1
    -0.0563 -0.2899 -0.6109  0.1410
s.e.    0.1458  0.1148  0.1419  0.0942

sigma^2 = 8.791: log likelihood = -296.73
AIC=603.46 AICc=604 BIC=617.36

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.01061095 2.902541 1.297262 21.67481 33.23949 0.6837977 0.007727035

```

En efecto, en el resumen del modelo indica un SARIMA tal como ya se estuvo mencionando en los previos análisis. Este modelo contempla los valores  $p=2$ ,  $d=1$ ,  $q=1$ ,  $P=0$ ,  $D=0$  y  $Q=1$ ; el modelo tiene una parte regular y otra estacional. En la parte regular contiene una parte autorregresiva de orden 2, 1 diferenciación y en media móvil es de orden 1. En cuanto a la parte estacional contempla solo un término de media móvil; no cuenta con términos autorregresivos ni de diferenciación.

Al costado de los parámetros explicados se halla entre corchetes el número 12 el cual indica la frecuencia dada que es de 12 meses por año.

En la siguiente línea se tiene la sección de coeficientes, se tiene los dos términos autorregresivos (ar1 y ar2), una media móvil regular (ma1) y una media móvil estacional (sma1). Se observa que todos a excepción de la media móvil estacionario son negativos lo cual sugiere que estos tendrán un impacto negativo en el valor. Luego se tiene las siglas s.e. el cual indica el valor del error estándar para cada coeficiente.

Lo siguiente a explicar es la  $\sigma^2 = 8.791$  el cual es la varianza residual del modelo e indica la dispersión de los errores. Además, a su lado se halla el Log likelihood el cual es un valor bajo que indicaría la probabilidad del modelo en el conjunto de datos.

Después se tiene un AIC bajo el cual indicaría también un posible mejor modelo. Al igual que el AIC, los valores de AICc y BIC son otros criterios de información cuyos valores bajos nos otorgan un mejor modelo penalizando por la complejidad.

En la siguiente sección se indica las medidas de error del conjunto de datos de entrenamiento. Se observa que el error medio es muy bajo lo cual es adecuado ya que significa que el modelo no tiene un sesgo significativo. Para las otras medidas se tiene la misma concepción de que mientras menores valores se obtenga es mejor el modelo; en el caso de ACF este indica la autocorrelación en el lag 1 de los residuos y en nuestro modelo resultó con 0.0077 el cual es un valor muy cercano a 0 e indica una clara no correlación significativa en los residuos.

Por ello se puede considerar el modelo con los parámetros dados como un modelo óptimo. Sin embargo, hace falta ver la funcionalidad de este por lo que en la siguiente parte se realizará un pronóstico para los siguientes 12 meses.

## 8. Ajuste del modelo SARIMA

Al graficar la autocorrelación de los residuos (Gráfica 8) se observa que tiene una vertical “significativa” pero ya que está en el lag 0 esto es común hallar porque cada valor está estrictamente correlacionado consigo mismo. Luego por el lag 0.5 se halla dos verticales que sobresalen ligeramente del intervalo de confianza; por ser ligero el desborde se considera que no hay una autocorrelación fuerte entre ellos.

