



---

# **Series de tiempo**

## **Volumen del Aguacate**

---

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

Volumen semanal de aguacate convencional vendido en  
EEUU

Noviembre 2020

Equipo:

- 6. Ávila Argüello Carlos
- 10. Bonilla Cruz José Armando
- 39. Luna Gutierrez Yanelly
- 64. Reyes González Belén
- 67. Rivera Mata Dante Tristán

## Índice

---

1. Introducción.....	2
2. Análisis Descriptivo.....	2
3. Análisis estadístico .....	5
3.1 Modelos propuestos	
3.2 Cuadro comparativo	
3.3 Mejor modelo y Predicciones	
4. Conclusiones.....	14
5. Bibliografía.....	14

## Introducción

---

La información que se genera por medio de la realidad se puede tratar de replicar a futuro siempre y cuando se tengan modelos matemáticos adecuados. En este proyecto nos encargaremos de tratar de adaptar un modelo de series de tiempo que pueda describir a futuro el comportamiento del volumen del aguacate convencional vendido en Estados Unidos del 4 de enero de 2015 al 25 de marzo de 2018.

En general, tratar información (datos) es una herramienta muy útil para toda clase de áreas, puede ser para empresas, laboratorios, cuestiones políticas o bien poblacionales, pero no es importante sólo la recabación de información y su descripción, lo que puede ser de gran interés es la forma en la que se interpretan y usan.

En nuestro caso decidimos analizar los datos relacionados al volumen de aguacate vendido en los Estados Unidos, nuestro objetivo; modelar el volumen de aguacate pues según la USDA (United States Department of Agriculture) la demanda de aguacates (2018, USDA)<sup>1</sup> se ha triplicado desde el 2001 hasta el 2018, por lo que nos gustaría dar un modelo del volumen de aguacate.

## Análisis Descriptivo

---

Nuestra bases de datos se compone de 13 columnas, entre las que se encuentran; indice, Date, Average Price, Total Volume, 4046, 4225, 4770, Total Bags, Small Bags, Large Bags, XLarge Bags, type, year, region.

Algunas de las columnas más importantes son:

# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**Date** - Fecha de observación, semanal del 2015/01/04 al 2018/03/25

**AveragePrice** - precio medio de un aguacate

**type** - conventional o organic

**year** - año

**Región** - la ciudad o región de observación

**Total Volumen** - Total de aguacates vendidos (libras)

**4046** - Total de aguacates vendidos con PLU (Price look up code) 4046

**4225** - Total de aguacates vendidos con PLU 4225

**4770** - Total de aguacates vendidos con PLU 4770

A pesar de que la base original tiene información que puede ser de gran interés, este trabajo busca adaptar un modelo de series de tiempo al volumen de aguacate, es por ello que se aplicarán diferentes filtros a la "base de datos" para obtener la información requerida. El primer filtro que aplicamos fue en la columna Type extrayendo solamente los datos correspondientes a conventional (tipo de aguacate convencional), pasamos de tener 18,249 filas a 9,126 esto porque para la misma semana se tiene registrada la venta de aguacates de tipo orgánico y convencional, sin embargo la proporción de aguacate orgánico con respecto a la venta total es menor al 3%, por tanto no es representativo de la venta total.

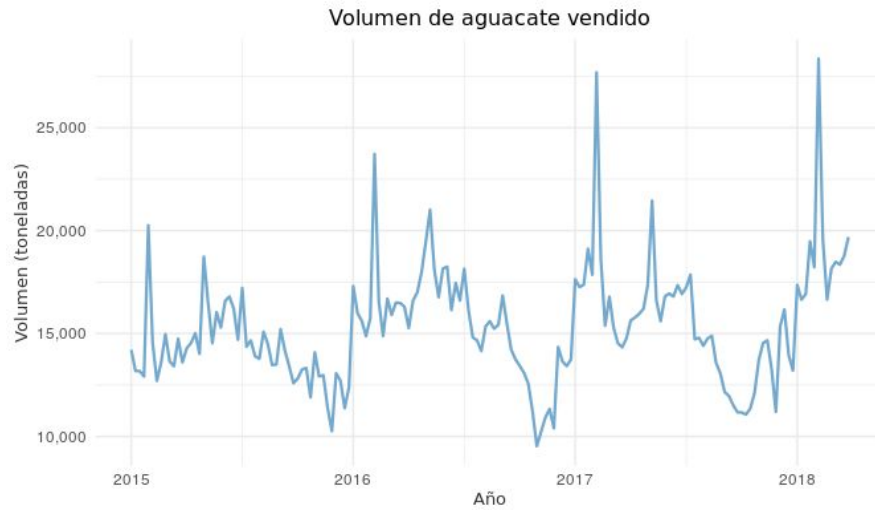
En la columna región podemos encontrar la zona geográfica de la cual provienen los datos (volumen vendido de aguacate) como: New York, California, ... y Estados Unidos. Como nuestro análisis considerará el volumen de aguacate convencional vendido en todo el país (EU), filtramos nuestra base bajo ese criterio y pasamos de los 9,126 renglones a 169 lo cual no quiere decir que hubo pérdida de información sino que hubo agrupación de la venta -se considera en esta información el volumen de venta por cada estado pero sin hacer la diferencia entre cada uno-.

