

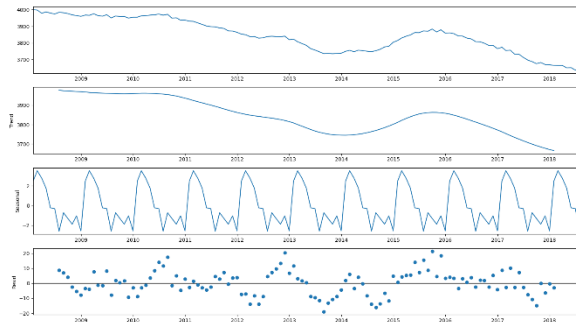
## FUNDAMENTOS DE ANALITICA 2 – MAESTRIA EN CIENCIA DE DATOS

### PARCIAL FINAL PRACTICO

Andres Candelo - Víctor Cardona

Para el proyecto se asumió que los datos tenían una temporalidad mensual para ambos productos y se dividieron en train (120 registros) y test (7 registros) para probar el modelo.

#### PRODUCTO 1

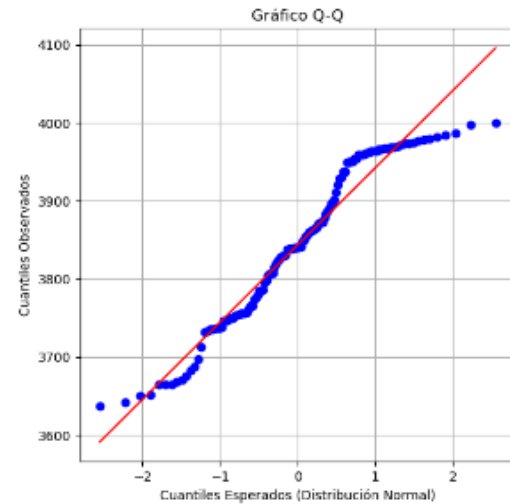


Se realiza una descomposición de la serie donde se puede observar una tendencia bajista, no presenta estacionalidad y los residuales están muy cercanos a cero.

Luego para evaluar si la serie es un ruido blanco, se realiza prueba de rachas propuesta por Wald y Wolfowitz para determinar la autocorrelación de la serie la cual arrojo resultados permiten rechazar la no autocorrelación ( $p\text{-value}=4.3171e-27$ ) y por lo tanto se puede concluir que existe autocorrelación en la serie. Se realiza la prueba de Box-Pierce y la modificación de Ljung-Box para los primeros 20 rezagos, y arrojo que la serie no parece comportarse como un ruido blanco y si presenta autocorrelación.

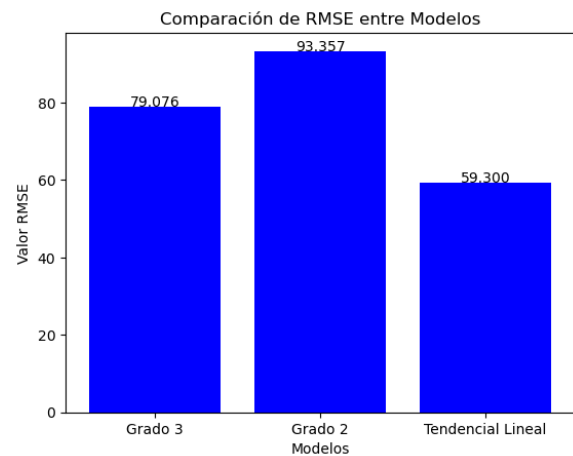
Se realizan pruebas de Homocedasticidad y Heteroscedasticidad de la serie utilizando la prueba de Ljung-Box lo cual arrojo que todos los valores  $p$  son menores al umbral por tanto hay comportamiento ARCH o GARCH.

Prueba de normalidad se evidencia que los datos no siguen una distribución normal, por lo tanto, se recomienda hacer un tipo de muestreo.