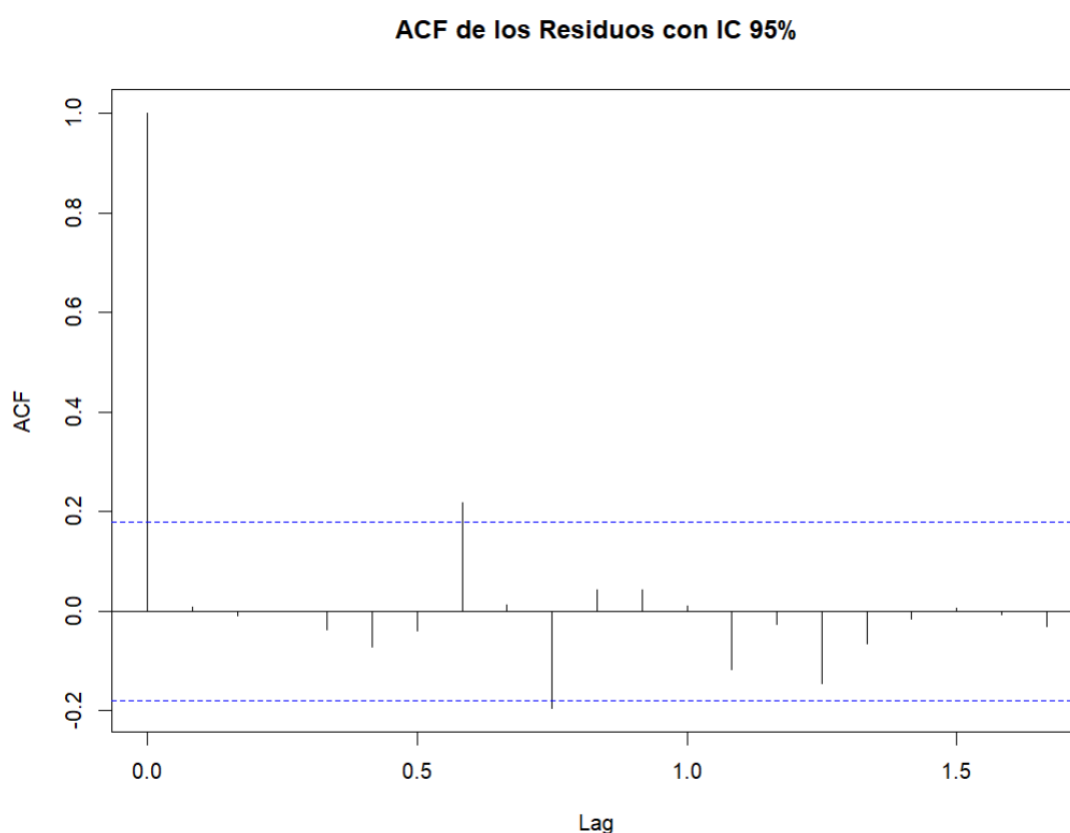


Gráfico 8. Función de autocorrelación ACF de los residuos

En la gráfica 9 se observa que los residuos no muestran autocorrelación parcial que sea significativa lo cual es un buen indicador que sugiere que el modelo está incluyendo las principales dependencias de los datos. A pesar de ver que dos verticales se excluyen ligeramente de los intervalos se puede considerar que es una autocorrelación débil ya que se encuentran en los lags menores que es hasta antes del lag 