El objetivo de este análisis es darle una explicación al volumen total de aguacates convencionales de todo Estados Unidos de América, por lo que sólo necesitaremos la columna de fechas y la de volumen total. De tal forma que la representación gráfica de nuestra base de datos es la siguiente:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
9530	13500	14966	15302	16795	28352

En la imagen anterior observamos que la media es algo mayor a la mediana y que existe una amplia diferencia entre el mínimo de toneladas de aguacate y el máximo de 18,822. Otra cosa que podemos observar es que el primer cuartil y el tercer cuartil no son tan distantes, lo que nos podría indicar un comportamiento leptocúrtico de la distribución. NOTA: Los datos de las semanas que manejamos a lo largo del periodo de estudio está completo, es decir, no tenemos datos faltantes, por lo que no fue necesario utilizar un método de imputación de datos.

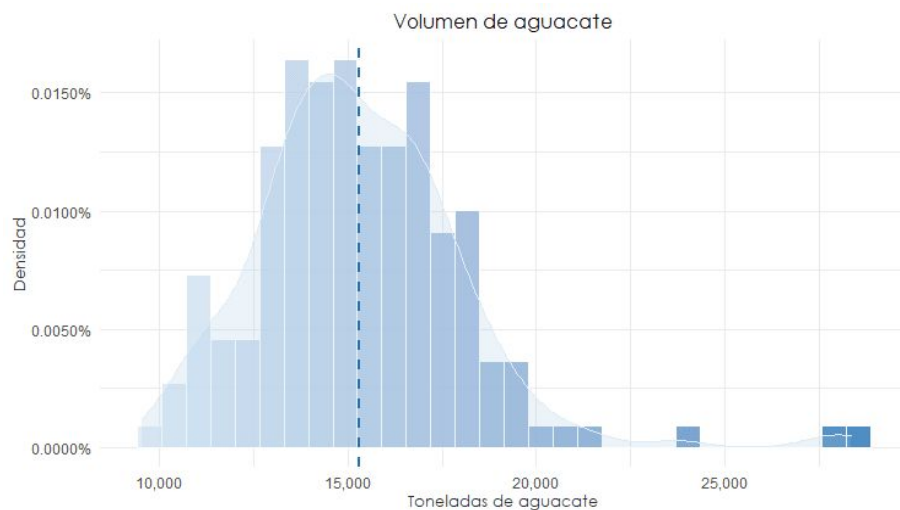
## Serie de tiempo



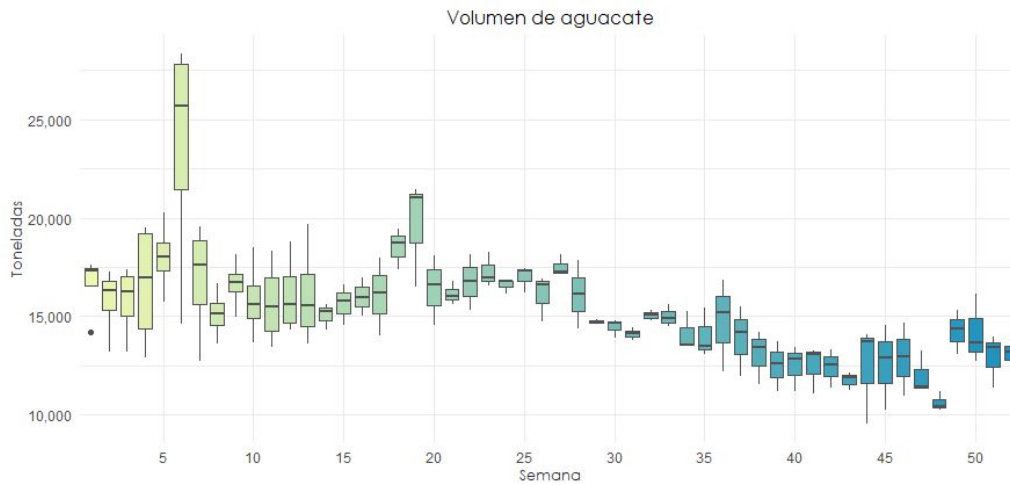
Al graficar nuestra serie de tiempo, podemos notar que la parte cíclica parece repetirse cada año (cada 52 semanas). Identificar la parte de la tendencia tampoco es sencillo en este caso, porque se podría confundir con la estacionalidad, aunque si nos imaginamos una regresión lineal sobre nuestros datos podremos darnos cuenta que efectivamente tiene una parte tendencial pequeña. Otra cosa que nos podríamos preguntar es si cuenta con varianza constante o no constante. No es evidente si hay varianza creciente. Del análisis anterior, podemos observar que necesitamos otras herramientas para poder describir y reafirmar las observaciones acerca de las componentes de esta serie de tiempo.

## Histograma

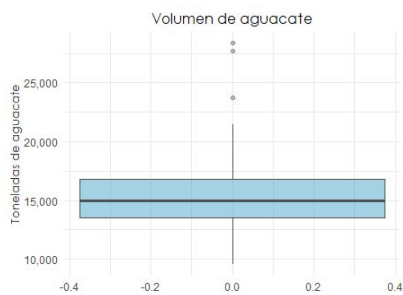
Del histograma podemos observar que tenemos una asimetría positiva, eso significa que la media está a la derecha de la mediana, lo que se debe a la presencia de valores extremos (ver serie de tiempo). Por otro lado, tenemos una distribución leptocúrtica, es decir, muchos de nuestros datos se concentran alrededor de la moda.



