Agrofaz – Journal of Environmental and Agroecological Sciences vol 8 (1): 68-75

Agrofaz
Journal of Environmental and Agroecological Sciences

Modelo estadístico para pronóstico de precios del melón cantaloupe (*cucumis melo* L.)

Statistical model for forecasting cantaloupe (Cucumis melo L.) prices

Damián Torres-Hernández[1*], Luis Maconetzín Isidro Requejo[1], Yasmín Ileana Chew Medinaveitia[1], José de Jesús Espinoza Arellano[2], Arturo Gaytán Mascorro[3], Antonio Anaya Salgado[1]

[1]Campo Experimental La Laguna-INIFAP. Boulevard José Santos Valdés, 1200 poniente, Matamoros, Coahuila, México. CP. 27460.
[2]Universidad Autónoma de Coahuila. Boulevard Revolución 153 oriente, Torreón, Coahuila. CP. 27000.
[3]Universidad Autónoma Agraria Antonio Narro, Unidad Laguna. Periférico Raúl López Sánchez, Torreón, Coahuila. CP. 27054.

*Autor para correspondencia: torres.damian@inifap.gob.mx

RESUMEN

En la presente investigación estadística se utiliza la metodología de Box y Jenkins para modelar series de tiempo con inestabilidad en la varianza y media no estacionaria. El modelo se resume en tres etapas de análisis: identificación, estimación y pronóstico, evaluado mediante contrastes de raíces unitarias, métricas de bondad de ajuste y medidas de error del pronóstico, también se empleó la metodología de entrenamiento y validación (70% y 30% de los datos observados, respectivamente). Los datos corresponden al precio promedio mensual del melón cantaloupe # 9, al mayoreo en central de abasto, expresados en pesos nominales mexicanos por kilogramo ($MXN kg-1), publicados por el Sistema Nacional de Información e Integración de Mercados (SNIIM). El modelo estacional seleccionado fue un ARIMA(1,1,1)X(1,0,0)_(S=12).

Palabras clave: Economía, Agrícola, Estadística, Modelo, Series de tiempo.

ABSTRACT

This statistical study employs the Box-Jenkins methodology to model time series characterized by variance instability and a non-stationary mean. The modeling process comprises three analytical stages: identification, estimation, and forecasting, and is evaluated using unit root tests, goodness-of-fit metrics, and forecast error measures; a training/validation split (70% and 30% of the observed data, respectively) was also utilized. The data represent the monthly average wholesale price of size #9 cantaloupes at central markets, expressed in nominal Mexican pesos per kilogram ($MXN kg^{-1}), published by the National System for Market Information and Integration (SNIIM). The selected seasonal model was an ARIMA model $(1,1,1)\mathrm{X}(1,0,0)_{S=12}$.

Keywords: Economics, Agricultural, Statistics, Model, Time series.

INTRODUCCIÓN

El melón se considera como uno de los principales cultivos de la agricultura comercial en la Región Lagunera. De acuerdo con datos del Servicio de Información Agroalimentaria y Pesquera (SIAP), durante 2025 se cosecharon 3,254 hectáreas con un rendimiento de 37 toneladas por hectárea. La derrama económica que genera es de 599 millones de pesos, además de que cada hectárea genera hasta 200 jornales por temporada, lo cual nos da una suma de 650,800 jornales, que equivale a 205 millones de pesos que sirve de ingreso a más de nueve mil familias productoras de melón (SIAP, 2026). La mayor parte de la producción se destina al abasto del mercado nacional, lo que ocurre principalmente en los meses de mayo a julio, aunque puede prolongarse hasta noviembre. La situación anterior hace relevante el monitoreo de precios de mercado en beneficio de los productores agrícolas de la región para continuar con la producción. El objetivo

del presente análisis fue aplicar un modelo estadístico de series de tiempo para pronosticar el precio nominal mensual del melón cantaloupe.

MATERIALES Y MÉTODOS

Los datos considerados corresponden al precio promedio mensual del melón cantaloupe # 9, al mayoreo en central de abasto, expresados en pesos nominales mexicanos por kilogramo ($MXN kg^{-1}), publicados por el Sistema Nacional de Información e Integración de Mercados (SNIIM). El periodo de análisis es de noviembre de 2021 a diciembre del 2025, generando una serie de tiempo con 50 observaciones mensuales, divida en dos conjuntos de datos, considerando el primero a partir de noviembre de 2021 hasta septiembre del 2024 para entrenamiento (70% de los datos observados) del modelo, y el segundo a partir de octubre del 2024 hasta diciembre de 2025 para la validación (30% de los datos observados) del modelo, ambos conjuntos expresados en pesos por kilogramo calculado ($/kg^{-1}). La modelación estadística se hizo con el programa R.

