

## INTRODUCCION

La cerveza llegó a Australia a principios de la colonización británica en el HMS Endeavour dirigido por el capitán James Cook a mediados de 1770.

Por aquel entonces la enfermedad más común entre los marineros que se aventuraban a viajes largos era el escorbuto; ocasionada por una deficiencia en el consumo de vitamina C al no tener acceso a alimentos frescos, algunos médicos de la época creían que podía ser curada a través de un cambio en la dieta que incluía el consumo de dosis diarias de *wort* (mosto de cerveza) o cualquier licor que los marineros bebieran mezclado con jugos de cítricos. Es por esto que, por recomendación de Nathaniel Hulme<sup>1</sup>, en todas las expediciones (en total se realizaron 3) se cargó el navío con barriles que contuvieran diferentes preparados a base de cítricos; además Nathaniel también le brindó a instrucciones a Joseph Banks<sup>2</sup> sobre cómo preparar cerveza cuando el agua que llevaban abordo no estuviera en condiciones óptimas para ser consumida: se debía mezclar el agua con melaza, un poco de la mejor trementina -también conocida como aguarrás; pese a ser tóxica, en el área médica se utilizaba para tratar heridas e infecciones nasales y/o de garganta- de Quios y levadura, así cualquier peste que habitara en el agua se vería eliminada gracias al proceso de fermentación.

Si bien durante gran parte del primer siglo de asentamiento se dice que la bebida preferida fue el ron, la elaboración de cerveza también tuvo bastante apoyo ya que se vio que era una opción mucho más sana que el whisky, el ron y otras bebidas que ellos llamaban “espirituosas “. Los primeros cerveceros oficiales de los que se tiene información son John Boston en 1796 y James Squire en 1804; el primero la fabricaba de manera similar a como lo hacían los marineros, mientras que el segundo logró cultivar lúpulo con éxito (nadie lo había conseguido aún en Australia) y así hacer que su producto tenga un sabor diferente.

Aun siendo una de las primeras bebidas en el territorio, la cerveza no tomó popularidad sino hasta que Australia empezó a convertirse en una ciudad y sociedad más urbana y cosmopolita. Fue en la década de los 80 donde comenzó a tener un mayor auge su industria ya que la tecnología se vio mejorada con creces y permitió que los productores pudieran aumentar su capacidad de producción. En los '90 ya empezó a ser una parte prácticamente indispensable del entretenimiento australiano dado que por un lado los bares comenzaron a atraer nuevos consumidores al ofrecer variedades de estilos y marcas; y por

---

<sup>1</sup> (1732-1807) Médico británico que defendía el uso de la cerveza y el wort como antiescorbúticos.

<sup>2</sup> (1743-1820) Naturalista y botánico a bordo del HMS Endeavour.

otro, los eventos deportivos y musicales habían empezado a ganar popularidad lo que creó así una cultura de celebración y socialización alrededor de la bebida ambarina. En este sentido, la cerveza se asoció con la identidad australiana: fresca, directa y sin pretensiones.

Es por esto que se considera interesante el análisis de la producción cervecera en Australia durante los años 1988 y 1993.

### **ORIGEN DE LOS DATOS**

Los datos para este trabajo fueron obtenidos a través de la plataforma Kaggle, brindados por el usuario ShenbagaKumarS.

## ANALISIS DESCRIPTIVO

Como se mencionó en la sección anterior, este trabajo se centrará en analizar la serie de producción de cerveza en Australia para así poder determinar un modelo que se amolde de manera adecuada a los datos y predecir valores de producción futuros. En particular, la información está registrada de manera mensual abarcando de Enero de 1988 a Diciembre de 1993.

Se examina la serie acompañada por su tabla de Buy-Ballot para determinar su comportamiento.