1. Luego, en adelante se observa que ya ningún vertical sobresale los intervalos por lo que deduce que no están correlacionadas de manera parcial fuerte.



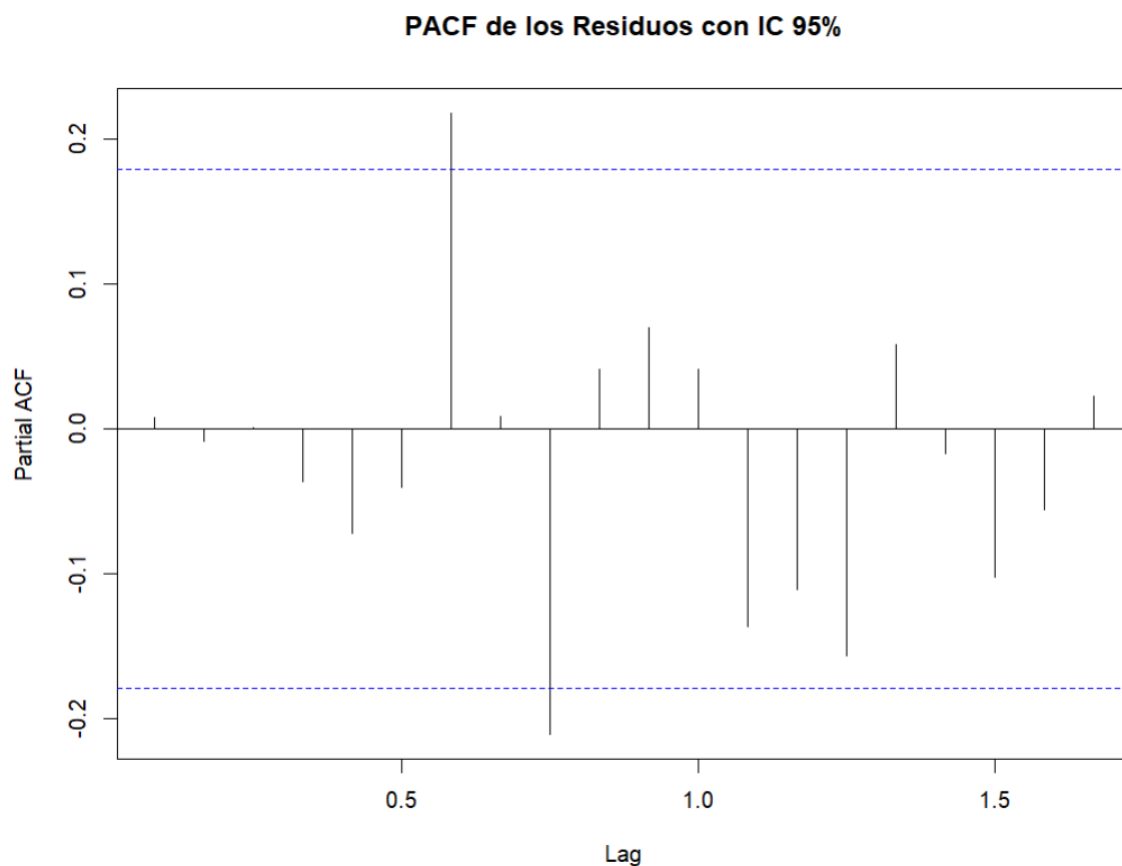


Gráfico 9. Función de autocorrelación parcial de los residuos

### 9. Pronóstico del modelo

Se observa (Gráfico 10) la predicción de ventas en el año 2020 a partir de aplicar el modelo SARIMA con los parámetros ya indicados anteriormente. Se observa que alrededor de años impares hay dos o más picos que descienden y