## Box Plot y Box Plot cíclico



De los los diagramas "Box Plot" podemos identificar que efectivamente hay datos atípicos, que son correspondientes a la segunda semana de febrero, lo que podría coincidir con la fecha del Super Bowl, evento deportivo significativo en la cultura estadounidense donde el guacamole mexicano ha tomado protagonismo, y por lo tanto se necesita un gran volumen de este producto agrario.



El impacto de las fechas del Super Bowl se podría verse claramente en las puntas más grandes de nuestra serie de tiempo. Por otra parte, podemos notar que a mediados de año, el volumen de aguacates no varía mucho, este fenómeno se puede deber a que la temporada de cosecha y venta a temporada alta de cosecha se produce entre los meses de octubre y febrero, mientras que la temporada media es entre marzo y mayo, y la baja entre junio y septiembre.

## Análisis estadístico

En series de tiempo se busca trabajar con series estacionarias o en su defecto con series débilmente estacionarias, es decir, series con varianza constante, media constante y sin tendencia.

Sin embargo, nuestra serie no es, al menos, débilmente estacionaria, por lo que nos auxiliamos de algunas funciones de R como *ndiffs* y *nsdiff* que nos ayudan a dar un estimado de cuántas diferencias tenemos que aplicar a nuestra serie para hacerla estacionaria eliminando la parte tendencial y para la tendencia estacional. Si la serie, luego de las diferencias aplicadas, es estacionaria, entonces podremos determinar que los números de diferencias obtenidas serán parámetros útiles para la parte de diferencias (ya sea para la tendencia o la parte estacional).

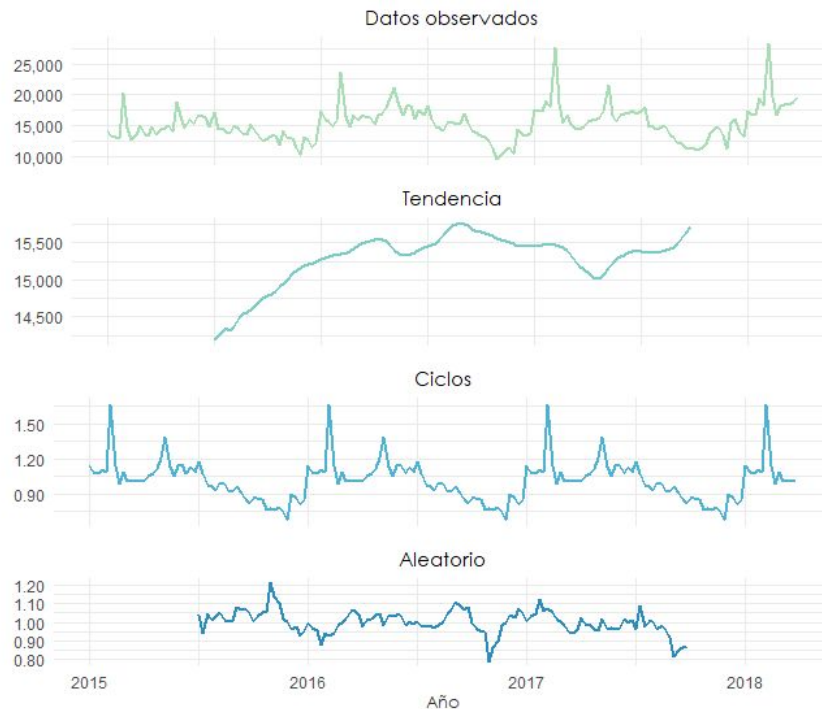
La primero fue analizar los correlogramas de la serie y aplicar diferencias para así dar algunas opciones para los parámetros que necesitamos para ajustar algún tipo de modelo, ya sea un  $AR(p)$ , un  $MA(q)$ , un  $ARMA(p,q)$  o incluso un  $SARIMA(p,d,q)(P,D,Q)(s)$  que toma en cuenta parámetros para la parte de estacionalidad de la serie original.

Para atacar nuestro problema hay varios caminos que podemos seguir, pero en esencia nos vamos a basar en lo siguiente: el análisis de la serie sin transformación alguna, con la transformación logaritmo y con una transformación Box Cox. Como se busca proponer varios modelos y después elegir el que tenga mejores resultados no se mencionarán detalles sobre los supuestos del modelo ni sobre la significancia de sus parámetros hasta el final del trabajo en una tabla resumen.

### Análisis de la Serie “Original” (sin aplicar transformaciones a los datos)

Existen tres puntos que tenemos que revisar de nuestra serie original, el primero es la tendencia, segundo los ciclos estacionales y tercero la parte aleatoria. Para obtener la componente de la tendencia y la parte cíclica utilizamos filtros lineales, en nuestro caso, al tener observaciones semanales y tres años de observaciones, decidimos darle una ventana a los promedios de 26 semanas, es decir que se calcularan los promedios con 26 semanas antes y después de la semana en cuestión.