#### MODELOS PREDICTIVOS

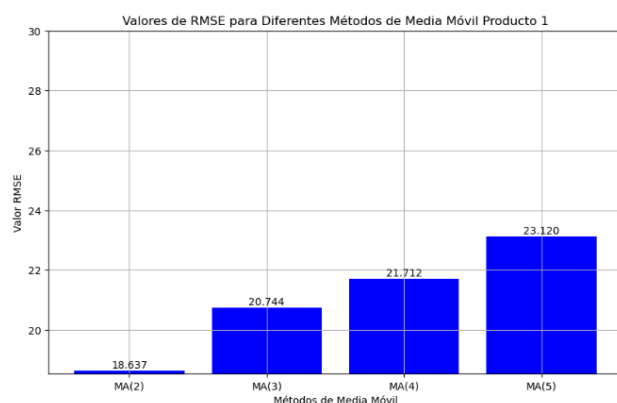
Se prueba un modelo con una tendencia lineal, polinómica de grado 2 y un modelo polinómico de grado 3, los RMSE para estos modelos fueron:



El modelo que dio menor RMSE fue el de tendencia lineal con un  $R^2$  de 0.719. En relación a los supuestos se obtiene que según el test de Durbin-Watson que evalúa la autocorrelación en los residuos del modelo ( $vp=0.039$ ) se concluye que no existe autocorrelación de los residuos, por otro lado el test de Jarque-Bera que evalúa normalidad en los residuos ( $vp=0.499$ ) lo que se concluye que los residuales si siguen una

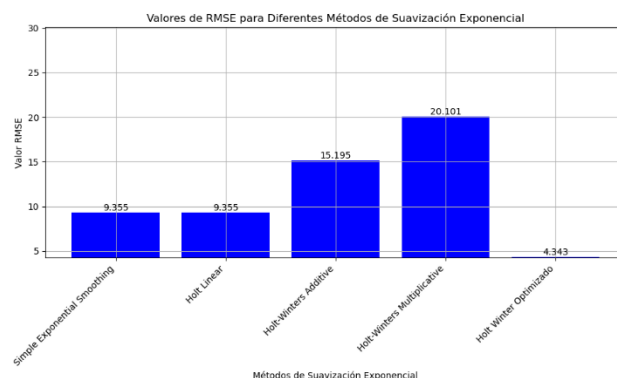
distribución normal y por último se realiza el test de Breuschpagan que evalúa heterocedasticidad en los residuales ( $vp= 6.333 \text{ e-}07$ ) se concluye que Se rechaza  $H_0$  y que hay homocedasticidad en los residuales. Por otro lado, no se evidencia multicolinealidad en el modelo lineal, pero en los otros dos modelos polinómicos de grado 2 y grado 3 si se evidencia. En conclusión, no se recomienda usar el modelo lineal por que no se cumplen los supuestos por ende pierde capacidad para generalizar con nuevos datos.

Se realiza un promedio móvil con ventanas de tiempo de 2, 3, 4 y 5 y se calculan los RMSE



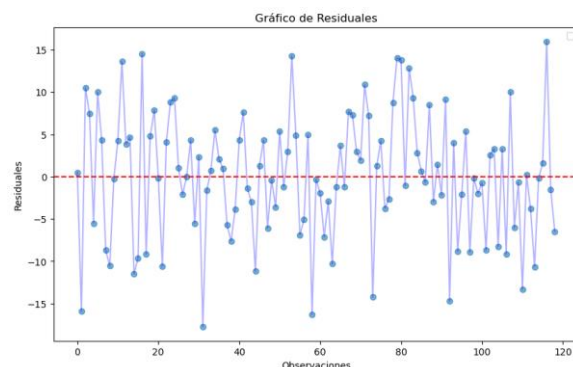
Se obtuvo que el mejor modelo es el promedio móvil con una ventana de 2 obteniendo un RMSE de 18.63

Se probaron métodos de suavización exponencial simple, Holt Linear, Holt-Winter Additive y Holt-Winters Multiplicativa y los rmse fueron:



El mejor modelo fue holt Winter aplicando optimización bayesiana con un RMSE de 4.3431

Se realiza el modelo Autoarima y se obtuvo un RMSE de 5.111, se observa que los residuales fluctúan alrededor de cero



Se obtuvo que el mejor modelo era un SARIMAX SARIMAX(2, 1, 2) con una seasonal order de (0,0,0,0).