al llegar a un año par hay un ligero incremento de las ventas. La predicción de ventas del 2020 está entre 80 y 100 lo cual está dentro de las variaciones que se observa en ventas de años pasado. Además, en la Gráfica 11 se observa que esta predicción se encuentra dentro del intervalo de confianza( líneas azules).

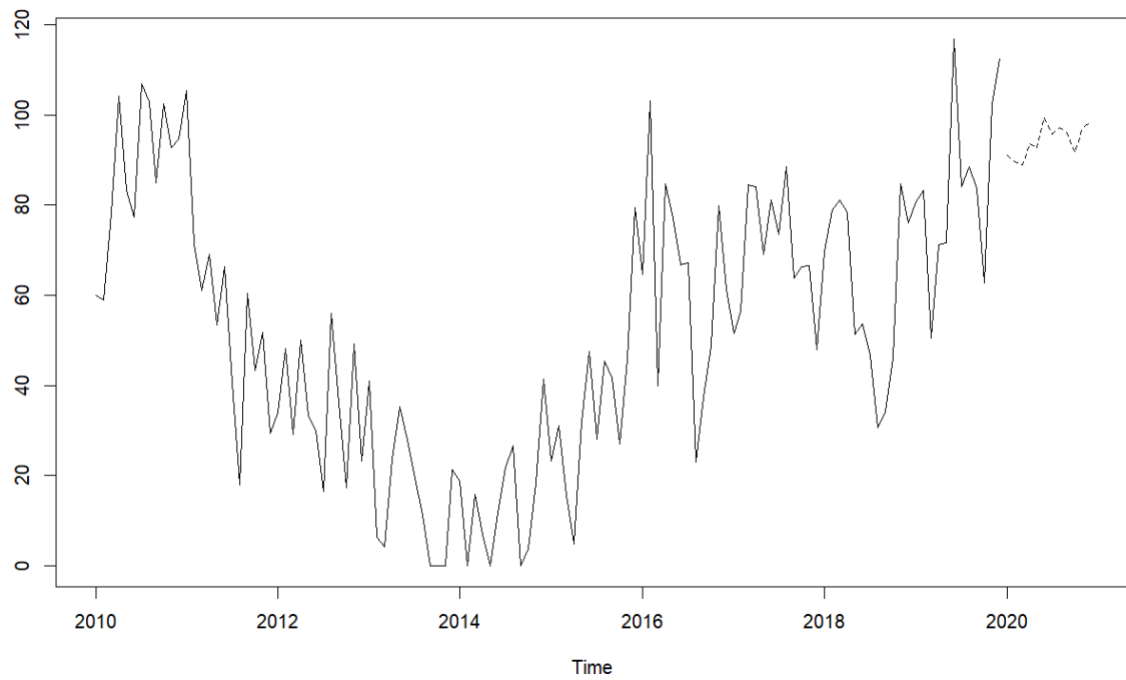


Gráfico 10. Pronóstico del modelo SARIMA

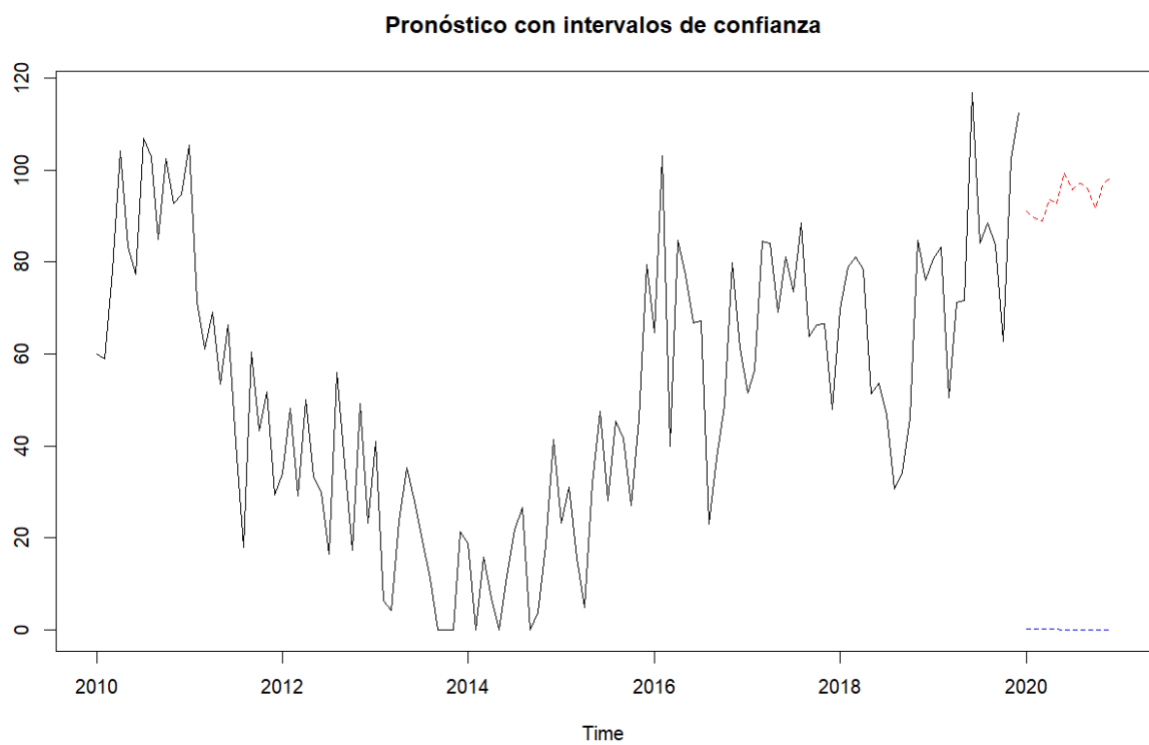


Gráfico 11. Pronóstico del modelo SARIMA con los intervalos de confianza