

Un Modelo de Series de Tiempo para la Evolución de la Matrícula del Nivel Medio Superior de la UAGro

A Time Series Model for the Evolution of High School's Enrolment of the UAGro

Jorge Luis Yopán-Fajardo

Universidad Autónoma de Guerrero

Víctor Manuel Alvarado-Castro

Universidad Autónoma de Guerrero

alvarado@cimat.mx

Resumen

Se utilizó el enfoque de Box-Jenkins, para ajustar un modelo Autorregresivo Integrado y de Medias Móviles (ARIMA), para el número de la matrícula de Nivel Medio Superior (NMS) de la Universidad Autónoma de Guerrero (UAGro), en el periodo 1960 a 2013. Se utilizó el modelo ajustado para predecir el número de la matrícula de NMS para los años 2014-2019 usando el enfoque iterativo 1-stead ahead, calculando intervalos del 95 % de confianza para las predicciones.

Palabras clave: Enfoque Box-Jenkins, ARIMA, número de la matrícula, función de autocorrelación.

Abstract

The Box-Jenkins approach was used to fit an Autoregressive Integrated Moving Average (ARIMA) model to the number of high school's enrolment (NMS) of the Autonomous University of Guerrero (UAGro), from 1960 to 2013. We use our fitted model to predict the enrolment of NMS for the years 2014 to 2019 using the iterative approach 1-stead ahead, 95% confidence intervals were computed.

Key words: Approach Box-Jenkins, ARIMA, enrolment, autocorrelation function.

Introducción

De acuerdo con la Ley General de la Educación de México, la educación media superior comprende el nivel de bachillerato, los demás niveles equivalentes a éste, así como la educación profesional que no requiere bachillerato o sus equivalentes (artículo 37)¹; esto es, el nivel educativo posterior a la educación básica y previo a la licenciatura, dirigido a los jóvenes de entre 15 y 18 años de edad. Es en este nivel educativo en el cual se plantea, en sus opciones terminales, la transición de la escuela a la vida productiva.

En el ciclo escolar 2011-2012 se atendieron, a nivel nacional, 4.3 millones de estudiantes en el sistema escolarizado. Así, 60.4 % (2.619 millones) de la matrícula cursó estudios de bachillerato general, 30.7 % (1.331 millones) bachillerato tecnológico y 8.9 % (383.5 miles) en escuelas de profesionales técnicos. La matrícula atendida en educación media superior representó el 12.4 % del total del Sistema Educativo Nacional^{2,3}.

La Universidad Autónoma de Guerrero (UAGro), en el ciclo escolar 2013-2014 contó con una matrícula de 44,925 estudiantes de NMS en el sistema escolarizado, distribuidos en 44 Unidades Académicas Preparatorias⁴. En 1960, cuando se fundó la UAGro, sólo atendió a 138 estudiantes de NMS, en la preparatoria No. 1 (actualmente Unidad Académica Preparatoria No. 1); este número fue triplicado para el siguiente ciclo escolar 1961-1962 y atendido en tan sólo dos Unidades Académicas. Desde entonces, tanto el número de la matrícula de NMS y el número de Unidades Académicas Preparatorias ha venido en ascenso en la UAGro⁵.

Los Modelos Autorregresivos Integrados y de Medias Móviles (ARIMA), desde que Box-Jenkins en 1970 desarrollaron los elementos teóricos y prácticos en que se apoyan^{6,7}, son ampliamente usados en el análisis de series de tiempo de fenómenos naturales (temperatura, humedad, velocidad del viento), en economía, en la proyección poblacional e incremento de

matrícula universitaria^{8,9,10,11}. En el caso del incremento poblacional Alonso, A., et al (2005), destaca la importancia de los modelos ARIMA como una herramienta útil para modelar la evolución de la mortalidad en España¹⁰. Fernández et al (2000), utilizan un modelo ARIMA para la proyección de la matrícula universitaria¹¹.

En este estudio, usamos un modelo ARIMA para ajustar y predecir el número de matrícula del NMS de la UAGro. En particular, utilizamos el enfoque de Box-Jenkins⁶ para ajustar los datos de la matrícula del NMS en el periodo 1960-2010. Para las predicciones se utilizó el enfoque iterativo; la predicción se comparó con el valor observado del ciclo escolar 2014-2015, no considerado en el ajuste, para medir la potencia predictiva del modelo y se obtuvieron predicciones con intervalos del 95 % de confianza para los años 2015-2019.