Al tener las componentes anteriores, se puede obtener la parte aleatoria de la serie de tiempo simplemente haciendo la resta de la tendencia y parte cíclica a la serie original. Lo anterior lo podemos resumir en la siguiente gráfica que descompone nuestra serie de forma multiplicativa: (el siguiente gráfico es propio NO se utilizó la función decompose de Rstudio)



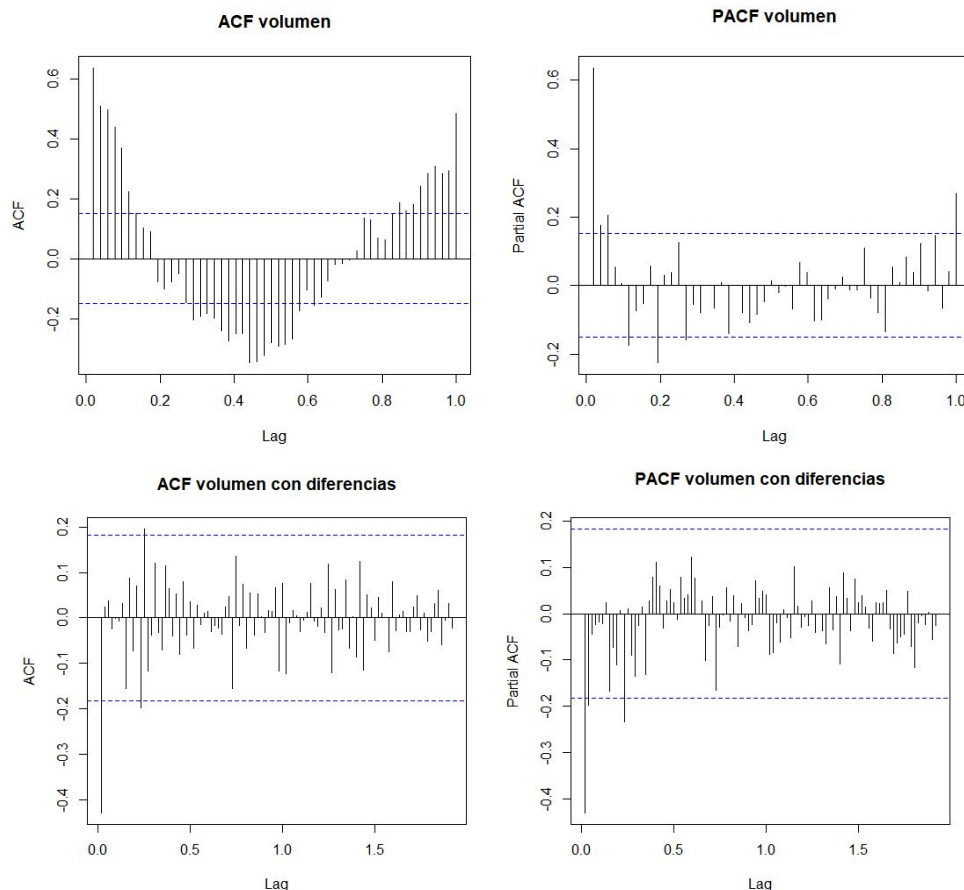
Notemos que considerando una ventana semestral podremos ver una tendencia que inicialmente es creciente hasta el 2017, es decir que en general el volumen de la venta de aguacate convencional en EEUU fue creciendo desde 2015 hasta 2017 donde se dio un cambio de tendencia.

Por otra parte, en los ciclos estacionales podemos notar que se repiten anualmente teniendo un claro aumento en febrero cuando se presenta el super Bowl y un decremento en diciembre. Esta tendencia y ciclos son los que podemos esperar al momento de hacer la predicción.

Para verificar la estacionariedad de la serie original utilizamos diversas pruebas estadísticas como: Dickey-Fuller, de la cual obtuvimos la aprobación de la hipótesis nula que corresponde a que **la serie original no es estacionaria**. Pero para confirmar lo anterior se procedió a utilizar los métodos de Phillips-Perron Unit Root Test y Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test las cuales confirmaron que se cumplía estacionariedad. Por lo tanto diremos que la serie original si cumple con la estacionariedad débil.

## Análisis Serie Original con diferencias

Aplicamos dos diferencias a la serie original: una para quitar la tendencia y otra con lag = 52 para quitar la estacionalidad, ya que los ciclos se repiten anualmente y nuestra información es semanal. A esta nueva serie le aplicamos las tres pruebas para estacionariedad mencionadas anteriormente y obtuvimos que en todas se acepta que la serie es estacionaria a un nivel de significancia de 0.05.



En los correlogramas anteriores podemos observar que para la serie original no es posible identificar un modelo que describa las autocorrelaciones pues la serie tiene tendencia y ciclos, lo que dificulta proponer un modelo. Sin embargo, podemos observar que a partir del lag 0.2 (que equivale a 10 semanas) las correlaciones comienzan a ser negativas, es decir, están asociadas inversamente con las observaciones de 10 semanas posteriores.

Además, este ACF marca la estacionalidad de nuestros datos, lo cual podría explicarse porque el aguacate es un producto agrícola que tiene temporadas de cosecha y de no producción. Lo anterior se puede confirmar ya que el ACF muestra una gran correlación con sus datos cercanos y a "lags" de mitad de año (26 semanas) correlaciones negativas pues ya no son temporadas de cosecha.

Por otro lado, cuando aplicamos las diferencias para quitar la tendencia y los ciclos estacionales observamos que la mayoría de las autocorrelaciones (incluyendo las primeras) se quedan dentro de la banda de confianza, a excepción de las que tienen "lags" de 10 y 11 semanas.

En las autocorrelaciones parciales notamos que para el lag de una semana estamos fuera de las bandas de confianza, por lo que podríamos proponer un modelo MA(1) ya que la altura de las correlaciones que siguen quedan dentro de las bandas de confianza, es decir, son estadísticamente cero, a excepción nuevamente del "lag" correspondiente a la semana 10.