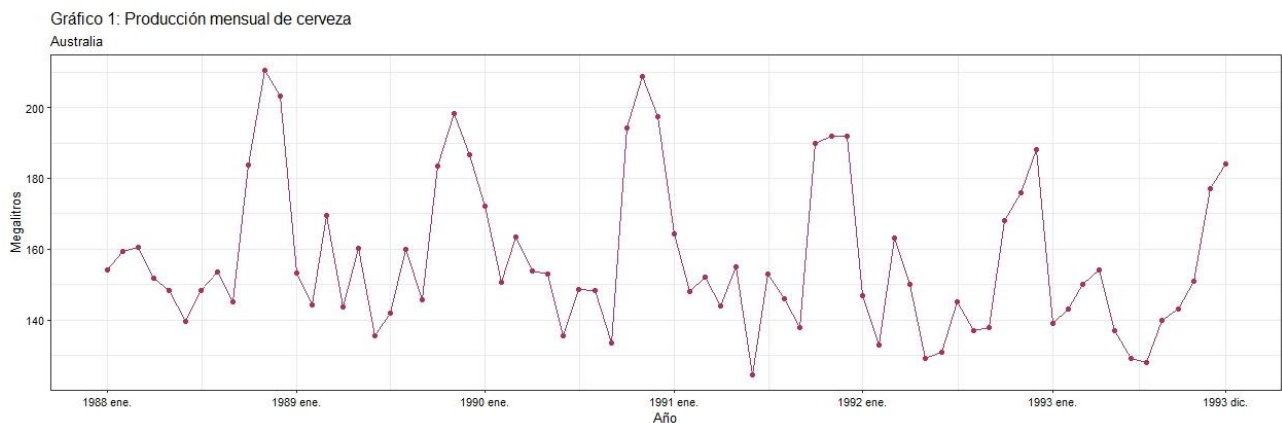


Tabla 1: Tabla de Buy - Ballot

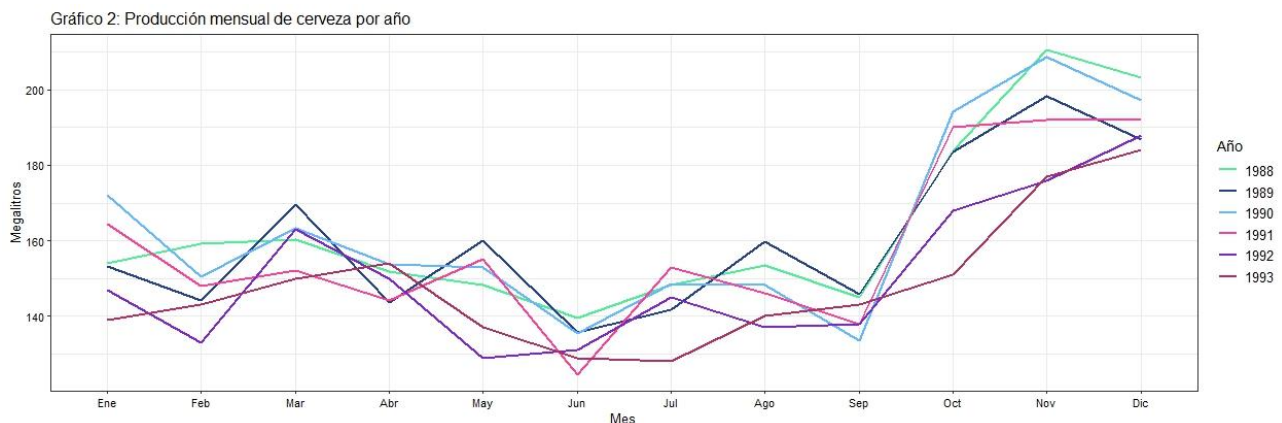
Mes\Año	1988	1989	1990	1991	1992	1993	Promedio	Total
Enero	154.1	153.3	172	164.4	147	139	154.9	929.8
Febrero	159.3	144.3	150.6	148.1	133	143	146.3	878.3
Marzo	160.4	169.6	163.3	152	163	150	159.7	958.3
Abril	151.9	143.7	153.7	144.1	150	154	149.5	897.4
Mayo	148.4	160.1	152.9	155	129	137	147	882.4
Junio	139.6	135.6	135.5	124.5	131	129	132.5	795.2
Julio	148.2	141.8	148.5	153	145	128	144	864.5
Agosto	153.5	159.9	148.4	146	137	140	147.4	884.8
Septiembre	145.1	145.7	133.6	138	138	143	140.5	843.4
Octubre	183.7	183.5	194.1	190	168	151	178.3	1070.3
Noviembre	210.5	198.2	208.6	192	176	177	193.7	1162.3
Diciembre	203.3	186.8	197.3	192	188	184	191.9	1151.4
Promedio	163.1	160.2	163.2	158.2	150.4	147.9		
Total	1958	1922.5	1958.5	1899.1	1805	1775		

Se puede ver que en los meses más calurosos es cuándo se produce más cerveza ya que al ir finalizando cada año la serie toma valores altos. En específico, parece ser que Noviembre se lleva el record en elaboración; esto puede deberse a que durante ese mes se desarrolla la *Melbourne Cup*, una carrera de caballos tan importante que dicen que “paraliza a todo el país”, así las reuniones en bares son accionar frecuente al igual que las apuestas y ¿Quién no se toma una pinta para avivar el espíritu competitivo? Esto sería indicio de un comportamiento estacional de periodo 12.