MATERIAL Y MÉTODOS

La UAGro es una institución pública y autónoma de educación media superior y superior, cuyos campos se encuentran distribuidos en diferentes ciudades del estado de Guerrero. La UAGro actualmente está conformada por 80 Unidades Académicas, de las cuales 37 ofrecen estudios de Nivel Superior y 43 ofrecen estudios de NMS⁴. La UAGro se funda en el año 1960, ofertando 5 Programas Educativos, 4 de Nivel Superior y tan sólo uno de NMS. Actualmente, en la UAGro existen 61 Programas Educativos ofertados por 37 Unidades Académicas de Nivel Superior y 43 Unidades Académicas Preparatorias⁴. En la actualidad, los Programas Educativos están agrupados por Dependencia de Educación Superior (DES), conformando 13 DES de acuerdo con las áreas de conocimiento que clasifica la Secretaría de Educación Pública (SEP). Actualmente, la Institución cuenta con 71 Programas Educativos del Nivel Superior acreditados como de buena calidad, lo que representa un total de 91 % de la matrícula de nuestra Universidad. En cuanto al Nivel Medio Superior, también se tienen indicadores de calidad, ya que del total de las Unidades Académicas 13 ya se encuentran en el Sistema Nacional de Bachillerato, lo que significa que aprobaron la evaluación que la SEP realiza a todas las instituciones que brindan este subsistema a nivel nacional¹².

Con la finalidad de realizar una planeación adecuada de la distribución de recursos para las Unidades Académicas que conforman la UAGro, es de suma importancia realizar proyecciones del incremento de la matrícula. Particularmente, debido al mayor número de demanda que la UAGro tiene en el NMS, es necesario estimar el número de la matrícula que se tendrá en generaciones futuras.

Usando los datos del número de matrícula del NMS de la UAGro de 1960 a 2013, ajustamos un modelo ARIMA para el incremento de la matrícula, y posteriormente se utilizó el modelo ajustado para predecir datos que no fueron considerados en la muestra (observación del año 2014). Se utilizó el enfoque de Box-Jenkins para proponer modelos ARIMA que expliquen el comportamiento de una serie de tiempo, el cual consiste básicamente en cuatro pasos^{6,8,13,14}: Primero, evaluamos la necesidad de una transformación de los datos para estabilizar la varianza usando el gráfico promedio-rango. Segundo, determinamos el orden de los parámetros de no-periodicidad (p, d, q) y periodicidad (P, D, Q)⁵² para un modelo ARIMA usando como herramientas: 1) El gráfico del número de matrícula del NMS, el cual proporciona información sobre la periodicidad y tendencia de la serie; 2) Las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), las cuales indican la presencia de una dependencia temporal en las series de tiempo estacionarias; 3) El criterio de información Akaike(AIC), el cual asiste en el buen ajuste del modelo; 4) La prueba de Ljung-Box, la cual mide la ACF de los residuales; y 5) la significancia de los parámetros, la cual debe ser estadísticamente diferente de cero. Tercero, estimamos los parámetros del modelo ARIMA usando máxima verosimilitud. Finalmente, se compararon gráficamente valores ajustados del modelo con los datos observados para comprobar si efectivamente se modela el número de matrícula del NMS.

Se utilizó el modelo ARIMA identificado para pronosticar el número de matrícula de NMS y sus intervalos del 95 % de confianza para datos anuales no considerados en la muestra, empleando el enfoque iterativo. Esto implica el uso del modelo ARIMA para predecir 1-step ahead, es decir, el valor predicho del año siguiente, año 3 (2014) del lustro once. Luego, a medida que se obtuvo el valor observado del año 4 del lustro once, se actualizaron los datos hasta este año, reestimado los parámetros del modelo ARIMA, y así se calculó el próximo 1-step ahead valor predicho, año 5 lustro once. Este procedimiento continuó hasta el año 4 del lustro doce (año 2019). Para medir la

potencia predictiva del modelo, el Root Mean Square Error (RMSE) fue calculado. El software estadístico R (ver 2.14.0) fue usado para todos los análisis estadísticos^{13,15,16,17}.

