LICENCIATURA EN ESTADÍSTICA

Trabajo Práctico

"Evaluación de las Tendencias de Ventas Mensuales en una Empresa de Logística Mediante Series de Tiempo"

Autor: Santini, Franco

Docentes: Mendez, Fernanda - Sigal, Facundo

2024

Tabla de contenidos

Introducción

En este trabajo se busca estudiar las unidades vendidas (bebidas con y sin alcohol) en Hectolitro[1] (HL) de una importante empresa de logística en la ciudad de Casilda, Santa Fe. El estudio se realiza con 80 observaciones, medidas mensualmente en el período (01/01/2018 - 31/08/2024).

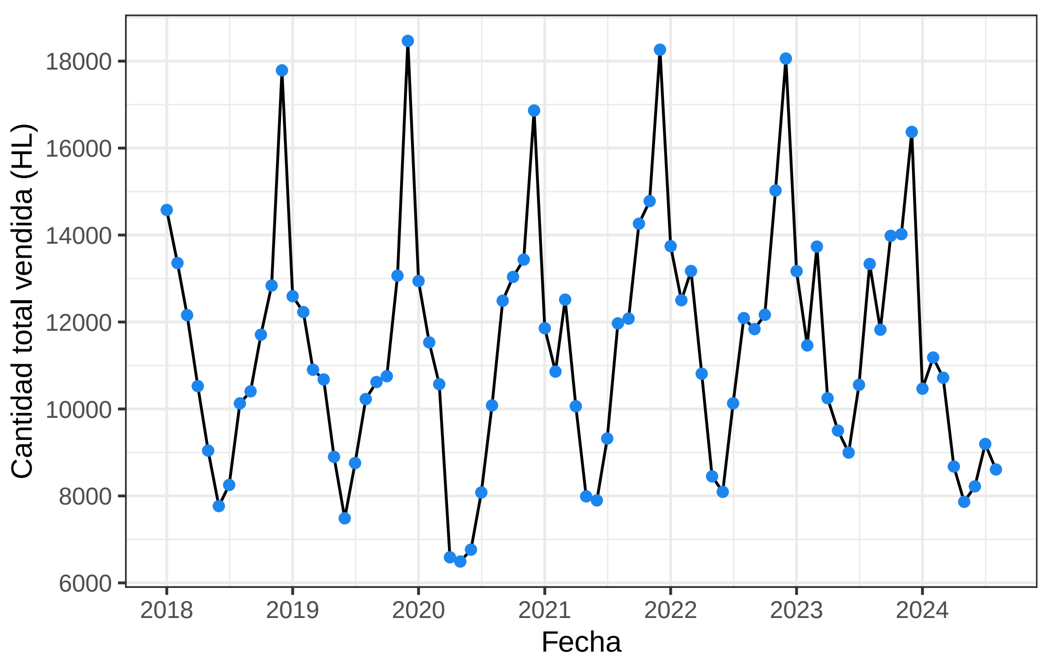
Cabe aclarar que en el período de estudio, atravesamos un evento muy catastrófico en todas las áreas, particularmente en el sector de logística, la pandemia del COVID-19. Al afectar principalmente a las personas, Argentina implementó el Aislamiento Social, Preventivo y Obligatorio (ASPO) como medida para combatir la pandemia de COVID-19. Debido a esta medida, a fines de marzo del 2020 - principios de abril del 2020, el sector de logística fue afectado y en general se produjo una caída en las ventas con respecto al año anterior.

Teniendo esto en cuenta, resulta de interés pronosticar a futuro las unidades vendidas en (HL) de la empresa, mediante modelos de *series de tiempo*.

Análisis descriptivo

En primera instancia, comenzamos con el gráfico de la serie para observar el comportamiento de las ventas a través del período en estudio.

Figura 1: Unidades vendidas mensuales en (HL)