También se tiene que en Marzo se realiza el Festival de Moomba para celebrar la identidad y la variedad cultural que está presente en la ciudad de Melbourne, lo cual explicaría por qué, pese a no ser un mes tan cálido, la producción de la bebida ambarina es mayor.

Finalmente se observa que el mínimo de producción se da entre los meses de Mayo y Julio, siendo Junio el mes en el que por lo general menos cerveza se produce; destaca Septiembre del '90 por ser el mes de ese año que menos producción tuvo.

Si se analiza la serie por año, se observa que en 1990 se tuvo el pico más alto de producción tanto total como promedio mientras que 1993 fue el año en que menos se produjo cerveza. Se presenta a continuación el gráfico anual de la serie para ver esto con más detalle:



Ahora bien, antes de proponer modelos para esta serie se deben determinar dos cosas importantes: (1) si debe hacer una transformación; (2) si se debe hacer alguna diferenciación.

En primer lugar se hará un análisis de dispersión de la serie:

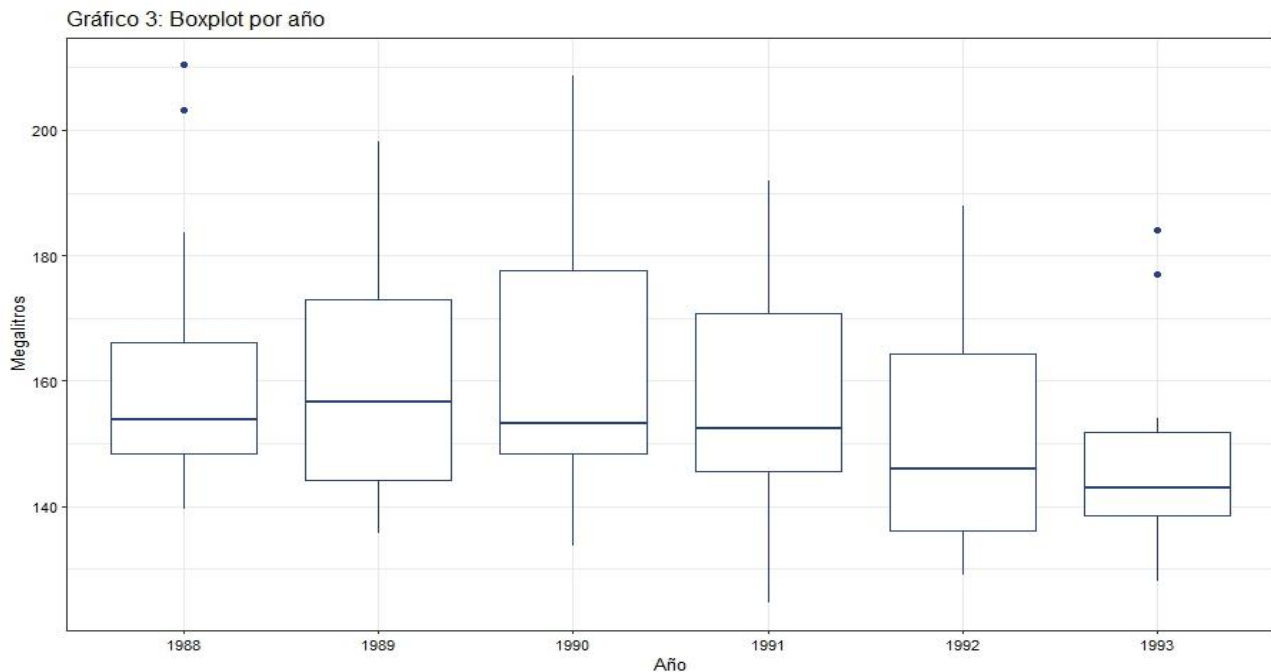


Tabla 2: Desvíos estándares

Año	1988	1989	1990	1991	1992	1993
Desvío	23.2	20.2	24.6	22.1	19.1	17.2

Tabla 3: Coeficientes de variación

$\lambda$	-4	-3	-2	-1	0	1	2
Coeficiente	$\approx 0$	$\approx 0$	$\approx 0$	0.125	0.025	0.135	0.282

Si bien los coeficientes de variación obtenidos a partir de transformaciones de Box-Cox realizadas a la variable “producción” indican que una transformación estaría siendo necesaria y se observa que cada año tiene una variabilidad distinta, las medianas y los desvíos de cada año no parecen distar mucho; es decir, sí son distintas pero se mantienen dentro de un rango común de valores. Es por esto que, para no complicar las interpretaciones, se decide no realizar ninguna transformación.