RESULTADOS

Se encontró necesario estabilizar la varianza del incremento de la matrícula del NMS de la UAGro, por lo que se calculó el logaritmo natural a los datos^{6,17}. La transformación de la matrícula del NMS muestra menos dispersión (Figura 1). Todos los procedimientos estadísticos, descriptivos y analíticos, se realizaron sobre el logaritmo natural de la matrícula del NMS.

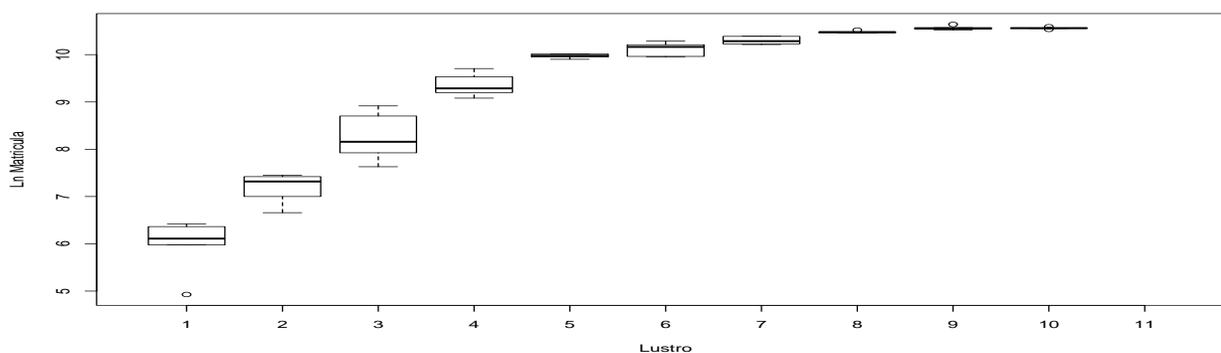


Figura 1. Logaritmo natural de la matrícula de NMS de la UAGro para el periodo 1960-2010.

El gráfico del logaritmo natural de la matrícula de NMS muestra una clara tendencia creciente (Figura 2). Los gráficos de las ACF y PACF muestrales describen una estructura de dependencia temporal y sugieren la inclusión de parámetros de no-estacionalidad y estacionalidad en el modelo (Figura 2). Un aumento significativo en el lag 1 en la ACF sugiere un componente no periódico AM(1), pero la falta de un incremento en el lag 5 en la ACF sugiere que no se considere un elemento periódico. Así, se consideraron como primeros modelos: el ARIMA (1,0,0) (1,0,0)₅ y ARIMA (0,1,0) (0,1,0)₅. También se analizaron los modelos: ARIMA (1,1,0)

(1,1,0)₅, ARIMA (0,1,1) (0,1,1)₅, ARIMA (1,1,1) (1,1,1)₅, ARIMA (1,2,1) (0,0,1)₅, ARIMA (0,2,1) (0,0,1)₅ y ARIMA (0,2,1) (0,1,1)₅. El mejor modelo fue el ARIMA (1,2,1) (0,0,1)₅, (AIC=-33.03, Table 1). Los gráficos de las ACF y PACF de los residuos no muestran correlación temporal y en su mayoría los spikes se encuentran dentro de los límites de significancia (Figure 3), por lo que los residuos pueden considerarse como ruido blanco.