Se realizan métodos de validación prueba el modelo para evaluar la capacidad de predecir en los meses siguientes y se obtienen los siguientes RMSE:

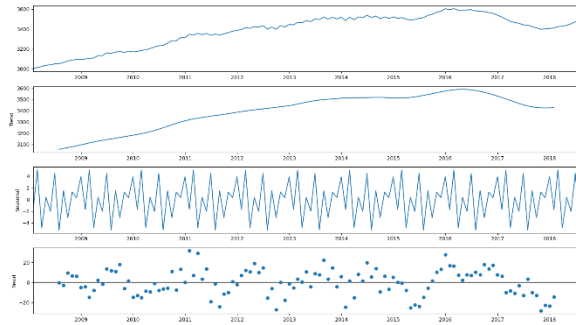
Pronostico 1: RMSE = 0.09598048425687011  
Pronostico 2: RMSE = 0.5531473656928938  
Pronostico 3: RMSE = 0.5678479327830246  
Pronostico 4: RMSE = 14.296779733413132  
Pronostico 5: RMSE = 13.390149418093188  
Pronostico 6: RMSE = 28.130979645733078  
Pronostico 7: RMSE = 23.18774638656305

En términos generales luego de realizar una iteración para probar el modelo por cada uno de los datos vemos se observa que tiene buena capacidad de predicción en los primeros 3 meses y a medida que se va alejando tiende a aumentar lo que quiere decir que se debe de realizar un reentrenamiento para así garantizar un error más bajo. Se realizan los test para validar los supuestos en los residuales:

Test	Hipótesis resultado	Respuesta
Durbin Watson	No autocorrelación	Cumple supuesto
Box-Pierce	No autocorrelación	Cumple supuesto
ljungbox	Homocedasticidad	Cumple supuesto
Shapiro-Wilk	Normalidad	Cumple supuesto

Se recomienda utilizar los modelos de Holt Winter Optimizado o el modelo SARIMAX dado que dieron las mejores métricas de predicción y para el caso del SARIMAX se cumplen los supuestos en los residuales

## PRODUCTO 2

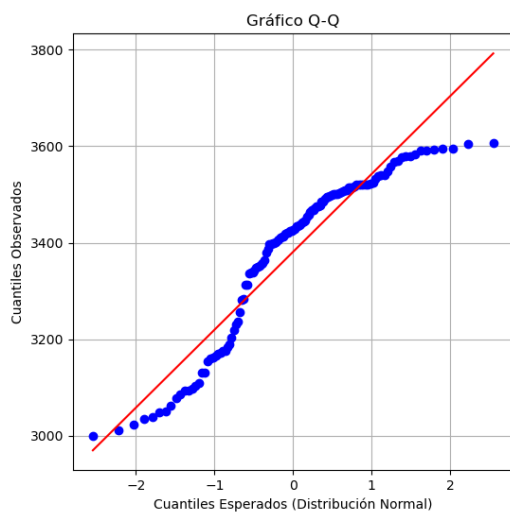


Se realiza una descomposición de la serie donde se puede observar una tendencia alcista los valores de la estacionalidad fluctúan entre -4 y 4 y los residuales rondan entre -20 y 20.

Se comprueba si la serie es ruido blanco:

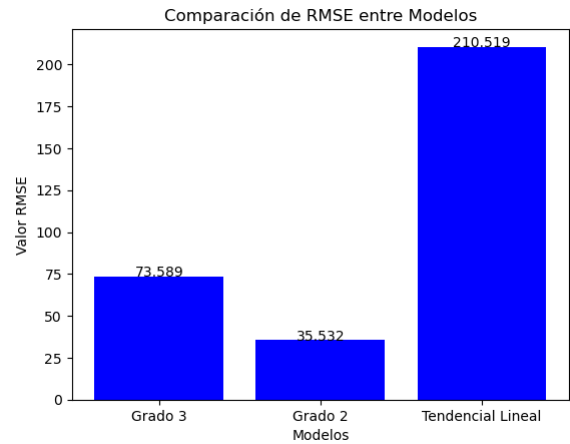
Test	Hipótesis resultada	Respuesta
Box-Pierce y la modificación de Ljung-Box	Autocorrelación	Cumple supuesto
jungbox	No autocorrelación de las varianzas	Cumple Supuesto
Shapiro Wilk y QQ-Plot	Normalidad	No cumple supuesto

Se puede decir que la serie no es ruido blanco solo tiene problemas de normalidad pero esto se puede solucionar con un bootstrapping de la serie.