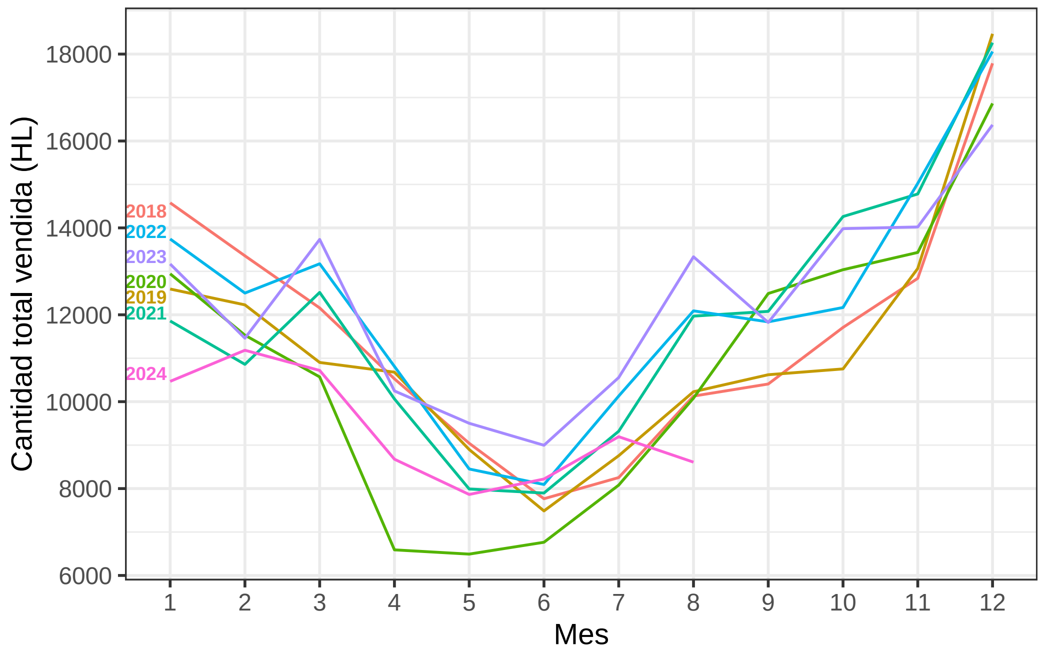


En la Figura 1 se puede observar un comportamiento estacional de las unidades vendidas, disminuye considerablemente en los meses (mayo, junio y julio) y luego aumenta considerablemente en los meses

[1]Unidad de volúmen equivalente a 100 litros.

(noviembre - diciembre - enero) en casi todos los años, exceptuando el año 2020 que puede deberse a la pandemia.

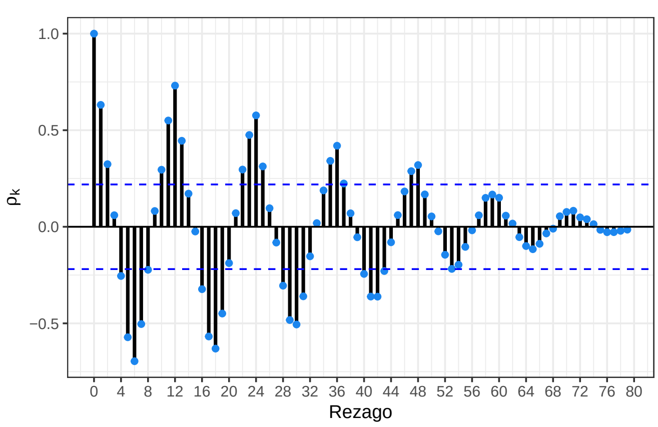
Figura 2: Comportamiento anual de las unidades vendidas en (HL)



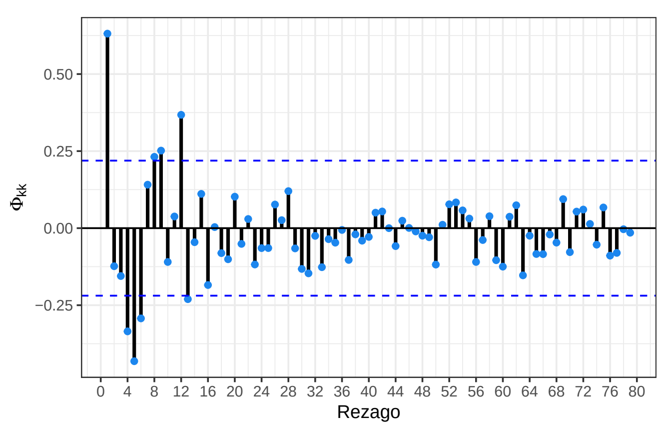
La Figura 2 podemos observar que claramente hay un comportamiento estacional de las unidades vendidas, pero se descarta cualquier tipo de tendencia creciente o decreciente de las mismas, dado que no se observa ni un crecimiento, ni un decrecimiento año a año de las unidades vendidas mensuales.