Por lo tanto, trabajando la serie original e incluyendo pruebas de estacionariedad, diferencias, y correlogramas propondremos los siguientes tres modelos:

- 1) El modelo que propone el método de R auto arima con el fin de comparar los parámetros con nuestros parámetros propuestos: SARIMA(0,1,0)(1,1,0)[52]
- 2) ARMA(1,1). Este modelo se propuso al graficar el ACF y el PACF de la parte aleatoria que obtuvimos de la función *decompose*.

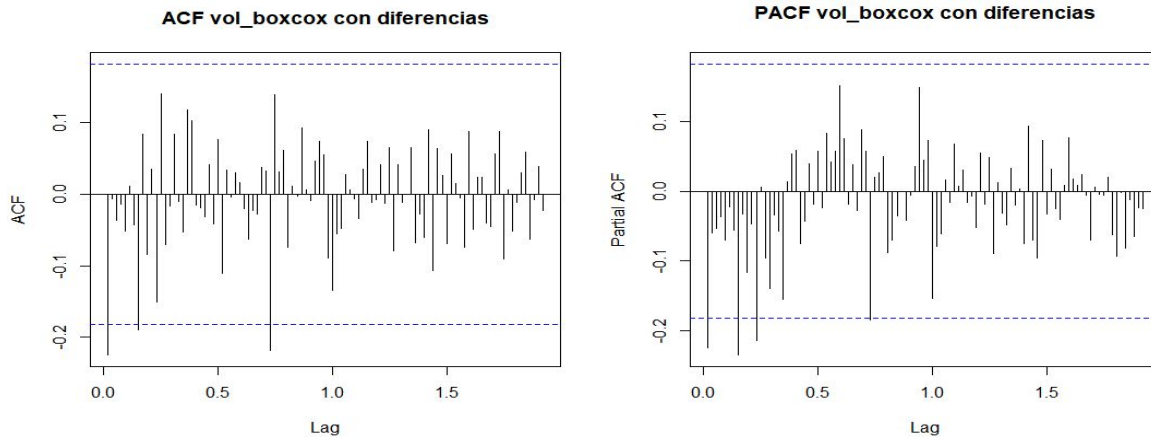
### **Análisis de la Serie con Transformación Box Cox**

Ahora trabajaremos la serie pero con la transformación que nos indique el método de Box Cox. Utilizando el comando de R para Box Cox obtenemos el valor de  $\lambda = -0.9$ , por lo que a nuestra serie original le sacaremos el inverso multiplicativo y trabajaremos con esta nueva serie transformada.

#### **Análisis Serie Box Cox con diferencias**

El primer acercamiento fue utilizar las funciones antes mencionadas (*ndiff* y *nsdiff*) para ver posibles diferencias por hacer, en este caso, se recomendaban no realizar diferencias para la tendencia y una diferencia para la parte estacional.





Con el diagrama de autocorrelación de las diferencias de la serie de tiempo con la transformación inversa, podemos ver que existe relación con 5 o 6 semanas y que a partir de ese se encuentran dentro de la banda de confianza. También se puede entender con el diagrama de autocorrelación parcial, que muestra relación con semanas próximas y después las líneas están dentro de las bandas de confianza.

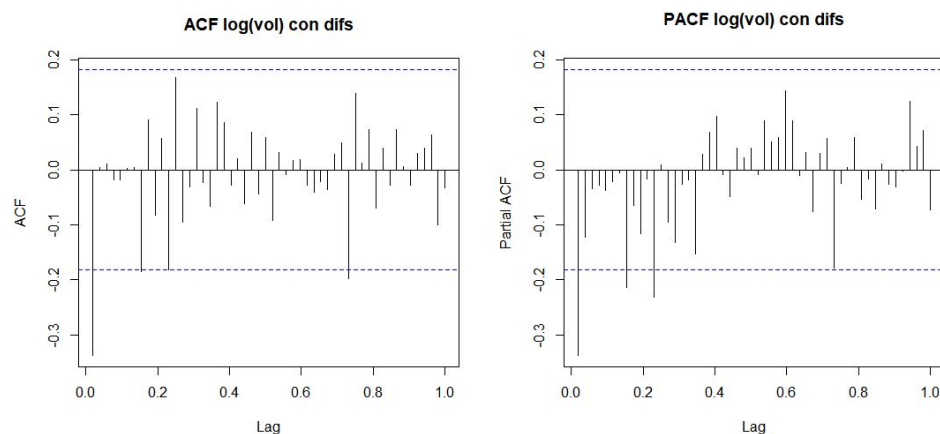
Lo anterior, sumado a que se nota un ciclo en nuestros datos podemos sospechar sobre los parámetros para modelos  $AR(p)$  y  $MA(q)$  pueden ser de valor 1. Aunado a lo previamente visto, notamos que las diferencias si están aportando información para la parte de la tendencia y para la parte cíclica, dejándonos más ruido blanco.