Una serie de tiempo $\{Y_t\}_{t=1}^{T}$ es un proceso estocástico $Y = \{Y_t(w) | t \in C\}$. Este proceso caracteriza una sucesión de variables aleatorias (estocásticas) que evolucionan en función de otra variable, generalmente el tiempo. Cada una de las variables aleatorias del proceso tiene su propia función de distribución de probabilidad ($FX(x) = P(X \leq x)$). En la mayoría de los procesos estocásticos cada resultado depende de lo que sucedió en etapas anteriores del proceso. Asumiendo los precios promedio mensuales del melón como una serie temporal, estos pasan a ser un conjunto de Y_t observaciones de un proceso discreto ordenadas y equiespaciadas en el tiempo, que para cierto periodo de tiempo t, la observación Y_{t-h} se llama el $h-ésimo$ desfase de Y_t.

Uno de los enfoques más utilizados para analizar series de tiempo es el método de George E. P. Box y Gwilym Jenkins (1976), mediante un modelo Autorregresivo e Integrado de Media Móvil (ARIMA). Este método puede resumirse en tres etapas de análisis: identificación, estimación y pronóstico (Martínez-Damián & Brambila-Paz, 2023). El modelo ARIMA se alimenta de la serie de tiempo de mínimo 50 datos (Box and Jenkins, 1976) para generar predictores que replican los datos longitudinales y permite la predicción de valores futuros de la serie histórica sin necesidad de tomar en cuenta otros factores que afectan la evolución de la variable en el tiempo, característica que los clasifica como modelos no estructurales (Sánchez, 2008).

La construcción del modelo tiene dos componentes siendo el primero autorregresivo (AR) de orden p, con parámetro ϕ, el cual se expresa de la siguiente forma:

$$Y_t = \mu + \sum_{i=1}^{p} +\phi_i Y_{t-i} + \varepsilon_t$$

El segundo componente es la media móvil (MA) de orden q, con parámetro θ, el cual se expresa de la siguiente forma:

$$Y_t = \mu + \sum_{i=1}^{q} +\theta_i Y_{t-i} + \varepsilon_t$$

Ambos componentes para la serie de tiempo $\{Y_t\}_{t=1}^{T}$ satisfacen un proceso ARIMA (p, d, q), expresado como $\phi_p(B)\nabla^d Y_t = \theta_q(B)\varepsilon_t$. Dicho proceso requiere que la serie de tiempo no tenga tendencia determinística (en la media) y estocástica (en la varianza), ni estacionalidad (conseguir estacionariedad), de lo contrario será necesario efectuar diferenciación no estacional (d) seguida de diferenciación estacional (D). Bajo esta característica surge el modelo ARIMA estacional, expresado por $(p, d, q)X(P, D, Q)_S$ $\phi(B)\Phi(B^S)\nabla_S^D\nabla^d Y_t = c + \theta(B)\Theta(B^S)\varepsilon_t$ (Guerrero, 1991).

Donde B es el operador de retraso ($B^r Y_t = Y_{t-1}$), Y_t es la serie de tiempo de la variable Y en el tiempo t, el polinomio autorregresivo (AR) de la parte no estacional, de orden "p", se simboliza $\phi(B)$ y se desarrolla: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$; $\theta(B)$ es el polinomio de promedios móviles (MA) de la parte no estacional, de orden "q", siendo $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$; mientras que $\Phi(B)$ es el polinomio autorregresivo (AR) de la parte estacional, de orden "P", esto es: $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_P B^P$; y $\Theta(B)$ es el polinomio de promedios móviles (MA) de la parte estacional, de orden "Q", y se desarrolla $\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_Q B^Q$; $\nabla^d =$

$(\mathrm{Y_t} - \mathrm{Y_{t-1}})^{\mathrm{d=1}}$ refiere al número de diferencias (d) para la lograr la estacionariedad de la serie en la parte no estacional (p, d, q); y $\nabla^{\mathrm{D}}\mathrm{Y_t^S} = (1 - \mathrm{B^S})^{\mathrm{D=1}}$ representa el número de diferencias (D) en la parte estacional del modelo $(P, D, Q)_S$; ε_t hace referencia a los términos de error aleatorio (factores adyacentes que el modelo no explica) o también referido como la representación de la combinación lineal de procesos de ruido blanco de variables aleatorias independientes e idénticamente distribuidas con media igual a cero y varianza constante $\varepsilon_\mathrm{t} \sim \mathrm{i.i.d.}\,(0, \sigma^2)$.