Figura 3: Funciones de autocorrelación muestral

(a) FACM



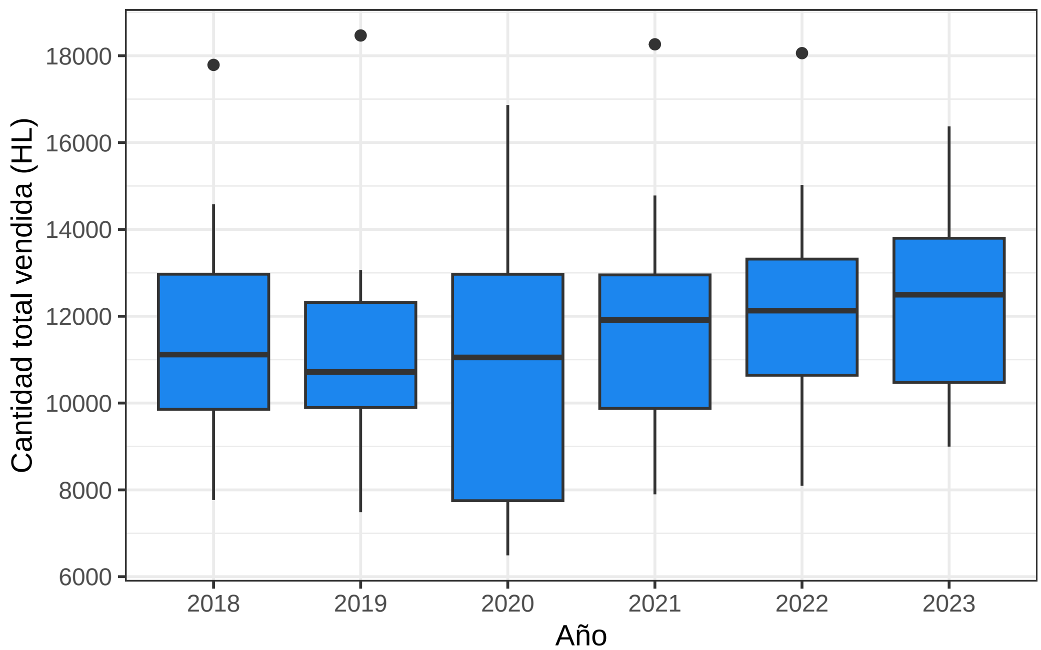
(b) FACPM



De acuerdo con lo mostrado en Figura 3, se puede observar que la FACM tiene un decrecimiento lento y sinusoidal, notando también que en los rezagos 12, 24, 36 y 48 son significativos esto da un indicio de que la serie tiene un comportamiento estacional, mientras que la FACPM no parece tener un patrón claro.

Analisis de dispersion

Figura 4: Boxplots de las unidades vendidas en (HL) por año



Se puede observar en la Figura 4 que la variabilidad no parece ser constante año a año, por lo que puede estar dando una advertencia de que la serie no es estacionaria en variancia, además se puede apreciar la presencia de outliers en los años 2018, 2019, 2021 y 2022. Se excluyó el año 2024 de la Figura 4 porque no estaban los datos completos.

Una vez observados los Boxplots, procederemos a realizar la transformación de Box-Cox. Dependiendo del valor de λ obtenido, determinaremos si es necesario transformar la serie o si puede mantenerse en su forma original. Además, puede ser de utilidad calcular el coeficiente de variación según los valores de λ arrojados por la transformación de Box-Cox, de esta manera, podremos identificar si una transformación de potencia es adecuada para estabilizar la variancia de la serie. Aquel valor de λ para el cuál se minimice el coeficiente de variación, será la transformación que tendremos que realizar.

Tabla 1: Coeficiente de variación

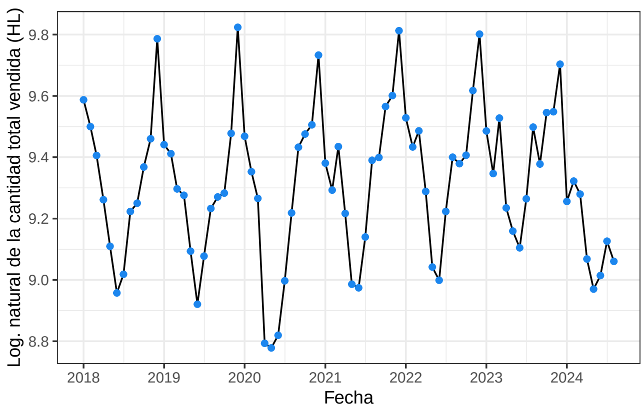
λ	Coeficiente de variación
-2.0	0.5001
-1.0	0.2433
-0.5	0.1206
0.0	0.0258
0.5	0.1198
1.0	0.2405
2.0	0.4909

Observando la Tabla 1, el coeficiente de variación mínimo esta asociado a un $\lambda = 0$, por lo que es apropiado aplicar la transformacion $y^{(\lambda)} = ln(y)$.

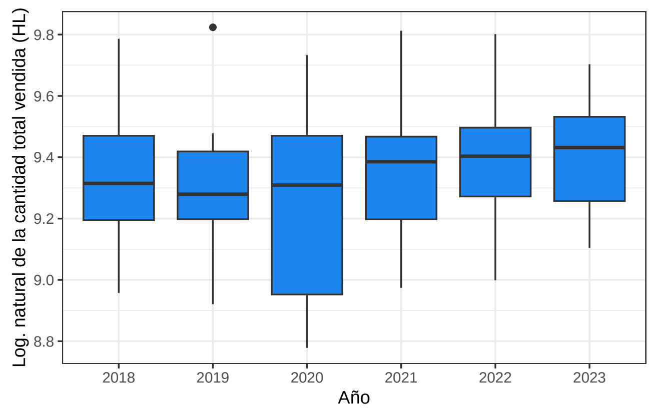
Identificación del modelo

Figura 5: Resúmen de la Serie transformada

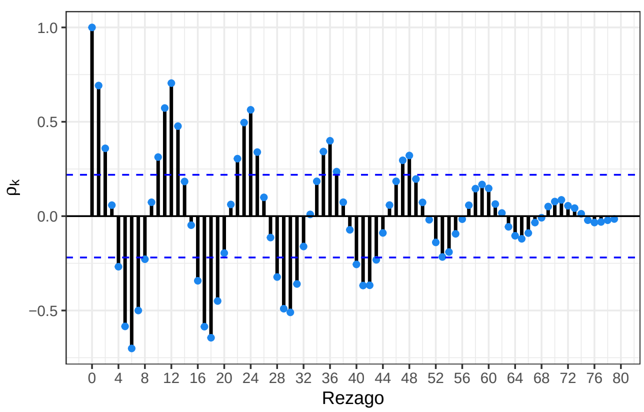
(a) Log. natural de las unidades vendidas mensuales en (HL)