Por otra parte, se realizan test para ver si tanto la parte regular como la estacional necesitan ser diferenciadas.

#### 1) Test para la parte regular

Se aplica el test de *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) para evaluar estacionariedad (la hipótesis nula es que la serie es estacionaria); la probabilidad asociada a la estadística del test resultó ser  $< 0.1$ , por lo que no es necesaria una diferenciación regular.

## 2) “Test” para la parte estacional

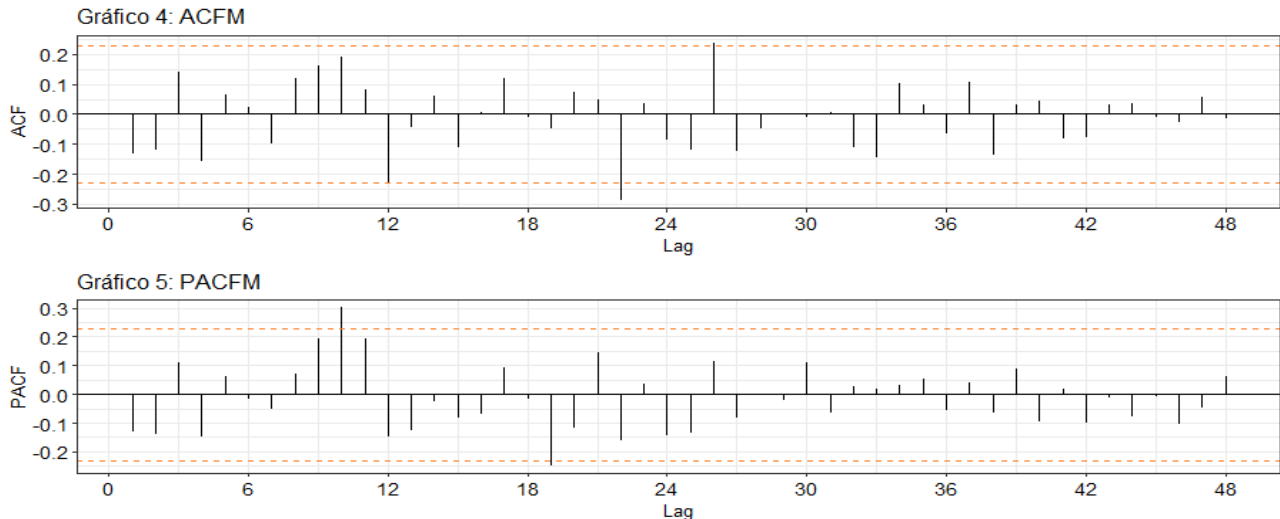
Se calcula  $F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right)^3$  y se concluye realizar una diferenciación estacional si es mayor a 0.64; en este caso, resultó que  $F_s = 0.904$  por lo que se hace una diferenciación. Se vuelve a calcular  $F_s$ , a la variable diferenciada, y al ser  $< 0.64$  se determina que no es necesaria ninguna otra diferenciación.

---

<sup>3</sup> Definida como “strength of seasonality”,  $S_t$  y  $R_t$  son parte de la descomposición de una serie:  $Y_t = T_t + S_t + R_t$  (Hyndman y Athanasopoulos, 3ra. Edición – Cap.4)

## **IDENTIFICACIÓN DEL MODELO**

Para plantear modelos candidatos se analizarán las características de las funciones de autocorrelación muestral y autocorrelación parcial muestral de la variable “producción” diferenciada.



La parte estacional se determina mirando, valga la redundancia, los rezagos estacionales; como el periodo de esta serie es 12, se miran los rezagos 12, 24, 36 y 48. En la ACF se destaca el rezago 12 por ser significativo al límite, mientras que en la PACF los rezagos estacionales decaen exponencialmente; esto lleva a que se proponga un proceso autorregresivo de orden 1, también se probará no poner ningún proceso en la parte estacional debido a lo antes mencionado sobre el rezago 12.

Para la parte regular se analizan los rezagos del 1 al 11, se observa en la PACF que el rezago 10 es significativo pero esto puede deberse a la interacción entre lo regular y lo estacional. Ya que no se observa ninguna característica destacable de los procesos conocidos para poder definir cuál asignar a esta parte, se decide probar combinaciones y ver cuál se ajusta mejor.

A continuación se muestran los modelos candidatos para los datos y sus criterios de información; a la par de estos modelos también se ajustó uno a través de un método automático y resultó coincidir con uno de los propuestos -el último-.

Tabla 4: Modelos candidatos.