La estacionariedad o no estacionariedad estadística planteada por Dickey y Fuller (1981), en la serie de tiempo se verifica mediante la prueba de raíz unitaria de Dickey Fuller Aumentada (ADF) para evaluar la existencia de una autocorrelación serial bajo las hipótesis H0: la serie no es estacionaria y tiene raíz unitaria vs Ha: la serie es estacionaria y no tiene raíz unitaria (Pankratz, 1983).

La parsimonia del modelo elegido se basa en los valores menores del Criterio de Información de Akaike (AIC) y el Criterio Bayesiano de Shwartz (BIC). Una vez elegido el modelo y hecha la estimación de parámetros, se realiza la verificación de supuestos mediante el análisis estadístico de los residuales y así tener la seguridad estadística al realizar los pronósticos. Para constatar lo anterior, la prueba de Jarque-Bera (1980), se utiliza para probar hipótesis de normalidad teniendo como hipótesis nula la distribución normal sobre el ruido blanco de los datos, mientras que la hipótesis alternativa señala lo opuesto. La implicación de este supuesto es garantizar que los intervalos para los pronósticos sean confiables.

La prueba de Ljung-Box (1978) estima un valor de probabilidad teniendo que ser mayor al nivel de significancia ($\boldsymbol{\alpha}$=$\mathbf{0.05}$) en cada uno de sus rezagos para no rechazar la hipótesis nula, señalando que los datos se distribuyen de forma independiente (no hay autocorrelación de los residuos). La prueba de Brock, Dechert y Scheinkman (1996), se utiliza para contrastar la hipótesis nula de que la secuencia es aleatoria, independiente e idénticamente distribuida (iid).

En la presente investigación se desarrolla un modelo ARIMA estacional para pronosticar precios de mercado, aplicado a una serie temporal con tendencia determinística y estocástica, así como patrones estacionales. Para dicho modelo se utilizan los precios nominales mensuales del melón, evaluando el poder predictivo mediante estacionalidad de la serie (contrastes de raíces unitarias), pruebas de bondad de ajuste, así como y medidas de error del pronóstico, así como la metodología de entrenamiento y validación (70% y 30% de los datos observados, respectivamente) para evaluar el modelo en el programa estadístico R.

RESULTADOS Y DISCUSIÓN

La serie de tiempo anualizada (Figura 1) muestra diferentes niveles de precios de mercado teniendo un comportamiento distinto mes con mes, ya que el incremento de los precios se observa a partir del mes septiembre alcanzando los precios más altos en los meses de noviembre y diciembre para posterior comenzar un descenso de los mismos a partir de enero, siendo los meses de mayo hasta el mes de agosto los precios más bajos del año. Lo anterior permite determinar que la serie tiene inestabilidad en la varianza y media no estacionaria.