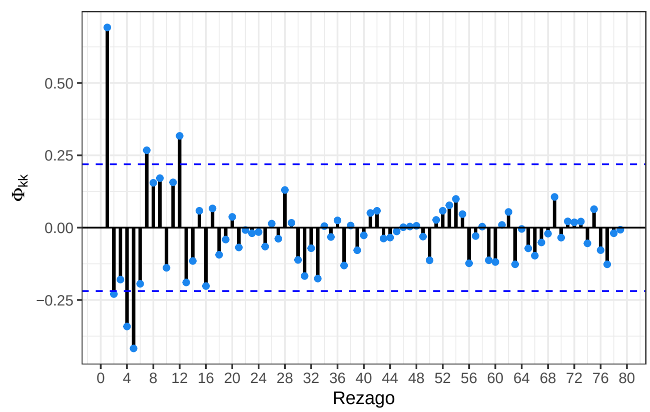
(b) Boxplots del Log. natural de las unidades vendidas en (HL) por año



(c) FACM



(d) FACPM



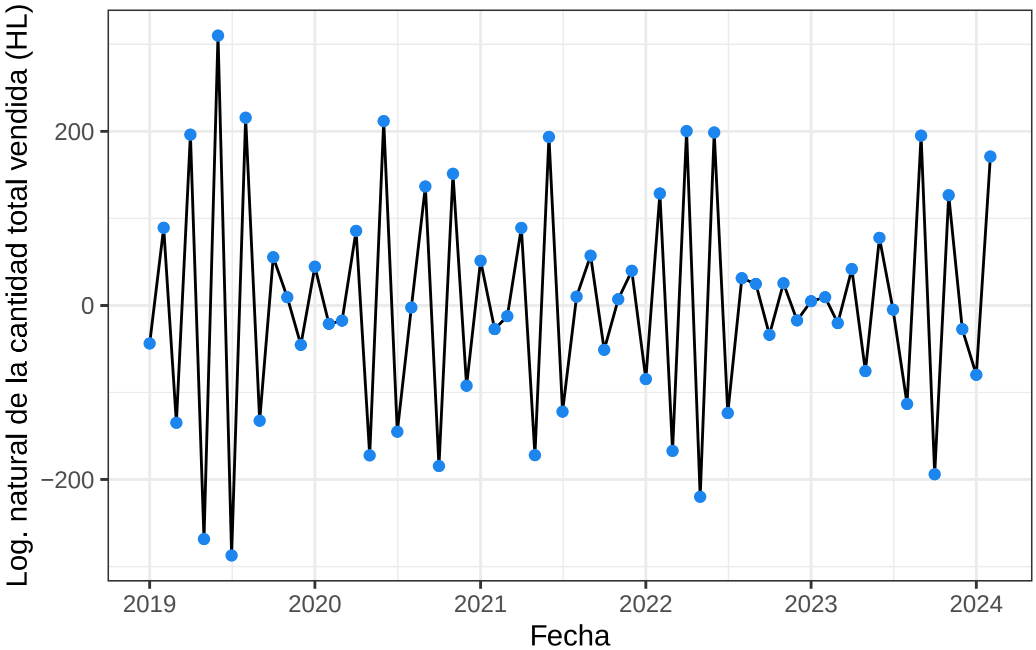
Luego de transformar a la serie, se realizó la Figura 5 que contiene un resúmen acerca del comportamiento de la serie. Notando, que ahora en la Figura (5b) se puede observar que la variabilidad parece ser un poco más constante año a año.

Diferenciación de la serie

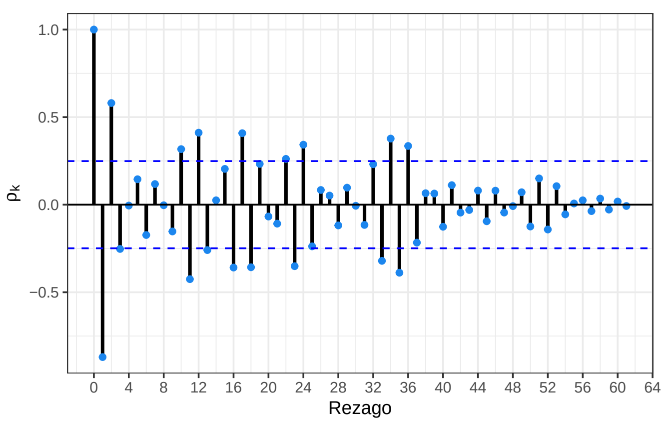
Una vez realizado el análisis descriptivo, se notó el comportamiento estacional de la serie, por lo qué, se realiza una diferenciación en la parte estacional de la misma, dado que lo rezagos estacionales en la FACM observada en la Figura 5 parecen decrecer lentamente de forma lineal, pasando a trabajar con 62 observaciones.