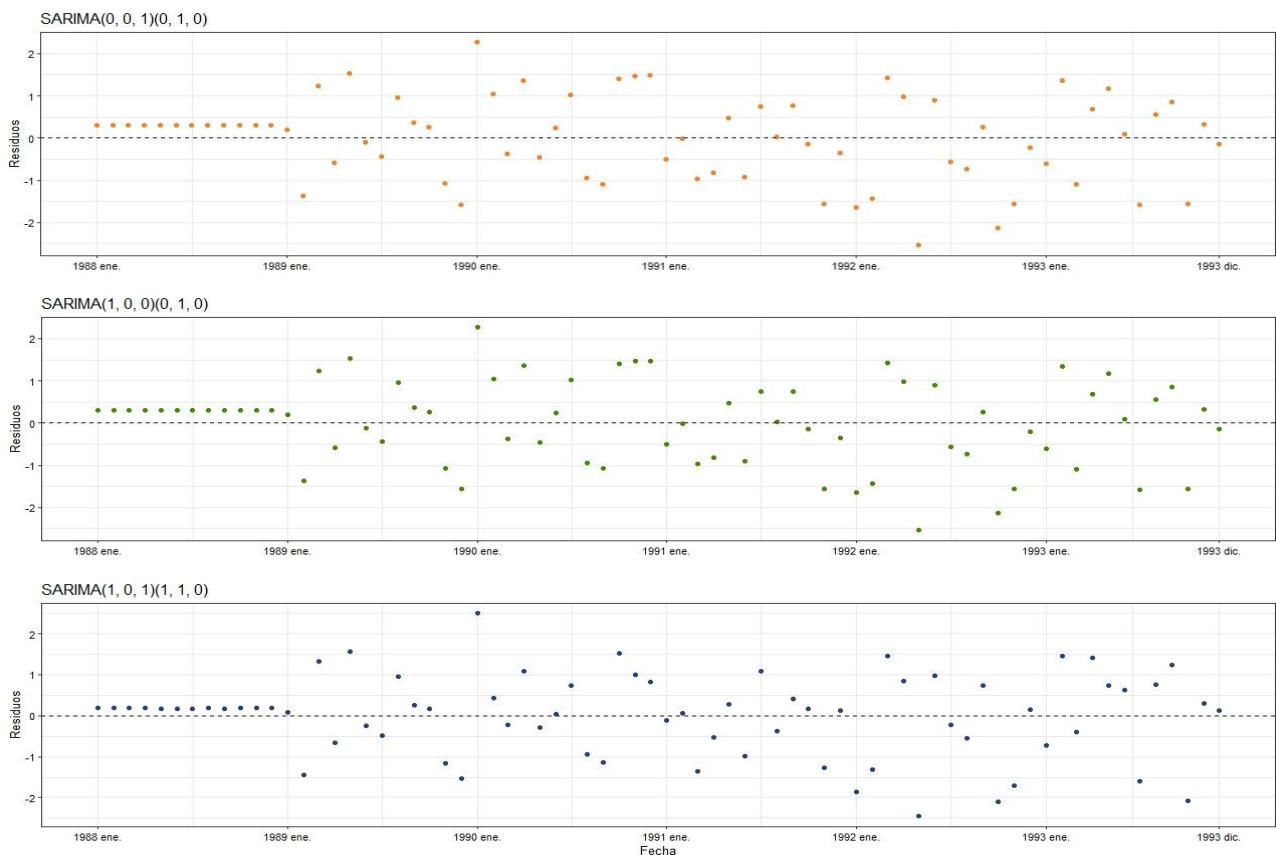
Modelo	AIC	AICc	BIC
$SARIMA(0, 0, 1)(0, 1, 0)_{12}$	454.36	454.58	458.55
$SARIMA(1, 0, 0)(0, 1, 0)_{12}$	454.37	454.58	458.56
$SARIMA(1, 0, 1)(0, 1, 0)_{12}$	456.36	456.79	462.65
$SARIMA(0, 0, 1)(1, 1, 0)_{12}$	455.16	455.58	461.44
$SARIMA(1, 0, 0)(1, 1, 0)_{12}$	455.16	455.58	461.44
$SARIMA(1, 0, 1)(1, 1, 0)_{12}$	451.48	452.21	459.86

Dado que son muchos modelos los que se propusieron, se decidió filtrar los 3 con menor valor en el criterio de información de Akaike para modelos ARIMA (también resultan ser los que tienen valores menores en los demás criterios) y analizar si ellos cumplen los supuestos necesarios para ser admisibles.