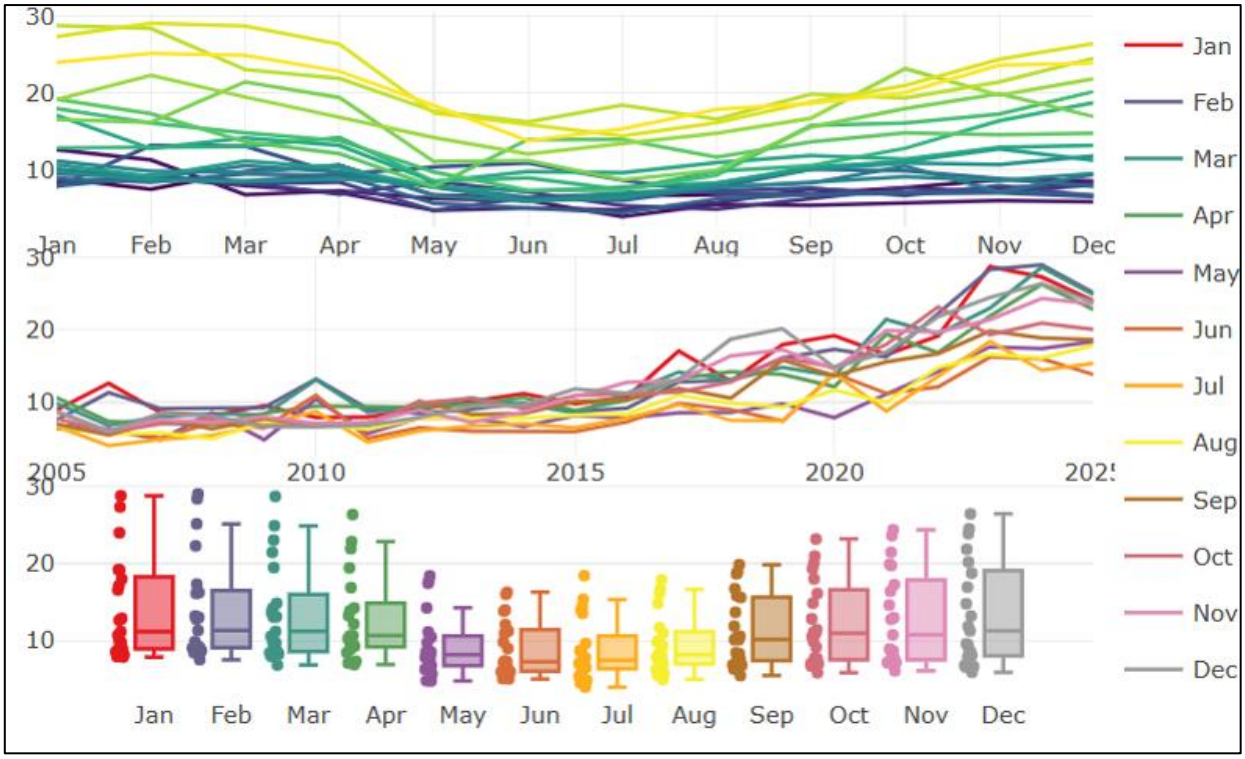


Figura 1. Precios promedio mensuales del melón cantaloupe al mayoreo en central de abasto.

Para constatar la inestabilidad en la varianza se aplicó la prueba de Cox-Box (1964), estimando un valor lambda λ =-0.26 el cual se ubica dentro del rango de −2 y 2 sugiriendo la transformación de la variable para tener una varianza constante. Lo anterior se realiza considerando una función de una familia de transformaciones de potencia que se utilizan para estabilizar la varianza y hacer que un conjunto de datos se aproxime más a una distribución normal, para este caso se utilizará la transformación logarítmica. Para aplicar la metodología de Box y Jenkins (1976) para series de tiempo, es necesario que la variable tenga una media estacionaria la cual se obtiene con al menos una diferenciación, con estos dos tratamientos estadísticos se mejora la aproximación a la distribución normal del proceso $\{\mathcal{Z}\}$.

La estacionariedad o no estacionariedad estadística planteada por Dickey y Fuller (1981), se verificó en la serie de tiempo mediante la prueba de raíz unitaria de Dickey Fuller Aumentada (ADF) para evaluar la existencia de una autocorrelación serial bajo las hipótesis H0: la serie no es estacionaria y tiene raíz unitaria vs Ha: la serie es estacionaria y no tiene raíz unitaria (Pankratz, 1983). Dado que el valor de probabilidad $(pvalue = 0.01)$ es menor al nivel de significancia $(\alpha = 0.05)$ se rechaza la hipótesis nula y se concluye que la serie no tiene raíz unitaria, siendo estacionaria, significando que la serie tiene media y varianza constante en el tiempo. Una vez cumplidas estas condiciones estadísticas de la serie de tiempo se realiza la modelación obteniendo los resultados que se muestran en el Cuadro1.

Cuadro 1. Identificación y selección del modelo ARIMA estacional.