Figura 6: Resúmen de la Serie diferenciada en la parte estacional

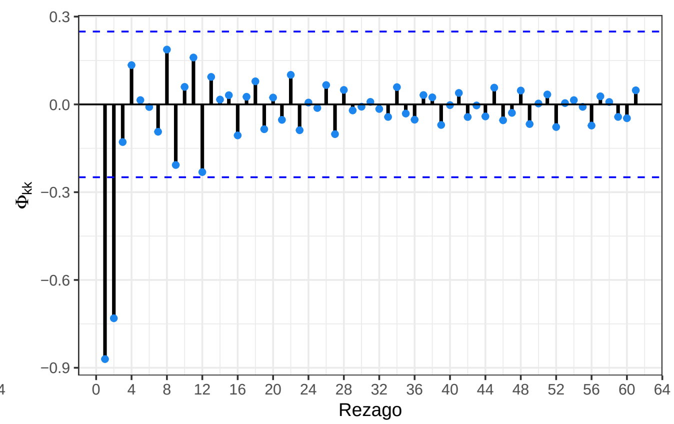
(a) Log. natural de las unidades vendidas mensuales en (HL)



(b) FACM



(c) FACPM



En la Figura 6 se observa, el gráfico de la serie diferenciada en la parte estacional y sus respectivas funciones de autocorrelación muestral y de autocorrelación parcial muestral. En la Figura (6a) se puede observar que la serie parece ser estacionaria, dado que varia de manera oscilante sobre un valor constante. Luego para realizar la identificación de los posibles modelos que se pueden ajustar a esta serie, utilizaremos las Figuras (6b) y (6c). En el gráfico de la FACM se puede observar que 2 rezagos son significativos y el tercer rezago está en el límite de ser significativo, viendo la parte estacional parece que los rezagos 12, 24 y 36 son significativos, persentando un decrecimiento exponencial. En el gráfico de la FACPM se puede observar que solo los primeros 2 rezagos son significativos, y el rezago 12 correspondiente a la parte estacional, parece estar al límite de ser significativo.

Por lo que teniendo esto en cuenta, planteamos los siguientes modelos que se pueden ajustar a nuestra serie:

- Modelo 1: $SARIMA(0,0,3)(1,1,0)_{12}$
- Modelo 2: $SARIMA(0,0,2)(1,1,0)_{12}$
- Modelo 3: $SARIMA(2,0,0)(1,1,0)_{12}$
- Modelo 4: $SARIMA(1,0,0)(0,1,1)_{12}$, selección del modelo de forma automática

Una vez planteado los modelos que vamos a utilizar para el pronóstico de la serie, se procede a estimarlos y compararlos.

Tabla 2: Comparación de los modelos

Modelos	$\hat{\sigma}^2$	Log_veros.	AIC	AICc	BIC
$SARIMA(1,0,0)(0,1,1)_{12}$	0.0087	57.1397	-108.2794	-107.8656	-101.8980
$SARIMA(2,0,0)(1,1,0)_{12}$	0.0100	55.0957	-102.1914	-101.4897	-93.6829
$SARIMA(0,0,2)(1,1,0)_{12}$	0.0104	53.9543	-99.9087	-99.2069	-91.4001
$SARIMA(0,0,3)(1,1,0)_{12}$	0.0104	54.5016	-99.0032	-97.9318	-88.3676

Se puede observar, que los modelos con menor AIC son:

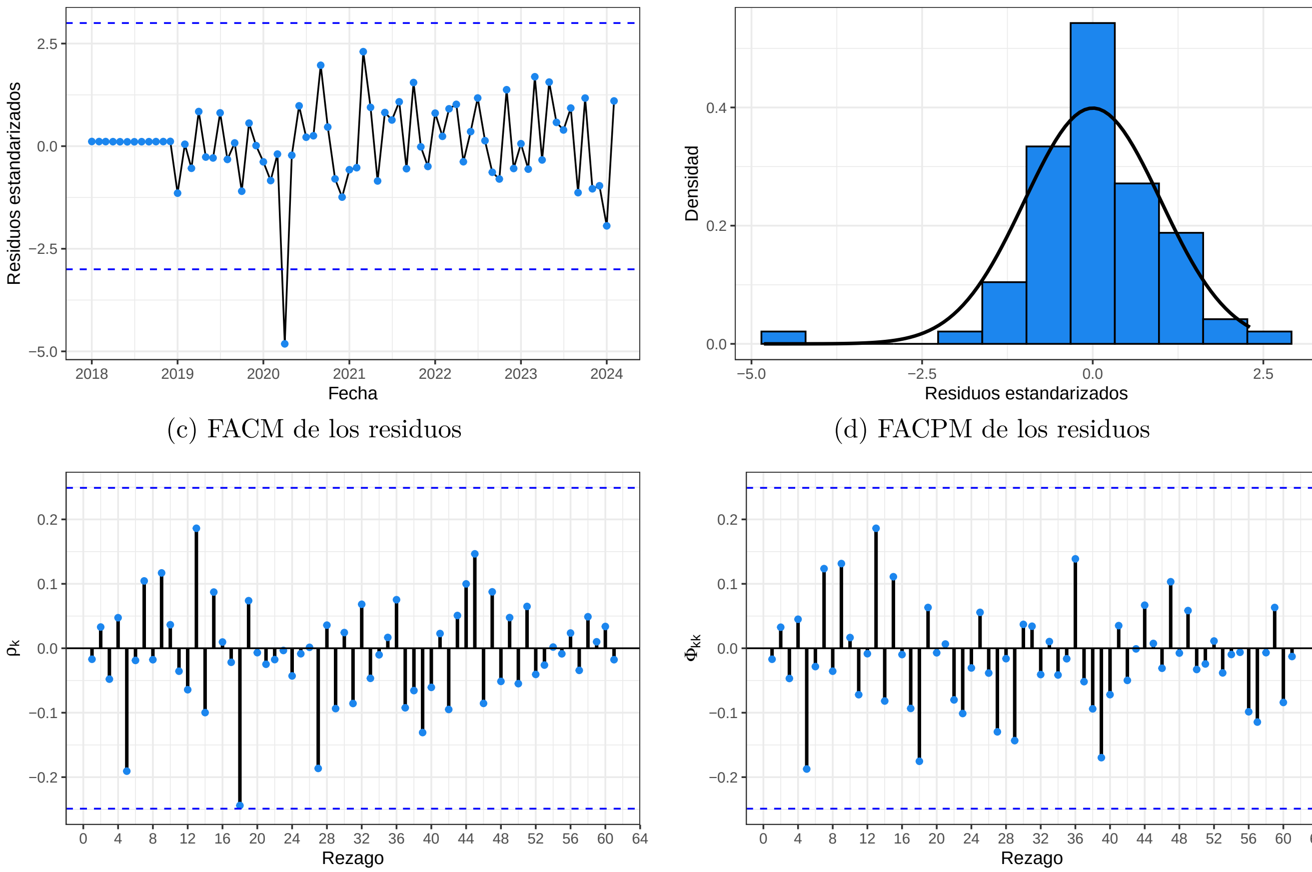
- Modelo 3 $SARIMA(2,0,0)(1,1,0)_{12}$
- Modelo 4 $SARIMA(1,0,0)(0,1,1)_{12}$