### → EVALUACIÓN DE SUPUESTOS

- 1) La esperanza de los residuos es cero
- 2) La variancia de los residuos es constante

Gráfico 6: Residuos estandarizados vs. Tiempo



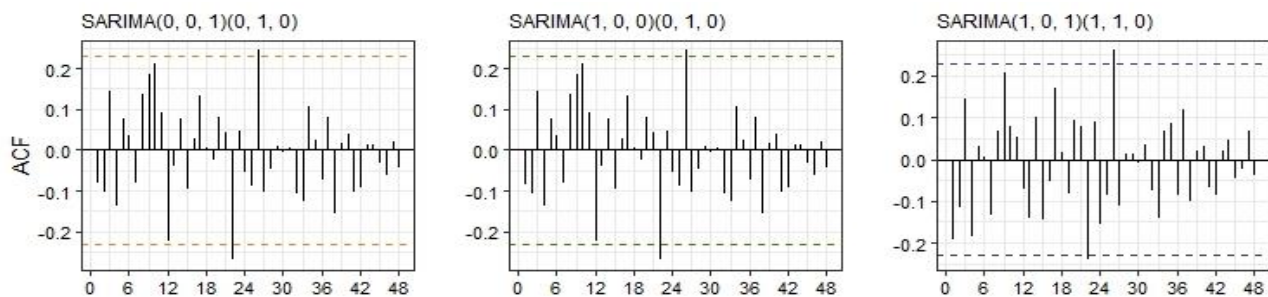


Se puede ver que los residuos de los 3 modelos se distribuyen alrededor del cero y no se observa ningún patrón en su comportamiento, con esto se puede decir que los primeros dos supuestos se cumplen adecuadamente.

### 3) Los residuos están incorrelados

Tabla 5: Test de Portmanteau					
Modelo	Valores\ Rezagos	6	12	18	24
<b>SARIMA(0, 0, 1)(0, 1, 0)<sub>12</sub></b>	Estadística	4.71	16.5	19	25.3
	Probabilidad asoc.	0.453	0.124	0.33	0.337
<b>SARIMA(1, 0, 0)(0, 1, 0)<sub>12</sub></b>	Estadística	4.74	16.5	19	25.3
	Probabilidad asoc.	0.448	0.123	0.328	0.336
<b>SARIMA(1, 0, 1)(1, 1, 0)<sub>12</sub></b>	Estadística	7.56	13.3	19.2	27.3
	Probabilidad asoc.	0.056	0.15	0.204	0.161

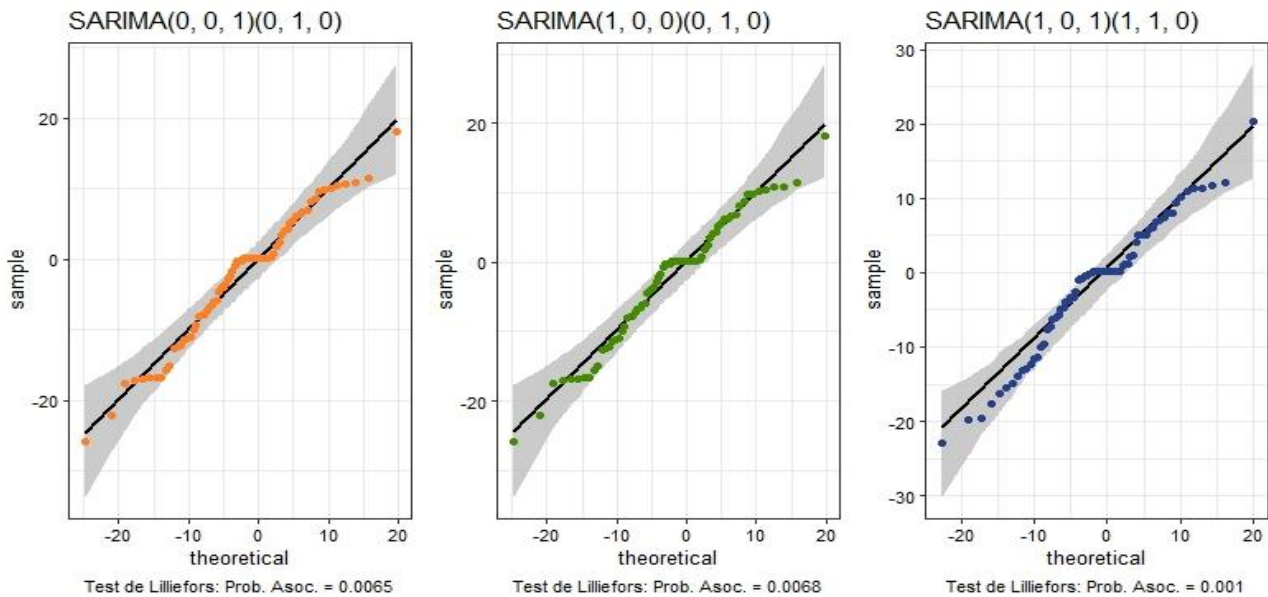
Gráfico 7: Funciones de autocorrelación y autocorrelación parcial