De la misma manera que en el tratamiento anterior pudimos determinar algunos modelos propuestos:

- 1) El modelo que propone el método de R auto arima con el fin de comparar los parámetros con nuestros parámetros propuestos:  
SARIMA(0,0,0)(0,1,0)(52)
- 2) SARIMA(0,0,1)(0,1,0)(52)
- 3) SARIMA(0,1,1)(0,1,0)(52)

## Análisis de la Serie con Logaritmo Natural

Para poder determinar cómo son los parámetros de nuestra serie transformada por medio del logaritmo natural, veamos los correlogramas de autocorrelación y autocorrelación parcial:



De estos correlogramas no podemos inferir muchas cosas, pareciera que son provenientes de ruido blanco, salvo unos cuantos puntos donde la correlación no parece ser cero, el problema es que estos puntos donde la correlación no es cero son de los primeros lags, por lo que no intentamos ajustar modelos con  $p,q=0$  o a lo más 1.

Por lo tanto, tratando la serie transformada con la función logaritmo, incluyendo pruebas de estacionariedad, diferencias, y correlogramas proponemos los siguientes tres modelos:

- 4) El modelo que propone el método de R auto arima con el fin de comparar los parámetros con nuestros parámetros propuestos SARIMA(0,1,0)(1,1,0)[52]
- 5) SARIMA(0,1,1)(0,1,0)[52]
- 6) SARIMA(1,1,1)(0,1,0)[52]

### Comparación de modelos

Los criterios de decisión para elegir alguno de los ocho modelos propuestos así como sus resultados se plasman a continuación.

Transforma -ción	Modelos propuestos SARIMA(p,d,q)(P,D,Q)[s]	Información sobre los modelos											
		Datos		Significancia					Supuestos				
		AIC	BIC	p	d	q	P	D	Q	Norma- lidad	No correlación	Homoced- asticidad	Media cero
Datos sin transformar	SARIMA(0,1,0)(1,1,0)[52]	2094	2099	.	Sí	.	No	Sí	.	No	No	Sí	Sí
	ARMA(1,1)	2069	2082	No	.	No	.	.	.	No	No	Sí	Sí
BoxCox	SARIMA(0,0,0)(0,1,0)[52]	-2355	-2350	.	.	.	.	Sí	.	No	No	No	Sí
	SARIMA(0,1,1)(0,1,0)[52]	-2404	-2396	.	Sí	Sí	.	Sí	.	No	Sí	Sí	Sí
	SARIMA(0,0,1)(0,1,0)[52]	-2400	-2393	.	Sí	Sí	.	.	.	No	No	No	No
Logaritmo Natural	SARIMA(0,1,0)(1,1,0)[52]	-160	-155	.	Sí	.	No	Sí	.	No	No	Sí	Sí
	SARIMA(0,1,1)(0,1,0)[52]	-178	-171	.	.	Sí	.	Sí	.	No	Sí	Sí	Sí
	SARIMA(1,1,1)(0,1,0)[52]	-176	-166	No	.	No	.	Sí	.	No	Sí	Sí	Sí

De todos nuestros modelos propuestos, ninguno cumple con la normalidad, esto se puede deber a que tenemos una muestra con tres años de observaciones, aunado a que utilizamos la prueba más potente para normalidad.

Podemos observar que casi todos los modelos cumplen con el supuesto de media cero, supuesto que comprobamos con el “*t.test*”, esto se da principalmente por la construcción de los modelos, pues si podemos observar, casi todos tienen diferencias tendenciales. La homocedasticidad tampoco causa mucho problema en nuestro análisis debido a que en la serie original se tenían pocos datos atípicos, de hecho tres (chechar BoxPlot).

El supuesto de no correlación sólo lo cumplen tres modelos, por lo que nos centraremos en estos. Es importante este supuesto ya que el hecho de su no cumplimiento (es decir que haya correlación entre los errores) implicaría que el modelo no está considerando la dependencia “total” entre las observaciones de la serie.

Observando la significancia de los modelos mencionados podemos excluir al SARIMA(1,1,1)(0,1,0)[52], que no cumple que sus parámetros (los coeficientes asociados a la parte autorregresiva y de promedios móviles) sean significativos.

A partir de los criterios previamente descritos optamos por los siguientes modelos: el SARIMA(0,1,1)(0,1,0)[52] con transformación Box Cox que tiene un AIC = -2404 y el de SARIMA(0,1,1)(0,1,0)[52] con transformación Logaritmo Natural que tiene un AIC mayor. Si comparamos estos dos modelos, optaríamos por el Box Cox dados sus resultados, pero este tiene un pequeño problema al hacer la predicción y es que, cuando hacemos la predicción de las bandas de confianza, estas toman valores negativos, algo que no queremos (porque el volumen del aguacate "negativo" no tiene sentido "real"), además al hacer la transformación inversa tenemos predicciones "muy malas", por lo que optamos por el otro modelo.

**Por tanto** nuestro **modelo propuesta es** un **SARIMA(0,1,1)(0,1,0)[52]**, esto quiere decir que tenemos un proceso estocástico  $Y_t$  tal que :