Modelo	(1,1,1)*(1,0,0)	(1,1,1)*(0,0,1)	(1,1,1)*(1,0,1)	(1,1,1)*(0,1,0)
AIC	237.05	243.27	239.05	286.05
BIC	244.62	250.84	248.51	290.88
AR ϕ	0.713801	-0.8531	0.7139191	0.399826
Pr(>|z|)	2.172e-11 ***	3.725e-09 ***	2.422e-11 ***	0.04984 *

MA1θ	-1.0000	0.95148	-0.9999992	-0.860746
Pr(>|z|)	< 2.2e-16 ***	< 2.2e-16 ***	< 2.2e-16 ***	< 2e-16 ***
SAR1Φ_{12}	0.553981	-	0.5593844	-
Pr(>|z|)	5.385e-06 ***	-	0.1269	-
SMA1Θ_{12}	-	0.4811	-0.0078727	-
Pr(>|z|)	-	0.00804 **	0.9875	-

El Cuadro 1, muestra cuatro posibles modelos identificados para la serie de tiempo, sin embargo, el modelo seleccionado para ajustar los precios del melón fue un ARIMA$(1,1,1)\mathrm{X}(1,0,0)_{S=12}$, con comportamiento estacional repetido cada 12 meses $(S = 12)$. Lo anterior significa que se aplicó una primera diferencia $(d = 1)$, se estimaron el parámetro autorregresivo AR1ϕ=1, así como el componente de promedio móvil MA1θ=1, ambos para la parte no estacional; mientras que los parámetros autorregresivos SAR1Φ_{12}=1, y el componente de promedio móvil SMA1Θ_{12}=0, se calcularon para la parte estacional sin diferenciar $(D = 0)$. La parsimonia del modelo elegido se basó en los valores menores del Criterio de Información de Akaike (AIC) y el Criterio Bayesiano de Shwartz (BIC).

Derivada del modelo seleccionado, la ecuación estimada de pronósticos se expresa de la siguiente forma:

$$\Delta Y_t = 0.713801 Y_{t-1} + 0.553981 Y_{t-12} + 1.0 Y_{t-1} + \varepsilon_t$$

Posterior a la estimación de parámetros mediante máxima verosimilitud, se realiza la verificación de supuestos mediante el análisis estadístico de los residuales para finalmente realizar los pronósticos. Los residuales estandarizados para este modelo se analizaron mediante el valor de probabilidad $(pvalue)$ para las pruebas de ruido blanco, mostrando una secuencia independiente e idénticamente distribuida con una media cero y una varianza constante.

Para constatar lo anterior, la prueba de Ljung-Box (1978) estima un valor de probabilidad $(pvalue = 0.5362)$ mayor al nivel de significancia $(\alpha = 0.05)$ en cada uno de sus rezagos, no rechazando la hipótesis nula, señalando que los datos se distribuyen de forma independiente (no hay autocorrelación de los residuos).

La prueba de Brock, Dechert y Scheinkman (1996), se utiliza para contrastar la hipótesis nula de que la secuencia es aleatoria, independiente e idénticamente distribuida (iid). El estadístico mencionado para el presente modelo estima valores de probabilidad mayores al nivel de significancia $(pvalue = 1.1807, 2.3613, 3.542, 4.7226 > \alpha = 0.05)$, lo cual descarta la no linealidad en los residuos. De manera que se reafirma la posibilidad de generar pronósticos mediante el uso del modelo. En la prueba de Jarque-Bera (1980), la hipótesis nula señala que se observa la distribución normal sobre el ruido blanco de los datos y la alternativa señala lo opuesto. La implicación de este supuesto es garantizar que los intervalos para los pronósticos sean confiables. El estadístico de prueba aplicado al presente modelo indica que el valor de probabilidad $(pvalue = 0.7143)$ es mayor al nivel de significancia $(\alpha = 0.05)$, por lo que no rechazamos la hipótesis nula.

El ajuste del modelo con respecto a los datos observados se muestra en la Figura 2.