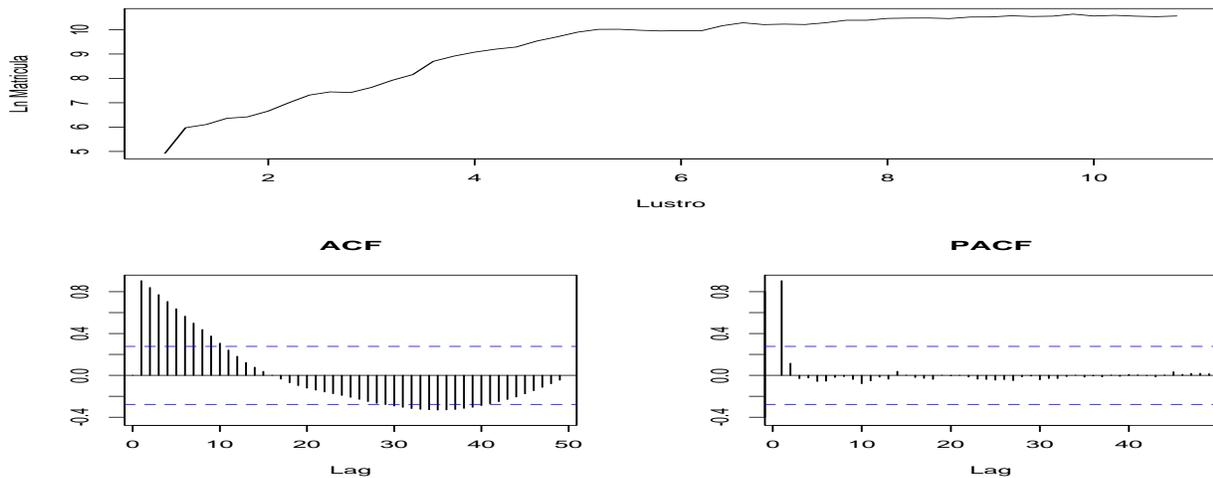


Figura 2. Panel superior: logaritmo natural de la matricula de NMS de la UAGro para el periodo 1960-2013. Paneles inferiores: función de autocorrelación muestral (ACF, izquierdo) y función de autocorrelación parcial (PACF, derecho); eje X indica el número de lags en lustros. Las líneas punteadas indican intervalos del 95 % de confianza.

Tabla 1. Coeficientes y error estándar de los parámetros* del modelo ARIMA (1,2,1) (0,0,1)₅.

Parámetros	Coeficiente	Error Estandar
AR(1) no-estacionalidad	0.0603	0.2440
AM(1) no-estacionalidad	-0.8178	0.1088
AM(1) estacionalidad	03656	0.1913

* Parámetros estimados por verosimilitud máxima.

Sigma² estimated as 0.0231, AIC = -38.02, AICc = -37.15

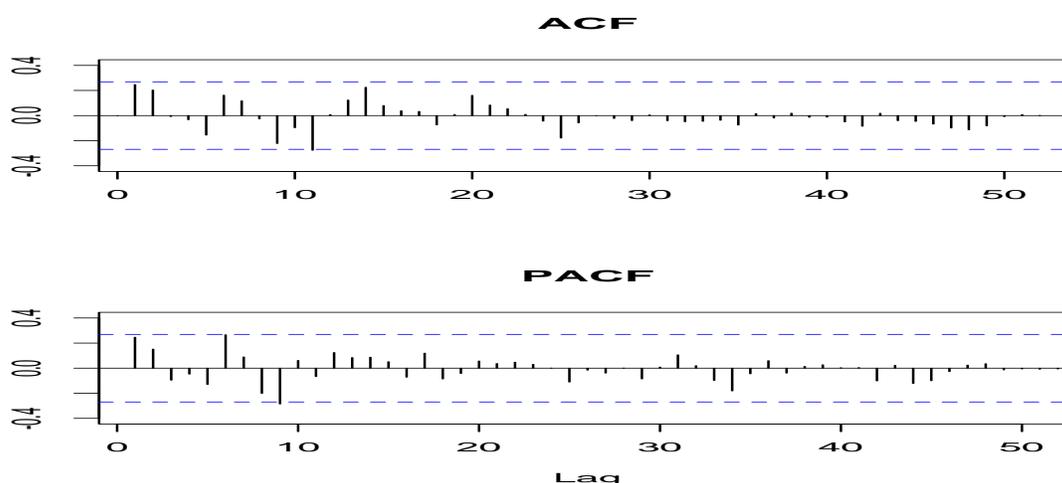


Figura 3. Función de autocorrelación (ACF, superior) y función de autocorrelación parcial (PACF, inferior) de los residuales del modelo ARIMA(0,2,1)(0,0,1)₅ ajustado al logaritmo natural de la matrícula del NMS de la UAGro periodo 1960-2013.