Donde  $Z_t$  es la observación (del volumen de aguacate) a tiempo  $t$

$$Y_t = (1 - B)(1 - B^{52})X_t$$

$$X_t = \ln(Z_t)$$

con  $X_t$  un proceso estacional multiplicativo  $ARMA(0, 1) \times ARMA(0, 0)_{52}$ , lo cual es equivalente a un MA(1) pues la parte estacional tiene coeficientes iguales a 0 es decir:

$$Y_t = (1 + B)w_t$$

Entonces,

$$(1 - B)(1 - B^{52})X_t = (1 + B)w_t$$

$$(1 - B - B^{52} + B^{53})X_t = (1 + B)w_t$$

$$X_t - X_{t-1} - X_{t-52} + X_{t-53} = w_t + w_{t-1}$$

$$X_t = X_{t-1} + X_{t-52} - X_{t-53} + w_t + w_{t-1}$$

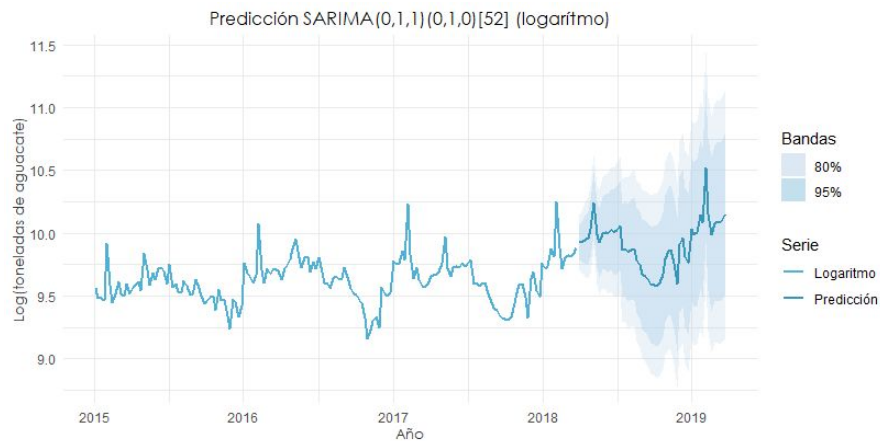
Estimamos un valor de  $\hat{w} = -0.3844$ , por lo que el modelo propuesto es:

$$X_t = X_{t-1} + X_{t-52} - X_{t-53} + w_t - 0.3844w_{t-1}$$

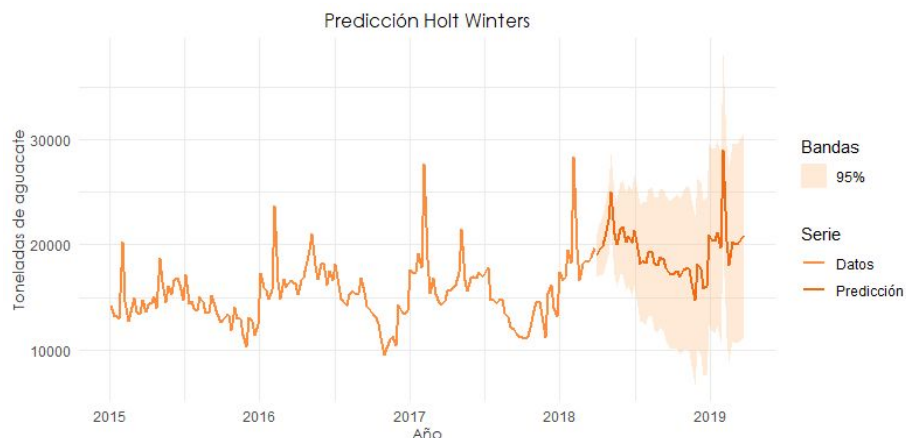
### Predicciones de Modelos Propuestos

A continuación observaremos las predicciones tanto del modelo SARIMA(0,1,1)(0,1,0)(52) como del modelo propuesto por el método de Holt Winters.

En ambas gráficas se observa la serie original hasta un tercio del año 2018 y después una año de predicción, eso quiere decir que observamos 52 semanas después del primer tercio del año del 2018.



Podemos observar de esta predicción utilizando el logaritmo de los datos tiene un gran área en las bandas de confianza para los valores que predécimos, estas bandas tan grandes se pueden explicar porque el intervalo de confianza de nuestro parámetro es  $(-0.57, -0.27)$  y también se puede justificar ya que no estamos cumpliendo el supuesto de normalidad en nuestros datos (supuesto que se utiliza en el cálculo de las bandas de confianza).



De esta gráfica sobre la predicción con Holt-Winters podemos notar que las bandas de confianza son más reducidas a comparación de la gráfica anterior. Lo anterior se debe a que este método no toma en cuenta la varianza del modelo (creciente o decreciente) ni los supuestos de sus residuales (distribución normal).

## Comparación de Predicciones de Modelos Propuestos

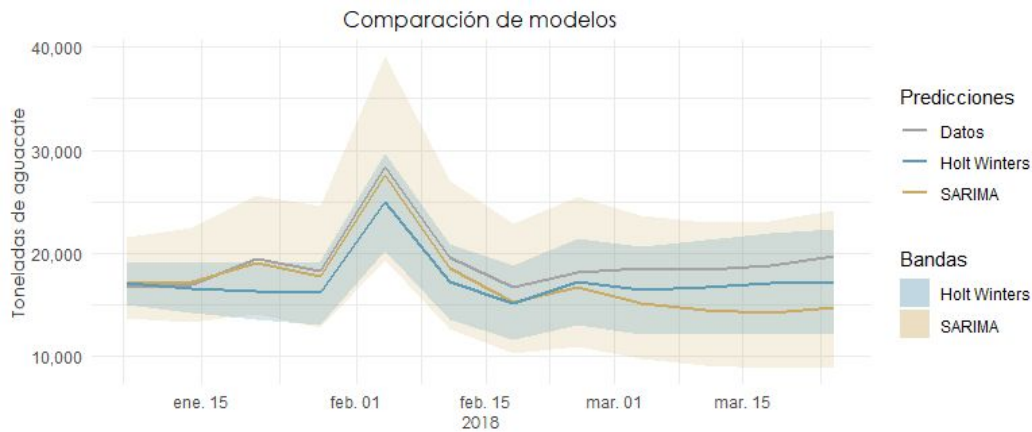
Una vez que comparamos los modelos propuestos con la tabla anterior, se eligieron como candidatos el modelo SARIMA(0,1,1)(0,1,0)(52) del logaritmo y el modelo por medio de Holt Winters (pues este no necesita cumplir con el supuesto de normalidad).