El modelo 3 tiene 3 parámetros de los cuáles ϕ_2 no es significativo, se puede observar esto en la tabla 4 del anexo. Por esta razón, se elimina un parámetro en la parte regular del modelo. Así, el Modelo 3 se reemplaza por un modelo $SARIMA(1,0,0)(1,1,0)_{12}$. Con respecto al modelo 4, sus dos parámetros son significativos, teniendo además una estimación de $\hat{\sigma}^2$ menor que la del modelo 3. Del mismo modo, continuaré trabajando con ambos modelos y decidiré cuál utilizar para la predicción, eligiendo aquel que demuestre una mejor capacidad predictiva, teniendo en cuenta la verificación de los supuestos de los mismos.

Evaluación de los supuestos de los modelos

Esta instancia se centra en la verificación de los supuestos de ambos modelos mencionados anteriormente. Se hará un análisis exhaustivo para el modelo $SARIMA(1,0,0)(0,1,1)_{12}$, mientras que para el modelo $SARIMA(1,0,0)(1,1,0)_{12}$, al ser un procedimiento análogo, se dejarán en el anexo, en la Figura 9.

Figura 7: Análisis de residuos $SARIMA(1,0,0)(0,1,1)_{12}$



(a) Residuos estandarizados vs fecha

(b) Histograma de los residuos estandarizados

(c) FACM de los residuos

(d) FACPM de los residuos

En la parte (a) y (b) de la Figura 7 se puede observar que hay una observación atípica, correspondiente al año 2020, lo cuál se puede deber a alguna situación inusual ocurrida en pandemia, además los residuos varían alrededor del cero y parecen tener una variancia constante. También se puede apreciar, en la Figura (7b) que los residuos parecen ser normales, obviamente sin incluir al outlier. Luego en las Figuras (7c) y (7d), da la impresión de que los residuos son incorrelados, en la tabla 5 que se encuentra en el anexo, se incluye el test de Ljung-Box para comprobar que los residuos son incorrelados. Por lo que podemos concluir que el modelo $SARIMA(1,0,0)(0,1,1)_{12}$ cumple con los supuestos.

Para el modelo $SARIMA(1,0,0)(1,1,0)_{12}$ se realizó lo mismo obteniendo resultados análogos al modelo anterior, se puede observar en la Figura 9.

Evaluación de la capacidad predictiva

Dado que los dos modelos mencionados anteriormente satisfacen los supuestos teóricos, la manera de decidir cuál utilizar para realizar las predicciones es en base a la capacidad predictiva que tienen, es decir, nos quedaremos con aquél modelo que tenga menor error de predicción. Para ello haremos uso de varias métricas, haciendo especial énfasis sobre el *MAPE*, dado que es expresado como un porcentaje lo cuál facilita las interpretaciones y las comparaciones.