Si bien en los 3 gráficos se observa que los rezagos 22 y 26 son significativos, el resto se mantiene dentro del rango de significación; además el test de Portmanteau muestra que los residuos no se encontrarían relacionados, de esta forma se verifica el tercer supuesto.

#### 4) Los residuos se distribuyen Normal

Gráfico 8: Gráfico probabilístico normal – Test de Lilliefors



Pese a que los residuos se mantienen dentro de las bandas de confianza, no siguen de manera “estricta” la línea teórica y el test de Kolmogorov-Smirnov-Lilliefors indica que no son normales; por lo tanto el supuesto no se cumple. Aun así, se sigue adelante con estos modelos ya que el único inconveniente que se presenta ante esta falta es que los intervalos de confianza de los estimadores de los parámetros serán más amplios, es decir se tendrán desvíos estándares mayores que los que se obtendrían bajo el cumplimiento de la normalidad.

#### → ESTADO DE LOS PARAMETROS

Otro detalle a tener en cuenta a la hora de determinar si un modelo es aceptable o no es la significación de sus parámetros y si estos cumplen con las condiciones de invertibilidad y/o estacionariedad.

Tabla 6: Comprobación de parámetros			
Modelo	$SARIMA(0, 0, 1)(0, 1, 0)_{12}$	$SARIMA(1, 0, 0)(0, 1, 0)_{12}$	$SARIMA(1, 0, 1)(1, 1, 0)_{12}$
Estimaciones	$\hat{\theta}_1 = -0.0346$	$\hat{\phi}_1 = -0.0335$	$\hat{\phi}_1 = 0.988$ $\hat{\theta}_1 = -0.9312$ $\hat{\Phi}_1 = -0.2886$
¿Significativo?	No	No	Si Si Si
Condiciones	MA invertible	AR estacionario	ARMA invertible y estacionario $AR_{estacional}$ estacionario
Correlaciones	-	-	$(\hat{\phi}_1, \hat{\theta}_1) = -0.76$ $(\hat{\phi}_1, \hat{\Phi}_1) = 0.025$ $(\hat{\theta}_1, \hat{\Phi}_1) = -0.198$

Antes que nada hay que aclarar dos puntos importantes: (1) No se muestran las correlaciones de los primeros 2 modelos ya que tienen un único parámetro y la correlación entre un parámetro y él mismo es 1, por lo que no aporta información indispensable; (2) La no significancia de las estimaciones de los primeros dos modelos puede estar asociada a la falta de normalidad de ambos.

Siguiendo con el análisis, es sabido que correlaciones elevadas -tanto positivas como negativas- entre parámetros diferentes son sinónimos de sobreajuste; en este caso, pese a la alta correlación entre  $\hat{\phi}_1$  y  $\hat{\theta}_1$  ambos resultan ser significativos, por ello se decide simplemente aceptar el valor y continuar.

## PREDICCIONES

Ahora que se vio que -casi- todo está en orden, hay que determinar cuál modelo será el elegido y para esto se evalúa la capacidad predictiva de cada uno de ellos: aquel con el menor error porcentual absoluto medio se llevará la victoria.

Tabla 7: Error porcentual absoluto medio		
Modelo	MAPE	Variancia
<b><i>SARIMA(0, 0, 1)(0, 1, 0)<sub>12</sub></i></b>	8.85	108.32
<b><i>SARIMA(1, 0, 0)(0, 1, 0)<sub>12</sub></i></b>	8.85	108.33
<b><i>SARIMA(1, 0, 1)(1, 1, 0)<sub>12</sub></i></b>	6.35	97.46