Recordemos que Holt Winters, es un modelo no paramétrico que cuenta con unas ecuación de pronósticos y otras tres ecuaciones las cuales son para la tendencia,

estacionalidad y el nivel de confianza. Este modelo HW es muy útil pues no tomó en cuenta la varianza de la serie de tiempo además de que su aplicación se restringe a cuatro líneas de código.

Para poder comparar las predicciones de los modelos propuestos decidimos recortar la serie de tiempo original hasta el 2018 y con estos datos construir el modelo, mientras que las observaciones restantes servirán para comparar las predicciones de ambos modelos con los datos observados. Es decir que utilizamos el método de "Validación Cruzada".

Tanto con el modelo SARIMA(0,1,1)(0,1,0)(52) como con el método de Holt-Winter se realizaron las predicciones de las observaciones posteriores de la serie, obteniendo así las siguientes gráficas:



De esta gráfica podemos observar que la predicción del modelo SARIMA(0,1,1)(0,1,0)(52) aproxima de mejor manera el mes de febrero donde se realiza el Super Bowl pero pierde precisión al pasar esta fecha. Por otro lado, se observa que la predicción con el método de Holt-Winters no replica tan bien la punta del Super Bowl pero se aproxima de mejor manera a las observaciones reales después de esta fecha.

Aunque con la observación de las gráficas, pareciera que el método de Holt-Winters es una mejor estimación. utilizaremos el método de errores medios absolutos para verificar cual es una mejor predicción, para eso presentamos las siguientes fórmulas:

$$Error_{SARIMA} = \frac{1}{n} \sum_{i=1}^n |y_{obs} - y_{pred_{SARIMA}}|$$

$$Error_{HW} = \frac{1}{n} \sum_{i=1}^n |y_{obs} - y_{pred_{HW}}|$$

Donde  $y_{obs}$  : Son las observaciones de la serie original a partir del 2018

Donde  $y_{pred}$  : son las predicciones de los modelos

Donde  $n$  : es el número de observaciones a predecir

**ERROR\_HW=1685.5**

**ERROR\_SARIMA= 1786.9**

De los resultados obtenidos de los errores medios absolutos, se puede confirmar que las predicciones por el método de Holt-Winter son mejores que por el modelo SARIMA propuesto. Lo anterior se puede explicar ya que las predicciones del modelo SARIMA suponen normalidad de los residuales y es algo con lo que no cuenta nuestro modelo. Sin embargo la diferencia de predicción entre ambos modelos no es más de 100 toneladas de diferencia, que entre cientos de miles que se observan, esto no es tan penalizable.

## Conclusiones

---

Con el análisis descriptivo y el análisis estadístico sobre la serie de tiempo del volumen de aguacate en Estados Unidos pudimos conocer, durante el periodo estudiado, el comportamiento del volumen de aguacate vendido, lo que incluye su tendencia, varianza y correlación. Además, este estudio nos permitió notar que es una serie de tiempo que tiene un ciclo estacional marcado a lo largo del tiempo. Lo anterior nos permitió confirmar los impactos que los eventos deportivos en Estados Unidos tienen sobre el consumo y venta de este producto agrícola.

En primera instancia notamos que todos los correlogramas de las transformaciones al aplicarles las diferencias eran de poca ayuda para poder proponer modelos, porque parecieran provenir de ruidos blancos. Pero no era así, ya que al aplicar estas diferencias nos dimos cuenta que tanto "D" como "d" tendrían que ser 1, y eran parámetros que explicaban muy bien el modelo y sólo tendríamos que "jugar" con los valores de  $p, q$ .

Podemos concluir que es una serie con varianza no constante y que tiene cierta tendencia, lo cual se puede deber a que a medida que pasa el tiempo, el evento "Super Bowl" es más visto en Estados Unidos y el guacamole es más cotizado, por ende se demanda un mayor volumen de aguacate.

Otro aspecto que podemos observar es que es una serie algo complicada porque no cumple con algunos supuestos (para los modelos estudiados en clase), principalmente no cumple la normalidad de sus residuales en ninguno de los modelos propuestos. Lo anterior es importante porque la normalidad es el supuesto en que nos basamos para hacer predicciones y lo más importante en una serie de tiempo es poder hacer pronósticos del comportamiento de esta.

Por último, observamos que existen series de tiempo que no son tan fáciles de tratar, ya sea por su varianza, significancias de parámetros, datos atípicos o incluso los supuestos de sus residuales, por lo cual es necesario tener métodos/modelos alternativos que nos permitan analizar los datos. En este caso el método de Holt-Winters fue la manera alternativa en la que pudimos trabajar la serie de tiempo y sus predicciones, las cuales fueron un poco más acertadas que las de nuestro modelo propuesto con el logaritmo natural de nuestros datos.

## Bibliografía

---

Serie obtenida de

<https://www.kaggle.com/neuromusic/avocado-prices>