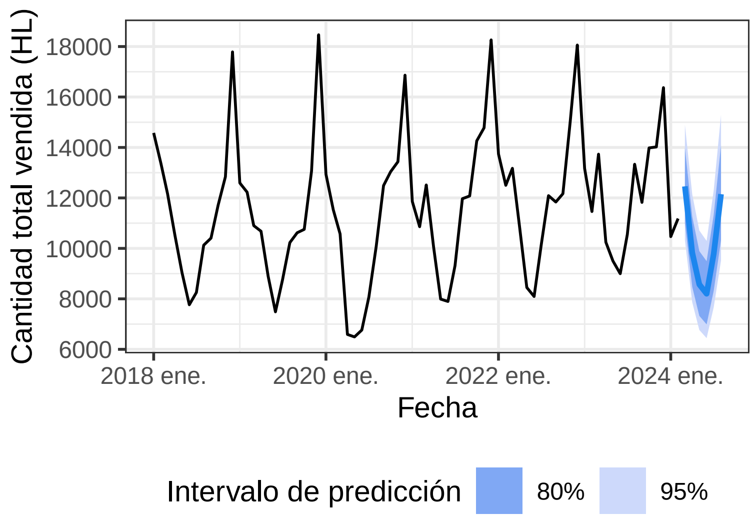
Tabla 3: Evaluación de la capacidad predictiva

Modelos	RMSE	MAE	MAPE
$SARIMA(1,0,0)(0,1,1)_{12}$	1720.094	1291.462	14.3804
$SARIMA(1,0,0)(1,1,0)_{12}$	2198.078	1817.721	20.2913

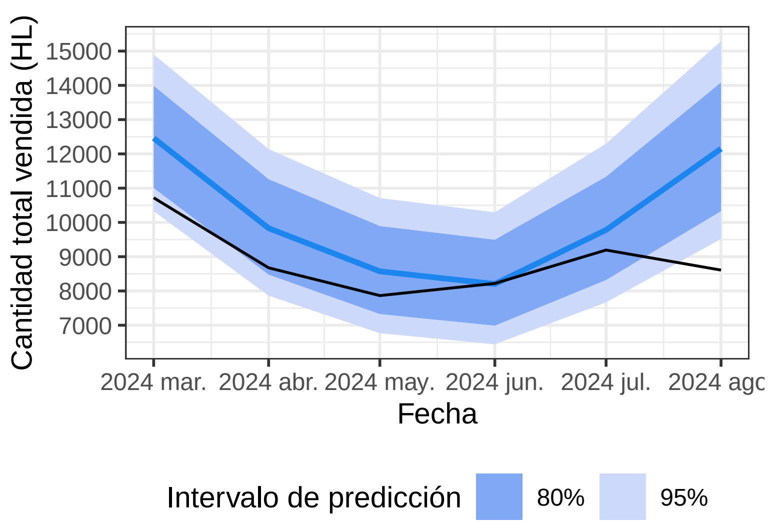
En la tabla 3, se puede verificar que el modelo $SARIMA(1,0,0)(0,1,1)_{12}$ tiene todas las medidas de error de pronóstico más bajas, por lo cuál, dicho modelo se utilizará para realizar las predicciones.

Figura 8: Comparación del pronóstico vs el valor real

(a) Pronósticos de las unidades vendidas en (HL) desde 03/2024 - 08/2024



(b) Pronósticos de las unidades vendidas en (HL) desde 03/2024 - 08/2024 vs unidades vendidas en (HL) reales



En la Figura 8 se pueden observar las predicciones realizadas utilizando el modelo seleccionado anteriormente, en la Figura (8b) se puede ver una comparación entre el valor real (línea negra) y las predicciones realizadas (línea azul), observando que los valores reales en su mayoría se encuentran dentro del intervalo de predicción. Lo que sucede en agosto es que la empresa a la que se le analizaron los datos tuvo una decaída en las ventas sujeto a un problema en las bonificaciones por falta de presupuesto.

Conclusión

El análisis de las ventas mensuales reveló un comportamiento estacional sin tendencias a través de los años, el cuál no tiene mayores inconvenientes además de que la variancia no es constante en el tiempo y el outlier encontrado en el año 2020, sujeto al escenario atípico que se vivió en dicho año. En base a esto, se ajusto un modelo SARIMA no muy complicado con una diferencia en la parte estacional, una parte autorregresiva de primer órden en la parte regular y una parte promedio móvil de primer orden en la parte estacional, el cuál satisface todos los supuestos teóricos.

Luego, en la parte de los resultados se vió que el modelo ajustado tenía un $MAPE$ del 14.4% aproximadamente, lo cuál significa que el error promedio de los pronósticos se desvían un 14.4% respecto a las unidades vendidas reales, lo cuál es aceptable pero lo ideal sería un error más pequeño.

Por último, es importante considerar el pronóstico para agosto de 2024, ya que durante ese mes ocurrió un problema con las bonificaciones en la empresa debido a la falta de presupuesto. Este factor podría influir en el desempeño del modelo, y es posible que el error estimado sea menor al observado.

Anexo

Todo el código con el cuál se desarrollo el trabajo práctico se encuentra en mi repostiorio de Github, al cuál se puede acceder haciendo click aquí.