Como se puede ver, el menor error lo tiene el modelo 3 y también es el de menor variancia. ¿Qué tan buenas serán sus predicciones? Se procede a pronosticar 6 meses hacia adelante desde la última fecha que utilizó en la serie con el modelo:

$$(1 + 0.28B^{12})(1 - 0.98B)(1 - B^{12})Z_t = (1 + 0.93B)a_t$$

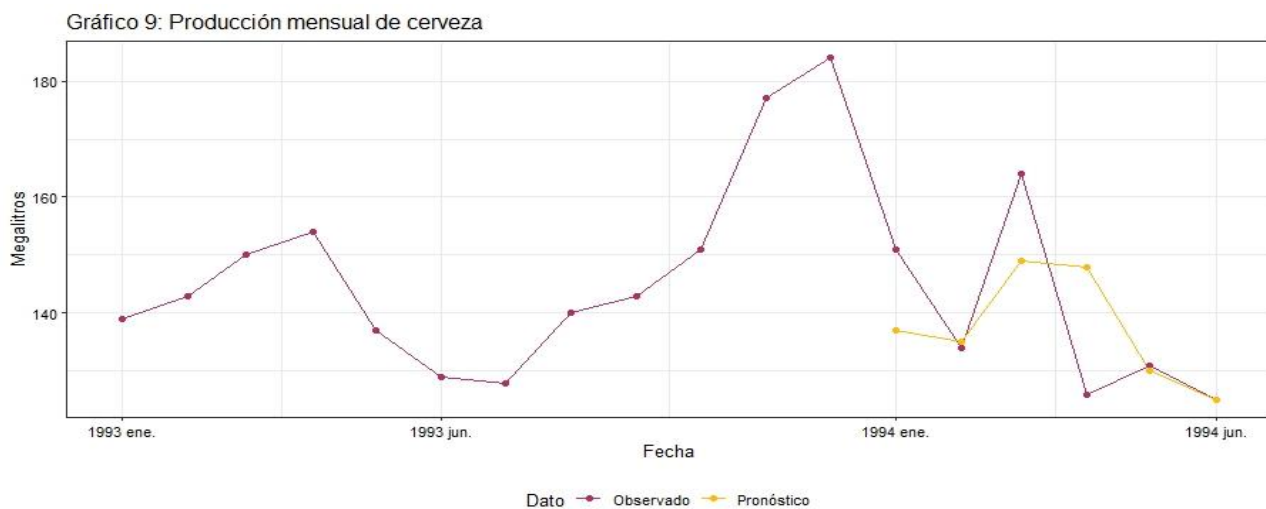


Tabla 8: Pronósticos y sus intervalos			
Mes	Observado	Pronóstico	Intervalo
<b>Enero</b>	151	137	(117.3; 155.9)
<b>Febrero</b>	134	135	(116.1; 154.8)
<b>Marzo</b>	164	149	(129.8; 168.6)
<b>Abril</b>	126	148	(128.9; 167.7)
<b>Mayo</b>	131	130	(110.7; 149.7)
<b>Junio</b>	125	125	(105.6; 144.6)

Teniendo una producción media de 138.5 megalitros para los meses que fueron pronosticados, se puede observar que toda aquella observación que sea mayor a ese valor medio es subestimada por el modelo, mientras que aquellas -la del mes de abril- que sea menor al valor medio es sobreestimada. De todas formas, los intervalos -calculados con un 95% de confianza- cubren el valor real de cada mes pero cada uno con un mayor desvío.

## **CONCLUSION**

Tomando el período de 1988 a 1993 se advirtió que la serie presentaba una tendencia decreciente de producción promedio anual, salvo por la crecida en los '90, junto con un comportamiento estacional de 12 meses donde se presentaba un crecimiento en la producción en los últimos meses de cada año y un decrecimiento en los meses de frío.

Todo el análisis se realizó con el fin de realizar predicciones respecto a la producción cervecera de los meses que siguen; en específico, el interés estuvo en pronosticar los primeros 6 meses de 1994. Las mismas -las predicciones- fueron efectuadas con un modelo  $SARIMA(1, 0, 1)(1, 1, 0)_{12}$  que no solo verificó todos los supuestos importantes de este tipo de procesos sino que, de los modelos postulados y con los que se lo comparó, fue el que menor MAPE obtuvo: 6.35%.

Cabe recordar que no se realizó ninguna transformación para estabilizar la variancia pese a los resultados brindados por el test de Box-Cox. Si bien esto no conllevó ningún conflicto destacable en el resto del examen de la serie, existe la posibilidad de que no haber omitido los resultados del test guiara a mejores métricas y, por ende, mejores resultados que los obtenidos; por ello se sugiere al público que considere esta opción en futuros análisis.

## **BIBLIOGRAFÍA**

- **WEI, W. W. S.** (2006): "*Time Series Analysis, Univariate and Multivariate Methods*". (2nd edition) Addison-Wesley Publishing Company.
- **HYNDMAN, R.J., & ATHANASOPOULOS, G.** (2021). "*Forecasting: principles and practice*". (3rd edition). Melbourne, Australia