Se aplicó la prueba estadística Ljung-Box y se encontró que se rechaza la hipótesis nula de la independencia en los residuos de la serie de tiempo (P value = 0.003). El modelo ajustado (1960-2010) y los valores predichos (años 2011-2019) siguen el comportamiento del incremento de la matrícula de NMS (Figura 4).

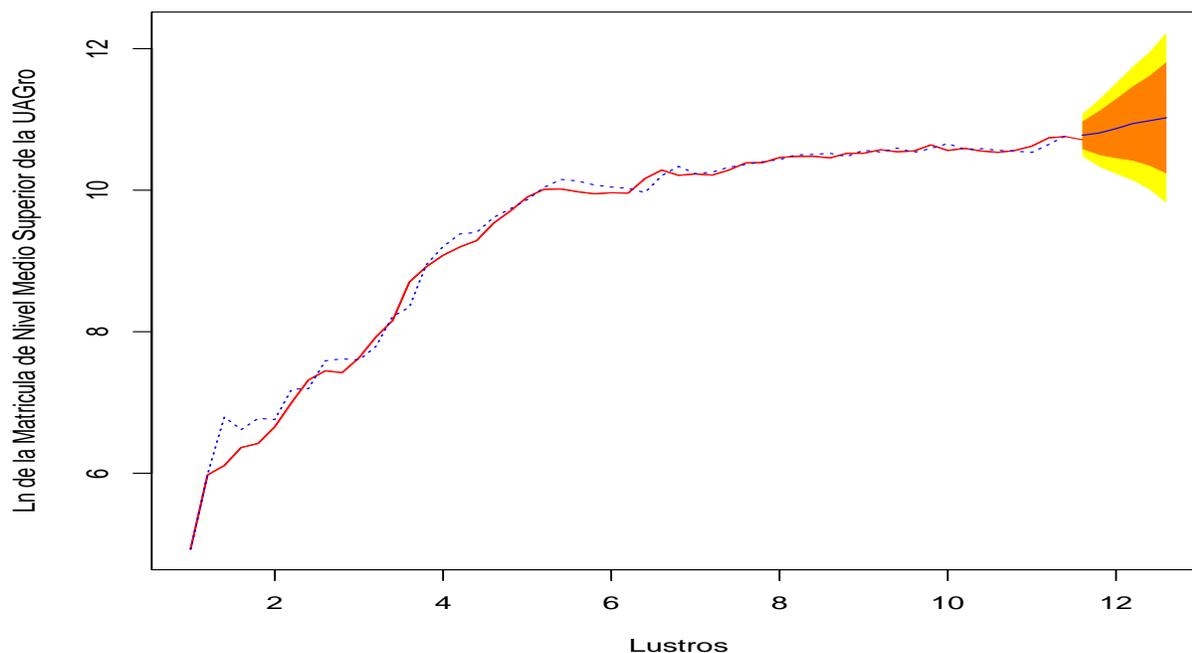


Figura 4. Línea sólida: logaritmo natural de la matrícula del NMS de la UAGro en el periodo 1960-2013. Línea punteada: Modelo Autorregresivo Integrado y de Medias Móviles (ARIMA) $ARIMA(0,2,1)(0,0,1)_5$ ajustado a los datos de la matrícula de NMS y valores proyectados 1-step (del 2014 al 2019) con intervalos del 95% de confianza.

Gráficamente, los valores predichos se estimaron utilizando el enfoque iterativo, se usó el modelo $ARIMA(1,2,1)(0,0,1)_5$ para predecir 1-step ahead, es decir, el valor predicho del año siguiente, año 3 (2014) del lustro once. Luego, a medida que se obtuvo el valor observado del año 4 del lustro once, se actualizaron los datos hasta este año, reestimado los parámetros del modelo $ARIMA(1,2,1)(0,0,1)_5$, y así se calculó el próximo 1-step ahead valor predicho, año 5 lustro once. Este procedimiento continuó hasta el año 4 del lustro doce (año 2019). El número de matrículas estimadas son presentadas en la tabla 2.