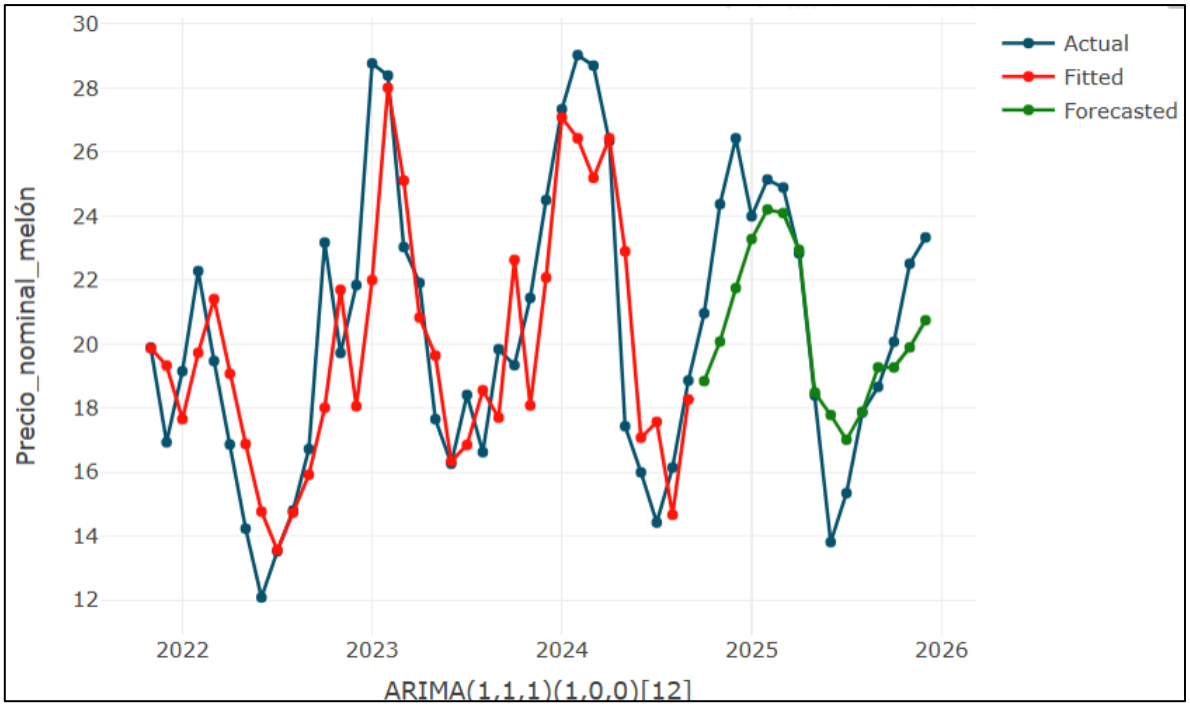


Figura 2. Entrenamiento y validación del modelo ARIMA estacional (1,1,1)X(1,0,0)_(S=12).

Además de las pruebas de bondad de ajuste, es conveniente realizar la evaluación del modelo con respecto a la capacidad de réplica, una vez que el modelo ha generado información con parámetros estimados. Lo anterior es posible dividiendo la serie de tiempo en conjuntos de entrenamiento y validación (70% y 30% de los datos observados, respectivamente), aplicando el modelo ARIMA estacional al conjunto de entrenamiento. Con esto, el modelo replica la serie en el mismo porcentaje (línea roja) y se realiza un pronóstico para el 30% de los datos restantes (línea verde). De esta manera, al graficar en conjunto los datos ajustados y pronosticados, la serie se compara con los datos reales (línea azul) permitiendo observar la capacidad de ajuste del modelo.

Además de la Figura 2, en el Cuadro 2, también se consideran las medidas de error de ajuste y pronóstico, como el error porcentual absoluto medio (MAPE, por sus siglas en ingles), el cual indica en cuántos puntos porcentuales se equivocan en promedio el ajuste y los pronósticos con respecto a los datos observados (Farrera, 2013). Para el conjunto de datos de entrenamiento el tamaño del error es de 10.61%, mientras que para la serie de validación el error es de 8.41%. Adicionalmente al MAPE, se muestran las medidas de precisión de pronóstico de los modelos ARIMA estacionales como el error porcentual de la raíz cuadrada del promedio de errores al cuadrado (RMSPE) y el error porcentual de la raíz cuadrada de la mediana de los errores al cuadrado (RMdSPE), así como las estadísticas U1 y U2 de Theil (Farrera, 2013). En todas las métricas de precisión de pronóstico valores cercanos a cero implican mejor capacidad predictiva, justificando así la significancia estadística al modelo ARIMA estacional (1,1,1)*(1,0,0)12, con 50 datos.