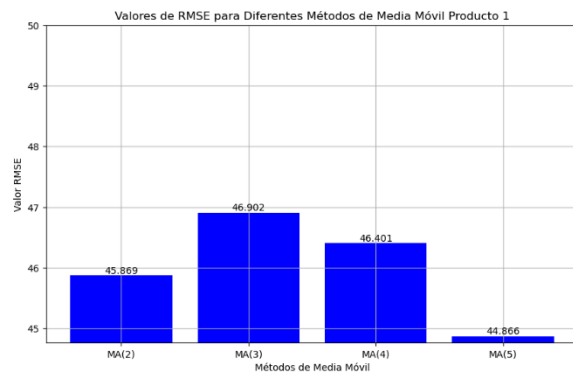
## MODELOS PREDICTIVOS

Se prueba un modelo con una tendencia lineal, polinómica de grado 2 y un modelo polinómico de grado 3, los RMSE para estos modelos fueron



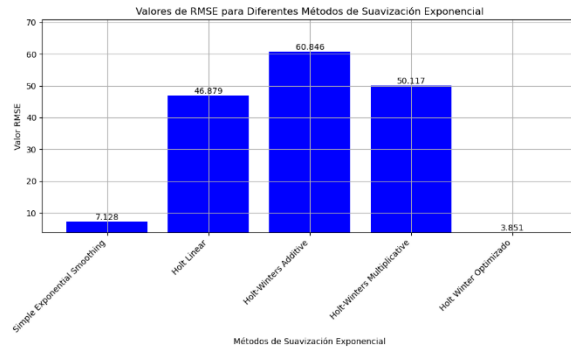
El modelo que obtuvo mejor RMSE fue el de grado dos con un RMSE de 35.532 pero presenta alto grado de multicolinealidad, por ende no es recomendable generar predicciones con este modelo.

Se realiza un promedio móvil con ventanas de tiempo de 2, 3, 4 y 5 y se calculan los RMSE



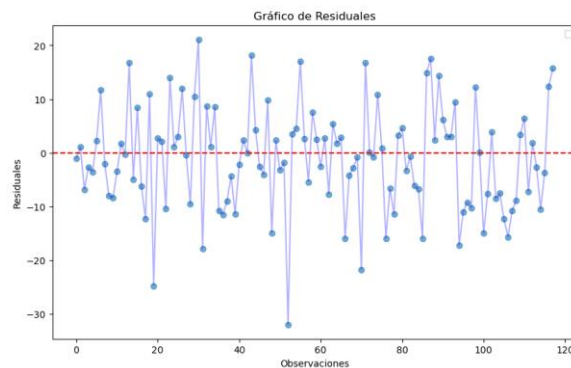
Se obtuvo que el mejor modelo es el promedio móvil con una ventana de 5 obteniendo un RMSE de 44.856

Se probaron métodos de suavización exponencial simple, Holt Linear, Holt-Winter Additive y Holt-Winters Multiplicative y los rmse fueron:



El mejor modelo fue holt Winter aplicando optimización bayesiana con un RMSE de 3.851.

Por último, se aplicó autoarima y se obtuvo RMSE de 43.47



Se obtuvo que el mejor modelo era un SARIMAX SARIMAX(2, 2, 1) con una seasonal order de (0,0,0,0).

Se realizan métodos de validación de prueba del modelo para evaluar la capacidad de predecir en cada uno de los meses y se obtienen los siguientes RMSE:

Pronostico 1: RMSE = 5.167232531978698  
Pronostico 2: RMSE = 18.672131149709003  
Pronostico 3: RMSE = 23.463062845638888  
Pronostico 4: RMSE = 31.269496106508996  
Pronostico 5: RMSE = 47.32531080494891  
Pronostico 6: RMSE = 66.6983737156188  
Pronostico 7: RMSE = 70.22812108628887

Se concluye que a medida que va aumentando los meses se va perdiendo la capacidad de predicción y el rmse tiende a aumentar. En términos generales el modelo no es bueno, Teniendo en cuenta que si la variable a predecir es venta del producto 2 se tiene que garantizar que al menos realice un buen pronóstico en el segundo mes.

Se realizan los test para validar los supuestos en los residuales:

Test	Hipótesis resultado	Respuesta
Durbin Watson	No autocorrelación	Cumple supuesto
Box-Pierce	No autocorrelación	Cumple supuesto
ljungbox	Homocedasticidad	Cumple supuesto
Shapiro-Wilk	Normalidad	Cumple supuesto

Se recomienda utilizar el modelo de Holt Winter Optimizado dado que dio la mejor métrica de predicción.