Tabla 2. Matrícula del NMS estimada utilizando el modelo ARIMA(0,2,1)(0,0,1)₅

Ciclo escolar	Matrícula del NMS	Intervalo del 95% de confianza
2014-2015	47889	(35559, 64495)
2015-2016	49424	(30742, 79459)
2016-2017	52591	(27743, 99694)
2017-2018	56507	(25303, 126292)
2018-2019	58803	(22285, 155163)
2019-2020	61292	(18412, 204038)

DISCUSIÓN Y CONCLUSIONES

El modelo de series de tiempo, según el enfoque de Box-Jekin, que mejor se ajustó a los datos es un ARIMA(0,2,1)(0,0,1)₅, el cual proyecta que el número de matrícula de NMS para los años 2014-2019 será: 47,889, 49,424, 52,591, 56,507, 58,803 y 61,292, respectivamente. La tendencia que nos arrojan los resultados, nos hace reflexionar que la UAGro en un futuro no muy lejano tendrá que programar nuevos espacios para albergar a la demandante cantidad de alumnos de NMS, ya que la infraestructura con la que cuenta en estos momentos quedará muy reducida. Para que se haga realidad esto, se tendrá que gestionar ante los tres niveles de gobierno para que puedan incrementar el presupuesto. Asimismo, la UAGro deberá elaborar proyectos y presentarlos a diversas estancias para gestionar y obtener recursos financieros extraordinarios que promuevan la construcción de nuevas Unidades Académicas, así como contratar personal académico y administrativo.

Las diferentes Unidades Académicas de Nivel Medio Superior tendrán que elaborar su propio diagnóstico (análisis FODA) para analizar el comportamiento tanto histórico como futuro de la demanda que tendrán y cuáles serán las consecuencias que tendrán que afrontar y cómo resolverlas.

Bibliografía

¹ Ley General de Educación, Art. 37 (2009). Diario Oficial de la Federación. México, D.F. última reforma publicada el 22 de junio de 2009.

² UNAM (2012). Plan educativo nacional. (Revisado en marzo de 2015, http://www.planeducativonacional.unam.mx/CAP_01/Text/01_05a.html)

³ SEP (2011). Principales Cifras Ciclo Escolar 2010-2011. Sistemas Educativos de Los Estados Unidos Mexicanos. México.

⁴ UAG (2014). Anuario Estadístico 2013-2014 de la Universidad Autónoma de Guerrero. Universidad Autónoma de Guerrero. México, D.F.

⁵ UAGro(2014). Evolución histórica en cifras de la UAGro 1960-2013. Universidad Autónoma de Guerrero. México, D.F.

⁶ Box, G.E.P., and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

⁷ De Gooijer, J.G. and Hyndman, R. J. (2006). 25 years of time series forecasting. International Journal of Forecasting, 22: 443-473.

⁸ Allard, R. (1998). Use of time-series analysis in infectious disease surveillance. Bull World Health Organ, 76: 327–333.

⁹ Helfenstein, U. (1986). Box-Jenkins modelling of some viral infectious diseases. Stat Med 5: 37–47.

¹⁰ Alonso, A., Peña, D. Y Rodríguez, J. (2005). Proyecciones de la población española. Serie de Estadística y Econometría. Madrid, España.

¹¹ Fernández, A., Firpo, C. Y Perera M. (2000). Proyección de la matrícula universitaria periodo: 2000-2030. Universidad de la República (UDELAR).

¹² UAGro (2013). Agenda de Información Estratégica 2001-2012. Universidad Autónoma de Guerrero. México, D.F.

¹³ Heiberger, R.M., and Holland, B. (2004). Statistical Analysis and Data Display: An Intermediate Course with Examples in S-plus, R, and SAS. New York: Springer.

¹⁴ Heiberger, R.M., and Teles, P. (2002). Displays for direct comparison of ARIMA models. Am Stat, 56: 131–138.

¹⁵ Hyndman, R.J. (2008). Forecast: Forecasting Functions for Time Series. R package version 4.8., (revised December 2014, <http://cran.r-project.org/web/packages/forecast/forecast.pdf>).

¹⁶ Hyndman, R.J., and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *J Statis Software*, 27(3).

¹⁷ Hyndman, R.J., and Athanasopoulos, G. (2013). *Forecasting: principles and practice*. (revised January 2015, <https://www.otexts.org/book/fpp>).