Cuadro 2. Medidas de la precisión de pronóstico

Modelo	MAPE (*entrenamiento*)	MAPE (*validación*)	RMSPE	RMDSPE	U1	U2
M2	10.61%	8.41%	0.1152042	0.03995368	0.05492997	0.1071779

De acuerdo con el modelo ARIMA estacional (1,1,1)X(1,0,0)_(S=12), los precios esperados a partir del mes de enero de 2026 se muestran en el Cuadro 3.

Cuadro 3. Pronóstico de precios promedio mensuales del melón de enero a diciembre 2026.

Año 2026	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
Pronósticos	22.03	22.70	22.59	21.46	19.02	16.49	17.34	18.74	19.19	19.97	21.33	21.78
Observados	*23.94*	*21.69*	*22.08*	*21.04*	*19.82*	*20.91*	*18.63*	*19.41*	-	-	-	-
Límite inferior 95%	17.21	16.74	16.11	14.72	12.13	9.52	10.33	11.71	12.14	12.91	14.26	14.71
Límite superior 95%	26.85	28.66	29.07	28.21	25.90	23.45	24.35	25.78	26.24	27.04	28.40	28.85

El Cuadro 3, muestra que los pronósticos tienen un comportamiento semejante a los que se muestran en la Figura 1, reflejando buena precisión para los siguientes 12 meses a partir de enero a diciembre de 2026, con los limites inferior y superior al 95% de confianza.

CONCLUSIÓN

El modelo para pronosticar los precios del melón es estadísticamente significativo, con buena capacidad de ajuste y confiable para los pronósticos. Cabe mencionar que los datos para el análisis fueron tomados en cuenta hasta diciembre de 2025, por lo que, mientras se realizó la presente investigación, para el año 2026 se registraron los precios nominales mensuales de $23.94, $21.69, $22.08, $21.04, $19.82, $20.91, $18.63 y $19.39 de enero al mes de agosto, respectivamente. Estos precios están dentro de los limites inferior y superior del 95% (bandas de confianza), lo cual permite constatar la afinidad del modelo para pronosticar el precio promedio mensual del melón cantaloupe # 9 por kilogramo.

LITERATURA CITADA

Box, G. E. P., and G. M. Jenkins. 1976: Time Series Analysis: Forecasting and Control. Ed. Holden-day. San Francisco, CA. 575 p.

Brock, W. A., W. D. Dechert, Scheinkman J, y B. LeBaron (1996). “A test for independence based on the correlation dimension”. Econometric Reviews 15(3):197- 235.

Cox, D. R., and G. E. Box. 1964. An analysis of transformations. J Royal Static. Soc. 26: 211-252.

Dickey, D. and W. Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica. 49: 1057-1072.

Farrera, G. A. 2013. Manual de pronósticos para la toma de decisiones. Editorial digital Tecnológico de Monterrey. Primera edición. 278 p.

Guerrero, G. V. M. 1991. Análisis estadístico de series de tiempo económicas. Ed. UAMIztapalapa. México. 309 p.

Jarque, C.M. and Bera, A.K. (1980) Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. Economics Letters, 6, 255-259.
https://doi.org/10.1016/0165-1765(80)90024-5

Ljung, G. M. and G. E. Box. 1978. On a measure of lack of fit in time series models. Econometrica. 65: 297-303.

Martínez-Damián, Miguel Ángel, & Brambila-Paz, José de Jesús. (2023). Modelación de predictores de precio nominal vs real aplicados a maíz, trigo y cebada en México. Revista mexicana de ciencias agrícolas, 14(2), 295-301. Epub 19 de junio de 2023.https://doi.org/10.29312/remexca.v14i2.2933

Pankratz, A. 1983. Forecasting with Univariate Box-Jenkins Models Concepts and Cases. Ed. John Wiley & Sons. USA. 566 p.

Sánchez, P. A. (2008). Cambios estructurales en series de tiempo: una revisión del estado del arte. Revista Ingenierías Universidad de Medellín, 7(12), 115-140.

SIAP. 2026. Servicio de Información Agroalimentaria y Pesquera. Datos anuales de producción y

valor de producción de melón. https://www.gob.mx/siap.

SNIIM, 2026. Sistema Nacional de Información e Integración de Mercados. Consulta de precios agrícolas de frutas y hortalizas en mercados nacionales. https://www.economia-sniim.gob.mx/nuevo/Home.aspx?opcion=Consultas/MercadosNacionales/PreciosDeMercado/Agricolas/ConsultaFrutasYHortalizas.aspx?SubOpcion=4|0