Tabla 4: Estimación de los parámetros y su significación estadística

(a) Modelo $SARIMA(2,0,0)(1,1,0)_{12}$

Parámetro	Estimación	Desvío	Estadística	P-Value
ϕ_1	0.6187	0.1281	23.3413	<0.0001
ϕ_2	0.0158	0.1319	0.0144	0.90461
Φ_1	-0.4172	0.1162	12.8959	0.00033

(b) Modelo $SARIMA(1,0,0)(1,1,0)_{12}$

Parámetro	Estimación	Desvío	Estadística	P-Value
ϕ_1	0.6284	0.0999	39.5259	<0.0001
Φ_1	-0.4167	0.1164	12.8195	0.00034

(c) Modelo $SARIMA(1,0,0)(0,1,1)_{12}$

Parámetro	Estimación	Desvío	Estadística	P-Value
ϕ_1	0.6368	0.0986	41.7324	<0.0001
Θ_1	-0.6041	0.1617	13.9588	0.00019

Tabla 5: Test de Ljung-Box cada 6 rezagos

(a) Modelo $SARIMA(1,0,0)(0,1,1)_{12}$

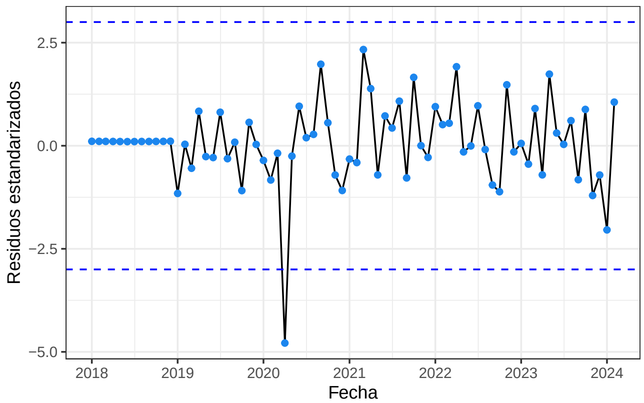
Rezago	Estadística	G.L	P-value
6	3.4630	4	0.4835
12	6.1953	10	0.7986
18	17.0846	16	0.3801
24	17.9538	22	0.7087
30	23.4495	28	0.7102
36	26.2276	34	0.8272

(b) Modelo $SARIMA(1,0,0)(1,1,0)_{12}$

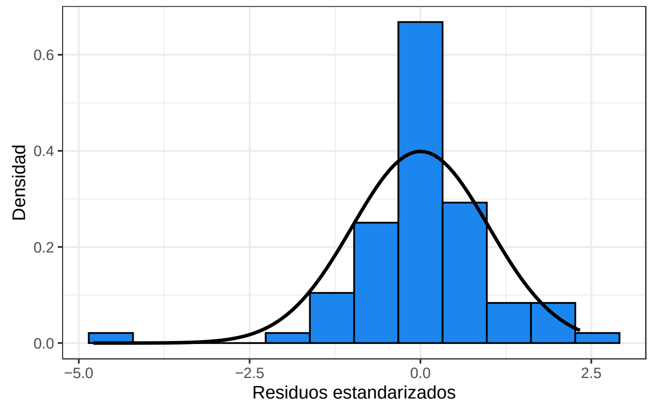
Rezago	Estadística	G.L	P-value
6	3.7688	4	0.4382
12	9.6763	10	0.4693
18	22.6349	16	0.1238
24	27.8891	22	0.1794
30	33.0589	28	0.2336
36	37.6119	34	0.3073

Figura 9: Análisis de residuos $SARIMA(1,0,0)(1,1,0)_{12}$

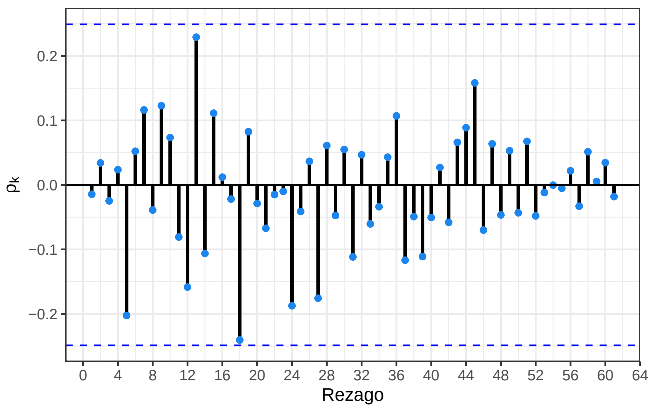
(a) Residuos estandarizados vs fecha



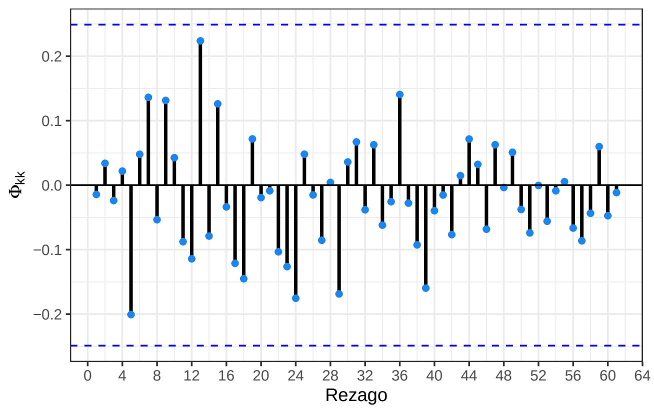
(b) Histograma de los residuos estandarizados



(c) FACM de los residuos



(d) FACPM de los residuos



Bibliografía

- Wei, W.S. (2006). Time Series Analysis: Univariate and Multivariate Methods, 2nd edition
- Hyndman, R.J., & Athanasopoulos, G. (2021). Forecasting: principles and practice, 3rd edition
- Guerrero, V.M. (1993). Time series analysis supported by power transformation